JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 8, NUMBER 10 OCTOBER 1967

Unitary Irreducible Representations of $SU(2, 2)$. I

TSU YAO
Department of Physics, Rutgers, The State University, New Brunswick, N.J.

(Received 30 March 1967)

This paper, which is the first of three, is concerned with the general properties of the noncompact group $SU(2, 2)$, the Lie algebra of which is isomorphic to the Dirac algebra. In the course of our study of the unitary representations, we first obtained all the finite dimensional irreducible representations. The associated Young diagrams are shown to have simple properties; the degenerate Young diagrams always denote degenerate representations. Through a theorem of Harish-Chandra, which relates the finite representations to the unitary representations in the discrete series, we are able to obtain explicitly all the unitary infinite-dimensional irreducible representations in this series, both degenerate and nondegenerate. The notion of multiplicity for nondegenerate representations is introduced and discussed in connection with a new operator F_3, which is required for a complete labeling of states.

I. INTRODUCTION

RECENTLY there has been much activity in the area of noncompact groups.[1] The motivation for such studies among physicists is the attempt to understand the hadron spectrum and its interactions. The two groups that have received the most attention are $SL(6, c)$ and $SU(6, 6)$,[2] since they can be regarded as "relativistic versions" of $SU(6)$, which has a certain amount of success. Therefore, the primary task in the study of these noncompact groups is the understanding and classification of unitary irreducible representations which are all infinite dimensional. This is a very ambitious and complicated program, and no complete answer is yet in sight.

In this paper, we plan to study the much simpler group $SU(2, 2)$, the noncompact version of $SU(4)$. This group appears in the decomposition of

$$SU(6, 6) \supset SU(3) \times SU(2, 2),$$

and its Lie algebra is isomorphic to the Dirac algebra of 15γ matrices. Furthermore, this group contains two important subgroups, $0(3, 1)$, the homogeneous Lorentz group, and $SU(1, 1)$, the noncompact version of $SU(2)$, which are the first two noncompact groups studied exhaustively by Bargmann[3] and by Gel'fand and Naimark.[4] Since $SU(2, 2)$ is also the covering group of $0(4, 2)$, the conformal group in six dimensions, it contains the two deSitter groups

[1] See for instance, *Proceedings of the Conference on Non-Compact Groups in Particle Physics*, Y. Chow, Ed. (Milwaukee, May 1966) (W. A. Benjamin, Inc., New York, 1966), *Proceedings of the Conference on High-Energy Physics and Elementary Particles* (Trieste, May, 1965); (International Atomic Energy Agency, Vienna, 1965); Y. Dothan, M. Gell-Mann and Y. Neeman, Phys. Letters **17**, 148 (1965).

[2] C. Fronsdal, ICTP, Trieste, preprint IC/66/51 (1966); W. Ruhl, Nuovo Cimento **44**, 572 (1966); A. Salam and J. Strathdee, ICTP, Trieste, preprint IC/66/5 (1966).

[3] V. Bargmann, Ann. Math. **48**, 568 (1947).

[4] I. M. Gel'fand and M. A. Naimark, Izv. Akad. Nauk SSSR, Ser. Matem. **11**, 411 (1947); M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon Press, Inc., New York, 1964).

0(4, 1) and 0(3, 2) as subgroups. Therefore, we see that the group $SU(2, 2)$ has a very rich structure, and a knowledge of its unitary irreducible representations is highly desirable.

Several years ago Murai made the first study of the representations of this group.[5] Unfortunately, he *omitted* one operator in the specification of states within an irreducible representation. Consequently, he was able to study only the so-called degenerate unitary irreducible representations; the nondegenerate representations were completely ignored. Furthermore, even for degenerate representations, Murai did not obtain *all* the unitary irreducible representations. Subsequent workers on this subject seem to realize some of Murai's inadequacies, but no clear statement on this deficiency was made. Esteve and Sona[6] applying Graev's general theorem[7] obtained the three fundamental series of unitary irreducible representations; the discrete series were, however, not studied. Furthermore, the "physical content" was not given, i.e., the decomposition into unitary irreducible representations of the maximal compact subgroup. Within the past year, several papers have appeared dealing with the representations of $SU(2, 2)$.[8] Unfortunately, because of the complications involved, these authors all *limited* themselves to special cases. There is some overlap between these new results and Murai's work, but to what extent is not very clear.

In the present paper, we start by studying the structure of the Lie algebra of $SU(2, 2)$, Sec. II, in which the generators of the group are defined, commutation relations are written down, and the maximal compact subgroup is introduced. Section III deals with the complete set of commuting operators; it is seen that besides the three Casimir operators of $SU(2, 2)$, we need a fourth operator, called F_3, and the five operators of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ to uniquely specify a state within an irreducible representation. The omission of F_3 leads directly to the consideration of degenerate representations only. In this connection, the important idea of multiplicity is introduced; this is the number of times an irreducible representation of the maximal compact subgroup appears in an irreducible representation of $SU(2, 2)$. For degenerate representations, all multiplicities are zero or one, while for nondegenerate representations, some of the multiplicities may be greater than one. In Sec. IV, we discuss the general expressions for the three Casimir operators and F_3, and several recursion relations among the "raising" and "lowering" functions are derived. Section V deals with the finite irreducible representations which are not unitary. However, they play an important role in the classification of unitary irreducible representations in the *discrete* series. Several interesting results in the classification of degenerate and nondegenerate finite representations are obtained and they are summarized as theorems in this section. Section VI deals with the unitary irreducible representations in the discrete series. Here a theorem due to Harish-Chandra is used, from which we know that for every finite irreducible representation, there exist two unitary irreducible representations in the discrete series with the same values for the Casimir operators. This correspondence holds true for both degenerate and nondegenerate cases. An explicit enumeration of this correspondence is given.

Due to the lengthiness of this paper, we have decided to present our results in three parts. This paper consists of Secs. I–VI described above. Paper II[9] will be devoted to a general study of the degenerate case, in which we present *all* the unitary irreducible representations belonging to the several discrete and continuous series. Paper III[10] will be concerned with the much more difficult study of the nondegenerate representations in which the operator F_3 plays a central role.

The reason we have presented the representations of the discrete series in this paper is our feeling that this series may be of particular physical interest. Harish-Chandra's theorem gives an intimate connection between this series and the finite nonunitary irreducible representations; this fact may be of special relevance when application to particle physics is attempted. Here, we do not engage in any speculation of that nature, but restrict ourselves to the algebraic aspect of the problem.

II. STRUCTURE OF THE LIE ALGEBRA

The group $SU(2, 2)$ is defined as the group of transformations on a four-dimensional complex space leaving invariant the indefinite quadratic form $|Z_1|^2 + |Z_2|^2 - |Z_3|^2 - |Z_4|^2$. The compact version is the group $SU(4)$ which leaves invariant the positive definite quadratic form $|Z_1|^2 + |Z_2|^2 + |Z_3|^2 + |Z_4|^2$. Let A_α^β be the canonical generators of $SU(4)$, $\alpha, \beta =$

[5] Y. Murai, Progr. Theoret. Phys. (Kyoto) 9, 147 (1953).
[6] A. Esteve and P. G. Sona, Nuovo Cimento 32, 473 (1964).
[7] M. I. Graev, Tr. Mosk. Math. Obs. 7, 335 (1958); M. I. Graev, Dokl. Akad. Nauk SSSR 98, 517 (1954).
[8] A. Kihlberg, V. F. Muller, and F. Halbwachs, Commun. Math. Phys. (Germany) 3, 194 (1966); R. Raczka and J. Fischer, ICTP (1966), preprints IC/66/16, IC/66/36; I. T. Todorov, ICTP, Trieste (1966) preprint IC/66/71.
[9] Tsu Yao, J. Math. Phys. (to be published).
[10] Tsu Yao, J. Math. Phys. (to be published).

1, 2, 3, 4, then the canonical commutation relations are

$$[A_\alpha^\beta, A_\gamma^\delta] = \delta_\gamma^\beta A_\alpha^\delta - \delta_\alpha^\delta A_\gamma^\beta, \quad \alpha, \beta, \gamma, \delta = 1, 2, 3, 4. \tag{2.1}$$

Because of the unimodular condition $\sum_{\alpha=1}^4 A_\alpha^\alpha = 0$, we have 15 independent generators.

A. Definition of the Generators of the Group

Since $SU(2, 2)$ has $SU(2) \times SU(2) \times U(1)$ as its maximal compact subgroup, it is convenient to define the generators of $SU(2, 2)$ slightly differently. We start with the two $SU(2)$ subgroups, and define

$$J_+ = A_1^2, \quad J_- = A_2^1, \quad J_3 = \tfrac{1}{2}(A_1^1 - A_2^2), \tag{2.2}$$

$$K_+ = A_3^4, \quad K_- = A_4^3, \quad K_3 = \tfrac{1}{2}(A_3^3 - A_4^4). \tag{2.3}$$

Next, we have four noncompact $SU(1, 1)$ subgroups,

$$P_+ = iA_1^3, \quad P_- = iA_3^1, \quad P_0 = \tfrac{1}{2}(A_1^1 - A_3^3), \tag{2.4}$$

$$Q_+ = iA_2^4, \quad Q_- = iA_4^2, \quad Q_0 = \tfrac{1}{2}(A_2^2 - A_4^4), \tag{2.5}$$

$$S_+ = iA_1^4, \quad S_- = iA_4^1, \quad S_0 = \tfrac{1}{2}(A_1^1 - A_4^4), \tag{2.6}$$

$$T_+ = iA_2^3, \quad T_- = iA_3^2, \quad T_0 = \tfrac{1}{2}(A_2^2 - A_3^3). \tag{2.7}$$

The noncompact character of these subgroups is obvious from the definition of their generators, e.g., the subgroup whose generators are P_+, P_-, and P_0 is the group of transformations which leaves invariant the quadratic form $|Z_1|^2 - |Z_3|^2$.

Equations (2.2)–(2.7) give us 18 generators. Since there are only 15 independent ones, we have the following three relations:

$$J_3 - K_3 = P_0 - Q_0, \tag{2.8}$$

$$J_3 + K_3 = S_0 - T_0, \tag{2.9}$$

$$P_0 + Q_0 = S_0 + T_0 \equiv R_0. \tag{2.10}$$

We have introduced the operator R_0 in Eq. (2.10), and together with $J_\pm, J_3, K_\pm, K_3, P_\pm, Q_\pm, S_\pm, T_\pm$, they are considered as the 15 generators of the group. However, for convenience we keep on using P_0, Q_0, S_0, and T_0.

B. Commutation Relations

From the canonical commutation relations (2.1), we can write immediately all the commutation relations which define the Lie algebra of the group:

$$[J_3, J_+] = J_+, \ [J_3, J_-] = -J_-, \ [J_+, J_-] = 2J_3; \tag{2.11}$$

$$[K_3, K_+] = K_+, \ [K_3, K_-] = -K_-, \ [K_+, K_-] = 2K_3; \tag{2.12}$$

$$[P_0, P_+] = P_+, \ [P_0, P_-] = -P_-, \ [P_+, P_-] = -2P_0; \tag{2.13}$$

$$[Q_0, Q_+] = Q_+, \ [Q_0, Q_-] = -Q_-, \ [Q_+, Q_-] = -2Q_0; \tag{2.14}$$

$$[S_0, S_+] = S_+, \ [S_0, S_-] = -S_-, \ [S_+, S_-] = -2S_0; \tag{2.15}$$

$$[T_0, T_+] = T_+, \ [T_0, T_-] = -T_-, \ [T_+, T_-] = -2T_0; \tag{2.16}$$

$$[J_i, K_j] = 0, \quad i, j = +, -, 0; \tag{2.17}$$

$$[J_+, P_+] = 0, \ [J_+, P_-] = -T_-, \ [J_+, P_0] = -\tfrac{1}{2}J_+,$$
$$[J_-, P_+] = T_+, \ [J_-, P_-] = 0, \ [J_-, P_0] = \tfrac{1}{2}J_-,$$
$$[J_3, P_+] = \tfrac{1}{2}P_+, \ [J_3, P_-] = -\tfrac{1}{2}P_-, \ [J_3, P_0] = 0; \tag{2.18}$$

$$[J_+, Q_+] = S_+, \ [J_+, Q_-] = 0, \ [J_+, Q_0] = \tfrac{1}{2}J_+,$$
$$[J_-, Q_+] = 0, \ [J_-, Q_-] = -S_-, \ [J_-, Q_0] = -\tfrac{1}{2}J_-,$$
$$[J_3, Q_+] = -\tfrac{1}{2}Q_+, \ [J_3, Q_-] = \tfrac{1}{2}Q_-, \ [J_3, Q_0] = 0; \tag{2.19}$$

$$[J_+, S_+] = 0, \ [J_+, S_-] = -Q_-, \ [J_+, S_0] = -\tfrac{1}{2}J_+,$$
$$[J_-, S_+] = Q_+, \ [J_-, S_-] = 0, \ [J_-, S_0] = \tfrac{1}{2}J_-,$$
$$[J_3, S_+] = \tfrac{1}{2}S_+, \ [J_3, S_-] = -\tfrac{1}{2}S_-, \ [J_3, S_0] = 0; \tag{2.20}$$

$$[J_+, T_+] = P_+, \ [J_+, T_-] = 0, \ [J_+, T_0] = \tfrac{1}{2}J_+,$$
$$[J_-, T_+] = 0, \ [J_-, T_-] = -P_-, \ [J_-, T_0] = -\tfrac{1}{2}J_-,$$
$$[J_3, T_+] = -\tfrac{1}{2}T_+, \ [J_3, T_-] = \tfrac{1}{2}T_-, \ [J_3, T_0] = 0; \tag{2.21}$$

$$[K_+, P_+] = -S_+, \ [K_+, P_-] = 0, \ [K_+, P_0] = \tfrac{1}{2}K_+,$$
$$[K_-, P_+] = 0, \ [K_-, P_-] = S_-, \ [K_-, P_0] = -\tfrac{1}{2}K_-,$$
$$[K_3, P_+] = -\tfrac{1}{2}P_+, \ [K_3, P_-] = \tfrac{1}{2}P_-, \ [K_3, P_0] = 0; \tag{2.22}$$

$$[K_+, Q_+] = 0, \ [K_+, Q_-] = T_-, \ [K_+, Q_0] = -\tfrac{1}{2}K_+,$$
$$[K_-, Q_+] = -T_+, \ [K_-, Q_-] = 0, \ [K_-, Q_0] = \tfrac{1}{2}K_-,$$
$$[K_3, Q_+] = \tfrac{1}{2}Q_+, \ [K_3, Q_-] = -\tfrac{1}{2}Q_-, \ [K_3, Q_0] = 0; \tag{2.23}$$

$$[K_+, S_+] = 0, \ [K_+, S_-] = P_-, \ [K_+, S_0] = -\tfrac{1}{2}K_+,$$
$$[K_-, S_+] = -P_+, \ [K_-, S_-] = 0, \ [K_-, S_0] = \tfrac{1}{2}K_-,$$
$$[K_3, S_+] = \tfrac{1}{2}S_+, \ [K_3, S_-] = -\tfrac{1}{2}S_-, \ [K_3, S_0] = 0; \tag{2.24}$$

$$[K_+, T_+] = -Q_+, \ [K_+, T_-] = 0, \ [K_+, T_0] = \tfrac{1}{2}K_+,$$
$$[K_-, T_+] = 0, \ [K_-, T_-] = Q_-, \ [K_-, T_0] = -\tfrac{1}{2}K_-,$$
$$[K_3, T_+] = -\tfrac{1}{2}T_+, \ [K_3, T_-] = \tfrac{1}{2}T_-, \ [K_3, T_0] = 0; \tag{2.25}$$

$$[P_i, Q_j] = 0 \quad i, j = +, -, 0; \tag{2.26}$$

$$[P_+, S_+] = 0,\ [P_+, S_-] = K_-,\ [P_+, S_0] = -\tfrac{1}{2}P_+,$$
$$[P_-, S_+] = -K_+,\ [P_-, S_-] = 0,\ [P_-, S_0] = \tfrac{1}{2}P_-,$$
$$[P_0, S_+] = \tfrac{1}{2}S_+,\ [P_0, S_-] = -\tfrac{1}{2}S_-,\ [P_0, S_0] = 0; \tag{2.27}$$

$$[P_+, T_+] = 0,\ [P_+, T_-] = -J_+,\ [P_+, T_0] = -\tfrac{1}{2}P_+,$$
$$[P_-, T_+] = J_-,\ [P_-, T_-] = 0,\ [P_-, T_0] = \tfrac{1}{2}P_-,$$
$$[P_0, T_+] = \tfrac{1}{2}T_+,\ [P_0, T_-] = -\tfrac{1}{2}T_-,\ [P_0, T_0] = 0; \tag{2.28}$$

$$[Q_+, S_+] = 0,\ [Q_+, S_-] = -J_-,\ [Q_+, S_0] = -\tfrac{1}{2}Q_+,$$
$$[Q_-, S_+] = J_+,\ [Q_-, S_-] = 0,\ [Q_-, S_0] = \tfrac{1}{2}Q_-,$$
$$[Q_0, S_+] = \tfrac{1}{2}S_+,\ [Q_0, S_-] = -\tfrac{1}{2}S_-,\ [Q_0, S_0] = 0; \tag{2.29}$$

$$[Q_+, T_+] = 0,\ [Q_+, T_-] = K_+,\ [Q_+, T_0] = -\tfrac{1}{2}Q_+,$$
$$[Q_-, T_+] = -K_-,\ [Q_-, T_-] = 0,\ [Q_-, T_0] = \tfrac{1}{2}Q_-,$$
$$[Q_0, T_+] = \tfrac{1}{2}T_+,\ [Q_0, T_-] = -\tfrac{1}{2}T_-,\ [Q_0, T_0] = 0; \tag{2.30}$$

$$[S_i, T_j] = 0 \quad i, j = +, -, 0. \tag{2.31}$$

C. Maximal Compact Subgroup

The maximal compact subgroup

$$SU(2) \times SU(2) \times U(1)$$

has $J_\pm$, J_3, $K_\pm$, K_3, and R_0 as its generators, and a unitary irreducible representation of this group is specified uniquely by the three members j, k, λ, where $\mathbf{J}^2 = j(j+1)$, $J_3 = -j, -j+1, \cdots, j$, $j = 0, \frac{1}{2}, 1, \cdots$

$$\mathbf{K}^2 = k(k+1),\ K_3 = -k, -k+1, \cdots, k,$$
$$k = 0, \tfrac{1}{2}, 1, \cdots \tag{2.32}$$
$$R_0 = \lambda.$$

From the definition of R_0 in Eq. (2.10), and using Eqs. (2.8) and (2.9), we see that for $j + k$ = integer, λ = integer, and for $j + k$ = half-integer, λ = half-integer.

We rewrite some of the commutation relations from Eqs. (2.11)–(2.31) as follows:

$$\begin{bmatrix} J_3 & \\ K_3, & P_\pm \\ R_0 & \end{bmatrix} = \pm \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix} P_\pm,$$

$$\begin{bmatrix} J_3 & \\ K_3, & Q_\pm \\ R_0 & \end{bmatrix} = \pm \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix} Q_\pm,$$

$$\begin{bmatrix} J_3 & \\ K_3, & S_\pm \\ R_0 & \end{bmatrix} = \pm \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix} S_\pm,$$

$$\begin{bmatrix} J_3 & \\ K_3, & T_\pm \\ R_0 & \end{bmatrix} = \pm \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix} T_\pm, \tag{2.33}$$

where

$$\pm \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix},$$

etc., are the roots.

Now let $|j, \mu; k, \nu; \lambda\rangle$ be a canonical basis which span a unitary irreducible representation of the compact subgroup $SU(2) \times SU(2) \times U(1)$. The basis vectors are so normalized that

$$J_+ |j, \mu; k, \nu; \lambda\rangle = [(j+\mu+1)(j-\mu)]^{\frac{1}{2}} |j, \mu+1; k, \nu; \lambda\rangle,$$
$$J_- |j, \mu; k, \nu; \lambda\rangle = [(j+\mu)(j-\mu+1)]^{\frac{1}{2}} |j, \mu-1; k, \nu; \lambda\rangle,$$
$$J_3 |j, \mu; k, \nu; \lambda\rangle = \mu |j, \mu; k, \nu; \lambda\rangle,$$
$$\mu = -j, -j+1, \cdots, j,$$
$$K_+ |j, \mu; k, \nu; \lambda\rangle = [(k+\nu+1)(k-\nu)]^{\frac{1}{2}} |j, \mu; k, \nu+1; \lambda\rangle,$$
$$K_- |j, \mu; k, \nu; \lambda\rangle = [(k+\nu)(k-\nu+1)]^{\frac{1}{2}} |j, \mu; k, \nu-1; \lambda\rangle,$$
$$K_3 |j, \mu; k, \nu; \lambda\rangle = \nu |j, \mu; k, \nu; \lambda\rangle,$$
$$\nu = -k, -k+1, \cdots, k,$$
$$R_0 |j, \mu; k, \nu; \lambda\rangle = \lambda |j, \mu; k, \nu; \lambda\rangle. \tag{2.34}$$

D. Determination of the Operators $P_\pm$, $Q_\pm$, $S_\pm$, and $T_\pm$

Let

$$|n\rangle \equiv P_+ |j, \mu; k, \nu; \lambda\rangle. \tag{2.35}$$

Then, from Eq. (2.33), we have

$$J_3 |n\rangle = J_3 P_+ |j, \mu; k, \nu; \lambda\rangle$$
$$= (\mu + \tfrac{1}{2}) P_+ |j, \mu; k, \nu; \lambda\rangle = (\mu + \tfrac{1}{2}) |n\rangle,$$
$$K_3 |n\rangle = (\nu - \tfrac{1}{2}) |n\rangle,$$
$$R_0 |n\rangle = (\lambda + 1) |n\rangle. \tag{2.36}$$

From Eqs. (2.18) and (2.21), we see that P_+ and T_+ transform as a J-spin doublet;

$$[J_3, P_+] = \tfrac{1}{2}P_+,\ [J_3, T_+] = -\tfrac{1}{2}T_+,$$
$$[J_+, P_+] = 0,\ [J_+, T_+] = P_+,$$
$$[J_-, P_+] = T_+,\ [J_-, T_+] = 0. \tag{2.37}$$

Similarly, we see that $-S_+$ and P_+ transform as a K-spin doublet;

$$[K_3, S_+] = \tfrac{1}{2}S_+, \quad [K_3, P_+] = -\tfrac{1}{2}P_+,$$
$$[K_+, S_+] = 0, \quad [K_+, P_+] = -S_+,$$
$$[K_-, S_+] = -P_+, \quad [K_-, P_+] = 0. \tag{2.38}$$

Therefore, under the compact subgroup $SU(2) \times SU(2) \times U(1)$, P_+ transforms like

$$P_+ \sim |j = \tfrac{1}{2}, \mu = \tfrac{1}{2}; k = \tfrac{1}{2}, \nu = -\tfrac{1}{2}; \lambda = 1\rangle. \tag{2.39}$$

The symbol $\sim$ here means transform like. Similarly,

$$Q_+ \sim -|j = \tfrac{1}{2}, \mu = -\tfrac{1}{2}; k = \tfrac{1}{2}, \nu = \tfrac{1}{2}; \lambda = 1\rangle,$$
$$S_+ \sim -|j = \tfrac{1}{2}, \mu = \tfrac{1}{2}; k = \tfrac{1}{2}, \nu = \tfrac{1}{2}; \lambda = 1\rangle,$$
$$T_+ \sim \ \ |j = \tfrac{1}{2}, \mu = -\tfrac{1}{2}; k = \tfrac{1}{2}, \nu = -\tfrac{1}{2}; \lambda = 1\rangle. \tag{2.40}$$

We can immediately write the following expressions:

$$\begin{aligned} P_+|j, \mu; k, \nu; \lambda\rangle = {} & [(j+\mu+1)(k-\nu+1)]^{\frac{1}{2}} a_1(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu+\tfrac{1}{2}; k+\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j+\mu+1)(k+\nu)]^{\frac{1}{2}} a_2(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu+\tfrac{1}{2}; k-\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j-\mu)(k-\nu+1)]^{\frac{1}{2}} a_3(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu+\tfrac{1}{2}; k+\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j-\mu)(k+\nu)]^{\frac{1}{2}} a_4(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu+\tfrac{1}{2}; k-\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle, \end{aligned} \tag{2.41a}$$

$$\begin{aligned} Q_+|j, \mu; k, \nu; \lambda\rangle = {} & -[(j-\mu+1)(k+\nu+1)]^{\frac{1}{2}} a_1(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu-\tfrac{1}{2}; k+\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j-\mu+1)(k-\nu)]^{\frac{1}{2}} a_2(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu-\tfrac{1}{2}; k-\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j+\mu)(k+\nu+1)]^{\frac{1}{2}} a_3(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu-\tfrac{1}{2}; k+\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle \\ & - [(j+\mu)(k-\nu)]^{\frac{1}{2}} a_4(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu-\tfrac{1}{2}; k-\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle, \end{aligned} \tag{2.41b}$$

$$\begin{aligned} S_+|j, \mu; k, \nu; \lambda\rangle = {} & -[(j+\mu+1)(k+\nu+1)]^{\frac{1}{2}} a_1(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu+\tfrac{1}{2}; k+\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j+\mu+1)(k-\nu)]^{\frac{1}{2}} a_2(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu+\tfrac{1}{2}; k-\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle \\ & - [(j-\mu)(k+\nu+1)]^{\frac{1}{2}} a_3(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu+\tfrac{1}{2}; k+\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j-\mu)(k-\nu)]^{\frac{1}{2}} a_4(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu+\tfrac{1}{2}; k-\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda+1\rangle, \end{aligned} \tag{2.41c}$$

$$\begin{aligned} T_+|j, \mu; k, \nu; \lambda\rangle = {} & [(j-\mu+1)(k-\nu+1)]^{\frac{1}{2}} a_1(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu-\tfrac{1}{2}; k+\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle \\ & + [(j-\mu+1)(k+\nu)]^{\frac{1}{2}} a_2(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu-\tfrac{1}{2}; k-\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle \\ & - [(j+\mu)(k-\nu+1)]^{\frac{1}{2}} a_3(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu-\tfrac{1}{2}; k+\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle \\ & - [(j+\mu)(k+\nu)]^{\frac{1}{2}} a_4(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu-\tfrac{1}{2}; k-\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda+1\rangle. \end{aligned} \tag{2.41d}$$

We have defined our phase convention in such a way that we have all positive signs in Eq. (2.41a). Equations (2.41b), (2.41c), and (2.41d) can be obtained from Eq. (2.41a) by using the commutation relations. The four functions $a_i(j, k, \lambda)$ are functions of j, k, and λ and depend on the irreducible representation of $SU(2, 2)$ with which we are dealing.

In an analogous manner, we can write the results when P_-, Q_-, S_-, and T_- are applied to the state $|j, \mu; k, \nu; \lambda\rangle$. From Eq. (2.33), we observe that Q_- behaves like P_+ under J-spin and K-spin transformations. Similarly, we also have

$$\begin{aligned} Q_- &\sim P_+, \\ P_- &\sim Q_+, \\ T_- &\sim S_+, \\ S_- &\sim T_+; \end{aligned} \tag{2.42}$$

$$\begin{aligned} Q_-|j, \mu; k, \nu; \lambda\rangle = {} & [(j+\mu+1)(k-\nu+1)]^{\frac{1}{2}} b_1(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu+\tfrac{1}{2}; k+\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda-1\rangle \\ & + [(j+\mu+1)(k+\nu)]^{\frac{1}{2}} b_2(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu+\tfrac{1}{2}; k-\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda-1\rangle \\ & + [(j-\mu)(k-\nu+1)]^{\frac{1}{2}} b_3(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu+\tfrac{1}{2}; k+\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda-1\rangle \\ & + [(j-\mu)(k+\nu)]^{\frac{1}{2}} b_4(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu+\tfrac{1}{2}; k-\tfrac{1}{2}, \nu-\tfrac{1}{2}; \lambda-1\rangle, \end{aligned} \tag{2.43a}$$

$$\begin{aligned} P_-|j, \mu; k, \nu; \lambda\rangle = {} & -[(j-\mu+1)(k+\nu+1)]^{\frac{1}{2}} b_1(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu-\tfrac{1}{2}; k+\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda-1\rangle \\ & + [(j-\mu+1)(k-\nu)]^{\frac{1}{2}} b_2(j,k,\lambda)\,|j+\tfrac{1}{2}, \mu-\tfrac{1}{2}; k-\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda-1\rangle \\ & + [(j+\mu)(k+\nu+1)]^{\frac{1}{2}} b_3(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu-\tfrac{1}{2}; k+\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda-1\rangle \\ & - [(j+\mu)(k-\nu)]^{\frac{1}{2}} b_4(j,k,\lambda)\,|j-\tfrac{1}{2}, \mu-\tfrac{1}{2}; k-\tfrac{1}{2}, \nu+\tfrac{1}{2}; \lambda-1\rangle, \end{aligned} \tag{2.43b}$$

$$T_-|j,\mu;k,\nu;\lambda\rangle = [(j+\mu+1)(k+\nu+1)]^{\frac{1}{2}}b_1(j,k,\lambda)\,|j+\tfrac{1}{2},\mu+\tfrac{1}{2};k+\tfrac{1}{2},\nu+\tfrac{1}{2};\lambda-1\rangle$$
$$-[(j+\mu+1)(k-\nu)]^{\frac{1}{2}}b_2(j,k,\lambda)\,|j+\tfrac{1}{2},\mu+\tfrac{1}{2};k-\tfrac{1}{2},\nu+\tfrac{1}{2};\lambda-1\rangle$$
$$+[(j-\mu)(k+\nu+1)]^{\frac{1}{2}}b_3(j,k,\lambda)\,|j-\tfrac{1}{2},\mu+\tfrac{1}{2};k+\tfrac{1}{2},\nu+\tfrac{1}{2};\lambda-1\rangle$$
$$-[(j-\mu)(k-\nu)]^{\frac{1}{2}}b_4(j,k,\lambda)\,|j-\tfrac{1}{2},\mu+\tfrac{1}{2};k-\tfrac{1}{2},\nu+\tfrac{1}{2};\lambda-1\rangle, \quad (2.43c)$$

$$S_-|j,\mu;k,\nu;\lambda\rangle = -[(j-\mu+1)(k-\nu+1)]^{\frac{1}{2}}b_1(j,k,\lambda)\,|j+\tfrac{1}{2},\mu-\tfrac{1}{2};k+\tfrac{1}{2},\nu-\tfrac{1}{2};\lambda-1\rangle$$
$$-[(j-\mu+1)(k+\nu)]^{\frac{1}{2}}b_2(j,k,\lambda)\,|j+\tfrac{1}{2},\mu-\tfrac{1}{2};k-\tfrac{1}{2},\nu-\tfrac{1}{2};\lambda-1\rangle$$
$$+[(j+\mu)(k-\nu+1)]^{\frac{1}{2}}b_3(j,k,\lambda)\,|j-\tfrac{1}{2},\mu-\tfrac{1}{2};k+\tfrac{1}{2},\nu-\tfrac{1}{2};\lambda-1\rangle$$
$$+[(j+\mu)(k+\nu)]^{\frac{1}{2}}b_4(j,k,\lambda)\,|j-\tfrac{1}{2},\mu-\tfrac{1}{2};k-\tfrac{1}{2},\nu-\tfrac{1}{2};\lambda-1\rangle. \quad (2.43d)$$

From the arbitrary nature of the basis

$$|j,\mu;k,\nu;\lambda\rangle,$$

we may define a new basis,

$$|j,\mu;k,\nu;\lambda\rangle' = \omega(j,k,\lambda)\,|j,\mu;k,\nu;\lambda\rangle, \quad (2.44)$$

where $\omega(j,k,\lambda)$ is an arbitrary function of j, k, and λ. Then, we have

$$\omega(j,k,\lambda)a_1(j,k,\lambda) = \omega(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)a_1'(j,k,\lambda),$$
$$\omega(j,k,\lambda)a_2(j,k,\lambda) = \omega(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)a_2'(j,k,\lambda),$$
$$\omega(j,k,\lambda)a_3(j,k,\lambda) = \omega(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)a_3'(j,k,\lambda),$$
$$\omega(j,k,\lambda)a_4(j,k,\lambda) = \omega(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)a_4'(j,k,\lambda), \quad (2.45)$$

and

$$\omega(j,k,\lambda)b_1(j,k,\lambda) = \omega(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)b_1'(j,k,\lambda),$$
$$\omega(j,k,\lambda)b_2(j,k,\lambda) = \omega(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)b_2'(j,k,\lambda),$$
$$\omega(j,k,\lambda)b_3(j,k,\lambda) = \omega(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)b_3'(j,k,\lambda),$$
$$\omega(j,k,\lambda)b_4(j,k,\lambda) = \omega(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)b_4'(j,k,\lambda). \quad (2.46)$$

Therefore, we have

$$a_1(j,k,\lambda)b_4(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1) = a_1'(j,k,\lambda)b_4'(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1),$$
$$a_2(j,k,\lambda)b_3(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1) = a_2'(j,k,\lambda)b_3'(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1),$$
$$a_3(j,k,\lambda)b_2(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1) = a_3'(j,k,\lambda)b_2'(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1),$$
$$a_4(j,k,\lambda)b_1(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1) = a_4'(j,k,\lambda)b_1'(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1), \quad (2.47)$$

and

$$\frac{a_1(j,k,\lambda)}{b_4(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)} = \left[\frac{\omega(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)}{\omega(j,k,\lambda)}\right]^2 \frac{a_1'(j,k,\lambda)}{b_4'(+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)},$$

$$\frac{a_2(j,k,\lambda)}{b_3(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)} = \left[\frac{\omega(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)}{\omega(j,k,\lambda)}\right]^2 \frac{a_2'(j,k,\lambda)}{b_3'(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)},$$

$$\frac{a_3(j,k,\lambda)}{b_2(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)} = \left[\frac{\omega(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)}{\omega(j,k,\lambda)}\right]^2 \frac{a_3(j,k,\lambda)}{b_2'(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)},$$

$$\frac{a_4(j,k,\lambda)}{b_1(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)} = \left[\frac{\omega(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)}{\omega(j,k,\lambda)}\right]^2 \frac{a_4'(j,k,\lambda)}{b_1'(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)}. \quad (2.48)$$

Consequently, from the very beginning, we have the freedom to choose

$$a_1(j,k,\lambda) = b_4(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1),$$
$$a_2(j,k,\lambda) = b_3(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1),$$
$$a_3(j,k,\lambda) = b_2(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1),$$
$$a_4(j,k,\lambda) = b_1(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1) \quad (2.49)$$

by defining the ω's appropriately.[4]

So far, the development of the subject is very simple and straightforward, and we have essentially recapitulated what Murai had done in the first part of his paper. However, Eqs. (2.41) and (2.43) are not correct, since an implicit oversimplification has been assumed. In the next section dealing with the problem of state labeling, we show why another operator besides $\mathbf{J}^2$, J_3, $\mathbf{K}^2$, K_3, and R_0 is needed to uniquely specify a state within an irreducible representation of $SU(2,2)$, and what the correct version of Eqs. (2.41) and (2.43) should be.

III. COMPLETE SET OF COMMUTING OPERATORS

A. Casimir Operators

We start out by considering the three Casimir operators of the group. There are three Casimir

operators, since $SU(2, 2)$ is a group of rank 3. From expressions like

$$\sum_{\alpha,\beta} A_\alpha^\beta A_\beta^\alpha, \quad \sum_{\alpha,\beta,\gamma} A_\alpha^\beta A_\beta^\gamma A_\gamma^\alpha,$$

and

$$\sum_{\substack{\alpha,\beta \\ \gamma,\delta}} A_\alpha^\beta A_\beta^\gamma A_\gamma^\delta A_\delta^\alpha,$$

we can construct the three Casimir operators C_2, C_3, and C_4, which commute with all 15 generators of the group. We write

$$C_2 \equiv 2(\mathbf{J}^2 + \mathbf{K}^2) + R_0(R_0 + 4) - 2\{P_-P_+ + Q_-Q_+ + S_-S_+ + T_-T_+\} \tag{3.1a}$$

$$= 2(\mathbf{J}^2 + \mathbf{K}^2) + R_0(R_0 - 4) - 2\{P_+P_- + Q_+Q_- + S_+S_- + T_+T_-\}, \tag{3.1b}$$

$$\begin{aligned} C_3 \equiv & -(R_0 + 2)(\mathbf{J}^2 - \mathbf{K}^2) \\ & + \{J_+(P_-T_+ + S_-Q_+) + J_-(T_-P_+ + Q_-S_+) + J_3(P_-P_+ - Q_-Q_+ + S_-S_+ - T_-T_+)\} \\ & + \{K_+(S_-P_+ + Q_-T_+) + K_-(P_-S_+ + T_-Q_+) + K_3(P_-P_+ - Q_-Q_+ - S_-S_+ + T_-T_+)\} \end{aligned} \tag{3.2a}$$

$$\begin{aligned} = & -(R_0 - 2)(\mathbf{J}^2 - \mathbf{K}^2) \\ & + \{J_+(T_+P_- + Q_+S_-) + J_-(P_+T_- + S_+Q_-) + J_3(P_+P_- - Q_+Q_- + S_+S_- - T_+T_-)\} \\ & + \{K_+(P_+S_- + T_+Q_-) + K_-(S_+P_- + Q_+T_-) + K_3(P_+P_- - Q_+Q_- - S_+S_- + T_+T_-)\}, \end{aligned} \tag{3.2b}$$

$$\begin{aligned} C_4 \equiv & \tfrac{1}{2}R_0(R_0 + 2)C_2 - \tfrac{1}{4}R_0(R_0 - 2)(R_0 + 2)(R_0 + 4) + 4\mathbf{J}^2\mathbf{K}^2 \\ & + 2(R_0 + 1)\{-(R_0 + 2)(\mathbf{J}^2 + \mathbf{K}^2) \\ & \qquad + [J_+(P_-T_+ + S_-Q_+) + J_-(T_-P_+ + Q_-S_+) + J_3(P_-P_+ - Q_-Q_+ + S_-S_+ - T_-T_+)] \\ & \qquad - [K_+(S_-P_+ + Q_-T_+) + K_-(P_-S_+ + T_-Q_+) + K_3(P_-P_+ - Q_-Q_+ - S_-S_+ + T_-T_+)]\} \\ & + 4(P_-Q_- - S_-T_-)(P_+Q_+ - S_+T_+) \\ & + 4\{[J_+K_+S_-T_+ + J_+K_-P_-Q_+ + J_-K_+Q_-P_+ + J_-K_-T_-S_+] \\ & \quad + J_3[K_+(S_-P_+ - Q_-T_+) + K_-(P_-S_+ - T_-Q_+)] + K_3[J_+(P_-T_+ - S_-Q_+) + J_-(T_-P_+ - Q_-S_+)] \\ & \quad + J_3K_3[P_-P_+ + Q_-Q_+ - S_-S_+ - T_-T_+]\} \end{aligned} \tag{3.3a}$$

$$\begin{aligned} = & \tfrac{1}{2}R_0(R_0 - 2)C_2 - \tfrac{1}{4}R_0(R_0 - 2)(R_0 + 2)(R_0 - 4) + 4\mathbf{J}^2\mathbf{K}^2 \\ & + 2(R_0 - 1)\{-(R_0 - 2)(\mathbf{J}^2 + \mathbf{K}^2) \\ & \qquad + [J_+(T_+P_- + Q_+S_-) + J_-(P_+T_- + S_+Q_-) + J_3(P_+P_- - Q_+Q_- + S_+S_- - T_+T_-)] \\ & \qquad - [K_+(P_+S_- + T_+Q_-) + K_-(S_+P_- + Q_+T_-) + K_3(P_+P_- - Q_+Q_- - S_+S_- + T_+T_-)]\} \\ & + 4(P_+Q_+ - S_+T_+)(P_-Q_- - S_-T_-) \\ & + 4\{[J_+K_+T_+S_- + J_+K_-Q_+P_- + J_-K_+P_+Q_- + J_-K_-S_+T_-] \\ & + J_3[K_+(P_+S_- - T_+Q_-) + K_-(S_+P_- - Q_+T_-)] + K_3[J_+(T_+P_- - Q_+S_-) + J_-(P_+T_- - S_+Q_-)] \\ & + J_3K_3[P_+P_- + Q_+Q_- - S_+S_- - T_+T_-]\}. \end{aligned} \tag{3.3b}$$

In the following, we use C_2, C_3, C_4 to denote both the Casimir operators and their eigenvalues.

B. Problem of Labeling States

A given irreducible representation of the group is uniquely specified, if the eigenvalues C_2, C_3, and C_4 are given.[11] The question now is, are the states within a given irreducible representation uniquely specified when j, μ, k, ν, and λ are given? The answer to this question is negative, and the reason is very simple. Let us consider the descending chain of subgroups

$$SU(2, 2) \supset SU(2, 1) \supset SU(2).$$

We need C_2, C_3, and C_4 to specify an irreducible representation of $SU(2, 2)$, C_2', C_3', [the Casimir operators of $SU(2, 1)$] to specify an irreducible representation of $SU(2, 1)$, and $C_2'' = \mathbf{J}^2$ to specify $SU(2)$. In addition, we need the "magnetic" quantum members μ, ν, and λ. We therefore see that in place of the two operators C_2' and C_3', we have only $\mathbf{K}^2$ when

[11] This statement is correct for compact groups only. For non-compact groups this is not true. In fact, as we shall see, for a given set of eigenvalues C_2, C_3, and C_4, there may exist several inequivalent irreducible representations.

we consider the chain

$$SU(2, 2) \supset SU(2) \times SU(2) \times U(1).$$

Hence, our job is to construct an operator, called F_3 (presumably it is a cubic operator to replace C_3') which commutes with $\mathbf{J}^2$, J_3, $\mathbf{K}^2$, K_3, and R_0. This is an easy task, and we proceed systematically.

1. We observe that there does not exist an independent quadratic operator apart from C_2 which commutes with $\mathbf{J}^2$, J_3, $\mathbf{K}^2$, K_3, and R_0.

2. We observe that any cubic operator that commutes with $\mathbf{J}^2$, J_3, $\mathbf{K}^2$, K_3, and R_0 also commutes with J_+, J_-, K_+, and K_-, and there are two such operators,

$$G_3 = J_+(P_-T_+ + S_-Q_+) + J_-(T_-P_+ + Q_-S_+) + J_3(P_-P_+ - Q_-Q_+ + S_-S_+ - T_-T_+), \quad (3.4a)$$

$$H_3 = K_+(S_-P_+ + Q_-T_+) + K_-(P_-S_+ + T_-Q_+) + K_3(P_-P_+ - Q_-Q_+ - S_-S_+ + T_-T_+); \quad (3.4b)$$

$$[G_3, \mathbf{J}] = [G_3, \mathbf{K}] = [G_3, R_0] = 0, \quad (3.5a)$$

$$[H_3, \mathbf{J}] = [H_3, \mathbf{K}] = [H_3, R_0] = 0. \quad (3.5b)$$

Since $G_3 + H_3$ appears in C_3, there is actually only one independent cubic operator. For symmetry reasons, we define F_3 as

$$\begin{aligned} F_3 \equiv & -(R_0 + 2)(\mathbf{J}^2 + \mathbf{K}^2) \\ & + \{J_+(P_-T_+ + S_-Q_+) + J_-(T_-P_+ + Q_-S_+) + J_3(P_-P_+ - Q_-Q_+ + S_-S_+ - T_-T_+)\} \\ & - \{K_+(S_-P_+ + Q_-T_+) + K_-(P_-S_+ + T_-Q_+) + K_3(P_-P_+ - Q_-Q_+ - S_-S_+ + T_-T_+)\} \quad (3.6a) \\ = & -(R_0 - 2)(\mathbf{J}^2 + \mathbf{K}^2) \\ & + \{J_+(T_+P_- + Q_+S_-) + J_-(P_+T_- + S_+Q_-) + J_3(P_+P_- - Q_+Q_- + S_+S_- - T_+T_-)\} \\ & - \{K_+(P_+S_- + T_+Q_-) + K_-(S_+P_- + Q_+T_-) + K_3(P_+P_- - Q_+Q_- - S_+S_- + T_+T_-)\}. \quad (3.6b) \end{aligned}$$

3. We observe that any quartic operator that commutes with $\mathbf{J}^2$, J_3, $\mathbf{K}^2$, K_3, and R_0 also commutes with J_+, J_-, K_+, K_-, and there is only one independent operator, apart from R_0F_3, which can be chosen to be $(P_-Q_- - S_-T_-)(P_+Q_+ - S_+T_+)$. Indeed,

$$\begin{aligned} [P_-Q_- - S_-T_-, \mathbf{J}] &= [P_+Q_+ - S_+T_+, \mathbf{J}] = 0, \\ [P_-Q_- - S_-T_-, \mathbf{K}] &= [P_+Q_+ - S_+T_+, \mathbf{K}] = 0, \end{aligned} \quad (3.7)$$

but

$$\begin{aligned} [R_0, P_-Q_- - S_-T_-] &= -2(P_-Q_- - S_-T_-), \\ [R_0, P_+Q_+ - S_+T_+] &= 2(P_+Q_+ - S_+T_+), \end{aligned}$$

and

$$[R_0, (P_-Q_- - S_-T_-)(P_+Q_+ - S_+T_+)] = 0. \quad (3.8)$$

Now we have a complete set of commuting operators C_2, C_3, C_4; F_3, $\mathbf{J}^2$, $\mathbf{K}^2$, J_3, K_3, R_0, and every state is uniquely specified when the nine eigenvalues are given. Therefore, in Eq. (2.34), we add one additional label α,

$$\begin{aligned} F_3 \,|j, \mu; k, \nu; \lambda; \alpha\rangle &= \alpha \,|j, \mu; k, \nu; \lambda; \alpha\rangle, \\ \mathbf{J}^2 \,|j, \mu; k, \nu; \lambda; \alpha\rangle &= j(j+1) \,|j, \mu; k, \nu; \lambda; \alpha\rangle, \\ J_3 \,|j, \mu; k, \nu; \lambda; \alpha\rangle &= \mu \,|j, \mu; k, \nu; \lambda; \alpha\rangle, \\ \mathbf{K}^2 \,|j, \mu; k, \nu; \lambda; \alpha\rangle &= k(k+1) \,|j, \mu; k, \nu; \lambda; \alpha\rangle, \\ K_3 \,|j, \mu; k, \nu; \lambda; \alpha\rangle &= \nu \,|j, \mu; k, \nu; \lambda; \alpha\rangle, \\ R_0 \,|j, \mu; k, \nu; \lambda; \alpha\rangle &= \lambda \,|j, \mu; k, \nu; \lambda; \alpha\rangle. \end{aligned} \quad (3.9)$$

Now Eqs. (2.41) and (2.43) have to be modified. The functions $a_i(j, k, \lambda)$ are now matrices with elements $a_i(j, k, \lambda)_{\alpha\beta}$. In general, $a_i(j, k, \lambda)$ are *not* square matrices. Equation (2.41a) now looks like

$$\begin{aligned} & P_+ \,|j, \mu; k, \nu; \lambda; \alpha\rangle \\ & = [(j + \mu + 1)(k - \nu + 1)]^{\frac{1}{2}} \sum_\rho a_1(j, k, \lambda)_{\alpha\rho} \\ & \quad \times |j + \tfrac{1}{2}, \mu + \tfrac{1}{2}; k + \tfrac{1}{2}, \nu - \tfrac{1}{2}; \lambda + 1; \rho\rangle \\ & \quad + [(j + \mu + 1)(k + \nu)]^{\frac{1}{2}} \sum_\rho a_2(j, k, \lambda)_{\alpha\rho} \\ & \quad \times |j + \tfrac{1}{2}, \mu + \tfrac{1}{2}; k - \tfrac{1}{2}, \nu - \tfrac{1}{2}; \lambda + 1; \rho\rangle \\ & \quad + [(j - \mu)(k - \nu + 1)]^{\frac{1}{2}} \sum_\rho a_3(j, k, \lambda)_{\alpha\rho} \\ & \quad \times |j - \tfrac{1}{2}, \mu + \tfrac{1}{2}; k + \tfrac{1}{2}, \nu - \tfrac{1}{2}; \lambda + 1; \rho\rangle \\ & \quad + [(j - \mu)(k + \nu)]^{\frac{1}{2}} \sum_\rho a_4(j, k, \lambda)_{\alpha\rho} \\ & \quad \times |j - \tfrac{1}{2}, \mu + \tfrac{1}{2}; k - \tfrac{1}{2}, \nu - \tfrac{1}{2}; \lambda + 1; \rho\rangle. \end{aligned} \quad (3.10)$$

Equations (2.41b)–(2.41d) and (2.43a)–(2.43d) can be written in the same way. In Eq. (3.10), the index ρ is summed over. For example, $a_1(j, k, \lambda)$ may be a $n \times m$ matrix, where we have $\alpha_1, \alpha_2, \cdots \alpha_n$, and $\rho_1, \rho_2, \cdots, \rho_m$; n is called the *multiplicity* of the state $|j, k, \lambda\rangle$, and m, the multiplicity of the state $|j + \frac{1}{2}, k + \frac{1}{2}, \lambda + 1\rangle$. Equation (2.49) now appears as

$$\begin{aligned} a_1(j, k, \lambda)_{\alpha\beta} &= b_4(j + \tfrac{1}{2}, k + \tfrac{1}{2}, \lambda + 1)_{\beta\alpha}, \\ a_2(j, k, \lambda)_{\alpha\beta} &= b_3(j + \tfrac{1}{2}, k - \tfrac{1}{2}, \lambda + 1)_{\beta\alpha}, \\ a_3(j, k, \lambda)_{\alpha\beta} &= b_2(j - \tfrac{1}{2}, k + \tfrac{1}{2}, \lambda + 1)_{\beta\alpha}, \\ a_4(j, k, \lambda)_{\alpha\beta} &= b_1(j - \tfrac{1}{2}, k - \tfrac{1}{2}, \lambda + 1)_{\beta\alpha}. \end{aligned} \quad (3.11)$$

C. Multiplicity and Degeneracy

Let us define the multiplicity to be $M(j, k, \lambda)$. It is the number of times the irreducible representation (j, k, λ) of $SU(2) \dot{\times} SU(2) \times U(1)$ appears in an irreducible representation of $SU(2, 2)$. $M(j, k, \lambda) = 0$ means the irreducible representation (j, k, λ) does not occur at all; $M(j, k, \lambda) = 1$ means (j, k, λ) occurs only once, and so on. If $M(j, k, \lambda) = 0$ or 1 for all j, k, λ in an irreducible representation of $SU(2, 2)$, then we say that we are dealing with a degenerate representation. In general, some $M(j, k, \lambda) > 1$ for some j, k, λ, and we have a non-degenerate representation.

For the degenerate representations, F_3 becomes a redundant label, and hence can be omitted. Since Murai neglected the operator F_3 from the beginning of his work, he dealt exclusively with degenerate representations.

IV. GENERAL EXPRESSIONS FOR THE CASIMIR OPERATORS AND F_3

We now express the Casimir operators C_2, C_3, C_4 and also F_3 in terms of the functions $a_i(j, k, \lambda)_{\alpha\beta}$ and $b_i(j, k, \lambda)_{\alpha\beta}$. Using the matrix version of Eqs. (2.41a–d) and (2.43a–d) in conjunction with Eqs. (3.1), (3.2), (3.3), and (3.6), we obtain for the diagonal elements [after some straightforward calculation in which Eq. (3.11) is used]

$$\begin{aligned} &C_2 - 2j(j+1) - 2k(k+1) - \lambda(\lambda+4) \\ &\quad = 8\sum_\rho \{(j+1)(k+1)a_1^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad - (j+1)ka_2^2(j,k,\lambda)_{\alpha\rho} - j(k+1)a_3^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + jka_4^2(j,k,\lambda)_{\alpha\rho}\}, \end{aligned} \tag{4.1a}$$

$$\begin{aligned} &C_2 - 2j(j+1) - 2k(k+1) - \lambda(\lambda-4) \\ &\quad = 8\sum_\rho \{(j+1)(k+1)b_1^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad - (j+1)kb_2^2(j,k,\lambda)_{\alpha\rho} - j(k+1)b_3^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + jkb_4^2(j,k,\lambda)_{\alpha\rho}\}, \end{aligned} \tag{4.1b}$$

$$\begin{aligned} &C_3 + (\lambda+2)(j-k)(j+k+1) \\ &\quad = 4\sum_\rho \{-(j+1)(k+1)(j-k)a_1^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + (j+1)k(j+k+1)a_2^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad - j(k+1)(j+k+1)a_3^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + jk(j-k)a_4^2(j,k,\lambda)_{\alpha\rho}\}, \end{aligned} \tag{4.2a}$$

$$\begin{aligned} &C_3 + (\lambda-2)(j-k)(j+k+1) \\ &\quad = 4\sum_\rho \{(j+1)(k+1)(j-k)b_1^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad -(j+1)k(j+k+1)b_2^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + j(k+1)(j+k+1)b_3^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad - jk(j-k)b_4^2(j,k,\lambda)_{\alpha\rho}\}, \end{aligned} \tag{4.2b}$$

$$\begin{aligned} &C_4 - \tfrac{1}{2}\lambda(\lambda+2)C_2 + \tfrac{1}{4}\lambda(\lambda-2)(\lambda+2)(\lambda+4) \\ &\quad + 2(\lambda+1)(\lambda+2)[j(j+1)+k(k+1)] \\ &\quad - 4j(j+1)k(k+1) \\ &\quad - 4\{(P_-Q_- - S_-T_-)(P_+Q_+ - S_+T_+)\}_{\alpha\alpha} \\ &= 8\sum_\rho \{(j+1)(k+1)[2jk - (\lambda+1)(j+k)] \\ &\qquad \times a_1^2(j,k,\lambda)_{\alpha\rho} + (j+1)k[2j(k+1) \\ &\qquad + (\lambda+1)(j-k-1)]a_2^2(j,k,\lambda)_{\alpha\rho} + j(k+1) \\ &\qquad \times [2(j+1)k - (\lambda+1)(j-k+1)]a_3^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + jk[2(j+1)(k+1) + (\lambda+1)(j+k+2)] \\ &\qquad \times a_4^2(j,k,\lambda)_{\alpha\rho}\}, \end{aligned} \tag{4.3a}$$

$$\begin{aligned} &C_4 - \tfrac{1}{2}\lambda(\lambda-2)C_2 + \tfrac{1}{4}\lambda(\lambda-2)(\lambda+2)(\lambda-4) \\ &\quad + 2(\lambda-1)(\lambda-2)[j(j+1)+k(k+1)] \\ &\quad - 4j(j+1)k(k+1) \\ &\quad - 4\{(P_+Q_+ - S_+T_+)(P_-Q_- - S_-T_-)\}_{\alpha\alpha} \\ &= 8\sum_\rho \{(j+1)(k+1)[2jk + (\lambda-1)(j+k)] \\ &\qquad \times b_1^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + (j+1)k[2j(k+1) - (\lambda-1)(j-k-1)] \\ &\qquad \times b_2^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + j(k+1)[2(j+1)k + (\lambda-1)(j-k+1)] \\ &\qquad \times b_3^2(j,k,\lambda)_{\alpha\rho} \\ &\qquad + jk[2(j+1)(k+1) - (\lambda-1)(j+k+2)] \\ &\qquad \times b_4^2(j,k,\lambda)_{\alpha\rho}\}, \end{aligned} \tag{4.3b}$$

where

$$\begin{aligned} &\{(P_-Q_- - S_-T_-)(P_+Q_+ - S_+T_+)\}_{\alpha\alpha} \\ &\qquad = \sum_\sigma (P_+Q_+ - S_+T_+)_{\alpha\sigma}(P_-Q_- - S_-T_-)_{\sigma\alpha}, \end{aligned}$$

$$\begin{aligned} &\{(P_+Q_+ - S_+T_+)(P_-Q_- - S_-T_-)\}_{\alpha\alpha} \\ &\qquad = \sum_\tau (P_-Q_- - S_-T_-)_{\alpha\tau}(P_+Q_+ - S_+T_+)_{\tau\alpha}, \end{aligned}$$

and (from Appendix A)

$$\begin{aligned} &(P_+Q_+ - S_+T_+)_{\alpha\sigma} \\ &\quad = -2(2k+1)\sum_\rho \{(j+1)a_1(j,k,\lambda)_{\alpha\rho} \\ &\qquad \times a_4(j+\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\sigma} \\ &\qquad - ja_3(j,k,\lambda)_{\alpha\rho}a_2(j-\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\sigma}\}, \end{aligned}$$

$$(P_-Q_- - S_-T_-)_{\sigma\alpha} = (P_+Q_+ - S_+T_+)_{\alpha\sigma},$$

$$\begin{aligned} &(P_-Q_- - S_-T_-)_{\alpha\tau} \\ &\quad = -2(2k+1)\sum_\rho \{(j+1)b_1(j,k,\lambda)_{\alpha\rho} \\ &\qquad \times b_4(j+\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda-1)_{\rho\tau} \\ &\qquad - jb_3(j,k,\lambda)_{\alpha\rho}b_2(j-\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda-1)_{\rho\tau}\}, \end{aligned}$$

$$(P_+Q_+ - S_+T_+)_{\tau\alpha} = (P_-Q_- - S_-T_-)_{\alpha\tau}. \tag{4.4}$$

For the operator F_3, we have

$$\begin{aligned}\alpha + (\lambda+2)[j(j+1)+k(k+1)] &= 4\sum_\rho \{-(j+1)(k+1)(j+k)a_1^2(j,k,\lambda)_{\alpha\rho} \\ &\quad + (j+1)k(j-k-1)a_2^2(j,k,\lambda)_{\alpha\rho} \\ &\quad - j(k+1)(j-k+1)a_3^2(j,k,\lambda)_{\alpha\rho} \\ &\quad + jk(j+k+2)a_4^2(j,k,\lambda)_{\alpha\rho}\}, \end{aligned} \tag{4.5a}$$

$$\begin{aligned}\alpha + (\lambda-2)[j(j+1)+k(k+1)] &= 4\sum_\rho \{(j+1)(k+1)(j+k)b_1^2(j,k,\lambda)_{\alpha\rho} \\ &\quad - (j+1)k(j-k-1)b_2^2(j,k,\lambda)_{\alpha\rho} \\ &\quad + j(k+1)(j-k+1)b_3^2(j,k,\lambda)_{\alpha\rho} \\ &\quad - jk(j+k+2)b_4^2(j,k,\lambda)_{\alpha\rho}\}. \end{aligned} \tag{4.5b}$$

For completeness, the off-diagonal elements are given in Appendix B, and the commutation relations between F_3 and $P_\pm$, $Q_\pm$, $S_\pm$, and $T_\pm$ are included in Appendix C.

Equations (4.1)–(4.5) are the most important and useful equations in this paper, and they are used over and over again. We observe that the problem of solving for $a_i(j,k,\lambda)_{\alpha\beta}$ and $b_i(j,k,\lambda)_{\alpha\beta}$ is extremely complicated when we deal with the nondegenerate representations, even though we have the many relations in (3.11), Appendices A and B. The functions $a_i(j,k,\lambda)_{\alpha\beta}$ and $b_i(j,k,\lambda)_{\alpha\beta}$ are coupled in a very intricate manner, which makes a systematic approach rather difficult. However, because of the close relationship between the finite irreducible representations which are nonunitary and the unitary irreducible representations in the discrete series, much information about the discrete series can be obtained. Before we proceed to a study of the finite representations, we first derive the recursion relations among the $a_i(j,k,\lambda)$ and $b_i(j,k,\lambda)$.

D. RECURSION RELATIONS

From Eqs. (4.1a), (4.1b), (4.2a), and (4.2b) we can express

$$\sum_\rho a_2^2(j,k,\lambda)_{\alpha\rho}, \quad \sum_\rho a_3^2(j,k,\lambda)_{\alpha\rho}, \quad \sum_\rho b_2^2(j,k,\lambda)_{\alpha\rho},$$

and

$$\sum_\rho b_3^2(j,k,\lambda)_{\alpha\rho}$$

as follows:

$$\begin{aligned}2(j+1)k(j+k+1)\sum_\rho a_2^2(j,k,\lambda)_{\alpha\rho} &= \sum_\rho \{(j+1)(k+1)(2j+1)a_1^2(j,k,\lambda)_{\alpha\rho} \\ &\quad + jk(2k+1)a_4^2(j,k,\lambda)_{\alpha\rho}\} \\ &\quad + \tfrac{1}{4}\{C_3 + (\lambda+2)(j-k)(j+k+1)\} \\ &\quad - \tfrac{1}{8}(j+k+1) \\ &\quad \times \{C_2 - 2j(j+1) - 2k(k+1) - \lambda(\lambda+4)\}, \end{aligned} \tag{4.6}$$

$$\begin{aligned}2j(k+1)(j+k+1)\sum_\rho a_3^2(j,k,\lambda)_{\alpha\rho} &= \sum_\rho \{(j+1)(k+1)(2k+1)a_1^2(j,k,\lambda)_{\alpha\rho} \\ &\quad + jk(2j+1)a_4^2(j,k,\lambda)_{\alpha\rho}\} \\ &\quad - \tfrac{1}{4}\{C_3 + (\lambda+2)(j-k)(j+k+1)\} \\ &\quad - \tfrac{1}{8}(j+k+1) \\ &\quad \times \{C_2 - 2j(j+1) - 2k(k+1) - \lambda(\lambda+4)\}, \end{aligned} \tag{4.7}$$

$$\begin{aligned}2(j+1)k(j+k+1)\sum_\rho b_2^2(j,k,\lambda)_{\alpha\rho} &= \sum_\rho \{(j+1)(k+1)(2j+1)b_1^2(j,k,\lambda)_{\alpha\rho} \\ &\quad + jk(2k+1)b_4^2(j,k,\lambda)_{\alpha\rho}\} \\ &\quad - \tfrac{1}{4}\{C_3 + (\lambda-2)(j-k)(j+k+1)\} \\ &\quad - \tfrac{1}{8}(j+k+1) \\ &\quad \times \{C_2 - 2j(j+1) - 2k(k+1) - \lambda(\lambda-4)\}, \end{aligned} \tag{4.8}$$

$$\begin{aligned}2j(k+1)(j+k+1)\sum_\rho b_3^2(j,k,\lambda)_{\alpha\rho} &= \sum_\rho \{(j+1)(k+1)(2k+1)b_1^2(j,k,\lambda)_{\alpha\rho} \\ &\quad + jk(2j+1)b_4^2(j,k,\lambda)_{\alpha\rho}\} \\ &\quad + \tfrac{1}{4}\{C_3 + (\lambda-2)(j-k)(j+k+1)\} \\ &\quad - \tfrac{1}{8}(j+k+1) \\ &\quad \times \{C_2 - 2j(j+1) - 2k(k+1) - \lambda(\lambda-4)\}. \end{aligned} \tag{4.9}$$

Substituting Eqs. (4.6)–(4.9) into Eqs. (4.3a) and (4.3c), we obtain a relation among

$$\sum_\rho a_1^2(j,k,\lambda)_{\alpha\rho}, \quad \sum_\rho a_4^2(j,k,\lambda)_{\alpha\rho}, \quad \sum_\rho b_1^2(j,k,\lambda)_{\alpha\rho},$$

and

$$\sum_\rho b_4^2(j,k,\lambda)_{\alpha\rho}.$$

Similarly, from Eqs. (4.5a) and (4.5b), we obtain another relation. From these two relations, we can solve for $\sum_\rho a_1^2(j,k,\lambda)_{\alpha\rho}$ and $\sum_\rho b_1^2(j,k,\lambda)_{\alpha\rho}$ in terms of $\sum_\rho a_4^2(j,k,\lambda)_{\alpha\rho}$ and $\sum_\rho b_4^2(j,k,\lambda)_{\alpha\rho}$:

$$\begin{aligned}(j+k)(2j+1)(2k+1)(j+1)(k+1)\sum_\rho a_1^2(j,k,\lambda)_{\alpha\rho} &= jk(2j+1)(2k+1) \\ &\quad \times \sum_\rho \{-a_4^2(j,k,\lambda)_{\alpha\rho} + (j+k+1)b_4^2(j,k,\lambda)_{\alpha\rho}\} \\ &\quad + \tfrac{1}{8}(j+k)(j+k+1)C_2 - \tfrac{1}{4}(j-k)C_3 \\ &\quad - \tfrac{1}{8}(j+k)(j+k+1) \\ &\quad \times \{(\lambda+2j+2)(\lambda+2k+2) \\ &\quad + 2(j+k-1)(j+k+2)\}, \end{aligned} \tag{4.10}$$

$$(j+k)(2j+1)(2k+1)(j+1)(k+1)\sum_{\rho} b_1^2(j,k,\lambda)_{\alpha\rho}$$
$$= jk(2j+1)(2k+1)$$
$$\times \sum_{\rho}\{(j+k+1)a_4^2(j,k,\lambda)_{\alpha\rho} - b_4^2(j,k,\lambda)_{\alpha\rho}\}$$
$$+ \tfrac{1}{8}(j+k)(j+k+1)C_2 + \tfrac{1}{4}(j-k)C_3$$
$$- \tfrac{1}{8}(j+k)(j+k+1)$$
$$\times \{(\lambda-2j-2)(\lambda-2k-2)$$
$$+ 2(j+k-1)(j+k+2)\}. \qquad (4.11)$$

Therefore, once $\sum_\rho a_4^2(j,k,\lambda)_{\alpha\rho}$ and $\sum_\rho b_4^2(j,k,\lambda)_{\alpha\rho}$ are known, $\sum_\rho a_i^2(j,k,\lambda)_{\alpha\rho}$ and $\sum_\rho b_i^2(j,k,\lambda)_{\alpha\rho}$, $i = 1, 2, 3$, can be calculated. Since

$$a_4(j,k,\lambda)_{\alpha\beta} = b_1(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\beta\alpha}$$

and

$$b_4(j,k,\lambda)_{\alpha\beta} = a_1(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda-1)_{\beta\alpha},$$

(notice the transpose nature of the relations,) Eqs. (4.10) and (4.11) are *almost* recursion relations in $a_1^2(j,k,\lambda)$ and $b_1^2(j,k,\lambda)$. Unfortunately, in Eqs. (4.10) and (4.11), the *second* index ρ is summed over, this fact prevents these two equations from being true recursion relations. However, we may sum over α in Eqs. (4.10) and (4.11), but now the terms that do not depend on α and ρ must be multiplied by the multiplicity of the state $|j,k,\lambda\rangle$, $M(j,k,\lambda)$. For clarity, we rewrite Eqs. (4.10) and (4.11):

$$(j+k)(2j+1)(2k+1)(j+1)(k+1)\sum_{\alpha,\rho} a_1^2(j,k,\lambda)_{\alpha\rho}$$
$$= jk(2j+1)(2k+1)$$
$$\times \sum_{\alpha,\rho}\{-a_4^2(j,k,\lambda)_{\alpha\rho} + (j+k+1)b_4^2(j,k,\lambda)_{\alpha\rho}\}$$
$$+ M(j,k,\lambda)[\tfrac{1}{8}(j+k)(j+k+1)C_2$$
$$- \tfrac{1}{4}(j-k)C_3 - \tfrac{1}{8}(j+k)(j+k+1)$$
$$\times \{(\lambda+2j+2)(\lambda+2k+2)$$
$$+ 2(j+k-1)(j+k+2)\}], \qquad (4.12)$$

$$(j+k)(2j+1)(2k+1)(j+1)(k+1)\sum_{\alpha,\rho} b_1^2(j,k,\lambda)_{\alpha\rho}$$
$$= jk(2j+1)(2k+1)$$
$$\times \sum_{\alpha,\rho}\{(j+k+1)a_4^2(j,k,\lambda)_{\alpha\rho} - b_4^2(j,k,\lambda)_{\alpha\rho}\}$$
$$+ M(j,k,\lambda)[\tfrac{1}{8}(j+k)(j+k+1)C_2$$
$$+ \tfrac{1}{4}(j-k)C_3 - \tfrac{1}{8}(j+k)(j+k+1)$$
$$\times \{(\lambda-2j-2)(\lambda-2k-2)$$
$$+ 2(j+k-1)(j+k+2)\}], \qquad (4.13)$$

and

$$\sum_{\alpha,\rho} a_4^2(j,k,\lambda)_{\alpha\rho} = \sum_{\alpha,\rho} b_1^2(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\alpha},$$

$$\sum_{\alpha,\rho} b_4^2(j,k,\lambda)_{\alpha\rho} = \sum_{\alpha,\rho} a_1^2(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda-1)_{\rho\alpha}.$$

For degenerate representations $M(j,k,\lambda) = 1$, Eqs. (4.10) and (4.11) are truly recursion relations, and the equations can be solved easily. In the general case, $M(j,k,\lambda)$ is unknown and depends on j, k, λ, this fact acts as the major obstacle in the solution of the problem. (A detailed discussion will be given in Paper III.) For the unitary irreducible representations in the discrete series, it is shown that we are able to solve the coupled diaphantine recursion relations (4.12) and (4.13).

V. FINITE IRREDUCIBLE REPRESENTATIONS

Since the representation is finite dimensional, there must exist a state $|j_m, k_m, \lambda_m\rangle$, (we omit the labels μ, ν, and α,) such that

$$a_1(j_m,k_m,\lambda_m) = a_2(j_m,k_m,\lambda_m),$$
$$= a_3(j_m,k_m,\lambda_m) = a_4(j_m,k_m,\lambda_m) = 0, \qquad (5.1)$$

and all the allowed λ satisfies $\lambda \leq \lambda_m$ [actually $|\lambda| \leq \lambda_m$, because of Eq. (5.5)]. From Eqs. (4.1a), (4.2a), and (4.3a), we get

$$C_2 = 2j_m(j_m+1) + 2k_m(k_m+1) + \lambda_m(\lambda_m+4), \qquad (5.2)$$

$$C_3 = -(\lambda_m+2)(j_m-k_m)(j_m+k_m+1), \qquad (5.3)$$

$$C_4 = \tfrac{1}{4}[(\lambda_m+2)^2 - 4j_m(j_m+1)]$$
$$\times [(\lambda_m+2)^2 - 4k_m(k_m+1)] - (\lambda_m+2)^2. \qquad (5.4)$$

By symmetry, we observe that there must exist the state $|j = k_m, k = j_m, \lambda = -\lambda_m\rangle$, such that

$$b_1(k_m,j_m,-\lambda_m) = b_2(k_m,j_m,-\lambda_m)$$
$$= b_3(k_m,j_m,-\lambda_m) = b_4(k_m,j_m,-\lambda_m) = 0, \qquad (5.5)$$

and Eqs. (5.2)–(5.4) satisfy Eqs. (4.1b), (4.2b), and (4.3b). Therefore, j_m, k_m, and λ_m uniquely determines the Casimir operators C_2, C_3, and C_4; the converse is also true. Hereafter, we use the symbol $((j_m, k_m, \lambda_m))$ to denote a finite irreducible representation of $SU(2, 2)$. From Eq. (4.5a), we see that

$$\alpha(j_m,k_m,\lambda_m)$$
$$= -(\lambda_m+2)[j_m(j_m+1) + k_m(k_m+1)].$$

Since α is uniquely determined above, the state $|j_m, k_m, \lambda_m\rangle$ occurs only once, $M(j_m, k_m, \lambda_m) = 1$. We see shortly that all states on the "boundary" have multiplicity one.

Let

$$p = j + k, \quad q = j - k,$$

and

$$|p,q,\lambda) \equiv |j,k,\lambda\rangle, \quad |p_m,q_m,\lambda_m) \equiv |j_m,k_m,\lambda_m\rangle.$$

From the state $|p_m, q_m, \lambda_m)$, we reach (along the boundary by applying P_-, Q_-, etc.)

$$|p_m + 1, q_m, \lambda_m - 1)$$

and

$$|p_m - 1, q_m, \lambda_m - 1).$$

In general, we have $|p_m + s, q_m, \lambda_m - s)$, $s = 0, 1, \cdots, s_m$, and $|p_m - t, q_m, \lambda_m - t)$, $t = 0, 1, \cdots, t_m$. These states have the following properties:

$$a_1\left(j_m + \frac{s}{2}, k_m + \frac{s}{2}, \lambda_m - s\right) = a_2\left(j_m + \frac{s}{2}, k_m + \frac{s}{2}, \lambda_m - s\right) = a_3\left(j_m + \frac{s}{2}, k_m + \frac{s}{2}, \lambda_m - s\right) = 0, \quad s = 0, 1, \cdots, s_m,$$

$$a_4\left(j_m + \frac{s}{2}, k_m + \frac{s}{2}, \lambda_m - s\right) \neq 0, \quad s = 1, 2, \cdots s_m, \tag{5.6}$$

$$a_2\left(j_m - \frac{t}{2}, k_m - \frac{t}{2}, \lambda_m - t\right) = a_3\left(j_m - \frac{t}{2}, k_m - \frac{t}{2}, \lambda_m - t\right) = a_4\left(j_m - \frac{t}{2}, k_m - \frac{t}{2}, \lambda_m - t\right) = 0, \quad t = 0, 1, 2, \cdots, t_m,$$

$$a_1\left(j_m - \frac{t}{2}, k_m - \frac{t}{2}, \lambda_m - t\right) \neq 0, \quad t = 1, 2, \cdots, t_m. \tag{5.7}$$

Equations (5.6) and (5.7) follow from Eqs. (5.1) and the uniqueness of the state $|j_m, k_m, \lambda_m\rangle$. Using Eqs. (4.1a) and (5.2), we can solve for

$$a_1\left(j_m - \frac{t}{2}, k_m - \frac{t}{2}, \lambda_m - t\right)$$

and

$$a_4\left(j_m + \frac{s}{2}, k_m + \frac{s}{2}, \lambda_m - s\right),$$

$$a_1^2\left(j_m - \frac{t}{2}, k_m - \frac{t}{2}, \lambda_m - t\right) = \frac{t(\lambda_m + j_m + k_m - t + 3)}{(2j_m - t + 2)(2k_m - t + 2)}, \quad t = 0, 1, 2, \cdots, t_m, \tag{5.8a}$$

and using Eq. (3.11), we get

$$b_4^2\left(j_m - \frac{t}{2}, k_m - \frac{t}{2}, \lambda_m - t\right) = \frac{(t + 1)(\lambda_m + j_m + k_m - t + 2)}{(2j_m - t + 1)(2k_m - t + 1)}, \quad t = 0, 1, 2, \cdots t_m - 1, \tag{5.8b}$$

where $t_m = 2k_m$, if $j_m > k_m$

$= 2j_m$, if $k_m > j_m$,

and

$$b_4^2\left(j_m - \frac{t_m}{2}, k_m - \frac{t_m}{2}, \lambda_m - t_m\right) = 0,$$

since $j \geq 0$, $k \geq 0$.

Similarly,

$$a_4^2\left(j_m + \frac{s}{2}, k_m + \frac{s}{2}, \lambda_m - s\right) = \frac{s(\lambda_m - j_m - k_m - s + 1)}{(2j_m + s)(2k_m + s)}, \quad s = 0, 1, 2, \cdots, s_m, \tag{5.9a}$$

and

$$b_1^2\left(j_m + \frac{s}{2}, k_m + \frac{s}{2}, \lambda_m - s\right) = \frac{(s + 1)(\lambda_m - j_m - k_m - s)}{(2j_m + s + 1)(2k_m + s + 1)}, \quad s = 0, 1, 2, \cdots, s_m. \tag{5.9b}$$

For finite representations, there must exist a finite s_m, such that

$$b_1^2(j_m + s_m/2, k_m + s_m/2, \lambda_m - s_m) = 0.$$

From Eq. (5.9b), we see therefore

$$s_m = \lambda_m - j_m - k_m, \tag{5.10}$$

where s_m is an integer.

Now, we have a complete classification of finite irreducible representations. Since the Casimir operators C_2, C_3, C_4 are determined by j_m, k_m, and $s_m(\lambda_m = j_m + k_m + s_m)$, we obtain *all* finite representations when j_m, k_m, and s_m take on all possible values:

$$j_m = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots,$$
$$k_m = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots,$$
$$s_m = 0, 1, 2, 3, \cdots. \tag{5.11}$$

In passing, we give without proof the formula for the dimensionality of the finite irreducible representation $((j_m, k_m, \lambda_m))$:

$$D(j_m, k_m, s_m) = (3!\,2!)^{-1}(2j_m + 1)(2k_m + 1)(s_m + 1) \times (2j_m + s_m + 2)(2j_m + s_m + 2) \times (2j_m + 2k_m + s_m + 3). \tag{5.12}$$

Before proceeding to a study of the problem of multiplicity and degeneracy, we wish to complete

the discussion of the states $|p_m + s, q_m, \lambda_m - s)$ and $|p_m - t, q_m, \lambda_m - t)$. These states are said to be on the "boundary" of a $p - \lambda$ diagram. (See Fig. 1.) There are, in general, six boundaries given by the following equations:

$$p + \lambda = p_m + \lambda_m, \tag{5.13a}$$
$$p - \lambda = p_m - \lambda_m, \tag{5.13b}$$
$$p = p_m + s_m \neq \lambda_m, \tag{5.13c}$$
$$p = p_m - t_m = |j_m - k_m|. \tag{5.13d}$$

Because of symmetry, the states

$$|p_m + s, -q_m, -\lambda_m + s)$$

and

$$|p_m - t, -q_m, -\lambda_m + t)$$

also exist and are on the boundary, and we have

$$p - \lambda = p_m + \lambda_m, \tag{5.13e}$$
$$p + \lambda = p_m - \lambda_m. \tag{5.13f}$$

It is obvious that on the boundary, $\alpha(j, k, \lambda)$ is uniquely determined by C_2, C_3, j, k, and λ, and $M(j, k, \lambda) = 1$. For clarity, we state this fact as a theorem.

Theorem 1: All states on the boundary have multiplicity $M(j, k, \lambda) = 1$.

Proof: 1. For states on the boundaries (5.13a) and (5.13b), we have Eqs. (5.6) and (5.7) , hence, $\alpha(j, k, \lambda)$ is uniquely determined through Eqs. (4.1a) and (4.5a).

2. Similarly, for states on the boundaries (5.13e) and (5.13f), $\alpha(j, k, \lambda)$ is determined through Eqs. (4.1b) and (4.5b).

3. For states on the boundary (5.13d),

$$a_4(j_m - t_m/2, k_m - t_m/2, \lambda) = 0,$$
$$a_2(j_m - t_m/2, k_m - t_m/2, \lambda) = 0, \quad \text{if } j_m > k_m,$$
$$a_3(j_m - t_m/2, k_m - t_m/2, \lambda) = 0, \quad \text{if } k_m > j_m,$$

therefore, $\alpha(j, k, \lambda)$ is determined through (4.1a) and (4.2a).

4. For states on the boundary (5.13c),

$$a_1(p = \lambda_m, q, \lambda) = 0, \quad b_1(p = \lambda_m, q, \lambda) = 0,$$

then $\alpha(j, k, \lambda)$ is determined through (4.6)–(4.11).

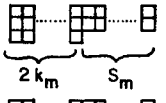
A. Degeneracy and Multiplicity

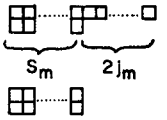
Our task now is to classify the irreducible representations of Eq. (5.11) into degenerate and nondegenerate representations, find the associated Young diagrams, the multiplicity of the states, and the decomposition of each representation with respect to the maximal compact subgroup. The results of this investigation are quite simple, in a sense even intuitively obvious, and we state them in the form of several theorems. However, the algebraic steps involved in this derivation are rather lengthy, and are relegated to Appendix D.

Theorem 1: The finite irreducible representations of the group $SU(2, 2)$, and the associated Young diagrams are classified as follows:

(a) Degenerate representation:

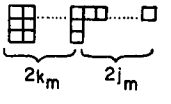
Class I. $s_m = 1, 2, 3, \cdots,$

i. $j_m = 0, k_m = \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots,$ 

ii. $k_m = 0, j_m = \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots,$ 

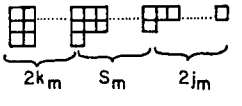
iii. $j_m = k_m = 0,$ 

Class II. $s_m = 0,$

$j_m, k_m = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots;$ 

(b) Nondegenerate representations:

Class III. $s_m = 1, 2, 3, \cdots,$

$j_m, k_m = \frac{1}{2}, 1, \frac{3}{2}, 2, \cdots.$ 

Proof: See Appendix D. We see that the Young diagram for nondegenerate representations is the most general diagram. When j_m, k_m, and s_m are *all* nonvanishing, we have Class III. When $j_m = 0$, we have Class I, (i); when $k_m = 0$, we have Class I, (ii); when $j_m = k_m = 0$, we have Class I, (iii); when $s_m = 0$, we have Class II; these are all degenerate representations. Therefore, for finite representations, degenerate also means that the Young diagram is degenerate.

The relation between a Young diagram and j_m, k_m, s_m can be shown very simply. When we decompose $SU(2, 2)$ with respect to $SU(2) \times SU(2) \times U(1)$, we

have for the three basic Young diagrams

$$\square = (j = 0, k = \tfrac{1}{2}, \lambda = -\tfrac{1}{2}) + (j = \tfrac{1}{2}, k = 0, \lambda = \tfrac{1}{2}) = ((j_m = \tfrac{1}{2}, k_m = 0, \lambda_m = \tfrac{1}{2})),$$

$$\begin{array}{|c|}\hline\\\hline\\\hline\end{array} = (j = \tfrac{1}{2}, k = 0, \lambda = -\tfrac{1}{2}) + (j = 0, k = \tfrac{1}{2}, \lambda = \tfrac{1}{2}) = ((j_m = 0, k_m = \tfrac{1}{2}, \lambda_m = \tfrac{1}{2})),$$

$$\begin{array}{|c|}\hline\\\hline\end{array} = (j = 0, k = 0, \lambda = -1) + (j = \tfrac{1}{2}, k = \tfrac{1}{2}, \lambda = 0) + (j = 0, k = 0, \lambda = 1) = ((j_m = 0, k_m = 0, \lambda_m = 1)),$$

where we recall (j, k, λ) is an irreducible representation of $SU(2) \times SU(2) \times U(1)$ and $((j_m, k_m, \lambda_m))$ is a finite irreducible representation of $SU(2, 2)$. Therefore, for a given Young diagram which has $2j_m$ columns of $\square$, s_m columns of $\begin{array}{|c|}\hline\\\hline\end{array}$, and $2k_m$ columns of $\begin{array}{|c|}\hline\\\hline\\\hline\end{array}$, the state of maximum weight (maximal λ) is constructed from

$$2j_m \times (j = \tfrac{1}{2}, k = 0, \lambda = \tfrac{1}{2}) + 2k_m \times (j = 0, k = \tfrac{1}{2}, \lambda = \tfrac{1}{2}) + s_m \times (j = 0, k = 0, \lambda = 1) = (j = j_m, k = k_m, \lambda = j_m + k_m + s_m = \lambda_m).$$

We now consider the problem of the decomposition of a finite irreducible representation of $SU(2, 2)$ into unitary irreducible representations of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$, and the multiplicities of these representations.

Theorem 2: The decomposition of a finite irreducible representation $((j_m, k_m, \lambda_m))$ of $SU(2, 2)$ into unitary irreducible representations (j, k, λ) of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ and the multiplicities $M(j, k, \lambda)$ of these representations are given as follows:

Let $p = j + k$, $q = j - k$, $p_m = j_m + k_m$, then $|q| \leq p$ always, and for $p \geq p_m$, $|q| \leq p_m$.

(a) Degenerate representations: $M(j, k, \lambda) = 1$ for all allowed values of j, k, and λ.

Class I. $s_m = 1, 2, 3, \cdots$.

(i) $j_m = 0, k_m = \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots$.

$$\left.\begin{array}{r} p = p_m \\ -s_m \leq q + \lambda \leq s_m \\ -2p_m - s_m \leq q - \lambda \leq 2p_m + s_m \end{array}\right\} \left.\begin{array}{r} p = p_m + 1 \\ -s_m + 1 \leq q + \lambda \leq s_m - 1 \\ -2p_m - s_m + 1 \leq q - \lambda \leq 2p_m + s_m - 1 \end{array}\right\} \cdots \left.\begin{array}{r} p = p_m + s \\ -s_m + s \leq q + \lambda \leq s_m - s \\ -2p_m - s_m + s \leq q - \lambda \leq 2p_m + s_m - s \end{array}\right\} \cdots \left.\begin{array}{r} p = p_m + s_m \\ q + \lambda = 0 \\ -2p_m \leq q - \lambda \leq 2p_m \end{array}\right\}, \quad (5.14)$$

where $s = 0, 1, 2, \cdots, s_m$. Here, we have the general restriction,

$-p_m \leq q \leq p_m$. $-s_m + s \leq q + \lambda \leq s_m - s$, means

$$q + \lambda = -s_m + s, \quad -s_m + s + 2, \quad -s_m + s + 4, \cdots, s_m - s;$$

and

$-2p_m - s_m + s \leq q - \lambda \leq 2p_m + s_m - s$ means

$$q - \lambda = -2p_m - s_m + s, \quad -2p_m - s_m + s + 2, \quad -2p_m - s_m + s + 4, \cdots, 2p_m + s_m - s,$$

(ii) $k_m = 0$, $j_m = \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \cdots$ is the same as (5.14) with $q \to -q$.

(iii) $j_m = k_m = 0$ is the same as (5.14) with $p_m = 0$, $q = 0$.

Class II. $s_m = 0$, $j_m, k_m = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots$.

(i) $j_m > k_m$

$j_m - k_m =$ integer, or half-integer

$$\left.\begin{array}{r} p = j_m - k_m \\ -2(j_m - k_m) \leq q + \lambda \leq 2(j_m - k_m) \\ q - \lambda = 0 \end{array}\right\} \cdots \left.\begin{array}{r} p = j_m - k_m + \gamma \\ -2(j_m - k_m) - \gamma \leq q + \lambda \leq 2(j_m - k_m) + \gamma \\ -\gamma \leq q - \lambda \leq \gamma \end{array}\right\} \cdots \left.\begin{array}{r} p = j_m + k_m \\ -2j_m \leq q + \lambda \leq 2j_m \\ -2k_m \leq q - \lambda \leq 2k_m \end{array}\right\}, \quad (5.15)$$

where $\gamma = 0, 1, 2, \cdots, 2k_m$. $-\gamma \leq q - \lambda \leq \gamma$ means that $q - \lambda = -\gamma, -\gamma + 2, -\gamma + 4, \cdots, \gamma$.

(ii) $k_m > j_m$ is the same as (5.15) with $j_m \leftrightarrow k_m$, $q \to -q$.
(iii) $j_m = k_m$ is the same as (5.15) with $j_m = k_m$.
(b) Nondegenerate representations $M(j, k, \lambda) \geq 1$ for all allowed values of j, k, λ.
Class III. $s_m = 1, 2, 3, \cdots, j_m, k_m = \frac{1}{2}, 1, \frac{3}{2}, \cdots$.
(i) $j_m > k_m$,

$$\left.\begin{array}{c} p = j_m - k_m, \\ -2(j_m - k_m) - s_m \leq q + \lambda \leq 2(j_m - k_m) + s_m \\ -s_m \leq q - \lambda \leq s_m \\ |q| \leq p \end{array}\right\} \cdots \Bigg\}$$

$$\left.\begin{array}{c} p = j_m - k_m + s \\ -2(j_m - k_m) - s_m - s \leq q + \lambda \leq 2(j_m - k_m) + s_m + s \\ -(s_m + s) \leq q - \lambda \leq s_m + s \\ |q| \leq p \end{array}\right\} \cdots \Bigg\}$$

$$\left.\begin{array}{c} p = p_m = j_m + k_m \\ -(2j_m + s_m) \leq q + \lambda \leq 2j_m + s_m \\ -(2k_m + s_m) \leq q - \lambda \leq 2k_m + s_m \\ |q| \leq p = p_m \end{array}\right\} \cdots \Bigg\} \quad \left.\begin{array}{c} p = j_m + k_m + \gamma \\ -(2j_m + s_m - \gamma) \leq q + \lambda \leq 2j_m + s_m - \gamma \\ -(2k_m + s_m - \gamma) \leq q - \lambda \leq 2k_m + s_m - \gamma \\ |q| \leq p_m \end{array}\right\} \cdots \Bigg\}$$

$$\left.\begin{array}{c} p = j_m + k_m + s_m \\ -2j_m \leq q + \lambda \leq 2j_m \\ -2k_m \leq q - \lambda \leq 2k_m \\ |q| \leq p_m \end{array}\right\}, \tag{5.16}$$

where $s = 0, 1, 2, \cdots, 2k_m$, and $\gamma = 0, 1, 2, \cdots, s_m$.
The multiplicities of (j, k, λ) are given as follows:

(a) $p = j_m - k_m + s, \quad s = 0, 1, 2, \cdots, 2k_m$

$$\left.\begin{array}{l} q = j_m - k_m + \gamma, \\ \lambda = j_m - k_m \pm (s + s_m - \gamma - 2t), \\ M\left(j = j_m - k_m + \frac{s}{2} + \frac{\gamma}{2}, k = \frac{s}{2} - \frac{\gamma}{2}, \lambda\right) = \min\{t + 1, 2k_m + s_m - s + 1, s - \gamma + 1\}, \\ \text{where} \quad t = 0, 1, 2, \cdots, \dfrac{s + s_m - \gamma}{2} \qquad (s + s_m - \gamma = \text{even}) \\ \qquad\qquad , \dfrac{s + s_m - \gamma - 1}{2} \quad (s + s_m - \gamma = \text{odd}) \\ \qquad \gamma = 0, 1, 2, \cdots, s. \end{array}\right\} \cdots \Bigg\},$$

$$\left.\begin{array}{l} q = j_m - k_m, \\ \lambda = j_m - k_m \pm (s + s_m - 2t), \\ M\left(j = j_m - k_m + \frac{s}{2}, k = \frac{s}{2}, \lambda\right) = \min\{t + 1, 2k_m + s_m - s + 1, s + 1\}, \\ \text{where} \quad t = 0, 1, 2, \cdots, \dfrac{s + s_m}{2} \qquad (s + s_m = \text{even}) \\ \qquad\qquad , \dfrac{s + s_m - 1}{2} \quad (s + s_m = \text{odd}). \end{array}\right\} \cdots \Bigg\}$$

$$\left.\begin{aligned}
&q = j_m - k_m - \gamma,\\
&\lambda = j_m - k_m - \gamma \pm (s + s_m - 2t),\\
&M\left(j + j_m - k_m + \frac{s}{2} - \frac{\gamma}{2}, k = \frac{s}{2} + \frac{\gamma}{2}, \lambda\right) = \min\{t+1, 2k_m + s_m - s + 1, s+1\},\\
&\text{where}\quad t = 0, 1, 2, \cdots, \frac{s + s_m}{2} \qquad (s + s_m = \text{even})\\
&\qquad\qquad , \frac{s + s_m - 1}{2} \qquad (s + s_m = \text{odd})\\
&\qquad\quad \gamma = 0, 1, 2, \cdots, j_m - k_m \qquad (j_m - k_m = \text{integer})\\
&\qquad\qquad , j_m - k_m - \tfrac{1}{2} \qquad (j_m - k_m = \text{half integer}).
\end{aligned}\right\}. \tag{5.17}$$

Due to symmetry, when $j \rightleftarrows k$, $q \to -q$, $\lambda \to -\lambda$, $M(j, k, \lambda) = M(k, j, -\lambda)$.

(b) $p = j_m + k_m + s, \quad s = 0, 1, 2, \cdots, s_m$

$$\left.\begin{aligned}
&q = j_m - k_m + \gamma,\\
&\lambda = j_m - k_m \pm (2k_m - s + s_m - \gamma - 2t),\\
&M\left(j = j_m + \frac{s}{2} + \frac{\gamma}{2}, k = k_m + \frac{s}{2} - \frac{\gamma}{2}, \lambda\right) = \min\{t+1, 2k_m + s - \gamma + 1, 2k_m + 1, s_m - s + 1\}\\
&\text{where}\quad t = 0, 1, 2, \cdots, \frac{2k_m - s + s_m - \gamma}{2} \qquad (2k_m - s + s_m - \gamma = \text{even})\\
&\qquad\qquad , \frac{2k_m - s + s_m - \gamma - 1}{2} \qquad (2k_m - s + s_m - \gamma = \text{odd})
\end{aligned}\right\}\cdots\Bigg\}$$

$$\gamma = 0, 1, 2, \cdots, 2k_m$$

$$\left.\begin{aligned}
&q = j_m - k_m\\
&\lambda = j_m - k_m \pm (2k_m - s + s_m - 2t)\\
&M\left(j = j_m + \frac{s}{2}, k = k_m + \frac{s}{2}, \lambda\right) = \min\{t+1, 2k_m + 1, s_m - s + 1\}\\
&\text{where}\quad t = 0, 1, 2, \cdots, \frac{2k_m - s + s_m}{2} \qquad (2k_m - s + s_m = \text{even})\\
&\qquad\qquad , \frac{2k_m - s + s_m - 1}{2} \qquad (2k_m - s + s_m = \text{odd})
\end{aligned}\right\}\cdots\Bigg\}$$

$$\left.\begin{aligned}
&q = j_m - k_m - \gamma,\\
&\lambda = j_m - k_m - \gamma \pm (2k_m - s + s_m - 2t),\\
&M\left(j = j_m + \frac{s}{2} - \frac{\gamma}{2}, k = k_m + \frac{s}{2} + \frac{\gamma}{2}, \lambda\right) = \min\{t+1, 2k_m + 1, s_m - s + 1\},\\
&\text{where}\quad t = 0, 1, 2, \cdots, \frac{2k_m - s + s_m}{2} \qquad (2k_m - s + s_m = \text{even})\\
&\qquad\qquad , \frac{2k_m - s + s_m - 1}{2} \qquad (2k_m - s + s_m = \text{odd})\\
&\qquad\quad \gamma = 0, 1, 2, \cdots, j_m - k_m \qquad (j_m - k_m = \text{integer})\\
&\qquad\qquad , j_m - k_m - \tfrac{1}{2} \qquad (j_m - k_m = \text{half integer}).
\end{aligned}\right\}. \tag{5.18}$$

Again, when $j \rightleftarrows k$, $q \to -q$, $\lambda \to -\lambda$. $M(j, k, \lambda) = M(k, j, -\lambda)$.

(ii) $k_m > j_m$, is the same as (5.16)–(5.18) with $j_m \leftrightarrow k_m, q \to -q$.

(iii) $j_m = k_m$ is the same as (5.16)–(5.18) with $j_m = k_m$.

Proof: See Appendix E. We remark that for nondegenerate representations, the multiplicity $M(j, k, \lambda)$ is given by the minimum of four numbers. Three of them are given by [see (5.17) and (5.18)] $p = j + k$, $q = j - k$, $|q| \leq p$, $|q| \leq p_m$);

$$\text{(a) } t + 1, \tag{5.19a}$$

$$\text{(b) } p - p_0 + 1 \quad (|q| \leq p_0, p_0 = |j_m - k_m|), \tag{5.19b}$$

$$p - |q| + 1 = 2j + 1 (k \geq j) \; (|q| \geq p_0), \tag{5.19c}$$

$$= 2k + 1 (j \geq k),$$

$$\text{(c) } p_m + s_m - p + 1. \tag{5.19d}$$

They are all "distances" in the p–λ diagram: (a) for given j, k the allowed λ ranges from $\lambda = \lambda_1$, $\lambda_1 + 2$, $\lambda_1 + 4, \cdots, \lambda_2$, and $2t = \min\{\lambda - \lambda_1, \lambda_2 - \lambda\}$. (b) for $|q| \leq p_0$, $p - p_0$ = distance to the boundary $(p = p_0)$; for $|q| \geq p_0$, $p - |q| =$ distance to $(p = |q|$, the minimal value of p where q first appears). (c) $p_m + s_m - p$ = distance to the boundary $(p = p_m + s_m)$.

Equation (5.19c) gives the *maximal* possible value for $M(j, k, \lambda)$. This is the result of a theorem due to Harish-Chandra[12] and Godement.[13] For completeness, in Appendix E we state this theorem and apply it to our problem.

The fourth number is

$$\text{(d) } p_m - p_0 + 1 = 2j_m + 1 \quad (k_m \geq j_m) \tag{5.19e}$$

$$= 2k_m + 1 \quad (j_m \geq k_m),$$

and is the absolute upper bound for *all* $M(j, k, \lambda)$ in the irreducible representation $((j_m, k_m, \lambda_m))$.

To summarize the above remarks, we have

$$M(j, k, \lambda) = \min\{t + 1, p - p_0 + 1, p - |q| + 1, p_m + s_m - p + 1, p_m - p_0 + 1\}. \tag{5.20}$$

VI. UNITARY IRREDUCIBLE REPRESENTATIONS IN THE DISCRETE SERIES

We now proceed to the study of unitary representations. Let

$$f(j, \mu; k, \nu; \lambda; \alpha) = |j, \mu; k, \nu; \lambda; \alpha\rangle$$

be vectors (states) in a unitary representation. Then, we have

$$(f(j + \tfrac{1}{2}, \mu + \tfrac{1}{2}; k + \tfrac{1}{2}, \nu - \tfrac{1}{2}; \lambda + 1; \beta), P_+ f(j, \mu; k, \nu; \lambda; \alpha)) = (P_- f(j + \tfrac{1}{2}, \mu + \tfrac{1}{2}; k + \tfrac{1}{2}, \nu - \tfrac{1}{2}; \lambda + 1; \beta), f(j, \mu; k, \nu; \lambda; \alpha)) \tag{6.1}$$

and using Eqs. (2.41a) and (2.43b), Eq. (6.1) reduces to

$$a_1(j, k, \lambda)_{\alpha\beta} = -b_4^*(j + \tfrac{1}{2}, k + \tfrac{1}{2}, \lambda + 1)_{\beta\alpha}. \tag{6.2}$$

Using Eq. (2.49), we finally have

$$a_1(j, k, \lambda)_{\alpha\beta} = -a_1^*(j, k, \lambda)_{\alpha\beta}. \tag{6.3}$$

Therefore, $a_1(j, k, \lambda)_{\alpha\beta}$ is purely imaginary. Similarly, we can show that $a_2(j, k, \lambda)_{\alpha\beta}$, and $a_3(j, k, \lambda)_{\alpha\beta}$ are purely real, and $a_4(j, k, \lambda)_{\alpha\beta}$ is purely imaginary; also $b_1(j, k, \lambda)_{\alpha\beta}$ and $b_4(j, k, \lambda)_{\alpha\beta}$ are purely imaginary, and $b_2(j, k, \lambda)_{\alpha\beta}$ and $b_3(j, k, \lambda)_{\alpha\beta}$ are purely real. Hence,

$$a_1^2(j, k, \lambda)_{\alpha\beta}, \; a_4^2(j, k, \lambda)_{\alpha\beta}, \; b_1^2(j, k, \lambda)_{\alpha\beta}, \; b_4^2(j, k, \lambda)_{\alpha\beta} \leq 0 \tag{6.4}$$

and

$$a_2^2(j, k, \lambda)_{\alpha\beta}, \; a_3^2(j, k, \lambda)_{\alpha\beta}, \; b_2^2(j, k, \lambda)_{\alpha\beta}, \; b_3^2(j, k, \lambda)_{\alpha\beta} \geq 0. \tag{6.5}$$

Equations (6.4) and (6.5) are the only conditions imposed on the functions $a_i^2(j, k, \lambda)_{\alpha\beta}$ and $b_i^2(j, k, \lambda)_{\alpha\beta}$ from the unitary nature of the representation. They are sufficient to determine all degenerate unitary irreducible representations. For the nondegenerate unitary irreducible representations, due to the difficulties mentioned in Sec. IV, the general solution of the problem is much more difficult to obtain. In this paper, we study the unitary irreducible representations in the discrete series, where we have been able to classify all degenerate and nondegenerate representations.

First, we state an important theorem due to Harish-Chandra.[14]

Theorem: (*Harish-Chandra*) If the rank of a semisimple noncompact group G is equal to that of its maximal compact subgroup, then for any finite-dimensional representation F of G, there exist two irreducible unitary representations in the discrete series of G which have the same infinitesimal character as F and can be represented over the space of holomorphic functions on G/A, where A is the maximal Abelian subgroup of G.

[12] Harish-Chandra, Trans. Am. Math. Soc. **75**, 185 (1953); **76**, 26 (1954); **76**, 234 (1954).

[13] R. Godement, Trans. Am. Math. Soc. **73**, 496 (1952).

[14] Harish-Chandra, Acta Math. **113**, 241 (1965); **116**, 1 (1966).

In our case $G = SU(2, 2)$, $SU(2) \times SU(2) \times U(1)$ is its maximal compact subgroup. Both groups are of rank 3, and the theorem is applicable. For the finite irreducible representations, we have

$$C_2 = 2j_m(j_m + 1) + 2k_m(k_m + 1) + \lambda_m(\lambda_m + 4), \tag{5.2}$$

$$C_3 = -(\lambda_m + 2)(j_m - k_m)(j_m + k_m + 1), \tag{5.3}$$

$$C_4 = \tfrac{1}{4}[(\lambda_m + 2)^2 - 4j_m(j_m + 1)] \times [(\lambda_m + 2)^2 - 4k_m(k_m + 1)] - (\lambda_m + 2)^2. \tag{5.4}$$

According to Harish-Chandra's theorem, for a given finite irreducible representation $((j_m, k_m, \lambda_m))$, there exist two inequivalent unitary irreducible representations in the discrete series which take on the same eigenvalues for C_2, C_3, and C_4 as given in Eqs. (5.2)–(5.4).

There are actually two discrete series, namely the D^- series, and the D^+ series. The unitary irreducible representations in D^- are semi-infinite dimensional, where all the allowed values of λ are negative. For such a unitary representation in D^-, there exists a state $|j = J_m, k = K_m, \lambda = \Lambda_m\rangle$, (we omit the labels μ, ν, and α), such that

$$a_1(J_m, K_m, \Lambda_m) = a_2(J_m, K_m, \Lambda_m) = a_3(J_m, K_m, \Lambda_m) = a_4(J_m, K_m, \Lambda_m) = 0, \tag{6.6}$$

and all the allowed λ satisfies

$$\lambda \leq \Lambda_m.$$

We see, therefore, that if (Refs. 9 and 10)

$$J_m = k_m, \quad K_m = j_m, \quad \text{and } \Lambda_m = -\lambda_m - 4, \tag{6.7}$$

Eqs. (4.1a)–(4.3a) give us

$$C_2 = 2J_m(J_m + 1) + 2K_m(K_m + 1) + \Lambda_m(\Lambda_m + 4) = 2j_m(j_m + 1) + 2k_m(k_m + 1) + \lambda_m(\lambda_m + 4), \tag{6.8}$$

$$C_3 = -(\Lambda_m + 2)(J_m - K_m)(J_m + K_m + 1) = -(\lambda_m + 2)(j_m - k_m)(j_m + k_m + 1), \tag{6.9}$$

$$C_4 = \tfrac{1}{4}[(\Lambda_m + 2)^2 - 4J_m(J_m + 1)] \times [(\Lambda_m + 2)^2 - 4K_m(K_m + 1)] - (\Lambda_m + 2)^2 = \tfrac{1}{4}[(\lambda_m + 2)^2 - 4j_m(j_m + 1)] \times [(\lambda_m + 2)^2 - 4k_m(k_m + 1)] - (\lambda_m + 2)^2. \tag{6.10}$$

Similarly, we next consider the D^+ series. The unitary representations in D^+ are also semi-infinite dimensional, and all the allowed values of λ are positive. For a given unitary irreducible representation in D^+, the state

$$|j = K_m = j_m, k = J_m = k_m, \lambda = -\Lambda_m = \lambda_m + 4\rangle$$

exists, such that

$$b_1(j_m, k_m, \lambda_m + 4) = b_2(j_m, k_m, \lambda_m + 4) = b_3(j_m, k_m, \lambda_m + 4) = b_4(j_m, k_m, \lambda_m + 4) = 0, \tag{6.11}$$

and all the allowed λ satisy

$$\lambda \geq \lambda_m + 4.$$

We see, therefore, that Eqs. (4.1b), (4.2b), and (4.3b) also give Eqs. (6.8)–(6.10). It is interesting to note the complete symmetry and complementarity between Eq. (6.6) and Eq. (5.5), and also Eq. (6.11) and Eq. (5.1).

By changing $j \leftrightarrow k$, $\lambda \to -\lambda$, we go from a unitary irreducible representation in the D^- series to a unitary irreducible representation in the D^+ series. Therefore, in the following, we need to study only unitary representations in the D^- series.

In contrast to the six boundaries of finite-dimensional representations, (5.13a)–(5.13f), unitary representations in the D^- series have only three boundaries, given by the equations,

$$p + \lambda = J_m + K_m + \Lambda_m, \tag{6.12a}$$

$$p - \lambda = J_m + K_m - \Lambda_m, \tag{6.12b}$$

$$p = p_0 = |J_m - K_m|. \tag{6.12c}$$

When $J_m = 0$, or $K_m = 0$, then the boundary (6.12b) is absent, and we have what will be called degenerate boundaries. We see shortly that degenerate boundaries imply degenerate representations. In fact, the last remark applies also to finite-dimensional representations, where we see that degenerate boundaries imply degenerate Young diagrams which imply degenerate representations. (see Fig. 2).

Equation (5.11) gives us *all* the finite representations, and through Eq. (6.7) therefore, gives us *almost*[15] *all* the unitary representations in the D^- series. On the boundaries (6.12a) and (6.12b), we have, similar to Eqs. (5.8)–(5.9),

$$a_1^2\left(J_m - \frac{t}{2}, K_m - \frac{t}{2}, \Lambda_m - t\right) = \frac{t(\Lambda_m + J_m + K_m - t + 3)}{(2J_m - t + 2)(2K_m - t + 2)} = -\frac{t(t + s_m + 1)}{(2J_m - t + 2)(2K_m - t + 2)} \leq 0,$$

$$t = 0, 1, 2, \cdots, t_m = 2 \times \min(J_m, K_m), \tag{6.13a}$$

[15] In Theorem 3, (a), $s_m = 0, 1, 2, 3, \cdots$. However, from a study in Ref. 9, $s_m = -2, -1$ are also allowed.

and

$$b_4^2\left(J_m-\frac{t}{2},K_m-\frac{t}{2},\Lambda_m-t\right)$$
$$=-\frac{(t+1)(t+s_m+2)}{(2J_m-t+1)(2K_m-t+1)}<0,$$
$$t=0,1,2,\cdots,t_m-1,$$
$$b_4^2\left(J_m-\frac{t_m}{2},K_m-\frac{t_m}{2},\Lambda_m-t_m\right)=0; \tag{6.13b}$$
$$a_4^2\left(J_m+\frac{s}{2},K_m+\frac{s}{2},\Lambda_m-s\right)$$
$$=\frac{s(\Lambda_m-J_m-K_m-s+1)}{(2J_m+s)(2K_m+s)}$$
$$=-\frac{s(2J_m+2K_m+s_m+s+3)}{(2J_m+s)(2K_m+s)}\leq 0, \tag{6.14a}$$

and

$$b_1^2\left(J_m+\frac{s}{2},K_m+\frac{s}{2},\Lambda_m-s\right)$$
$$=-\frac{(s+1)(2J_m+2K_m+s_m+s+4)}{(2J_m+s+1)(2K_m+s+1)}<0,$$
$$s=0,1,2,\cdots. \tag{6.14b}$$

We observe that Eqs. (6.13) and (6.14) are all negative, satisfying Eq. (6.4). We further notice that Eq. (6.14b) never vanishes, and s increases without bound, in contrast to Eq. (5.9b), where s has the upper bound s_m. The infinite-dimensional nature of the unitary representations is therefore obvious. Consequently, in the decomposition of a unitary irreducible representation, $M(j,k,\lambda)$ may *a priori* be finite or infinite. But from a theorem of Harish-Chandra and Godement (see Appendix E), $M(j,k,\lambda)\leq 2\times\min(j,k)+1$ and is always finite.

From now on, a unitary irreducible representation in the D^- series which satisfies the conditions Eqs. (6.6)–(6.10) will be denoted by $D^-(J_m,K_m,\Lambda_m)$. The decomposition of $D^-(J_m,K_m,\Lambda_m)$ into an infinite sum of (j,k,λ) with multiplicity $M(j,k,\lambda)$ is now considered, and the results are stated in the following theorem.

Theorem 3: The decomposition of an unitary irreducible representation $D^-(J_m,K_m,\Lambda_m)$ in the discrete series of $SU(2,2)$ into unitary irreducible representations (j,k,λ) of the maximal compact subgroup $SU(2)\times SU(2)\times U(1)$ with multiplicity $M(j,k,\lambda)$ is given as follows:

Let $p=j+k$, $q=j-k$, $p_m=J_m+K_m$.

(*a*) *Degenerate representations:*

$M(j,k,\lambda)=1$ for all allowed values of j,k,λ.

$$\Lambda_m=-J_m-K_m-s_m-4,$$

where $s_m=0,1,2,3,\cdots$, (from Part II we also have $s_m=-2,-1$).

(i) $J_m=0$, $K_m=\frac{1}{2},1,\frac{3}{2},\cdots$

$$\left.\begin{aligned}p&=p_m\\q&=-p_m+\gamma\\\lambda&=\Lambda_m-\gamma-2t\end{aligned}\right\}\left.\begin{aligned}p&=p_m+1\\q&=-p_m+\gamma\\\lambda&=\Lambda_m-\gamma-2t-1\end{aligned}\right\}\cdots\Bigg\}$$
$$\left.\begin{aligned}p&=p_m+s\\q&=-p_m+\gamma\\\lambda&=\Lambda_m-\gamma-2t-s\end{aligned}\right\}\cdots, \tag{6.15}$$

where $0\leq\gamma\leq 2p_m$, $s,t=0,1,2,\cdots$.

(ii) $K_m=0$, $J_m=\frac{1}{2},1,\frac{3}{2},\cdots$

$$\left.\begin{aligned}p&=p_m\\q&=p_m-\gamma\\\lambda&=\Lambda_m-\gamma-2t\end{aligned}\right\}\left.\begin{aligned}p&=p_m+1\\q&=p_m-\gamma\\\lambda&=\Lambda_m-\gamma-2t-1\end{aligned}\right\}\cdots\Bigg\}$$
$$\left.\begin{aligned}p&=p_m+s\\q&=p_m-\gamma\\\lambda&=\Lambda_m-\gamma-2t-s\end{aligned}\right\}\cdots,$$

where $0\leq\gamma\leq 2p_m$, $s,t=0,1,2,\cdots$.

(iii) $J_m=K_m=0$

$$\left.\begin{aligned}p&=0\\q&=0\\\lambda&=\Lambda_m-2t\end{aligned}\right\}\left.\begin{aligned}p&=1\\q&=0\\\lambda&=\Lambda_m-2t-1\end{aligned}\right\}\cdots\Bigg\}$$
$$\left.\begin{aligned}p&=s\\q&=0\\\lambda&=\Lambda_m-2t-s\end{aligned}\right\}\cdots,$$

where $s,t=0,1,2,\cdots$.

(*b*) *Nondegenerate representations:*

$M(j,k,\lambda)\geq 1$ for all allowed values of j,k,λ.

$$\Lambda_m=-J_m-K_m-s_m-4,$$

where $s_m=0,1,2,3,\cdots$ [from Eq. (6.13ab) we also have $s_m=-1$].

(i) $J_m \geq K_m$, $K_m = \frac{1}{2}, 1, \frac{3}{2}, \cdots$

(a) $p = J_m - K_m + s$, $s = 0, 1, 2, \cdots, 2K_m$,

$$\left.\begin{aligned} &q = J_m - K_m + \gamma, \\ &\lambda = \Lambda_m - 2K_m + s - \gamma - 2t, \\ &M(j,k,\lambda) = \min\{t+1, 2K_m + t - s + 1, \\ &\qquad s - \gamma + 1\}, \\ &\text{where} \\ &t = 0, 1, 2, 3, \cdots, \\ &\gamma = 0, 1, 2, \cdots, s. \end{aligned}\right\} \cdots\Bigg\},$$

$$\left.\begin{aligned} &q = J_m - K_m, \\ &\lambda = \Lambda_m - 2K_m + s - 2t, \\ &M(j,k,\lambda) = \min\{t+1, 2K_m + t - s + 1, \\ &\qquad s + 1\} \\ &\text{where} \\ &t = 0, 1, 2, 3, \cdots. \end{aligned}\right\} \cdots\Bigg\},$$

$$\left.\begin{aligned} &q = J_m - K_m - \gamma, \\ &\lambda = \Lambda_m - 2K_m + s - \gamma - 2t, \\ &M(j,k,\lambda) = \min\{t+1, 2K_m + t - s + 1, \\ &\qquad s + 1, 2J_m - 2K_m + s - \gamma + 1\} \\ &\text{where} \\ &t = 0, 1, 2, 3, \cdots, \\ &\gamma = 0, 1, 2, \cdots, 2(J_m - K_m) + s. \end{aligned}\right\},$$

(b) $p = J_m + K_m + s$, $s = 0, 1, 2, 3, \cdots$,

$$\left.\begin{aligned} &q = J_m - K_m + \gamma, \\ &\lambda = \Lambda_m - s - \gamma - 2t, \\ &M(j,k,\lambda) = \min\{t+1, 2K_m + s - \gamma + 1, \\ &\qquad 2K_m + 1\}, \\ &\text{where} \\ &t = 0, 1, 2, \cdots, \\ &\gamma = 0, 1, 2, \cdots, 2K_m. \end{aligned}\right\} \cdots\Bigg\},$$

$$\left.\begin{aligned} &q = J_m - K_m, \\ &\lambda = \Lambda_m - s - 2t, \\ &M(j,k,\lambda) = \min\{t+1, 2K_m + 1\}, \\ &\text{where} \\ &t = 0, 1, 2, \cdots. \end{aligned}\right\} \cdots\Bigg\},$$

$$\left.\begin{aligned} &q = J_m - K_m - \gamma, \\ &\lambda = \Lambda_m - s - \gamma - 2t, \\ &M(j,k,\lambda) = \min\{t+1, 2K_m + 1\}, \\ &\text{where} \\ &t = 0, 1, 2, \cdots, \\ &\gamma = 0, 1, 2, \cdots, 2J_m. \end{aligned}\right\}. \quad (6.16)$$

(ii) $K_m > J_m$, $J_m = \frac{1}{2}, 1, \frac{3}{2}, \cdots$ is the same as (i) with the change $J_m \leftrightarrow K_m$, $q \to -q$.

Proof: For clarity of presentation the proof for degenerate representations will be given in Paper II of this series, and that for nondegenerate representations in Paper III.

Similarly, for the D^+ series, we immediately have

Theorem 4: The decomposition of an unitary irreducible representation $D^+(K_m, J_m, -\Lambda_m)$, defined in Eq. (6.11), in the discrete series of $SU(2,2)$ into unitary irreducible representations (j,k,λ) of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$ with multiplicity $M(j,k,\lambda)$ is given by the results from Theorem 3 with the change $j \leftrightarrow k$ $(q \to -q)$ and $\lambda \to -\lambda$.

ACKNOWLEDGMENT

This paper was started while the author was visiting the Laboratory for Nuclear Science, Massachusetts Institute of Technology. He wishes to thank Professor Francis E. Low for his interest in the subject and for his hospitality.

APPENDIX A. RELATIONS AMONG $a_i(j,k,\lambda)$ AND $b_i(j,k,\lambda)$

From the commutation relations (2.26)–(2.31), we obtain the following relations:

$$(k+1)\sum_\rho a_1(j,k,\lambda)_{\alpha\rho} a_2(j+\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta} = k\sum_\rho a_2(j,k,\lambda)_{\alpha\rho} a_1(j+\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta}, \quad \text{(A1a)}$$

$$(j+1)\sum_\rho a_1(j,k,\lambda)_{\alpha\rho} a_3(j+\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta} = j\sum_\rho a_3(j,k,\lambda)_{\alpha\rho} a_1(j-\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta}, \quad \text{(A2a)}$$

$$(j+1)\sum_\rho a_2(j,k,\lambda)_{\alpha\rho} a_4(j+\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta} = j\sum_\rho a_4(j,k,\lambda)_{\alpha\rho} a_2(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta}, \quad \text{(A3a)}$$

$$(k+1)\sum_\rho a_3(j,k,\lambda)_{\alpha\rho} a_4(j-\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta} = k\sum_\rho a_4(j,k,\lambda)_{\alpha\rho} a_3(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta}, \quad \text{(A4a)}$$

$$(j+1)\sum_\rho \{a_1(j,k,\lambda)_{\alpha\rho} a_4(j+\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta} + a_2(j,k,\lambda)_{\alpha\rho} a_3(j+\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta}\} = j\sum_\rho \{a_3(j,k,\lambda)_{\alpha\rho} a_2(j-\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta} + a_4(j,k,\lambda)_{\alpha\rho} a_1(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta}\}, \quad \text{(A5a)}$$

$$(k+1)\sum_\rho \{a_1(j,k,\lambda)_{\alpha\rho} a_4(j+\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta} + a_3(j,k,\lambda)_{\alpha\rho} a_2(j-\tfrac{1}{2}, k+\tfrac{1}{2}, \lambda+1)_{\rho\beta}\} = k\sum_\rho \{a_2(j,k,\lambda)_{\alpha\rho} a_3(j+\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta} + a_4(j,k,\lambda)_{\alpha\rho} a_1(j-\tfrac{1}{2}, k-\tfrac{1}{2}, \lambda+1)_{\rho\beta}\}. \quad \text{(A6a)}$$

Similarly, we have

$$(k+1)\sum_{\rho} b_1(j,k,\lambda)_{\alpha\rho}b_2(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta} = k\sum_{\rho} b_2(j,k,\lambda)_{\alpha\rho}b_1(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}. \quad \text{(A1b)}$$

Equations (A2b)–(A6b) follow in the same manner.

We also have the following relations among the $a_i(j,k,\lambda)$ and $b_i(j,k,\lambda)$:

$$\sum_{\rho} a_1(j,k,\lambda)_{\alpha\rho}b_1(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} = \sum_{\rho} b_1(j,k,\lambda)_{\alpha\rho}a_1(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}, \quad \text{(A7)}$$

$$\sum_{\rho} a_2(j,k,\lambda)_{\alpha\rho}b_2(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} = \sum_{\rho} b_2(j,k,\lambda)_{\alpha\rho}a_2(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}, \quad \text{(A8)}$$

$$\sum_{\rho} a_3(j,k,\lambda)_{\alpha\rho}b_3(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} = \sum_{\rho} b_3(j,k,\lambda)_{\alpha\rho}a_3(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}, \quad \text{(A9)}$$

$$\sum_{\rho} a_4(j,k,\lambda)_{\alpha\rho}b_4(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} = \sum_{\rho} b_4(j,k,\lambda)_{\alpha\rho}a_4(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}, \quad \text{(A10)}$$

$$\sum_{\rho}\{a_1(j,k,\lambda)_{\alpha\rho}b_2(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_2(j,k,\lambda)_{\alpha\rho}b_1(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta}\} = \sum_{\rho}\{b_1(j,k,\lambda)_{\alpha\rho}a_2(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_2(j,k,\lambda)_{\alpha\rho}a_1(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\}, \quad \text{(A11)}$$

$$\sum_{\rho}\{a_3(j,k,\lambda)_{\alpha\rho}b_4(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_4(j,k,\lambda)_{\alpha\rho}b_3(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta}\} = \sum_{\rho}\{b_3(j,k,\lambda)_{\alpha\rho}a_4(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_4(j,k,\lambda)_{\alpha\rho}a_3(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\}, \quad \text{(A12)}$$

$$\sum_{\rho}\{a_1(j,k,\lambda)_{\alpha\rho}b_3(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_3(j,k,\lambda)_{\alpha\rho}b_1(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} = \sum_{\rho}\{b_1(j,k,\lambda)_{\alpha\rho}a_3(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_3(j,k,\lambda)_{\alpha\rho}a_1(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}\}, \quad \text{(A13)}$$

$$\sum_{\rho}\{a_2(j,k,\lambda)_{\alpha\rho}b_4(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_4(j,k,\lambda)_{\alpha\rho}b_2(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta}\} = \sum_{\rho}\{b_2(j,k,\lambda)_{\alpha\rho}a_4(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_4(j,k,\lambda)_{\alpha\rho}a_2(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\}, \quad \text{(A14)}$$

$$(k+1)\sum_{\rho}\{a_1(j,k,\lambda)_{\alpha\rho}b_2(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_1(j,k,\lambda)_{\alpha\rho}a_2(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}\} = k\sum_{\rho}\{a_2(j,k,\lambda)_{\alpha\rho}b_1(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_2(j,k,\lambda)_{\alpha\rho}a_1(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\}, \quad \text{(A15)}$$

$$(k+1)\sum_{\rho}\{a_3(j,k,\lambda)_{\alpha\rho}b_4(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_3(j,k,\lambda)_{\alpha\rho}a_4(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}\} = k\sum_{\rho}\{a_4(j,k,\lambda)_{\alpha\rho}b_3(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_4(j,k,\lambda)_{\alpha\rho}a_3(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\}, \quad \text{(A16)}$$

$$(j+1)\sum_{\rho}\{a_1(j,k,\lambda)_{\alpha\rho}b_3(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_1(j,k,\lambda)_{\alpha\rho}a_3(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}\} = j\sum_{\rho}\{a_3(j,k,\lambda)_{\alpha\rho}b_1(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_3(j,k,\lambda)_{\alpha\rho}a_1(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}\}, \quad \text{(A17)}$$

$$(j+1)\sum_{\rho}\{a_2(j,k,\lambda)_{\alpha\rho}b_4(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_2(j,k,\lambda)_{\alpha\rho}a_4(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\} = j\sum_{\rho}\{a_4(j,k,\lambda)_{\alpha\rho}b_2(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_4(j,k,\lambda)_{\alpha\rho}a_2(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\}. \quad \text{(A18)}$$

Equations (A7)–(A18) are all homogeneous relations. Now, we also have two inhomogeneous relations,

$$(j+1)\sum_{\rho}\{a_1(j,k,\lambda)_{\alpha\rho}b_4(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_2(j,k,\lambda)_{\alpha\rho}b_3(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_1(j,k,\lambda)_{\alpha\rho}a_4(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_2(j,k,\lambda)_{\alpha\rho}a_3(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\} - j\sum_{\rho}\{a_3(j,k,\lambda)_{\alpha\rho}b_2(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_4(j,k,\lambda)_{\alpha\rho}b_1(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_3(j,k,\lambda)_{\alpha\rho}a_2(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_4(j,k,\lambda)_{\alpha\rho}a_1(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\} + \delta_{\alpha\beta} = 0, \quad \text{(A19)}$$

$$(k+1)\sum_{\rho}\{a_1(j,k,\lambda)_{\alpha\rho}b_4(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_3(j,k,\lambda)_{\alpha\rho}b_2(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_1(j,k,\lambda)_{\alpha\rho}a_4(j+\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_3(j,k,\lambda)_{\alpha\rho}a_2(j-\tfrac{1}{2},k+\tfrac{1}{2},\lambda-1)_{\rho\beta}\} - k\sum_{\rho}\{a_2(j,k,\lambda)_{\alpha\rho}b_3(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + a_4(j,k,\lambda)_{\alpha\rho}b_1(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda+1)_{\rho\beta} + b_2(j,k,\lambda)_{\alpha\rho}a_3(j+\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta} + b_4(j,k,\lambda)_{\alpha\rho}a_1(j-\tfrac{1}{2},k-\tfrac{1}{2},\lambda-1)_{\rho\beta}\} + \delta_{\alpha\beta} = 0. \quad \text{(A20)}$$

Finally, we have the relation

$$\sum_{i=1}^{4}\sum_{\rho} a_i(j,k,\lambda)_{\alpha\rho}a_i(j,k,\lambda)_{\beta\rho} = \sum_{i=1}^{4}\sum_{\rho} b_i(j,k,\lambda)_{\alpha\rho}b_i(j,k,\lambda)_{\beta\rho}. \quad \text{(A21)}$$

APPENDIX B. OFF-DIAGONAL ELEMENTS OF C_2, C_3, C_4, AND F_3

For $\alpha \neq \beta$, we have from Eqs. (3.1)–(3.3), and (3.6)

$$\sum_\rho \{(j+1)(k+1)a_1(j.\,k,\lambda)_{\alpha\rho}a_1(j,k,\lambda)_{\beta\rho} - (j+1)ka_2(j,k,\lambda)_{\alpha\rho}a_2(j,k,\lambda)_{\beta\rho} - j(k+1)a_3(j,k,\lambda)_{\alpha\rho}a_3(j,k,\lambda)_{\beta\rho} + jka_4(j,k,\lambda)_{\alpha\rho}a_4(j,k,\lambda)_{\beta\rho}\} = 0, \quad \text{(B1)}$$

$$\sum_\rho \{(j+1)(k+1)(j-k)a_1(j,k,\lambda)_{\alpha\rho}a_1(j,k,\lambda)_{\beta\rho} - (j+1)k(j+k+1)a_2(j,k,\lambda)_{\alpha\rho}a_2(j,k,\lambda)_{\beta\rho} + j(k+1)(j+k+1)a_3(j,k,\lambda)_{\alpha\rho}a_3(j,k,\lambda)_{\beta\rho} - jk(j-k)a_4(j,k,\lambda)_{\alpha\rho}a_4(j,k,\lambda)_{\beta\rho}\} = 0, \quad \text{(B2)}$$

$$jk(j+1)(k+1)\sum_{i=1}^{4}\sum_\rho a_i(j,k,\lambda)_{\alpha\rho}a_i(j,k,\lambda)_{\beta\rho} = -\tfrac{1}{4}\{(P_-Q_- - S_-T_-)(P_+Q_+ - S_+T_+)\}_{\alpha\beta}, \quad \text{(B3)}$$

where $\{(P_-Q_- - S_-T_-)(P_+Q_+ - S_+T_+)\}_{\alpha\beta}$ is given by Eq. (4.4). Finally,

$$\sum_\rho \{(j+1)(k+1)(j+k)a_1(j,k,\lambda)_{\alpha\rho}a_1(j,k,\lambda)_{\beta\rho} - (j+1)k(j-k-1)a_2(j,k,\lambda)_{\alpha\rho}a_2(j,k,\lambda)_{\beta\rho} + j(k+1)(j-k+1)a_3(j,k,\lambda)_{\alpha\rho}a_3(j,k,\lambda)_{\beta\rho} - jk(j+k+2)a_4(j,k,\lambda)_{\alpha\rho}a_4(j,k,\lambda)_{\beta\rho}\} = 0. \quad \text{(B4)}$$

Similarly, we can write identical expressions for the $b_i(j,k,\lambda)$.

APPENDIX C. COMMUTATION RELATIONS BETWEEN F_3 AND $P_\pm$, $Q_\pm$, $S_\pm$, AND $T_\pm$

$$[F_3,P_+] = \{-\tfrac{1}{2}C_2 + \tfrac{1}{2}(R_0-2)(R_0+3) + (\mathbf{J}^2+\mathbf{K}^2) - 2(J_3-\tfrac{3}{2})(K_3+\tfrac{3}{2})\}P_+ - 2(J_3-\tfrac{3}{2})K_-S_+ - 2(K_3+\tfrac{3}{2})J_+T_+ - 2K_-J_+Q_+ + 2Q_-(S_+T_+ - P_+Q_+), \quad \text{(C1)}$$

$$[F_3,P_-] = \{\tfrac{1}{2}C_2 - \tfrac{1}{2}(R_0+2)(R_0-3) - (\mathbf{J}^2+\mathbf{K}^2) + 2(J_3+\tfrac{3}{2})(K_3-\tfrac{3}{2})\}P_- + 2(J_3+\tfrac{3}{2})K_+S_- + 2(K_3-\tfrac{3}{2})J_-T_- + 2J_-K_+Q_- - 2Q_+(S_-T_- - P_-Q_-). \quad \text{(C2)}$$

The other six commutation relations can similarly be written down. However, they provide no new information, and are omitted here.

The matrix elements of the commutation relations (C1) and (C2) are very important. They will play a vital role in the discussion of Paper III. Equation (C1) gives us four relations,

$$a_1(j,k,\lambda)_{\alpha\beta}\{\beta-\alpha+\tfrac{1}{2}C_2-\tfrac{1}{2}\lambda(\lambda+3) - (j+k+1)^2+\tfrac{3}{2}\} = -2\sum_\sigma (P_+Q_+ - S_+T_+)_{\alpha\sigma}b_1(j,k,\lambda+2)_{\sigma\beta}, \quad \text{(C3)}$$

$$a_2(j,k,\lambda)_{\alpha\beta}\{\beta-\alpha+\tfrac{1}{2}C_2-\tfrac{1}{2}\lambda(\lambda+3) - (j-k)^2+\tfrac{3}{2}\} = -2\sum_\sigma (P_+Q_+ - S_+T_+)_{\alpha\sigma}b_2(j,k,\lambda+2)_{\sigma\beta}, \quad \text{(C4)}$$

$$a_3(j,k,\lambda)_{\alpha\beta}\{\beta-\alpha+\tfrac{1}{2}C_2-\tfrac{1}{2}\lambda(\lambda+3) - (j-k)^2+\tfrac{3}{2}\} = -2\sum_\sigma (P_+Q_+ - S_+T_+)_{\alpha\sigma}b_3(j,k,\lambda+2)_{\sigma\beta}, \quad \text{(C5)}$$

$$a_4(j,k,\lambda)_{\alpha\beta}\{\beta-\alpha+\tfrac{1}{2}C_2-\tfrac{1}{2}\lambda(\lambda+3) - (j+k+1)^2+\tfrac{3}{2}\} = -2\sum_\sigma (P_+Q_+ - S_+T_+)_{\alpha\sigma}b_4(j,k,\lambda+2)_{\sigma\beta}. \quad \text{(C6)}$$

Similarly, Eq. (C2) gives us

$$b_1(j,k,\lambda)_{\alpha\gamma}\{\gamma-\alpha-\tfrac{1}{2}C_2+\tfrac{1}{2}\lambda(\lambda-3) + (j+k+1)^2-\tfrac{3}{2}\} = 2\sum_\tau (P_-Q_- - S_-T_-)_{\alpha\tau}a_1(j,k,\lambda-2)_{\tau\gamma}, \quad \text{(C7)}$$

$$b_2(j,k,\lambda)_{\alpha\gamma}\{\gamma-\alpha-\tfrac{1}{2}C_2+\tfrac{1}{2}\lambda(\lambda-3) + (j-k)^2-\tfrac{3}{2}\} = 2\sum_\tau (P_-Q_- - S_-T_-)_{\alpha\tau}a_2(j,k,\lambda-2)_{\tau\gamma}, \quad \text{(C8)}$$

$$b_3(j,k,\lambda)_{\alpha\gamma}\{\gamma-\alpha-\tfrac{1}{2}C_2+\tfrac{1}{2}\lambda(\lambda-3) + (j-k)^2-\tfrac{3}{2}\} = 2\sum_\tau (P_-Q_- - S_-T_-)_{\alpha\tau}a_3(j,k,\lambda-2)_{\tau\gamma}, \quad \text{(C9)}$$

$$b_4(j,k,\lambda)_{\alpha\gamma}\{\gamma-\alpha-\tfrac{1}{2}C_2+\tfrac{1}{2}\lambda(\lambda-3) + (j+k+1)^2-\tfrac{3}{2}\} = 2\sum_\tau (P_-Q_- - S_-T_-)_{\alpha\tau}a_4(j,k,\lambda-2)_{\tau\gamma}. \quad \text{(C10)}$$

We observe that Eq. (C3) is similar in structure with Eq. (C6); also (C4) with (C5), (C7) with (C10), and (C8) with (C9). These properties will be utilized extensively in Paper III.

APPENDIX D. PROOFS FOR THEOREMS 1 AND 2

The relation between a Young diagram and j_m, k_m, s_m has already been mentioned in the text following Theorem 1. There we see that the state of maximum weight $|j_m,k_m,\lambda_m\rangle$ exists such that Eq. (5.1) holds and $\lambda_m = j_m + k_m + s_m$, which is Eq. (5.10). The three classes of finite representations listed in Theorem 1 are the results, when we enumerate all the possible values of j_m, k_m, and s_m. The problem of degeneracy and multiplicity is the content of Theorem 2, which we discuss next.

The discussion and proof of Theorem 2 is much facilitated if we introduce the idea of a p–λ diagram.

Class III:

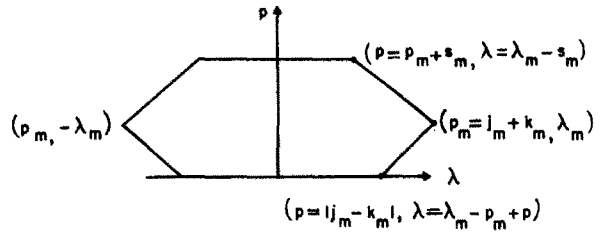


FIG. 1. Boundaries of a finite irreducible representation in the p–λ plane.

It is a diagram giving the boundaries of a finite irreducible representation in the p–λ plane. In the decomposition of a finite representation $((j_m, k_m, \lambda_m))$ of $SU(2, 2)$ into unitary irreducible representations (j, k, λ) of the maximal compact subgroup $SU(2) \times SU(2) \times U(1)$, the allowed values of j, k, λ are such that $p = j + k$ and λ are all *inside* or *on* the boundaries. The most general p–λ diagram is that of Class III, where $j_m \neq 0$, $k_m \neq 0$, $s_m \neq 0$.

The p–λ diagrams associated with Class I and Class II of the finite representations are the degenerate forms of Fig. 1, and are given in Fig. 2.

The p–λ diagrams in Figs. 1 and 2 are the diagrammatic representations of Eqs. (5.13a)–(5.13f). The allowed (j, k, λ) of $SU(2) \times SU(2) \times U(1)$ in $((j_m, k_m, \lambda_m))$ of $SU(2, 2)$ can now be written down immediately, the results are stated in a compact form in Eqs. (5.14)–(5.16). The derivation of this part is quite simple, and involves only straightforward enumeration of all the states (j, k, λ).

Class I: $s_m \neq 0$
(1) $j_m = 0, k_m \neq 0$
(2) $k_m = 0, j_m \neq 0$
(3) $j_m = k_m = 0$

Class II: $s_m = 0$
(1) $j_m \neq 0, k_m \neq 0, j_m \neq k_m$
(2) $j_m = k_m \neq 0$
(3) $j_m \neq 0, k_m = 0$
or
$j_m = 0, k_m \neq 0$

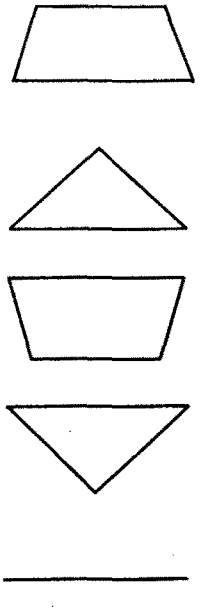


FIG. 2. Degenerate forms of Fig. 1.

The problem of multiplicity is much more complicated. For degenerate representations, we put $M(j, k, \lambda) = 1$ in the coupled recursion relations (4.12) and (4.13), and solve for the $a_i(j, k, \lambda)$ and $b_i(j, k, \lambda)$. Then, we demand that these solutions satisfy the appropriate conditions on the boundaries Eqs. (5.6), (5.7), etc. In this way, we find that the degenerate representations have only the allowed p–λ diagrams given in Fig. 2. In this connection, we remark that representations of Class II require more caution on our part in the study of Eqs. (4.12) and (4.13), since the boundaries (5.13b) and (5.13f) "intrude" there.

Since the only type of p–λ diagram left now is that of Fig. 1, we know it must be nondegenerate. Now we have the coupled diophantine recursion relations (4.12) and (4.13) to solve, where $M(j, k, \lambda)$ are unknown integers. However, from a detailed study of Eqs. (C3)–(C10), which will be presented in Paper III in our discussion of nondegenerate unitary irreducible representations, we have a knowledge of the multiplicity $M(j, k, \lambda)$ of the states (j, k, λ). Once $M(j, k, \lambda)$ is known, Eqs. (4.12) and (4.13) can be solved for $\sum_{\alpha,\rho} a_i(j, k, \lambda)_{\alpha\rho}$ and $\sum_{\alpha,\rho} b_i(j, k, \lambda)_{\alpha\rho}$, which can be shown to satisfy the appropriate conditions on the boundaries Eqs. (5.6), (5.7), etc. Again the complication, due to the "intrusion" of the boundaries (5.13b) and (5.13f), has to be noted.

So far, only a sketch of the proof to Theorems 1 and 2 has been indicated. A detailed presentation of the algebraic steps involved would occupy several pages and is therefore omitted. However, in Papers II and III when we discuss the unitary irreducible representations in the discrete series, we shall present the relevant intermediate steps in the proof which is in every detail parallel that for the finite representations.

We conclude this appendix by giving the p–λ diagram for the unitary irreducible representations in the discrete series.

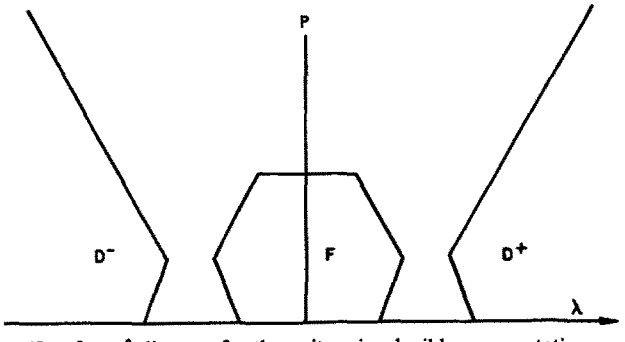


FIG. 3. p–λ diagram for the unitary irreducible representations in the discrete series.

APPENDIX E. THEOREM OF HARISH-CHANDRA AND GODEMENT

Let G be the noncompact group under consideration, and g its Lie algebra.[16] The Cartan decomposition gives $g = k + p$, where k is the maximal compact subalgebra of g and p, a vector subspace. Let a denote a maximal Abelian subspace of p which can be contained in a Cartan subalgebra h of g. The Iwasawa decomposition now gives $g = k + a + n$, where n is a nilpotent subalgebra of g. Let the corresponding subgroups of G be K, A, and N, then $G = KAN$, in the sense that every element of G can be uniquely decomposed. Finally, let m be the centralizer of a in k, $m = h \cap k$. The corresponding subgroup M of K consists of elements which commute with A. Now denote D as an irreducible representation of K, and let $N(D)$ = maximum number of times an irreducible representation of M occurs in restricting the representation D to M.

Theorem: If $x \to T_x$ is an irreducible representation of G, then D occurs at most $N(D)$ times in the restriction of T to K.

We now apply this important theorem to the group $SU(2, 2)$ under consideration. Here $G = SU(2, 2)$, $K = SU(2) \times SU(2) \times U(1)$. The Abelian subalgebra a is two dimensional, and we may choose P_+ and Q_+ as the two elements of this algebra, $[P_+, Q_+] = 0$. The Cartan subalgebra h now contains the three elements P_+, Q_+, and $J_3 + K_3$; the subalgebra m is one dimensional and consists of $J_3 + K_3$ only.

We now illustrate the application of the theorem of Harish-Chandra and Godement with an example. Let $D = (j, k, \lambda)$ be an irreducible representation of K, then $N(D) = 2 \times \min(j, k) + 1$ is the number of times the irreducible representation $J_3 + K_3 = |j - k|$ occurs in (j, k, λ). The theorem now tells us that (j, k, λ) occurs in an irreducible representation of G not more than $N(D)$ times.

[16] For general reference on noncompact groups and Cartan and Iwasawa decomposition see: S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962). R. Hermann, *Lie Groups for Physicists* (W. A. Benjamin, Inc., New York, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Representation Theory of Local Current Algebras*

ERIC H. ROFFMAN
Institute for Advanced Study, Princeton, New Jersey

(Received 27 February 1967)

We prove the following theorem: Every finite-dimensional irreducible representation of the local current algebra

$$[F_\rho(\mathbf{k}_1), F_\sigma(\mathbf{k}_2)] = e^\tau_{\rho\sigma} F_\tau(\mathbf{k}_1 + \mathbf{k}_2),$$

where the $\{e^\tau_{\rho\sigma}\}$ are the structure constants of a semisimple Lie algebra L, has the form

$$F_\rho(\mathbf{k}) \to \sum_{i=1}^{N} e^{i\mathbf{k}x_i} T^i_\rho,$$

where T^i_ρ is a finite-dimensional representation of L, and for $i \neq j$, $x_i \neq x_j$; T^i_ρ commutes with T^j_σ for all ρ and σ.

IF the local current algebras are assumed to satisfy the commutation relations

$$[F_\rho(\mathbf{k}_1), F_\sigma(\mathbf{k}_2)] = e^\tau_{\rho\sigma} F_\tau(\mathbf{k}_1 + \mathbf{k}_2), \tag{1}$$

where the $\{e^\tau_{\rho\sigma}\}$ are the structure constants of a semisimple Lie algebra L, then it is of interest to study the representations of the system (1) consistent with Lorentz invariance. It may be hoped, and there is reason to believe, that the space of states of elementary particles is closely related to the space of such a representation.[1]

We do not consider Lorentz invariance, which may

* Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR NR 42-65.

[1] See R. F. Dashen and M. Gell-Mann, Phys. Rev. Letters 17, 340 (1966); also M. Gell-Mann, "Relativistic Quark Model as Representation of Current Algebra," (lecture notes from International School of Physics "Ettore Majorana"), (to be published).

APPENDIX E. THEOREM OF HARISH-CHANDRA AND GODEMENT

Let G be the noncompact group under consideration, and g its Lie algebra.[16] The Cartan decomposition gives $g = k + p$, where k is the maximal compact subalgebra of g and p, a vector subspace. Let a denote a maximal Abelian subspace of p which can be contained in a Cartan subalgebra h of g. The Iwasawa decomposition now gives $g = k + a + n$, where n is a nilpotent subalgebra of g. Let the corresponding subgroups of G be K, A, and N, then $G = KAN$, in the sense that every element of G can be uniquely decomposed. Finally, let m be the centralizer of a in k, $m = h \cap k$. The corresponding subgroup M of K consists of elements which commute with A. Now denote D as an irreducible representation of K, and let $N(D)$ = maximum number of times an irreducible representation of M occurs in restricting the representation D to M.

Theorem: If $x \to T_x$ is an irreducible representation of G, then D occurs at most $N(D)$ times in the restriction of T to K.

We now apply this important theorem to the group $SU(2, 2)$ under consideration. Here $G = SU(2, 2)$, $K = SU(2) \times SU(2) \times U(1)$. The Abelian subalgebra a is two dimensional, and we may choose P_+ and Q_+ as the two elements of this algebra, $[P_+, Q_+] = 0$. The Cartan subalgebra h now contains the three elements P_+, Q_+, and $J_3 + K_3$; the subalgebra m is one dimensional and consists of $J_3 + K_3$ only.

We now illustrate the application of the theorem of Harish-Chandra and Godement with an example. Let $D = (j, k, \lambda)$ be an irreducible representation of K, then $N(D) = 2 \times \min(j, k) + 1$ is the number of times the irreducible representation $J_3 + K_3 = |j - k|$ occurs in (j, k, λ). The theorem now tells us that (j, k, λ) occurs in an irreducible representation of G not more than $N(D)$ times.

[16] For general reference on noncompact groups and Cartan and Iwasawa decomposition see: S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962). R. Hermann, *Lie Groups for Physicists* (W. A. Benjamin, Inc., New York, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Representation Theory of Local Current Algebras*

ERIC H. ROFFMAN
Institute for Advanced Study, Princeton, New Jersey

(Received 27 February 1967)

We prove the following theorem: Every finite-dimensional irreducible representation of the local current algebra

$$[F_\rho(\mathbf{k}_1), F_\sigma(\mathbf{k}_2)] = e^\tau_{\rho\sigma} F_\tau(\mathbf{k}_1 + \mathbf{k}_2),$$

where the $\{e^\tau_{\rho\sigma}\}$ are the structure constants of a semisimple Lie algebra L, has the form

$$F_\rho(\mathbf{k}) \to \sum_{i=1}^{N} e^{i\mathbf{k}x_i} T^i_\rho,$$

where T^i_ρ is a finite-dimensional representation of L, and for $i \neq j$, $x_i \neq x_j$; T^i_ρ commutes with T^j_σ for all ρ and σ.

IF the local current algebras are assumed to satisfy the commutation relations

$$[F_\rho(\mathbf{k}_1), F_\sigma(\mathbf{k}_2)] = e^\tau_{\rho\sigma} F_\tau(\mathbf{k}_1 + \mathbf{k}_2), \tag{1}$$

where the $\{e^\tau_{\rho\sigma}\}$ are the structure constants of a semisimple Lie algebra L, then it is of interest to study the representations of the system (1) consistent with Lorentz invariance. It may be hoped, and there is reason to believe, that the space of states of elementary particles is closely related to the space of such a representation.[1]

We do not consider Lorentz invariance, which may

* Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR NR 42-65.

[1] See R. F. Dashen and M. Gell-Mann, Phys. Rev. Letters 17, 340 (1966); also M. Gell-Mann, "Relativistic Quark Model as Representation of Current Algebra," (lecture notes from International School of Physics "Ettore Majorana"), (to be published).

impose additional restrictions on the representations of Eq. (1).[2] We prove the following.

Theorem: Every representation of Eq. (1) which is finite dimensional and irreducible, has the form

$$F_\rho(\mathbf{k}) \to \sum_{i=1}^{N} e^{ikx_i} 1 \otimes \cdots \otimes T'^i_\rho \otimes \cdots \otimes 1, \qquad (2)$$

where T'^i_ρ is a finite-dimensional irreducible representation of L and $x_i \neq x_j$, if $i \neq j$.

Finite-dimensional representations of Eq. (1) are of interest in quark models. We may call the T^i_ρ of the abstract [which is the same as the

$$1 \otimes \cdots \otimes T'^i_\rho \otimes \cdots \otimes 1$$

of Eq. (2)] the representation of a quark at x_i. It is interesting that the "trivial" representation ($N = 1$) is not the only irreducible representation, but that nothing more complicated than Eq. (2) can occur. The example at the end of the paper shows the typical sort of nontriviality which can arise.

It is likely that infinite-dimensional representations of the current algebra are the ones realized in nature. The theorem we prove may be a step toward a theory for that case. In particular, the infinite analog of Eq. (2) is a nontrivial irreducible (formal) representation of Eq. (1). Also, the parameters in Eq. (2) are viewed there as scalars. If they are considered to be operators, or variables defining a space on which space translations and rotations can act, then the representations (2) become infinite-dimensional. The difficulty is proving a uniqueness theorem.

We first define precisely what we mean by a structure of the form (1) and by a representation of such a structure. Then we analyze the representation for some details, and prove the theorem. Finally, we note some properties of the representation (2).

It is useful to remark that the group of vectors $(\mathbf{k}_1, \mathbf{k}_2) \to \mathbf{k}_1 + \mathbf{k}_2$ can be replaced by an arbitrary commutative group.

Definition: Let L be a semisimple Lie algebra, let G be a commutative group, then the set of pairs

$$\{(l, g) \in L \times G\} \qquad (3a)$$

with the multiplication law

$$(l_1, g_1) \cdot (l_2, g_2) = ([l_1, l_2], g_1 g_2) \qquad (3b)$$

will be called a group-extended Lie algebra, and denoted L_G.

[2] See F. Coester and G. Roepstorff, "Current Algebras at Infinite Momentum," Argonne National Laboratory, Report.

Definition: By a representation T of a group-extended Lie algebra L_G, we mean a map

$$T: L_G \to R_E, \qquad (4)$$

where R_E is an algebra of linear operators on a linear topological space E, which satisfies

$$T(l_1 + a l_2, g) = T(l_1, g) + a T(l_2, g) \qquad (5a)$$

and

$$[T(l_1, g_1), T(l_2, g_2)] = T([l_1, l_2], g_1 g_2). \qquad (5b)$$

Proposition 1: There is a basis

$$E_\alpha, H_i, \quad \begin{matrix} \alpha = 1 \cdots n \\ i = 1 \cdots m \end{matrix}$$

for the Lie algebra L, such that for any representation T

$$T(l, g) = \sum_\alpha l_\alpha T(E_\alpha, g) + \sum_i l_i T(H_i, g), \qquad (6)$$

where $l_\alpha, l_i \in \mathbb{C}$ and

$$[T(H_i, g_1), T(H_j, g_2)] = 0, \qquad (7a)$$

$$[T(H_i, g_1), T(E_\alpha, g_2)] = r_i(\alpha) T(E_\alpha, g_1 g_2), \qquad (7b)$$

$$[T(E_\alpha, g_1), T(E_{-\alpha}, g_2)] = \sum_i r_i(\alpha) T(H_i, g_1 g_2), \qquad (7c)$$

$$[T(E_\alpha, g_1), T(E_\beta, g_2)] = N_{\alpha\beta} T(E_{\alpha+\beta}, g_1 g_2). \qquad (7d)$$

Proof: Proposition 1 follows from the corresponding properties for the Lie algebra L, and the linearity [Eq. (5a)] and commutation relations [Eq. (5b)] of a representation.

Proposition 2: Let T denote an irreducible representation on the finite-dimensional space E and let $|\xi\rangle, |\eta\rangle \in E$ denote basis elements, then:

(a) The basis elements may be identified with eigenstates of the operators $T(H_i, g)$.

(b) The basis elements may also be identified with the basis elements for a representation of the Lie algebra

$$\{(l, e)\} \simeq L \qquad (8)$$

such that the eigenvalues $\eta_i(g)$ of $|\eta\rangle$ have the property that $\eta_i(e)$ is the weight of the state $|\eta\rangle$ in the representation $(l, e) \to T(l, e)$ of L.

(c) The matrix elements $\langle\eta|\, T(E_\alpha, g)\, |\xi\rangle$ satisfy

$$\langle\eta|\, T(E_\alpha, g)\, |\xi\rangle = \chi^\alpha_{\eta\xi}(g)\, \langle\eta|\, T(E_\alpha, e)\, |\xi\rangle, \qquad (9)$$

where $\chi^\alpha_{\eta\xi}(g)$ is a one-dimensional representation of G. (It need not be the same for each choice of η, ξ, and α.)

(d) Suppose that $\eta(g)$ has a Fourier integral expansion

$$\eta(g) = \int_X d\chi \chi(g) \eta(\chi). \qquad (10)$$

Then, $\eta(\chi_0) = 0$, unless $\chi_0(g) = \chi^\alpha_{\eta\xi}(g)$ for some choice of α, η, ξ.

Proof: Equation (7a) shows the $T(H_i, g)$ are a commuting family, and the rest of Eqs. (7) show they are maximal. This proves (a).

Setting $g = e$, the identity of G, in Eqs. (7), shows that the $T(l, e)$ are a representation of L. In particular, $T(H_i, e)$ is a maximal commuting family of operators for this representation of L. This shows (b). (c) is shown by taking matrix elements of Eq. (7b). We must assume the representation is continuous. (d) follows from Eq. (7c). The Fourier expansion of $\eta(g)$ is given explicitly by Eq. (7c) as a sum of terms of the form (9).

We can now prove our Theorem.

Theorem 1: Given a finite-dimensional irreducible representation T of L_G there are a finite number of representations

$$(l, g) \to \chi_i(g) T_{\chi_i}(l, e) \tag{11}$$

of L_G which satisfy $\chi_i \neq \chi_j$, if $i \neq j$, and

$$T(l, g) = \sum_{i=1}^{N} \chi_i(g) T_{\chi_i}(l, e), \tag{12a}$$

$$l \to T_{\chi_i}(l, e) \quad \text{is a representation of } L, \tag{12b}$$

$$[T_{\chi_i}(l', e), T_{\chi_j}(l'', e)] = 0. \tag{12c}$$

if $\chi_i \neq \chi_j$. That is,

$$T(l, g) = \sum_i \chi_i(g) 1 \otimes \cdots \otimes T'_{\chi_i}(l, e) \otimes \cdots \otimes 1, \tag{13}$$

and T is irreducible if and only if each $T'_{\chi_i}(l, e)$ is irreducible.

Proof: Let $|\xi\rangle$ $|\eta\rangle$ be basis elements, and

$$\sum_\rho |\rho\rangle\langle\rho| = 1$$

a sum over a complete set of states. Consider the matrix element

$$\langle\xi|\, T(l, g)\, |\eta\rangle. \tag{14a}$$

From Eqs. (9) and (10), there is at most one χ_i for each choice of ξ, η of basis elements and each root α, which can occur in the Fourier expansion of Eq. (14a). Since the representation is finite dimensional and a Lie algebra has a finite number of roots, Eq. (14a) has the expansion (with $N < \infty$),

$$\sum_{i=1}^{N} \chi_i(g) \langle\xi|\, T_{\chi_i}(l, e)\, |\eta\rangle. \tag{14b}$$

Equation (14b) defines $\langle\xi|\, T_{\chi_i}(l, e)\, |\eta\rangle$.

A matrix element of Eq. (5b) has the form

$$\langle\xi|\, [T(l_1, g_1), T(l_2, g_2)]\, |\eta\rangle = \langle\xi|\, T([l_1, l_2], g_1 g_2)\, |\eta\rangle. \tag{15}$$

We insert a complete set of states, using Eq. (14). Linear independence of the characters now proves Eqs. (12) except for the irreducibility of the T'_{χ_i}.

On each irreducible subspace $E_{1\mu}$ of the representation T_{χ_1}, the operators $T_{\chi_2} \cdots T_{\chi_N}$ must be the identity representation by Schur's lemma. Let $\xi_i \sim \xi_j$, if they are in the same irreducible subspace of T_{χ_1}, and let φ_1 denote the equivalence relation. Continuing, we find T_{χ_N} acts on $((E/\varphi_1) \cdots /\varphi_{N-1})$. If T_{χ_i} is reducible on this space, T will be reducible. Similarly each T_{χ_N} must be irreducible on

$$[((E/\varphi_1) \cdots /\varphi_{i-1})/\varphi_{i+1}] \cdots /\varphi_m.$$

On $[((E/\varphi_1) \cdots /\varphi_{i-1})/\varphi_{i+1}] \cdots /\varphi_n T_{\chi_i}$ is isomorphic to T'_{χ_i}. That is, $T_{\chi_i} = 1 \otimes \cdots T'_{\chi_i} \otimes \cdots \otimes 1$.

Conversely, if T has the form (13), then Eqs. (12) hold. To show T is irreducible, we assume each T'_{χ_i} is irreducible; then any operator A not a multiple of the identity which commutes with T must be a permutation of the χ_i, since it must be a multiple of the identity on each irreducible subspace of each representation T_{χ_i}. But a linear combination of permutations cannot commute with T, because of the linear independence of the χ_i. Therefore, an operator A commutes with T, if and only if it is a multiple of the identity. Hence, T is irreducible.

This theorem can probably be extended in a similar, but not necessarily identical form, to the case of infinite-dimensional representations by making use of properties of distributions and infinite-tensor products.

We now give an example to show that the representations (2) need not be entirely trivial.

Let

$$L = SU(2),$$
$$G = \mathbb{R}^3, \quad \mathbf{k} \in G,$$
$$N = 2,$$
$$l \to T_{\chi_i}(l) \quad \text{be a } J = \tfrac{1}{2} \text{ representation,}$$

then

$$T(l, \mathbf{k}) = e^{i\mathbf{k}x_1}[T_{J=\frac{1}{2}}(l) \otimes 1] + e^{i\mathbf{k}x_2}[1 \otimes T_{J=\frac{1}{2}}(l)].$$

Note that

$$T_{J=\frac{1}{2}}(l) \otimes 1 = [T_{r=\frac{1}{2}}(l) \oplus T_{J=\frac{1}{2}}(l)]$$

and

$$T(l, 0) = T_{J=1}(l) \oplus T_{J=0}(l),$$

so that "at x_i," the representation is reducible ($= \frac{1}{2} \oplus \frac{1}{2}$); the integrated current algebra representation $T(l, 0)$ is also reducible ($= 1 \oplus 0$) and different from the representation at x_i, and yet the representation of the current algebra itself is irreducible.

ACKNOWLEDGMENTS

The author would like to express his great appreciation to Dr. B. W. Lee for introducing him to this problem, and to thank Dr. G. Dell'Antonio for many very helpful discussions, and Dr. M. E. Mayer for much helpful correspondence.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8 NUMBER 10 OCTOBER 1967

Interpolation Formulas in the Angular Momentum Plane

PIERRE C. SABATIER
Laboratoire de Physique Théorique et Hautes Energies, Orsay, France

(Received 20 February 1967)

When the potential is an analytic function of a rational power of r, analytical properties of the wavefunctions enable one to derive interpolation formulas in the angular-momentum plane. In these formulas, products of the wavefunctions which correspond to two different potentials are expanded in terms of their values at an infinite discrete set of points on the real axis. These expansions hold uniformly in z. Properties of the expansion coefficients are investigated. They allow the author to construct generating functions, and to obtain several interpolation formulas for functions of mathematical physics. Results are applied to the inverse scattering problem at fixed energy. They provide further information on an "angular-momentum dispersion relation" obeyed by the Jost function. A semiclassical approximation is given for the interpolation coefficients, and its meaning is investigated. Finally, an elementary derivation of the interpolation formulas is sketched and suggests a curious symbolic writing of the wavefunctions.

INTRODUCTION

In a recent paper,[1] we constructed a method to study the class $\mathcal{A}$ of potentials which are analytical functions of a rational power of the variable.[2] We applied that method to the inverse scattering problem at fixed energy, for which it gives a more general approach than that of a previous method due to Newton[3] and extensively studied by the author.[4,5] As a by-product, that method enables one to study easily the analytical properties of the wavefunctions and related quantities. These properties allow us to obtain for these functions interpolation formulas in the λ complex plane. The values of the functions in the λ plane are given by expansions in terms of their values at points of an infinite discrete set on the real axis. This set extends from $-\infty$ to $+\infty$. However, it is possible to limit the set to the positive real axis if we introduce an infinite discrete set of real coefficients, the "interpolation coefficients." The obtaining of the interpolation formulas, and properties of the interpolation coefficients are the subject of the present paper.

Such a study may have both a mathematical and a physical interest. From the mathematical point of view, it enables one to extend many formulas and properties, which were known only for Bessel functions, to solutions of the Schrödinger radial equation, considered as functions of the angular momentum λ $(= l + \frac{1}{2})$. Most of these formulas involve the interpolation coefficients, which are characteristic of the potential (at a given energy), but are unknown in most cases. However, the formulas are still interesting by the simplicity of their form. In the special case of the Coulombian potential, the interpolation coefficients can be calculated exactly and yield for the Coulombian wavefunctions an interpolation formula which seems to be new. From the physical point of view, this study allows us to get at a deeper understanding of the meaning of those formulas which enabled us to solve the inverse scattering problem in the class $\mathcal{C}$ of potentials (which is only a subclass of $\mathcal{A}$) and to enlighten the limitations of $\mathcal{C}$. That method suggests also a new interpretation of the so-called "asymptotic" or JWKB approximation. Besides, in the development of the method, many connections with the Fredholm equation theory are present. In a forthcoming paper, we shall study more extensively these connections and their application to the approximation theory in the inverse scattering problem at fixed energy.

The ambiguous characteristics of such a study are also apparent in the division of the text. Putting apart the introductory Sec. 1 and the concluding Sec. 7, we give two sections (4 and 5), of essentially physical interest. Although Sec. 2 may lead one to physical applications, we think it is mostly mathematical and so are the short Secs. 3 and 6. However, it should be noticed that all the mathematical results concern functions of current physical interest.

In Sec. 1, both in order to introduce notation and to put previous results into a more general form necessary for the following, we sketch a generalization of the method previously given by the author.[1] We take this opportunity to give some by-products of the method, for instance, a series expansion for the resolvent of the Fredholm equation there involved.

[1] P. C. Sabatier, J. Math. Phys. 8, 905 (1967). In the following, this paper will be referred to as I.
[2] The method given in Ref. 1 is slightly more general, and applies to cases which would correspond in the present paper to a continuous sequence of interpolation coefficients. Some arguments of Sec. 6 can be extended to these cases.
[3] R. G. Newton, J. Math. Phys. 3, 75 (1962).
[4] P. C. Sabatier, J. Math. Phys. 7, 1515 (1966).
[5] P. C. Sabatier, J. Math. Phys. 7, 2079 (1966).

Interpolation formulas[6] for the wavefunctions in the λ plane are derived in Sec. 2. These formulas involve products of wavefunctions corresponding at a given energy to two different potentials, and interpolation coefficients which depend on the two potentials. However, a very remarkable additive property enables us to express every coefficient as a difference between two terms corresponding respectively to each of the potentials (at the given energy). Examples are given for Bessel functions and Coulombian functions (which are known to be only a special writing of the confluent hypergeometric functions).

In Sec. 3, we show how the interpolation coefficients enable one to construct, for the wavefunctions, a generating function whose wavefunction is simply the Mellin transform.

Interpolation formulas for the Jost functions are investigated in Sec. 4. These formulas can be obtained if the potential belongs to the class $\mathcal{C}$; we previously used them in order to solve the inverse scattering problem. It is shown here that the formula used in that problem is equivalent to the data of some special analytical properties of the Jost function. A physical investigation of the potentials for which this can be true is given.

In Sec. 5, a JWKB approximation of the interpolation coefficients is derived. The "asymptotic" interpolation coefficients thus obtained are not asymptotic approximations for the interpolation coefficients of the starting potential, but of a potential of class $\mathcal{C}$ equivalent to it as regards the crosssection.

A tedious but elementary derivation of the interpolation formulas is given in Sec. 6. It suggests an expansion of the wavefunctions in terms of products of linear operators acting on 1. A symbolic writing of the wavefunction is given as a mathematical curiosity.

In Sec. 7, we remark that the interpolation formulas given in this paper are only examples of the many interpolation formulas which could be obtained in similar ways. Some extensions are suggested.

1. SOME SERIES EXPANSIONS FOR RADIAL WAVEFUNCTIONS

1.1. Generalization of Previous Results[7]

We first give in a more general form[8] some properties we previously obtained.[1] We study the wavefunctions corresponding to a potential $V(r)$ such that $zV(z)$ can be continued inside a circle $|z| < \Gamma(V)$ as an analytic function.

Let us introduce the operators

$$\begin{cases} D_0(z) = z^2((\partial^2/\partial z^2) + 1) \\ D_V(z) = D_0(z) - z^2V(z). \end{cases} \tag{1.1}$$

Now, let $\psi_\lambda(z)$ be the regular solution of the equation

$$[D_V(z) + \tfrac{1}{4}]\psi_\lambda(z) = \lambda^2\psi_\lambda(z) \tag{1.2}$$

normalized in such a way that

$$\bar{\psi}_\lambda(z) = (\tfrac{1}{2}z)^{-\lambda}\Gamma(1 + \lambda)(\tfrac{1}{2}\pi z)^{-\frac{1}{2}}\psi_\lambda(z) \to 1 \quad \text{as} \quad z \to 0. \tag{1.3}$$

From the results of the Frobenius method,[9] we know that the functions $\bar{\psi}_\lambda(z)$ are analytic in a circle $\Omega(V)$ with its center at the origin and a radius[10] $\Gamma(V) - \epsilon$.

For $V = 0$, the functions $\psi_\lambda(z)$ reduce to

$$v_\lambda(z) = (\tfrac{1}{2}\pi z)^{\frac{1}{2}}J_\lambda(z). \tag{1.4}$$

In I, we have studied the solution of the radial Eq. (1.2) starting from the functions $v_\lambda(z)$. Let $W(r)$ be a potential such that $zW(z)$ can be continued analytically inside a circle centered at the origin. In the present section, we study the solutions of the radial equations corresponding to $W(z)$, starting from the functions $\psi_\lambda(z)$. Let us introduce for that a set of numbers $\gamma_\mu(W, V)$, where μ is taken in the sequence S:

$$\mu \in S \Leftrightarrow 2\mu \quad \text{is a positive integer.} \tag{1.5}$$

We define a function $f_V^W(z, z')$ by the expansion

$$f_V^W(z, z') = \sum_{\mu \in S} \gamma_\mu(W, V)\psi_\mu(z)\psi_\mu(z'). \tag{1.6}$$

From a previous study[11] we can derive for the functions $\psi_\lambda(z)$ the following majorization[12]:

$$|\psi_\lambda(z)| < |\tfrac{1}{2}\pi z|^{\frac{1}{2}}\,|\tfrac{1}{2}z|^{\lambda}\,|\Gamma(1 + \lambda)|^{-1} \times C(1 - [\Gamma(V)]^{-1}\,|z|)^{-1}. \tag{1.7}$$

Let us now assume that the coefficients $\gamma_\mu(W, V)$ are bounded as follows:

$$|\gamma_\mu(W, V)| < C\,|\Gamma(1 + \mu)|^2\,(\tfrac{1}{2}R)^{-2\mu}. \tag{1.8}$$

It follows from (1.6), (1.7), and (1.8) that $f_V^W(z, z')$ is an analytic function of z and z', provided that z and

[6] A prepublication of some of the results concerning even potentials has been given as a note to the French Academy of Sciences. P. C. Sabatier, Compt. Rend. **263**, 788 (1966).

[7] All the proofs in this generalization are very similar to those given in Sec. 1 of Ref. 1. The reader should remark that the variable λ used here is not equal to the variable ν used in Ref. 1, but to $(\nu + \frac{1}{2})$.

[8] R. G. Newton suggests a similar generalization of his method in Ref. 3 and in his book *Scattering Theory of Waves and Particles* (McGraw-Hill Book Company, Inc., New York, 1966).

[9] See Appendix I of Ref. 1.

[10] Throughout this paper, by ϵ we mean a positive number which can be made arbitrarily small but not equal to zero. It is not meant to have the same value every time it is used, even inside a given formula. In some cases, indices will be used to avoid confusion.

[11] See in particular formula (A10) of Ref. 9.

[12] We use C as a general constant, which is not meant to have the same value every time it is used.

z' lie in the circle $\mathfrak{D}$ centered at the origin and whose radius is equal to the smaller of the numbers $\Gamma - \epsilon$ and $R - \epsilon$. Besides, it is easy to see from (1.3) and (1.5) that $f_V^W(z, z')$ is a solution of the partial differential equation

$$\begin{cases}[D_V(z) - D_V(z')]f_V^W(z, z') = 0 \\ f_V^W(z, 0) = f_V^W(0, z') = 0. \end{cases} \tag{1.9}$$

For real values of z and z', we now define $K_V^W(z, z')$ from $f_V^W(z, z')$ as the solution of the integral equation:

$$K_V^W(z, z') = f_V^W(z, z') - \int_0^z d\zeta \zeta^{-2} K_V^W(z, \zeta) f_V^W(\zeta, z'). \tag{1.10}$$

$K_V^W(z, z')$ can be related to the resolvent of a Fredholm equation associated with (1.10)[13]:

$$K(z, z') = (zz')\sigma_z(z, z'), \tag{1.11}$$

where $\sigma_z(z, z')$ is a particular value of the solution of

$$\sigma_z(z'', z') = \phi(z'', z') - \int_0^z d\zeta \sigma_z(z'', \zeta)\phi(\zeta, z'), \tag{1.12}$$

where

$$\phi(z'', z') = (z''z')^{-1} f(z'', z'). \tag{1.13}$$

The Fredholm method yields readily $\sigma_z(z, z')$:

$$\sigma_z(z, z') = [\mathfrak{D}(z)]^{-1}\mathfrak{D}_z(z, z'), \tag{1.14}$$

where

$$\mathfrak{D}(z) = 1 + \sum_{m=1}^{\infty} (m!)^{-1} \int_0^z \cdots \int_0^z \times \Phi\begin{pmatrix} z_1, z_2, \cdots, z_m \\ z_1, z_2, \cdots, z_m \end{pmatrix} dz_1 \cdots dz_m, \tag{1.15}$$

$$\mathfrak{D}_z(z, z') = -\sum_{m=0}^{\infty} (m!)^{-1} \int_0^z \cdots \int_0^z \times \Phi\begin{pmatrix} z, z_1, \cdots, z_m \\ z', z_1, \cdots, z_m \end{pmatrix} dz_1 \cdots dz_m, \tag{1.16}$$

where the notation Φ is used for the Fredholm determinants associated with the kernel $\phi(z, z')$. As we did in I, we can use formulas (1.11), (1.14), (1.15), and (1.16) in order to continue analytically $K(z, z')$ as z and z' take complex values in $\mathfrak{D}$. It is clear that $K(z, z')$ is an analytic function of z' and a meromorphic function of z as z and z' belong to $\mathfrak{D}$. It may be of interest to remark that from (1.15) and (1.16) we derive the simple formula

$$\mathfrak{D}_z(z, z) = -\mathfrak{D}'(z). \tag{1.17}$$

This is similar to a result obtained in the Gel'fand–Marshenko method.[14] It enables one to relate the determinant of Eq. (1.12) to the potential $W(z)$ as defined below. Manipulations of (1.10) similar to those used in I enable us to show that

$$\begin{cases} D_W(z)K_V^W(z, z') - D_V(z')K_V^W(z, z') = 0 \\ K_V^W(z, 0) = K_V^W(0, z') = 0, \end{cases} \tag{1.18}$$

where $D_W(z)$ is defined as in (1.1) with the potential $W(z)$ instead of $V(z)$, $W(z)$ being defined in turn by

$$W(z) = V(z) - 2z^{-1}(d/dz)z^{-1}K_V^W(z, z). \tag{1.19}$$

Introducing now the functions $\chi_\lambda(z)$, which are associated with the potential $W(z)$ as the functions $\psi_\lambda(z)$ to $V(z)$, and proceeding as in I, we easily derive the formula

$$\chi_\lambda(z) = \psi_\lambda(z) - \int_0^z d\zeta \zeta^{-2} K_V^W(z, \zeta)\psi_\lambda(\zeta) \quad (\lambda > 0). \tag{1.20}$$

This formula is the key of all the series expansions which we now study.

1.2. Series Expansions

From (1.20) and (1.10), it is easy to derive the bilinear expansion of $K_V^W(z, z')$:

$$K_V^W(z, z') = \sum_{\mu \in S} \gamma_\mu(W, V)\chi_\mu(z)\psi_\mu(z'). \tag{1.21}$$

The method is the same as given previously.[1,15] Taking into account (1.19) and the behavior of the functions near $z = 0$, we see that $K_V^W(z, z)$ is unambiguously related to the difference of the potentials

$$K_V^W(z, z) = -\tfrac{1}{2}z \int_0^z \rho(W(\rho) - V(\rho))\, d\rho. \tag{1.22}$$

Let us now consider the bilinear expansion of $K_V^W(z, z)$:

$$K_V^W(z, z) = \sum_{\mu \in S} \gamma_\mu(W, V)\chi_\mu(z)\psi_\mu(z). \tag{1.23}$$

With the same arguments as in I, we can easily show the following points:

(a) The expansion of $K_V^W(z, z)$ can be formally constructed from the data of $W(z)$ and $V(z)$ by comparing the Taylor expansions of the functions there involved.

(b) The expansion converges uniformly in any circle centered at the origin in which both $zW(z)$ and $zV(z)$ are analytic.

(c) The expansion coefficients $\gamma_\mu(W, V)$ are unique for a given couple of potentials $V(z)$, $W(z)$.

Inserting (1.21) in (1.20) yields the expansion of $\chi_\lambda(z)$

$$\chi_\lambda(z) = \psi_\lambda(z) - \sum_{\mu \in S} {}^V\Lambda_\lambda^\mu(z)\gamma_\mu(W, V)\chi_\mu(z), \tag{1.24}$$

[13] From Eqs. (1.11) to (1.17), we drop the indices V, W aside the functions.

[14] We thank Dr. Cornille who drew our attention to this point.

[15] The derivation is also analogous to (and simpler than) the derivation of $S_z(z_1, z_2)$, as given below.

where

$$^{V}\Lambda^{\mu}_{\lambda}(z) = [\psi_{\lambda}(z)\psi'_{\mu}(z) - \psi_{\mu}(z)\psi'_{\lambda}(z)]/(\mu^2 - \lambda^2). \quad (1.25)$$

From (1.22) and (1.23) we obtain easily the antisymmetric relation

$$\gamma_{\mu}(W, V) = -\gamma_{\mu}(V, W) \quad (1.26)$$

from which we derive, with the help of (1.24),

$$\chi_{\lambda}(z) = \psi_{\lambda}(z) - \sum_{\mu \in S} {}^{W}\Lambda^{\mu}_{\lambda}(z)\gamma_{\mu}(W, V)\psi_{\mu}(z). \quad (1.27)$$

It is interesting to find similar expansions[16] for the resolvent of Eq. (1.10):

$$S_z(z_1, z_2) = z_1 z_2 \sigma_z(z_1, z_2), \quad (1.28)$$

which is a solution of the equation

$$S_z(z_1, z_2) = f(z_1, z_2) - \int_0^z d\rho \rho^{-2} S_z(z_1, \rho) f(\rho, z_2), \quad (1.29)$$

where $f(z_1, z_2)$, $S_z(z_1, z_2)$ is a symmetric function of z_1 and z_2. We try to find an expansion of $S_z(z_1, z_2)$ in the form

$$S_z(z_1, z_2) = \sum_{\mu \in S} \gamma_{\mu} \mathcal{K}_{\mu}(z \mid z_1)\psi_{\mu}(z_2). \quad (1.30)$$

Inserting (1.30) in (1.29) leads us to the equation, for all $\nu \in S$:

$$\mathcal{K}_{\nu}(z \mid z_1) = \psi_{\nu}(z_1) - \sum_{\mu \in S} \gamma_{\mu} \mathcal{K}_{\mu}(z \mid z_1) {}^{V}\Lambda^{\mu}_{\nu}(z). \quad (1.31)$$

Differentiating (1.31) yields the equation

$$\frac{d}{dz}\mathcal{K}_{\nu}(z \mid z_1) = \psi_{\nu}(z) F(z \mid z_1) - \sum_{\mu \in S} \gamma_{\mu} {}^{V}\Lambda^{\mu}_{\nu}(z) \frac{d}{dz}\mathcal{K}_{\mu}(z \mid z_1), \quad (1.32)$$

where

$$\begin{aligned} F(z \mid z_1) &= -z^{-2} \sum_{\mu \in S} \gamma_{\mu} \mathcal{K}_{\mu}(z \mid z_1)\psi_{\mu}(z) \\ &= -z^{-2} S_z(z_1, z) \\ &= -z^{-2} S_z(z, z_1) \\ &= -z^{-2} K(z, z_1). \end{aligned} \quad (1.33)$$

Let us now put

$$(d/dz)\mathcal{K}_{\nu}(z \mid z_1) = F(z \mid z_1)\chi_{\nu}(z \mid z_1). \quad (1.34)$$

Inserting (1.34) in (1.32) shows that $\chi_{\nu}(z \mid z_1)$ is a solution of (1.24), analytic in z, with the same value for $z = 0$ as $\chi_{\nu}(z)$. Now, the infinite linear system obtained by giving successively to λ all the values of S in (1.24) is equivalent to (1.10). Besides, from the same study we previously made,[17] we know that there is only one solution of (1.10) analytic in z. It follows that $\chi_{\nu}(z \mid z_1)$ is equal to $\chi_{\nu}(z)$. Insertion of this result of (1.34) and a glance at (1.31) yield the value of $\mathcal{K}_{\nu}(z \mid z_1)$:

$$\begin{aligned} \mathcal{K}_{\nu}(z \mid z_1) &= \psi_{\nu}(z_1) - \int_0^z K(\rho, z_1)\chi_{\nu}(\rho)\rho^{-2}\, d\rho \\ &= \psi_{\nu}(z_1) - \sum_{\mu \in S} \gamma_{\mu}\psi_{\mu}(z_1) {}^{W}\Lambda^{\mu}_{\nu}(z). \end{aligned} \quad (1.35)$$

It is clear, from (1.27), that $\mathcal{K}_{\nu}(z \mid z)$ reduces to $\chi_{\nu}(z)$. Formulas (1.35) and (1.30) yield the bilinear expansion for the resolvent:

$$S_z(z_1, z_2) = \sum_{\mu \in S} \gamma_{\mu}\psi_{\mu}(z_1)\psi_{\mu}(z_2) - \sum_{\mu \in S} \gamma_{\mu}\psi_{\mu}(z_2) \sum_{\mu' \in S} \gamma_{\mu'}\psi_{\mu'}(z_1) {}^{W}\Lambda^{\mu'}_{\mu}(z) \quad (1.36a)$$

or equivalently,

$$S^{W}_{Vz}(z_1, z_2) = f^{W}_{V}(z_1, z_2) - \int_0^z K^{W}_{V}(\tau, z_1) \times K^{W}_{V}(\tau, z_2)\tau^{-2}\, d\tau. \quad (1.36b)$$

The expansion formulas we have obtained show the close connection between Fredholm equations and the Schrödinger equation. This connection can be studied still further by putting a parameter—say σ—as a multiplicative coefficient in front of the integral in (1.10). Equation (1.10) takes then the usual form of an integral equation whose solutions are studied as functions of σ. It is remarkable that the connection with the Schrödinger equation can be kept—with some simple dependence on σ in the connection formulas. This will be studied thoroughly in a forthcoming paper, together with the problem of approximations in both Fredholm equations and the inverse scattering problem at fixed energy.

1.3. Generalization of the Method

Let m be a positive integer, and let us define a determination ζ of $z^{m^{-1}}$, for instance, the determination which is real and positive for z real and positive. The method we have described for analytic potentials can straightforwardly be extended for potentials $V(r)$ such that $z^2V(z)$ is an analytic function of ζ, equal to zero for ζ equal to zero. The method is derived from the method introduced previously,[1] exactly in the same way as we did in Sec. 1.1 for analytic potentials. Results are similar. The expansions which are obtained are analytic functions of ζ in some circle centered at the origin. We can generalize readily in this way formulas (1.18) to (1.36). The only difference is that now the sequence S includes all numbers μ equal to

$$\mu = k/2m \Leftrightarrow \mu \in S_m \quad (k \text{ positive integer}). \quad (1.37)$$

[16] From (1.28) to (1.36a), we drop the indices V, W aside the functions.

[17] See in particular Sec. 1 of Ref. 5.

1.4. Particular Potentials

A series expansion like (1.24) enables us to continue analytically $\chi_\lambda(z)$ in the λ complex plane, provided we know the continuation of ${}^V\Lambda^\mu_\lambda(z)$. This is achieved for those potentials for which the Schrödinger equation is exactly solvable. The most remarkable case is $V(z) = +1$. We have then

$$\psi^1_\lambda(z) \equiv s_\lambda(z) = (\tfrac{1}{2}\pi z)^{\frac{1}{2}}[\Gamma(1+\lambda)]^{-1}(\tfrac{1}{2}z)^\lambda. \quad (1.38)$$

In this case, $K_1^W(z, z')$ appears to be a generating function of the functions $[\chi_\mu(z) - s_\mu(z)]$:

$$(\tfrac{1}{2}\pi t)^{-\frac{1}{2}}K_1^W(z,t) = \sum_{\mu\in S}\gamma_\mu(W,1)\chi_\mu(z)\frac{(\frac{1}{2}t)^\mu}{\Gamma(1+\mu)} \quad (1.39)$$

and the functions $\chi_\lambda(z)$ are generated from $K_1^W(z, t)$ by a Mellin transform:

$$\chi_\lambda(z) = s_\lambda(z) - \int_0^z d\zeta\zeta^{-2}K_1^W(z,\zeta)s_\lambda(\zeta). \quad (1.40)$$

The coefficients ${}^1\Lambda^\mu_\lambda(z)$ are very simple:

$${}^1\Lambda^\mu_\lambda(z) = \frac{\pi}{\Gamma(1+\lambda)\Gamma(1+\mu)}(\tfrac{1}{2})^{\lambda+\mu+1}\frac{z^{\lambda+\mu}}{\lambda+\mu}. \quad (1.41)$$

For a given sequence S, that is, for a given set of fixed μ, and for variable λ, these functions are meromorphic with poles for $\lambda = -\mu$, except if μ is an integer; q being a positive integer, we have

$${}^1\Lambda^q_{-q}(z) = -(\pi/2q)(-1)^q. \quad (1.42)$$

Other particular potentials are the constant potentials and the Coulombian potentials. Needless to say, we use here the word "potential" with a very large meaning, without any assumption concerning, for instance, the asymptotic behavior. Assumptions of this kind will only be introduced later (Sec. 4, below). The functions ${}^V\Lambda^\mu_\lambda(z)$ are given for the constant potentials by formula (1.25), where $\psi_\lambda(z)$ reduces to a Bessel function [see (1.44) below]. $\psi_\lambda(z)$ is then an entire function of λ. $\Lambda^\mu_\lambda(z)$ is therefore an analytic function of λ, unless $\lambda = \mu$ if μ is not an integer, and is an entire function of λ if μ is an integer. For the zero potential, it is easy to show the reflection formula

$${}^0\Lambda^p_{\lambda=-q} = (-1)^q[{}^0\Lambda^p_{\lambda=q}(z) - (\pi/2p)\delta^q_p] \quad (1.43)$$

by making use of the following expression of ${}^0\Lambda^\mu_\lambda(z)$:

$${}^0\Lambda^\mu_\lambda(z) = \frac{\sin(\lambda-\mu)\frac{1}{2}\pi}{\lambda^2-\mu^2} - \int_z^\infty v_\lambda(t)v_\mu(t)t^{-2}\,dt, \quad (1.44)$$

which is valid for $\mathrm{Re}\,(\lambda+\mu) > 0$ and can be continued outside this domain.

1.5. Properties of the Wavefunctions in the λ Plane

Formula (1.24), where S can be defined as in (1.37), enables us to continue $\chi_\lambda(z)$ in the λ plane provided we know the continuation of ${}^V\Lambda^\mu_\lambda(z)$. For this we have to take a particular potential V. We shall call it the base potential, or base. Taking a constant potential as a base, we see that, for any z for which expansion(1.24) holds, $\chi_\lambda(z)$ is a meromorphic function of λ, with poles at the points of S, except the integers. Using (1.42) and (1.43) yields the reflection formulas

$$\chi_{-q}(z) = (\pi/2q)(-1)^q\gamma_q(W,1)\chi_q(z), \quad (1.45)$$

$$\chi_{-q}(z) = (-1)^q[1 + (\pi/2q)\gamma_q(W,0)]\chi_q(z). \quad (1.46)$$

Using the zero potential as a base, it is also easy to obtain the asymptotic behavior of $\chi_\lambda(z)$ as $|\lambda|$ tends to infinity:

$$-\frac{\chi_\lambda(z)}{(\frac{1}{2}\pi z)^{\frac{1}{2}}(\frac{1}{2}z)^\lambda\Gamma(-\lambda)}\frac{\pi}{\sin\pi\lambda} = \frac{\Gamma(1+\lambda)\chi_\lambda(z)}{(\frac{1}{2}\pi z)^{\frac{1}{2}}(\frac{1}{2}z)^\lambda}$$
$$= 1 - \lambda^{-1}\left[(\tfrac{1}{2}z)^2 + z^{-1}\sum_{\mu\in S}\gamma_\mu(W,0)\chi_\mu(z)v_\mu(z)\right] + 0(\lambda^{-2}) \quad (|\mathrm{Im}\,\lambda| > \epsilon). \quad (1.47)$$

2. INTERPOLATION FORMULAS

2.1. Even Potentials

We first assume that $V(z)$ is zero, whereas $W(z)$ is an even analytic function inside the circle $|z| < \Gamma$. We know from a result of I that the coefficients $\gamma_{p+\frac{1}{2}}(W, 0)$ are then equal to zero. For the sake of simplicity, we use the notation δ_p instead of $\gamma_p(W, 0)$. Let now Ω be a circle $|z| < \Gamma - \epsilon$, the wavefunction $\chi_\lambda(z)$ is therefore given, as z lies in Ω, by the expansion

$$\chi_\lambda(z) = v_\lambda(z) - \sum_{p=1}^\infty {}^0\Lambda^p_\lambda(z)\delta_p\chi_p(z). \quad (2.1)$$

In order to obtain interpolation formulas for the functions $\chi_\lambda(z)$, we now use a method previously used by Chadan in order to obtain interpolation formulas in the energy plane.[18] The key for this method is the Lagrange–Valiron theorem[19]:

Theorem: Any entire function $f(z)$ of order 1 and finite type τ which is bounded on the real axis has the representation:

$$f(z) = \frac{f'(0)}{\tau}\sin\tau z + \frac{f(0)}{\tau z}\sin\tau z + \tau z\sin\tau z\sum_{-\infty}^{+\infty}{}'\frac{(-1)^n f[n(\pi/\tau)]}{n\pi(\tau z - n\pi)}, \quad (2.2)$$

[18] K. Chadan, Nuovo Cimento **39**, 697 (1965).

[19] R. P. Boas, Jr., *Entire Functions* (Academic Press Inc., New York, 1954).

where the prime on $\sum$ means that one has to exclude $n = 0$.

Let us now recall the well-known expansion

$$\frac{-v_\lambda(z)}{(\frac{1}{2}\pi z)^{\frac{1}{2}}(z/2)^\lambda\Gamma(-\lambda)} = \frac{\sin\pi\lambda}{\pi}\sum_0^\infty \frac{(-1)^{2n}(z/2)^n\Gamma(\lambda+1)}{n!\,\Gamma(\lambda+1+n)}. \quad (2.3)$$

Using in (2.1) formula (2.3) and the formula obtained for $v'_\lambda(z)$ by differentiating (2.3), we see that

$$\bar{\chi}_\lambda \equiv -\chi_\lambda(z)/(\tfrac{1}{2}\pi z)^{\frac{1}{2}}(z/2)^\lambda\Gamma(-\lambda) \quad (2.4)$$

is an entire function of order 1 and type π, bounded on the real axis. Applying the above theorem, and taking into account (1.46), we obtain the interpolation formula

$$\left(\frac{z}{2}\right)^{-\lambda}\Gamma(1+\lambda)\chi_\lambda(z) = \chi_0(z) + \sum_1^\infty \frac{\lambda}{\lambda+n} \times \left(1+\frac{\pi}{2n}\delta_n\right)\left(\frac{z}{2}\right)^n\frac{\chi_n(z)}{n!}. \quad (2.5)$$

For $\delta_n \equiv 0$, the functions $\chi_\lambda(z)$ reduce to the Bessel functions (1.4) and (2.5) reduces to a well-known formula.[20] From (2.5), it is possible to find other formulas by letting $|\lambda| \to \infty$ in (2.5) (avoiding the negative real axis for the sake of simplicity), and comparing the coefficients of $|\lambda|^{-n}$ on both sides of (2.5). The coefficients in the right-hand side are given in (1.47). The results for the first order are[21]

$$1 = (\tfrac{1}{2}\pi z)^{-\frac{1}{2}}\left\{\chi_0(z) + \sum_1^\infty\left(1+\frac{\pi}{2n}\delta_n\right)\left(\frac{z}{2}\right)^n\frac{\chi_n(z)}{n!}\right\}, \quad (2.6)$$

$$\left[\left(\frac{z}{2}\right)^2 + z^{-1}\sum_p \delta_p\chi_p(z)v_p(z)\right] = \tfrac{1}{2}\int_0^z \tau[1-W(\tau)]\,d\tau = (\tfrac{1}{2}\pi z)^{-\frac{1}{2}}\sum_1^\infty n\left(1+\frac{\pi}{2n}\delta_n\right)\left(\frac{z}{2}\right)^n\frac{\chi_n(z)}{n!}. \quad (2.7)$$

Let us now introduce the functions $\psi_\lambda(z)$, associated with the potential $V(z)$, and let us assume that $V(z)$ is itself an even analytic potential, whose coefficients $\gamma_p(V, 0)$ will be denoted by γ_p. It is easy to see that the function $\lambda\psi_\lambda(z)\chi_{-\lambda}(z)$ is an entire function of order 1, type π, bounded on the real axis. Application of formula (2.3) leads us to the interpolation formula

$$\frac{\pi\lambda}{\sin\pi\lambda}\psi_\lambda(z)\chi_{-\lambda}(z) = \psi_0(z)\chi_0(z) + \sum_1^\infty \psi_n(z)\chi_n(z) \times \left\{\frac{\lambda}{\lambda-n}\left(1+\frac{\pi}{2n}\delta_n\right) + \frac{\lambda}{\lambda+n}\left(1+\frac{\pi}{2n}\gamma_n\right)\right\}. \quad (2.8)$$

Comparison of the powers of $|\lambda|^{-1}$ as $|\lambda| \to \infty$ leads us, for the zeroth order, to the formula,

$$1 = (\tfrac{1}{2}\pi z)^{-1}\left\{\psi_0(z)\chi_0(z) + \sum_1^\infty \psi_n(z)\chi_n(z) \times [2 + (\pi/2n)(\gamma_n+\delta_n)]\right\}. \quad (2.9)$$

When γ_n and δ_n are equal to zero, ψ_n and χ_n reduce to v_n and (2.9) reduces to a well-known formula.[22] Comparison of the coefficients of $|\lambda|^{-1}$ in (2.8) yields the formula,

$$\sum_1^\infty \delta_p\chi_p(z)v_p(z) - \sum_1^\infty \gamma_p\psi_p(z)v_p(z) = \sum_1^\infty(\delta_p-\gamma_p)\chi_p(z)\psi_p(z). \quad (2.10)$$

Since the right-hand side of (2.10) is nothing but the function $K_V^W(z, z)$, we derive from (2.10) the very remarkable formula:

$$\gamma_p(W, V) = \gamma_p(W, 0) - \gamma_p(V, 0) \quad (2.11)$$

from which elementary algebra leads us to

$$\gamma_p(W, V) = \gamma_p(W, V_0) - \gamma_p(V, V_0). \quad (2.12)$$

This formula is very important. From (1.18), we might think that the set of coefficients $\gamma_p(W, V)$ is characteristic of a *couple* of potentials W, V. From (2.12) we see that, at least for even potentials, it is possible, by taking an arbitrary reference potential V_0, to consider the γ_p's as a set of coefficients characteristic of *a potential* V. It follows also from (2.12) that, if the asymptotic behavior of the γ_p's gives rise to difficulties in the study of a potential W, it may be possible to use a known potential $V(z)$ such that $\gamma_p(W, V)$ has a good asymptotic behavior, so that the comparison is easy.

2.2. Example of Even Potentials

The simplest example is the constant potential, e.g., $(1 - k^2)$. Solutions of (1.2) are then

$$\chi_\lambda(z) = (\tfrac{1}{2}\pi z)^{\frac{1}{2}}k^{-\lambda}J_\lambda(kz) \quad (2.13)$$

so that

$$k^{2p} = 1 + (\pi/2p)\delta_p. \quad (2.14)$$

[20] Bateman Manuscript Project, *Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953), formula 7.15 (10).

[21] For a more detailed proof, please refer to the Appendix.

[22] Ref. 20, formula 7.15 (38) for $\phi = 0$.

Equation (2.8) reduces to

$$\frac{\pi\lambda}{\sin\pi\lambda}\psi_\lambda(z)k^\lambda J_{-\lambda}(kz) = \psi_0(z)J_0(kz) + \sum_1^\infty \psi_n(z)J_n(kz) \times \left\{\frac{\lambda}{\lambda-n}k^n + \frac{\lambda}{\lambda+n}\left(1+\frac{\pi}{2n}\gamma_n\right)k^{-n}\right\}. \quad (2.15)$$

From (2.15), it is easy to derive (2.5) by setting $k \to 0$ and $\psi_\lambda = \chi_\lambda$. Replacing ψ_λ by (2.13) and setting k equal to 1 yields the formula

$$J_\lambda(z)J_{-\lambda}(z) = \frac{\sin\pi\lambda}{\pi\lambda}\left\{J_0^2(z) + 2\sum_1^\infty J_n^2(z)(\lambda^2-n^2)^{-1}\lambda^2\right\}. \quad (2.16)$$

2.3. Analytic Potentials

We now assume that $W(z)$ is an analytic function inside the circle $|z| < \Gamma$, and we use for its coefficients the notation δ_μ. The wavefunction $\chi_\lambda(z)$ is given in Ω by the expansion

$$\chi_\lambda(z) = v_\lambda(z) - \sum_{\mu\in S}{}^0\Lambda_\lambda^\mu(z)\delta_\mu\chi_\mu(z), \quad (2.17)$$

where S includes all the μ's such that 2μ is a positive integer. This function is meromorphic, having poles for half-integral values of λ. We define therefore the entire function of λ,

$$\chi_\lambda^1(z) = \cos\pi\lambda\chi_\lambda(z). \quad (2.18)$$

For any positive integer p or q, the function $\chi_\lambda^1(z)$ satisfies the following relations:

$$\begin{aligned}\chi_{-q}^1(z) &= (-1)^q\chi_q^1(z)[1+(\pi/2q)\delta_q] \\ &= \chi_q(z)[1+(\pi/2q)\delta_q], \\ \chi_{p-\frac12}^1(z) &= 0, \\ \chi_{-(p-\frac12)}^1(z) &= -\pi/[2(p-\tfrac12)]\delta_{p-\frac12}\chi_{p-\frac12}.\end{aligned} \quad (2.19)$$

As previously, we see that

$$\bar\chi_\lambda^1(z) = \frac{-\chi_\lambda^1(z)}{(\frac12\pi z)^{\frac12}(z/2)^\lambda\Gamma(-\lambda)}$$

is an entire function of order 1 and type 2π, bounded on the real axis. Applying formulas (2.3) and (2.19), we obtain the interpolation formula

$$\begin{aligned}&\Gamma(1+\lambda)\left(\frac{z}{2}\right)^{-\lambda}\chi_\lambda(z) \\ &= \chi_0(z) + \sum_1^\infty \frac{\lambda}{\lambda+n}\left(1+\frac{\pi}{2n}\delta_n\right)\left(\frac{z}{2}\right)^n\frac{\chi_n(z)}{n!} \\ &\quad + \sum_1^\infty \frac{\lambda}{\lambda+n-\frac12}\frac{\pi}{2(n-\frac12)}\delta_{n-\frac12}\left(\frac{z}{2}\right)^{n-\frac12}\frac{\chi_{n-\frac12}(z)}{\Gamma(n+\frac12)}.\end{aligned} \quad (2.20)$$

Formulas similar to (2.6) or (2.7) can easily be derived from (2.20). Let us now introduce the functions $\psi_\lambda(z)$, associated to a potential $V(z)$ analytic in Ω. It is easy to see that the function

$$\lambda\cos\pi\lambda\psi_\lambda(z)\chi_{-\lambda}(z)$$

is an entire function, of order 1, type 2π, bounded on the real axis. Application of (2.3) yields

$$\begin{aligned}\frac{\pi\lambda}{\sin\pi\lambda}\psi_\lambda(z)\chi_{-\lambda}(z) &= \psi_0(z)\chi_0(z) + \sum_1^\infty \psi_n(z)\chi_n(z) \\ &\quad + \left[\frac{\lambda}{\lambda-n}\left(1+\frac{\pi}{2n}\delta_n\right) + \frac{\lambda}{\lambda+n}\left(1+\frac{\pi}{2n}\gamma_n\right)\right] \\ &\quad + \sum_1^\infty \psi_{n-\frac12}(z)\chi_{n-\frac12}(z)\frac{\pi}{2n-1} \\ &\quad \times\left(\frac{\lambda}{\lambda-n+\frac12}\delta_{n-\frac12} + \frac{\lambda}{\lambda+n-\frac12}\gamma_{n-\frac12}\right).\end{aligned} \quad (2.21)$$

This formula is a straightforward generalization of (2.8). Considerations of the powers of $|\lambda|^{-1}$ leads us to the generalization of (2.12),

$$\gamma_\mu(W, V) = \gamma_\mu(W, V_0) - \gamma_\mu(V, V_0) \quad \forall\mu\in S. \quad (2.22)$$

2.4. Examples

The simplest example is the Coulombian potential

$$zV(z) = 2\eta = 2(Z_1Z_2e^2/\hbar v), \quad (2.23)$$

where v is the initial speed, Z_1e and Z_2e the charges of the incident particle and the target. From well-known results,[23] we obtain easily the value of $\psi_\lambda(z)$:

$$\begin{aligned}\psi_\lambda(z) &= (\tfrac12\pi z)^{\frac12}\left(\frac{z}{2}\right)^\lambda\frac{e^{iz}}{\Gamma(1+\lambda)} \\ &\quad \times {}_1F_1(\lambda+\tfrac12+i\eta; 2\lambda+1; -2iz) \quad (2.24) \\ &= (\tfrac12\pi z)^{\frac12}\left(\frac{z}{2}\right)^\lambda\frac{e^{-iz}}{\Gamma(1+\lambda)} \\ &\quad \times {}_1F_1(\lambda+\tfrac12-i\eta; 2\lambda+1; 2iz) \quad (2.25)\end{aligned}$$

by Kummer's transformation.

Comparison between $\psi_p(z)$ and $\psi_{-p}(z)$ yields, after some algebra, the value of the coefficients for integer p:

$$1+\frac{\pi}{2p}\gamma_p = \frac{\Gamma(p+\frac12-i\eta)\Gamma(p+\frac12+i\eta)}{\Gamma(p+\frac12)\Gamma(p+\frac12)}\cosh\pi\eta \quad (2.26)$$

which reduces to 1 for $p = 0$, and goes to $\cosh\pi\eta$ as $p \to \infty$. The coefficients γ_p increase therefore like p for large values of p.

[23] See, for example, A. Messiah, *Mécanique Quantique* (Dunod Cie., Paris, 1959), Vol. 1, p. 413.

Comparison between $\cos \pi\lambda\psi_{-\lambda}(z)$ and $\psi_\lambda(z)$ for $\lambda = p - \frac{1}{2}$ yields, after some algebra, the value of the coefficients with half-integer index

$$-\frac{\pi}{2(p-\frac{1}{2})}\delta_{p-\frac{1}{2}} = \frac{\Gamma(p+i\eta)\Gamma(p-i\eta)}{\Gamma(p)\Gamma(p)}\sinh \pi\eta. \tag{2.27}$$

2.5. Further Generalization

If we assume now that S is defined as in (1.37), we can still define an entire function of λ

$$\chi_\lambda^m(z) = \frac{\sin 2m\pi\lambda}{2m\sin\pi\lambda}\chi_\lambda(z). \tag{2.28}$$

This function satisfies the following relations:

$$\begin{cases} \chi_{-q}^m(z) = (-1)^q\chi_q^m\left(1+\dfrac{\pi}{2q}\delta_q\right) \\ \qquad\qquad = \chi_q(z)\left(1+\dfrac{\pi}{2q}\delta_q\right), \\ \chi_{k/2m}^m(z) = 0, \quad \text{for integer} \quad k \neq 2qm, \\ \chi_{-k/2m}^m(z) = (-1)^k(\pi m/k)\delta_{k/2m}\chi_{k/2m}(z). \end{cases} \tag{2.29}$$

We can easily see that

$$\lambda(\sin 2m\pi\lambda/2m\sin\pi\lambda)\psi_\lambda(z)\chi_{-\lambda}(z)$$

is an entire function of order 1, type $2m\pi$, bounded on the real axis, and obtain the general interpolation formula

$$\begin{aligned} &\frac{\pi\lambda}{\sin\pi\lambda}\psi_\lambda(z)\chi_{-\lambda}(z) \\ &= \psi_0(z)\chi_0(z) + \sum_{\substack{\mu\in S\\ \mu\neq n}}\frac{\pi}{2\mu}\psi_\mu(z)\chi_\mu(z)\left(\frac{\lambda}{\lambda-\mu}\delta_\mu + \frac{\lambda}{\lambda+\mu}\gamma_\mu\right) \\ &\quad + \sum_{n=1}^{\infty}\psi_n(z)\chi_n(z)\left[\frac{\lambda}{\lambda-n}\left(1+\frac{\pi}{2n}\delta_n\right)\right. \\ &\quad \left. + \frac{\lambda}{\lambda+n}\left(1+\frac{\pi}{2n}\gamma_n\right)\right]. \end{aligned} \tag{2.30}$$

Formula (2.22) is also generalized straightforwardly. Consideration of it and comparison between (1.45) and (1.46) show that

$$\begin{aligned} \gamma_\mu(0,1) &= 2\mu/\pi, \quad \text{for positive integer } \mu \\ &= 0, \qquad \text{otherwise.} \end{aligned} \tag{2.31}$$

Taking 1 as a reference potential yields therefore the simplest form of the interpolation formula (2.30):

$$\begin{aligned} &\frac{\pi\lambda}{\sin\pi\lambda}\psi_\lambda(z)\chi_{-\lambda}(z) \\ &= \psi_0(z)\chi_0(z) + \sum_{\mu\in S}\frac{\pi}{2\mu}\psi_\mu(z)\chi_\mu(z) \\ &\quad \times\left\{\frac{\lambda}{\lambda-\mu}\gamma_\mu(W,1) + \frac{\lambda}{\lambda+\mu}\gamma_\mu(V,1)\right\}. \end{aligned} \tag{2.32}$$

3. GENERATING FUNCTIONS

As we have already noticed, $K_V^W(z,z')$ may be, in a certain sense, considered as a generating function for the functions $\chi_\lambda(z)$, through formula (1.20). It is, however, possible to find another function, for which the dependence of the functions $\chi_\lambda(z)$ does not involve an extra term.

For a given difference of potentials $W(z) - V(z)$, there is one and only one function $K_V^W(z,z)$, as defined by (1.22). If we add to $z^{-1}K_V^W(z,z)$ another function, Eq. (1.19) leads us to a different potential unless this additional function is a constant. This remark suggests the study of an operator $L_V^W(z,z')$ which would be a solution to the partial differential equation (1.18) and which would reduce to Cz as $z = z'$.

From Eq. (2.9), or more exactly, from the generalization of (2.9) obtained from (2.32), it is clear that

$$\begin{aligned} L_V^W(z,z') &= 2\pi^{-1}\chi_0(z)\psi_0(z') \\ &\quad + \sum_{\mu\in S}\chi_\mu(z)\psi_\mu(z')(2\mu)^{-1}[\gamma_\mu(W,1) + \gamma_\mu(V,1)] \end{aligned} \tag{3.1}$$

and

$$L_V^W(z,z) = z. \tag{3.2}$$

Let us now define the function

$$\chi_\lambda^*(z) = \int_0^z d\rho\rho^{-2}L_V^W(z,\rho)\psi_\lambda(\rho) \quad \text{Re}\,\lambda > 0. \tag{3.3}$$

Applying $D_W(z)$ on both sides of (3.3), together with straightforward integrations by parts, and taking into account (3.3), we see that $\chi_\lambda^*(z)$, for $\text{Re}\,\lambda > 0$, is a solution of the equation

$$[D_W(z) + \tfrac{1}{4}]\chi_\lambda^*(z) = \lambda^2\chi_\lambda^*(z) \tag{3.4}$$

with the boundary condition

$$\chi_\lambda^*(z) \sim (\tfrac{1}{2}\pi z)^{\frac{1}{2}}(\tfrac{1}{2}z)^\lambda\frac{\lambda^{-1}}{\Gamma(1+\lambda)} \quad \text{as} \quad z\to 0 \quad \text{Re}\,\lambda > 0. \tag{3.5}$$

It follows from (3.4), (3.5), and (1.3) that

$$\chi_\lambda(z) = \lambda\int_0^z d\rho\rho^{-2}L_V^W(z,\rho)\psi_\lambda(\rho) \quad \text{Re}\,\lambda > 0. \tag{3.6}$$

When V is equal to 1, $\psi_\lambda(\rho)$ reduces to $s_\lambda(\rho)$ and (3.6) is a Mellin transform. (3.1) reduces to a power expansion and $L(z,z')$ to a standard generating function for the functions $\chi_\lambda(z)$.

Example: The simplest example concerns potential 1 for the base and potential 0 for the wavefunctions to be studied. The wavefunctions reduce, respectively, to $s_\lambda(z)$ given by (1.38) and to $v_\lambda(z)$ given by (1.4),

whereas the interpolation coefficients are given by (2.31). Using well-known formulas,[24] we therefore obtain for $K_1^0(z, z')$ and $L_1^0(z, z')$

$$K_1^0(z, z') = z \sum_1^\infty n(\tfrac{1}{2}z')^n \frac{1}{n!} J_n(z)$$

$$= \tfrac{1}{2}(zz')^{\frac{1}{2}} \frac{z' J_1\{z[1 - (z'/z)]^{\frac{1}{2}}\}}{[1 - (z'/z)]^{\frac{1}{2}}} \quad (3.7)$$

$$L_1^0(z, z') = (zz')^{\frac{1}{2}} J_0\{z[1 - (z'/z)]^{\frac{1}{2}}\}. \quad (3.8)$$

It is easy to verify (3.3) where $\psi_\lambda(z)$ reduce to $s_\lambda(z)$ and the integral in the right-hand side is nothing but Sonine's first integral.

4. INTERPOLATION FORMULAS FOR JOST FUNCTIONS

We now define the class C of potentials for which the coefficients $\gamma_\mu(W, 0)$ are bounded by $C\mu^{\frac{1}{2}}$ for any μ. If the potential belongs to C, it follows from a study previously done by the author[25] that series (1.24) converges uniformly as $z \to \infty$. We define the Jost functions through the asymptotic behavior of the regular wavefunction:

$$\chi_\lambda(r) \sim (2i)^{-1}[f_1(\lambda)e^{ir} - f_2(\lambda)e^{-ir}]. \quad (4.1)$$

Replacing the functions in (1.24) by their asymptotic behavior, we obtain, for the Jost functions, interpolation formulas which we have already written in I:

$$f_1(\lambda) = e^{-i\frac{1}{2}\pi(\lambda-\frac{1}{2})} - \sum_{\mu \in S} \frac{\sin(\lambda - \mu)\pi/2}{\lambda^2 - \mu^2} \gamma_\mu(W, 0) f_1(\mu), \quad (4.2)$$

$$f_2(\lambda) = e^{i\frac{1}{2}\pi(\lambda-\frac{1}{2})} - \sum_{\mu \in S} \frac{\sin(\lambda - \mu)\pi/2}{\lambda^2 - \mu^2} \gamma_\mu(W, 0) f_2(\mu). \quad (4.3)$$

Other interpolation formulas can be derived, with the same assumption, through formula (2.32) or derived formulas. Let us prove now that these interpolation formulas can be derived from the following assumptions, in which we are interested, for instance, in $f_1(\lambda)$.

4.1. Assumptions

I. The only singularities of $f_1(\lambda)$ are simple poles for negative rational values of λ.

II. It is possible to find a number m equal to $\frac{1}{2}$ or to a positive integer, such that

$$F(\lambda) = \frac{\sin 2m\pi\lambda}{2m \sin \pi\lambda} f_1(\lambda) \quad (4.4)$$

is an entire function.

III. $F(\lambda)$ is of order 1 and finite type, equal to $(2m - \frac{1}{2})\pi$.

IV. $F(\lambda)$ is bounded on the real axis.

V. For positive integer n and not multiple of $2m$, the ratio $F(-n/2m)/f_1(n/2m)$ is real, and goes to zero faster than $Cn^{-\epsilon}$ as $n \to \infty$.

VI. For positive integer q, the ratio $F(-q)/(-1)^{2mq} f_1(q)$ is real and goes to 1 faster than $1 + Cq^{-\epsilon}$ as $q \to \infty$.

VII. From the previous assumptions, we prove that

$$f^+(\lambda) \equiv \frac{f_1(\lambda) + f_1(-\lambda)}{2 \cos(\lambda\pi/2)} \to C_1 \quad \text{as} \quad |\operatorname{Im}\lambda| \to \infty, \quad (4.5)$$

$$f^-(\lambda) \equiv \frac{f_1(\lambda) - f_1(-\lambda)}{2 \sin(\lambda\pi/2)} \to C_2 \quad \text{as} \quad |\operatorname{Im}\lambda| \to \infty. \quad (4.6)$$

Assumption VII is

$$C_1 = iC_2 = e^{i\pi/4}. \quad (4.7)$$

4.2. Derivation of the Formula

From I, II, V, and VI, we can deduce the following formulas, where n is not a multiple of $2m$

$$\begin{cases} F(-n/2m) = (-1)^n s(n/2m) f_1(n/2m), \\ F(-q) = (-1)^{2mq}(1 + s(q)) f_1(q), \\ F(n/2m) = 0, \\ F(q) = (-1)^{(2m-1)q} f_1(q), \end{cases} \quad (4.8)$$

where $s(\mu)$ is a real function, bounded by $C\mu^{-\epsilon}$ as $\mu \to \infty$.

Assumptions I to IV enable us to apply now the Lagrange–Valiron formula (2.2) to the function

$$f(\lambda) \equiv (2m)^{-1} \sin 2m\pi\lambda f^-(\lambda) \quad (4.9)$$

which furthermore is an odd function, of order $2m\pi$. Dividing both sides of the result by $(2m)^{-1} \times \sin 2m\pi\lambda$, we get

$$f^-(\lambda) = \frac{f'(0)}{\pi} + \frac{f(0)}{\pi\lambda} + 2m\lambda \sum_{n \neq 0} \frac{(-1)^n f(n/2m)}{n\pi(\lambda - n/2m)}. \quad (4.10)$$

Taking into account the parity, and calculating $f(n/2m)$ with the help of (4.8), we obtain the following formula:

$$f^-(\lambda) = \frac{f'(0)}{\pi} - 4m\lambda^2 \sum_{n=1}^\infty \frac{\cos(\pi n/4m) s(n/2m) f_1(n/2m)}{n\pi(\lambda^2 - n^2/4m^2)}. \quad (4.11)$$

Let $\operatorname{Im}\lambda \to \infty$. The bounds of $s(\mu)$ enable us to state

[24] Ref. 20, formula 7.10 (16).
[25] See in particular Sec. 2 of Ref. 5.

that the right-hand side of (4.11) goes to a definite limit.[21] We have therefore shown (4.5), and formula (4.7) yields $f'(0)$,

$$\pi^{-1}f'(0) = -ie^{i\pi/4} + 4m\sum_{n=1}^{\infty}\frac{\cos(n\pi/4m)s(n/2m)f_1(n/2m)}{n\pi}. \tag{4.12}$$

Inserting (4.12) in (4.11) gives the interpolation formula of $f^-(\lambda)$

$$f^-(\lambda) = -ie^{i\pi/4} - \sum_{n=1}^{\infty}\frac{n\cos(\pi n/4m)s(n/2m)f_1(n/2m)}{m\pi(\lambda^2 - n^2/4m^2)}. \tag{4.13}$$

A similar argument can be used to derive the interpolation formula of $f^+(\lambda)$

$$f^+(\lambda) = e^{i\pi/4} + \sum_{n=1}^{\infty}\frac{n\sin(\pi n/4m)s(n/2m)f_1(n/2m)}{m\pi(\lambda^2 - n^2/4m^2)}. \tag{4.14}$$

Combining (4.13) and (4.14) yields readily (4.2) provided that

$$\gamma_\mu(W, 0) = 2_\mu\pi^{-1}s(\mu) \quad (\mu = n/2m). \tag{4.15}$$

The analytic properties attached to the Jost functions of potentials *belonging to* $\mathcal{C}$ are therefore a key to formulas shown through a tedious method of Sec. 1.[26] They also enable us to answer some questions. For instance, is it possible, in the case of even potentials, to find a Jost function with a finite number N of zeros? The answer is no. Since the Jost function is entire, of order 1, type $\pi/2$, Hadamard's theorem would yield for it a representation of the form

$$\exp[P_1(\lambda)]P_N(\lambda), \tag{4.16}$$

where $P_n(\lambda)$ stands for a polynomial of order n. A glance at the asymptotic behavior of $f_{1,2}(\lambda)$ shows that (4.16) necessarily reduces to $Ce^{\mp i\pi/2\lambda}$, which holds only for a potential equal to zero.

For any potential it is also very easy to show, from the above Assumptions, the well-known symmetry relation

$$f_1(\lambda)f_1^*(-\lambda^*) - f_1(-\lambda)f_1^*(\lambda^*) = -2i\sin\pi\lambda. \tag{4.17}$$

So as to prove this point, we only need to construct the function

$$(\sin 2m\pi\lambda/\sin\pi\lambda)[f_1(\lambda)f_1^*(-\lambda^*) - f_1(-\lambda)f_1^*(\lambda^*)]$$

and see that this entire function is of order 1 and type $2m\pi$, bounded on the real axis, and, according to (4.8), is equal to zero at all the points $n/2m$ so that [taking into account its asymptotic behavior and formula (2.2)], it is just equal to $-2i\sin 2m\pi\lambda$.

Entire function properties can still be used in order to derive an integral representation of the Jost function. This follows from the theorem of Paley and Wiener.[27]

Theorem: The entire function $g(z)$ is of exponential type τ and belongs to L^2 on the real axis if and only if

$$g(z) = \int_{-\tau}^{\tau} e^{izt}\phi(t)\,dt. \tag{4.18}$$

When $h(z)$ fulfills conditions of application of the above theorem but is only bounded on the real axis, the theorem can be applied to $g(z) = z^{-1}[h(z) - h(0)]$ (this remark is in fact the starting point of the Lagrange–Valiron theorem). It is easy to infer from (4.8) and (2.2) the expression of $\phi(t)$

$$\phi(t) = \frac{1}{2\tau}\sum_{-\infty}^{+\infty} e^{in\pi/\tau}g\left(\frac{n\pi}{\tau}\right) \tag{4.19}$$

or

$$\phi(t) = \frac{1}{2\tau}\sum_{n\neq 0} e^{in\pi/\tau}\left(\frac{h(n\pi/\tau) - h(0)}{n\pi/\tau}\right) + \frac{1}{2\tau}h'(0). \tag{4.20}$$

These results can be applied to various functions studied above. Application to $F(\lambda)$ yields in the case of even potentials ($m = \frac{1}{2}$) a particularly simple formula:

$$f_1(\lambda) = f_1(0) + \lambda\int_{-\pi/2}^{\pi/2} e^{i\lambda t}\phi(t)\,dt, \tag{4.21}$$

where $\phi(t)$ is given by (4.19).

It is interesting to investigate the classes of potentials for which some of the above assumptions hold. For any potential, the Jost function can be defined as the Wronskian of the regular solution and the Jost solution. If $zV(z)$ is analytic in some circle centered at $z = 0$, the Jost solution is an even entire function of λ, whereas the regular solution is meromorphic, with poles at half-negative integers. Assumptions I, II, V, VI, and VII hold, but we have no information on III and IV.

Unfortunately, we have only heuristic arguments for the behavior of the Jost function as $\mathrm{Re}\,\lambda \to -\infty$. Application of these arguments to potentials of the Yukawa type is sketched, for instance, in the book of Newton,[28] and gives an asymptotic behavior for the Jost functions. For finite sums or integrals of Yukawas,

[26] In a recent preprint, R. G. Newton gives also a direct derivation of formula (4.2) starting from the assumption that a function associated to the Jost functions has a Mittag–Leffler expansion without remainder. We are indebted to Professor Newton for sending us a preprint of his work.

[27] Ref. 19, p. 103.

[28] R. G. Newton, *The Complex j-Plane* (W. A. Benjamin, Inc., New York, 1964), p. 44.

$zV(z)$ is analytic, and if the quoted arguments are correct, the Jost function is of order 1 and finite type. Severe assumptions are necessary to ensure a type equal to π, and the boundedness on the real axis can occur only for special potentials. We conclude from this that the common part of the class $\mathcal{C}$ and the potentials of Yukawa type reduces very likely to quite special potentials, if any.

5. ASYMPTOTIC INTERPOLATION COEFFICIENTS

We now derive the interpolation coefficients obtained if we *take for granted* the so-called "asymptotic" or "JWKB" approximation. Let us first recall[29] some results of this method. We first assume that λ is a real positive number. The Schrödinger equation can be written in the form

$$d^2y/dz^2 + [\phi^2(z) + \chi(z)]y = 0, \tag{5.1}$$

where

$$\begin{cases} \phi^2(z) = 1 - V(z) - \lambda^2 z^{-2} \\ \chi(z) = \frac{1}{4}z^{-2}. \end{cases} \tag{5.2}$$

Let us define z_0 as the positive solution of the equation

$$z_0[1 - V(z_0)]^{\frac{1}{2}} = \lambda. \tag{5.3}$$

We assume that z_0 is correctly and unambiguously defined by equation (5.3), that is to say, we restrict our study to the values of λ for which this is true. Let $\mathcal{K}(\lambda)$ be this interval. Since $V(z)$ goes to zero as $z \to \infty$, it is clear that $\mathcal{K}(\lambda)$ extends from a positive value λ_0 to $+\infty$. For a given attractive and bounded static potential, we know that, if the energy is larger than a "critical energy" E_c, λ_0 reduces to zero, although this precise value may be excluded from $\mathcal{K}(\lambda)$. Let us now define a function $\bar{z}$ of z through the following equation:

$$\int_\lambda^{\bar{z}} \bar{\phi}(\tau)\, d\tau = \int_{z_0}^{z} \phi(\tau)\, d\tau, \tag{5.4}$$

where

$$\bar{\phi}(\tau) = (1 - \lambda^2\tau^{-2})^{\frac{1}{2}}. \tag{5.5}$$

The determinations of $\phi(\tau)$ and $\bar{\phi}(\tau)$ are chosen in such a way as to give the same function for a potential equal to zero. For instance, we may state that, as z is larger than z_0, $\phi(\tau)$ and $\bar{\phi}(\tau)$ are positive, and they get a phase $+\pi/2$ as z decreases and crosses z_0.

It is easy to derive from (5.1) the following equation:

$$(d^2/d\bar{z}^2)\bar{y}(\bar{z}) + [\bar{\phi}^2(\bar{z}) + \tfrac{1}{4}\bar{z}^{-2} + R(\bar{z})]\bar{y}(\bar{z}) = 0, \tag{5.6}$$

where

$$\bar{y}(\bar{z}) = (d\bar{z}/dz)^{\frac{1}{2}}y(z) = [\phi(z)/\bar{\phi}(\bar{z})]^{\frac{1}{2}}\, y(z) \tag{5.7}$$

and

$$R(\bar{z}) = \tfrac{1}{2}\{z, \bar{z}\} + \tfrac{1}{4}[z^{-2}(dz/d\bar{z})^2 - \bar{z}^{-2}]. \tag{5.8}$$

For large energy E, or for large values of λ^2, $R(z)$ is of the order of E^{-1}, or λ^{-2}. Equation (5.6) can then be considered as a perturbed equation, the perturbation being $R(\bar{z})$, and the unperturbed equation,

$$(d^2/d\bar{z}^2)\bar{y}_0(\bar{z}) + \{1 - [(\lambda^2 - \tfrac{1}{4})/\bar{z}^2]\}\bar{y}_0(\bar{z}) = 0. \tag{5.9}$$

The Liouville method enables one to transform (5.6) into a Volterra equation involving two independent solutions of (5.9) and the perturbation $R(\bar{z})$. It is easy to study Eq. (5.8) in this way for real positive λ. Outside the real positive axis in the λ plane, the nonuniform limits of Bessel functions make this study very difficult if we are not satisfied with heuristic arguments or with a study restricted to very special classes of potentials. We conclude that the validity of the "asymptotic approximation," obtained by using the regular solution of (5.9) in place of the regular solution of (5.6), is difficult to justify for real negative values of λ. We therefore use the result of this approximation only as a starting point for giving a set of coefficients c_l, and we investigate their relation with the potential by a different method. For simplicity's sake we limit our study to the case of potentials analytic and even in a circle centered at the origin. It is not difficult to make a slightly more intricate study of noneven analytic potentials.

In order to compute the interpolation coefficients, we use formula (1.46). We should therefore be able to continue analytically the regular wavefunction from λ to $-\lambda$. Let $z_0(\lambda)$ be the inverse function of $\lambda(z_0)$ as defined by (5.3). We have to assume that $z_0(\lambda)$ is analytic in some connected domain including λ and $-\lambda$; that is to say, we assume that $\mathcal{K}(\lambda)$ extends in the λ plane in such a way as to include the values of λ larger than λ_0 on the positive real axis and the values smaller than $-\lambda_0$ on the negative one. With the parity assumption on the potential, this is certainly true for $E > E_c$ provided that the potential can be continued analytically in a strip including the real axis. Even for $E < E_c$, this assumption is very weak. Let us now define the function $H(x)$ equal to

$$H(x) = \int_0^x \left[\frac{1}{t} - \frac{z_0'(t)}{z_0(t)}\right] dt = \log\left[\frac{x}{z_0(x)}\right]. \tag{5.10}$$

With the assumptions on the potential, $H(x)$ is an even analytic function in a strip including the whole real axis, and has the same asymptotic behavior as the potential, as $x \to \pm\infty$. Use of the function $H(x)$ enables us to put (5.4) in the form

$$\int_{\lambda(z)}^{\lambda} (\lambda^2 - x^2)^{\frac{1}{2}}H'(x)\, dx = \int_{\lambda(z)}^{\bar{z}} (\lambda^2 - x^2)^{\frac{1}{2}}x^{-1}dx \tag{5.11}$$

[29] For a detailed study of the asymptotic approximation as given in this form, see, for example, P. C. Sabatier, Nuovo Cimento 37, 1180 (1965).

or

$$\int_{\lambda}^{\lambda(z)} (x^2 - \lambda^2)^{\frac{1}{2}} H'(x)\, dx = \int_{\bar{z}}^{\lambda(z)} (x^2 - \lambda^2)^{\frac{1}{2}} x^{-1}\, dx. \tag{5.12}$$

Using (5.11) it is easy to show that

$$z/\bar{z} \to N(\lambda), \quad \text{as} \quad z \to 0, \tag{5.13}$$

where

$$N(\lambda) = \exp\left[-\lambda^{-1}\int_0^{\lambda} (\lambda^2 - x^2)^{-\frac{1}{2}} H(x) x\, dx\right]. \tag{5.14}$$

From (5.13) and (5.14) it follows that $\psi_\lambda(z)$, normalized as in (1.3), is equal to

$$\psi_\lambda(z) = [N(\lambda)]^\lambda \left(\frac{\pi}{2}\bar{z}\right)^{\frac{1}{2}} \left[\frac{\lambda^2 \bar{z}^{-2} - 1}{\lambda^2 z^{-2} + V(z) - 1}\right]^{\frac{1}{4}} J_\lambda(\bar{z}). \tag{5.15}$$

$\bar{z}$ is an even function of λ. Using this remark and the symmetry relations of Bessel functions, it is very easy to compare $\psi_l(z)$ and $\psi_{-l}(z)$ and to derive the value of γ_l:

$$1 + \frac{\pi}{2l}\gamma_l = \exp\left[2\int_0^l (l^2 - x^2)^{-\frac{1}{2}} H(x) x\, dx\right]. \tag{5.16}$$

Equation (5.12) enables us to show that $(z - \bar{z})$ goes to a constant as $z \to \infty$. Taking into account this result and (5.15) yields the Jost function

$$f_l(\lambda) \sim \exp\left[-\int_0^{\lambda} (\lambda^2 - x^2)^{-\frac{1}{2}} H(x) x\, dx + i\int_{\lambda}^{\infty} (x^2 - \lambda^2)^{-\frac{1}{2}} H(x) x\, dx - i(\lambda - \tfrac{1}{2})\frac{\pi}{2}\right]. \tag{5.17}$$

5.1. Physical Meaning of the Asymptotic Coefficients

Now, what is the meaning of the "asymptotic" coefficients (5.16)? In order to justify the semiclassical approximation, we need only weak assumptions on the potential: analytical properties near the real axis, and bounds for the first or the two first derivatives. Potentials may have all kinds of singularities outside the real axis and still fulfill these assumptions. For those potentials, the coefficients $\gamma_l(W, 0)$ may increase faster than $(2l)!$ as $l \to \infty$, whereas the "asymptotic" coefficients are clearly bounded by $Cl^{1-\epsilon}$, provided the potential be bounded by $Cr^{-(1+\epsilon)}$ as $r \to \infty$. The "asymptotic" coefficients cannot therefore be considered as asymptotic approximations to the actual coefficients.

Let us then study the relation between the "asymptotic" coefficients and the JWKB phase shifts. We limit our study to the so-called "first-order JWKB approximation," in which it is possible to use the linear approximation for the exponentials in (5.16) and (5.17). This supplementary approximation is valid if $H(\lambda)$ is small enough, for instance if the energy is large, and it can be considered as a Born approximation for smooth potentials.

Using linear approximations in formula (4.2) yields the following connection between the δ_l and the interpolation coefficients of the even potentials in class $\mathcal{C}$ which can be constructed from them:

$$\delta_l = \sum_{l'} N_l^{l'} \gamma_{l'}, \tag{5.18}$$

where $N_l^{l'} = [l'^2 - l^2]^{-1}$ for $|l - l'|$ odd, 0 otherwise.

Use of the matrix inverse of N, given previously by Newton[3] and by the author,[30] yields the value of the coefficient γ_l:

$$\gamma_{2p} = \frac{16}{\pi^2}(2p)^2 v_{2p} \sum_n \frac{1}{(2p)^2 - (2n+1)^2}\, \delta_{2n+1}, \tag{5.19}$$

$$\gamma_{2p+1} = \frac{16}{\pi^2}(2p+1)^2 \sum_n \frac{v_{2n}}{(2p+1)^2 - (2n)^2}\, \delta_{2n}, \tag{5.20}$$

where

$$v_{2n} = 1 - \tfrac{1}{2}\delta_0^n. \tag{5.21}$$

From (5.17), using a well-known formula,[31] we can write the phase shifts in the form

$$\delta_{l'} = \int_0^{\infty} \cos l' t\, dt \int_0^{\infty} J_0(xt) H(x) x\, dx. \tag{5.22}$$

Insertion of (5.22) in (5.19) and (5.20) leads us to Fourier series, which are easily seen to be, respectively, the Fourier series of the continuous even functions $[(\sin 2pt)/2p]\,\text{sign}\, t$ and $[(\sin(2p+1)t)/2p + 1]\,\text{sign}\, t$. (This result is valid, up to a multiplicative constant, inside the interval $-\pi, \pi$.) We therefore obtain

$$\gamma_l = \frac{4l}{\pi}\left\{\int_0^{\pi} - \int_{\pi}^{2\pi} + \cdots (-1)^n \times \int_{n\pi}^{(n+1)\pi} + \cdots\right\} \sin lt\, dt\, \mathcal{K}(t), \tag{5.23}$$

where

$$\mathcal{K}(t) = \int_0^{\infty} J_0(xt) H(x) x\, dx. \tag{5.24}$$

Now, it follows[32] from the assumptions on the semiclassical approximation that $H(x)$ being a smooth function, $\mathcal{K}(t)$ decreases rapidly and goes to zero, so that we can replace in (5.23) the integrals by $\int_0^{\infty}$, and

[30] See Ref. 4. In order to have $\gamma_0 = 0$, we use a prescription similar to that used in Ref. 1, Sec. 2.2.

[31] Bateman Manuscript Project, *Integral Transforms* (McGraw-Hill Book Company, Inc., New York, 1954), formula 1.12 (1).

[32] See Ref. 29, Sec. 1, for a description of the function $H(x)$.

get, using a well-known formula,[33]

$$\gamma_l = \frac{4l}{\pi}\int_0^l (l^2 - x^2)^{-\frac{1}{2}} H(x)\,dx, \qquad (5.25)$$

which is precisely the first-order approximation for (5.16).

The γ_l's defined by (5.16) are therefore, in the first-order approximation, the interpolation coefficients of an even potential V_1 of class $\mathcal{C}$ which leads to phase shifts equal, in the same approximation, to the phase shifts given by the actual potential V. Since the inverse scattering problem has one solution and only one in the even potentials of $\mathcal{C}$, there exists a potential V_2 which yields exactly the JWKB phase shifts and which differs from V_1 at most by second-order terms. If the initial potential belongs to class $\mathcal{C}$, V_2 is an approximation of it up to the second order in the asymptotic parameter ($E^{-\frac{1}{2}}$ or λ^{-1}).

Suppose now we solve the inverse scattering problem in the framework of the semiclassical theory, using an even entire interpolation function for the phase shifts and computing the potential V_3 as the fractional derivative of this interpolation. In the first-order approximation, V_3 is equal to V_2, and we know from a previous study[34] that V_3 is an asymptotic approximation of V. We have therefore shown that if the conditions of the first-order JWKB approximation are satisfied, there exists a potential (V_2) which yields exactly the JWKB phase shifts and which is an approximation of the initial potential. Let us recall that we came to the same result by assuming both the validity of the Born approximation and a sufficiently large energy. It would be interesting to extend our result to the general JWKB (and not only first-order) approximation. We found heuristic arguments for this extension, but we cannot be satisfied with them.

6. ELEMENTARY DERIVATION OF THE INTERPOLATION FORMULAS

The interpolation formulas obtained in Sec. 2 for products of wavefunctions can be obtained by elementary manipulations of the wave equation (1.2). The simplest case occurs when the functions $\chi_{-\lambda}(z)$ in (2.30) reduce to powers of z. In the case of even potentials, the interpolation formula reduces then to (2.5). In this section we derive (2.5) by a very simple method. This method enables us also to derive exact formulas for the solutions of Eq. (1.2) and for the interpolation coefficients. As expected, these formulas cannot be used for practical computations. It is then shown how the interpolation formulas of type (2.5) can be obtained for potentials analytic in $z^{1/m}$. To end this section, we sketch an elementary approach to the general interpolation formula (2.30).

In order to show our first point, let us remark that the following equations are derived straightforwardly from (1.2):

$$[T_z - 2lz^{-1}(d/dz)]z^{-1}\chi_l(z)s_l(z) = 0 \qquad (6.1)$$

$$[T_z + 2\lambda z^{-1}(d/dz)]z^{-1}\chi_\lambda(z)s_{-\lambda}(z) = 0, \qquad (6.2)$$

where

$$T_z \equiv (d^2/dz^2) + \mathbf{W}(z) + z^{-1}(d/dz) \quad [\text{with } \mathbf{W}(z) = 1 - W(z)] \qquad (6.3)$$

and the functions $s_\lambda(z)$ are defined by (1.38).

Assume now that a sequence $\{a_n\}$ can be constructed (with $a_0 = 1$), for which the relation

$$\frac{d}{dz}\, z^{-1}\sum_{n=0}^{\infty} a_n\chi_n(z)s_n(z) = 0 \qquad (6.4)$$

holds inside a circle ω centered at the origin and with a nonvanishing radius R. We assert that the following relation holds in ω:

$$\frac{\pi\lambda}{\sin\lambda\pi}\chi_\lambda(z)s_{-\lambda}(z) = \sum_{n=0}^{\infty}\frac{\lambda}{\lambda+n}\, a_n\chi_n(z)s_n(z) \qquad (6.5)$$

or

$$\Gamma(1+\lambda)\left(\frac{z}{2}\right)^{-\lambda}\chi_\lambda(z) = \sum_{n=0}^{\infty}\frac{\lambda}{\lambda+n}\,\frac{a_n}{n!}\left(\frac{z}{2}\right)^n\chi_n(z). \qquad (6.6)$$

The proof is made in two steps. First, applying $[T_z + 2\lambda z^{-1}(d/dz)]$ to the right-hand side of (6.5), using (6.4), and taking into account the boundary conditions, we easily see that provided the right-hand side of (6.5) be an analytic function, it is equal to the left-hand side. The problem reduces therefore to constructing $\{a_n\}$ and proving the existence of ω. Now, we know from a previous analysis[9] that if $W(z)$ is an even analytic function of z inside a circle Ω centered at the origin and with a radius Γ, the functions $z^{-1}\chi_n(z)s_n(z)$ are even analytic functions in this circle and have the expansion

$$(\tfrac{1}{2}\pi z)^{-1}\chi_n(z)s_n(z) = N(n)\sum_{q=n}^{\infty}\chi_n^q z^{2q}, \qquad (6.7)$$

where

$$N(n) = (2^n n!)^{-2} \quad \text{and} \quad \chi_n^n = 1. \qquad (6.8)$$

In (6.7), the expansion coefficients are bounded by

$$|\chi_n^q| < M(R)q^{-1}R^{-2(q-n)}, \quad q > n, \qquad (6.9)$$

where R is any positive number smaller than Γ, i.e.,

$$R = \Gamma - \epsilon.$$

[33] Ref. 31, formula 2.12 (1).

[34] We suggested this study in Ref. 29 and we performed it in detail, up to the third order of asymptotic parameters, in our doctoral dissertation (Faculté des Sciences, Orsay No. 153, 1966).

It is easy to calculate the coefficients a_n by induction from the following equations which one deduces from (6.7):

$$\sum_0^n a_l N(l)\chi_l^n = \delta_0^n, \tag{6.10}$$

where δ_0^n is the Kronecker symbol. Obviously a_n can be calculated exactly as an uninteresting Cramer's determinant whose elements are taken in the χ_l^n. If positive numbers $\bar{\chi}_l^n$ are upper bounds for the $|\chi_n^l|$'s, then the following relations define upper bounds $\bar{a}_l$ for the $|a_l|$'s:

$$\bar{a}_0 = 1 \quad N(n)\bar{a}_n = \sum_0^{n-1} N(l)\bar{a}_l\bar{\chi}_l^n. \tag{6.11}$$

Using (6.9) for $\bar{\chi}_n^l$, and solving (6.11), we get

$$\begin{cases} \bar{a}_0 = 1 \\ N(n)\bar{a}_n = R^{-2n}\prod_{p=0}^{n-1}\dfrac{p + N(R)}{p+1} < CR^{-2n}. \end{cases} \tag{6.12}$$

From (6.9), we easily obtain the bound for $\chi_n(z)s_n(z)$,

$$|(\tfrac{1}{2}\pi z)^{-1}\chi_n(z)s_n(z)| < N(n)\,|z|^{2n}(1 - R^{-2}\,|z|^2)^{-1}. \tag{6.13}$$

The series (6.4) and (6.5) are therefore uniformly convergent and define an analytic function inside the circle $|z| < R$. Q.E.D.

6.1. Formal Expansions for the Partial Wavefunctions

We use hereafter for $(\frac{1}{2}\pi z)^{-1}\chi_n(z)s_n(z)$ the condensed notation $\xi_n(z^2)$ or, defining a new variable x equal to z^2, $\xi_n(x)$. We use also for convenience the function

$$f(x) = \tfrac{1}{4}\mathbf{W}(z) \tag{6.14}$$

and define the operator Δ which transforms any analytic function $g(x)$ according to the following scheme

$$\Delta g(x) \equiv x\frac{d}{dx}g(x) + \int_0^x f(t)g(t)\,dt. \tag{6.15}$$

From (6.1), it is easy to see that $\xi_n(x)$ is an eigenfunction of Δ, which corresponds to the eigenvalue n:

$$\Delta\xi_n(x) = n\xi_n(x). \tag{6.16}$$

Since it follows from (6.4) that

$$1 = \sum_{n=0}^{\infty} a_n\xi_n(x), \tag{6.17}$$

we see that $\Delta 1$ is equal to

$$\sum_{n=0}^{\infty} na_n\xi_n(x) \quad \text{or to} \quad \int_0^x f(t)\,dt,$$

a result equivalent to (2.7).

It is important to notice that any number of applications of Δ to a function $\xi_n(x)$ leads us to a function whose zero for $x = 0$ is at least of order n. As a result, it is easy to derive, from (6.16) and (6.17), the formulas

$$a_nN(n) = \left[\frac{\Delta(\Delta-1)\cdots(\Delta-n+1)1}{n!}\right]_n, \tag{6.18}$$

$$a_pN(p)\alpha_p^n = \left[\frac{\Delta\cdots(\Delta-p+1)(\Delta-p-1)\cdots(\Delta-n)1}{(-1)^{n-p}p!(n-p)!}\right]_n,$$

where the subscript n outside the bracket means that we take the coefficient of x^n in the Taylor expansion.[35] The same result could be obtained by calculating $[1]_n, [\Delta 1]_n, \cdots, [\Delta^n 1]_n$, with the help of (6.17), and solving the Cramer's system thus obtained. The solution is easy to obtain, since the determinants are of the Van der Monde type.

Let us now define the following functions:

$$g_n(x) = \frac{\Delta(\Delta-1)\cdots(\Delta-n+1)1}{n!} = \frac{\Gamma(-\Delta+n)1}{\Gamma(-\Delta)n!}(-1)^n, \tag{6.19}$$

where the last formula is only a formal way of writing. From (6.16) and (6.17), we derive straightforwardly

$$g_n(x) = \sum_{l=n}^{\infty}\frac{\Gamma(l+1)}{\Gamma(l+1-n)n!}a_l\xi_l(x) \tag{6.20}$$

which holds at least in the circle $x < \Gamma$. Let us then calculate

$$\sum_{q=p}^{\infty}\frac{\Gamma(1+q)}{\Gamma(1-p+q)}(-1)^q g_q(x). \tag{6.21}$$

Inserting (6.20) in (6.21) and performing the algebraic sums with the help of Gauss theorem yield the result which is proportional to $\xi_p(x)$. Using (6.19) we write

$$a_p\xi_p = \frac{(-1)^p}{p!}\sum_{q=p}^{\infty}\frac{1}{\Gamma(1+q-p)}\frac{\Gamma(-\Delta+q)}{\Gamma(-\Delta)}1. \tag{6.22}$$

This formula should be considered in general as meaning only equality between the Taylor expansions of both sides.

If we notice that the $g_n(x)$ are equivalent to x^n as $x \to 0$, it is easy to derive formula (6.18) from (6.22). Formal manipulations of (6.22) may lead one to interesting results—to be proved in another way.

[35] Notice that (6.18) expresses the interpolation coefficients as polynomials of powers of the potential strength. Similar formulas and results can be obtained for analytic potentials.

Algebraic manipulations yield, for instance

$$h_\lambda(x) = \sum_0^\infty \frac{\lambda}{\lambda + n} a_n \xi_n(x) = (\lambda + \Delta)^{-1}\lambda. \quad (6.23)$$

The result is therefore a solution of the equation

$$(\lambda + \Delta)h_\lambda(x) = \lambda. \quad (6.24)$$

Differentiating both sides of (6.24) and taking into account the boundary conditions lead us to (6.5). Bessel functions hypergeometric series can also be easily obtained from (6.22) using the appropriate Δ.

6.2. Extension of the Results for Analytic Potentials of $z^{1/m}$

Equation (6.1) thus reads

$$\begin{cases} \left[\frac{d^2}{dz^2} + z^{-1}\frac{d}{dz} + V(z^{m^{-1}})\right]\xi_\lambda(z) = 2\lambda z^{-1}\frac{d}{dz}\xi_\lambda(z), \\ \xi_\lambda(z) \sim z^{2\lambda} D(\lambda) \quad \text{as} \quad z \to 0. \end{cases} \quad (6.25)$$

Using the new variable x equal to $z^{1/(2m)}$, we obtain the equation

$$\begin{cases} \left[\left(\frac{d^2}{dx^2} + x^{-1}\frac{d}{dx}\right) + 4m^2 x^{4m-2} V(x^2)\right]\xi_\lambda(x^{2m}) \\ \qquad = 4\lambda m x^{-1}\frac{d}{dx}\xi_\lambda(x^{2m}) \quad (6.26) \\ \xi_\lambda(x^{2m}) \sim D(\lambda) x^{4\lambda m} \quad \text{as} \quad x \to 0. \end{cases}$$

It follows from (6.26) that if $\xi_\mu(x)$ are the solutions of (6.26) corresponding to the eigenvalue μ with the usual normalization, we have

$$\xi_\lambda(x^{2m}) = \xi_{2\lambda m}(x) \cdot \frac{N(\lambda)}{N(2\lambda m)}. \quad (6.27)$$

The solutions $\xi_\lambda(x^{2m})$ can therefore be related to the solutions for an even potential, for which the previous interpolation holds. This yields readily the generalization of (6.6) valid when the potential is an analytic function of $z^{1/m}$.

6.3. Elementary Approach to the General Formula

The general interpolation formula (2.32) can be derived by elementary but somewhat tedious methods which we only sketch here. From the equations

$$\begin{cases} \psi_\lambda'' = [(\lambda^2/r^2) - \mathbf{W}_1]\psi_\lambda \\ \chi_\lambda'' = [(\lambda^2/r^2) - \mathbf{W}_2]\chi_\lambda \end{cases}, \quad (6.28)$$

it is easy to prove that the function y equal to $\psi_{\pm\lambda}\chi_{\pm\lambda}$ satisfies the following integrodifferential equation:

$$y''' + 2(\mathbf{W}_1 + \mathbf{W}_2)y' + (\mathbf{W}_1' + \mathbf{W}_2')y + (\mathbf{W}_2 - \mathbf{W}_1)\left[C + \int_0^r (\mathbf{W}_2 - \mathbf{W}_1)y(\rho)\,d\rho\right] = (4\lambda^2/r^2)(y' - r^{-1}y), \quad (6.29)$$

where the constant C is equal to zero for a solution equal to zero at the origin, to λ for $\psi_\lambda\,\chi_{-\lambda}$.

Equation (6.29) can be as well written in the form

$$\mathcal{E}y \equiv (4\lambda^2/r^2)[y' - r^{-1}y] = \lambda^2 \mathcal{F} y. \quad (6.30)$$

Assume now that it is possible to define two sequences $\{a_n\}$ and $\{b_n\}$ such that

$$\frac{\pi}{2} z = \psi_0\chi_0 + \sum_1^\infty \psi_n\chi_n b_n, \quad (6.31)$$

$$\frac{\pi}{2} K_V^W(z, z) = \sum_1^\infty \psi_n\chi_n a_n. \quad (6.32)$$

Then applying $\mathcal{E}$ to both sides of

$$\mathcal{Y} = \psi_0\chi_0 + \sum_1^\infty \psi_n\chi_n\left[\frac{\lambda}{\lambda^2 - n^2} a_n + \frac{\lambda^2}{\lambda^2 - n^2} b_n\right] \quad (6.33)$$

and using (6.31) and (6.32), it is easy to show that $\mathcal{Y}$ is a solution of (6.29) and that the boundary conditions prove that it is equal to

$$(\lambda/\sin \pi\lambda)\psi_\lambda(z)\chi_{-\lambda}(z).$$

Identification with formula (2.32) requires the following relations:

$$\left.\begin{aligned} b_\mu &= (\pi/2\mu)[\gamma_\mu(W, 1) + \gamma_\mu(V, 1)] \\ a_\mu &= \tfrac{1}{2}\pi[\gamma_\mu(W, 1) - \gamma_\mu(V, 1)] \end{aligned}\right\}. \quad (6.34)$$

7. CONCLUDING REMARKS

We would like to emphasize the fact that the various formulas given in this paper are only *examples* of the interpolation formulas which can be obtained for the wavefunctions or related quantities.

First, the conditions of the Lagrange–Valiron theorem enable one to translate the set of indices in the right-hand side of (2.2), since $f(z + \alpha)$ has obviously the same properties as $f(z)$. It would be easy to derive, for instance, an interpolation formula in which only the half-integral values of λ are included in the set. However, since only the positive half-integral values of λ are "physical," this formula is not very interesting.

From (2.32) or from a formula derived from (2.32) and from a definition of the irregular wavefunction for integer λ analogous to the one of the irregular Bessel function $Y_n(z)$, it is easy to derive an expansion of the irregular wavefunction in terms of the regular wavefunctions and interpolation coefficients.

As a third kind of derived formulas, we notice the exact summation of expansions like

$$\left.\begin{aligned} S_1 &= \sum_\mu a_\mu(\psi_\mu\chi_\mu' - \psi_\mu'\chi_\mu) \\ S_2 &= \sum_\mu b_\mu(\psi_\mu\chi_\mu' - \psi_\mu\chi_\mu) \end{aligned}\right\}, \quad (7.1)$$

where the a_μ's and the b_μ's are defined as in (6.34), and $b_0 = 1$. Use of the Wronskian properties, together with formulas like (6.31) and (6.32) lead us straightforwardly to the results,

$$\left.\begin{aligned} S_1 &= -\frac{\pi}{8}\left\{\int_0^z (W - V)\rho\, d\rho\right\}^2 \\ S_2 &= \frac{\pi}{2}\int_0^z (W - V)\rho\, d\rho \end{aligned}\right\}. \tag{7.2}$$

If the interpolation coefficients are bounded as in Sec. 4, it is possible to obtain sums of Jost functions by letting $z \to \infty$ in those formulas. Results of the same kind are easy to derive either by applying the operator Δ n times to (6.31), (6.32), (7.1), or by using next orders in the asymptotic expansion (1.47).

ACKNOWLEDGMENT

It is a pleasure for the author to thank Dr. K. Chadan for interesting discussions.

APPENDIX

Several times in this paper we identified the coefficients of a given power of $|\lambda|^{-1}$ in two asymptotic expansions which are valid only outside the real axis. This may seem to be a shortcoming since the function involved is *not* an analytic function of λ^{-1} at $\lambda^{-1} = 0$, and since, in no case, we studied thoroughly the right-hand side asymptotic expansion. To ensure our point, let us recall that the expansions to be compared are of the form:

(1) on the left-hand side,

$$a_0 + a_1\lambda^{-1} + \cdots a_n\lambda^{-n} + O(\lambda^{-n-\epsilon}) \quad (|\mathrm{Im}\ \lambda| > \epsilon_0), \tag{A1}$$

where ϵ_0 is any fixed positive number, the remainder $O(\lambda^{-n-\epsilon})$ depending on ϵ_0;

(2) on the right-hand side,

$$\sum_1^\infty \frac{b_q}{\nu^2 - q^2}, \quad \text{where} \quad |b_q| < C\,|q|^{-2r-\epsilon_1}. \tag{A2}$$

Actually the series we encountered are of the form

$$\sum_1^\infty \frac{c_q}{\nu - q}, \tag{A3}$$

but it is easy to reduce their study to that of (A2) with the help of a prescription similar to the one sketched in (6.33) and (6.34). Using now a well-known identity, we can write (A2) in the form

$$\sum_{p=0}^{r-1}\sum_{q=1}^\infty b_q \frac{1}{\nu^2}\left(\frac{q^2}{\nu^2}\right)^p + \sum_1^\infty \frac{1}{\nu^2 - q^2}\, b_q\left(\frac{q^2}{\nu^2}\right)^r. \tag{A4}$$

In a previous paper,[5] we showed that a series

$$S = \sum_{q=1}^\infty \frac{\gamma_q}{\nu^2 - q^2}, \tag{A5}$$

where the $|\gamma_q|$'s are bounded by $Cq^{-\beta}(-1 < \beta < 1)$ is bounded, as $|\nu|$ goes to ∞ outside the real axis, by $(C\,|\nu|^{-1-\beta+\epsilon} + C\,|\nu|^{-2})$, where the constants depend on $|\mathrm{Im}\ \nu|$. Expression (A4) is therefore an asymptotic expansion of form (A1), provided the inequality in (A2) be satisfied. Identification of the coefficients in two expansions of form (A1) is easy to do. Making $|\lambda| \to \infty$ leads us to the identification of a_0. After suppressing a_0 and multiplying both sides by λ, the same device leads us to the identification of a_1, etc.

In Sec. 2 and 6, where we used these results, the coefficients b_q are functions of z. They are bounded, according to (1.7), as z lies in the circle $\mathfrak{D}$, in such a way that $|b_q| < C/(q!)^2$. As a result we can take $r = \infty$ in (A2). However, if we try to use the identification process after letting $z \to \infty$, we have to test inequalities carefully. This has been taken into account in Sec. 4.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

General Space–Momentum Commutation Relation

DAVID M. ROSENBAUM
Institute for Defense Analyses, Arlington, Virginia

(Received 29 July 1966)

This paper derives the general commutator $[q^m, p^n]$ for all integers m and n, positive or negative. Two new identities for the binomial coefficients are derived using the general commutator. A simple example of the use of the general commutation relation is given in the expansion of $(Ap + Bq)^n$ (A, B, constants) in normal form for integer n.

I. INTRODUCTION

A RECENT article in this journal[1] has given a method of rearranging functions of the operators q and p so that all the q's lie to the left of all the p's. The method has two limitations. First, it involves solving an associated eigenvalue equation. And, second, it is proved only for functions of positive powers of the operators. The importance of considering negative powers of operators is clear from the fact that the canonical conjugate of most operators of interest involve negative powers of operators. Negative powers of operators do not exist in Hilbert space, of course, but do in a generalized function space.

It is true that all of the results in this paper could be derived in a representation. However, in addition to being aesthetically less pleasing, this would make the results representation-dependent and one would then have to prove separately that they were independent of the representation. The fact that mixed representations as well as q and p representations are of interest in generalized function spaces would make such a proof more difficult.

This paper derives the general commutator $[q^m, p^n]$ for all integer m and n, positive or negative. Two new identities for the binomial coefficients are derived using the general commutator. A simple example of the use of the general commutation relations is given in the expansion of $(Ap + Bq)^n$ (A, B, constants) in normal form for integer n.

II. DERIVATION OF THE GENERAL SPACE–MOMENTUM COMMUTATION RELATION

We consider only the one-dimensional case. The extension to higher dimensions is trivial. We assume there exist two operators which obey the commutation relation

$$[p, q] = -i\hbar. \tag{1}$$

Since this is our only assumption and since (1) is invariant under the transformation

$$q \to p; \quad p \to -q, \tag{2}$$

all the resulting equations also have this invariance. For the same reason all the resulting equations are invariant under the transformation

$$q \to p; \quad p \to q; \quad \hbar \to -\hbar. \tag{3}$$

We wish to find the operator function $F(q, p)$ such that

$$[p^\alpha, q^\beta] = F(q, p), \tag{4}$$

where in general α and β are allowed to be any integers, positive or negative. We say that an operator $F(q, p)$ is in the standard form when it is expressed as an infinite series, each term of which has all the q's to the left of all the p's. The notation $F(q \mid p)$ means $F(q, p)$ expressed in its standard form. Thus,

$$F(q \mid p) = \sum_{j,k=-\infty}^{\infty} a_{jk} q^j p^k. \tag{5}$$

We demand that $F(q, p)$ in Eq. (4) be expressed in its standard form.

A. Positive Powers of the Operators, $\alpha > 0$; $\beta > 0$

First consider $[p^n, q]$. If

$$[p^n, q] = -i\hbar n p^{n-1} \quad (n \geq 1), \tag{6}$$

then

$$\begin{aligned}[p^{n+1}, q] &= p[p^n, q] + [p, q]p^n \\ &= -i\hbar n p^n - i\hbar p^n = -i\hbar(n + 1)p^n. \end{aligned} \tag{7}$$

Thus, if Eq. (6) is true for one choice of n, $(n \geq 1)$, it must be true for all larger choices. Since it is true for $n = 1$, it is true for all n, $(n \geq 1)$.

Applying the transformation (2) immediately gives

$$[p, q^n] = -i\hbar n q^{n-1} \quad (n \geq 1). \tag{8}$$

Now consider $[p^\alpha, q^\beta]$. If,

$$[p^\alpha, q^\beta] = \sum_{m=1}^{L(\alpha,\beta)} \binom{\alpha}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m} p^{\alpha-m} \quad (\alpha \geq 1; \quad \beta \geq 1), \tag{9}$$

where $L(\alpha, \beta)$ is equal to α or β (whichever is least),

[1] Leon Cohen, J. Math. Phys. 7, 244 (1966).

and $\binom{\alpha}{m}$ and $\binom{\beta}{m}$ are binomial coefficients,[2] then

$$
\begin{aligned}
&[p^{\alpha+1}, q^{\beta}] \\
&= p[p^{\alpha}, q^{\beta}] + [p, q^{\beta}]p^{\alpha} \\
&= -i\hbar\beta q^{\beta-1}p^{\alpha} + \sum_{m=1}^{L(\alpha,\beta)} \binom{\alpha}{m}\binom{\beta}{m}(-i\hbar)^m m!\, pq^{\beta-m}p^{\alpha-m} \\
&= -i\hbar\beta q^{\beta-1}p^{\alpha} \\
&\quad + \sum_{m=1}^{L(\alpha,\beta)} \binom{\alpha}{m}\binom{\beta}{m}(-i\hbar)^m m!\left(q^{\beta-m}p + [p, q^{\beta-m}]\right)p^{\alpha-m} \\
&= -i\hbar\beta q^{\beta-1}p^{\alpha} + \sum_{m=1}^{L(\alpha,\beta)} \binom{\alpha}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m}p^{\alpha-m+1} \\
&\quad + \sum_{m=1}^{L(\alpha,\beta)} \binom{\alpha}{m}\binom{\beta}{m}(-i\hbar)^{m+1} m!\,(\beta - m)q^{\beta-m-1}p^{\alpha-m} \\
&= -i\hbar\beta q^{(\beta-1)}p^{\alpha} \\
&\quad + \sum_{m=1}^{L(\alpha,\beta)} \binom{\alpha}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m}p^{\alpha-m+1} \\
&\quad + \sum_{m=2}^{L(\alpha,\beta)+1} \binom{\alpha}{m-1}\binom{\beta}{m-1}(-i\hbar)^m (m-1)! \\
&\quad \times (\beta - m + 1)q^{\beta-m}p^{\alpha+1-m} \\
&= -i\hbar(\alpha+1)\beta q^{\beta-1}p^{\alpha} \\
&\quad + \sum_{m=2}^{L(\alpha,\beta)} \binom{\alpha+1}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m}p^{\alpha+1-m} \\
&\quad + \binom{\alpha}{L(\alpha,\beta)}\binom{\beta}{L(\alpha,\beta)}(-i\hbar)^{(L(\alpha,\beta)+1)}[L(\alpha,\beta)]! \\
&\quad \times (\beta - L(\alpha,\beta))q^{(\beta-L(\alpha,\beta)-1)}p^{(\alpha-L(\alpha,\beta))} \\
&= \sum_{m=1}^{L(\alpha,\beta)} \binom{\alpha+1}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m}p^{\alpha+1-m} \\
&\quad + \binom{\alpha}{L(\alpha,\beta)}\binom{\beta}{L(\alpha,\beta)}(-i\hbar)^{(L(\alpha,\beta)+1)}[L(\alpha,\beta)]! \\
&\quad \times (\beta - L(\alpha,\beta))q^{(\beta-L(\alpha,\beta)-1)}p^{(\alpha-L(\alpha,\beta))}. \qquad (10)
\end{aligned}
$$

If $\beta > \alpha$, then $L(\alpha, B) = \alpha$ and the last term in (10) is

$$
\begin{aligned}
&\binom{\beta}{\alpha}(-i\hbar)^{(\alpha+1)}\alpha!\,(\beta - \alpha)q^{\beta-\alpha-1} \\
&\qquad = \binom{\beta}{\alpha+1}(-i\hbar)^{\alpha+1}(\alpha+1)!\, q^{\beta-(\alpha+1)}, \qquad (11)
\end{aligned}
$$

which gives

$$
[p^{\alpha+1}, q^{\beta}] = \sum_{m=1}^{\alpha+1} \binom{\alpha+1}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m}p^{\alpha+1-m} \qquad (\beta > \alpha). \qquad (12)
$$

If $\alpha \geq \beta$, then $L(\alpha, \beta) = \beta$ and the last term in Eq. (10) vanishes. Then,

$$
[p^{\alpha+1}, q^{\beta}] = \sum_{m=1}^{\beta} \binom{\alpha+1}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m}p^{\alpha+1-m} \qquad (\alpha \geq \beta). \qquad (13)
$$

Combining Eqs. (12) and (13) we get

$$
[p^{\alpha+1}, q^{\beta}] = \sum_{m=1}^{L(\alpha+1,\beta)} \binom{\alpha+1}{m}\binom{\beta}{m}(-i\hbar)^m m!\, q^{\beta-m}p^{\alpha+1-m}. \qquad (14)
$$

Applying the transformation (2) to (14) and then exchanging the labels $(\alpha + 1)$ and β gives

$$
[q^{\beta}, p^{\alpha+1}] = \sum_{m=1}^{L(\alpha+1,\beta)} \binom{\alpha+1}{m}\binom{\beta}{m}(i\hbar)^m m!\, p^{\alpha+1-m}q^{\beta-m}. \qquad (15)
$$

Applying Eq. (3) to Eq. (15) and then exchanging α and β gives

$$
[p^{\alpha}, q^{\beta+1}] = \sum_{m=1}^{L(\alpha+1,\beta)} \binom{\alpha}{m}\binom{\beta+1}{m}(-i\hbar)^m m!\, q^{\beta+1-m}p^{\alpha-m}. \qquad (16)
$$

Equations (14) and (16) together imply that if Eq. (9) is true for any values of α ($\alpha \geq 1$) and β ($\beta \geq 1$), it is true for all larger values. Since it is true for $(\alpha = 1; \beta = 1)$, it is true for $\alpha \geq 1; \beta \geq 1$.

B. Inverse Operators

1. Definition and Basic Commutation Relations Involving Inverse Operators

We define the inverse of an operator q by the relation

$$
q\frac{1}{q} = \frac{1}{q}q = 1. \qquad (17)
$$

$\frac{1}{q}$, as defined in (17), exists as an operator in a generalized function space.

Note then, that

$$
p\left[\frac{1}{p}, q\right] = q - pq\frac{1}{p} = q - (qp - i\hbar)\frac{1}{p} = i\hbar\frac{1}{p}. \qquad (18)
$$

Thus,

$$
\left[\frac{1}{p}, q\right] = \frac{i\hbar}{p^2}. \qquad (19)
$$

Transforming (19) with (3) gives

$$
\left[p, \frac{1}{q}\right] = \frac{i\hbar}{q^2}. \qquad (20)
$$

Also,

$$
\begin{aligned}
qp\left[\frac{1}{p}, \frac{1}{q}\right] &= 1 - qp\frac{1}{q}\frac{1}{p} = 1 - (pq + i\hbar)\frac{1}{q}\frac{1}{p} \\
&= -i\hbar\frac{1}{q}\frac{1}{p}, \qquad (21)
\end{aligned}
$$

[2] This result has been previously derived in the q-representation by J. R. Shewell, Am. J. Phys. 27, 161 (1959).

$$pq\left[\frac{1}{p},\frac{1}{q}\right]=pq\frac{1}{p}\frac{1}{q}-1=(qp-i\hbar)\frac{1}{p}\frac{1}{q}-1$$
$$=-i\hbar\frac{1}{p}\frac{1}{q}. \tag{22}$$

From Eqs. (21) and (22) we find

$$\left[\frac{1}{p},\frac{1}{q}\right]=-i\hbar\frac{1}{p}\frac{1}{q^2}\frac{1}{p}=-i\hbar\frac{1}{q}\frac{1}{p^2}\frac{1}{q}. \tag{23}$$

Equation (23) is not in the standard form. We later put it in this form.

2. $\alpha>0;\ \beta\le 0$

We define $\gamma=-\beta$. Consider $\left[p,\frac{1}{q^\gamma}\right]$. If

$$\left[p,\frac{1}{q^\gamma}\right]=i\hbar\gamma\frac{1}{q^{\gamma+1}}, \tag{24}$$

then

$$\left[p,\frac{1}{q^{\gamma+1}}\right]=\frac{1}{q}\left[p,\frac{1}{q^\gamma}\right]+\left[p,\frac{1}{q}\right]\frac{1}{q^\gamma}$$
$$=i\hbar\gamma\frac{1}{q^{\gamma+2}}+i\hbar\frac{1}{q^{\gamma+2}}=i\hbar(\gamma+1)\frac{1}{q^{\gamma+2}}. \tag{25}$$

Thus, since Eq. (24) is true for $(\gamma=0)$, it is true for all $\gamma\ge 0$. Using Eq. (3) on (24) we get

$$\left[\frac{1}{p^\gamma},q\right]=i\hbar\gamma\frac{1}{p^{\gamma+1}}\quad(\gamma\ge 0). \tag{26}$$

Now we consider $\left[p^\alpha,\frac{1}{q^\gamma}\right]$. If

$$\left[p^\alpha,\frac{1}{q^\gamma}\right]=\sum_{m=1}^{\alpha}\binom{\alpha}{m}\binom{\gamma+m-1}{m}(i\hbar)^m m!\,\frac{1}{q^{\gamma+m}}\,p^{\alpha-m}, \tag{27}$$

then

$$\left[p^{\alpha+1},\frac{1}{q^\gamma}\right]$$
$$=p\left[p^\alpha,\frac{1}{q^\gamma}\right]+\left[p,\frac{1}{q^\gamma}\right]p^\alpha$$
$$=\sum_{m=1}^{\alpha}\binom{\alpha}{m}\binom{\gamma+m-1}{m}(i\hbar)^m m!\,p\frac{1}{q^{\gamma+m}}p^{\alpha-m}$$
$$+i\hbar\gamma\frac{1}{q^{\gamma+1}}p^\alpha$$
$$=\sum_{m=1}^{\alpha}\binom{\alpha}{m}\binom{\gamma+m-1}{m}(i\hbar)^m m!\,\frac{1}{q^{\gamma+m}}p^{\alpha+1-m}$$
$$+\sum_{m=2}^{\alpha+1}\binom{\alpha}{m-1}\binom{\gamma+m-2}{m-1}(i\hbar)^m(m-1)!$$
$$\times(\gamma+m-1)\frac{1}{q^{\gamma+m}}p^{\alpha+1-m}+i\hbar\gamma\frac{1}{q^{\gamma+1}}p^\alpha$$
$$=\sum_{m=1}^{\alpha+1}\binom{\alpha+1}{m}\binom{\gamma+m-1}{m}(i\hbar)^m m!\,\frac{1}{q^{\gamma+m}}p^{\alpha+1-m}. \tag{28}$$

Thus, since we have shown that Eq. (27) is true for $\alpha=1$, for any $\gamma\ge 0$, it is true for all $\alpha\ge 1$ and any $\gamma\ge 0$.

An analogous proof gives

$$\left[\frac{1}{p^\gamma},q^\beta\right]=\sum_{m=1}^{\beta}\binom{\beta}{m}\binom{\gamma+m-1}{m}(i\hbar)^m m!\,q^{\beta-m}\,\frac{1}{p^{\gamma+m}}, \tag{29}$$

where $\gamma\ge 0;\ \beta>0$.

3. $\alpha\le 0;\ \beta\le 0$

We define $\gamma=-\alpha$, $\epsilon=-\beta$, so that γ and ϵ are nonnegative integers.

Consider the identity $(\gamma>0)$

$$-i\hbar\gamma\frac{1}{p^{\gamma+1}}$$
$$=-\sum_{m=2}^{\infty}\binom{\gamma+m-1}{m}(-i\hbar)^m m!\,\frac{1}{q^{m-1}}\frac{1}{p^{\gamma+m}}$$
$$+\sum_{m=1}^{\infty}\binom{\gamma+m-1}{m}(-i\hbar)^m m!\,\frac{1}{q^{m-1}}\frac{1}{p^{\gamma+m}}$$
$$=-\sum_{m=2}^{\infty}\binom{\gamma+m-2}{m-1}(-i\hbar)^m(m-1)!$$
$$\times(\gamma+m-1)\frac{1}{q^{m-1}}\frac{1}{p^{\gamma+m}}$$
$$+\sum_{m=1}^{\infty}\binom{\gamma+m-1}{m}(-i\hbar)^m m!\,\frac{1}{q^{m-1}}\frac{1}{p^{\gamma+m}}$$
$$=-\sum_{m=1}^{\infty}\binom{\gamma+m-1}{m}(-i\hbar)^{m+1}m!\,(\gamma+m)\frac{1}{q^m}\frac{1}{p^{\gamma+m+1}}$$
$$+\sum_{m=1}^{\infty}\binom{\gamma+m-1}{m}(-i\hbar)^m m!\,\frac{1}{q^{m-1}}\frac{1}{p^{\gamma+m}}$$
$$=\sum_{m=1}^{\infty}\binom{\gamma+m-1}{m}(-i\hbar)^m m!$$
$$\times\frac{1}{q^m}\left(q\frac{1}{p^{\gamma+m}}+i\hbar(\gamma+m)\frac{1}{p^{\gamma+m+1}}\right)$$

$$= \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \times \frac{1}{q^m}\left\{q\frac{1}{p^{\gamma+m}} + \left[\frac{1}{p^{\gamma+m}}, q\right]\right\}$$

$$= \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \frac{1}{q^m}\frac{1}{p^{\gamma+m}} q. \tag{30}$$

Thus,

$$-i\hbar\gamma \frac{1}{p^{\gamma+1}}\frac{1}{q} = \left\{\frac{1}{p^\gamma}q - \left[\frac{1}{p^\gamma}, q\right]\right\}\frac{1}{q} - \frac{1}{p^\gamma} = q\frac{1}{p^\gamma}\frac{1}{q} - \frac{1}{p^\gamma}$$

$$= \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \frac{1}{q^m}\frac{1}{p^{\gamma+m}}. \tag{31}$$

Therefore for $\gamma > 0$,

$$\left[\frac{1}{p^\gamma}, \frac{1}{q}\right] = \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \frac{1}{q^{1+m}}\frac{1}{p^{\gamma+m}}. \tag{32}$$

The special case $\gamma = 1$ gives the standard form for Eq. (23). Similarly one can show that for $\epsilon > 0$,

$$\left[\frac{1}{p}, \frac{1}{q^\epsilon}\right] = \sum_{m=1}^{\infty} \binom{\epsilon+m-1}{m} (-i\hbar)^m m! \frac{1}{q^{\epsilon+m}}\frac{1}{p^{1+m}}. \tag{33}$$

For $\gamma > 0$ and $\epsilon > 0$, if

$$\left[\frac{1}{p^\gamma}, \frac{1}{q^\epsilon}\right] = \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m}\binom{\epsilon+m-1}{m} \times (-i\hbar)^m m! \frac{1}{q^{\epsilon+m}}\frac{1}{p^{\gamma+m}}, \tag{34}$$

then

$$\left[\frac{1}{p^\gamma}, \frac{1}{q^{\epsilon+1}}\right]$$

$$= \left[\frac{1}{p^\gamma}, \frac{1}{q^\epsilon}\right]\frac{1}{q} + \frac{1}{q^\epsilon}\left[\frac{1}{p^\gamma}, \frac{1}{q}\right]$$

$$= \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m}\binom{\epsilon+m-1}{m} (-i\hbar)^m m! \frac{1}{q^{\epsilon+m}}\frac{1}{p^{\gamma+m}}\frac{1}{q} + \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \frac{1}{q^{\epsilon+1+m}}\frac{1}{p^{\gamma+m}}$$

$$= \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m}\binom{\epsilon+m-1}{m} (-i\hbar)^m m! \times \frac{1}{q^{\epsilon+m+1}}\left\{\frac{1}{q}\frac{1}{p^{\gamma+m}} + \left[\frac{1}{p^{\gamma+m}}, \frac{1}{q}\right]\right\} + \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \frac{1}{q^{\epsilon+1+m}}\frac{1}{p^{\gamma+m}}$$

$$= \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \times \left\{\left[\begin{matrix}\epsilon+m-1\\ m\end{matrix}\right] + 1\right\}\frac{1}{q^{\epsilon+m+1}}\frac{1}{p^{\gamma+m}} + \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m}\binom{\epsilon+m-1}{m} m! \times \sum_{M=m+1}^{\infty} \binom{\gamma+M-1}{M-m} \times (-i\hbar)^M (M-m)! \frac{1}{q^{\epsilon+M+1}}\frac{1}{p^{\gamma+M}}$$

$$= \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \left\{\binom{\epsilon+m-1}{m} + 1\right\} \times \frac{1}{q^{\epsilon+m+1}}\frac{1}{p^{\gamma+m}} + \sum_{m=1}^{\infty} \binom{\epsilon+m-1}{m} \times \sum_{M=m+1}^{\infty} \binom{\gamma+M-1}{M} M! (-i\hbar)^M \frac{1}{q^{\epsilon+M+1}}\frac{1}{p^{\gamma+M}}$$

$$= -i\hbar\gamma(\epsilon+1)\frac{1}{q^{\epsilon+1}}\frac{1}{p^{\gamma+1}} + \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m} (-i\hbar)^m m! \left\{\binom{\epsilon+m-1}{m} + 1 + \sum_{n=1}^{m-1} \binom{\epsilon+n-1}{n}\right\}\frac{1}{q^{\epsilon+m+1}}\frac{1}{p^{\gamma+m}}. \tag{35}$$

If we define $S = \epsilon + m - 1$, then

$$\sum_{n=1}^{m-1} \binom{\epsilon+n-1}{n} = \sum_{S=\epsilon}^{\epsilon+m-2} \binom{S}{\epsilon-1} = \sum_{S=\epsilon-1}^{\epsilon+m-2} \binom{S}{\epsilon-1} - 1. \tag{36}$$

It is a well-known identity[3] that

$$\sum_{S=\epsilon-1}^{\epsilon+m-2} \binom{S}{\epsilon-1} = \binom{\epsilon+m-1}{\epsilon}. \tag{37}$$

Putting Eqs. (36) and (37) into Eq. (35), we find that if Eq. (34) is true then

$$\left[\frac{1}{p^\gamma}, \frac{1}{q^{\epsilon+1}}\right] = \sum_{m=1}^{\infty} \binom{\gamma+m-1}{m}\binom{\epsilon+m}{m} \times (-i\hbar)^m m! \frac{1}{q^{\epsilon+1+m}}\frac{1}{p^{\gamma+m}}. \tag{38}$$

A similar proof shows that if Eq. (34) is true then

$$\left[\frac{1}{p^{\gamma+1}}, \frac{1}{q^\epsilon}\right] = \sum_{m=1}^{\infty} \binom{\gamma+m}{m}\binom{\epsilon+m-1}{m} \times (-i\hbar)^m m! \frac{1}{q^{\epsilon+m}}\frac{1}{p^{\gamma+1+m}}. \tag{39}$$

Thus, since we have previously shown that Eq. (34) is true for $\gamma = 1$ or $\epsilon = 1$, it is true for all $\gamma \geq 1$; $\epsilon \geq 1$.

[3] *Handbook of Mathematical Tables* (The Chemical Rubber Company, Cleveland, 1964), p. 388.

We can combine the four subcases (9), (27), (29), and (34) by introducing the following notation:

$$\theta(q) \equiv \begin{cases} 0 & \text{for } q < 0 \\ 1 & \text{for } q > 0 \end{cases};$$

$$\sigma(\alpha, \beta) = \theta(\alpha)\theta(\beta) + \theta(-\alpha)\theta(-\beta);$$

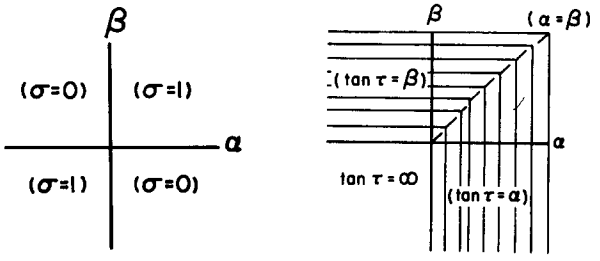
$$\tau(\alpha, \beta) \equiv \theta(\alpha)\theta(\beta)\{\alpha\theta(\beta - \alpha) + \beta\theta(\alpha - \beta) + \tfrac{3}{2}\} + \theta(\alpha)\{\alpha\theta(-\beta) - 1\} + \theta(\beta)\{\beta\theta(-\alpha) - 1\} + \frac{1}{\theta(\alpha) + \theta(\beta)}. \quad (40)$$

$\theta(q)$ is the usual step function.

$$\sigma(\alpha, \beta) = \begin{cases} 1 & \text{when } \alpha \text{ and } \beta \text{ have the } \textit{same} \text{ signs} \\ 0 & \text{when } \alpha \text{ and } \beta \text{ have } \textit{different} \text{ signs} \end{cases}.$$

$\tau(\alpha, \beta) =$

$$\begin{cases} \text{The smaller of the } \textit{two}, \text{ if both } \alpha \text{ and } \beta \text{ are positive} \\ \text{The positive } \textit{one}, \text{ if } \alpha \text{ and } \beta \text{ are of different signs} \\ \infty, \text{ if } \alpha \text{ and } \beta \text{ are both negative} \end{cases}.$$



With this notation, we can now write for any integers α and β, positive or negative,

$$[p^\alpha, q^\beta] = (1 - \delta_{\alpha,0})(1 - \delta_{\beta,0}) \sum_{m=1}^{\tau(\alpha,\beta)} \times \binom{|\alpha| + (m-1)\theta(-\alpha)}{m} \times \binom{|\beta| + (m-1)\theta(-\beta)}{m} \times (i\hbar)^m (-1)^{\sigma(\alpha,\beta)m} m!\, q^{\beta - m} p^{\alpha - m}. \quad (41)$$

This result can be extended to the case of two operators whose commutator $[p, q] = K$, an arbitrary constant, by merely letting $i\hbar \to -K$ in Eq. (41).

III. TWO NEW IDENTITIES SATISFIED BY THE BINOMIAL COEFFICIENTS

Multiplying both sides of (34) by q^ϵ we get

$$q^\epsilon \frac{1}{p^\gamma}\frac{1}{q^\epsilon} - \frac{1}{p^\gamma} = \left(\frac{1}{p^\gamma} q^\epsilon - \left[\frac{1}{p^\gamma}, q^\epsilon\right]\right)\frac{1}{q^\epsilon} - \frac{1}{p^\gamma}$$

$$= -\sum_{m=1}^{\epsilon} \binom{\epsilon}{m} \binom{\gamma + m - 1}{m} (i\hbar)^m m!\, q^{\epsilon - m} \frac{1}{p^{\gamma+m}} \frac{1}{q^\epsilon}$$

$$= \sum_{m=1}^{\infty} \binom{\gamma + m - 1}{m} \binom{\epsilon + m - 1}{m} (-i\hbar)^m m! \frac{1}{q^m} \frac{1}{p^{\gamma+m}}. \quad (42)$$

Thus,

$$-\sum_{m=1}^{\epsilon} \binom{\epsilon}{m} \binom{\gamma + m - 1}{m} (i\hbar)^m m!\, q^{\epsilon - m} \frac{1}{p^{\gamma+m}}$$

$$= \sum_{m=1}^{\infty} \binom{\gamma + m - 1}{m} \binom{\epsilon + m - 1}{m} (-i\hbar)^m m! \frac{1}{q^m} \frac{1}{p^{\gamma+m}} q^\epsilon$$

$$= \sum_{m=1}^{\infty} \binom{\gamma + m - 1}{m} \binom{\epsilon + m - 1}{m} (-i\hbar)^m m! \frac{1}{q^m} \times \left(q^\epsilon \frac{1}{p^{\gamma+m}} + \left[\frac{1}{p^{\gamma+m}}, q^\epsilon\right]\right)$$

$$= \sum_{m=1}^{\infty} \binom{\gamma + m - 1}{m} \binom{\epsilon + m - 1}{m} (-i\hbar)^m m!\, q^{\epsilon - m} \frac{1}{p^{\gamma+m}} + \sum_{m=1}^{\infty} \binom{\gamma + m - 1}{m} \binom{\epsilon + m - 1}{m} (-1)^m m! \sum_{n=1}^{\epsilon} \binom{\epsilon}{n} \times \binom{\gamma + m + n - 1}{n} (i\hbar)^{m+n} n!\, q^{\epsilon - (m+n)} \frac{1}{p^{\gamma+(m+n)}}. \quad (43)$$

Defining $G(\alpha, \beta)$ to be α or β whichever is larger, (43) can be written

$$\sum_{m=1}^{\infty} \binom{\gamma + m - 1}{m} \binom{\epsilon + m - 1}{m} (-i\hbar)^m m!\, q^{\epsilon - m} \frac{1}{p^{\gamma+m}} + \sum_{m=1}^{\epsilon} \binom{\epsilon}{m} \binom{\gamma + m - 1}{m} (i\hbar)^m m!\, q^{\epsilon - m} \frac{1}{p^{\gamma+m}} + \sum_{m=2}^{\infty} \binom{\gamma + m - 1}{m} m!\, (i\hbar)^m \sum_{n=G(1,m-\epsilon)}^{m-1} (-1)^n \times \binom{\epsilon + n - 1}{n} \binom{\epsilon}{m - n} q^{\epsilon - m} \frac{1}{p^{\gamma+m}} = 0. \quad (44)$$

Since (44) is an operator equation, the coefficient of $q^{\epsilon-\alpha}(1/p^{\gamma+\alpha})$ must vanish identically for each choice of α. The choice $\alpha = 1$ gives a trivial identity. For $\epsilon \geq \alpha \geq 2$, $G(1, \alpha - \epsilon) = 1$ and we get the identity

$$\sum_{n=0}^{\alpha} (-1)^n \binom{n + \epsilon - 1}{n} \binom{\epsilon}{\alpha - n} = 0. \quad (45)$$

Swapping the labels α and ϵ in (44), we get for $\alpha \geq 1$; $\epsilon \geq 2$; and $\epsilon > \alpha$,

$$\sum_{n=0}^{\alpha} (-1)^n \binom{n + \epsilon - 1}{\alpha - 1} \binom{\alpha}{n} = 0. \quad (46)$$

These two identities can be combined to give

$$\sum_{n=G(0,\alpha-\epsilon)}^{\alpha} (-1)^n \binom{n + \epsilon - 1}{n} \binom{\epsilon}{\alpha - n} = 0. \quad (47)$$

IV. EXPANSION OF $(Ap + Bq)^n$ IN NORMAL FORM FOR INTEGER n

$$\text{Let}\quad \Psi(n) \equiv \begin{cases} \dfrac{n}{2} & \text{for } n \text{ even}, \\ \dfrac{n-1}{2} & \text{for } n \text{ odd}. \end{cases} \tag{48}$$

$$\text{Let}\quad p!! \equiv \begin{cases} 1 & \text{for } p \leq 0 \\ 1\cdot 3\cdot 5\cdots p & \text{for } p > 0, p \text{ odd} \\ 2\cdot 4\cdot 6\cdots p & \text{for } p > 0, p \text{ even}. \end{cases} \tag{49}$$

Assume that

$$(Ap + Bq)^n = \sum_{j=0}^{\Psi(n)} (-i\hbar AB)^j (2j-1)!! \binom{n}{2j} \times \sum_{r=0}^{n-2j} \binom{n-2j}{r} (Bq)^{(n-2j-r)} (Ap)^r, \tag{50}$$

where A and B are numbers and p and q are operators satisfying

$$[p, q] = -i\hbar. \tag{51}$$

From Eq. (6) we see that

$$[p^r, q] = -i\hbar r p^{r-1}. \tag{52}$$

Multiplying both sides of Eq. (50), on the right by $(Ap + Bq)$, and applying (52) to the result, we get

$$\begin{aligned} (Ap + Bq)^{n+1} = {} & \sum_{j=0}^{\Psi(n)} (-i\hbar AB)^j (2j-1)!! \binom{n}{2j} \times \sum_{r=0}^{n-2j} \binom{n-2j}{r} (Bq)^{(n-2j-r)} (Ap)^{r+1} \\ & + \sum_{j=0}^{\Psi(n)} (-i\hbar AB)^j (2j-1)!! \binom{n}{2j} \times \sum_{r=0}^{n-2j} \binom{n-2j}{r} (Bq)^{(n+1-2j-r)} (Ap)^r \\ & + \sum_{j=0}^{\Psi(n)} (-i\hbar AB)^{(j+1)} (2j-1)!! \binom{n}{2j} \times \sum_{r=0}^{n-2j} r \binom{n-2j}{r} (Bq)^{(n-2j-r)} (Ap)^{(r-1)}. \end{aligned} \tag{53}$$

In the last term of (53), the term where $r = 0$ equals zero and thus can be dropped from the summation. In the case where n is even, however, $\Psi(n) = \frac{1}{2}n$ and the term $j = \Psi(n)$ contains only the $r = 0$ term, and thus, must be dropped simultaneously. The last term in (53) can be then written as

$$\sum_{j=0}^{\Phi(n)} (-i\hbar AB)^{(j+1)} (2j-1)!! \binom{n}{2j} \times \sum_{r=1}^{n-2j} r \binom{n-2j}{r} (Bq)^{n-2j-r} (Ap)^{r-1}, \tag{54}$$

where

$$\Phi(n) \equiv \begin{cases} \dfrac{n}{2} - 1 & \text{for } n \text{ even} \\ \dfrac{n-1}{2} & \text{for } n \text{ odd}. \end{cases} \tag{55}$$

Making the transformations $r \to (r-1)$ and $j \to (j+1)$ in (54) and simplifying, we get for the last term in (53)

$$\sum_{j=1}^{\Phi(n)+1} (-i\hbar AB)^j (2j-1)!! \binom{n+1}{2j} \left[\frac{2j}{n+1}\right] \times \sum_{r=0}^{n+1-2j} \binom{n+1-2j}{r} (Bq)^{(n+1-2j-r)} (Ap)^r. \tag{56}$$

Making the transformation $r \to (r+1)$ in the first term on the right side of (53) and combining it with the second term on the right side of (53), we get for the first two terms

$$\sum_{j=0}^{\Psi(n)} (-i\hbar AB)^j (2j-1)!! \binom{n}{2j} \times \sum_{r=0}^{n+1-2j} \binom{n+1-2j}{r} (Bq)^{(n+1-2j-r)} (Ap)^r. \tag{57}$$

Then combining Eqs. (56) and (57), we get

$$\begin{aligned} & (Ap + Bq)^{n+1} \\ & = \sum_{r=0}^{\Psi(n)} (-i\hbar AB)^j (2j-1)!! \binom{n+1}{2j} \times \sum_{r=0}^{n+1-2j} \binom{n+1-2j}{r} (Bq)^{(n+1-2j-r)} (Ap)^r \\ & \quad + (1 + \Phi(n) - \Psi(n)) (-i\hbar AB)^{(n+1)/2} n!!. \end{aligned} \tag{58}$$

Note that

$$1 + \Phi(n) - \Psi(n) = \begin{cases} 0 & \text{for } n \text{ even} \\ 1 & \text{for } n \text{ odd}. \end{cases} \tag{59}$$

Since

$$\Psi(n+1) = \begin{cases} \Psi(n) & \text{for } n \text{ even} \\ \Psi(n) + 1 & \text{for } n \text{ odd}, \end{cases} \tag{60}$$

(58) can be written

$$\begin{aligned} & (Ap + Bq)^{n+1} \\ & = \sum_{r=0}^{\Psi(n+1)} (-i\hbar AB)^j (2j-1)!! \binom{n+1}{2j} \times \sum_{r=0}^{n+1-2j} \binom{n+1-2j}{r} (Bq)^{(n+1-2j-r)} (Ap)^r. \end{aligned} \tag{61}$$

Thus, if Eq. (50) is true for n, it is true for $n + 1$. Since it is true for $n = 0$, it is true for all $n \geq 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Phase-Integral Approximation in Momentum Space and the Bound States of an Atom

MARTIN C. GUTZWILLER
IBM Watson Laboratory, Columbia University, New York, New York 10025

(Received 13 February 1967)

The phase integral approximation of the Green's function in momentum space is investigated for an electron of negative energy (corresponding to a bound state) which moves in a spherically symmetric potential. If the propagator rather than the wavefunction is considered, all classical orbits enter into the formulas, rather than only the ones which satisfy certain quantum conditions, and the separation of variables can be avoided. The distinction between classically accessible and classically inaccessible regions does not arise in momentum space, because any two momenta can be connected by a classical trajectory of given negative energy for a typical atomic potential. Three approaches are discussed: the Fourier transform of the phase integral approximation in coordinate space, the approximate solution of Schrödinger's equation in momentum space by a WKB ansatz, and taking the limit of small Planck's quantum in the Feynman-type functional integral which was recently proposed by Garrod for the energy–momentum representation. In particular, the last procedure is used to obtain the phase jumps of $\pi/2$ which occur every time neighboring classical trajectories cross one another. These extra phase factors are directly related to the signature of the second variation for the action function, and provide a physical application of Morse's calculus of variation in the large. The phase integral approximation in momentum space is then applied to the Coulomb potential. The location of the poles on the negative energy axis gives the Bohr formula for the bound-state energies, and the residues of the approximate Green's function are shown to yield all the exact wavefunctions for the bound states of the hydrogen atom.

I. INTRODUCTION

THE present investigation was undertaken with the ultimate goal of finding analytic (as opposed to numerical), approximate expressions for single electron wavefunctions of bound states in atoms or simple molecules. The phase-integral approximation, sometimes called the WKB method, provides such expressions. However, it turns out that a somewhat unusual approach working in momentum space is more appropriate than the well-known form involving Hamilton's action function in coordinate space. Actually, we construct a phase-integral approximation for the propagator, or Green's function, $F(p'' p' E)$, in terms of the initial momentum p', the final momentum p'', and the energy E. The singularities of F along the negative E axis give the approximate wavefunctions. This procedure is tested for the Coulomb potential, where it is found to yield the exact wavefunctions for all the bound states.

Although this last result seems better than expected, there are good reasons to believe that the present scheme is indeed more efficient than the usual ones, at least in the case of bound states for typical atomic potentials. First, the connection between classical and quantum mechanics is much simpler for the propagator than for the individual wavefunctions. The construction of the approximate propagator requires the knowledge of all the classical paths which go from the initial to the final point, whereas an approximate wavefunction requires a special class of classical orbits for which there is no simple and general description. Second, at a given negative energy, $E < 0$, any two momenta p' and p'' can be connected by a classical trajectory in the case of a typical atomic or molecular potential. But two positions, q' and q'', can be connected by a classical trajectory only if they lie both in the region where the potential energy $V(q)$ is smaller than the total energy E. The propagator $F(p'' p' E)$ will, therefore, be approximated by an expression $\tilde{F}(p'' p' E)$ with smoother behavior than the propagator $G(q'' q' E)$ whose approximation $\tilde{G}(q'' q' E)$ has some rather artificial singularities around $V(q') = E$ or $V(q'') = E$. Third, it appears that the common procedure of separating the variables in a problem of spherical symmetry has an adverse effect upon the phase-integral approximation. The well-known difficulty in obtaining Bohr's formula for the hydrogen levels vanishes entirely if we construct either $\tilde{F}$ or $\tilde{G}$ in three dimensions without bothering to separate variables.

The general formula for $\tilde{F}(p'' p' E)$ is easily written down, but its derivation does not satisfy a mathematician's requirement for rigor. Even the phase-integral approximation $\tilde{K}(q'' q' t)$ for the propagator $K(q'' q' t)$ from position q' to position q'' in the given time t has not yet been established with the desirable degree of accuracy and generality for singular potentials such as the Coulomb potential, although $\tilde{K}$ is certainly better understood than $\tilde{G}$ which in turn is better known than $\tilde{F}$. The author found it helpful to

arrive at $\tilde{F}$ by several different methods. The emphasis of this report is, therefore, not on the applications for which the method was originally devised, but on the more basic problems concerned with the phase-integral approximation. In particular, it is necessary to obtain $\tilde{F}$ directly from the path-integral expression for F which was recently discovered by Garrod[1] as a generalization of the Feynman integral for K. Apart from the very difficult question of justifying Feynman-type integrals and deriving their limits for vanishing Planck's quantum, certain results from the classical calculus of variation are needed, in particular the character of the second variation. As far as $\tilde{K}$ is concerned, these results have been obtained by Morse,[2] but the class of problems arising from $\tilde{F}$ and $\tilde{G}$ (which one may legitimately call isoperimetric) has apparently not been considered as yet and their solution is only conjectured for the purpose at hand.

The discussion of the various topics is presented in the following manner. Section IIA summarizes some of the well-known results about the propagator $K(q''t'', q't')$, and proposes a general formula for the limit $\hbar \to 0$ when $t'' - t'$ is arbitrarily long. The crucial phase jumps at a focal point are related to Morse's theory of the second variation, which in turn arises quite naturally if one goes to vanishing $\hbar$ in Feynman's path integral for K. Section IIB discusses the same ideas for the Green's function $G(q'' q' E)$, although our mathematical background in this instance is much poorer. Three different ways to obtain the limit of G for small $\hbar$ are presented, by taking the Fourier transform of K, by solving the inhomogeneous Schrödinger equation, and by letting $\hbar$ vanish in Garrod's path-integral expression. A second variation is again needed, except that the variational quantity is not covered by Morse's theory, and certain conjectures have to be made. Section IIC carries the arguments over into the study of $F(p'' p' E)$, in particular the three methods for going to the limit $\hbar \to 0$. The investigation of Garrod's path integral and the study of the second variation for the action integral are now particularly important, because Schrödinger's equation is not local anymore, and the phase jumps cannot be obtained in the customary manner.

Since the formulas for the limits $\tilde{G}(q'' q' E)$ and $\tilde{F}(p'' p' E)$ as $\hbar$ vanishes are completely analogous, any detailed calculations can be carried out in either case. The more familiar $\tilde{G}$ is chosen in Sec. IIIA to exhibit the simplifications due to a spherically symmetric potential. Section IIIB establishes the same result by performing explicitly the limiting process in Garrod's path integral for $G(q'' q' E)$; this feat has only been possible for a spherically symmetric potential, although the formula for $\tilde{G}$ is believed to be valid more generally. The various results are listed in Sec. IIIC for $\tilde{F}$, as they are needed for the Coulomb problem.

The Kepler orbits in momentum space are discussed in Sec. IVA. Since they are circles, their geometry is much easier to understand than in coordinate space, and simplifies all explicit calculations. The phase-integral approximation $\tilde{F}$ is worked out in Sec. IVB on this basis. In particular the phase jumps at focal points are obtained, and compared with those of another famous problem, the linear oscillator. The resulting approximate Green's function is shown in Sec. IVC to have poles at the negative values of E in agreement with Bohr's formula. The residues are worked out and are compared with the residues in the exact Green's function which has recently been established by various authors. The complete agreement confirms our original expectation that bound states are best described by the phase-integral method in momentum space.

II. GENERAL FORMULAS

A. Time and Space Coordinates

Consider a simple physical system without spin, e.g., an electron in a given electromagnetic field. Its coordinates are given by a vector q, and its momentum by a vector p. In case the components of q or p have to be specified, they are indicated by an upper index, such as q^j or p^j. The propagation function $K(q''t'', q't')$ for this system depends on the initial coordinates q' and time t', as well as the final coordinates q'' and time $t'' > t'$. K is found from the requirements that

$$i\hbar(\partial K/\partial t'') - H_{\mathrm{op}}(p'' q'' t)K = 0, \tag{1}$$

$$\lim_{t'' \to t'} K(q''t'', q't') = \delta(q'' - q'), \tag{2}$$

where $H_{\mathrm{op}}(p\, q\, t)$ is the Hamiltonian operator. H_{op} is obtained formally from the classical Hamiltonian $H(p\, q\, t)$ if p is replaced by the operator $-i\hbar\partial/\partial q$. Planck's quantum divided by 2π is written as $\hbar$. Equation (1) is Schrödinger's equation, and the initial condition (2) appears quite naturally if one tries to solve the initial value problem for (1).

After a suggestion by Dirac, it was demonstrated

[1] C. Garrod, Rev. Mod. Phys. **38**, 483 (1966).

[2] M. Morse, *The Calculus of Variations in the Large* (American Mathematical Society, Providence, Rhode Island, 1935). For more contemporary presentations, cf. J. Milnor, *Morse Theory* (Princeton University Press, Princeton, New Jersey, 1962); H. M. Edwards, Ann. Math., 2nd Ser. **80**, 22 (1964); S. Smale, J. Math. Mech. **14**, 1049 (1965).

by Feynman that K can be represented as an integral over all possible trajectories from q' at t' to q'' at t'' in the following manner. Let the time interval from t' to t'' be subdivided into N subintervals by inserting $t_1, t_2, \cdots, t_{N-1}$, and define a discrete path from q' to q'' by inserting the intermediate points $q_1, q_2, \cdots, q_{N-1}$. An action integral R_N along this path is given by

$$R_N = \sum_1^N (t_n - t_{n-1}) L\left(\frac{q_n - q_{n-1}}{t_n - t_{n-1}}, q_n, t_n\right), \quad (3)$$

where $q' = q_0$, $t' = t_0$, $q'' = q_N$, and $t'' = t_N$. Also, we have introduced the classical Lagrangian

$$L(\dot{q}\, q\, t) = \sum_j p^j \frac{\partial H}{\partial p^j} - H, \quad (4)$$

where the momenta p^j are eliminated on the right-hand side with the help of the relation $\dot{q}^j = dq^j/dt = \partial H/\partial p^j$. Feynman's formula is then given by

$$K = \lim_{N\to\infty} \prod_1^N \left[\frac{m}{2\pi i\hbar(t_n - t_{n-1})}\right]^{\frac{3}{2}} \times \int d^3q_1 \cdots \int d^3q_{N-1} \exp\,[iR_N/\hbar]. \quad (5)$$

For definiteness we have assumed a 3-dimensional q-space, and a particle of mass m. The physical content of (5) is discussed in a recent monograph by Feynman and Hibbs.[3] The constant in front of the $(N-1)$-fold integration has been chosen mainly to obtain the relation

$$\int d^3q K(q''t'', qt)K(qt, q't') = K(q''t'', q't'). \quad (6)$$

Nelson[4] has recently discussed Feynman's formula as an analytic continuation of Wiener's formula[5] for Brownian motion, but we would like to start directly from (5).

Pauli[6] investigated the limit of K for small time intervals $t'' - t'$. The result can be written in terms of the action integral

$$R(q''t'', q't') = \int_{t'}^{t''} L(\dot{q}\, q\, t)\, dt, \quad (7)$$

calculated along the classical trajectory which carries the particle from q' at time t' to q'' at time t''. The approximate value $\tilde{K}$ is given by

$$\tilde{K}(q''t'', q't') = (2\pi i\hbar)^{-\frac{3}{2}}(D_R)^{\frac{1}{2}} \exp\,[iR(q''t'', q't')/\hbar], \quad (8)$$

where D_R is the determinant of the mixed derivatives

$$D_R = (-1)^3 \det\,|(\partial^2 R)/(\partial q' \partial q'')|. \quad (9)$$

Since the initial momentum p' along the classical trajectory is given by $p' = -\partial R/\partial q'$, one can interpret D_R as the Jacobian $\partial(p')/\partial(q'')$ between the range d^3p' of initial momenta and the volume d^3q'' covered by the endpoints.

The validity of (8) has been established by Choquard[7] for potentials without singularities. But even for the Coulomb potential, one has always at least two classical trajectories connecting any given pair of points q' and q'' in a given time $t'' - t'$. For a short time interval $t'' - t'$, one trajectory follows quite closely the straight line from q' to q'', whereas the other trajectory heads first for the center of attraction, then turns around it in a sharp twist, and goes to the final point following an almost radial path again. The formula (8) for $\tilde{K}$ will, therefore, not be sufficient for a typical atomic potential. Actually, the singularity (2) in K at $t'' = t'$ follows from (8) if we evaluate R for the direct path from q' to q''. Formula (8) remains presumably valid for sufficiently small $|q'' - q'|$ if it is applied only to the direct trajectory, since the contribution from the indirect trajectory would remain finite.

Pauli showed that $\tilde{K}$ as given by (8) satisfies Schrödinger's Eq. (1) up to a remainder which is proportional to $\hbar^2$. It is, therefore, reasonable to expect that the limit of K, as $\hbar$ goes to zero, has an appearance very much like (8). Feynman's formula (5) shows that there is a contribution to the limit of vanishing $\hbar$ from every path $q' = q_0, q_1, \cdots, q_N = q''$ for which R_N is stationary. Thus we expect in general a sum of terms like (8), one for each classical trajectory from q' at t' to q'' at t''. The continuity of the result requires that each term in this sum takes the exact form (8) as q'' approaches q' along the direct path while $t'' - t'$ is sufficiently small. As t'' increases from t', and q'' runs along a given classical trajectory, the amplitude $(D_R)^{\frac{1}{2}}$ becomes infinite every time q'' passes a focal point. A detailed examination of Schrödinger's equation in its neighborhood shows that (8) remains valid even beyond the focal point, if we take the amplitude $(|D_R|)^{\frac{1}{2}}$ and insert a special phase factor $\exp(-i\pi/2)$ for every reduction by 1 in the rank of the Jacobian $\partial(q'')/\partial(p') = 1/D_R$ at the focal point. Thus, we obtain

$$\tilde{K}(q''t'', q't') = (2\pi i\hbar)^{-\frac{3}{2}} \sum_{\text{classical paths}} (D_R)^{\frac{1}{2}} \times \exp\left[\frac{iR}{\hbar} + \text{phases}\right], \quad (10)$$

as the limit of $K(q''t'', q't')$ for vanishing $\hbar$.

[3] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill Book Company, Inc., New York, 1965).
[4] E. Nelson, J. Math. Phys. **5**, 332 (1964).
[5] N. Wiener, J. Math. Phys. **2**, 131 (1923).
[6] W. Pauli, Ausgewählte Kapitel aus der Feldquantisierung, Lecture Notes, Zurich, 1951.
[7] Ph. Choquard, Helv. Phys. Acta **28**, 89 (1955).

This last expression was obtained by solving Eq. (1) up to terms in $\hbar^2$, by imposing the initial condition (2), and by forcing the result to be continuous. The relation (6) can be checked for (10) if the integral over q is computed by the stationary phase method. However, it would help our understanding of similar approximations for $G(q''\,q'\,E)$ and $F(p''\,p'\,E)$ if the expression (10) could be directly derived from Feynman's integral (5). We shall indicate the necessary steps, although we realize that there are many gaps to be filled before any mathematical rigor can be claimed.

Let N be large enough so that the particular classical path of interest can be adequately described by a sequence $q' = \bar{q}_0, \bar{q}_1, \cdots, \bar{q}_N = q''$ corresponding to the times $t' = t_0, t_1, \cdots, t_N = t''$. The approximate action R_N given by (3) is stationary for $q_1 = \bar{q}_1, \cdots, q_{N-1} = \bar{q}_{N-1}$. If we write $q_j = \bar{q}_j + \delta q_j$ for $j = 1, \cdots, N-1$, we find

$$R_N = R(q''t'', q't') + \tfrac{1}{2}\sum_{jl} R_{jl}\delta q_j \delta q_l + \cdots, \quad (11)$$

where the omitted terms are of order either $(\delta q)^3$ or $1/N$. Since δq_j is a vector, the matrix R_{jl} has more elements than are actually suggested by (11). The integrations over $q_1, \cdots, q_{N-1}$ are easily performed if the exponent in (5) is replaced by the two terms in (11). The matrix R_{jl} has to be transformed to principal axes, so that one gets $3N-3$ Fresnel integrals. Thus we find an amplitude equal to

$$(2\pi\hbar)^{\frac{1}{2}(3N-3)} \cdot (|\det R_{jl}|)^{-\frac{1}{2}}$$

and a phase factor

$$\exp\,[iR/\hbar + (3N-3)i\pi/4 - iM\pi/2],$$

where M is the number of negative eigenvalues of R_{jl}. In order to show agreement with (10) we have to establish that

$$\lim_{N\to\infty} \prod_1^N \left[\frac{m}{(t_n - t_{n-1})}\right]^3 \cdot (|\det R_{jl}|)^{-1} = |D_R|, \quad (12)$$

and that M equals the number of focal points between q' and q'', each counted according to the rank of D_R^{-1}. The proof of (12) can be accomplished without difficulty in the case of a spherically symmetrical potential because R_{jl} simplifies greatly in polar coordinates and its determinant can be evaluated by writing out the appropriate recursion formulas (cf. Gel'fand and Yaglom,[8] as well as Montroll[9]). Unfortunately, such a procedure has not been successful in the case of a nonseparable potential. On the other hand, the relation between M and the focal points is a well-known result of the calculus of variation in the large, as worked out by Morse[2] in a classic monograph. Morse's results can, therefore, be interpreted physically in terms of the extra phase which a wave loses at a caustic due to its spilling over into the classically forbidden region.

The results of Morse have not received any attention in the textbooks of classical mechanics. Yet, in every course there is at least one student to ask whether, indeed, the integral $\int L\,dt$ becomes minimal along the classical trajectory. If the answer might have seemed unimportant because there has been no physical application for it so far, it is all the more interesting to find such an application in the transition from classical to quantum mechanics. The fact that $\int L\,dt$ becomes minimal for a sufficiently short path gives Morse's theory a simplicity which will not be matched by the later examples of a second variation (cf. Secs. IIB and IIC).

B. Energy and Space Coordinates

In order to describe stationary states of a physical system, one has to know the propagator at constant energy. We assume from now on that H is independent of t, so that K depends only on the difference $t'' - t'$. The Green's function $G(q''\,q'\,E)$ is defined as

$$G(q''\,q'\,E) = \frac{1}{i\hbar}\int_0^\infty dt K(q''t, q'0)\exp\left[\frac{iEt}{\hbar}\right], \quad (13)$$

where E is in the upper half of a complex E plane. The homogeneous differential equation (1) and the initial condition (2) are now combined into the inhomogeneous equation

$$[E - H_{op}(p''q'')]G(q''q'E) = \delta(q'' - q'). \quad (14)$$

If the homogeneous equation $[E - H_{op}(pq)]\psi = 0$ has no acceptable solution for an interval of real values of E, then Eq. (14) has a solution which is, moreover, symmetric in q' and q''. Green's function G can then be continued analytically into the lower half of the complex E plane by putting

$$G(q''\,q'\,E^*) = [G(q'\,q''\,E)]^*. \quad (15)$$

Thus, the behavior of G along the real E axis is directly related to the existence of solutions for the homogeneous equation which corresponds to (14). The details of this relation are discussed in any modern textbook on Green's functions.

The expression for G in terms of an integral over all paths from q' to q'' has only been discovered very recently by Garrod.[1] The crucial step is to consider

[8] I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. 1, 48 (1960).
[9] E. W. Montroll, Commun. Pure Appl. Math. 5, 415 (1952).

all possible paths in "phase space" rather than coordinate space only. Thus one introduces a sequence of coordinates $q' = q_0, q_1, \cdots, q_N = q''$ (as before), but in addition a sequence of momenta $p_{\frac{1}{2}}, p_{\frac{3}{2}}, \cdots, p_{N-\frac{1}{2}}$. A path in "phase space" is described by the combined sequence $q' = q_0, p_{\frac{1}{2}}, q_1, p_{\frac{3}{2}}, \cdots, q_{N-1}, p_{N-\frac{1}{2}}, q_N = q''$, and the mean energy $\mathcal{E}$ along this path is defined by

$$\mathcal{E} = \frac{1}{N}\sum_{\frac{1}{2}}^{N-\frac{1}{2}} \frac{p_n^2}{2m} + \frac{1}{N}\sum_{0}^{N} V(q_n). \tag{16}$$

For simplicity's sake the Hamiltonian has been assumed to consist only of the usual kinetic energy $p^2/2m$ and the potential energy $V(q)$. But, both a relativistic kinetic energy and a vector potential could equally well have been included. The Green's function now becomes

$$G = \lim_{N\to\infty} (2\pi\hbar)^{-3N} \int \prod_{1}^{N-1} d^3q_n \int \prod_{\frac{1}{2}}^{N-\frac{1}{2}} d^3p_n \times \exp\left[\frac{i}{\hbar} S_N\right] \Big/ (E - \mathcal{E}), \tag{17}$$

where S_N is the action along the path $q_0, p_{\frac{1}{2}}, q_1, \cdots, p_{N-\frac{1}{2}}, q_N$ in phase space,

$$S_N = \sum_{1}^{N} p_{n-\frac{1}{2}}(q_n - q_{n-1}). \tag{18}$$

The few formal steps from (5) to (17) are explained in Appendix A, because our definition of G differs slightly from Garrod's.

There are three ways to finding the approximation $\tilde{G}$ of G for small $\hbar$. First, one can simply insert (10) into (13) and evaluate the integral over t by the stationary phase method. Second, the inhomogeneous wave equation (14) can be solved in the limit of vanishing $\hbar$. Third, the limit of (17) can be found as $\hbar$ goes to zero.

The first method is the most straightforward and is carried out in Appendix B. Its result is expressed in terms of the classical action

$$S(q'' q' E) = \int_{q'}^{q''} p\, dq, \tag{19}$$

evaluated along the classical path which leads from q' to q'' at the given energy $H(p q) = E$. The phase integral approximation $\tilde{G}$ becomes

$$\tilde{G}(q'' q' E) = -\frac{1}{2\pi\hbar^2} \sum_{\text{classical paths}} (|D_S|)^{\frac{1}{2}} \times \exp\left[\frac{iS}{\hbar} + \text{phases}\right], \tag{20}$$

where the determinant D_S now contains not only the second mixed derivatives with respect to q' and q'', but also with respect to E,

$$D_S = \begin{vmatrix} \dfrac{\partial^2 S}{\partial q' \partial q''} & \dfrac{\partial^2 S}{\partial q' \partial E} \\ \dfrac{\partial^2 S}{\partial E \partial q''} & \dfrac{\partial^2 S}{\partial E^2} \end{vmatrix}. \tag{21}$$

Actually, the element $\partial^2 S/\partial E^2$ might just as well be replaced by 0, because the 3×3 determinant $|\partial^2 S/\partial q' \partial q''|$ vanishes. The phases in (20) are the same as in (10), except when $\partial^2 R/\partial t^2 = -\partial E/\partial t < 0$, i.e., a higher-energy orbit leads to a longer transit time. The interpretation of D_S can be made as follows: Consider the family of classical trajectories which leave q' with the initial momentum p' in the neighborhood $d^3p' = d\Omega'\, dE$; their endpoints lie in a neighborhood $d^3q'' = d\Omega''\, dt$ of q''; D_S is then the Jacobian $d\Omega'/d\Omega''$. The phases in (20) are again $-i\pi/2$ times the reduction in rank of the 2×2 matrix associated with $d\Omega''/d\Omega'$ at a focal point.

The second method for obtaining $\tilde{G}$ has been studied extensively, e.g., by Avila and Keller,[10] in the case where $E - V(q)$ is positive and bounded for all q. This situation corresponds to the scattering of particles by a potential without singularities, whereas we are interested in particles which are trapped in a singular potential such as the Coulomb potential. Nevertheless, the general considerations are similar; in particular, the discussion of caustics, as in the work of Ludwig,[11] can be taken over directly. But one will not have an infinity of trajectories from q' to q'' if $E - V(q)$ is bounded and positive. Kohn and Sham[12] have obtained $\tilde{G}$ in one dimension with the help of the well-known expression for G in terms of a Wronskian. Their method has not been generalized to more than one dimension; formula (20) leads exactly to their result.

The singularity of $\tilde{G}$ for a small distance $|q'' - q'|$ can be obtained directly from the inhomogeneous equation (14). It is found that up to terms in $|q'' - q'|^2$

$$\tilde{G}(q'' q' E) \cong -\frac{m}{2\pi\hbar^2 |q'' - q'|} \times \exp\{i |q'' - q'| [2m(E - V(q))]^{\frac{1}{2}}/\hbar\}, \tag{22}$$

where $q = \frac{1}{2}(q' + q'')$. This expression corresponds to limiting the expansion (20) to the shortest trajectory from q' to q'', and evaluating $S(q'' q' E)$ in powers of $|q'' - q'|$. The approximation (22) for $\tilde{G}$ is completely equivalent to the Thomas–Fermi approximation,

[10] C. S. S. Avila and J. B. Keller, Comm. Pure Appl. Math. **16**, 363 (1963).
[11] D. Ludwig, Comm. Pure Appl. Math. **19**, 215 (1966).
[12] W. Kohn and L. J. Sham, Phys. Rev. **137**, A1697 (1965).

which appears usually as the Fourier transform $\int d^3(q''-q')\exp[-ip(q''-q')/\hbar]$ of (22), namely

$$G_{\mathrm{TF}}(pqE)=[E-(p^2/2m)-V(q)]^{-1}; \quad (23)$$

cf. Baraff and Borowitz.[13]

The third method seems to be the most interesting because it leads to a new viewpoint in classical mechanics and to some new problems for the calculus of variations in the large. It is natural to perform the integrations in (17) in two steps. First, one integrates over the variables p_n and q_n on a hypersurface of constant average energy $\mathcal{E}$, as given by (16). Second, $\mathcal{E}$ is integrated from $-\infty$ to $+\infty$. As $\hbar$ goes to zero, one is, therefore, faced with finding the stationary path $q'=q_0, p_{\frac{1}{2}}, q_1, \cdots, p_{N-\frac{1}{2}}, q_N=q''$ for S_N under the subsidiary condition (16). In the limit of large N, one has to solve the "isoperimetric" problem: Find the curves $p(t)$, $q(t)$ in phase space for which $\int p\,dq$ is stationary, given the endpoints q' and q'', as well as the average energy $\mathcal{E}=\int H(pq)\,dt/(t''-t')$. The Euler equations of this problem are the Hamilton equations of motion, but the usual variational principle at constant energy demands that $\int p\,dq$ be stationary for given endpoints q' and q'', while $H(pq)=\mathcal{E}$ at each point $p(t)$, $q(t)$; cf. Whittaker.[14]

Garrod[1] noticed this novel variational principle. For the purpose of finding $\tilde{G}$, one has to go one step further, since the second variation of S_N is needed. Let N again be large enough to describe a particular classical path in phase space by a sequence $q'=\bar{q}_0, \bar{p}_{\frac{1}{2}}, \bar{q}_1, \cdots, \bar{p}_{N-\frac{1}{2}}, \bar{q}_N=q''$ taken at equal time intervals. With $q_n=\bar{q}_n+\delta q_n$ and $p_n=\bar{p}_n+\delta p_n$, where $n=\frac{1}{2}, 1, \cdots, N-\frac{1}{2}$, one expands in powers of δq_n and δp_n,

$$\begin{aligned} S_N &= S(q''q'\mathcal{E})+\delta^1 S+\delta^2 S+\cdots \\ \mathcal{E} &= \frac{1}{(t''-t')}\int_{t'}^{t''} H(pq)\,dt+\delta^1\mathcal{E}+\delta^2\mathcal{E}+\cdots, \end{aligned} \quad (24)$$

where the omitted terms are either of order $1/N$ or of third order in δq_n and δp_n. The classical path is stationary if the condition

$$\delta^1 S-\tau\delta^1\mathcal{E}=0 \quad (25)$$

is identically fulfilled in all δq_n and δp_n for a parameter τ such as to satisfy (16). The second variation of the exponent in (17) subject to (16) becomes

$$[\delta^2 S]_{\mathcal{E}}=[\delta^2 S-\tau\delta^2\mathcal{E}]_{\delta^1\mathcal{E}=0}, \quad (26)$$

i.e., the variables δq_n and δp_n in the quadratic form $\delta^2 S-\tau\delta^2\mathcal{E}$ are subject to the linear constraint $\delta^1\mathcal{E}=0$.

The further steps in the integration over δq_n and δp_n, with (26) inserted into the exponent of (17), are straightforward. One finds an amplitude $(2\pi\hbar)^{3N-2}$ times $A(q''\,q'\,\mathcal{E})\,d\mathcal{E}$, where A contains the determinant of the matrix associated with (26) and a Jacobian, because the integration uses internal coordinates for (26) in addition to $\mathcal{E}$, rather than δq_n and δp_n. The phase factor is simply $\exp[iS(q''\,q'\,\mathcal{E})/\hbar-iM\pi/4]$, where M can be called the index of the classical trajectory. M is equal to the number of negative eigenvalues in (26) minus the number of positive ones. In the case of a spherically symmetric potential, it will be shown in Sec. IIIB that the amplitude $A(q''\,q'\,\mathcal{E})$ equals $(|D_S|)^{\frac{1}{2}}$, where E is replaced by $\mathcal{E}$. As in the case of $\tilde{K}$ and Eq. (12), we have not been able to show this identity in the more general case of an arbitrary potential, but we shall assume it henceforth. The index M starts out with a value 2 for the most direct trajectory from q' to a nearby endpoint q''. As can be observed from the sign of D_S, the index M changes at every focal point. We conjecture that M increases at every focal point by twice as much as the rank of $d\Omega''/d\Omega'$ is reduced. There does not seem to exist a simple relationship between $\int L\,dt$ on one hand, and $\int p\,dq$ at constant average energy on the other, although the equations of motion for stationary trajectories are identical.

Since the amplitude $A(q''\,q'\,\mathcal{E})$ does not depend on $\hbar$, the main variation in the integral over $\mathcal{E}$ comes from the phase factor $\exp[iS(q''\,q'\,\mathcal{E})/\hbar-iM\pi/4]$. Therefore, A is pulled out of the integral with $\mathcal{E}$ replaced by E, and $S(q''\,q'\,\mathcal{E})$ is expanded around E to first power in $\mathcal{E}-E$. The remaining integral becomes

$$\int_{-\infty}^{+\infty}\frac{d\mathcal{E}}{E-\mathcal{E}}\,e^{it(\mathcal{E}-E)/\hbar}=\begin{cases}-2\pi i & \text{for } t>0,\\ 0 & \text{for } t<0,\end{cases} \quad (27)$$

where $t=\partial S(q''q'E)/\partial E$ is the transit time for the particle to go from q' to q''. The denominator in (17) automatically limits the contributions from the various paths in phase space to the ones which correspond to going forward in time, provided the imaginary part of E is positive. Thus we find again the approximation (20), but this time on the basis of Garrod's formula (17).

C. Energy and Momentum

The propagator $F(p''\,p'\,E)$ for a particle to start out with a momentum p' and end up with a momentum p'' while propagating with the energy E, is

[13] G. E. Baraff and S. Borowitz, Phys. Rev. **121**, 1704 (1961).

[14] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge University Press, Cambridge, England, 1937), 4th ed., p. 247.

defined by

$$F(p'' p' E) = (2\pi\hbar)^{-3}\int d^3q''\int d^3q' G(q'' q' E) \times \exp[i(p'q' - p''q'')/\hbar]. \quad (28)$$

The inhomogeneous Schrödinger equation (14) becomes

$$[E - H_{op}(p''q'')]F(p'' p' E) = \delta(p'' - p'), \quad (29)$$

where the Hamilton operator is the integral operator

$$H \cdot F = (p''^2/2m)F(p'' p' E) + \int d^3p V(p''p)F(p p' E). \quad (30)$$

$V(p'' p)$ is the Fourier transform of the potential $V(q)$, i.e.,

$$V(p''p') = (2\pi\hbar)^{-3}\int d^3q V(q) \exp[-i(p'' - p')q/\hbar]. \quad (31)$$

The path integral expression (17) can directly be inserted into (28) to yield the formula

$$F = \lim_{N\to\infty}(2\pi\hbar)^{-3N-3}\int\prod_0^N d^3q_n\int\prod_{\frac{1}{2}}^{N-\frac{1}{2}} d^3p_n \times \exp\left[-\frac{i}{\hbar}T_N\right]\Big/(E - \mathcal{E}), \quad (32)$$

where T_N is the action along the path $p' = p_{-\frac{1}{2}}$, $q_0, p_{\frac{1}{2}}, q_1, \cdots, p_{N-\frac{1}{2}}, q_N, p_{N+\frac{1}{2}} = p''$ in phase space given by

$$T_N = \sum_0^N q_n(p_{n+\frac{1}{2}} - p_{n-\frac{1}{2}}), \quad (33)$$

and the average energy $\mathcal{E}$ is given by the same formula (16).

The semiclassical or WKB method has been used occasionally in momentum space. Kohn[15] describes the motion of electrons in a solid in this manner. Goldman *et al.*[16] discuss the transformation between WKB wavefunctions in coordinate and in momentum space for one dimension. Schiller[17] writes the equations for the phase and the amplitude in a time-dependent situation. But none of these authors has investigated the Green's function F in the semiclassical approximation, nor were they interested in bound states, even for a spherically symmetric potential. There are again the three ways to finding the approximation $\tilde{F}$ of F for small $\hbar$ which were discussed in the preceding section.

The first method consists in applying the Fourier transform (28) to the formula (20) for $\tilde{G}$, and evaluating the integral by the stationary phase method. The procedure corresponds very closely to the calculations in Appendix B. The result involves the classical action

$$T(p'' p' E) = \int_{p'}^{p''} q\, dp, \quad (34)$$

calculated along the classical path in momentum space which leads from p' to p'' at the given energy $H(pq) = E$. The phase-integral approximation $\tilde{F}$ is given, in complete analogy to (20), by

$$\tilde{F}(p'' p' E) = -\frac{1}{2\pi\hbar^2}\sum_{\text{classical paths}}(|D_T|)^{\frac{1}{2}} \times \exp\left[-\frac{iT}{\hbar} + \text{phases}\right], \quad (35)$$

where the 4×4 determinant D_T contains again the second mixed derivatives of T with respect to p' and p'' as well as E,

$$D_T = \begin{vmatrix} \frac{\partial^2 T}{\partial p'\partial p''} & \frac{\partial^2 T}{\partial p'\partial E} \\ \frac{\partial^2 T}{\partial E\partial p''} & \frac{\partial^2 T}{\partial E^2} \end{vmatrix}. \quad (36)$$

Again, the element $\partial^2 T/\partial E^2$ may be replaced by zero, because the 3×3 determinant $|\partial^2 T/\partial p'\partial p''|$ vanishes. The determinant (36) has a completely analogous interpretation to (21), in terms of the family of classical trajectories which go from p' into the neighborhood of p'' at the given energy E. Presumably the phases in (35) are similarly related to the caustics which are generated by this family of trajectories in momentum space. But it is important to realize that the two families, one in coordinate space and the other in momentum space, are not simply the same set of curves in different representations. This fact becomes especially apparent if one studies the character of the focal points along the classical trajectory. Thus, a Kepler orbit in coordinate space has two singly counting focal points followed by the doubly counting starting point, whereas in momentum space there is one doubly counting focal point followed by the doubly counting starting point to which all trajectories of the family return.

The second method of deriving $\tilde{F}$ consists in using a trial solution of the type

$$B(p'' p' E)\exp[-iT(p'' p' E)/\hbar]$$

in order to solve the inhomogeneous Schrödinger equation (29) to first order in $\hbar$. The potential-energy term in (30) is evaluated in Appendix C with the help of the stationary phase method. The Hamiltonian

[15] W. Kohn, Proc. Phys. Soc. (London) **72**, 1147 (1958), cf. also E. I. Blount, Phys. Rev. **126**, 1636 (1962).

[16] I. I. Goldman, V. D. Krivchenko, V. I. Kogan, and V. M. Galitskii, *Problems in Quantum Mechanics* (Academic Press Inc. New York, 1960), pp. 11 and 92.

[17] R. Schiller, Phys. Rev. **125**, 1100 and 1109 (1962).

operating on $B \exp[-iT/\hbar]$ becomes

$$H_{\mathrm{op}}F = \exp\left[\frac{iT}{\hbar}\right] \cdot \left\{\left[\frac{p''^2}{2m} + V(q'')\right]B + i\hbar\left[\frac{\partial V}{\partial q''}\frac{\partial B}{\partial p''} + \frac{B}{2}\frac{\partial^2 V}{\partial q''\partial q''} \cdot \frac{\partial^2 T}{\partial p''\partial p''}\right] + \cdots\right\}, \tag{37}$$

where $q'' = \partial T/\partial p''$, and the terms in the last line are sums over the components of p'' and q''. The remainder is of order $\hbar^2$. If this expression is inserted into the left-hand side of (29), it is not evident at all which terms in (37) are to be matched by the $\delta(p'' - p')$ on the right-hand side of (29). Obviously this inhomogeneous term in (29) determines the amplitude B, exactly as the $\delta(q'' - q')$ on the right-hand side of (14) determines the amplitude A of $\tilde{G}$, whose behavior for small $|q'' - q'|$ is expressed in (22).

Upon closer examination, the following is found. The terms $p''^2/2m + V(q'')$ in the first line of (37) are equal to E, provided T is the appropriate action function (34). The square bracket in the second line of (37) vanishes if B is proportional to $(D_T)^{\frac{1}{2}}$ and $p'' \neq p'$. The behavior of $(D_T)^{\frac{1}{2}}$ as p'' approaches p' can be most easily investigated if one starts with the Thomas–Fermi approximation

$$S \cong |q'' - q'| \cdot \{2m[E - V(\tfrac{1}{2}(q' + q''))]\}^{\frac{1}{2}}$$

as in (22), and examines the transformation into momentum space, $p'' = \partial S/\partial q''$ and $p' = -\partial S/\partial q'$. The Jacobian of this transformation is given in the limit of $q' = q''$ by the expression $m^4\nu/2m(E - v(q))$ with

$$\nu = -\begin{vmatrix} V_{11} & V_{12} & V_{13} & V_1 \\ V_{21} & V_{22} & V_{23} & V_2 \\ V_{31} & V_{32} & V_{33} & V_3 \\ V_1 & V_2 & V_3 & 0 \end{vmatrix}, \tag{38}$$

where $V_{ij} = \partial^2 V/\partial q^i \partial q^j$ and $V_j = \partial V/\partial q^j$. The value of $-(2\pi\hbar^2)^{-1}(|D_T|)^{\frac{1}{2}} \exp(-iT/\hbar)$ for small $|p'' - p'|$ is obtained as

$$-\frac{1}{2\pi\hbar^2\,|p'' - p'|} \cdot \frac{|\partial V/\partial q|}{(\nu)^{\frac{1}{2}}} \exp\left[-\frac{i}{\hbar}|p'' - p'| \cdot |q|\right], \tag{39}$$

where q is chosen on the surface $2m[E - V(q)] = (p'' + p')^2/4$ such that the direction of $-\partial V/\partial q$ coincides with the direction of $p'' - p'$. Since the square bracket in the second line of (37) contains only first derivatives, it will not lead to a singularity $\delta(p'' - p')$ if we insert (39). This is, in fact, what one has to expect, since the factor $1/(2\pi\hbar^2)$ in (39), together with the factor $i\hbar$ in (37), yields a term of order $\hbar^{-1}$, whereas the right-hand side of (29) is of order $\hbar^0$. The inhomogeneous term in (29) is, therefore, not generated by the formula (35) for $\tilde{F}$ [as the $\delta(q'' - q')$ in (14) is generated by the formula (20) for $\tilde{G}$]; it would come about only by going to the next term in the expansion (37) for $H_{\mathrm{op}}F$. If the singularity at $p'' = p'$ is to be included explicitly in an approximation for F, one would have to write

$$\frac{\delta(p'' - p')}{E - p^2/2m} + \frac{V(p''p')}{(E - p''^2/2m)(E - p'^2/2m)} - \frac{1}{2\pi\hbar^2}\sum_{\text{classical paths}} (|D_T|)^{\frac{1}{2}} \exp\left[-i\frac{T}{\hbar} + \text{phases}\right], \tag{40}$$

where the first two terms are obtained from an expansion of F in powers of $V(p''p')$. For Coulomb-like potentials, these terms are of order $\hbar^0$ and $\hbar^{-1}$.

The preceding discussion shows that, contrary to $\tilde{G}(q'' q' E)$, the Green's function $\tilde{F}(p'' p' E)$ is not easily obtained by solving the inhomogeneous Schrödinger equation. It seems very hard to get higher-order terms in the expansion (37) for $H_{\mathrm{op}}F$. Also, the behavior of $\tilde{F}$ near a caustic and the extra phase factor cannot be determined from (29), because Schrödinger's equation is an integral equation in momentum space. The expansion of $\tilde{G}(q'' q' E)$ near a caustic, however, is based on finding solutions to Schrödinger's equation which are only valid in a small neighborhood. If $\tilde{F}$ is derived directly from the path integral formula (32), i.e., by the third method, the procedure is absolutely identical with the derivation of $\tilde{G}$ from (17). The discussion at the end of the preceding section can be repeated exactly with p and q, as well as T and S, interchanged. A detailed examination of (17) or (32) in the limit of small $\hbar$ appears, therefore, quite worthwhile.

III. SPHERICALLY SYMMETRIC POTENTIAL

A. Approximate Green's Function in Polar Coordinates

The classical orbit going from q' to q'' lies in the plane which is determined by q', q'', and the center of force at the origin. The action $S(q'' q' E)$ depends only on the absolute values r' and r'' of q'' and q', and on the angle φ between q'' and q'. By straightforward calculation, one finds for the determinant (21) that

$$D_S = \frac{S_\varphi}{r'^2 r''^2 \sin\varphi} \cdot \begin{vmatrix} S_{r'r''} & S_{r'\varphi''} & S_{r'E} \\ S_{\varphi' r''} & S_{\varphi'\varphi''} & S_{\varphi' E} \\ S_{Er''} & S_{E\varphi''} & S_{EE} \end{vmatrix}. \tag{41}$$

φ' and φ'' are the polar angles of q' and q'' in the plane of the orbit. The indices on S indicate the derivatives of S with respect to these quantities. The determinant in (41) is obtained by finding the orbit corresponding to r', r'', φ, and E in a plane.

The equations of motion can be solved by quadratures if we know the angular momentum M of the orbit. It is, therefore, advisable to use M as a third parameter, besides r' and r'', rather than φ. The connection between φ and M follows immediately if we combine the two conservation laws for angular momentum and for energy,

$$mr^2\, d\varphi/dt = M, \tag{42}$$

$$(dr/dt)^2 + r^2(d\varphi/dt)^2 = 2m[E - V(r)]. \tag{43}$$

φ increases always if $M > 0$, even if r sometimes increases and sometimes decreases. Therefore, the integrand in

$$\varphi = \varphi'' - \varphi' = \int_r^{r''} \frac{M\, dr}{r^2[2m(E - V(r) - M^2/2mr^2)]^{\frac{1}{2}}} \tag{44}$$

has to be interpreted as making positive contributions, even if r is made to run back and forth between certain maximum and minimum values, $r_{\max}$ and $r_{\min}$, before reaching the limits of integration, r' and r''. In the same sense we find that

$$S = \int_{r'}^{r''} dr \frac{2m[E - V(r)]}{[2m(E - V(r) - M^2/2mr^2)]^{\frac{1}{2}}}. \tag{45}$$

If the derivatives with respect to φ' and φ'' in (41) are now expressed as derivatives with respect to M, one obtains finally

$$D_S = \frac{M}{r'^2 r''^2 \sin\varphi} \cdot \frac{m}{[2m(E - V(r') - M^2/2mr'^2)]^{\frac{1}{2}}} \times \frac{m}{[2m(E - V(r'') - M^2/2mr''^2)]^{\frac{1}{2}}} \left(\frac{\partial\varphi}{\partial M}\right)^{-1}. \tag{46}$$

The last factor can be expressed formally as an integral over r with the help of (44), namely

$$\frac{\partial\varphi}{\partial M} = \int_{r'}^{r''} dr \frac{2m[E - V(r)]}{r^2[2m(E - V(r) - M^2/2mr^2)]^{\frac{3}{2}}}. \tag{47}$$

It is important to notice certain special cases of (46). If φ tends to zero while M tends to a nonvanishing limit, the approximation (22) is obtained after inserting (45) and (46) into (20). If r'' approaches either $r_{\max}$ or $r_{\min}$, where $E - V(r) - M^2/2mr^2$ vanishes, the amplitude D_s stays finite. Formally, this comes about because the integral (47) diverges while the denominator in (46) vanishes. Physically, it means simply that the orbits do not crowd one another at their point of greatest or smallest distance from the origin. However, D_s becomes infinite wherever $\partial\varphi/\partial M = 0$. A plot of r'' vs φ reveals immediately a caustic for the family of classical trajectories in the same plane which leave q' with different angular momenta M. Similarly, the vanishing of $\sin\varphi$ in the denominators of (41) and (46) indicates a focal point for the family of trajectories which leave q' in different planes, but with the same absolute value of angular momentum. Each occurrence contributes a phase $-i\pi/2$ to the formula (20). These two types of focal points can coincide, such as in the Coulomb potential where all trajectories of a given energy E return to the initial point q', independently of the direction or the magnitude of their angular momentum.

The formulas (45) and (46) can be inserted into (21) in order to yield $\tilde{G}(q''\, q'\, E)$, provided one can solve Eq. (44) so as to find the angular momentum M in terms of the distances r' and r'', and the angle φ. For the Coulomb potential this problem should not be too hard to treat explicitly. But we shall not go into these details here, because a more interesting example of the same calculation is given in the last three sections.

B. Garrod–Feynman Integral in the Limit of Small $\hbar$

The phase-integral approximation in coordinate space at a given energy E can be obtained from the results in the preceding section in the case of a spherically symmetric potential. The same formulas are gotten directly from the path-integral expression (17) for $G(q''\, q'\, E)$ by going to the limit of small $\hbar$. This second derivation is important because it can be used equally well to find the limit of small $\hbar$ for the path-integral expression (32) of $F(p''\, p'\, E)$. It also yields the phase jumps at the focal points and gives new insights into the Garrod–Feynman integrals, (17) and (32).

The first task is to rewrite (17) as well as the original Feynman formula (5) in polar coordinates. Edwards and Gulyaev[18] have discussed this transformation for $K(q''t'', q't')$ in the case of a free particle. But their arguments are greatly simplified for our purpose by the following remarks. The propagator K satisfies Schrödinger's Eq. (1), exactly as the transition probability in Brownian motion satisfies the Fokker–Planck equation (cf. Wang and Uhlenbeck[19]). The

[18] S. F. Edwards and Y. V. Gulyaev, Proc. Roy. Soc. (London) 279, 229 (1964).

[19] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. 17, 323 (1945).

only quantities of importance are, therefore, the various momenta of K in the limit of vanishing $t'' - t'$. For a nonrelativistic particle of mass m in a potential $V(q)$, one has the relations

$$\lim_{t''\to t'}(t''-t')^{-1}\left\{\int K(q''t'', q't')\,d^3q'' - 1\right\} = V(q')/i\hbar, \quad (48)$$

$$\lim_{t''\to t'}(t''-t')^{-1}\int (q^{j''}-q^{j'})K(q''t'', q't')\,d^3q'' = 0, \quad (49)$$

$$\lim_{t''\to t'}(t''-t')^{-1}\int (q^{j''}-q^{j'})(q^{l''}-q^{l'})K\,d^3q'' = (i\hbar/m)\delta_{jl}. \quad (50)$$

It can be shown, by straightforward computation, that these relations are satisfied, not only by the kernel

$$[m/2\pi i\hbar(t''-t')]^{\frac{3}{2}} \times \exp i\{(q''-q')^2/2m(t''-t') - (t''-t')V(q')\}/\hbar$$

in Cartesian coordinates, but also by the expression

$$\left[\frac{m}{2\pi i\hbar(t''-t')}\right]^{\frac{3}{2}} \times \exp\frac{i}{\hbar}\left\{\frac{1}{2m(t''-t')}[(r''-r')^2 + r'r''(\theta''-\theta')^2 + r'r''\sin\theta'\sin\theta''(\varphi''-\varphi')^2] - (t''-t')V(q') + \frac{\hbar^2}{2mr'r''}\left(\frac{1}{4}+\frac{1}{4\sin\theta'\sin\theta''}\right)(t''-t')\right\}, \quad (51)$$

where we have used polar coordinates by putting $q^1 = r\sin\theta\cos\varphi$, $q^2 = r\sin\theta\sin\varphi$, $q^3 = r\cos\theta$ for both the initial and final points. The symmetric occurrence of the single and the double primed coordinates in the first part of the exponent is essential in order to guarantee the relations (49) and (50). The last term in the exponent looks like an additional potential, and has to be inserted if (48) is to be satisfied. The expression (51) is now used to generate the propagator K, i.e., the action function R_N in (5) is written in polar coordinates as

$$R_N = \sum_1^N (t_n - t_{n-1})\left\{\frac{1}{2m(t_n-t_{n-1})^2}[(r_n-r_{n-1})^2 + r_nr_{n-1}(\theta_n-\theta_{n-1})^2 + r_nr_{n-1}\sin\theta_n\sin\theta_{n-1}(\varphi_n-\varphi_{n-1})^2] - V(q_n) + \frac{\hbar^2}{8mr_nr_{n-1}} + \frac{\hbar^2}{8mr_nr_{n-1}\sin\theta_n\sin\theta_{n-1}}\right\}. \quad (52)$$

The elements of integration d^3q_n in (5) are replaced by $r_n^2\sin\theta_n\,dr_n\,d\theta_n\,d\varphi_n$.

The second task is to transform the propagator K into Green's function G in the manner of Appendix A. Three momenta $s_{n-\frac{1}{2}}$, $L_{n-\frac{1}{2}}$, $M_{n-\frac{1}{2}}$ are inserted between the coordinate triples $(r_{n-1}, \theta_{n-1}, \varphi_{n-1})$ and $(r_n, \theta_n, \varphi_n)$. The average energy $\mathcal{E}$ is now defined by

$$\mathcal{E} = \frac{1}{2mN}\sum_{\frac{1}{2}}^{N-\frac{1}{2}}\left[s_n^2 + \frac{L_n^2 - \hbar^2/4}{r_{n+\frac{1}{2}}r_{n-\frac{1}{2}}} + \frac{M_n^2-\hbar^2/4}{r_{n+\frac{1}{2}}r_{n-\frac{1}{2}}\sin\theta_{n+\frac{1}{2}}\sin\theta_{n-\frac{1}{2}}}\right] + \frac{1}{N}\sum_0^N V(q_n), \quad (53)$$

instead of the Cartesian formula (16). Green's function is given by

$$G = \lim_{N\to\infty}(2\pi\hbar)^{-3N}\int\prod_1^{N-1} dr_n\,d\theta_n\,d\varphi_n \times \int\prod_{\frac{1}{2}}^{N-\frac{1}{2}} ds_n\,dL_n\,dM_n\cdot(r_0^2r_n^2\sin\theta_0\sin\theta_N)^{-\frac{1}{2}} \times \exp\left[\frac{i}{\hbar}S_N\right]\Big/(E-\mathcal{E}), \quad (54)$$

and the action S_N along the path in phase space by

$$S_N = \sum_1^N [s_{n-\frac{1}{2}}(r_n - r_{n-1}) + L_{n-\frac{1}{2}}(\theta_n - \theta_{n-1}) + M_{n-\frac{1}{2}}(\varphi_n - \varphi_{n-1})], \quad (55)$$

instead of the Cartesian formulas (17) and (18). s_n is naturally associated with the projection of p_n onto the direction of q_n, whereas M_n and L_n correspond to the components of the angular momentum parallel and perpendicular to the z axis.

The third task is to apply the procedure at the end of Sec. 2 to the energy $\mathcal{E}$ and the action S_N given by (53) and (55). The equations of motion for the classical trajectory follow from (25), and are the following:

$$N(r_n - r_{n-1}) = \tau s_{n-\frac{1}{2}}/m,$$
$$N(\theta_n - \theta_{n-1}) = \tau L_{n-\frac{1}{2}}/mr_nr_{n-1},$$
$$N(\varphi_n - \varphi_{n-1}) = \tau M_{n-\frac{1}{2}}/mr_nr_{n-1}\sin\theta_n\sin\theta_{n-1}; \quad (56)$$

$$N(s_{n-\frac{1}{2}} - s_{n+\frac{1}{2}}) = \tau\frac{\partial}{\partial r_n}\left\{V(r_n) + \frac{L_{n-\frac{1}{2}}^2 - \hbar^4}{2mr_nr_{n-1}} + \frac{L_{n+\frac{1}{2}}^2 - \hbar^2/4}{2mr_nr_{n+1}} + \frac{M_{n-\frac{1}{2}}^2 - \hbar^2/4}{2mr_nr_{n-1}\sin\theta_n\sin\theta_{n-1}} + \frac{M_{n+\frac{1}{2}}^2 - \hbar^2/4}{2mr_nr_{n+1}\sin\theta_n\sin\theta_{n+1}}\right\},$$

$$N(L_{n-\frac{1}{2}} - L_{n+\frac{1}{2}}) = -\frac{\tau\cos\theta_n}{2mr_n\sin^2\theta_n}\left\{\frac{M_{n-\frac{1}{2}}^2 - \hbar^2/4}{r_{n-1}\sin\theta_{n-1}} + \frac{M_{n+\frac{1}{2}}^2 - \hbar^2/4}{r_{n+1}\sin\theta_{n+1}}\right\},$$

$$N(M_{n-\frac{1}{2}} - M_{n+\frac{1}{2}}) = 0. \quad (57)$$

The further calculations are greatly simplified if the coordinate system is chosen such that $\theta' = \theta_0 = \frac{1}{2}\pi = \theta_N = \theta''$ because this implies that $\theta_n = \frac{1}{2}\pi$ and $L_n = 0$. Also, we find that all M_n are equal to some constant M. The additional terms $\hbar^2/4$, which appear in the last two Eqs. (57), can be neglected compared to the classical quantities M_n and L_n. In the limit of infinite N, the remaining Eqs. (56) and (57) can be reduced to (42) and (43) with the help of (53).

In order to compute the second variation (26), we notice that the subsidiary condition $\delta^1\mathcal{E} = 0$ does not involve the variations $\delta L_{\frac{1}{2}}, \delta\theta_1, \delta L_{\frac{3}{2}}, \cdots, \delta\theta_{N-1}, \delta L_{N-\frac{1}{2}}$, and that the quadratic form $\delta^2 S - \tau\delta^2\mathcal{E}$ does not couple them to the other variations. The quadratic form (26) decays, therefore, into a sum, of which the first term has the matrix

$$-\frac{1}{2}\begin{bmatrix} \frac{M\tau}{Nmr_0r_1} & -1 & 0 & 0 & 0 & \cdots \\ -1 & \frac{M\tau}{2Nmr_1}\left(\frac{1}{r_0}+\frac{1}{r_2}\right) & 1 & 0 & 0 & \cdots \\ 0 & 1 & \frac{M\tau}{Nmr_1r_2} & -1 & 0 & \cdots \\ 0 & 0 & -1 & \frac{M\tau}{2Nmr_2}\left(\frac{1}{r_1}+\frac{1}{r_3}\right) & 0 & \cdots \\ 0 & 0 & 0 & 1 & \frac{M\tau}{Nmr_2r_3} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{bmatrix}, \tag{58}$$

in terms of the variations

$$\left(\frac{\delta L_{\frac{1}{2}}}{M^{\frac{1}{2}}}, M^{\frac{1}{2}}\delta\theta_1, \frac{\delta L_{\frac{3}{2}}}{M^{\frac{1}{2}}}, M^{\frac{1}{2}}\delta\theta_2, \frac{\delta L_{\frac{5}{2}}}{M^{\frac{1}{2}}}, \cdots\right). \tag{59}$$

The normalization in (59) with the help of M has been chosen such as to make the matrix (58) dimensionless. The eigenvalues and the determinant of (58) will be discussed in Appendix D. This part of the second variation (26) can be fully understood without any difficulty.

The remainder can be simplified if we integrate immediately over the variations $\delta\varphi_n$ and then over all but one of the variations δM_n. This provides a factor $(2\pi\hbar)^{N-1}$ to the integral (54) and reduces the second variation (26) because all $\delta\varphi_n$ have been eliminated and all δM_n have been replaced by a single one δM. Thus, the second part of the quadratic form (26) can be represented by the matrix

$$-\frac{1}{2}\begin{bmatrix} \sum_1^N \frac{\tau}{Nmr_{n-1}r_n} & 0 & -\frac{2M\tau}{Nmr_1^3} & 0 & -\frac{2M\tau}{Nmr_2^3} & \cdots \\ 0 & \tau/Nm & -1 & 0 & 0 & \cdots \\ -\frac{2M\tau}{Nmr_1^3} & -1 & \tau\bar{V}_{11}/N & 1 & \tau\bar{V}_{12}/N & \cdots \\ 0 & 0 & 1 & \tau/Nm & -1 & \cdots \\ -\frac{2M\tau}{Nmr_2^3} & 0 & \tau\bar{V}_{21}/N & -1 & \tau\bar{V}_{22}/N & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{bmatrix} \tag{60}$$

in terms of the variations

$$(\delta M, \delta s_{\frac{1}{2}}, \delta r_1, \delta s_{\frac{3}{2}}, \delta r_2, \cdots). \tag{61}$$

The quantities $\bar{V}_{jl}$ are defined as

$$\bar{V}_{jl} = \frac{\partial^2}{\partial r_j \partial r_l}\left\{\sum_0^N V(r_n) + \sum_{\frac{1}{2}}^{N-\frac{1}{2}} \frac{M^2}{2m r_{n-\frac{1}{2}} r_{n+\frac{1}{2}}}\right\}. \tag{62}$$

The variation δM can be eliminated from the quadratic form (60) with the help of the subsidiary condition $\delta^1 \mathcal{E} = 0$, i.e.,

$$\delta M \sum_1^N \frac{M}{m r_n r_{n-1}} + \sum_{\frac{1}{2}}^{N-\frac{1}{2}} \frac{s_n}{m} \delta s_n + \sum_1^{N-1} \bar{V}_n \delta r_n = 0, \tag{63}$$

where $\bar{V}_n$ is the same kind of derivative as (62). The eigenvalues and the determinant of the resulting matrix are obtained in Appendix E. With the help of (27) the formula (54) for G is, therefore, reduced to

$$\bar{G} = \frac{-2\pi i}{r_0 r_N (2\pi\hbar)^N} \cdot e^{iS/\hbar} \int \prod_1^{N-1} d\theta_n \int \prod_{\frac{1}{2}}^{N-\frac{1}{2}} dL_n \exp \frac{i}{\hbar} [(58)]$$
$$\cdot (2\pi\hbar)^{-N-1} \int \prod_1^{N-1} dr_n \int \prod_{\frac{1}{2}}^{N-\frac{1}{2}} ds_n \left(\frac{\partial M}{\partial E}\right)_{r_j, s_l}$$
$$\times \exp \frac{i}{\hbar} [(60) \text{ with } (63)]. \tag{64}$$

The derivative $\partial M/\partial E$ at constant r_j and s_l is obtained from (53) with $\mathcal{E} = E$ after neglecting the $\hbar^2/4$ terms and setting $L_n = 0$, $\theta_n = \frac{1}{2}\pi$, as well as $M_n = M$. Compared to the exponentials, the variation of $\partial M/\partial E$ with r_j and s_l is slow, so that it can be evaluated for the classical trajectory and pulled out of the integral. If we insert the results of Appendixes D and E into (64), we find the formula (20) with the amplitude given by (46). This completes the discussion of the Garrod–Feynman integral for small $\hbar$ in the case of a spherically symmetric potential.

C. Polar Coordinates in Momentum Space

For a spherically symmetric potential, the classical trajectory in momentum space lies again in the plan which is spanned by the initial and the final momentum. The action $T(p'' p' E)$ depends only on the absolute values ρ' and ρ'' of p' and p'', and on the angle η between p' and p''. The determinant (36) in the amplitude of $\tilde{F}(p'' p' E)$ is now given by

$$D_T = \frac{T_\eta}{\rho'^2 \rho''^2 \sin\eta} \begin{vmatrix} T_{\rho'\rho''} & T_{\rho'\eta''} & T_{\rho' E} \\ T_{\eta'\rho''} & T_{\eta'\eta''} & T_{\eta' E} \\ T_{E\rho''} & T_{E\eta''} & T_{EE} \end{vmatrix}, \tag{65}$$

where η' and η'' are the polar angles of p' and p'' in the plane of the classical trajectory. The indices on T indicate the derivatives of T with respect to these quantities. The determinant in (65) is found from the orbit which corresponds to ρ', ρ'', η, and E in a plane.

The variables which are conjugate to ρ and η are the projection σ of the position vector q onto the direction of motion and the angular momentum M. The radial distance r is given by

$$r^2 = (\partial T/\partial\rho)^2 + \rho^{-2}(\partial T/\partial\eta)^2 = \sigma^2 + M^2/\rho^2 \tag{66}$$

so that the Hamilton–Jacobi equation becomes

$$\rho^2/2m + V[(\partial T/\partial\rho)^2 + \rho^{-2}(\partial T/\partial\eta)^2]^{\frac{1}{2}} = E. \tag{67}$$

A more familiar-looking equation is obtained by introducing the inverse $r(V)$ of $V(r)$. Such an inverse exists for the typical potentials where the force of attraction increases monotonically as the distance from the center decreases. The new equation

$$(\partial T/\partial\rho)^2 + \rho^{-2}(\partial T/\partial\eta)^2 - r^2(E - \rho^2/2m) = 0 \tag{68}$$

looks like an ordinary Hamilton–Jacobi equation of a fictitious particle with polar coordinates ρ and η, at zero energy, in a radial potential given by

$$-\tfrac{1}{2} \cdot r^2(E - \rho^2/2m).$$

As ρ increases indefinitely, this radial potential vanishes; but since the energy is zero, the very large values of ρ become accessible. For $\rho = 0$, the potential has the value $-\frac{1}{2}r^2(E)$ where $r(E)$ is the maximum distance of the real particle with the energy $E < 0$.

In analogy to the formulas (44) through (47), one has now

$$\eta = \eta'' - \eta' = \int_{\rho'}^{\rho''} \frac{M\, d\rho}{\rho^2[r^2(E - \rho^2/2m) - M^2/\rho^2]^{\frac{1}{2}}}, \tag{69}$$

$$T = \int_{\rho'}^{\rho''} d\rho \frac{r^2(E - \rho^2/2m)}{[r^2(E - \rho^2/2m) - M^2/\rho^2]^{\frac{1}{2}}}, \tag{70}$$

$$D_T = \frac{M}{\rho'^2 \rho''^2 \sin\eta} \cdot \frac{\partial}{\partial E}\left[r^2\left(E - \frac{\rho''^2}{2m}\right) - \frac{M^2}{\rho''^2}\right]^{\frac{1}{2}}$$
$$\cdot \frac{\partial}{\partial E}\left[r^2\left(E - \frac{\rho'^2}{2m}\right) - \frac{M^2}{\rho'^2}\right]^{\frac{1}{2}} \Big/ \frac{\partial\eta}{\partial M}, \tag{71}$$

$$\frac{\partial\eta}{\partial M} = \int_{\rho'}^{\rho''} d\rho \frac{r^2(E - \rho^2/2m)}{\rho^2[r^2(E - \rho/2m) - M^2/\rho^2]^{\frac{3}{2}}}. \tag{72}$$

All remarks concerning the critical points in coordinate space apply again to (71) with respect to momentum space. $\tilde{F}$ can be computed according to (35). But, as explained earlier, it is of great interest to arrive at this result directly as the limit of small $\hbar$ of the Garrod–Feynman integral (32) for a spherically symmetric potential.

The first task is again to rewrite (32) in polar coordinates. The detailed correspondence between coordinate and momentum space is maintained by the coordinate transformation

$$\begin{aligned} p_n^1 &= \rho_n \sin\zeta_n \cos\eta_n, \\ p_n^2 &= \rho_n \sin\zeta_n \sin\eta_n, \\ p_n^3 &= \rho_n \cos\zeta_n, \end{aligned} \tag{73}$$

for half-integer n, and

$$\begin{aligned} q_n^1 &= \left(\sigma_n \overline{\sin\zeta} + \frac{L_n}{\bar\rho}\overline{\cos\zeta}\right)\cos\bar\eta - \frac{M_n}{\bar\rho \cdot \overline{\sin\zeta}}\sin\bar\eta, \\ q_n^2 &= \left(\sigma_n \overline{\sin\zeta} + \frac{L_n}{\bar\rho}\overline{\cos\zeta}\right)\sin\bar\eta + \frac{M_n}{\bar\rho \cdot \overline{\sin\zeta}}\cos\bar\eta, \\ q_n^3 &= \sigma_n \overline{\cos\zeta} - \frac{L_n}{\bar\rho}\overline{\sin\zeta}, \end{aligned} \tag{74}$$

for integer n, where $\bar\rho = (\rho_{n-\frac{1}{2}}\rho_{n+\frac{1}{2}})^{\frac{1}{2}}$, $\overline{\cos\zeta} = (\cos\zeta_{n-\frac{1}{2}}\cos\zeta_{n+\frac{1}{2}})^{\frac{1}{2}}$, $\overline{\sin\zeta} = (\sin\zeta_{n-\frac{1}{2}}\sin\zeta_{n+\frac{1}{2}})^{\frac{1}{2}}$, and $\bar\eta = \frac{1}{2}(\eta_{n-\frac{1}{2}} + \eta_{n+\frac{1}{2}})$. The integral (32) is then transformed into

$$\begin{aligned} F = \lim_{N\to\infty} &(2\pi\hbar)^{-3N-3} \\ &\times \int \prod_0^N d\sigma_n\, dL_n\, dM_n \int \prod_{\frac{1}{2}}^{N-\frac{1}{2}} d\rho_n\, d\eta_n\, d\zeta_n \\ &\times (\rho'^2\rho''^2 \sin\zeta' \sin\zeta'')^{-\frac{1}{2}} \exp\left[-(i/\hbar)T_N\right]/(E-\mathcal{E}), \end{aligned} \tag{75}$$

where the action T_N along the path in phase space is given by

$$\begin{aligned} T_N = \sum_0^N &[\sigma_n(\rho_{n+\frac{1}{2}} - \rho_{n-\frac{1}{2}}) + L_n(\zeta_{n+\frac{1}{2}} - \zeta_{n-\frac{1}{2}}) \\ &+ M_n(\eta_{n+\frac{1}{2}} - \eta_{n-\frac{1}{2}})] \\ &+ [\text{terms at least quadratic in } (\rho_{n+\frac{1}{2}} - \rho_{n-\frac{1}{2}}), \\ &\qquad (\zeta_{n+\frac{1}{2}} - \zeta_{n-\frac{1}{2}}), (\eta_{n+\frac{1}{2}} - \eta_{n-\frac{1}{2}})], \end{aligned} \tag{76}$$

and the average energy $\mathcal{E}$ can be written as

$$\mathcal{E} = \frac{1}{N}\sum_{\frac{1}{2}}^{N-\frac{1}{2}} \frac{\rho_n^2}{2m} + \frac{1}{N}\sum_0^N V(|q_n|) \tag{77}$$

with

$$\begin{aligned} |q_n|^2 = \sigma_n^2 &+ \frac{L_n^2}{\rho_{n-\frac{1}{2}}\rho_{n+\frac{1}{2}}} + \frac{M_n^2}{\rho_{n-\frac{1}{2}}\rho_{n+\frac{1}{2}} \sin\zeta_{n-\frac{1}{2}} \sin\zeta_{n+\frac{1}{2}}} \\ &- \left(\sigma_n^2 + \frac{L_n^2}{\rho_{n-\frac{1}{2}}\rho_{n+\frac{1}{2}}}\right)(1 - \cos(\zeta_{n+\frac{1}{2}} - \zeta_{n-\frac{1}{2}})). \end{aligned} \tag{78}$$

One would like to get rid of the last terms in (76) and (78). Obviously, they can not simply be neglected, since even in the expression (54) for G additional terms, $-\hbar^2/8mr_{n+\frac{1}{2}}r_{n-\frac{1}{2}}$ and

$$-\hbar^2/8mr_{n+\frac{1}{2}}r_{n-\frac{1}{2}} \sin\theta_{n+\frac{1}{2}} \sin\theta_{n-\frac{1}{2}},$$

had to be inserted into $\mathcal{E}$. The arguments of the previous section are not applicable because F satisfies an integral equation, rather than a Fokker–Planck-like Schrödinger equation. More than just the zero, first, and second moment of the propagator for small times are needed in momentum space.

It is not clear whether simple formulas like (53), (54), (55) can be found for F in polar coordinates as $N \to \infty$. As $\hbar$ goes to zero, however, the variations in the coordinate differences $(\rho_{n+\frac{1}{2}} - \rho_{n-\frac{1}{2}})$, $(\sigma_n - \sigma_{n-1})$, etc., become small. It is sufficient to keep only the first parts of (76) and (78). The correspondence between the formulas (53) through (55) for G and the formulas (75) through (78) for F is complete again. $V(r_n)$ in (53) is replaced by $\rho_n^2/2m$ in (77), and the kinetic energy term, $[s_n^2 + \cdots]/2m$, in (53) is replaced by the potential energy, $V([\sigma_n^2 + \cdots]^{\frac{1}{2}})$, in (77). In order to apply the arguments of the previous section to the discussion of (75) in the limit of vanishing $\hbar$, they have to be sufficiently general so as to include a kinetic energy which is not simply the square of the momentum. The Appendices D and E treat this general case, and are, therefore, immediately applicable to the formulas (75) through (78), after the preliminary steps corresponding to the formulas (56) through (64) for $\tilde G$ have been completed.

In this manner we are ultimately again lead to the expression (35) for $\tilde F$ with the expression (71) for D_T and the phase jumps at focal points which were discussed earlier. It is evident from the arguments in the Appendices D and E that the rotational invariance has been used extensively, so that the limit of small $\hbar$, in the Garrod–Feynman integral has been established only for potentials of spherical symmetry.

IV. PHASE-INTEGRAL APPROXIMATION FOR THE COULOMB PROBLEM IN MOMENTUM SPACE

A. Classical Kepler Orbits in Momentum Space

The orbits in momentum space can be obtained in a straightforward manner if one computes the integral (69) with the Coulomb potential

$$V(r) = -e^2/r. \tag{79}$$

It seems, however, more appealing to describe these orbits in a geometric manner, particularly because they turn out to be so simple.

Starting from the trajectory in coordinate space

$$r = (M^2/me^2)(1 + \epsilon\cos\varphi)^{-1}, \tag{80}$$

$$\epsilon = [1 + 2M^2E/me^4]^{\frac{1}{2}}, \tag{81}$$

one gets immediately the momenta

$$p_1 = m\frac{dq_1}{dt} = m\frac{dq_1}{d\varphi}\cdot\frac{d\varphi}{dt} = -\frac{me^2}{M}\sin\varphi,$$
$$p_2 = m\frac{dq_2}{dt} = m\frac{dq_2}{d\varphi}\cdot\frac{d\varphi}{dt} = \frac{me^2}{M}(\epsilon + \cos\varphi). \tag{82}$$

[The Cartesian components of p and q are called (p_1, p_2) and (q_1, q_2) in this section.] If φ is eliminated between the last two equations, we get the equation of a circle in momentum space

$$p_1^2 + [p_2 - (me^2/M)\epsilon]^2 = (me^2/M)^2, \tag{83}$$

with the radius me^2/M and the center at a distance $\epsilon me^2/M$ from the origin. All orbits in momentum space intersect a circle of radius $(-2mE)^{\frac{1}{2}}$ around the origin at diametrically opposite points, as can be recognized from the solution $p_1 = \pm(-2mE)^{\frac{1}{2}}$ and $p_2 = 0$ of (83). Conversely, for any circle in momentum space which intersects the circle of radius $(-2mE)^{\frac{1}{2}}$ around the origin, we can find a value M between 0 and $(-me^4/2E)^{\frac{1}{2}}$ such that its radius is given by me^2/M and the distance of its center from the origin by $(2mE + m^2e^4/M^2)^{\frac{1}{2}}$.

Let us now find the locus of the centers of all such circles which pass through a given point, say $(\rho, 0)$, for a given energy $E < 0$. Suppose that one such circle goes through the point

$$[(-2mE)^{\frac{1}{2}}\cos\alpha, (-2mE)^{\frac{1}{2}}\sin\alpha].$$

Its center (p_1, p_2) lies, therefore, on the bisectrix given by

$$(p_1 - \rho)^2 + p_2^2 = (p_1 - (-2mE)^{\frac{1}{2}}\cos\alpha)^2 + (p_2 - (-2mE)^{\frac{1}{2}}\sin\alpha)^2,$$

as well as on the straight line through the origin and perpendicular to the direction $(\cos\alpha, \sin\alpha)$, i.e.,

$$p_1\cos\alpha + p_2\sin\alpha = 0.$$

If we eliminate α from these two equations, we find

$$p_1 = \tfrac{1}{2}[\rho + (2mE/\rho)]. \tag{84}$$

The locus of the centers of all orbits through $(\rho, 0)$ is the straight line perpendicular to $(\rho, 0)$ at a distance $\frac{1}{2}[\rho + (2mE/\rho)]$ from the origin. For $\rho < (-2mE)^{\frac{1}{2}}$ the quantity (84) is negative so that the origin lies on the same side of the locus as the point $(\rho, 0)$.

It is now easy to find the center of the orbit in momentum space which passes through two given momenta, $\mathbf{p}'$ and $\mathbf{p}''$. We have only to intersect the two loci for the centers of the circles through $\mathbf{p}'$ and through $\mathbf{p}''$. Since these loci are straight lines, there is exactly one intersection. We find, therefore, the important statement that: *For given $E < 0$ there is exactly one classical orbit in momentum space which connects a given initial momentum $\mathbf{p}'$ with a given final momentum $\mathbf{p}''$.* The exception to this statement arises in the special case where p' and p'' are "opposite" each other with respect to the circle of radius $(-2mE)^{\frac{1}{2}}$ around the origin, i.e., $\mathbf{p}'' = 2mE\mathbf{p}'/|p'|^2$. In that case all orbits through $\mathbf{p}'$ go also through $\mathbf{p}''$.

This last configuration is of particular interest because it turns out that all the classical trajectories starting from a momentum $\mathbf{p}'$ intersect one another at the "opposite" momentum, and nowhere else. Again this situation is much simpler than for the Coulomb potential in coordinate space where all the classical trajectories of a given energy $E < 0$ starting from a position $\mathbf{q}'$ touch one another along a caustic. In momentum space this caustic has seemingly contracted into a point.

The action function (34) can be obtained from (80) and (82) by writing

$$T = -\int_{\varphi'}^{\varphi''}\left(r\cos\varphi\frac{dp_1}{d\varphi} + r\sin\varphi\frac{dp_2}{d\varphi}\right)d\varphi$$
$$= \int_{\varphi'}^{\varphi''}\frac{L\,d\varphi}{1 + \epsilon\cos\varphi} = \left(-\frac{me^4}{2E}\right)^{\frac{1}{2}}(u'' - u'), \tag{85}$$

where u is the "eccentric anomaly" which is given by

$$u = 2\arctan[(1 - \epsilon)/(1 + \epsilon)]^{\frac{1}{2}}\tan\varphi/2. \tag{86}$$

The "true anomaly" φ is measured from the point of closest approach, the perihelion.

B. Phase Integral Approximation

In order to find explicit expressions for the approximate Green's function $\tilde{F}(p'' p' E)$ as given by (35), one has to find a relation between the polar angle η in momentum space and the angular momentum M which occurs in the formulas (70) and (71). In terms of the quantity

$$P = \frac{1}{2}\left[\frac{|\mathbf{p}|}{(-2mE)^{\frac{1}{2}}} - \frac{(-2mE)^{\frac{1}{2}}}{|\mathbf{p}|}\right]$$
$$= \frac{1}{2}\left[\frac{\rho}{(-2mE)^{\frac{1}{2}}} - \frac{(-2mE)^{\frac{1}{2}}}{\rho}\right], \tag{87}$$

one finds after some obvious algebra that

$$M = (me^4/2E)^{\frac{1}{2}}\sin\eta[P''^2 - 2P'P''\cos\eta + P'^2 + \sin^2\eta]^{-\frac{1}{2}}. \tag{88}$$

The determinant D_T in (35) is then obtained from (71) as

$$D_T = \frac{me^8}{-2E\rho'^2\rho''^2(-E + \rho'^2/2m)(-E + \rho''^2/2m)} \times (P''^2 - 2P'P''\cos\eta + P'^2 + \sin^2\eta). \tag{89}$$

The denominator vanishes only if $P'' \to P'$ at the same time as $\eta \to 0$, or if $P'' \to -P'$ at the same time as $\eta \to \pi$. The latter case corresponds to $\mathbf{p}''$ being "opposite" to $\mathbf{p}'$. For the action function T, we find from (86) that

$$T = \left(-\frac{me^4}{2E}\right)^{\frac{1}{2}} \times \arctan \frac{[P''^2 - 2P'P'' \cos\eta + P'^2 + \sin^2\eta]^{\frac{1}{2}}}{(P'P'' + \cos\eta)}. \tag{90}$$

The arctan as well as the root in its argument are uniquely defined in the range $(0, \pi)$ for $0 < \rho' < \infty$, and $0 < \eta < \pi$. As any one of the three independent variables in (90) reaches the end of its domain, there is always a natural definition for T to preserve its continuity. It suffices to construct the corresponding classical trajectory which does not go through the "opposite" momentum in order to find the correct value of T.

The action T from a momentum $\mathbf{p}'$ to its opposite is always given by $\pi(-me^4/2E)^{\frac{1}{2}}$. As we follow the trajectory through the opposite momentum to a final momentum $\mathbf{p}''$, the total action accumulated is given by

$$T = \left(-\frac{me^4}{2E}\right)^{\frac{1}{2}} \times \left\{2\pi - \arctan \frac{[P''^2 - 2P'P'' \cos\eta + P'^2 + \sin^2\eta]^{\frac{1}{2}}}{P'P'' + \cos\eta}\right\}; \tag{91}$$

η is again the angle between $\mathbf{p}'$ and $\mathbf{p}''$ measured as in the case of (90) and restricted to the interval $0 < \eta < \pi$. If we follow the orbit any further, the total action can be obtained from (90) or (91) by adding as many times $2\pi(-me^4/2E)^{\frac{1}{2}}$ as full orbits have been completed.

In order to apply the formula (35), we have to determine the extra phase factors which come from the critical points along the classical trajectory. For the Kepler orbits in momentum space the two kinds of critical points discussed in Sec. 4 coincide, since all trajectories of given energy E leaving a given momentum $\mathbf{p}'$ meet again at the opposite momentum whatever the direction or the absolute value of their angular momentum. A factor $\exp(-i\pi) = -1$ is picked up for each traversal of such a doubly critical point. The same factor enters into (35) when the trajectory goes through the initial point $\mathbf{p}'$ again.

Since both (90) and (91) are expressed in terms of the angle η which is defined by $\mathbf{p}' \cdot \mathbf{p}'' = \rho'\rho'' \cos\eta$, it seems appropriate to use this scalar product in (89), (90), and (91) rather than η. It should be noticed that the amplitude $(D_T)^{\frac{1}{2}}$ stays the same for all the trajectories which go from $\mathbf{p}'$ to $\mathbf{p}''$, because according to (36) only the derivatives of T with respect to $\mathbf{p}'$ and $\mathbf{p}''$ are needed, whereas, the actions along different trajectories from $\mathbf{p}'$ to $\mathbf{p}''$ differ only by multiples of $2\pi(-me^4/2E)^{\frac{1}{2}}$ and possibly a sign. The summation over all trajectories reduces to the geometric series of the powers of $\exp[2\pi i(-me^4/2E\hbar^2)^{\frac{1}{2}}]$. After some rearranging we can finally write for $\tilde{F}(p'' p' E)$ the expression

$$-\frac{4m^2e^4}{\pi\hbar^2(\rho'^2 - 2mE)(\rho''^2 - 2mE)\,|\mathbf{p}'' - \mathbf{p}'|\,[(\rho'^2 - 2mE)(\rho''^2 - 2mE) + 2mE(\mathbf{p}'' - \mathbf{p}')^2]^{\frac{1}{2}}} \cdot \sin\left\{2\left(\frac{-me^4}{2E\hbar^2}\right)^{\frac{1}{2}} \arctan\left[\frac{(\rho'^2 - 2mE)(\rho''^2 - 2mE) + 2mE(\mathbf{p}'' - \mathbf{p}')^2}{-2mE(\mathbf{p}'' - \mathbf{p}')^2}\right]^{\frac{1}{2}}\right\} \Big/ \sin\pi\left(\frac{-me^4}{2E\hbar^2}\right)^{\frac{1}{2}}. \tag{92}$$

The arctan varies between 0, when $\mathbf{p}'$ and $\mathbf{p}''$ are opposite each other, and $\frac{1}{2}\pi$, when $\mathbf{p}' = \mathbf{p}''$. Therefore, the amplitude of (92) becomes infinite as $\mathbf{p}''$ approaches $\mathbf{p}'$, but it stays finite as $\mathbf{p}''$ goes through the opposite momentum of $\mathbf{p}'$. The only singularity in (92) is the one which was described by the formula (39).

The above results for the approximate Green's function of the Coulomb problem in momentum space is so simple because the caustics have shrunk to points. The corresponding function $\tilde{G}(q'' q' E)$ in coordinate space is expected to be more complicated, although it will have the same denominator. In this connection, the three-dimensional harmonic oscillator is of interest, because it combines the features of the Coulomb problem in momentum as well as in coordinate space. To each initial point, $\mathbf{q}'$ or $\mathbf{p}'$ corresponds an "opposite" point, $-\mathbf{q}'$ or $-\mathbf{p}'$, where all trajectories through $\mathbf{q}'$ or $\mathbf{p}'$ meet again. But in between, the trajectories belonging to one plane touch one another along an envelope. There will be effectively a total of six critical points for each full oscillator orbit, whereas there are only four critical points for each Kepler orbit, whether in momentum or in coordinate space. The quantum condition of Bohr and Sommerfeld requires, therefore, half-integer quantum numbers for the three-dimensional oscillator (with a minimum of $\frac{3}{2}$); but for the Coulomb problem, the quantum numbers are integer, as shown above.

C. Comparison with the Exact Green's Function

The main purpose for studying the phase-integral approximation in momentum space was to find wavefunctions for bound states. The formula (92) is, therefore, of greatest interest for negative values of the energy. We shall put $-2mE = \gamma^2$ with a real $\gamma > 0$, whenever the Green's function is examined along the negative E axis. The expression under the root in (92) can then be written as

$$(\rho'^2 + \gamma^2)(\rho''^2 + \gamma^2) - \gamma^2(\mathbf{p}'' - \mathbf{p}')^2 = \rho'^2\rho''^2 + \gamma^4 + 2\gamma^2\rho'\rho'' \cos\eta,$$

which vanishes if and only if $\rho'\rho'' = \gamma^2$ and $\cos\eta = -1$. The only singularities in (92) along the negative real E axis arise from the zeros of the denominator. The corresponding poles are at

$$E = -me^4/2\hbar^2n^2, \tag{93}$$

where n is a positive integer. The Bohr formula is obtained without any gimmickry. It would have resulted with similar ease from $\tilde{G}(q'' q' E)$.

The residues at the poles (93) can be most conveniently expressed in terms of the Bohr momentum $\gamma_n = me^2/n\hbar$ and the angular variable

$$\beta = 2 \arctan \left[\frac{(\rho'^2 + \gamma^2)(\rho''^2 + \gamma^2) - \gamma^2(\mathbf{p}'' - \mathbf{p}')^2}{\gamma^2(\mathbf{p}'' - \mathbf{p}')^2}\right]^{\frac{1}{2}}. \tag{94}$$

The residue at the pole (93) is found to be

$$(-1)^{n+1}(8n/\pi^2)\gamma_n^5(\rho'^2 + \gamma_n^2)^{-2}(\rho''^2 + \gamma_n^2)^{-2} \times (\sin n\beta_n/\sin\beta_n). \tag{95}$$

The quotient $\sin n\beta/\sin\beta$ is a rational function of the momenta $\mathbf{p}'$ and $\mathbf{p}''$, as is seen immediately if $\sin n\beta$ is expanded in terms of $\sin\beta$ and $\cos\beta$ and the elementary formulas

$$\sin(2 \arctan\alpha) = 2\alpha/(1 + \alpha^2),$$
$$\cos(2 \arctan\alpha) = (1 - \alpha^2)/(1 + \alpha^2) \tag{96}$$

are used together with (94).

As an example let us put $n = 1$ in (95), which gives

$$(8/\pi^2)\gamma_1^5(\rho'^2 + \gamma_1^2)^{-2}(\rho''^2 + \gamma_1^2)^{-2}. \tag{97}$$

According to Bethe and Salpeter,[20] this expression is just the product of the two normalized $1s$ functions of the hydrogen spectrum, with variable ρ'^2 and ρ''^2, respectively. Similarly, we obtain for $n = 2$ from (95) and (96)

$$(32/\pi^2)\gamma_2^5(\rho'^2 + \gamma_2^2)^{-3}(\rho''^2 + \gamma_2^2)^{-3} \times \{(\rho'^2 - \gamma_2^2)(\rho''^2 - \gamma_2^2) + 4\gamma_2^2(\mathbf{p}', \mathbf{p}'')\}. \tag{98}$$

If we expand the scalar product $(\mathbf{p}', \mathbf{p}'') = p_1'p_1'' + p_2'p_2'' + p_3'p_3''$, we are left with four terms in the braces. Together with the factors in front, each term is again a product of normalized hydrogen wavefunctions, the first term in the braces providing the $2s$ function and the last three terms the three $2p$ functions.

Instead of comparing further the residues (95) with the known hydrogen wavefunctions in momentum space, it is more efficient to compare directly the approximate Green's function (92) with the exact one. The latter has been obtained in closed form by Okubo[21] and has recently been discussed by other authors.[22] Along the negative real E axis we can write

$$F(p'' p' E) = -\frac{2m\delta(\mathbf{p}'' - \mathbf{p}')}{\rho'^2 + \gamma^2} - \frac{4m^2e^2}{2\pi^2\hbar(\rho'^2 + \gamma^2)(\rho''^2 + \gamma^2)(\mathbf{p}'' - \mathbf{p}')^2} - \frac{8m^3e^4\gamma}{\pi^2\hbar^2(\rho'^2 + \gamma^2)(\rho''^2 + \gamma^2)} \int_1^\infty d\zeta\zeta^{me^2/\hbar\gamma} \times [(\zeta - 1)^2(\rho'^2 + \gamma^2)(\rho''^2 + \gamma^2) + 4\zeta\gamma^2(\mathbf{p}'' - \mathbf{p}')^2]^{-1}. \tag{99}$$

The first two terms are obtained from the integral equation (29) by a formal expansion of F in powers of the potential. The last term is formally of the same power in $\hbar$ as the approximation (35) for the Green's function which led us earlier to consider the expression (40) as being possibly superior to (35), especially for small $|\mathbf{p}'' - \mathbf{p}'|$.

The last term in (99) can be regarded as a Mellin transform with $me^2/\hbar\gamma$ as the new variable instead of ζ. Since the function of ζ differs from zero only in the interval from 1 to ∞ where it can also be expressed in powers of $1/\zeta$, the integral over ζ presents no great difficulties. Thus, we can write the last term in (99) for $1 > me^2/\hbar\gamma$ as

$$\frac{8m^3e^4\gamma}{\pi^2\hbar^2(\rho'^2 + \gamma^2)^2(\rho''^2 + \gamma^2)^2} \int_1^\infty d\zeta\zeta^{me^2/\hbar\gamma} \times \sum_{n=1}^\infty \frac{(-1)^{n+1}}{\zeta^{n+1}} \cdot \frac{\sin n\beta}{\sin\beta} = \frac{8m^3e^4\gamma}{\pi^2\hbar^2(\rho'^2 + \gamma^2)^2(\rho''^2 + \gamma^2)^2} \times \sum_{n=1}^\infty \frac{(-1)^{n+1}}{(me^2/\hbar\gamma) - n} \cdot \frac{\sin n\beta}{\sin\beta}. \tag{100}$$

The poles of (99) in the left-hand part of the complex E plane are correctly given by (100), but the expansion converges only on the negative real E axis and

[20] H. A. Bethe and E. E. Salpeter, "Quantum Mechanics of One- and Two-electron Systems," in *Encyclopedia of Physics* (Springer-Verlag, Berlin, 1957), p. 125.

[21] S. Okubo and D. Feldman, Phys. Rev. 117, 292 (1960).

[22] L. Hostler, J. Math. Phys. 5, 1235 (1964); J. J. Schwinger, J. Math. Phys. 5, 1606 (1964).

is, therefore, useless as a representation of the last term in (99). Indeed, the angle β becomes complex for values of E off the negative real axis, so that $\sin n\beta$ becomes exponentially larger for increasing n. The poles of (99) and their residues are shown by (100) to be the same as (93) and (95). The phase-integral approximation is thus shown to yield exactly all the bound states of the hydrogen atom.

Although we have thereby achieved the main goal of this paper, it may be of interest to discuss just a few points which are concerned with the approximate Green's function $\tilde{F}$ for positive values of the energy. If we insert $-2mE = \gamma^2 e^{-i\omega}$ into (92) and let ω grow from 0 to π, we get the analytic continuation of (92) from the negative to the positive real axis through the upper half-plane. In order to go from some point E' on the negative E axis to a point E'' on the positive E axis, we can either first adjust γ for $\omega = 0$ and then let ω grow to π, or we can first let ω go from 0 to π and then adjust γ^2. These two procedures give the same purely imaginary result, whatever the vectors $\mathbf{p}'$ and $\mathbf{p}''$. A well-defined discontinuity across the positive real axis is obtained between the results of the analytic continuation through the upper and through the lower half of the complex E plane. The formula (92) for $\tilde{F}$ has all the attributes of a well-behaved Green's function, which is all the more surprising because, for a given positive energy, certain parts of momentum space are classically inaccessible.

It would be interesting to compare the discontinuities across the positive E axis for the phase-integral approximation $\tilde{F}$ with those of the exact Green's function (99). This seems almost more difficult than the comparison of the bound states, because the latter form a countable set, whereas the former form a continuum which depends on the three variables ρ', ρ'', and $\cos\eta$. This problem is, therefore, not examined at this time, although it may be of interest for the discussion of scattering problems and virtual bound states. The original goal still seems of greater importance, i.e., the extension of the phase-integral method to bound states in general spherically symmetric potentials and eventually to simple molecular potentials such as in the diatomic molecules. The successful treatment of the hydrogen atom which was presented in this investigation, constitutes a crucial first step in this direction.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Dr. W. Schlup for many interesting discussions, and to Dr. F. Odeh and Dr. B. Weiss for enlightening remarks about the existence and the meaning of Morse theory.

APPENDIX A

The starting point is the identity

$$\left(\frac{m}{2\pi i\hbar t}\right)^{\frac{3}{2}} \exp\left[\frac{im(q''-q')^2}{2\hbar t}\right] = \frac{1}{(2\pi\hbar)^3}\int d^3p \exp\frac{i}{\hbar}\left[p(q''-q') - t\frac{p^2}{2m}\right], \tag{A1}$$

which can be used in the Feynman integral (5) between any two consecutive points q_j and q_{j+1}, if the Lagrangian has the classical form $L = m\dot{q}^2/2 - V(q)$. If one calls $p_{j+\frac{1}{2}}$ the momentum between q_j and q_{j+1}, the expression (5) becomes

$$K = \lim_{N\to\infty}(2\pi\hbar)^{-3N}\int\prod_{1}^{N-1} d^3q_n \int\prod_{\frac{1}{2}}^{N-\frac{1}{2}} d^3p_n \times \exp\frac{i}{\hbar}[S_N - (t''-t')\mathcal{E}], \tag{A2}$$

with the abbreviations S_N and $\mathcal{E}$ as defined by (16) and (18). The time intervals $t_{j+1} - t_j$ have all been chosen equal in order to simplify the definition of $\mathcal{E}$. The Fourier transform (13) immediately yields the expression (17) for $G(q''\,q'\,E)$, if the integration over t from 0 to ∞ is interpreted as a Laplace integral where E has a small positive imaginary part.

APPENDIX B

If (8) is inserted into (13) the exponent becomes $R(q''t, q'0) + Et$ apart from the factor $i/\hbar$. For given values of q'', q', and E this exponent is stationary for t_0 defined by the equation

$$-\partial R/\partial t = E. \tag{B1}$$

The exponent is now expanded in powers of $t - t_0$, so that

$$R(q''t, q'0) + Et = R(q''t_0, q'0) + Et_0 + \tfrac{1}{2}(t-t_0)^2(\partial^2R/\partial t^2)|_{t_0} + \cdots. \tag{B2}$$

The factor $(D_R)^{\frac{1}{2}}$ in (8) is assumed to vary slowly and can, therefore, be evaluated at $t = t_0$ without any further corrections. The integration over t is elementary and gives

$$\tilde{G}(q''\,q'\,E) = -(1/2\pi\hbar^2)(D_R)^{\frac{1}{2}}\left(\frac{\partial^2R}{\partial t^2}\bigg|_{t_0}\right)^{-\frac{1}{2}} \times \exp[iS(q''\,q\,E)/\hbar], \tag{B3}$$

$$S(q''\,q'\,E) = R(q''t_0, q'0) + Et_0, \tag{B4}$$

where t_0 is to be eliminated with the help of (B1).

The second derivatives of R have to be written in

terms of the derivatives of S. One finds that

$$\frac{\partial^2 R}{\partial q'' \partial q'} = \frac{\partial^2 S}{\partial q'' \partial q'} - \frac{\partial^2 S}{\partial q'' \partial E} \cdot \frac{\partial_2 S}{\partial q' \partial E} \Big/ \frac{\partial^2 S}{\partial E^2}, \quad \text{(B5)}$$

$$\frac{\partial^2 R}{\partial t^2} = -\left(\frac{\partial^2 S}{\partial E^2}\right)^{-1}, \quad \text{(B6)}$$

which leads immediately to the wave amplitude $(D_S)^{\frac{1}{2}}$ in (20).

The 3×3 determinant $|\partial^2 S/\partial q'' \partial q'|$ vanishes, because the equation $H(\partial S/\partial q'', q'') = E$ can be differentiated with respect to q', which gives three linear homogeneous equations in the quantities $\partial H/\partial p''$ with the matrix $\partial^2 S/\partial q' \partial q''$. An interesting special case of this remark arises in one dimension, where one has the well-known formula

$$S(q'' q' E) = \int_{q'}^{q''} [2m(E - V(q))]^{\frac{1}{2}} dq = S(q'' E) - S(q' E) + \text{const.} \quad \text{(B7)}$$

The expression (20) becomes, therefore, after adjusting the normalization

$$\frac{1}{2\pi\hbar} \sum_{\text{classical paths}} e^{i \cdot \text{const}/\hbar} \cdot \left(\frac{\partial^2 S}{\partial E \partial q''}\right)^{\frac{1}{2}} \times e^{iS''/\hbar} \left(\frac{\partial^2 S}{\partial E \partial q'}\right)^{\frac{1}{2}} e^{-iS'/\hbar}. \quad \text{(B8)}$$

The constant in (B7) which reappears in the exponent of (B8) is different for each classical path, depending on the number of cycles in the path. If these details are properly considered, one arrives at the formulas of Kohn and Sham.[12]

APPENDIX C

The main problem in obtaining the expansion (37) for the Hamiltonian in momentum space comes from the potential energy. The discussion is, therefore, limited to evaluating the potential term in (30), if the trial solution $B(p'' p' E) \exp[-iT(p'' p' E)/\hbar]$ is inserted. We shall present first a short and rather formal argument, and then attempt to give a more rigorous, although lengthy, proof.

The inverse Fourier transform of (31) is given by

$$V(q) = \int V(p'' p') \exp\left[\frac{i}{\hbar}(p'' - p')q\right] d^3p'', \quad \text{(C1)}$$

and the derivatives of $V(q)$ are written as

$$\frac{\partial V}{\partial q} = \frac{i}{\hbar} \int V(p'' p') \cdot (p'' - p') \times \exp\left[\frac{i}{\hbar}(p'' - p')q\right] d^3p'', \quad \text{(C2)}$$

etc. One can write, therefore, the following sequence of equations:

$$\int d^3p V(p'' p) \cdot B(p p' E) \exp\left[-\frac{i}{\hbar} T(p p' E)\right] = B(p'' p' E) \exp\left[-\frac{i}{\hbar} T(p'' p' E)\right] \quad \text{(C3)}$$

$$\int d^3p V(p'' p) \frac{B(p p' E)}{B(p'' p' E)} \exp\frac{i}{\hbar}[-T(p p' E) + T(p'' p' E)]$$

$$= B(p'') \exp\left[\frac{i}{\hbar} T(p'')\right] \int d^3p V(p'' p) \times \left\{\frac{B(p)}{B(p'')} \exp\frac{i}{\hbar}\left[-T(p) + T(p'') + (p - p'')\frac{\partial T}{\partial p''}\right]\right\} \times \exp\left[\frac{i}{\hbar}(p'' - p)q''\right],$$

where $q'' = +\partial T/\partial p''$ and the argument p' as well as E has not been written anymore in the last line. The next step is the questionable one, since it consists in simply expanding the terms inside $\{\ \}$ in powers of $p - p''$. Thus, one obtains

$$\{\ \} = 1 + \frac{1}{B(p'')}(p - p'')\frac{\partial B}{\partial p''} + \frac{i}{2\hbar}(p - p'')(p - p'')\frac{\partial^2 T}{\partial p'' \partial p''} + \cdots, \quad \text{(C4)}$$

where the neglected terms would all contribute to the order $\hbar^2$ and higher. The expansion (37) for the Hamiltonian follows immediately with the help of (C2) and if we assume that $V(p'' p')$ depends only on the differences $p'' - p'$. If the expansion (C4) is carried further, higher terms in (37) are obtained without difficulty.

A more careful procedure consists in applying Parseval's theorem to the integral in the second line of (C3). Apart from the factor

$$B(p'' p' E) \exp[(-i/\hbar)T(p'' p' E)],$$

the potential-energy term in the Hamiltonian becomes

$$\int d^3q V(q) \cdot \frac{1}{(2\pi\hbar)^3} \int d^3p \frac{B(p)}{B(p'')} \times \exp\frac{i}{\hbar}[-T(p) + T(p'') - (p - p'')q]. \quad \text{(C5)}$$

The factor which multiplies $V(q)$ can be regarded as a density function $\sigma(q)$ which weighs the various contributions of the potential $V(q)$. One finds, indeed, that $\int \sigma(q)\, d^3q = 1$ whatever $B(p)$ and $T(p)$. It is, therefore, reasonable to study $\sigma(q)$, its main peak and its spread, particularly in the limit of small $\hbar$.

Jones and Kline[23] have investigated the asymptotic expansion of multiple integrals by the method of stationary phase. The method used in this Appendix differs somewhat and treats only interior critical points, although boundary points would have to be examined in a more complete theory. Also, we assume analyticity for the functions $B(p)$ and $T(p)$.

Given q, the exponent in the integrand of (C5) becomes stationary for a value $p = \bar{p}$ which is obtained from the equation

$$\partial T/\partial p = q. \tag{C6}$$

If we assume $\bar{p}$ to be a regular point of T, a real linear transformation

$$p^j - \bar{p}^j = \sum_l \alpha_{jl}\tilde{p}^l \tag{C7}$$

can be found, such that we can write the expansion

$$\begin{aligned} T(p) - T(\bar{p}) + (p-\bar{p})q &= \tfrac{1}{2}\sum_j \epsilon_j \tilde{p}^{j2} + \tfrac{1}{6}\sum_{jlm}\beta_{jlm}\tilde{p}^j\tilde{p}^l\tilde{p}^m \\ &\quad + \tfrac{1}{24}\sum_{jlmn}\beta_{jlmn}\tilde{p}^j\tilde{p}^l\tilde{p}^m\tilde{p}^n + \cdots. \end{aligned} \tag{C8}$$

The coefficients β are symmetric in all their indices, and $\epsilon_j = +1$ or -1. By purely algebraic manipulations, we can define coefficients γ for a further expansion

$$\tilde{p}^j = \hat{p}^j + \sum_{lm}\gamma_{jlm}\hat{p}^l\hat{p}^m + \sum_{lmn}\gamma_{jlmn}\hat{p}^l\hat{p}^m\hat{p}^n + \cdots, \tag{C9}$$

such that we have to all orders in $\hat{p}$ the equality

$$T(p) - T(\bar{p}) + (p-\bar{p})q = \tfrac{1}{2}\sum_j \epsilon_j \hat{p}^{j2}. \tag{C10}$$

If we require that the γ's are symmetric in all their indices, they are uniquely determined in terms of the β's, namely

$$\epsilon_j\gamma_{jlm} = -\tfrac{1}{6}\beta_{jlm}, \tag{C11}$$

$$\begin{aligned} \epsilon_j\gamma_{jlmn} &= -\tfrac{1}{24}\beta_{jlmn} \\ &\quad + \tfrac{1}{16}\sum_k \epsilon_k[\beta_{jlk}\beta_{kmn} + \beta_{jmk}\beta_{kln} + \beta_{jnk}\beta_{klm}]. \end{aligned} \tag{C12}$$

These relations increase rapidly in complexity.

The variable of integration in (C5) is now changed from p to $\hat{p}$. The Jacobian $\partial(p)/\partial(\hat{p})$ as well as $B(p)$ can be expressed in terms of α, β, and γ, and then expanded in powers of $\hat{p}$. The integration over $\hat{p}$ is trivial because of (C10) and yields after some obvious manipulations

$$\begin{aligned} &\frac{1}{(2\pi\hbar)^{\frac{3}{2}}B(p'')}\prod_k e^{i\epsilon_k\pi/4}\det|\alpha_{jl}(\bar{p})| \\ &\quad\times\exp\frac{i}{\hbar}[-T(p) + T(p'') - (\bar{p}-p'')q] \\ &\quad\cdot\Big\{B(\bar{p}) - \frac{i\hbar}{2}\sum_{jlm}\epsilon_l\epsilon_m\frac{\partial B}{\partial\bar{p}^j}\alpha_{jl}\beta_{lmm} \\ &\quad + \frac{i\hbar}{2}\sum_{jlm}\epsilon_m\frac{\partial^2 B}{\partial\bar{p}^j\partial\bar{p}^l}\alpha_{jm}\alpha_{lm} + i\hbar B(\bar{p})\Big[-\frac{1}{\gamma}\sum_{jl}\epsilon_j\epsilon_l\beta_{jjll} \\ &\quad + \tfrac{1}{24}\sum_{jlm}\epsilon_j\epsilon_l\epsilon_m(3\beta_{jjl}\beta_{lmm} + 2\beta_{jlm}\beta_{jlm})\Big] + \cdots\Big\}. \end{aligned} \tag{C13}$$

The whole expression is to be considered as a distribution function $\sigma(q)$, with $\bar{p}$ related to q through (C6).

The result (C13) is now multiplied by $V(q)$ and integrated over q, again by the stationary phase method with $T(\bar{p}) - T(p'') + (\bar{p}-p'')q$ as the rapidly varying phase. This phase is stationary at $\bar{p} = p''$ or $q = \partial T/\partial p''$, provided the determinant of the second derivatives of T with respect to $\bar{p}$ does not vanish at $\bar{p} = p''$. This requirement makes it also possible to use $\bar{p}$ as variable of integration rather than q. Therefore, we can also multiply (C13) with

$$V(+\partial T/\partial\bar{p})\cdot\det|\partial^2 T/\partial\bar{p}\partial\bar{p}|,$$

replace q in the exponent by $\partial T/\partial\bar{p}$, and integrate over $\bar{p}$.

The phase $T(\bar{p}) - T(p'') - (\bar{p}-p'')\cdot\partial T/\partial\bar{p}$ is now treated with respect to $\bar{p} - p''$ exactly as the phase $T(p) - T(\bar{p}) - (p-\bar{p})\cdot\partial T/\partial\bar{p}$ was treated with respect to $p - \bar{p}$. There are some minor, though obvious, modifications because the latter phase is not the same function of $p - \bar{p}$ as the former of $\bar{p} - p''$. All the slowly varying quantities in (C13) have to be expanded in powers of $\bar{p} - p''$, although this is not necessary for the terms which are already of order $\hbar$. Finally, we can express the coefficients β in terms of the derivatives of T with respect to $\bar{p}$, and use such relations as

$$\sum_{jl}\alpha_{jm}\alpha_{ln}\frac{\partial^2 T}{\partial\bar{p}^j\partial\bar{p}^l} = \epsilon_m\delta_{mn}, \tag{C14}$$

in order to express everything in terms of T and its derivatives. All the complicated terms in (C13) are cancelled out, and one is left with the relatively simple expansion (37).

Whereas the derivation of (C13) can be made sufficiently rigorous, provided we include a discussion of the boundary points, the further integration over

[23] D. S. Jones and M. J. Kline, J. Math. & Phys. 37, 1 (1958).

q or $\bar{p}$ using (C13) may be much harder to justify in view of its complicated structure.

APPENDIX D

In order to find the eigenvalues and the determinant of the matrix (58), we first add a term $-\lambda\tau/N$ in the diagonal elements. Let U_n be the determinant which results from (58) after all rows and columns beyond n have been eliminated. The following recursion formulas are then obtained:

$$U_n = \frac{\tau}{N}\left[\frac{M}{mr_{n+\frac{1}{2}}r_{n-\frac{1}{2}}} - \lambda\right]U_{n-\frac{1}{2}} - U_{n-1}, \tag{D1}$$

$$U_n = \frac{\tau}{N}\left[\frac{M}{2mr_{n+1}r_n} + \frac{M}{2mr_nr_{n-1}} - \lambda\right]U_{n-\frac{1}{2}} - U_{n-1}, \tag{D2}$$

for half-integer and for integer n, respectively. The initial values are

$$U_{\frac{1}{2}} = \frac{\tau}{N}\left[\frac{M}{mr_0r_1} - \lambda\right], \tag{D3}$$

$$U_1 = -1 + \frac{\tau^2}{N^2}\left[\frac{M}{mr_0r_1} - \lambda\right]\left[\frac{M}{2mr_0r_1} + \frac{M}{2mr_1r_2} - \lambda\right]. \tag{D4}$$

An alternating sign has to be eliminated before going to the limit of large N. Therefore, we define

$$\begin{aligned} \tilde{U}_n &= (-1)^{n-\frac{1}{2}}U_n \quad \text{for half-integer } n, \\ \tilde{U}_n &= (-1)^{n-1}U_n \quad \text{for integer } n. \end{aligned} \tag{D5}$$

In the limit of large N with τ remaining constant, the recursion formulas (D1) and (D2) become

$$dW/dt = -[(M/mr^2) - \lambda]U, \tag{D6}$$

$$dU/dt = [(M/mr^2) - \lambda]W, \tag{D7}$$

where $W = \tilde{U}_n$ for half-integer n, and $U = \tilde{U}_n$ for integer n. The initial conditions (D3) and (D4) reduce to

$$W(0) = 0, \quad U(0) = -1. \tag{D8}$$

The discrete parameter n has been replaced by the continuous parameter $t = n\tau/N$. The consecutive values of the radial distance r_n are assumed to lie very close to corresponding values $r(t)$ along the classical orbit.

Because of (42), the solution of the initial-value problem (D6), (D7), and (D8) can be written immediately in terms of the polar angle φ along the orbit,

$$W = \sin(\varphi - \lambda t), \quad U = -\cos(\varphi - \lambda t). \tag{D9}$$

If N is sufficiently large, we find, therefore, the following approximate value for the determinant of (58),

$$\det|(58)| = (-1)^{N-1}\sin[(\varphi'' - \varphi') - \lambda(t'' - t')]. \tag{D10}$$

The eigenvalues are, therefore, given by

$$\lambda = (\nu\pi - \varphi'' + \varphi')/(t'' - t'),$$

where ν is any integer, positive or negative, larger than $-N$ and smaller than N.

APPENDIX E

The matrix (58) for the variations δL_n and $\delta\theta_n$ was easy to discuss because its determinant could be evaluated explicitly for large N even after including a term $-\lambda\tau/N$ in the diagonal. In the case of the matrix (60) with the subsidiary condition (63), such a direct procedure can again be devised; but it is important not to specialize the particular form of the Hamiltonian at an early stage, because the general features might easily be lost in the arithmetic. Also, the treatment of the phase-integral approximation in momentum space is equivalent to the treatment in coordinate space, only if the kinetic energy is allowed to be a more general function of momentum than the usual $p^2/2m$. Such a generalization would automatically include a relativistic Hamiltonian. We shall assume, therefore, that the kinetic energy is an arbitrary function $D(\rho)$ of $\rho = |p|$ and the potential energy an arbitrary function $V(r)$ of $r = |q|$, so that $H(pq) = D(\rho) + V(r)$.

In terms of the momenta s_n, L_n, and M_n, as well as the coordinates r_n, θ_n, φ_n, we have in the limit $\hbar \to 0$,

$$\rho_n^2 = s_n^2 + \frac{L_n^2}{r_{n-\frac{1}{2}}r_{n+\frac{1}{2}}} + \frac{M_n^2}{r_{n-\frac{1}{2}}r_{n+\frac{1}{2}}\sin\theta_{n-\frac{1}{2}}\sin\theta_{n-\frac{1}{2}}}, \tag{E1}$$

$$\mathcal{E} = \frac{1}{N}\sum_{\frac{1}{2}}^{N-\frac{1}{2}} D(\rho_n) + \frac{1}{N}\sum_{0}^{N} V(r_n) \tag{E2}$$

instead of (53). The formulas (77) and (78) are obtained from (E1) and (E2) by the formal replacements $D \leftrightarrow V$, $\rho \leftrightarrow r$, $s \leftrightarrow \sigma$, $\theta \leftrightarrow \zeta$, $\varphi \leftrightarrow \eta$, provided we neglect the last term in (78). A discussion of the second variation of S_N as given by (55) with a constant $\mathcal{E}$ as given by (E2) includes, therefore, a discussion of the second variation of T_N as given by (76) without the quadratic terms with a constant $\mathcal{E}$ as given by (77).

The equations of motion (56) and (57) as well as the matrix (58) are hardly affected by the new kinetic energy, and can be treated exactly as before. The matrix of the second variations in $(\delta M, \delta s_{\frac{1}{2}}, \delta r_1, \delta s_{\frac{3}{2}}, \delta r_2, \cdots)$ is most easily written out in terms of the function

$$\mathcal{H} = \sum_{\frac{1}{2}}^{N-\frac{1}{2}} D[s_n^2 + M^2/r_{n-\frac{1}{2}}r_{n-\frac{1}{2}}]^{\frac{1}{2}} + \sum_{0}^{N} V(r_n). \quad \text{(E3)}$$

Let lower indices on $\mathcal{H}$ indicate the corresponding derivatives, half-integers for s_n and integers for r_n. Instead of (60), one now has the matrix

$$-\frac{1}{2}\begin{vmatrix} \Delta\mathcal{H}_{MM} & \Delta\mathcal{H}_{M\frac{1}{2}} & \Delta\mathcal{H}_{M1} & \Delta\mathcal{H}_{M\frac{3}{2}} & \Delta\mathcal{H}_{M2} & \cdots \\ \Delta\mathcal{H}_{M\frac{1}{2}} & \Delta\mathcal{H}_{\frac{1}{2}\frac{1}{2}} & -(1-\Delta\mathcal{H}_{\frac{1}{2}1}) & 0 & 0 & \cdots \\ \Delta\mathcal{H}_{M1} & -(1-\Delta\mathcal{H}_{1\frac{1}{2}}) & \Delta\mathcal{H}_{11} & (1+\Delta\mathcal{H}_{1\frac{3}{2}}) & \Delta\mathcal{H}_{12} & \cdots \\ \Delta\mathcal{H}_{M\frac{3}{2}} & 0 & (1+\Delta\mathcal{H}_{\frac{3}{2}1}) & \Delta\mathcal{H}_{\frac{3}{2}\frac{3}{2}} & -(1-\Delta\mathcal{H}_{\frac{3}{2}2}) & \cdots \\ \Delta\mathcal{H}_{M2} & 0 & \Delta\mathcal{H}_{21} & -(1-\Delta\mathcal{H}_{2\frac{3}{2}}) & \Delta\mathcal{H}_{22} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{vmatrix}, \quad \text{(E4)}$$

where $\Delta = \tau/N$. The subsidiary condition (63) becomes

$$\delta M \cdot \mathcal{H}_M + \sum_{\frac{1}{2}}^{N-\frac{1}{2}} \mathcal{H}_n \delta s_n + \sum_{1}^{N-1} \mathcal{H}_n \delta r_n = 0. \quad \text{(E5)}$$

A comparison with (60) shows that a number of new off-diagonal terms arise from the more general energy (E2).

The $2N \times 2N$ matrix (E4) is reduced to a $2N-1 \times 2N-1$ matrix in $(\delta s_{\frac{1}{2}}, \delta r_1, \delta s_{\frac{3}{2}}, \delta r_2, \cdots, \delta s_{N-\frac{1}{2}})$ by eliminating the variation δM with the help of (E5). The second variation which goes into the exponent of (64) has, therefore (apart from a factor $-\frac{1}{2}$), the matrix

$$c_{ij} + \Delta\mathcal{H}_{MM}[(a_i + b_i)(a_j + b_j) - b_i b_j], \quad \text{(E6)}$$

where c_{ij} is the matrix (E4) without the first row and the first column. The quantities a_j and b_j are given by the derivatives of $\mathcal{H}$,

$$a_j = \mathcal{H}_j/\mathcal{H}_M, \quad b_j = -\mathcal{H}_{Mj}/\mathcal{H}_{MM}. \quad \text{(E7)}$$

The signature of the matrix (E6), i.e., the difference between the number of positive and the number of negative eigenvalues, as well as its determinant, have to be found. Let Γ_n be the determinant which is obtained from (E6) after all rows and columns beyond the index n have been deleted. According to a theorem from linear algebra, cf., Bôcher,[24] the signature equals the sum over sign $(\Gamma_{n-\frac{1}{2}}\Gamma_n)$ from $n = \frac{1}{2}$ to $n = N - \frac{1}{2}$ with $\Gamma_0 = 1$.

The determinant Γ_n can be written in terms of the determinants C_{kl}, which are obtained from the matrix c_{ij} by deleting all rows and columns before k and beyond l. In terms of $\tilde{C}_{kl} = (-1)^{k(2k-1)+l(2l-1)}C_{kl}$, one finds after some straightforward algebra that

$$(-1)^{n(2n-1)}\Gamma_n = \tilde{C}_{\frac{1}{2}n} - 2\Delta\mathcal{H}_{MM} \times \sum_{i<j<n} (a_i a_j + a_i b_j + b_i a_j)\tilde{C}_{\frac{1}{2}i-\frac{1}{2}}\tilde{C}_{j+\frac{1}{2}n} + 2(\Delta\mathcal{H}_{MM})^2 \sum_{i<j<k<l<n} \tilde{C}_{\frac{1}{2}i-\frac{1}{2}}\tilde{C}_{j+\frac{1}{2}k-\frac{1}{2}}\tilde{C}_{l+\frac{1}{2}n}\{\ \}, \quad \text{(E8)}$$

$$\{\ \} = a_i b_j b_k a_l + b_i a_j b_k a_l + b_i a_j a_k b_l + a_i b_j a_k b_l - 2a_i a_j b_k b_l - 2b_i b_j a_k a_l.$$

Certain terms of equal indices have been neglected because their contribution is only of order $1/N$ or smaller.

Since the elements of (c_{ij}) differ from zero only if they are close to the diagonal, one can easily derive recursion formulas for $\tilde{C}_{kl}$. In the limit of large N, one obtains ordinary linear differential equations in the parameter $t = n\tau/N = n\Delta$, provided the sequence of momenta s_n and distances r_n approximates the classical trajectory $s(t)$, $r(t)$ in phase space. With the Hamiltonian

$$H(sr) = D(s^2 + M^2/r^2)^{\frac{1}{2}} + V(r), \quad \text{(E9)}$$

this trajectory satisfies the equations of motion

$$ds/dt = -\partial H/\partial r, \quad dr/dt = \partial H/\partial s. \quad \text{(E10)}$$

The "initial" values are $r(0) = r'$, $r(\tau) = r''$, and the angular momentum M is determined such that

$$\int_0^\tau dt \frac{\partial H}{\partial M} = \int_0^\tau dt \frac{M D_\rho}{\rho r^2} = \varphi'' - \varphi', \quad \text{(E11)}$$

with $\rho = (s^2 + M^2/r^2)^{\frac{1}{2}}$; e.g., let $\tilde{C}_{kl} = W$ if l is half-integer, and $\tilde{C}_{kl} = U$ if l is integer. In terms of $t'' =$

[24] M. Bôcher, *Introduction to Higher Algebra* (The Macmillan Company, New York, 1907), p. 147.

$l\tau/N$, the recursion formulas with respect to l become

$$\frac{dW}{dt''} = \frac{\partial^2 H}{\partial s^2} U + \frac{\partial^2 H}{\partial s \partial r} W,$$
$$\frac{dU}{dt''} = -\frac{\partial^2 H}{\partial r^2} W - \frac{\partial^2 H}{\partial s \partial r} U. \tag{E12}$$

These are the Jacobi equations for the so-called assessory problem; cf. Caratheodory.[25]

Solutions to (E12) can be constructed if a family of solutions $s(t, \mu)$ and $r(t, \mu)$ to (E10) is known which depends on some parameter μ. One finds immediately that

$$U = \partial s/\partial \mu|_{t=t''}, \quad W = \partial r/\partial \mu|_{t=t''}. \tag{E13}$$

The initial conditions for U and W depend on k. One finds for half-integer k that $U(t') = 1$ and $W(t') = 0$, whereas $U(t') = 0$ and $W(t') = -1$ for integer k, with $t' = k\tau/N$. Since the angular momentum M is considered a constant in (E10) and (E12), the only parameter left to yield a family of solutions is the energy E.

The function U and W can be written as integrals over the classical trajectory in the following manner: in the case of half-integer k, the second Eq. (E10) is integrated by writing

$$t'' - t' = \int_{r'}^{r''} \frac{dr}{H_s}, \tag{E14}$$

where s in H_s is assumed to be eliminated with the help of $H(sr) = E$. The above equation is then differentiated with respect to E at constant t'', so that one obtains the relation

$$0 = \frac{\partial r''}{\partial E}\bigg|_{t''} \cdot \frac{1}{H_s''} - \int_{r'}^{r''} \frac{dr}{H_s^2} H_{ss} \frac{\partial s}{\partial E}\bigg|_r, \tag{E15}$$

where $\partial r'/\partial E$ was assumed to vanish in accordance with the initial condition $W(t') = 0$. The derivative $\partial s/\partial E$ follows from $H(sr) = E$, so that

$$\frac{\partial r''}{\partial E}\bigg|_{t''} = H_s'' \int_{r'}^{r''} dr \frac{H_{ss}}{H_s^3} = -H_s'' \int_{t'}^{t''} dt \left(\frac{1}{H_s}\right)_s. \tag{E16}$$

The lower indices always designate partial derivatives, whereas the prime or double prime indicate the time at which the quantity is to be evaluated. If the expression (E16) is inserted for W into the first Eq. (E12), the corresponding function U is obtained. After adjusting the result to the initial condition $U(t') = 1$, one finds that

$$U = \frac{H_s'}{H_s''} + H_s' H_r'' \int_{t'}^{t''} \left(\frac{1}{H_s}\right)_s,$$
$$W = -H_s' H_s'' \int_{t'}^{t''} dt \left(\frac{1}{H_s}\right)_s \tag{E17}$$

for half-integer k. Similarly, it follows that

$$W = -\frac{H_r'}{H_r''} + H_r' H_s'' \int_{t'}^{t''} dt \left(\frac{1}{H_r}\right)_r,$$
$$U = -H_r' H_r'' \int_{t'}^{t''} dt \left(\frac{1}{H_r}\right)_r \tag{E18}$$

for integer k.

The integrals in (E17) diverge when t passes a classical turning point where $dr/dt = H_s = 0$. But a close examination of U and W as t'' approaches such a time, shows that these functions approach well-defined, finite values and can be continued in a natural fashion without discontinuities. The same is true for (E18). With the help of identity

$$\int_{t'}^{t''} dt \left(\frac{1}{H_r}\right)_r = \int_{t'}^{t''} dt \left(\frac{1}{H_s}\right)_s + \frac{1}{H_s'' H_r''} - \frac{1}{H_s' H_r'}, \tag{E19}$$

the integrals (E18) can be written like the integrals (E17), and vice versa. Also, this identity shows how to avoid a divergent integral in the neighborhood of a turning point.

The calculation of Γ_n from (E8) presents no difficulties in the limit of large N. The coefficients a_j and b_j are written for half integer j as

$$a_j \cong H_s\, dt \Big/ \int_0^N H_M\, dt,$$
$$b_j \cong -H_{Ms}\, dt \Big/ \int_0^N H_{MM}\, dt, \tag{E20}$$

and for integer j as

$$a_j \cong H_r\, dt \Big/ \int_0^N H_M\, dt,$$
$$b_j \cong -H_{Mr}\, dt \Big/ \int_0^N H_{MM}\, dt. \tag{E21}$$

It seems advisable to obtain first the sums of the type $\sum (i<j) \bar{C}_{\frac{1}{2}i-\frac{1}{2}} a_i$, $\sum (i<j) a_j \bar{C}_{j+\frac{1}{2}n}$, etc., with the help of (E17) and (E18), then the sums of the type $\sum (i<j<k<l) a_j \bar{C}_{j+\frac{1}{2}k-\frac{1}{2}} a_k$, etc., and finally the complete sums as they occur in (E8). The various successive integrations can always be combined and simplified, although the procedure is very tedious and one suspects that there must be some shortcuts to avoid these lengthy computations. The result is

$$(-1)^{N-1}\Gamma_N = H_s^0 H_s^N \left\{ \int_0^N H_{MM}\, dt - \int_0^N \left(\frac{H_M^2}{H_s}\right)_s dt \right\} \times \left[(t_N - t_0) \Big/ \int_0^N H_M\, dt \right]^2. \tag{E22}$$

The integrals in braces can be shown to equal $\partial(\varphi_N - \varphi_0)/\partial M$ at constant r_0 and r_N, whereas the ratio $(t_N - t_0)/\int_0^N H_M\, dt$ is equal to $\partial M/\partial E$, i.e., the change in angular momentum which is necessary to accommodate a change in average energy while keeping the same orbit $s(t)$ and $r(t)$. The integral (64) is combined with (D10) and (E22) to give (20) with (46).

[25] C. Carathéodory, *Variationsrechnung und partielle Differentialgleichungen erster Ordnung* (B. G. Teubner, Leipzig, 1935), p. 260.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Compatibility Requirements in BBGKY Expansion

RICHARD L. LIBOFF AND GIOVANNI E. PERONA
School of Electrical Engineering, Cornell University, Ithaca, New York

(Received 29 March 1967)

The BBGKY equations are written in nondimensionalized form leaving two parameters α and β explicit. The coefficient α is a measure of the "strength" of the potential of interaction, while β is a measure of the "range" of the interaction potential. The s-particle distribution function is expanded about vanishing correlation with ϵ entering as a parameter of smallness. Nine cases are considered, depending on the order of magnitude of α and β (*viz.*, ϵ, 1, ϵ^{-1}). Lowest order terms, in each of the cases considered, give a kinetic equation, together with subsidiary conditions. In the "rigid sphere" limit, the Vlasov equation appears together with a second-order differential constraint. The Vlasov equation alone appears in the Rosenbluth–Rostoker limit. A critical survey is made of the compatibility problem which is inherent to expansions of the BBGKY sequence. For the most part, the problem reduces to the compatibility of "type-l" constraint equations. These constraint conditions cause any given order set of equations to be over determined.

I. INTRODUCTION

IN the recent past, a host of so-called "kinetic equations" have appeared in the literature. A kinetic equation is a single equation in space, time, and velocity, for the single unknown—the one-particle distribution function, F_1. To date, the set of kinetic equations include: The Boltzmann equation,[1] the Vlasov equation,[2] the Landau equation,[3] the Fokker–Planck equation,[4] and the Balescu–Lenard equation.[5]

Ideally, a kinetic equation should follow in a self contained manner from the Liouville equation, which is the equation of motion for the N-particle distribution function F_N. Equivalently, one may work with the BBGKY[6] sequence of N-coupled equations for the set of N s-particle distributions functions $\{F_s\}$. The first of these equations (hereafter called BY_1) is an equation which couples F_1 to F_2. The second couples F_2 to F_3, and so forth. If a relation of the form $F_2 = F_2(F_1)$ is known, then substitution into BY_1 gives a kinetic equation. Bogoliubov[7] argues that in the time sequence of events governing the relaxation of a gas, the mid-epoch (the "kinetic stage") is characterized by the functional relations

$$F_s = F_s(1, \cdots, s; F_1).$$

All of the s-particle distributions become functionals of the one-particle distribution. The notation is such that the number l denotes the coordinates and momenta of the lth particle.

The method of obtaining kinetic equations due to Prigogine[8] is also centered about the Liouville equation and employs the technique of the resolvant operator familiar to quantum mechanics. A diagrammatic approach is employed to represent the expansion of distribution functions. Ansatzen regarding the form of the exponential relaxation of the distributions permits one to collect different "order" diagrams, and kinetic equations emerge depending on the relaxation scale imposed.

The method of analysis of the BBGKY sequence, which we wish to examine in the paper, strongly resembles that Mayer–Mayer[9] cluster expansion familiar to equilibrium statistical mechanics. It has been employed by Rosenbluth and Rostoker[10] to extract the Vlasov equation from the BBGKY set and by Freeman[11] to formulate a theory of irreversible processes. The technique consists of expanding F_s about vanishing correlations. In general, if for all $s \leq N$,

$$F_s = \prod_{l=1}^{s} F_1(l),$$

the gas is completely correlationless. The deviation of F_s from the product form defines the correlation functions $\mathcal{C}_s$. For instance, F_2 appears as,

$$F_2(1, 2) = F_1(1)F_1(2) + \mathcal{C}_2(1, 2).$$

More generally, the first term in the expansion for F_s

[1] L. Boltzmann, *Lectures on Gas Theory*, translated by S. Brush University of California, Los Angeles, Calif., 1964).
[2] A. A. Vlasov, *Many Particle Theory and Its Application to Application to Plasma* (Gordon and Breach, Science Publishers Inc., New York, 1961).
[3] L. Landau, J. Phys. USSR **10**, 25 (1946).
[4] (a) A. Fokker, Ann. Physik **43**, 810 (1914).
(b) M. Planck, Sitz. Preuss. Acad. 324 (1917).
(c) S. Chandrasekhar, Rev. Mod. Phys. **15**, 31 (1945).
(d) J. Hubbard, Proc. Roy. Soc. (London) **261A**, 371 (1961).
[5] (a) R. Balescu, Phys. Fluids **3**, 52 (1960).
(b) A. Lenard, Ann. Phys. (N.Y.) **3**, 390 (1960).
[6] M. Born and N. S. Green, Proc. Roy. Soc. (London) **188A**, 10 (1946); **189A**, 103 (1947); **191A**, 168 (1947); J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946); **15**, 72 (1947); J. Yvon, *La Theorie Statistique dès Fuides et l'Equation* (d'Etat, Herman, Paris, 1935).
[7] N. N. Bogoliubov, Studies Stat. Mech. (Amsterdam) **1**, 1 (1962).
[8] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1962).
[9] J. E. Mayer and M. G. Mayer *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1940).
[10] N. Rostoker and M. N. Rosenbluth, Phys. Fluids **3**, 1 (1960).
[11] E. A. Freeman, J. Math. Phys. **4**, 410 (1963).

is the product form given above. These expansions merely serve as a transformation from the set of functions $\{F_s\}$ to the set $\{\mathcal{C}_s\}$. However, in the method of vanishing correlations, one sets $\mathcal{C}_s \sim \epsilon^s$ where ϵ is an infinitesimal. The lowest-order expression for F_s then becomes the product form $\prod^s F_1(l)$. Substitution of these lowest-order forms into the BBGKY sequence gives N equations for the single unknown F_1. For these equations to be meaningful, they must be compatible, e.g., redundant. This compatibility requirement is merely a property which the equations must have in order that the solution is in the trial product form given above.

The included analysis stems from a nondimensional form of the BBGKY sequence due to Sandri.[12] In this form, two parameters (α, β) emerge which contain the strength and range of the interparticle potential, and the temperature and number density of the system. Nine distinct cases are considered, depending on the magnitude of these parameters (i.e., ϵ, 1, ϵ^{-1}). Tables III–XI give the kinetic equations and constraint conditions which emerge. These tables are discussed in the following analysis section. The physical relevance of these results is discussed in the final section of this paper.

II. ANALYSIS

A. Preliminary Definitions

Our investigation stems from the Liouville equation,

$$\frac{\partial F_N}{\partial t} + [F_N, H] = 0, \tag{1}$$

where the brackets are the Poisson brackets, and H is the Hamiltonian,

$$H = \sum_{i=1}^{N} p_i^2/2m + \sum_{i<j}^{N} \Phi(\mathbf{x}_i, \mathbf{x}_j). \tag{2}$$

In this formula, $\mathbf{x}_i$ is the position of the ith particle, and Φ is the potential of interaction between two particles. Hereafter, we will set $\Phi(\mathbf{x}_i, \mathbf{x}_j) = \Phi_{ij}$. The N-particle joint probability function F_N is normalized to V^N, where V is the volume of the system,

$$\int F_N \, d1 \cdots dN = V^N. \tag{3}$$

Operating on Eq. (1) with $\int d1 \cdots ds$, and performing some detailed, but straightforward manipulations, gives the desired form,

$$\frac{\partial F_s}{\partial t} + \hat{K}_s F_s - \hat{\mathcal{H}}_s F_s = \left(\frac{N-s}{V}\right) \hat{I}_s F_{s+1}. \tag{4}$$

[12] G. Sandri, Ann. Phys. (N.Y.) **24**, 332 and 380 (1963).

The operators $\hat{K}$, $\hat{\mathcal{H}}$, and $\hat{I}$ are defined as follows:

$$\hat{K}_s \equiv \sum_{l=1}^{s} (\mathbf{p}_l/m) \cdot (\partial/\partial \mathbf{x}_l), \tag{5a}$$

$$\hat{\mathcal{H}}_s \equiv \sum_{i<j}^{s} \left(\frac{\partial \Phi_{ij}}{\partial \mathbf{x}_i} \cdot \frac{\partial}{\partial \mathbf{p}_i} + \frac{\partial \Phi_{ij}}{\partial \mathbf{x}_j} \cdot \frac{\partial}{\partial \mathbf{p}_j} \right), \tag{5b}$$

$$\hat{I}_s \equiv \sum_{i=1}^{s} \frac{\partial}{\partial \mathbf{p}_i} \cdot \int d(s+1) \mathbf{G}_{i,s+1}. \tag{5c}$$

The force of the ith particle on the $(s+1)$st particle is

$$\mathbf{G}_{i,s+1} = -\frac{\partial \Phi_{i,s+1}}{\partial \mathbf{x}_{s+1}} = +\frac{\partial \Phi_{i,s+1}}{\partial \mathbf{x}_i}.$$

For convenience of future reference, we shall call the sth BBGKY equation, BY_s.

The transformation from the set of distribution functions $\{F_s\}$ to the set of correlation functions $\{\mathcal{C}_s\}$ appears as

$$F_1(1) = F_1(1),$$

$$F_2(1, 2) = F_1(1)F_1(2) + \mathcal{C}_2(1, 2),$$

$$F_3(1, 2, 3) = F_1(1)F_1(2)F_1(3) + F_1(1)\mathcal{C}_2(2, 3) + F_1(2)\mathcal{C}_2(3, 1) + F_1(3)\mathcal{C}_2(1, 2) + \mathcal{C}_3(1, 2, 3);$$

$$\vdots$$

$$F_s = \prod_{k=1}^{s} F_1(k) + \sum_{P_2(l_1, \cdots, l_s)} F_1(l_1) \cdots F_1(l_{s-2})\mathcal{C}_2(l_{s-1}, l_s) + \sum_{P_3(l_1, \cdots, l_s)} F_1(l_1) \cdots F_1(l_{s-3})\mathcal{C}_3(l_{s-2}, l_{s-1}, l_s) + \cdots + \mathcal{C}_s. \tag{6}$$

In this latter formula, P_n is the set of permutations on the set of integers $(l_1, \cdots, l_s)$ counting all permutations amongst $(l_1, \cdots, l_{s-n})$ only once, similarly for all permutations among $(l_{s-n+1}, \cdots, l_s)$. There are $s!/n!(s-n)!$ distinct permutations in the sequence P_n. In passing, we note that an alternate way of writing the last equation appears as

$$F_s = \prod_{k=1}^{s} F_1(k) + \sum_{k_1<k_2}^{s} F_1(1) \cdots F_1(k_1 - 1) \times F(k_1 + 1) \cdots F_1(k_2 - 1)F_1(k_2 + 1) \cdots F(s)\mathcal{C}_2(k_1, k_2) + \cdots + \sum_{k_1<k_2<\cdots<k_{b-1}<k_b}^{s} F_1(1) \cdots F_1(k_1 - 1) F_1(k_1 + 1) \cdots F_1(k_2 - 1)F_1(k_2 + 1) \cdots F_1(k_{b-1} - 1) \times F_1(k_{b-1} + 1) \cdots F_1(k_b - 1)F_1(k_b + 1) \cdots F_1(s)\mathcal{C}_b(k_1, k_2, \cdots k_b) + \cdots + \mathcal{C}_s(1, 2, \cdots, s). \tag{6a}$$

The corresponding diagrammatic sum appears as

$$F_s = [0]^s + [0]^{s-2}[0-0] + [0]^{s-3}[0-0-0] + \overset{s\text{-cluster}}{\cdots} + [0-0-0-\cdots-0].$$

The nth cluster is a sum of $s!/n!(s-n)!$ terms. The correlation functions $\mathcal{C}_s$ are merely another representation for the distribution functions F_s. What is hoped is that this new choice of variables is more conducive to the investigation of systems which enjoy some element of "correlationlessness."

The first insertion of any physics into the scheme comes by way of the hypothesis that under proper conditions $\mathcal{C}_s \simeq \epsilon^{s-1}$, where ϵ is a parameter of smallness. The motivation is as follows. Consider the limit of infinite tenuity (of a gas). Certainly (for finite range potential of interaction), under such conditions all $\mathcal{C}_s$ vanish, and $F_s = \Pi F_1$. Now, let the volume decrease, so that the probability of two particles being in each other's sphere of influence is finite, while the probability that three particles simultaneously being so joined remains zero. It is consistent then to write F_3 in some manner as the product of F_1 and F_2, with $F_2 \neq F_1F_1$.

For any three particles, the probabilities relating to the state of a cluster of any one—and that of the remaining two, are independent probabilities. This property is satisfied, if in the second equation in the sequence 6 for F_3, $\mathcal{C}_3$ is set equal to zero. The inference is that as the specific density is increased, the higher-order correlation functions become significant.

The question arises as to the proper choice of ϵ in the expression $\mathcal{C}_s \simeq \epsilon^{s-1}$. To answer this question, we introduce a nondimensionalizing procedure, first proposed by Sandri.[12]

The physical variables which enter into BY_s (Eq. 4) are t, $\mathbf{x}$, $\mathbf{p}$, and Φ. These are nondimensionalized through the following equations. A variable which carries a tilde is a nondimensional variable,

$$\mathbf{x} = r_0\tilde{\mathbf{x}};\ \mathbf{p} = p_0\tilde{\mathbf{p}};\ \Phi = \Phi_0\tilde{\Phi};\ t = t_0\tilde{t},$$
$$p_0 = mC;\ mC^2 = KT;\ t_0 = r_0/C. \tag{7}$$

The "strength" of the potential is Φ_0, while r_0 is the "range" of the potential. In most cases, the choice of Φ_0 and r_0 is clearly implied by the form of Φ itself.

Inserting these changes into Eq. (4) gives [dropping the tilde notation and setting $(N-s)/V \cong N/V = n$]

$$\left(\frac{\partial}{\partial t} + \hat{K}_s\right)F_s - \alpha\,\hat{\mathfrak{H}}_s F_s = \alpha\beta\hat{I}_s F_{s+1}. \tag{8}$$

In this equation,

$$\alpha \equiv \Phi_0/mC^2,\quad \beta \equiv nr_0^3. \tag{9}$$

TABLE I. The $\alpha - \beta$ scheme diagram

α \ β	ϵ	1	ϵ^{-1}
	CASE I	CASE II	CASE III
ϵ^{-1}	ϵ^{-1}, ϵ	$\epsilon^{-1}, 1$	$\epsilon^{-1}, \epsilon^{-1}$
	CASE IV	CASE V	CASE VI
1	$1, \epsilon$	$1, 1$	$1, \epsilon^{-1}$
	CASE VII	CASE VIII	CASE IX
ϵ	ϵ, ϵ	$\epsilon, 1$	ϵ, ϵ^{-1}

All variables and functions, and parameters in Eq. (8) are nondimensional. Inasmuch as two parameters arise in the process of obtaining this equation, the choices of expansion about $\mathcal{C}_s = 0$, may be put in terms of a two-dimensional array as depicted in Table I.

This scheme of values for α and β gives rise to nine independent domains of expansion. In each case, the expansion is completely specified by the additional constraint

$$\mathcal{C}_s \propto \epsilon^{s-1}.$$

Having stated these prerequisite definitions, we are now prepared to move on to the analysis of these equations.

B. Expansion Technique and Compatibility Requirements

In order to find effective expressions for F_1, $\mathcal{C}_2$, $\mathcal{C}_3, \cdots$, the expansions of the distribution functions $F_1, F_2, \cdots, F_s$, as given by Eq. (6), are substituted in the BY_s equations, whose general forms are given by Eq. (8). These equations, together with the order-of-magnitude ansätz (9), generate a sequence of coupled equations. The equations so obtained depend very strongly on which of the nine limits (shown in Table I) is being considered. In each case, the coefficients (α, β) assume prescribed orders of magnitude. On the other hand, F_l as given by Eq. (6a), exhibits the structure

$$F_l = \prod^{l} F_1 + \epsilon \sum \left(\prod^{l-1} F_1\right)\tilde{\mathcal{C}}_2 + \cdots. \tag{10}$$

Substituting all of the F_l functions into the appropriate BY_s equations $(s \leq l)$ and expanding, gives the following set of equations:

$$BY_1^0 + \epsilon BY_1^{(1)} + \epsilon^2 BY_1^{(2)} + \cdots = 0,$$
$$BY_2^{(0)} + \epsilon BY_2^{(1)} + \epsilon^2 BY_2^{(2)} + \cdots = 0. \tag{11}$$
$$\vdots$$

TABLE II. Order of magnitude of type-l constraint equations

FUNCTIONS \ CASES	I	II	III	IV	V	VI	VII	VIII	IX
F_1	0–1	0–1	0–1–2	0–1	0	0–1	0–1–2	0–1	0–1
$\mathcal{C}_2$	1–2	1–2	1–2–3	1–2	1	1–2	1–2–3	1–2	1–2
$\mathcal{C}_3$	2–3	2–3	2–3–4	2–3	2	2–3	2–3–4	2–3	2–3

Equating coefficients of equal powers of ϵ to zero, gives

$$O(1) \quad BY_1^{(0)} = 0, \quad BY_2^{(0)} = 0, \cdots,$$
$$O(\epsilon) \quad BY_2^{(1)} = 0, \quad BY_2^{(1)} = 0, \cdots. \tag{12}$$

The $O(1)$ equations involve only F_1, and may be used to determine F_1. This solution, when substituted into the $O(\epsilon)$ equations, determines $\mathcal{C}_2$, since $O(\epsilon)$ equations involve only F_1 and $\mathcal{C}_2$. The $O(\epsilon^2)$ equations involve only F_1, $\mathcal{C}_2$, and $\mathcal{C}_3$, and may be used to calculate $\mathcal{C}_3$, once F_1 and $\mathcal{C}_2$ are known.

Although such an iterative technique clearly points the way to the construction of a solution, it is at this point that the problem of redundancy enters.

The redundancy problem is two fold. For example, there are N equations of the form $[BY_s^{(0)} = 0]$. All of these involve F_1 alone. Clearly, these equations must be compatible (not redundant, since we are dealing with partial differential equations and not algebraic equations).

The second type of redundancy condition which enters, is as follows. Suppose all $\{BY_s^{(0)} = 0\}$ are compatible. Now suppose $\{BY_s^{(1)} = 0\}$ contains equations which involve F_1 alone. Such equations are not of the form which contribute to the solution for $\mathcal{C}_2$. Instead, they are extra equations which F_1 must satisfy beyond the set $\{BY_s^{(0)} = 0\}$. Similarly, $\{BY_s^{(2)} = 0\}$ may contain equations which involve only F_1 and $\mathcal{C}_2$.

The question then arises, how many equations are there which involve F_1 alone, beyond the set $\{BY_s^{(0)} = 0\}$? How many equations are there which involve only F_1 and $\mathcal{C}_2$, beyond the set $\{BY_s^{(1)} = 0\}$?

Let us call equations which involve only (F_1, $\mathcal{C}_2, \cdots, \mathcal{C}_l$) with $\mathcal{C}_l$ explicit, "type-l constraint equations." In Table II, the order (k in $\{BY_s^{(k)} = 0\}$) of the constraint equations which enter for all of the cases considered, are presented. For instance, in Case II, the only type-2 constraint equations which enter are of the form $\{BY_s^{(1)} = 0\}$ and $\{BY_s^{(2)} = 0\}$ These type-l constraint equations are presented in the final tables of results, for each of the nine cases considered.

The schematic of solution is given in Fig. 1. The type-1 constraint equation (Box I) determines F_1 which in turn gives F_s to terms of $O(\epsilon)$:

$$F_s = \prod^{s} F_1 + 0(\epsilon). \tag{13}$$

The resulting F_1 is then used to calculate $\mathcal{C}_2$. This is accomplished by inserting F_1 into Box II. At each step of the procedure, F_s is determined to successively higher orders in ϵ. If an incompatibility exists, most likely this implies that solutions in the form of Eq. (6) do not exist.

From Fig. 1, it is clear that in the cases considered, there are, at most, three type-l classes of constraint equations (e.g., the box containing $BY_s^{(1)}$ in the first level represents a class of type-1 constraint equations). More generally, at any level in this schematic, there are at most three type-l classes of constraint equations. This follows from the fact that the orders of magnitude which are chosen for the coefficients α and β only range over three values.

In the following analysis the nine cases of Table X are separately studied.

C. Preliminary Ordering of Forms

Before constructing the systems of equations of different orders, it is useful at this point to rewrite the operators appearing in Eq. (8) in a more suitable form,

$$\hat{K}_m = \sum_{l=l_1,\cdots,l_m} \left(\frac{\mathbf{p}_l}{m}\right) \cdot \left(\frac{\partial}{\partial \mathbf{x}_l}\right). \tag{14}$$

In this expression, $\hat{K}_m$ operates on the phase variables $(l_1, \cdots, l_m)$.

With this generalization, the $O(\epsilon)$ term of the first term on the left-hand side of Eq. (8) appears as

$$\left(\frac{\partial}{\partial t} + \hat{K}_s\right) \sum_{P_2(l_1\cdots l_s)} F_1(l_1)\cdots F_1(l_{s-2})\mathcal{C}_2(l_{s-1}, l_s)$$
$$= \sum_{P_2(l_1\cdots l_s)} F_1(l_1)\cdots F_1(l_{s-2})\left[\frac{\partial}{\partial t} + \hat{K}_2\right]\mathcal{C}_2(l_{s-1}, l_s)$$
$$+ \sum_{P_3(l_1\cdots l_s)} F_1(l_1)\cdots F_1(l_{s-3})$$
$$\times \left\{\sum_{P_2(i,j,k)}^{s-2,s-1,s} \mathcal{C}_2(l_j, l_k)\left[\frac{\partial}{\partial t} + \hat{K}_1\right]F_1(l_i)\right\}, \tag{15}$$

where each of the integers i, j, k, ranges over the 3 values $s-2$, $s-1$, and s. In the same way, the

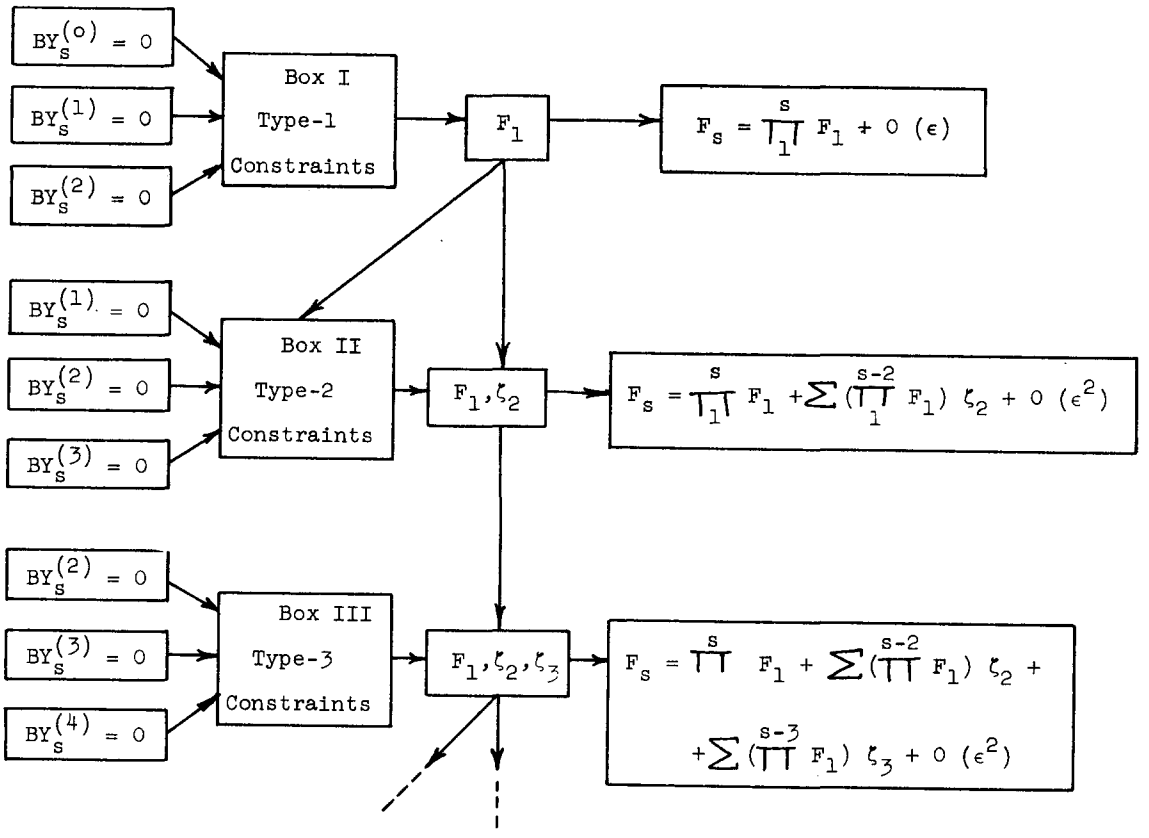


FIG. 1. Summary of type-*l* constraint equations.

$O(\epsilon^2)$ term is

$$\left(\frac{\partial}{\partial t} + \hat{K}_s\right) \sum_{P_3(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-3}) \mathcal{C}_3(l_{s-2}, l_{s-1}, l_s)$$
$$= \sum_{P_3(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-3}) \left[\frac{\partial}{\partial t} + \hat{K}_3\right] \mathcal{C}_3(l_{s-2}, l_{s-1}, l_s)$$
$$+ \sum_{P_4(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-4})$$
$$\times \left\{ \sum_{P_3(i,j,k,y)}^{s-3, \cdots, s} \mathcal{C}_3(l_j, l_k, l_y) \left[\frac{\partial}{\partial t} + \hat{K}_1\right] F_1(l_i) \right\}. \quad (16)$$

In the same way as for the operator $\hat{K}_m$ of Eq. (14), we can define an operator $\hat{\mathcal{H}}_m$, acting on a function of m phase variables, $l_1 \cdots l_m$, in the following way:

$$\hat{\mathcal{H}}_m = \sum_{\substack{b_1 < b_2 \\ b_1, b_2 = l_1, \cdots, l_m}} \left(\frac{\partial \Phi_{b_1 b_2}}{\partial \mathbf{x}_{b_1}} \cdot \frac{\partial}{\partial \mathbf{p}_{b_1}} + \frac{\partial \Phi_{b_1 b_2}}{\partial \mathbf{x}_{b_2}} \cdot \frac{\partial}{\partial \mathbf{p}_{b_1}} \right). \quad (17)$$

If the operator $\hat{\mathcal{H}}_n$ is intended to act on only n of the m variables of the operand, the n variables are explicitly indicated in parentheses at the right of $\hat{\mathcal{H}}_n$ [cf. Eq. (19)]. It is now possible to rewrite $\hat{\mathcal{H}}_s$ [Eq. (5b)] in a different form. Again, as before, the $O(\epsilon)$ term appears as

$$\hat{\mathcal{H}}_s \sum_{P_2(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-2}) \mathcal{C}_2(l_{s-1}, l_s)$$
$$= \sum_{P_2(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-2}) \hat{\mathcal{H}}_2 \mathcal{C}_2(l_{s-1}, l_s)$$
$$+ \sum_{P_4(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-4})$$
$$\times \left\{ \sum_{P_2(i,j,k,y)}^{s-3, \cdots, s} \mathcal{C}_2(l_k, l_y) \hat{\mathcal{H}}_2 F_1(l_i) F_1(l_j) \right\}$$
$$+ \sum_{P_3(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-3})$$
$$\times \left\{ \sum_{P_2(a_1, a_2, a_3)}^{s-2, s-1, s} \left\{ \sum_{j=2}^{3} \hat{\mathcal{H}}_2(l_{a_1}, l_{a_j}) F(l_{a_1}) \mathcal{C}_2(l_{a_2}, l_{a_3}) \right\} \right\}. \quad (18)$$

The $O(\epsilon^2)$ term is

$$\hat{\mathcal{H}}_s \sum_{P_3(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-3}) \mathcal{C}_3(l_{s-2}, l_{s-1}, l_s)$$
$$= \sum_{P_3(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-4}) \hat{\mathcal{H}}_3 \mathcal{C}_3(l_{s-2}, l_{s-1}, l_s)$$
$$+ \sum_{P_5(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-5})$$
$$\times \left\{ \sum_{P_3(i,j,k,y,z)}^{s-4, \cdots, s} \mathcal{C}_3(l_k, l_y, l_z) \hat{\mathcal{H}}_2 F_1(l_i) F_1(l_j) \right\}$$
$$+ \sum_{P_4(l_1 \cdots l_5)} F_1(l_1) \cdots F_1(l_{s-4})$$
$$\times \left\{ \sum_{P_3(a_1 \cdots a_4)}^{s-3, \cdots, s} \left\{ \sum_{j=2}^{4} \hat{\mathcal{H}}_2(l_{a_1}, l_{a_j}) F_1(l_{a_1}) \mathcal{C}_3(l_{a_2}, l_{a_3}, l_{a_4}) \right\} \right\}. \quad (19)$$

Next consider the operator $\hat{I}_s$ (Eq. 5c). In order to develop its generalization, we define,

$$\hat{I}_1(l_K) = \frac{\partial}{\partial \mathbf{p}_{l_K}} \cdot \int d(s+1)\mathbf{G}_{l_K,s+1} \tag{20}$$

so that,

$$\hat{I}_s = \sum_{i=1}^{s} \hat{I}_1(i). \tag{20a}$$

The remaining term on the right-hand side of Eq. (8) then appears as, i.e., the $O(\epsilon)$ term,

$$\begin{aligned}
&\hat{I}_s \sum_{P_2(l_1 \cdots l_{s+1})} F_1(l_1) \cdots F_1(l_{s-1}) \mathrm{C}_2(l_s, l_{s+1}) \\
&= \sum_{P_3(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-3}) \\
&\quad \times \left\{ \sum_{P_2(i,j,k)}^{s-2,s-1,s} \mathrm{C}_2(l_j, l_k) \hat{I}_1 F_1(s+1) F(l_i) \right\} \\
&\quad + \sum_{P_2(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-2}) \\
&\quad \times \left\{ \sum_{i=s-1}^{s} \hat{I}_1(l_i) F_1(s+1) \mathrm{C}_2(l_{s-1}, l_s) \right\} \\
&\quad + \sum_{P_2(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-2}) \\
&\quad \times \left\{ \sum_{P_1(i,j)}^{s-1,s} \hat{I}_1(l_i) F_1(l_i) \mathrm{C}_2(l_j, s+1) \right\} \\
&\quad + \sum_{P_1(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-2}) \hat{I}_1 \mathrm{C}_2(l_s, s+1).
\end{aligned} \tag{21}$$

With these generalizations of the operators, $\hat{K}$, $\hat{\mathrm{H}}$, and $\hat{I}$ at our disposal, we move on to the nine cases depicted in Table X.

Toward these ends, we consider the special Case VII ($\alpha = \epsilon$), ($\beta = \epsilon$).

In this case, there is no term of order ϵ^{-1}. The terms of order 1, from BY_1, appear as

$$\left(\frac{\partial}{\partial t} + \hat{K}_1\right) F_1(1) = 0. \tag{22}$$

Those from BY_s appear as

$$\left(\frac{\partial}{\partial t} + \hat{K}_s\right) F_1(1) \cdots F_1(s) = 0. \tag{23}$$

Equation (23) can also be written in the following way:

$$\sum_{P_1(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-1}) \left(\frac{\partial}{\partial t} + \hat{K}_1\right) F_1(l_s) = 0. \tag{24}$$

Therefore, Eq. (24) is satisfied if F_1 satisfies Eq. (22), so that the system of Eqs. (22) and (23) are compatible.

The first group of terms of order ϵ appears in BY_2;

$$\left(\frac{\partial}{\partial t} + \hat{K}_2\right) \mathrm{C}_2(1,2) - \alpha \hat{\mathrm{H}}_2 F_1(1) F_2(2) = 0. \tag{25}$$

Those from BY_s ($s > 2$) are

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} + \hat{K}_s\right) \sum_{P_2(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-2}) \mathrm{C}_2(l_{s-1}, l_s) \\
&\qquad - \alpha \hat{\mathrm{H}}_s F_1(1) \cdots F_1(s) = 0.
\end{aligned} \tag{26}$$

Using Eq. (15), the above expression can be rewritten

$$\begin{aligned}
&\sum_{P_3(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-3}) \\
&\quad \times \left\{ \sum_{P_2(i,j,k)}^{s-1,s-1,s} \mathrm{C}_2(l_j, l_k) \left[\frac{\partial}{\partial t} + \hat{K}_1\right] F_1(l_i) \right\} \\
&\quad + \sum_{P_2(l_1 \cdots l_s)} F_1(l_1) \cdots F_1(l_{s-2}) \\
&\quad \times \left\{ \left[\frac{\partial}{\partial t} + \hat{K}_2\right] \mathrm{C}_2(l_{s-1}, l_s) - \alpha_2 \hat{\mathrm{H}} F_1(l_{s-1}) F_1(l_s) \right\} = 0
\end{aligned} \tag{27}$$

which is satisfied if F_1 and C_2 are solutions to Eqs. (22) and (25). It follows that Eqs. (22) and (25) completely specify the solution to terms of $O(\epsilon)$.

The term of order ϵ^2, that appears in BY_1, is

$$\frac{\partial}{\partial \mathbf{p}_1} \cdot \int d\mathbf{z}_2 \mathbf{G}_{1,2} F_1(1) F_1(2) = 0, \tag{28}$$

and in BY_2

$$-\alpha \hat{\mathrm{H}}_2 \mathrm{C}_2(1,1) = \alpha\beta \sum_{i=1}^{2} \frac{\partial}{\partial \mathbf{p}_i} \cdot \int dz_3 \mathbf{G}_{1,3} F_1(1) F_1(2) F_1(3) \tag{29}$$

or simply,

$$\hat{\mathrm{H}}_2 \mathrm{C}_2(1,2) = 0, \tag{30}$$

since the right-hand term in Eq. (29) is equal to zero. Let us now consider BY_3;

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} + \hat{K}_3\right) \mathrm{C}_3(1,2,3) - \alpha \hat{\mathrm{H}}_3 \\
&\qquad \times \sum_{P_2(l_1, l_2, l_3)} F_1(l_1) \mathrm{C}_2(l_2, l_3) = 0.
\end{aligned} \tag{31}$$

In the above equation, the right-hand side vanished because of Eq. (28). This equation can be further simplified taking into account Eq. (30) and Eq. (18). There results,

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} + \hat{K}_3\right) \mathrm{C}_3(1,2,3) - \alpha \sum_{P_2(l_1, l_2, l_3)} \\
&\qquad \times \left\{ \sum_{j=2}^{3} \hat{\mathrm{H}}_2(l_1, l_j) F_1(l_1) \mathrm{C}_2(l_2, l_3) \right\} = 0.
\end{aligned} \tag{32}$$

The terms of order ϵ^2 in BY_4 are

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} + \hat{K}_4\right) \sum_{P_3(l_1 \cdots l_4)} F_1(l_1) \mathrm{C}_3(l_2, l_3, l_4) - \alpha \hat{\mathrm{H}}_4 \\
&\qquad \times \sum_{P_2(l_1 \cdots l_4)} F_1(l_1) F_1(l_2) \mathrm{C}_2(l_3, l_4) = 0
\end{aligned} \tag{33}$$

or using Eqs. (16), (18), (22), and (30),

$$\sum_{P_2(l_1\cdots l_4)} C_2(l_3, l_4)\hat{\mathfrak{H}}_2 F_1(l_1)F_1(l_2) = 0. \qquad (34)$$

Finally, it is possible to verify that all the terms of order ϵ^2 in the BY_s $(s > 4)$ equations vanish. Explicitly, these terms are

$$\left(\frac{\partial}{\partial t} + \hat{K}_s\right)\sum_{P_3(l_1\cdots l_s)} F_1(l_1)\cdots F_1(l_{s-3})C_3(l_{s-2}, l_{s-1}, l_s) - \alpha\hat{\mathfrak{H}}_s \sum_{P_2(l_1\cdots l_s)} F_1(l_1)\cdots F_1(l_{s-2})C_2(l_{s-1}, l_s) = 0 \qquad (35)$$

and using Eq. (16) and Eq. (18), this equation can be shown to be a linear combination of Eqs. (22), (30), (32), and (34).

Procedures, similar to the one described above, have been used in all the other nine cases. The results are tabulated below.

III. CONCLUSIONS

A. Compilation of Results

For each of the nine cases depicted in Table I, terms of at least three orders of magnitude from the largest[13] have been studied. In most cases the smallest-order terms included are of $O(\epsilon^2)$.

In the tables below, equations are grouped according to their order in ϵ. In each case, equations of $O(\epsilon^2)$ have been reduced using equations of $O(\epsilon^{s'})$ $(s' < s)$.

TABLE III.

Case I: $\alpha = 0(1/\epsilon)$; $\beta = 0(\epsilon)$.

Terms of order	Equations	Derived from
$1/\epsilon$	$\hat{\mathfrak{H}}_2 F_1(1)F_1(2) = 0$	BY_2
1	$\left(\frac{\partial}{\partial t} + K_1\right)F_1 = \alpha\beta\hat{I}_1 F_1(1)F_1(2)$	BY_1
	$\hat{\mathfrak{H}}_2 C_2(1, 2) = 0$	BY_2
	$\sum_{P_2(l_1,l_2,l_3)}\left\{\sum_{j=2}^{3}\hat{\mathfrak{H}}_2(l_1, l_j)F_1(l_1)C_2(l_2, l_3)\right\} = 0$	BY_3
ϵ	$\hat{I}_1 C_2(1, 2) = 0$	BY_1
	$\left(\frac{\partial}{\partial t} + K_2\right)C_2(1, 2) = \alpha\beta\left\{\sum_{i=1}^{2}\hat{I}_1(i)F_1(3)C_2(1, 2) + \sum_{P_1(l_1,l_2)}\hat{I}_1(l_1)F_1(l_1)C_2(l_2, 3)\right\}$	BY_3
	$\hat{\mathfrak{H}}_3 C_3(1, 2, 3) = 0$	
	$\sum_{P_3(l_1\ldots l_4)}\left\{\sum_{j=2}^{4}\hat{\mathfrak{H}}_2(l_1, l_j)F_1(l_1)C_3(l_2, l_3, l_4) = 0\right\}$	BY_4

TABLE IV.

Case II: $\alpha = 0(1/\epsilon)$; $\beta = 0(\epsilon)$.

Terms of order	Equations	Derived from
$1/\epsilon$	$\hat{I}_1 F_1(1)F_1(2) = 0$	BY_1
	$\hat{\mathfrak{H}}_2 F_1(1)F_1(2) = 0$	BY_2
1	$\left(\frac{\partial}{\partial t} + \hat{K}_1\right)F_1(1) = \alpha\beta\hat{I}_1 C_2(1, 2)$	BY_1
	$-\hat{\mathfrak{H}}_2 C_2(1, 2) = \beta\left\{\sum_{i=1}^{2}\hat{I}_1(i)F_1(3)C_2(1, 2) + \sum_{P_1(l_1,l_2)}\hat{I}_1(l_1)F_1(l_1)C_2(l_2, 3)\right\}$	BY_2
	$\sum_{P_2(l_1,l_2,l_3)}\left\{\sum_{j=2}^{3}\hat{\mathfrak{H}}_2(l_1, l_j)F_1(l_1)C_2(l_2, l_3)\right\} = 0$	BY_3

[13] The order of the largest order terms differs from case to case. In all, there are three possibilities. These are $O(\epsilon^{-2})$, $O(\epsilon^{-1})$, and $O(1)$.

TABLE IV (*Continued*).

	$\left(\frac{\partial}{\partial t}+\hat{K}_2\right)C_2(1,2)=\alpha\beta\sum_{i=1}^{2}\hat{I}_1(i)C_3(1,2,3)$	BY_2
ϵ	$\sum_{P_2(l_1\dots l_3)}C_2(l_2,l_3)\left(\frac{\partial}{\partial t}+K_1\right)F_1(l_1)-\alpha\hat{\mathfrak{H}}_3C_3(1,2,3)=\alpha\beta\left\{\sum_{i=1}^{3}\hat{I}_1(i)C_3(1,2,3)F_1(4)+\sum_{P_2(l_1\dots l_3)}\hat{I}_1(l_1)F_1(l_1)C_3(l_2,l_3,4)\right\}$	BU_3
	$\sum_{P_3(l_1\dots l_4)}\left\{\sum_{i=2}^{4}\hat{\mathfrak{H}}_2(l_1,l_i)F_1(l_1)C_3(l_2,l_3,l_4)\right\}=0$	BY_4

TABLE V.

Case III. $\alpha=0(1/\epsilon)$; $\beta=0(1/\epsilon)$.

Terms of order	Equations	Derived from
$1/\epsilon^2$	$\hat{I}_1F_1(1)F_1(2)=0$	BY_1
$1/\epsilon$	$\hat{I}_1C_2(1,2)=0$	BY_1
	$-\hat{\mathfrak{H}}_2F_1(1)F_1(2)=\beta\left\{\sum_{i=1}^{2}\hat{I}_1(i)F_1(3)C_2(1,2)+\sum_{P_1(l_1,l_2)}\hat{I}_1(l_1)F_1(l_1)C_2(l_2,3)\right\}$	BY_2
1	$\left(\frac{\partial}{\partial t}+\hat{K}_1\right)F_1(1)=0$	BY_1
	$-\alpha\hat{\mathfrak{H}}_2C_2(1,2)=\alpha\beta\sum_{i=1}^{2}\hat{I}_1(i)C_3(1,2,3)$	BY_2
	$-\sum_{P_2(l_1,l_2,l_3)}\left\{\sum_{i=1}^{3}\hat{\mathfrak{H}}_2(l_1,l_i)F_1(l_1)C_2(l_2,l_3)\right\}=\beta\sum_{i=1}^{3}\hat{I}_1(i)F_1(4)C_3(1,2,3)+\beta\sum_{P_2(l_1,l_2,l_3)}\hat{I}_1(l_1)F_1(l_1)C_3(l_2,l_3,4)$	BY_3
	$\sum_{P_2(l_1\dots l_4)}C_2(l_3,l_4)\hat{\mathfrak{H}}_2F_1(l_1)F_1(l_2)=0$	BY_4
ϵ	$\left(\frac{\partial}{\partial t}+\hat{K}_2\right)C_2(1,2)=0$	BY_2
	$-\hat{\mathfrak{H}}_3C_3(1,2,3)=\beta\sum_{i=1}^{3}\hat{I}_1(i)C_4(1,2,3,4)$	BY_3
	$-\sum_{P_3(l_1\dots l_4)}\sum_{i=2}^{4}\hat{\mathfrak{H}}_2(l_1,l_i)C_3(l_2,l_3,l_4)F_1(l_1)=\beta\sum_{i=1}^{4}\hat{I}_1(i)C_4(1,2,3,4)F_1(5)+\beta\sum_{P(l_1\dots l_4)}\hat{I}_1(l_1)F_1(l_1)C_4(l_2,l_3,l_4,5)$	BY_4
	$\sum_{P_3(l_1\dots l_5)}C_3(l_3,l_4,l_5)\hat{\mathfrak{H}}_2F_1(l_1)F_1(l_2)=0$	BY_5

TABLE VI.

Case IV: $\alpha=0(1)$; $\beta=0(\epsilon)$.

Terms of order	Equations	Derived from
1	$\left(\frac{\partial}{\partial t}+\hat{K}_1\right)F_1(1)=0$	BY_1
	$\hat{\mathfrak{H}}_2F_1(1)F_1(2)=0$	BY_2

TABLE VI (*Continued*).

	$\hat{I}_1 F_1(1)F_1(2) = 0$	BY_1
ϵ	$\left(\frac{\partial}{\partial t} + K_2\right)C_2(1,2) - \alpha\hat{\mathfrak{H}}_2 C_2(1,2) = 0$	BY_2
	$\sum_{P_2(l_1,l_2,l_3)} \sum_{i=2}^{3} \hat{\mathfrak{H}}_2(l_1, l_i)C_2(l_2, l_3)F_1(l_1) = 0$	BY_3
	$\hat{I}_1 C_2(1,2) = 0$	BY_1
	$\sum_{i=2}^{2} \hat{I}_1(i)C_2(1,2)F_1(3) + \sum_{P_1(l_1,l_2)} \hat{I}_1(l_1)F_1(l_1)C_2(l_2,3) = 0$	BY_2
ϵ^2	$\left(\frac{\partial}{\partial t} + K_3\right)C_3(1,2,3) - \alpha\hat{\mathfrak{H}}_3 C_3(1,2,3) = 0$	BY_3
	$\sum_{P_3(l_1 \ldots l_4)} \sum_{i=2}^{4} \hat{\mathfrak{H}}_2(l_1, l_i)F_1(l_1)C_3(l_2,l_3,l_4) = 0$	BY_4

TABLE VII.

Case V: $\alpha = 0(1)$; $\beta = 0(1)$.

Terms of order	Equations	Derived from
1	$\left(\frac{\partial}{\partial t} + \hat{K}_1\right)F_1(1) = \alpha\beta\hat{I}_1 F_1(1)F_1(2)$	BY_1
	$\hat{\mathfrak{H}}_2 F_1(1)F_1(2) = 0$	BY_2
	$\hat{I}_1 C_2(1,2) = 0$	BY_1
ϵ	$\left(\frac{\partial}{\partial t} + \hat{K}_2\right)C_2(1,2) - \alpha\hat{\mathfrak{H}}_2 C_2(1,2) = \alpha\beta\sum_{i=1}^{2} \hat{I}_1(i)F_1(3)C_2(1,2) + \alpha\beta\sum_{P_1(l_1,l_2)} \hat{I}_1(l_1)F_1(l_1)C_2(l_2,3)$	BY_2
	$\sum_{P_2(l_1,l_2,l_3)} \sum_{i=2}^{3} \hat{\mathfrak{H}}_2(l_1, l_i)F_1(l_1)C_2(l_1,l_3) = 0$	BY_3
	$\sum_{i=1}^{2} \hat{I}_1(i)C_3(1,2,3) = 0$	BY_2
ϵ^2	$\left(\frac{\partial}{\partial t} + K_3\right)C_3(1,2,3) - \alpha\hat{\mathfrak{H}}_3 C_3(1,2,3) = \alpha\beta\sum_{i=1}^{3} \hat{I}_1(i)F_1(4)C_3(1,2,3) + \alpha\beta\sum_{P_2(l_1,l_2,l_3)} \hat{I}_1(l_1)F_1(l_1)C_3(l_2,l_3,4)$	BY_3
	$\sum_{P_3(l_1 \ldots l_4)} \sum_{i=2}^{4} \hat{\mathfrak{H}}_2(l_1, l_i)F_1(l_1)C_3(l_2,l_3,l_4) = 0$	BY_4

TABLE VIII.

Case VI: $\alpha = 0(1)$; $\beta = 0(1/\epsilon)$.

Terms of order	Equations	Derived from
$1/\epsilon$	$\hat{I}_1 F_1(1)F_1(2) = 0$	BY_1
	$\left(\frac{\partial}{\partial t} + \hat{K}_1\right)F_1(1) = \alpha\beta\hat{I}_1 C_2(1,2)$	BY_1
1	$-\hat{\mathfrak{H}}_2 F_1(1)F_1(2) = \beta\sum_{i=1}^{2} \hat{I}_1(i)F_1(3)C_2(1,2) + \beta\sum_{P_1(l_2,l_3)} \hat{I}_1(l_1)F_1(l_1)C_2(l_2,3)$	BY_2

TABLE VIII (*Continued*).

	$\left(\frac{\partial}{\partial t}+\hat{K}_2\right)C_2(1,2)-\alpha\hat{\mathbb{H}}_2C_2(1,2)=\alpha\beta\sum_{i=1}^{2}\hat{I}_1(i)C_3(1,2,3)$	BY_2
ϵ	$\sum_{P_2(l_1,l_2,l_3)}C_2(l_2,l_3)\left(\frac{\partial}{\partial t}+\hat{K}_1\right)F_1(l_1)-\alpha\sum_{P_3(l_1,l_2,l_3)}\sum_{i=2}^{3}\hat{\mathbb{H}}_2(l_1,l_i)F_1(l_1)C_2(l_2,l_3)$	BY_3
	$=\alpha\beta\sum_{i=1}^{3}\hat{I}_1(i)F_1(4)C_3(1,2,3)+\alpha\beta\sum_{P_2(l_1,l_2,l_3)}\hat{I}_1(l_1)F_1(l_1)C_3(l_2,l_3,4)$	
	$\sum_{P_2(l_1\ldots l_4)}C_2(l_3,l_4)\hat{\mathbb{H}}_2F_1(l_1)F_1(l_2)=0$	BY_4

TABLE IX.

Case VII: $\alpha=0(\epsilon)$; $\beta=0(\epsilon)$.

Terms of order	Equations	Derived from
1	$\left(\frac{\partial}{\partial t}+\hat{K}_1\right)F_1(1)=0$	BY_1
ϵ	$\left(\frac{\partial}{\partial t}+\hat{K}_2\right)C_2(1,2)-\alpha\hat{\mathbb{H}}_2F_1(1)F_1(2)=0$	BY_2
	$\hat{I}_1F_1(1)F_1(2)=0$	BY_1
	$\hat{\mathbb{H}}_2C_2(1,2)=0$	BY_2
ϵ^2	$\left(\frac{\partial}{\partial t}+\hat{K}_3\right)C_3(1,2,3)-\alpha\sum_{P_2(l_1,l_2,l_3)}\sum_{i=2}^{3}\hat{\mathbb{H}}_2(l_1,l_i)F_1(l_1)C_2(l_2,l_3)=0$	BY_3
	$\sum_{P_2(l_1\ldots l_2)}C_2(l_3,l_4)\hat{\mathbb{H}}_2F_1(l_1)F_1(l_2)=0$	BY_4

TABLE X.

Case VIII: $\alpha=0(\epsilon)$; $\beta=0(1)$.

Terms of order	Equations	Derived from
1	$\left(\frac{\partial}{\partial t}+\hat{K}_1\right)F_1(1)=0$	BY_1
	$\hat{I}_1F_1(1)F_1(2)=0$	BY_1
ϵ	$\left(\frac{\partial}{\partial t}+\hat{K}_2\right)C_2(1,2)-\alpha\hat{\mathbb{H}}_2F_1(1)F_1(2)=0$	BY_2
	$\hat{I}_1C_2(1,2)=0$	BY_1
	$-\hat{\mathbb{H}}_2C_2(1,2)=\beta\left\{\sum_{i=1}^{2}\hat{I}_1(i)F_1(3)C_2(1,2)+\sum_{P_2(l_1,l_2)}\hat{I}_1(l_1)F_1(l_1)C_2(l_2,3)\right\}$	BY_2
ϵ^2	$\left(\frac{\partial}{\partial t}+\hat{K}_3\right)(1,2,3)-\alpha\sum_{P_2(l_1,l_2,l_3)}\sum_{i=2}^{3}\hat{\mathbb{H}}_2(l_2,l_i)F_1(l_1)C_2(l_2,l_3)=0$	BY_3
	$\sum_{P_2(l_1\ldots l_4)}C_2(l_3,l_4)\hat{\mathbb{H}}_2F_1(l_1)F_1(l_2)=0$	BY_4

TABLE XI.

Case IX: $\alpha = 0(\epsilon)$; $\beta = 0(1/\epsilon)$.

Terms of order	Equations	Derived from
1	$\left(\frac{\partial}{\partial t} + \hat{K}_1\right)F_1(1) = \alpha\beta \hat{I}_1 F_1(1)F_1(2)$	BY_1
	$\hat{I}_1 C_2(1,2) = 0$	BY_1
ϵ	$\left(\frac{\partial}{\partial t} + \hat{K}_2\right)C_2(1,2) - \alpha\hat{\mathfrak{H}}_2 F_1(1)F_1(2) = \alpha\beta \sum_{i=1}^{3} \hat{I}_1(i)C_2(1,2)F_1(3) + \alpha\beta \sum_{P_1(l_1,l_2)} \hat{I}_1(l_1)F_1(l_1)C_2(l_2,3)$	BY_2
	$-\hat{\mathfrak{H}}_2 C_2(1,2) = \beta \sum_{i=1}^{2} \hat{I}_1(i)C_3(1,2,3)$	BY_2
ϵ^2	$\left(\frac{\partial}{\partial t} + \hat{K}_3\right)C_3(1,2,3) - \alpha \sum_{P_2(l_1,l_2,l_3)} \sum_{i=2}^{3} \hat{\mathfrak{H}}_2(1,l_i)F_1(l_1)C_2(l_2,l_3)$	BY_3
	$= \alpha\beta \sum_{i=1}^{3} \hat{I}_1(i)F_1(4)C_3(1,2,3) + \alpha\beta \sum_{P_2(l_1,l_2,l_3)} \hat{I}_1(l_1)F_1(l_1)C_3(l_3,l_2,4)$	
	$\sum_{P_2(l_1 \ldots l_4)} C_2(l_3,l_4)\hat{\mathfrak{H}}_2 F_1(l_1)F_1(l_2) = 0$	BY_4

B. Comments on Results

The physical significance of each of the nine cases presented above is given in terms of the coefficients α and β. To realize this significance, we recall [Eq. (9)] that β is proportional to the third power of the ratio between the "range" r_0 of the potential, and r_a, the average distance between particles.

If the system is called rarefied when $(r_0/r_a) \ll 1$ and dense when $(r_0/r_a) \gg 1$, Cases I, IV, and VII correspond to rarefied systems and Cases III, VI, and IX to dense systems.

The parameter α is proportional to the strength of the potential and inversely proportional to the mean thermal energy per particle. It follows that Cases I, II, and III correspond to systems of either strong interaction or low temperature. In particular, Case I corresponds to the "rigid sphere" approximation. Case VII describes high temperature, rarefied systems and Case IX, high temperature dense systems.

From inspection of the above nine tables, we see that all the $\{BY_s^{(k)} = 0\}$ equations, at least in the cases considered ($k = 0, 1, 2$), can be reduced to a very small number of nonredundant (but hopefully compatible) equations. In particular $\{BY_s^{(o)} = 0\}$ can always be reduced to one, or at most, two equations.

Cases VII and VIII yield the "streaming" equation to lowest order. Both these cases correspond to systems which include vanishing interaction. Such a state is realized at either high temperature, or vanishing "strength" of potential. It is interesting to note that although Cases I and IV include states of vanishing density, in Case IV where the strength of the potential is comparable to the mean thermal energy, the system obeys a modified streaming equation; while in Case I, where the strength of the potential is much larger than the mean thermal energy, the streaming state is destroyed.

Case IX to lowest order, yields the Vlasov equation. This by far is the most revealing result of the present analysis. What is known from plasma kinetic theory studies[10] is that the Vlasov equation follows from the Liouville equation in the limit $(nd^3)^{-1} \to 0$, where d is the Debye distance. This is consistent with the above result inasmuch as $\alpha\beta = 1$ and $\beta \sim \epsilon^{-1}$ (Case IX) implies that r_0 is the Debye distance and β becomes (nd^3).

However, what has not been uncovered in such analyses is that the Vlasov equation also appears in two other cases implied by the single constraint $\alpha\beta = 1$. These cases are Case I and Case V. In Case I, which includes strong interaction and rare density (e.g., rigid sphere), the Vlasov equation appears in the $O(1)$ equations. The $O(\epsilon^{-1})$ equation is a differential constraint on F_1 and appears as

$$\hat{\mathfrak{H}}_2 F_1(1)F_1(2) = 0.$$

In Case V, where both α and β are of $O(1)$, the lowest-order equations are the Vlasov equation together with the same differential constraint Eq. (36), as appeared in Case I.

TABLE XII. Summary of type-1 constraint equations.

Case I	Case II	Case III
Modified Vlasov $\alpha\beta = 1$	Integral + differential. constraint $\alpha\beta = \epsilon^{-1}$	Integral constraint $\alpha\beta = \epsilon^{-2}$
Case IV	**Case V**	**Case VI**
Modified Streaming $\alpha\beta = \epsilon$	Modified Vlasov $\alpha\beta = 1$	Integral constraint $\alpha\beta = \epsilon^{-1}$
Case VII	**Case VIII**	**Case IX**
Modified Streaming $\alpha\beta = \epsilon^2$	Streaming $\alpha\beta = \epsilon$	Vlasov $\alpha\beta = 1$

These results suggest that a modified Vlasov equation (*viz.*, together with differential constraint) gives a proper kinetic description of a rigid sphere gas.

A summary of the above results is given in Table XII.

Although this study is clearly an introductory discourse, nevertheless, it is quite clear that the analysis already assumes a burdensome quality. Many outstanding problems are suggested. The most evident of these are:

A. Extension of the analysis to terms of $O(\epsilon^s)$, $s > 2$.

B. Determination of the existence of solutions. In this domain, particular attention must be given to the compatibility requirements discussed above, i.e., between type-*l* constraint equations.

ACKNOWLEDGMENT

This research was supported in part by the Physics Branch of the Office of Naval Research under contract Nonr 401-(60).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Analytic Properties of Certain Radial Matrix Elements

H. ALMSTRÖM AND P. OLSSON
Department of Mathematical Physics, Royal Institute of Technology, Stockholm, Sweden

(Received 6 November 1966)

The analytic properties of the Laplace transform of products of two confluent hypergeometric functions as function of certain parameters are investigated. The radial matrix element of multipole type that contain Coulomb–Dirac wavefunctions or special cases thereof are considered. These are investigated as functions of the momenta p of the particles. Branch cuts and the behavior along the branch cuts in the complex p plane are given. The formulas presented permit accurate calculation of the matrix elements also, when the parameters have values arbitrarily close to singular points. The prime concern, however, is to facilitate second- and higher-order perturbation calculations.

IN this paper we investigate the analytic properties of certain radial matrix elements of a type that occurs, e.g., in the theory of multipole radiation. Such a matrix may be written

$$\langle p_1|\ V\ |p_2\rangle = \int_0^\infty \Phi_i(p_1 r) e^{-sr} r^{\alpha-1} \Phi_k(p_2 r)\, dr, \qquad i, k = 1, 2, \quad (1)$$

where the functions Φ_1 and Φ_2 may be expressed in terms of confluent hypergeometric functions

$$ {}_1F_1(b, c, x) = \sum_{n=0}^{\infty} \frac{(b)_n x^n}{(c)_n n!},$$

$$\Psi(b, c, x) = \frac{\Gamma(1-c)}{\Gamma(b-c+1)}\, {}_1F_1(b, c, x) + \frac{\Gamma(c-1)}{\Gamma(b)}\, x^{1-c}\, {}_1F_1(b-c+1, 2-c, x), \quad (2)$$

as follows:

$$\begin{aligned} \Phi_1(pr) &= e^{ipr}\, {}_1F_1(\gamma - i\eta, 2\gamma, -2ipr), \\ \Phi_2(pr) &= e^{ipr}\Psi(\gamma - i\eta, 2\gamma, -2ipr). \end{aligned} \quad (3)$$

It can be shown that the functions

$$u_i(r) = r^\gamma \Phi_i(pr), \quad i = 1, 2,$$

are two independent solutions of the equation

$$\{(d^2/dr^2) + E - (\beta/r) - [\gamma(\gamma-1)/r^2]\} u(r) = 0, \quad (4)$$

provided that

$$p = E^{\frac{1}{2}}, \quad \eta = -\beta/2p;$$

and that the functions

$$f_i(r) = \frac{\alpha Z}{\kappa - \gamma} r^{\gamma-1}\left[E - 1 + \frac{\gamma - \kappa}{\alpha Z}\frac{d}{dr}\right]\Phi_i(pr),$$

$$g_i(r) = -r^{\gamma-1}\left[E + 1 + \frac{\gamma + \kappa}{\alpha Z}\frac{d}{dr}\right]\Phi_i(pr), \quad i = 1, 2,$$

are two independent solutions of the equations

$$\begin{aligned} \frac{df(r)}{dr} &= \frac{\kappa - 1}{r} f(r) - \left(E - 1 + \frac{\alpha Z}{r}\right) g(r), \\ \frac{dg(r)}{dr} &= \left(E + 1 + \frac{\alpha Z}{r}\right) f(r) - \frac{\kappa + 1}{r} g(r), \end{aligned} \quad (5)$$

provided

$$p = (E^2 - 1)^{\frac{1}{2}}, \quad \eta = E\alpha Z/p, \quad \gamma = (\kappa^2 - \alpha^2 Z^2)^{\frac{1}{2}},$$

where κ takes any of the values $\kappa = \pm 1, \pm 2, \pm 3, \cdots$.

Equations (4) and (5) are the Schrödinger and Dirac equations with a Coulomb potential included as they appear after a separation in spherical coordinates. In order to obtain the free solutions we put, of course, $\eta = 0$.

The solutions in (3) have been chosen because, accepting the convention $0 \leq \arg p \leq \pi$ for the square root of p, one solution is square integrable at the origin and one at infinity, since

$$\Phi_1(0) = 1,$$

and[1]

$$|\Phi_2(r)| \approx r^{-[\gamma + \mathrm{Im}(\eta)]} e^{-\mathrm{Im}(pr)}.$$

Clearly Φ_2 is square integrable for any nonreal value of E.

If we exclude the relativistic Coulomb wavefunctions the function $\Phi_1(pr)$ is an integral function of p. The point $p = 0$, for which the parameter $b = \gamma + i(\beta/2p)$ may be infinite, is no exception as can be verified by a closer investigation. Since the function $\Phi_2(pr)$ is defined by (2) and (3) it is an analytic function of p in the entire p plane except on a branch cut along the negative imaginary p axes which cut is caused by the factor x^{1-c} in (2). When $c = 2\gamma$ is an integer the

[1] Bateman Manuscript Project, *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, p. 278.

function Ψ is the limit of (2) as c approaches the integer value. In this limit a term $\ln x$ will appear in the Ψ function so that the cut is present whether c is integer or not.

In case $\Phi_1(pr)$ is an relativistic Coulomb wavefunction it depends on $E = (p^2 + 1)^{\frac{1}{2}}$ through the parameter η as it is defined in (5). The function is an integral function of p^2 and E since it has the expansion

$$\Phi_1(pr) = \sum_{n=0}^{\infty} c_n \frac{(2\alpha Zr)^n}{n!},$$

where the coefficients c_n are obtained from the recurrence relation

$$(n + 2\gamma)c_{n+1} + Ec_n + \frac{p^2}{4\alpha^2 Z^2} n \cdot c_{n-1} = 0, \quad c_0 = 1,$$

and since the series is uniformly convergent in p^2 and E. It is no longer an integral function of p, since it has the branch points $p = \pm i$ of $(p^2 + 1)^{\frac{1}{2}}$.

In the following arguments concerning the analyticity of the matrix elements (1) the relativistic Coulomb case is excluded throughout. Its inclusion simply implies that the branch cuts from $p = \pm i$ to infinity along the imaginary p axes are added to the branch cuts that are obtained in this paper.

After this preliminary discussion of the analytic properties of the standing and the outgoing wave solutions we proceed to the discussion of the matrix elements, beginning with

$$\langle p_1| V |p_2^+\rangle = \int_0^{\infty} e^{-(s - ip_1 - ip_2)r} r^{a-1} {}_1F_1(b_1, c_1, -2ip_1 r) \times \Psi(b_2, c_2, -2ip_2 r)\, dr, \quad (6)$$

where, for the sake of brevity, we have introduced

$$b_k = \gamma_k - i\eta_k, \quad c_k = 2\gamma_k, \quad k = 1, 2.$$

The integral exists when $\mathrm{Re}\,(a - c_2 + 1) > 0$ and $\mathrm{Re}\,(s \pm ip_1 - ip_2) > 0$ which follows from (2) and the asymptotic behavior of the confluent hypergeometric functions.[2] When it exists it is an analytic function of p_1 and also of p_2 if a cut is made along the negative imaginary p_2 axis where the integrand is nonanalytic.

Our aim is now to continue analytically the matrix element (6) as function of p_1 or p_2. This can be accomplished by expressing it in terms of an associated Appell function[3]

$$\langle p_1| V |p_2^+\rangle = \frac{\Gamma(a)\Gamma(a - c_2 + 1)}{\Gamma(a + b_2 - c_2 + 1)} (s - ip_1 - ip_2)^{-a} \times F_P(a, b_1, b_2, c_1, c_2, z_1, z_2), \quad (7)$$

where

$$z_k = -2ip_k/(s - ip_1 - ip_2), \quad k = 1, 2,$$

and the F_P function is defined by the expansion

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = \sum_{m=0}^{\infty} \frac{(a)_m (b_1)_m (a - c_2 + 1)_m z_1^m}{(c_1)_m (a + b_2 - c_2 + 1)_m m!} \times {}_2F_1(a + m, b_2, a + b_2 - c_2 + m + 1, 1 - z_2), \quad |z_1| < 1, \quad z_2 \neq 0. \quad (8)$$

The investigation of the analytic properties of the matrix elements (1) is, in fact, intimately connected with the theory of the Appell functions, which were defined by Appell as certain solutions of a set of partial differential equations.[4] We also need solutions which are not Appell functions. Such solutions are called associated Appell functions and are used in order to define the Appell functions for the whole range of their arguments.

In large the analytic properties of the F_P function are most easily obtained from integral representations. Choosing a suitable integral representation of the hypergeometric functions appearing in the terms of the series (8) (Ref. 5)

$${}_2F_1(a + m, b_2, a + b_2 - c_2 + m + 1, 1 - z_2) = \frac{\Gamma(a + b_2 - c_2 + 1)}{\Gamma(a)\Gamma(b_2 - c_2 + 1)} \frac{(a + b_2 - c_2 + 1)_m}{(a)_m} \times \int_0^1 t^{a+m-1}(1 - t)^{b_2 - c_2}[1 - t(1 - z_2)]^{-b_2}\, dt, \quad \mathrm{Re}\, a > 0, \quad \mathrm{Re}\,(b_2 - c_2 + 1) > 0,$$

and inserting it into (8) we obtain

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = \frac{\Gamma(a + b_2 - c_2 + 1)}{\Gamma(a)\Gamma(b_2 - c_2 + 1)} \sum_{m=0}^{\infty} \frac{(b_1)_m (a - c_2 + 1)_m}{(c_1)_m m!} \times \int_0^1 t^{a-1}(tz_1)^m (1 - t)^{b_2 - c_2}[1 - t(1 - z_2)]^{-b_2}\, dt.$$

When $|z_1| < 1$ we may sum under the sign of integration which gives

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = \frac{\Gamma(a + b_2 - c_2 + 1)}{\Gamma(a)\Gamma[b_2 - c_2 + 1)} \int_0^1 t^{a-1}(1 - t)^{b_2 - c_2} \times [1 - t(1 - z_2)]^{-b_2} {}_2F_1(a - c_2 + 1, b_1, c_1, tz_1)\, dt, \quad \mathrm{Re}\, a > 0, \quad \mathrm{Re}\,(b_2 - c_2 + 1) > 0.$$

Since the condition

$$\mathrm{Re}\,(b_2 - c_2 + 1) = \mathrm{Re}\,(-i\eta_2 - \gamma_2 + 1) > 0$$

[2] Ref. 1, p. 278.
[3] P. O. M. Olsson, Ark. Fysik 30, 187 (1965), especially p. 190.
[4] Ref. 1, p. 234.
[5] Ref. 1, p. 59.

in general is not fulfilled, we transform the integral with the aid of[6]

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = z_2^{1-c_2} F_P(a - c_2 + 1, b_1, b_2 - c_2 + 1, c_1, 2 - c_2, z_1, z_2),$$

and obtain

$$\frac{\Gamma(a - c_2 + 1)}{\Gamma(a + b_2 - c_2 + 1)} F_P(a, b_1, b_2, c_1, c_2, z_1 z_2) = \frac{z_2^{1-c_2}}{\Gamma(b_2)} \int_0^1 t^{a-c_2}(1 - t)^{b_2-1}[1 - t(1 - z_2)]^{c_2-b_2-1} \times {}_2F_1(a, b_1, c_1, tz_1)\, dt, \quad \mathrm{Re}\,(a - c_2 + 1) > 0, \quad \mathrm{Re}\, b_2 > 0. \quad (9)$$

The ${}_2F_1$ function is an integral function of b_1 and an analytic function of its argument $x = tz_1$ in a cut x plane from $x = 1$ to $x = +\infty$. Since the right-hand side of (9) is an analytic function of p_1 if p_1 takes values such that the singularities of the integrand are avoided along the path of integration, the possible nonanalytic domains are the branch cut of the ${}_2F_1$ function

$$tz_1 = 1 + \xi_1, \quad 0 \le \xi_1, \quad 0 \le t \le 1,$$

and the branch cut

$$1 - t(1 - z_2) = -\xi_2, \quad 0 \le \xi_2, \quad 0 \le t \le 1,$$

of the factor $[1 - t(1 - z_2)]^{c_2-b_2-1}$. This may be written

$$z_1 = 1 + \xi_1, \quad z_2 = -\xi_2,$$

where we have not bothered to introduce new notations for the parameters ξ as long as they have the same range. The points to be avoided are then

$$p_1 = -p_2(1 + \xi_1) - is,$$
$$p_1 = \pm(p_2 + is)(1 + \xi_2).$$

The point $p_1 = 0$, for which $b_1 = \gamma_1 - i\eta_1$ may be infinite, need not be considered since we know already from (6) that the matrix element is analytic when $\mathrm{Re}\,(s \pm ip_1 - ip_2) > 0$ which is always the case for sufficiently small values of $|p_1|$ since $s > 0$ and p_2 is assumed to be real.

The factor $(s - ip_1 - ip_2)^{-a}$ in (7) contributes with the cut

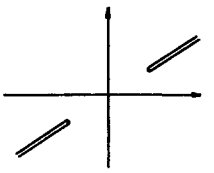
$$p_1 = -p_2 - i(s + \xi_3), \quad 0 \le \xi_3,$$

in case the exponent is noninteger. There are then three possible cuts ending at the singular point $p_1 = -p_2 - is$ but only one ending at $p_1 = p_2 + is$. Since

$$\Phi_1(pr) = \Phi_1(-pr),$$

[6] Ref. 3, p. 189.

FIG. 1. Analytic domain of $\langle p_1|\, V\, |p_2^+\rangle$ in the p_1 plane.



the matrix element considered is an even function of p_1 and the cuts

$$p_1 = \pm(p_2 + is)(1 + \xi_2), \quad 0 \le \xi_2$$

are the only ones that cannot be discarded. Hence the matrix element $\langle p_1|\, V\, |p_2^+\rangle$ is an analytic function of p_1 in a cut p_1 plane as in Fig. 1.

The properties of the matrix element as function of p_2 are not conveniently obtained from (9) since the existence of the integral requires $\mathrm{Re}\, b_2 = \gamma_2 + \mathrm{Im}\, \eta_2 > 0$. An integral representation which is more appropriate in this case can be obtained from the five parametric double-series expansion of the F_P function[7]

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = z_2^{-a} \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(a - c_2 + 1)_{m+n}(b_1)_m}{(a + b_2 - c_2 + 1)_{m+n}(c_1)_m m!\, n!} \times \left(\frac{z_1}{z_2}\right)^m \left(\frac{z_2 - 1}{z_2}\right)^n,$$

in the following way:

$$\sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(a - c_2 + 1)_{m+n}(b_1)_m}{(a + b_2 - c_2 + 1)_{m+n}(c_1)_m m!\, n!} \left(\frac{z_1}{z_2}\right)^m \left(\frac{z_2 - 1}{z_2}\right)^n$$
$$= \sum_{N=0}^{\infty} \frac{(a)_N(a - c_2 + 1)_N}{(a + b_2 - c_2 + 1)_N N!} \times \sum_{m=0}^{N} \frac{(b_1)_m N!}{(c_1)_m (N - m)!\, m!} \left(\frac{z_1}{z_2}\right)^m \left(\frac{z_2 - 1}{z_2}\right)^{N-m}$$
$$= \sum_{N=0}^{\infty} \frac{(a)_N(a - c_2 + 1)_N}{(a + b_2 - c_2 + 1)_N N!} \left(\frac{z_2' - 1}{z_2}\right)^N \times \sum_{m=0}^{N} \frac{(-N)_m(b_1)_m}{(c_1)_m m!} \left(\frac{z_1}{1 - z_2}\right)^m$$
$$= \sum_{N=0}^{\infty} \frac{(a)_N(a - c_2 + 1)_N}{(a + b_2 - c_2 + 1)_N N!} \left(\frac{z_2 - 1}{z_2}\right)^N \times {}_2F_1\left(-N, b_1, c_1, \frac{z_1}{1 - z_2}\right)$$
$$= \frac{\Gamma(c_1)}{\Gamma(b_1)\Gamma(c_1 - b_1)} \int_0^1 t^{b_1-1}(1 - t)^{c_1-b_1-1} \times {}_2F_1\left(a, a - c_2 + 1, a + b_2 - c_2 + 1, \frac{z_1 t + z_2 - 1}{z_2}\right) dt, \quad \mathrm{Re}\, b_1 > 0, \quad \mathrm{Re}(c_1 - b_1) > 0.$$

[7] Ref. 3, p. 187.

In the last step we have used

$$
{}_2F_1\left(-N, b_1, c_1, \frac{z_1}{1-z_2}\right)
= \frac{\Gamma(c_1)}{\Gamma(b_1)\Gamma(c_1-b_1)} \int_0^1 t^{b_1-1}(1-t)^{c_1-b_1-1} \times \left(1-\frac{z_1 t}{1-z_2}\right)^N dt,
$$

$$
\text{Re } b_1 > 0, \quad \text{Re } c_1 - b_1 > 0,
$$

and summed under the sign of integration. Hence

$$
F_P(a, b_1, b_2, c_1, c_2, z_1, z_2)
= \frac{\Gamma(c_1)}{\Gamma(b_1)\Gamma(c_1-b_1)} z_2^{-a} \int_0^1 t^{b_1-1}(1-t)^{c_1-b_1-1} \times {}_2F_1\left(a, a-c_2+1, a+b_2-c_2+1, \frac{z_1 t + z_2 - 1}{z_2}\right) dt,
$$

$$
\text{Re } b_1 > 0, \quad \text{Re}(c_1 - b_1) > 0, \quad (10)
$$

and

$$
\langle p_1| V |p_2^+\rangle = \frac{\Gamma(a)\Gamma(a-c_2+1)\Gamma(c_1)}{\Gamma(b_1)\Gamma(c_1-b_1)} (-2ip_2)^{-a} \times \int_0^1 t^{b_1-1}(1-t)^{c_1-b_1-1} \times \frac{{}_2F_1\left(a, a-c_2+1, a+b_2-c_2+1, \frac{z_1 t + z_2 - 1}{z_2}\right)}{\Gamma(a+b_2-c_2+1)} dt,
$$

$$
\text{Re } b_1 > 0, \quad \text{Re}\,(c_1 - b_1) > 0. \quad (11)
$$

The function

$$
{}_2F_1(a, b, c, x)/\Gamma(c)
$$

is an integral function of c for every regular point[8] x from which follows that the integral in (11) is an analytic function of p_2 except possibly at $p_2 = 0$, where b_2 may be infinite and points p_2 for which

$$
(z_1 t + z_2 - 1)/z_2 = 1 + \xi, \quad 0 \le \xi, \quad 0 \le t \le 1,
$$

or

$$
p_2 = [1/(1+2\xi)][p_1(2t-1) - is].
$$

This excludes a triangular domain with corners at the points $p_2 = 0$ and $p_2 = \pm p_1 - is$ in the p_2 plane.

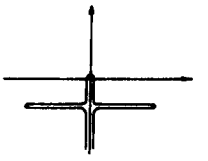


FIG. 2. Analytic domain of $\langle p_1| V |p_2^+\rangle$ in the p_2 plane.

In case the parameter a is noninteger, the factor $(-2ip_2)^{-a}$ contributes with a cut along the negative imaginary p_2 axes. Since we know already that the matrix element is analytic when $\text{Im}\, p_2 > -s$ (p_1 is now assumed to be real), except for the cut along the negative imaginary p_2 axes, we obtain the analytic domain in Fig. 2.

Either function in (6) may be a bound state wavefunction by letting p_1 or p_2 take the values on the positive imaginary axes which correspond to the eigenvalues of E.

We next consider the matrix element

$$
\langle p_1^+| V |p_2^+\rangle = \int_0^\infty e^{-(s-ip_1-ip_2)r} r^{a-1} \Psi(b_1, c_1, -2ip_1 r) \times \Psi(b_2, c_2, -2ip_2 r)\, dr. \quad (12)
$$

This integral exists when $\text{Re}\,(a - c_1 - c_2 + 2) > 0$ and $\text{Re}\,(s - ip_1 - ip_2) > 0$. There is obviously no distinction between the variables p_1 and p_2 as far as possible singular domains are concerned. The integral is uniformly convergent in p_1 when $s + \text{Im}\, p_1 > 0$ and since the integrand is an analytic function of p_1 in this domain, cut along the negative imaginary p_1 axes, the integral is analytic in the domain. Analytic continuations are obtained by expressing the matrix element (12) in terms of an associated Appell function[9]

$$
\langle p_1^+| V |p_2^+\rangle = (s - ip_1 - ip_2)^{-a} \times F_R(a, b_1, b_2, c_1, c_2, z_1, z_2).
$$

This function has an integral representation

$$
F_R(a, b_1, b_2, c_1, c_2, z_1, z_2) = \frac{\Gamma(a)}{\Gamma(b_1)\Gamma(b_2)} \int_0^\infty \int_0^\infty t_1^{b_1-1} t_2^{b_2-1}(1+t_1)^{c_1-b_1-1} \times (1+t_2)^{c_2-b_2-1}(1+t_1 z_1 + t_2 z_2)^{-a}\, dt_1\, dt_2,
$$

subject to convergence conditions, but may also be represented by the series expansion[9]

$$
F_R(a, b_1, b_2, c_1, c_2, z_1, z_2) = \Gamma(a)\Gamma(a-c_1+1)\Gamma(a-c_2+1) z_1^{1-c_1} z_2^{1-c_2} \sum_{n=0}^{\infty} \frac{(a)_n (b_1)_n (b_2)_n}{n!} \times \frac{{}_2F_1(a-c_1+1, b_1-c_1+1, a+b_1-c_1+n+1, 1-z_1)}{\Gamma(a+b_1-c_1+n+1)} \times \frac{{}_2F_1(a-c_2+1, b_2-c_2+1, a+b_2-c_2+n+1, 1-z_2)}{\Gamma(a+b_2-c_2+n+1)}. \quad (13)
$$

[8] Ref. 1, p. 68.

[9] P. O. M. Olsson, Ark. Fysik **28**, 113 (1964).

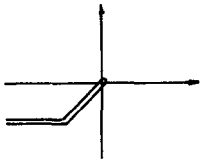


FIG. 3. Analytic domain of $\langle p_1^+| V |p_2^+\rangle$ in the p_1 plane.

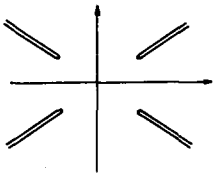


FIG. 4. Analytic domain of $\langle p_1| V |p_2\rangle$ in the p_1 or p_2 plane.

The series is absolutely convergent when

$$\mathrm{Re}\,(a - c_1 - c_2 + 2) > 0,$$

$z_1 \neq 0$ and $z_2 \neq 0$. As we have seen, the terms are analytic functions of p_1 except on the branch cuts of the ${}_2F_1$ functions

$$z_1 = -\xi_1, \quad 0 \leq \xi_1,$$
$$z_2 = -\xi_2, \quad 0 \leq \xi_2,$$

and the point $p_1 = 0$ for which b_1 may be infinite. This gives the two possible cuts

$$p_1 = (-p_2 - is)/(1 + \xi_1),$$
$$p_1 = -p_2(1 + \xi_2) - is,$$

which are drawn in Fig. 3.

Since we may write

$$\langle p_1^+| V |p_2^+\rangle = (-2ip_2)^{-a} z_2^a \times F_R(a, b_1, b_2, c_1, c_2, z_1, z_2),$$

in case p_1 is the variable, the matrix element has the same branch cuts.

It is possible to define the F_R function with a cut along the negative imaginary axes in the p_1 plane. This would agree with the fact that the integrand in (12) has a cut there but we prefer to define the F_R function by (13) with conventional concepts concerning the ${}_2F_1$ functions.

There remains now to investigate the matrix element

$$\langle p_1| V |p_2\rangle = \int_0^\infty e^{-(s - ip_1 - ip_2)r} r^{a-1} {}_1F_1(b_1, c_1, -2ip_1 r) \times {}_1F_1(b_2, c_2, -2ip_2 r)\, dr. \quad (14)$$

The integral exists when $\mathrm{Re}\, a > 0$ and

$$\mathrm{Re}\,(s \pm ip_1 \pm ip_2) > 0$$

and is then an analytic function of p_1 or p_2. Again there is no distinction between the variables as long as we are interested in possible singular domains only. It may be expressed in terms of the Appell function[10] F_2

$$\langle p_1| V |p_2\rangle = \Gamma(a)(s - ip_1 - ip_2)^{-a} \times F_2(a, b_1, b_2, c_1, c_2, z_1, z_2), \quad (15)$$

defined by the double series[11]

$$F_2(a, b_1, b_2, c_1, c_2, z_1, z_2) = \sum_{m,n=0}^{\infty} \frac{(a)_{m+n}(b_1)_m(b_2)_n z_1^m z_2^n}{(c_1)_m(c_2)_n m!\, n!}, \quad |z_1| + |z_2| < 1.$$

Analytic continuations may now conveniently be obtained by expressing the matrix element in terms of the previously discussed matrix elements $\langle p_1| V |p_2^+\rangle$. In order to obtain such an expression we replace one of the standing wave solutions in (14) by a combination of incoming and outgoing solutions, using[12]

$${}_1F_1(b, c, x) = \frac{\Gamma(c)}{\Gamma(c - b)} e^{i\pi b\, \mathrm{sip}\, x} \Psi(b, c, x) + \frac{\Gamma(c)}{\Gamma(b)} e^{-i\pi(c-b)\, \mathrm{sip}\, x} \Psi(c - b, c, -x).$$

Here sip x means the sign of the imaginary part of x. The occurrence of these discontinuous factors are due to the many-valuedness of the Ψ function. We obtain easily from (14) and (6)

$$\langle p_1| V |p_2\rangle = \frac{\Gamma(c_2)}{\Gamma(c_2 - b_2)} e^{-i\pi b_2\, \mathrm{srp}\,(p_2)} \langle p_1| V |p_2^+\rangle + \frac{\Gamma(c_2)}{\Gamma(b_2)} e^{i\pi(c_2 - b_2)\, \mathrm{srp}\,(p_2)} \langle -p_1| V |-p_2^+\rangle, \quad (16)$$

where srp (p_2) means the sign of the real part of p_2. We recall that in $\langle p_1| V |p_2^+\rangle$ we have found the branch cuts $p_1 = \pm(p_2 + is)(1 + \xi_1)$. Hence we have in $\langle -p_1| V |-p_2^+\rangle$ the cuts $p_1 = \pm(p_2 - is)(1 + \xi_2)$, which immediately gives us the analytic domain in Fig. 4.

We may as well, however, consider the matrix element as function of p_2 without changing the role of p_1 and p_2 in (16). As a function of p_2, the matrix element $\langle p_1| V |p_2^+\rangle$ was cut as in Fig. 2. The second term in (16) gives a corresponding cut in the upper half-plane. We know, however, that $\langle p_1| V |p_2\rangle$ is analytic in a strip $|\mathrm{Im}\, p_2| < s$ so that there remains T-shaped cuts in the lower and upper half-planes as is shown in Fig. 5. This is an alternate way of cutting the p planes.

We next investigate in detail the behavior of the various matrix elements at the end points of the branch cuts and assume throughout that $p_1 \approx \pm p_2$

[10] Bateman Manuscript Project, *Table of Integral Transforms*, A. Erdélyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, p. 224.

[11] Ref. 1, p. 224.

[12] Ref. 1, p. 259.

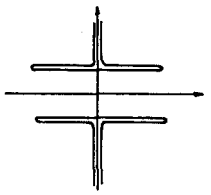


FIG. 5. An alternate way of cutting the p_1 or p_2 plane of $\langle p_1| V |p_2\rangle$.

and that $|p_1| \gg s$. Physically, this assumption corresponds to elastic scattering by long-ranged interactions or high-energy scattering with small energy transfers. In these cases a direct numerical calculation may be prohibitively difficult. Mathematically the assumption ensures that the arguments z_1 and z_2 of the Appell functions are both close to unity ($p_1 \approx p_2$) or infinity ($p_1 \approx -p_2$). These points are singular points of the functions.

Since all matrix elements encountered here may be expressed in terms of F_P functions it suffices to investigate this function near these points. However, the neighborhood of $|z_1| = |z_2| = \infty$ may be transformed into the neighborhood of $z_1 = z_2 = 1$ by using the result

$$\langle p_1| V |p_2^+\rangle = \langle -p_1| V |p_2^+\rangle.$$

Expressed by means of the F_P function this transformation reads

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = (1 - z_1)^{-a} \times F_P\left(a, c_1 - b_1, b_2, c_1, c_2, \frac{z_1}{z_1 - 1}, \frac{z_2}{1 - z_1}\right).$$

It suffices then to investigate the neighborhood of $p_1 = p_2 + is$, which gives $z_1 = 1$, $z_2 = p_2/(p_2 + is) \approx 1$.

The analytic continuation of the F_P function to this neighborhood has been derived recently,[13]

$$F_P(a, b_1, b_2, c_1, b_2, z_1, z_2) = C\left[u(z_1, z_2) + \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{u(tz_2)}{t - z_1}\, dt\right]. \quad (17)$$

Here the constant C is

$$C = \frac{\Gamma(a + b_2 - c_2 + 1)\Gamma(a + b_1 - c_1 - b_2)\Gamma(c_1)}{\Gamma(a)\Gamma(b_1)\Gamma(a - c_2 + 1)},$$

and $u(z_1, z_2)$ a function which can be expressed in terms of Appell's hypergeometric function[14] F_3

$$u(z_1, z_2) = z_1^{b_1 - c_1} z_2^{-b_2}(1 - z_1)^{c_1 - b_1 + b_2 - a} \times F_3\left(c_1 - b_1, b_2, 1 - b_1, b_2 - c_2 + 1, \times c_1 - b_1 + b_2 - a + 1, \frac{z_1 - 1}{z_1}, \frac{1 - z_1}{z_2}\right),$$

which function is defined by the double series

$$F_3(a_1, a_2, b_1, b_2, c, x_1, x_2) = \sum_{m,n=0}^{\infty} \frac{(a_1)_m(a_2)_n(b_1)_m(b_2)_n x_1^m x_2^n}{(c)_{m+n} m!\, n!}, \quad |x_1| < 1, \; |x_2| < 1.$$

It follows that $u(z_1, z_2)$ is regular in a neighborhood of $z_1 = z_2 = 1$ if a cut from $z_1 = 1$ to $z_1 = +\infty$ is made in order to remove the many valuedness of the factor $(1 - z_1)^{c_1 - b_1 + b_2 - a}$. In general this cut is necessary also when $c_1 - b_1 + b_2 - a$ is an integer since then a term $\ln(1 - z_1)$ appears in the F_P function and the formula (17) has to be modified. This so-called logarithmic case is not considered. It offers, it seems, no essential difficulties, and is in many cases best dealt with by disregarding it until the final result is obtained. We have in mind here perturbation theoretical applications not only to the first order.

The second function in (17) is, as we see below, regular at $z_1 = z_2 = 1$. If there is no Coulomb interaction the parameters η_1 and η_2 are zero and the dependence of p_1 and p_2 enters solely in the arguments z_1 and z_2. Then the right-hand side of (17) is an analytic function of p_1 or p_2 except on the cut $z_1 = 1 + \xi$, $\xi \geq 0$ of the factor $(1 - z_1)^{c_1 - b_1 + c_2 - a}$ which gives rise to the cuts in Figs. 1 and 4. The situation is not altered if the Coulomb interaction is included though the energy dependence then enters also via the parameters b_1 and b_2. The rather detailed and lengthy discussion that verifies this statement is omitted.

In order to obtain the explicit behavior of the F_P function at $z_1 \approx z_2 \approx 1$, we derive a series expansion of the second term in (17). The F_P function may be written

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2)_n(b_2 - c_2 + 1)_n}{(a + b_2 - c_2 + 1)_n n!}\left(\frac{z_2 - 1}{z_2}\right)^n \times {}_3F_2\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right), \quad (18)$$

where the ${}_3F_2$ function is defined by the series

$${}_3F_2\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right) = \sum_{m=0}^{\infty} \frac{(a)_m(b_1)_m(a - c_2 + 1)_m z_1^m}{(c_1)_m(a + b_2 - c_2 + n + 1)_m m!}, \quad |z_1| < 1,$$

and its analytic continuation.

The series (18) can be shown to converge when Re $z_2 > \frac{1}{2}$ for any z_1 except possibly $z_1 = 1$, for which

[13] P. O. M. Olsson, Ark. Fysik 29, 459 (1965).
[14] Ref. 1, p. 224.

value the F_P function, as we have seen, may be infinite. Replacing the ${}_3F_2$ functions in (18) by their analytic continuations to the neighborhood of[15] $z_1 = 1$,

$$
\begin{aligned}
&{}_3F_2\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right) \\
&\quad = F_I\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right) \\
&\qquad + F_R\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right),
\end{aligned}
$$

we obtain two series one of which can be identified with the function $u(z_1, z_2)$. The other series must then be an expansion of the integral in (17).

We have[16]

$$
\begin{aligned}
&F_I\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right) \\
&= C \cdot \frac{(a + b_2 - c_2 + 1)_n}{(c_1 - b_1 + b_2 - a + 1)_n} \\
&\quad \times z_1^{1-b_2-c_1}(1 - z_1)^{c_1-b_1+b_2-a}\left(\frac{z_1 - 1}{z_1}\right)^n \\
&\quad \times \sum_{m=0}^{\infty} \frac{(c_1 + c_2 - a - 1)_m (b_2 + n)_m}{(c_1 - b_1 + b_2 - a + n + 1)_m m!}\left(\frac{z_1 - 1}{z_1}\right)^m \\
&\quad \times {}_2F_1(1 - a, 1 - b_1, \\
&\qquad c_1 - b_1 + b_2 - a + m + n + 1, 1 - z_1), \\
&\qquad\qquad \operatorname{Re} z_1 > \tfrac{1}{2}.
\end{aligned}
$$

For large values of n the sum tends to

$$
\begin{aligned}
&\sum_{m=0} \frac{(c_1 + c_2 - a - 1)_m (b_2 + n)_m}{(c_1 - b_1 + b_2 - a + n + 1)_m m!}\left(\frac{z_1 - 1}{z_1}\right)^m \\
&\quad = {}_2F_1\left(c_1 + c_2 - a - 1, b_2 + n, \right. \\
&\qquad \left. c_1 - b_1 + b_2 - a + n + 1, \frac{z_1 - 1}{z_1}\right) \\
&\quad = z_1^{c_1+c_2-a-1}\, {}_2F_1(c_1 + c_2 - a - 1, c_1 - b_1 - a + 1, \\
&\qquad c_1 - b_1 + b_2 - a + n + 1, 1 - z_1), \\
&\qquad\qquad \operatorname{Re} z_1 > \tfrac{1}{2},
\end{aligned}
$$

since[17]

$$
{}_2F_1(a, b, c + n, z) \approx 1 + O\left(\frac{1}{n}\right), \quad z \neq 1.
$$

Hence also

$$
\begin{aligned}
&{}_2F_1(c_1 + c_2 - a - 1, c_1 - b_1 - a + 1, \\
&\qquad c_1 - b_1 + b_2 - a + n + 1, 1 - z_1) \\
&\qquad\qquad \approx 1 + O(1/n),
\end{aligned}
$$

and we obtain

$$
\begin{aligned}
&F_I\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right) \\
&\qquad \approx C \frac{(a + b_2 - c_2 + 1)_n}{(c_1 - b_1 + b_2 - a + 1)_n} \\
&\qquad\quad \times z_1^{c_2-b_2-a}(1 - z_1)^{c_1-b_1+b_2-a}\left(\frac{z_1 - 1}{z_1}\right)^n.
\end{aligned}
$$

Thus the series

$$
\begin{aligned}
&z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2)_n (b_2 - c_2 + 1)_n}{(a + b_2 - c_2 + 1)_n n!}\left(\frac{z_2 - 1}{z_2}\right)^n \\
&\qquad\qquad \times F_I\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right)
\end{aligned}
$$

converges for $\operatorname{Re} z_1 > \frac{1}{2}$ and $\operatorname{Re} z_2 > \frac{1}{2}$ and consequently the series

$$
\begin{aligned}
&z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2)_n (b_2 - c_2 + 1)_n}{(a + b_2 - c_2 + 1)_n n!}\left(\frac{z_2 - 1}{z_2}\right)^n \\
&\qquad\qquad \times F_R\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right)
\end{aligned}
$$

converges in the same domain since its terms are the differences between the terms of two convergent series. In the neighborhood of $z_1 = z_2 = 1$ the first series behaves as

$$
C(1 - z_1)^{c_1-b_1+b_2-a},
$$

and is thus identical with $Cu(z_1, z_2)$, since the second expansion is regular at the point since F_R is regular at $z_1 = 1$. We may now write

$$
\begin{aligned}
&F_{PR}(a, b_1, b_2, c_1, c_2, z_1, z_2) \\
&\quad = \frac{C}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{u(t, z_2)}{t - z_1}\, dt \\
&\quad = z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2)_n (b_2 - c_2 + 1)_n}{(a + b_2 - c_2 + 1)_n n!}\left(\frac{z_2 - 1}{z_2}\right)^n \\
&\qquad\qquad \times F_R\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right). \qquad (19)
\end{aligned}
$$

For the function F_R we have derived various expansions, e.g.,[18]

$$
\begin{aligned}
&F_R\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right) = \frac{\Gamma(c_1)\Gamma(a + b_2 - c_2 + n + 1)\Gamma(c_1 - b_1 + b_2 + a + n)}{\Gamma(a - c_2 + 1)\Gamma(c_1 + b_2 - a + n)\Gamma(c_1 - b_1 + b_2 + n)} \\
&\quad \times \sum_{m=0}^{\infty} \frac{(c_1 + c_2 - a - 1)_m (b_2 + n)_m (c_1 - b_1 + b_2 - a + n)_m}{(c_1 + b_2 - a + n)_m (c_1 - b_1 + b_2 + n)_m m!} \\
&\quad \times {}_2F_1(a, b_1, a + b_1 - c_1 - b_2 - n - m + 1, 1 - z_1), \\
&\qquad\qquad \operatorname{Re}(a - c_2 + 1) > 0, \quad \operatorname{Re} z_1 > \tfrac{1}{2}.
\end{aligned}
$$

[15] P. O. M. Olsson, J. Math. Phys. 7, 702 (1966).
[16] See Ref. 15.
[17] O. Perron, Sitz. Ber. Heidelberger Akad. Wiss. 9, 3 (1917).
[18] Ref. 15, p. 704.

This gives

$$F_{PR}(a, b_1, b_2, c_1, c_2, z_1, z_2) = \frac{\Gamma(c_1)\Gamma(a + b_2 - c_2 + 1)\Gamma(c_1 - b_1 + b_2 - a)}{\Gamma(a - c_2 + 1)\Gamma(c_1 + b_2 - a)\Gamma(c_1 + b_2 - b_1)} z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2 - c_2 + 1)_n}{n!} \left(\frac{z_2 - 1}{z_2}\right)^n$$
$$\times \sum_{m=0}^{\infty} \frac{(b_2)_{m+n}(c_1 + c_2 - a - 1)_m(c_1 - b_1 + b_2 - a)_{m+n}}{(c_1 + b_2 - a)_{m+n}(c_1 + b_2 - b_1)_{m+n} m!}$$
$$\times {}_2F_1(a, b_1, a + b_1 - c_1 - b_2 - m - n + 1, 1 - z_1),$$
$$\mathrm{Re}\,(a - c_2 + 1) > 0, \quad \mathrm{Re}\, z_1 > \tfrac{1}{2}, \quad \mathrm{Re}\, z_2 > \tfrac{1}{2}.$$

Putting $m + n = N$ and summing over N and n we obtain

$$F_{PR}(a, b_1, b_2, c_1, c_2, z_1, z_2) = \frac{\Gamma(c_1)\Gamma(a + b_2 - c_2 + 1)\Gamma(c_1 - b_1 + b_2 - a)}{\Gamma(a - c_2 + 1)\Gamma(c_1 + b_2 - a)\Gamma(c_1 + b_2 - b_1)} z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2 - c_2 + 1)_n}{n!} \left(\frac{z_2 - 1}{z_2}\right)^n$$
$$\times \sum_{m=0}^{\infty} \frac{(b_2)_N(c_1 - b_1 + b_2 - a)_N(c_1 + c_2 - a - 1)_{N-n}(-N)_n(-1)^n}{(c_1 + b_2 - a)_N(c_1 + b_2 - b_1)_N N!}$$
$$\times {}_2F_1(a, b_1, a + b_1 - c_1 - b_2 - N + 1, 1 - z_1)$$
$$= \frac{\Gamma(c_1)\Gamma(a + b_2 - c_2 + 1)\Gamma(c_1 - b_1 + b_2 - a)}{\Gamma(a - c_2 + 1)\Gamma(c_1 + b_2 - a)\Gamma(c_1 + b_2 - b_1)} z_2^{-b_2} \sum_{N=0}^{\infty} \frac{(b_2)_N(c_1 - b_1 + b_2 - a)_N(c_1 + c_2 - a - 1)_N}{(c_1 + b_2 - a)_N(c_1 + b_2 - b_1)_N N!}$$
$$\times {}_2F_1(a, b_1, a + b_1 - c_1 - b_2 - N + 1, 1 - z_1) \sum_{n=0}^{\infty} \frac{(-N)_n(b_2 - c_2 + 1)_n}{(a - c_1 - c_2 - N + 2)_n n!} \left(\frac{z_2 - 1}{z_2}\right)^n.$$

The last sum is the hypergeometric function

$${}_2F_1[-N, b_2 - c_2 + 1, a - c_1 - c_2 - N + 2, (z_2 - 1)/z_2].$$

Noting that

$$\frac{(c_1 + c_2 - a - 1)_N}{(c_1 + b_2 - a)_N} {}_2F_1\left(-N, b_2 - c_2 + 1, a - c_1 - c_2 - N + 2, \frac{z_2 - 1}{z_2}\right)$$
$$= {}_2F_1\left(-N, b_2 - c_2 + 1, c_1 + b_2 - a, \frac{1}{z_2}\right),$$

we finally obtain the single series

$$F_{PR}(a, b_1, b_2, c_1, c_2, z_1, z_2)$$
$$= \frac{\Gamma(c_1)\Gamma(a + b_2 - c_2 + 1)\Gamma(c_1 - b_1 + b_2 - a)}{\Gamma(a - c_2 + 1)\Gamma(c_1 + b_2 - a)\Gamma(c_1 + b_2 - b_1)}$$
$$\times z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2)_n(c_1 - b_1 + b_2 - a)_n}{(c_1 + b_2 - b_1)_n n!}$$
$$\times {}_2F_1(a, b_1, a + b_1 - c_1 - b_2 - n + 1, 1 - z_1)$$
$$\times {}_2F_1\left(-n, b_2 - c_2 + 1, c_1 + b_2 - a, \frac{1}{z_2}\right),$$
$$\mathrm{Re}\,(a - c_2 + 1) > 0, \quad \mathrm{Re}\, z_1 > \tfrac{1}{2}, \quad \mathrm{Re}\, z_2 > \tfrac{1}{2}. \tag{20}$$

Since the function is regular at $z_1 = z_2 = 1$, it may be expanded in a power series. Various power series expansions of the F_R function in (19), such as[19]

$$F_R\left(\begin{matrix} a, b_1, a - c_2 + 1, z_1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right)$$
$$= z_1^{-b_1} \sum_{m=0}^{\infty} \frac{(b_1)^m}{m!} \left(\frac{z_1 - 1}{z_1}\right)^m \times {}_3F_2\left(\begin{matrix} a, b_1 + m, a - c_2 + 1, 1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right),$$

may be used to accomplish this. We immediately obtain, using (19),

$$F_{PR}(a, b_1, b_2, c_1, c_2, z_1, z_2)$$
$$= z_1^{-b_1} z_2^{-b_2} \sum_{m,n=0}^{\infty} \frac{(b_1)_m(b_2)_n(b_2 - c_2 + 1)_n}{(a + b_2 - c_2 + 1)_n m!\, n!} \left(\frac{z_1 - 1}{z_1}\right)^m \left(\frac{z_2 - 1}{z_2}\right)^n$$
$$\times {}_3F_2\left(\begin{matrix} a, b_1 + m, a - c_2 + 1, 1 \\ c_1, a + b_2 - c_2 + n + 1 \end{matrix}\right). \tag{21}$$

In general, however, the single series (20) can be expected to be more convenient than the double series (21) in numerical applications. Its terms are simpler

[19] Ref. 15, p. 705.

to compute and are connected by simpler recursion relations.

Series containing products of ${}_2F_1$ functions in its terms exist for most Appell and associated Appell functions and seem, in fact, to be one of the best ways of representing them particularly from a numerical point of view.

We give an expansion of this type for the F_P function:

$$F_P(a, b_1, b_2, c_1, c_2, z_1, z_2) = \frac{\Gamma(a + b_2 - c_2 + 1)\Gamma(c_1)}{\Gamma(b_2 + c_1)\Gamma(a - c_2 + 1)} \times z_2^{-b_2} \sum_{n=0}^{\infty} \frac{(b_2)_n (c_1 + b_2 - a)_n}{(b_2 + c_1)_n n!} \times {}_2F_1(a, b_1, b_2 + c_1 + n, z_1) \times {}_2F_1\left(-n, b_2 - c_2 + 1, c_1 + b_2 - a, \frac{1}{z_2}\right). \quad (22)$$

The series is absolutely convergent when

$$\mathrm{Re}\,(a - c_2 + 1) > 0$$

and $\mathrm{Re}\, z_2 > \frac{1}{2}$ for all values of z_1 which can be verified by means of estimates of the ${}_2F_1$ functions for large values of n.[20] For $z_1 = 1$, however, a finite number of terms at the beginning of the series may be infinite, since[21]

$$\begin{aligned} &{}_2F_1(a, b_1, b_2 + c_1 + n, z_1) \\ &= \frac{\Gamma(b_2 + c_1 + n)\Gamma(c_1 - b_1 + b_2 - a + n)}{\Gamma(b_2 + c_1 - a + n)\Gamma(c_1 - b_1 + b_2 + n)} \\ &\quad \times {}_2F_1(a, b_1, a + b_1 - c_1 - b_2 - n + 1, 1 - z_1) \\ &\quad + \frac{\Gamma(b_2 + c_1 + n)\Gamma(a + b_1 - c_1 - b_2 - n)}{\Gamma(a)\Gamma(b_1)} \\ &\quad \times (1 - z_1)^{c_1 - b_1 + b_2 - a + n} \\ &\quad \times {}_2F_1(b_2 + c_1 - a + n, b_2 + c_1 - b_1 + n, \\ &\qquad c_1 - b_1 + b_2 - a + n + 1, 1 - z_1), \end{aligned}$$

where again we encounter the singular factor $(1 - z_1)^{c_1 - b_1 + b_2 - a}$. The proof of (22) is easily established by inserting the above expression in (22) which is an elementary way of deriving (17) in its explicit form. The knowledge of the ${}_3F_2$ function previously used in deriving it is no longer needed.

With this result we think that our investigation of the matrix elements considered here is fairly complete.

As far as we are aware, the singularities of the matrix elements have never been given explicitly though there has been a considerable amount of investigation in the field. For a selection of more recent investigations we refer the reader to a paper by Reynolds, Onley, and Biedenharn,[22] published when this work was at an early stage. These authors derive $\langle p|\, r^{1-L}\, |p\rangle$ which is essentially the Appell function F_2 of unit arguments.

ACKNOWLEDGMENT

The authors are indebted to Professor B. Ch. Nagel for a valuable suggestion.

[20] See Ref. 17.
[21] Ref. 1, p. 108.
[22] J. T. Reynolds, D. S. Onley, and L. C. Biedenharn, J. Math. Phys. 5, 411 (1964).

Evaluation of a Partial-Wave Unitarity Integral*

JAMES DILLEY
Department of Physics, Ohio University, Athens, Ohio

(Received 4 August 1966)

The integral $\int_{s_+}^{\infty}\rho(s)f(s)\,ds$, where $\rho(s) = (1/s)[(s-s_+)(s-s_-)]^{\frac{1}{2}}$, $0 \le s_- < s_+$, is evaluated for an arbitrary function $f(s)$ which is meromorphic in the complex s plane and has the behavior at infinity $|s^{1+\epsilon}f(s)| \xrightarrow[|s|\to\infty]{} 0$. Some applications to the N/D method of elementary particle physics are discussed.

IN the relativistic scattering theory of spinless particles, there arises integrals of the type

$$I = \int_{s_+}^{\infty} \rho(s)f(s)\,ds, \tag{1}$$

where

$$\rho(s) = (1/s)[(s-s_+)(s-s_-)]^{\frac{1}{2}}, \tag{2}$$

$$s_+ = (m_1+m_2)^2, \quad s_- = (m_1-m_2)^2, \tag{3}$$

and m_1, m_2 are the masses of the particles. It is our purpose here to evaluate (1) for an arbitrary function $f(s)$ which is meromorphic in the complex s plane with the behavior at infinity

$$|s^{1+\epsilon}f(s)| \xrightarrow[|s|\to\infty]{} 0. \tag{4}$$

In Eq. (2), $\rho(s)$ is positive for $s > s_+$, and can be analytically continued in s with cuts on the real axis $-\infty < s \le s_-$, $s_+ \le s < +\infty$. Taking the phases as indicated in Fig. 1, we have

$$(1/s)(\rho_+\rho_-)^{\frac{1}{2}}e^{\frac{1}{2}i(\phi_+ + \phi_-)}, \tag{5}$$

		$s+i\epsilon$	$s-i\epsilon$	$\Delta\rho \equiv \rho(s+i\epsilon)-\rho(s-i\epsilon)$
value of $\rho(s)$	$s > s_+$	$+\lvert\rho(s)\rvert$	$-\lvert\rho(s)\rvert$	$2\lvert\rho(s)\rvert$
	$s_- < s < s_+$	$+i\lvert\rho(s)\rvert$	$+i\lvert\rho(s)\rvert$	0
	$s < s_-$	$-\lvert\rho(s)\rvert$	$+\lvert\rho(s)\rvert$	$-2\lvert\rho(s)\rvert$.

For $s < s_-$, we have ignored the $1/s$ factor which changes sign for $s < 0$.

Now suppose that we have a function $F(s)$, suitably bounded at infinity, which has only the left-hand cut of $\rho(s)$, and the same discontinuity, i.e.,

$$\Delta F(s) \equiv F(s+i\epsilon) - F(s-i\epsilon) = \Delta\rho(s) = -2\lvert\rho(s)\rvert \tag{6}$$

so that $\rho(s) - F(s)$ has only the right-hand cut of ρ.

Then, integrating around the contour shown in Fig. 2, and discarding the terms on the semicircles,

$$\begin{aligned}\oint &[\rho(s)-F(s)]f(s)\,ds \\ &= \int_{s_+}^{\infty}[\rho(s+i\epsilon)-F(s)]f(s)\,ds - \int_{s_+}^{\infty}[\rho(s-i\epsilon)-F(s)]f(s)\,ds \\ &= \int_{s_+}^{\infty}\Delta\rho(s)f(s)\,ds \\ &= 2\int_{s_+}^{\infty}\rho(s)f(s)\,ds, \end{aligned}\tag{7}$$

which is the desired integral. We may therefore evaluate (1) by means of the Cauchy theorem

$$\int_{s_+}^{\infty}\rho(s)f(s)\,ds = \pi i \sum \text{residues}\,\{[\rho(s)-F(s)]f(s)\}, \tag{8}$$

where the sum is over the residues of the poles of $f(s)$ and the pole of $[\rho(s) - F(s)]$ at $s = 0$.

It is shown in Appendix A that the function

$$F(s) = -(1/\pi i)\tilde{\rho}(s)\ln g(s), \tag{9}$$

where

$$\tilde{\rho}(s) = (1/s)[(s-s_-)(s-s_+)]^{\frac{1}{2}}, \tag{10}$$

$$g(s) = \frac{(s-s_-)^{\frac{1}{2}} - (s-s_+)^{\frac{1}{2}}}{(s-s_-)^{\frac{1}{2}} + (s-s_+)^{\frac{1}{2}}} \tag{11}$$

satisfy the necessary conditions. It should be noted, as indicated in Appendix A, that $\rho(s)$ and $\tilde{\rho}(s)$ are not identical, since their branch cuts are taken in different ways. The function $\rho(s) - F(s)$ has a simple pole at

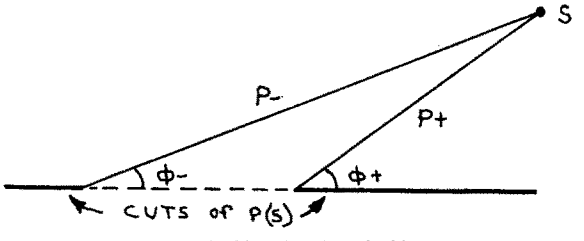


FIG. 1. Singularities of $\rho(s)$.

* Supported by U.S. Air Force Grant AF-AFOSR-198-67 and the Ohio University Fund.

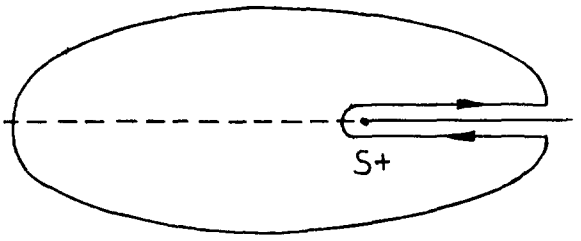


FIG. 2. Contour for evaluating Eq. (1).

$s = 0$ whose residue is

$$\lim_{s\to 0} s[\rho(s+i\epsilon) - F(s+i\epsilon)]$$
$$= -(s_+s_-)^{\frac{1}{2}} - \frac{1}{\pi i}(s_+s_-)^{\frac{1}{2}}\left[\ln\frac{s_+^{\frac{1}{2}} - s_-^{\frac{1}{2}}}{s_+^{\frac{1}{2}} + s_-^{\frac{1}{2}}} - i\pi\right]$$
$$= -\frac{1}{\pi i}(s_+s_-)^{\frac{1}{2}}\ln\frac{s_+^{\frac{1}{2}} - s_-^{\frac{1}{2}}}{s_+^{\frac{1}{2}} + s_-^{\frac{1}{2}}}. \quad (12)$$

We have, then, assuming that $f(s)$ has simple poles,

$$\int_{s_+}^{\infty}\rho(s)f(s)\,ds = -\left[(s_+s_-)^{\frac{1}{2}}\ln\frac{s_+^{\frac{1}{2}} - s_-^{\frac{1}{2}}}{s_+^{\frac{1}{2}} + s_-^{\frac{1}{2}}}\right] \times f(0) + \pi i\sum_j \beta_j[\rho(s_j) - F(s_j)], \quad (13)$$

where s_j are the positions of the poles of $f(s)$ and β_j the residues. For the equal-mass case $m_1 = m_2$, $s_- = 0$, and the first term on the right disappears. In this case, $\rho(s) = [(s-4m^2)/s]^{\frac{1}{2}}$ and there is no pole in $\rho(s)$.

The form of the other contributions from the poles of $f(s)$ depends upon the location of s_j. We write down several cases of particular interest..

1. $s_j > s_+$ [on top of the cut of $\rho(s)$]

Then $g(s_j)$, $\rho(s_j)$ and $\tilde{\rho}(s_j)$ are all real and positive,

$$\rho(s_j) - F(s_j) = \frac{1}{s_j}[(s_j - s_+)(s_j - s_-)]^{\frac{1}{2}} \times\left[1 + \frac{1}{\pi i}\ln\frac{(s_j - s_-)^{\frac{1}{2}} - (s_j - s_+)^{\frac{1}{2}}}{(s_j - s_-)^{\frac{1}{2}} + (s_j - s_+)^{\frac{1}{2}}}\right], \quad (14)$$

2. $s_- < s_j < s_+$.

In this case, $\rho(s_j)$ is positive imaginary and $F(s_j)$ has the form indicated in (A11),

$$\rho(s_j) - F(s_j) = i\frac{[(s_+ - s_j)(s_j - s_-)]^{\frac{1}{2}}}{s_j} \times\left[1 + \frac{1}{\pi}\arg\frac{(s_j - s_-)^{\frac{1}{2}} - (s_j - s_+)^{\frac{1}{2}}}{(s_j - s_-)^{\frac{1}{2}} + (s_j - s_+)^{\frac{1}{2}}}\right], \quad (15)$$

3. $s_j < s_-$

Both $\rho(s)$ and $F(s)$ have branch cuts in this region, but $\rho(s) - F(s)$ does not, so that there is no difficulty. If we look at $s_j + i\epsilon$ we see that $\rho(s)$ is negative and $F(s)$ has the form indicated in (A12),

$$\rho(s_j) - F(s_j) = \frac{-[(s_+ - s_j)(s_- - s_j)]^{\frac{1}{2}}}{s_j} - \frac{1}{\pi i}\frac{[(s_+ - s_j)(s_- - s_j)]^{\frac{1}{2}}}{s_j} \times\left[\ln\frac{(s_+ - s_j)^{\frac{1}{2}} - (s_- - s_j)^{\frac{1}{2}}}{(s_+ - s_j)^{\frac{1}{2}} + (s_- - s_j)^{\frac{1}{2}}} - i\pi\right]$$
$$= -\frac{1}{\pi i}\frac{[(s_+ - s_j)(s_- - s_j)]^{\frac{1}{2}}}{s_j} \times\ln\frac{(s_+ - s_j)^{\frac{1}{2}} - (s_- - s_j)^{\frac{1}{2}}}{(s_+ - s_j)^{\frac{1}{2}} + (s_- - s_j)^{\frac{1}{2}}}. \quad (16)$$

It should be noted that (15) and (16) can also be obtained directly from (14) by analytic continuation, and that (14) can actually be continued to arbitrary complex s_j.

We close by discussing some applications in the N/D method for spinless particles. Here, the scattering amplitude $A_l(s)$ is written as $A_l(s) = N_l(s)/D_l(s)$, where $D_l(s)$ has only the unitarity cut $(m_1 + m_2)^2 \leq s < \infty$ of $A_l(s)$, and $N_l(s)$ has the unphysical cuts L of $A_l(s)$.[1] Then, neglecting inelastic effects, the equations for $N_l(s)$ and $D_l(s)$ can be written as

$$N_l(s) = \frac{1}{\pi}\int_L \frac{ds' D_l(s')\,\mathrm{Im}\,A_l(s')}{s' - s}, \quad (17)$$

$$D_l(s) = 1 - \frac{1}{\pi}\int_{(m_1+m_2)^2}^{\infty}\frac{ds'\rho(s')N_l(s')}{s' - s}. \quad (18)$$

We have taken the subtraction of $D_l(s)$ at infinity for simplicity, but this restriction is in no way essential.

A. Effective Range Formulas[2]

The essential approximation is to replace the unphysical cuts of $A_l(s)$ by poles, or

$$\mathrm{Im}\,A_l(s) = \pi\sum_i \alpha_i\delta(s + a_i^2), \quad (19)$$

$$N_l(s) = -\sum_i \frac{\alpha_i D_l(-a_i^2)}{s + a_i^2} = -\sum_i \frac{\lambda_i}{s + a_i^2}, \quad (20)$$

$$\lambda_i = \alpha_i D_l(-a_i^2). \quad (21)$$

Then

$$D_l(s) = 1 + \sum_i \lambda_i\frac{1}{\pi}\int_{(m_1+m_2)^2}^{\infty}\frac{ds'\rho(s')}{(s' - s)(s' + a_i^2)}. \quad (22)$$

[1] G. F. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960).
[2] L. A. P. Balazs, Phys. Rev. **128**, 1939 (1962).

The integrals can be evaluated for all s and a_i^2 by the above methods, λ_i being determined by setting $s = -a_i^2$. The model is then exactly soluble, $A_l(s)$ being given in terms of the input parameters α_i, a_i^2.

We write down the integral for real $a_i^2 > 0$ and $s > (m_1 + m_2)^2$,

$$\int_{(m_1+m_2)^2}^{\infty} \frac{ds'\rho(s')}{(s'-s)(s'+a_i^2)}$$
$$= \frac{(s_+s_-)^{\frac{1}{2}}}{sa_i^2} \ln \frac{s_+^{\frac{1}{2}} - s_-^{\frac{1}{2}}}{s_+^{\frac{1}{2}} + s_-^{\frac{1}{2}}} + \frac{1}{s+a_i^2}\Bigg\{ - \frac{[(s_+ + a_i^2)(s_- + a_i^2)]^{\frac{1}{2}}}{a_i^2}$$
$$\times \ln \frac{(s_+ + a_i^2)^{\frac{1}{2}} - (s_- + a_i^2)^{\frac{1}{2}}}{(s_+ + a_i^2)^{\frac{1}{2}} + (s_+ + a_i^2)^{\frac{1}{2}}} + \frac{[(s-s_+)(s-s_-)]^{\frac{1}{2}}}{s}$$
$$\times \ln \frac{(s-s_-)^{\frac{1}{2}} - (s-s_+)^{\frac{1}{2}}}{(s-s_-)^{\frac{1}{2}} + (s-s_+)^{\frac{1}{2}}} + \frac{[(s-s_+)(s-s_-)]^{\frac{1}{2}}}{s}\pi i\Bigg\}, \tag{23}$$

with other cases being obtainable by analytic continuation.

B. Pagels's Approximation[3]

Pagels has given an approximate solution for the N/D equations, completely soluble algebraically for arbitrary input. The essential approximation is that on the unphysical cuts of $A_l(s)$, the function

$$\frac{1}{s} G(s) = \frac{1}{\pi} \int_{(m_1+m_2)^2}^{\infty} \frac{ds'\rho(s')}{(s')^2(s'-s)} \tag{24}$$

can be approximated by poles on the real axis, $a_j > (m_1 + m_2)^2$,

$$\frac{1}{s} G(s) = \sum_j \frac{c_j}{s - a_j}. \tag{25}$$

In evaluating (24), there is the slight complication of having the double pole at $s' = 0$ coincide with the simple pole of $[\rho(s) - F(s)]$ at the same point. We can then write for $s < (m_1 - m_2)^2$

$$\frac{1}{s} G(s) = - \frac{[(s_+ - s)(s_- - s)]^{\frac{1}{2}}}{s^3}$$
$$\times \ln \frac{(s_+ - s)^{\frac{1}{2}} - (s_- - s)^{\frac{1}{2}}}{(s_+ - s)^{\frac{1}{2}} + (s_- - s)^{\frac{1}{2}}}$$
$$- \frac{1}{2} \frac{d^2}{ds'^2} \Bigg\{ \frac{[(s_+ - s')(s_- - s')]^{\frac{1}{2}}}{s' - s}$$
$$\times \ln \frac{(s_+ - s')^{\frac{1}{2}} - (s_- - s')^{\frac{1}{2}}}{(s_+ - s')^{\frac{1}{2}} + (s - s')^{\frac{1}{2}}} \Bigg\}_{s'=0}. \tag{26}$$

[3] H. Pagels, Phys. Rev. **140**, B1599 (1965).

C. Integral Equation for $D_l(s)$

If one substitutes (17) into (18), then an integral equation for $D_l(s)$ is obtained,

$$D_l(s) = 1 + \frac{1}{\pi} \int_L ds'' D_l(s'') \operatorname{Im} A_l(s'') \frac{1}{\pi}$$
$$\times \int_{(m_1+m_2)^2}^{\infty} \frac{ds'\rho(s')}{(s'-s)(s'-s'')}. \tag{27}$$

The integral has exactly the same form as (23),

$$\int_{(m_1+m_2)^2}^{\infty} \frac{ds'\rho(s')}{(s'-s)(s'-s'')}$$
$$= - \frac{(s_+s_-)^{\frac{1}{2}}}{ss''} \ln \frac{s_+^{\frac{1}{2}} - s_-^{\frac{1}{2}}}{s_+^{\frac{1}{2}} + s_-^{\frac{1}{2}}} + \frac{1}{s-s''}$$
$$\times \Bigg\{ - \frac{[(s_+ - s'')(s_- - s'')]^{\frac{1}{2}}}{s''} \ln \frac{(s_+ - s'')^{\frac{1}{2}} - (s_- - s'')^{\frac{1}{2}}}{(s_+ - s'')^{\frac{1}{2}} + (s_- - s'')^{\frac{1}{2}}}$$
$$+ \frac{[(s-s_+)(s-s')]^{\frac{1}{2}}}{s} \ln \frac{(s-s_-)^{\frac{1}{2}} - (s-s_+)^{\frac{1}{2}}}{(s-s_-)^{\frac{1}{2}} + (s-s_+)^{\frac{1}{2}}}$$
$$+ \frac{[(s-s_+)(s-s_-)]^{\frac{1}{2}}}{s}\pi i \Bigg\}. \tag{28}$$

Again, this particular form has been written for real $s'' < (m_1 - m_2)^2$, $s > (m_1 + m_2)^2$, other cases being obtainable by continuation.

D. Inelasticity

One method of treating inelastic effects phenomenologically is to introduce a function $R_i(s) = \sigma$ total/σ elastic into (18) so that it becomes[4]

$$D_i(s) = 1 - \frac{1}{\pi} \int_{(m_1+m_2)^2}^{\infty} \frac{ds'\rho(s')R_i(s')N_i(s')}{s'-s}. \tag{29}$$

Then all of the integrals considered here can be formally evaluated in the same manner if one assumes that $R_i(s')$ can be analytically continued into the entire s' plane. We have no idea what such a procedure might mean and merely want to point out that the possibility exists.

APPENDIX A

Consider the function

$$F(s) = -(1/\pi i)\tilde{\rho}(s) \ln g(s), \tag{A1}$$

where

$$\tilde{\rho}(s) = \frac{1}{s}[(s-s_-)(s-s_+)]^{\frac{1}{2}}, \tag{A2}$$

$$g(s) = \frac{(s-s_-)^{\frac{1}{2}} - (s-s_+)^{\frac{1}{2}}}{(s-s_-)^{\frac{1}{2}} + (s-s_+)^{\frac{1}{2}}}. \tag{A3}$$

[4] M. Froissart, Nuovo Cimento **22**, 191 (1961).

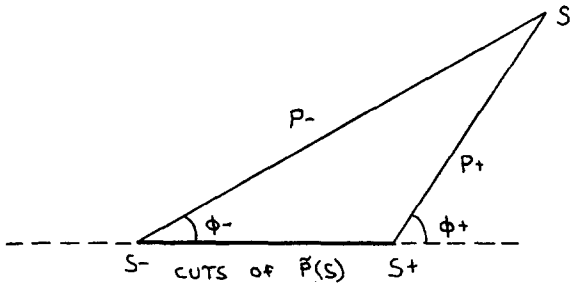


FIG. 3. Singularities of $\tilde{\rho}(s)$.

We take the cut of $\tilde{\rho}(s)$ on the real axis $s_- < s < s_+$ with the phase as shown in Fig. 3,

$$\tilde{\rho}(s)\begin{cases} s > s_+ & |\tilde{\rho}(s)| \\ s_- < s < s_+ & i\,|\tilde{\rho}(s)| \\ s < s_- & -|\tilde{\rho}(s)| \end{cases}\quad\begin{matrix} s+i\epsilon & s-i\epsilon \\ |\tilde{\rho}(s)| & \\ -i\,|\tilde{\rho}(s)| & \\ |\tilde{\rho}(s)|, & \end{matrix} \tag{A4}$$

where we have used $\tilde{\rho}(s) = (\rho_+\rho_-)^{\frac{1}{2}}e^{i/2(\phi_+ + \phi_-)}$. As in the case of $\rho(s)$, the last line does not include the change of sign due to $1/s$ for $s < 0$.

The function $g(s)$ also has branch points at $s_\pm$ and the cuts are taken to the left on the real axis. However, we need to keep track of the phase of $g(s)$ because of the logarithm, Writing

$$g(s) = \frac{\rho_-^{\frac{1}{2}}e^{\frac{1}{2}i\phi_-} - \rho_+^{\frac{1}{2}}e^{\frac{1}{2}i\phi_+}}{\rho_-^{\frac{1}{2}}e^{\frac{1}{2}i\phi_-} + \rho_+^{\frac{1}{2}}e^{\frac{1}{2}i\phi_+}} = |g|\,e^{i\phi}, \tag{A5}$$

it is seen that the phase is given by

$$\phi = \arg g\begin{cases} s > s_+ & 0 \\ s_- < s < s_+ & -\pi < \phi < 0 \\ s < s_- & -\pi \end{cases}\quad\begin{matrix} s+i\epsilon & s-i\epsilon \\ 0 & 0 \\ 0 < \phi < \pi & \\ \pi. & \end{matrix} \tag{A6}$$

In particular, for $s_- < s < s_+$ it is seen that $g(s)$ has the special properties

$$\arg g(s + i\epsilon) = -\arg g(s - i\epsilon), \tag{A7}$$

$$|g| = 1. \tag{A8}$$

Finally, there are logarithmic branch points where $g(s) = 0$ and $g(s) = \infty$. The first of these is as $s = \infty$. The second comes from the vanishing of

$$[(s - s_+)^{\frac{1}{2}} + (s - s_-)^{\frac{1}{2}}],$$

and is reached by passing through one of the square-root branch cuts to the second Riemann sheet and going to $s = \infty$. We can therefore take this logarithmic cut $-\infty < s < s_+$ so that, writing

$$\ln g(s) = \ln|g(s)| + i\arg g(s), \tag{A9}$$

we have

$$\ln g(s)\begin{cases} s > s_+ & \ln|g(s)| \\ s_- < s < s_+ & i\arg g(s+i\epsilon) \\ s < s_- & \ln|g(s)| - i\pi \end{cases}\quad\begin{matrix} s+i\epsilon & s-i\epsilon \\ \ln|g(s)| & \\ -i\arg g(s+i\epsilon) & \\ \ln|g(s)| + i\pi. & \end{matrix} \tag{A10}$$

It then follows from (A4) and (A10) that for $s_- < s < s_+$,

$$\begin{aligned}\Delta F(s) &\equiv F(s + i\epsilon) - F(s - i\epsilon) \\ &= -(\pi i)^{-1}\{[i\,|\rho(s)|][i\arg g(s + i\epsilon)] \\ &\qquad - [-i\,|\rho(s)|][-i\arg g(s + i\epsilon)]\} \\ &= 0,\end{aligned} \tag{A11}$$

and for $s < s$,

$$\begin{aligned}\Delta F(s) &= -(\pi i)^{-1}\{-|\tilde{\rho}(s)|\,[\ln|g(s)| - i\pi] \\ &\qquad + |\tilde{\rho}(s)|[\ln|g(s)| + i\pi]\} \\ &= -2\,|\rho(s)|\end{aligned} \tag{A12}$$

so that $F(s)$ is the required function.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Upper Bounds on Holomorphy Envelopes for Wightman Functions

A. C. MANOHARAN
Department of Physics, University of Wisconsin—Milwaukee, Wisconsin

(Received 3 March 1967)

As a consequence of Lorentz invariance, reasonable mass spectrum and local commutativity, the holomorphy envelopes $H(\mathfrak{S}_n^P)$ for the analytic functions arising from vacuum expectation values of products of n field operators are related in such a way that $H(\mathfrak{S}_n^P)$ furnishes upper bounds (in the coordinates) for all $H(\mathfrak{S}_N)$ with $N > n$.

1. INTRODUCTION

IN recent years, it has been found instructive to study some of the problems of relativistic quantum field theory in a general manner. Useful tools for such investigations have been the analytic functions, namely the Wightman functions, which have vacuum expectation values of field operators as their boundary values.[1,2] It is known that a quantum theory of fields can be reformulated in terms of a denumerable set of such analytic functions.[3,4] One would thus like to know the nature of the holomorphy domains for these functions. To date they have proved to be very difficult to determine.[5,6] We show in this note that one can make a general description, of how the holomorphy envelopes for functions of different orders are related.[7] It is hoped that this description will help towards solving problems of analytic completion. At the same time, this will set up a connection between such questions that are of interest to the physicist and the classical theory of several complex variables.[8,9]

The physical assumptions used are Lorentz invariance, reasonable mass spectrum, and local commutativity, but not the positive definiteness condition.

Section 2 summarizes the known facts that we require about Wightman functions. Section 3 deals with upper bounds for permuted extended "tubes." The discussion of domains over $\mathbb{C}^m$ begun in Sec. 4 is continued in the Appendix. Upper bounds for holomorphy envelopes are obtained in Sec. 5 and discussed in Sec. 6.

2. PRELIMINARIES

Our interest begins with the vacuum expectation values

$$\mathcal{W}_n(x_1, x_2, \cdots x_n) = \langle 0 | \phi_1(x_1) \cdots \phi_n(x_n) | 0 \rangle$$

of the products of n field operators transforming according to irreducible representations of the inhomogeneous $SL(2, C)$ group. (This is referred to as inhomogeneous Lorentz invariance.) N points in $(3 + 1)$ dimensional spacetime are shown as $x_1, \cdots x_n$. By the condition of a reasonable mass spectrum, which means it lies inside the closed forward light cone $\bar{V}_+$, these distributions $\mathcal{W}_n$ are boundary values of functions $W_n(z_1, \cdots z_n)$, holomorphic in the domain

$$\mathfrak{S}_n(z_1, \cdots z_n) = \{(z_1, \cdots z_n) \mid -\mathrm{Im}\,(z_i - z_{i+1}) \in V_+, \quad i = 1, \cdots n-1\}. \quad (2.1)$$

We notice that

$$\mathfrak{S}_n(z_1 + a, \cdots z_n + a) = \mathfrak{S}_n(z_1, \cdots z_n)$$

for all complex translations a. As a consequence of Lorentz invariance, $W_n(z_1, \cdots z_n)$ has a single valued continuation[1,10] into a larger (*extended*) domain

$$\mathfrak{S}'_n(z_1, \cdots z_n) = \bigcup_{\Lambda \in \mathfrak{L}_+(C)} \Lambda \mathfrak{S}_n(z_1, \cdots z_n) \quad (2.2)$$

where

$$\Lambda \mathfrak{S}_n(z_1, \cdots z_n) = \{(z'_1, \cdots z'_n) \mid z'_i = \Lambda z_i, \quad i = 1, \cdots n, \quad (z_1, \cdots z_n) \in \mathfrak{S}_n\}$$

and $\mathfrak{L}_+(C)$ is the identity component of the complex Lorentz group, i.e., Λ is any 4×4 matrix satisfying $\tilde{\Lambda} g \Lambda = g$, $\det \Lambda = 1$, where $g = \mathrm{diag}\ (1, -1, -1, -1)$. Thus $\mathfrak{S}'_n$ is invariant both under complex

[1] R. F. Streater and A. S. Wightman, *PCT Spin and Statistics and All That* (W. A. Benjamin, Inc., New York, 1964), and further references there.
[2] R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., 1965), and further references there.
[3] A. S. Wightman, Phys. Rev. **101**, 860 (1956).
[4] This reconstruction theorem can be extended to cover arbitrary spinor fields—A. S. Wightman (unpublished).
[5] The simplest nontrivial one is known, Ref. 6.
[6] G. Källén and A. S. Wightman, Mat. Fys. Skr. Dan. Vid. Selsk. **1**, No. 6 (1958).
[7] A. C. Manoharan, Bull. Am. Phys. Soc. **12**, No. 1, 116 (1967).
[8] H. Behnke and P. Thullen, *Theorie der Functionen Mehrerer Komplexer Veränderlichen* (Julius Springer-Verlag, Berlin, 1934), (reprinted by Chelsea).
[9] A. S. Wightman, in *Dispersion Relations and Elementary Particles* (John Wiley & Sons, Inc., New York, 1960), p. 229.
[10] D. Hall and A. S. Wightman, Kgl. Danske Videnskab. Selskab, Mat. fys. Medd. **31**, No. 5 (1957).

Lorentz transformations as well as complex translations. As a consequence of local commutativity, $W_n(z_1, \cdots z_n)$ has a single-valued analytic continuation[11,12] into the union of permuted extended domains

$$\mathfrak{S}_n^P = \bigcup_{\sigma \in \mathfrak{P}_n} (\sigma \mathfrak{S}_n'),$$

where

$$\mathfrak{S}_n^P = \mathfrak{S}_n^P(z_1, \cdots z_n),$$
$$\sigma \mathfrak{S}_n'(z_1, \cdots z_n) = \mathfrak{S}_n'(z_{\sigma(1)}, \cdots z_{\sigma(n)})$$

and σ denotes a permutation of the group $\mathfrak{P}_n$ on n symbols. $\mathfrak{S}_n^P$ is schlicht.

If $n > 2$, $\mathfrak{S}_n^P$ is not a natural domain of holomorphy and every function holomorphic in $\mathfrak{S}_n^P$ is also holomorphic in a larger domain. One wishes to determine the largest domain into which all functions holomorphic in $\mathfrak{S}_n^P$ may be analytically continued. This is the holomorphy envelope $H(\mathfrak{S}_n^P)$ of $\mathfrak{S}_n^P$. We refer to $H(\mathfrak{S}_n^P)$ as the holomorphy envelope for the n point function.

3. UPPER BOUNDS FOR UNIONS OF PERMUTED EXTENDED DOMAINS

Let us use the abbreviations

$$\mathbb{C}^{4r} = \mathbb{C}^{4r}(z_1, \cdots z_r), \quad \mathfrak{S}_{n-r}^P = \mathfrak{S}_{n-r}^P(z_{r+1}, \cdots z_n),$$

and let $\times$ denote the topological product. Thus, $\mathbb{C}^{4r}$ is the space of $4r$ complex variables $z_1, \cdots z_r$.

Lemma 1:

$$\mathfrak{S}_n^P \subset (\sigma \mathfrak{S}_{n-r}^P) \times (\sigma \mathbb{C}^{4r}), \quad \text{all} \quad \sigma \in \mathfrak{P}_n, \quad 0 < r < n. \tag{3.1}$$

Proof: We pick out r indices $\rho(1), \cdots \rho(r)$, in order, from the set $(1, 2, \cdots n)$ and let the remaining indices, in order, be denoted $\rho(r+1), \cdots \rho(n)$. Then from the definition (2.1) of the domain $\mathfrak{S}_n$, we see from the convexity of V_+ that

$$\mathfrak{S}_n(z_1, \cdots z_n) \subset \mathfrak{S}_{n-r}(z_{\rho(r+1)}, \cdots z_{\rho(n)}) \times \mathbb{C}^{4r}(z_{\rho(1)}, \cdots z_{\rho(r)}), \tag{3.2}$$

because,

$$-\mathrm{Im}\,(z_j - z_{j+1}) \in V_+, j = 1, \cdots n-1$$
$$\Rightarrow -\mathrm{Im}\,(z_{\rho(j)} - z_{\rho(j+1)}) \in V_+, j = r+1, \cdots n-1.$$

Statement (3.2) may also be written as

$$\mathfrak{S}_n \subset (\rho \mathfrak{S}_{n-r}) \times (\rho \mathbb{C}^{4r}). \tag{3.2}$$

Next, from the definition (2.2) of the extended domain $\mathfrak{S}_n'$ and from (3.2), we have

$$\mathfrak{S}_n' \subset \bigcup_{\Lambda \in \mathfrak{L}_+(C)} \Lambda[(\rho \mathfrak{S}_{n-r}) \times (\rho \mathbb{C}^{4r})]$$
$$= \bigcup_{\Lambda \in \mathfrak{L}_+(C)} [(\Lambda \rho \mathfrak{S}_{n-r}) \times (\rho \mathbb{C}^{4r})]$$
$$= \left(\bigcup_{\Lambda \in \mathfrak{L}_+(C)} \Lambda \rho \mathfrak{S}_{n-r} \right) \times (\rho \mathbb{C}^{4r})$$
$$= (\rho \mathfrak{S}_{n-r}') \times (\rho \mathbb{C}^{4r})$$

or

$$\mathfrak{S}_n' \subset (\rho \mathfrak{S}_{n-r}') \times (\rho \mathbb{C}^{4r}). \tag{3.3}$$

Let $\mathfrak{Q}_n$ denote the set of permutations of $(1, \cdots n)$ such that $\rho(r+1), \cdots \rho(n)$ remain in the same order but $\rho(1), \cdots \rho(r)$ are placed between or outside them in all possible ways. Then from (3.3),

$$\bigcup_{\tau \in \mathfrak{Q}_n} (\tau \mathfrak{S}_n') \subset (\rho \mathfrak{S}_{n-r}') \times (\rho \mathbb{C}^{4r}). \tag{3.4}$$

Let $\mathfrak{R}_n$ denote the set of permutations of $(1, \cdots n)$, such that the positions of only $\rho(r+1), \cdots \rho(n)$ in $(1, \cdots n)$ are permuted among themselves. Then,

$$\bigcup_{\mu \in \mathfrak{R}_n} \bigcup_{\tau \in \mathfrak{Q}_n} (\mu \tau \mathfrak{S}_n') \subset \bigcup_{\mu \in \mathfrak{R}_n} [(\mu \rho \mathfrak{S}_{n-r}') \times (\rho \mathbb{C}^{4r})]. \tag{3.5}$$

Since $\mathfrak{R}_n \mathfrak{Q}_n = \mathfrak{P}_n$, we can write (3.5) as

$$\mathfrak{S}_n^P \subset (\rho \mathfrak{S}_{n-r}^P) \times (\rho \mathbb{C}^{4r}). \tag{3.6}$$

This is true for all ρ. But disordering the indices $\rho(1), \cdots \rho(r)$ and/or the indices $\rho(r+1), \cdots \rho(n)$ does not change the right-hand side of (3.6). The result (3.1) follows from this.

We also have

Corollary:

$$\mathfrak{S}_n^P \subset \bigcap_{\sigma \in \mathfrak{P}_n} (\sigma \mathfrak{S}_{n-r}^P) \times (\sigma \mathbb{C}^{4r}). \tag{3.7}$$

4. DOMAINS OVER $\mathbb{C}^m$

We are concerned in the following with a certain generalization[8,9,13] to several complex variables of the concept of a Riemann surface for one complex variable. This will be within the framework of the classical theory of several complex variables. We shall consider a many sheeted object called a *domain over* $\mathbb{C}^m$ which does not include ramification points, whether uniformizable or not. Our restriction here to such a concept of a *Riemann surface* appears to be justified for the following reason. We deal in this note with Wightman functions having the original Lorentz vector variables as arguments and so begin with schlicht domains *in* $\mathbb{C}^m$ without ramification points and consider analytic continuations from such domains. To answer certain types of questions one

[11] D. Ruelle, see Ref. 2, p. 150.
[12] Y. Tomozawa, J. Math. Phys. 4, 1240 (1963).
[13] B. A. Fuks, *Introduction to the Theory of Analytic Functions of Several Complex Variables*, Am. Math. Soc. transl. (1963), Chap. II.

makes a holomorphic mapping from this space to the space of invariants, and one has to consider[10,14] a more refined notion of analyticity on an algebraic variety in place of analyticity on a manifold. By not going over to invariant space, and working in the flat space of vectors, we avoid at least some of these complications. A brief summary of definitions needed, regarding domains over $\mathbb{C}^m$, is given in the Appendix, which is a continuation of this section but may be skipped by the reader who is familiar with the notation used in Sec. 5.

5. UPPER BOUNDS FOR HOLOMORPHY ENVELOPES

As a consequence of the definition of $H(\mathfrak{D})$, we have:

Lemma 2:[13] If $\mathfrak{D}_1$, $\mathfrak{D}_2$ are domains over $\mathbb{C}^m$ and $\mathfrak{D}_1 < \mathfrak{D}_2$, then

$$H(\mathfrak{D}_1) < H(\mathfrak{D}_2). \tag{5.1}$$

This is because the functions holomorphic in $\mathfrak{D}_2$ are also holomorphic in $\mathfrak{D}_1$, but not necessarily vice versa.

Lemma 3:[15,16] Let $\mathfrak{D} = \mathfrak{D}(z_1, \cdots z_l)$ be a domain over $\mathbb{C}^l$, and $\mathfrak{E} = \mathfrak{E}(w_1, \cdots w_m)$ be a domain over $\mathbb{C}^m$. Then

$$H(\mathfrak{D} \times \mathfrak{E}) < H(\mathfrak{D}) \times H(\mathfrak{E}). \tag{5.2}$$

Proof: If S is the set of functions $f(z, w)$ holomorphic in $\mathfrak{D} \times \mathfrak{E}$, then

$$H(\mathfrak{D} \times \mathfrak{E}) = \bigcap_{f \in S} (\mathfrak{D} \times \mathfrak{E})_f. \tag{5.3}$$

Let S_1 be the set of functions $f_1(z, w)$ holomorphic in $\mathfrak{D} \times \mathfrak{E}$ which are such that $f_1(z, w) = g(z)h(w)$ where $g(z)$ has domain of holomorphy $H(\mathfrak{D})$ and $h(w)$ has domain of holomorphy $H(\mathfrak{E})$. Then $f_1(z, w)$ cannot be analytically continued into a domain $> H(\mathfrak{D}) \times H(\mathfrak{E})$. Thus,

$$\bigcap_{f_1 \in S_1} (\mathfrak{D} \times \mathfrak{E})_{f_1} < H(\mathfrak{D}) \times H(\mathfrak{E}). \tag{5.4}$$

On the other hand $S_1 \subset S$, and so we have

$$\bigcap_{f \in S} (\mathfrak{D} \times \mathfrak{E})_f < \bigcap_{f_1 \in S_1} (\mathfrak{D} \times \mathfrak{E})_{f_1}. \tag{5.5}$$

Combining statements (5.3), (5.4), and (5.5), we get the result (5.2).

Remark: By using Hartog's fundamental theorem on simultaneous holomorphy in several complex variables, one may show that[13] in fact

$$H(\mathfrak{D} \times \mathfrak{E}) = H(\mathfrak{D}) \times H(\mathfrak{E}).$$

Theorem:

$$H[\mathfrak{S}_n^P(z_1, \cdots z_n)] < \bigcap_{\sigma \in \mathfrak{S}_n} \{H[\mathfrak{S}_{n-r}^P(z_{\sigma(r+1)}, \cdots z_{\sigma(n)})] \times \mathbb{C}^{4r}(z_{\sigma(1)}, \cdots z_{\sigma(r)})\}, \quad 0 < r < n, \text{ relative to } H(\mathfrak{S}_n^P). \tag{5.6}$$

Proof: From Lemma 1 and Lemma 2, we have

$$H(\mathfrak{S}_n^P) < H[(\sigma\mathfrak{S}_{n-r}^P) \times (\sigma\mathbb{C}^{4r})] \quad \text{all} \quad \sigma \in \mathfrak{S}_n. \tag{5.7}$$

Using Lemma 3 we obtain

$$H(\mathfrak{S}_n^P) < H(\sigma\mathfrak{S}_{n-r}^P) \times (\sigma\mathbb{C}^{4r}) \quad \text{all} \quad \sigma \in \mathfrak{S}_n \tag{5.8}$$

and thus

$$H(\mathfrak{S}_n^P) < \bigcap_{\sigma \in \mathfrak{S}_n} \{H(\sigma\mathfrak{S}_{n-r}^P) \times (\sigma\mathbb{C}^{4r})\}, \quad \text{relative to} \quad H(\mathfrak{S}_n^P),$$

which is (5.6). Although the intersection is relative to $H(\mathfrak{S}_n^P)$, it is not necessary to know the latter to construct the intersection. Any point of $\mathfrak{S}_n^P$ is sufficient to do this.

6. DISCUSSION

If it turns out that the holomorphy envelopes involved are schlicht, then we can replace the $<$ by the set theoretic inclusion $\subset$ in (5.6).

In general, the projection (the image under the covering map) of the left-hand side of (5.6) must be contained in that of the right-hand side. For example, putting $r = n - 3$, the right-hand side is the intersection of permuted cylinders whose bases are the known[6,17] three-point function holomorphy envelopes in the appropriate variables. Thus, we have upper bounds, in the coordinates, for the holomorphy envelope of the $n > 3$ point function in terms of that of the three-point function. This is true within the so-called "linear program" i.e., before applying the positive definiteness condition in Hilbert space.

ACKNOWLEDGMENT

I thank Professor A. S. Wightman for his interest, and Professor E. Lubkin for commenting on the understandability of the manuscript. Some examples concerning this work were studied at the University of Wisconsin Computer Centers at Madison and Milwaukee and supported in part by the Research Committee.

APPENDIX

Domains Over $\mathbb{C}^m$ (Continued)

A detailed treatment of the theory is available elsewhere.[8,9,13] We only give a brief summary of the ideas. A *domain* $\mathfrak{D}$ *over* $\mathbb{C}^m$ consists of a connected

[14] K. Hepp, Helv. Phys. Acta **36**, 355 (1963); Math. Ann. **152**, 149 (1963).

[15] See Ref. 13. The definition of the topological product of domains over $\mathbb{C}^l$, $\mathbb{C}^m$ is also given here.

[16] F. Sommer and J. Mehring, Math. Ann. **131**, 1 (1956).

[17] D. Ruelle; Helv. Phys. Acta. **34**, 587 (1961).

Hausdorff topological space which is locally homeomorphic to an open subset of $\mathbb{C}^m$. One pictures $\mathfrak{D}$ as lying above its *projection* $\underline{\mathfrak{D}}$ into $\mathbb{C}^m$. The projection $\underline{P}$ of a point P of $\mathfrak{D}$ furnishes coordinates for that point. It is possible that several points of $\mathfrak{D}$ project into the same point $\underline{P}$ of $\underline{\mathfrak{D}}$. If the projection map is a global homeomorphism, then $\mathfrak{D}$ is a *schlicht domain* (single sheeted). If $\mathfrak{D}_1$, $\mathfrak{D}_2$ are two domains over $\mathbb{C}^m$ and $\mathfrak{D}_1$ can be mapped continuously into a subset of $\mathfrak{D}_2$ so that corresponding points have the same coordinates, we say $\mathfrak{D}_1$ *lies inside* (*liegt im innern*) $\mathfrak{D}_2$ and write $\mathfrak{D}_1 < \mathfrak{D}_2$. Although $\mathfrak{D}_1 < \mathfrak{D}_2$, it is possible that $\mathfrak{D}_1$ is more ramified (has more sheets) than $\mathfrak{D}_2$. For schlicht domains $\mathfrak{D}_1 < \mathfrak{D}_2$ means the same as $\mathfrak{D}_1 \subset \mathfrak{D}_2$.

Suppose $\mathfrak{G}$ is a domain over $\mathbb{C}^m$, such that $\mathfrak{G} < \mathfrak{D}_\alpha$ for each of a set $\{\mathfrak{D}_\alpha\}$ of domains over $\mathbb{C}^m$. Then, if P is a point of $\mathfrak{G}$, the projection preserving continuous map in the definition of $<$ gives an image point P_α in each domain $\mathfrak{D}_\alpha$. If $\mathfrak{E} < \mathfrak{D}_\alpha$, then there is associated, due to $<$, an image subdomain $\mathfrak{E}_\alpha$ of $\mathfrak{D}_\alpha$. One defines the intersection $\mathfrak{E} = \bigcap \mathfrak{D}_\alpha$ of the domains $\{\mathfrak{D}_\alpha\}$ relative to $\mathfrak{G}$ as follows[8,9]:

(a) $\mathfrak{E} < \mathfrak{D}_\alpha$, each α,

(b) $P_\alpha \in \mathfrak{E}_\alpha$, and

(c) every domain $\mathfrak{E}'$ satisfying (a) and (b) satisfies $\mathfrak{E}' < \mathfrak{E}$.

It can be shown that $\mathfrak{E}$ is a domain over ∞^m and that it does not depend on the choice of P,[8,9] provided $P \in \mathfrak{G}$.

Given a (single valued) holomorphic function f' in a domain $\mathfrak{D}'$, suppose there is a domain $\mathfrak{D} > \mathfrak{D}'$ and a function f holomorphic in $\mathfrak{D}$ such that $f = f'$ for corresponding points of the domains. Then f is an analytic continuation of f' from $\mathfrak{D}'$ into (the "larger" domain) $\mathfrak{D}$. Considering the development of the holomorphic functional germs of f', one obtains f, and in the process $\mathfrak{D}$ is generated with adequate ramification to have f single valued in it. If $\mathfrak{D}$ is such that f cannot be analytically continued into a domain $\mathfrak{E} > \mathfrak{D}$, then $\mathfrak{D}$ is the holomorphy domain $\mathfrak{D}_f$ for the function f.

Given a domain $\mathfrak{D}$, we define the holomorphy envelope $H(\mathfrak{D})$ of $\mathfrak{D}$ by $H(\mathfrak{D}) = \bigcap_{f \in S} \mathfrak{D}_f$, relative to $\mathfrak{D}$, where the intersection is taken over the set S of all functions holomorphic in $\mathfrak{D}$. We remark that although $\mathfrak{D}$ may be schlicht (as we have for $\mathfrak{S}_n^P$), its holomorphy envelope $H(\mathfrak{D})$ may not be schlicht. This explains why we have gone into the consideration of domains *over* $\mathbb{C}^m$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Invariants of Nearly Periodic Hamiltonian Systems

B. McNAMARA AND K. J. WHITEMAN
U.K.A.E.A. Research Group, Culham Laboratory, Abingdon, Berkshire, England

(Received 30 January 1967)

A new and simple method of finding an invariant J of a nearly periodic dynamical system is presented. The Hamiltonian is written as $H = p_1 + \epsilon\Omega(q_i p_i)$, where Ω is periodic in q_1 and $\epsilon \ll 1$. The first four terms of the invariant series are found explicitly in terms of Ω using Poisson bracket and averaging operators. This invariant is related to the adiabatic invariant and to various constants of motion discussed in celestial mechanics, such as Whittaker's adelphic integral. J is shown to be an asymptotic constant by using the rigorous methods of Kruskal to calculate the adiabatic invariant K; it is found that $K/\tau = H - \epsilon J$, where τ is the period in q_1. The adelphic integral has different functional forms depending on the presence of resonant denominators, but is shown to be always a function of H and J. The present method provides a single functional form which is applicable even when Ω is only almost periodic in q_1. It is also much simpler than the methods of adiabatic invariant theory.

1. INTRODUCTION

THIS paper is primarily concerned with dynamical systems that are nearly periodic in the following sense. The Hamiltonian of the system is time independent and also a function of a small parameter ϵ. When $\epsilon = 0$, all motions of the system execute closed orbits in phase space but for small values of ϵ the orbits are no longer exactly closed and slowly drift. One can show (see Sec. 2) that the Hamiltonian for such systems can be reduced to the form

$$H = p_1 + \epsilon\Omega(q_i; p_i), \tag{1.1}$$

where $(q_i; p_i)$ are some set of canonical coordinates and Ω is a periodic function of q_1. There is also a range of problems where Ω is only an almost-periodic function of q_1. Many of the results with Ω periodic can

Hausdorff topological space which is locally homeomorphic to an open subset of $\mathbb{C}^m$. One pictures $\mathfrak{D}$ as lying above its *projection* $\underline{\mathfrak{D}}$ into $\mathbb{C}^m$. The projection $\underline{P}$ of a point P of $\mathfrak{D}$ furnishes coordinates for that point. It is possible that several points of $\mathfrak{D}$ project into the same point $\underline{P}$ of $\underline{\mathfrak{D}}$. If the projection map is a global homeomorphism, then $\mathfrak{D}$ is a *schlicht domain* (single sheeted). If $\mathfrak{D}_1$, $\mathfrak{D}_2$ are two domains over $\mathbb{C}^m$ and $\mathfrak{D}_1$ can be mapped continuously into a subset of $\mathfrak{D}_2$ so that corresponding points have the same coordinates, we say $\mathfrak{D}_1$ *lies inside* (*liegt im innern*) $\mathfrak{D}_2$ and write $\mathfrak{D}_1 < \mathfrak{D}_2$. Although $\mathfrak{D}_1 < \mathfrak{D}_2$, it is possible that $\mathfrak{D}_1$ is more ramified (has more sheets) than $\mathfrak{D}_2$. For schlicht domains $\mathfrak{D}_1 < \mathfrak{D}_2$ means the same as $\mathfrak{D}_1 \subset \mathfrak{D}_2$.

Suppose $\mathfrak{G}$ is a domain over $\mathbb{C}^m$, such that $\mathfrak{G} < \mathfrak{D}_\alpha$ for each of a set $\{\mathfrak{D}_\alpha\}$ of domains over $\mathbb{C}^m$. Then, if P is a point of $\mathfrak{G}$, the projection preserving continuous map in the definition of $<$ gives an image point P_α in each domain $\mathfrak{D}_\alpha$. If $\mathfrak{E} < \mathfrak{D}_\alpha$, then there is associated, due to $<$, an image subdomain $\mathfrak{E}_\alpha$ of $\mathfrak{D}_\alpha$. One defines the intersection $\mathfrak{E} = \bigcap \mathfrak{D}_\alpha$ of the domains $\{\mathfrak{D}_\alpha\}$ relative to $\mathfrak{G}$ as follows[8,9]:

(a) $\mathfrak{E} < \mathfrak{D}_\alpha$, each α,

(b) $P_\alpha \in \mathfrak{E}_\alpha$, and

(c) every domain $\mathfrak{E}'$ satisfying (a) and (b) satisfies $\mathfrak{E}' < \mathfrak{E}$.

It can be shown that $\mathfrak{E}$ is a domain over ∞^m and that it does not depend on the choice of P,[8,9] provided $P \in \mathfrak{G}$.

Given a (single valued) holomorphic function f' in a domain $\mathfrak{D}'$, suppose there is a domain $\mathfrak{D} > \mathfrak{D}'$ and a function f holomorphic in $\mathfrak{D}$ such that $f = f'$ for corresponding points of the domains. Then f is an analytic continuation of f' from $\mathfrak{D}'$ into (the "larger" domain) $\mathfrak{D}$. Considering the development of the holomorphic functional germs of f', one obtains f, and in the process $\mathfrak{D}$ is generated with adequate ramification to have f single valued in it. If $\mathfrak{D}$ is such that f cannot be analytically continued into a domain $\mathfrak{E} > \mathfrak{D}$, then $\mathfrak{D}$ is the holomorphy domain $\mathfrak{D}_f$ for the function f.

Given a domain $\mathfrak{D}$, we define the holomorphy envelope $H(\mathfrak{D})$ of $\mathfrak{D}$ by $H(\mathfrak{D}) = \bigcap_{f \in S} \mathfrak{D}_f$, relative to $\mathfrak{D}$, where the intersection is taken over the set S of all functions holomorphic in $\mathfrak{D}$. We remark that although $\mathfrak{D}$ may be schlicht (as we have for $\mathfrak{S}_n^P$), its holomorphy envelope $H(\mathfrak{D})$ may not be schlicht. This explains why we have gone into the consideration of domains *over* $\mathbb{C}^m$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Invariants of Nearly Periodic Hamiltonian Systems

B. McNAMARA AND K. J. WHITEMAN
U.K.A.E.A. Research Group, Culham Laboratory, Abingdon, Berkshire, England

(Received 30 January 1967)

A new and simple method of finding an invariant J of a nearly periodic dynamical system is presented. The Hamiltonian is written as $H = p_1 + \epsilon\Omega(q_i p_i)$, where Ω is periodic in q_1 and $\epsilon \ll 1$. The first four terms of the invariant series are found explicitly in terms of Ω using Poisson bracket and averaging operators. This invariant is related to the adiabatic invariant and to various constants of motion discussed in celestial mechanics, such as Whittaker's adelphic integral. J is shown to be an asymptotic constant by using the rigorous methods of Kruskal to calculate the adiabatic invariant K; it is found that $K/\tau = H - \epsilon J$, where τ is the period in q_1. The adelphic integral has different functional forms depending on the presence of resonant denominators, but is shown to be always a function of H and J. The present method provides a single functional form which is applicable even when Ω is only almost periodic in q_1. It is also much simpler than the methods of adiabatic invariant theory.

1. INTRODUCTION

THIS paper is primarily concerned with dynamical systems that are nearly periodic in the following sense. The Hamiltonian of the system is time independent and also a function of a small parameter ϵ. When $\epsilon = 0$, all motions of the system execute closed orbits in phase space but for small values of ϵ the orbits are no longer exactly closed and slowly drift. One can show (see Sec. 2) that the Hamiltonian for such systems can be reduced to the form

$$H = p_1 + \epsilon\Omega(q_i; p_i), \tag{1.1}$$

where $(q_i; p_i)$ are some set of canonical coordinates and Ω is a periodic function of q_1. There is also a range of problems where Ω is only an almost-periodic function of q_1. Many of the results with Ω periodic can

be extended to the almost-periodic case and so this class of motions will also be considered.

Being time-independent, the Hamiltonian itself is one constant of the motion, but unless it exhibits some other symmetry, like possessing an ignorable coordinate, there are no other obvious constants. However, it is possible to find another integral of the motion, expressed as a power series in ϵ, associated with the fact that the system is nearly periodic. It is with this invariant that the paper is mainly concerned.

Invariants of this type have been discussed previously but in two quite different contexts, celestial mechanics and plasma physics. In celestial mechanics various authors[1–6] have studied the motion of a particle near a point of equilibrium and looked for another constant of the motion different from the Hamiltonian. For example, Whittaker[6] obtains an invariant that he calls an adelphic integral Φ. He solves Liouville's equation and finds Φ as a series of terms of decreasing order of magnitude. In his method several difficulties arise. He encounters the problem of small or vanishing denominators which entails using different techniques in a variety of special cases. As a result the adelphic integral has not one but several different analytical forms. Another drawback is that these methods do not allow any conclusion to be reached about the nature of the series, whether for example it is convergent or asymptotic. Applications of this invariant have been found in the three-body problem and the problem of the third integral of galactic motion.

In plasma physics the problem of particle containment has led to detailed study of adiabatic invariants. Originally they were regarded as quantities that remained virtually constant as the parameters of the system varied slowly in time. A more general invariant is obtained when the parameters are allowed to vary slowly in both space and time and a detailed discussion of this case has been given by Kruskal.[7] The invariant is found as an action integral and is evaluated as a series of terms which is shown to be an asymptotic series. Examples of the invariant are the three invariants of a charged particle in a magnetic field and the magnetic surfaces generated by nearly periodic magnetic field lines. Other examples arise in the motion of artificial satellites.

It is not immediately apparent and does not seem to have been remarked in the literature that the invariants obtained in these two situations are equivalent. The correspondence is most easily seen by considering the Hamiltonian in each case. The equations that Kruskal deals with are the equations of motion for a Hamiltonian of the form (1.1), but with Ω strictly periodic in q_1 with period τ. One can in general find an invariant as long as τ is a function of p_1 and a slowly varying function of $(q_2, q_3 \cdots q_n; p_2, p_3 \cdots p_n)$ and the time, t. The dynamical systems discussed by Whittaker also have Hamiltonians that can be reduced to the form (1.1). In his case Ω is the sum of several terms periodic in q_1 with periods τ_i $(i = 1, 2 \cdots)$ and where the τ_i are constants, independent of the coordinates and time. In this paper we limit ourselves to a discussion of dynamical systems having this latter form with Ω either

> (a) periodic in q_1 with period τ independent of space and time coordinates; or (b) almost periodic in the sense that it is the sum of several such terms with periods τ_i that are incommensurable.

The methods used below can however be generalized to other cases where the periods are functions of the coordinates.

Once the Hamiltonian has been reduced to the standard form an invariant can be calculated explicitly in terms of Ω and this is done by two methods. In the first of these (Sec. 3), which is the most straightforward, Liouville's equation is expanded in powers of ϵ and solved order by order. It differs from Whittaker's method in several important respects and as a result the problem of vanishing denominators no longer arises. One formalism embraces all the cases that previously needed separate treatment and a single expression is obtained for the invariant. The operator techniques used also allow several terms in the series to be evaluated with comparative ease. The second method is to carry out explicitly the procedure described in general terms by Kruskal. This is more tedious to apply but has compensating advantages. It is systematic and so allows one, in principle, to obtain the series to any number of terms. It also gives information about the asymptotic nature of the invariant. Kruskal's method and its application are described in Sec. 4. Although these two methods give essentially the same result, it is not obvious that they are equivalent to the adelphic integral. The relation between Whittaker's invariant and the forms obtained in this paper are discussed in Sec. 5.

---

[1] G. D. Birkhoff, *Dynamical Systems* (Am. Math. Soc., New York, 1927).
[2] J. M. Cherry, Proc. Cambridge Phil. Soc. **22**, 325 and 510 (1924).
[3] G. Contopoulos, Z. Astrophys. **49**, 273 (1960).
[4] G. Contopoulos, Astron. J. **68**, 763 (1963).
[5] G. Contopoulos and M. Moutsoulas, Astron. J. **70**, 817 (1965).
[6] E. T. Whittaker, *Analytical Dynamics of Particles and Rigid Bodies* (Cambridge University Press, Cambridge, England, 1937), 4th Ed.
[7] M. Kruskal, J. Math. Phys. **3**, 806 (1962).

2. REDUCTION OF THE HAMILTONIAN TO STANDARD FORM

The basic dynamical system considered in the following is n-dimensional and has a time-independent Hamiltonian of the form

$$H = H_0(q_i', p_i') + \epsilon H_1(q_i', p_i'), \tag{2.1}$$

where (q_i', p_i') are canonical coordinates. It is assumed that the equations of motion for the zero-order Hamiltonian H_0 can be solved and that all the trajectories are closed curves in phase space, having the same periodic time τ. This being so, one can use Hamilton–Jacobi theory[8] to transform to new coordinates $(q_i; p_i)$ such that H_0 is a function of the new momenta p_i only. In particular it is possible to choose $(q_i; p_i)$ so that $H_0 = p_1$. This is done by solving the Hamilton–Jacobi equation

$$H_0\left(q_i', \frac{\partial W}{\partial q_i'}\right) = p_1, \tag{2.2}$$

for Hamilton's characteristic function $W(q_i', p_i)$. It is then found that

$$q_1 = t + \beta_1, \quad \text{and} \quad q_i = \beta_i, \quad i \neq 1, \tag{2.3}$$

where the β_i are constants. The new momenta are given by the equations

$$p_i' = \frac{\partial W}{\partial q_i'}(q_i', p_i). \tag{2.4}$$

During the motion described by H_0, therefore, q_1 is proportional to the time and $(q_2 \cdots q_n; p_1 \cdots p_n)$ remain constant. A simplifying step in the following calculations is to transform to this coordinate system that displays the fundamental angle variable, q_1, of the zero-order motion.

Because of the periodicity the point $(q_1 \cdots q_n; p_1 \cdots p_n)$ is the same as the point $(q_1 + \tau, q_2 \cdots q_n; p_1 \cdots p_n)$. Since the Hamiltonian is a constant of the motion and a single valued function of position in phase space it follows that

$$H(q_1) = H(q_1 + \tau).$$

The Hamiltonian can therefore be written in the required form

$$H = p_1 + \epsilon\Omega(p_i, q_i) \tag{1.1}$$

with Ω periodic in q_1 period τ. Although it has not been stated explicitly, Ω can itself be a function of ϵ.

There are dynamical systems where the lowest-order motion is not strictly periodic with closed orbits but conditionally periodic, the orbits describing open Lissajous figures. For example, if the lowest-order Hamiltonian is of the form

$$H = \sum_{i=1}^{n} \tfrac{1}{2}(p_i'^2 + \alpha_i^2 q_i'^2), \tag{2.5}$$

the system represents n uncoupled simple harmonic oscillators. If all the ratios α_i/α_j are rational numbers, the phase–space trajectories are closed and the system periodic. If any ratio is an irrational number, however, then the motion is only conditionally periodic. In this case one can still transform to new coordinates such that the Hamiltonian takes the form (1.1), but one now finds that Ω is only almost periodic in q_1.

3. THE POISSON BRACKET METHOD

With the Hamiltonian in the reduced form, a new constant of the motion J can now be found different from H. Considering firstly the case with Ω periodic in q_1, an invariant is sought independent of time and periodic in q_1 with period τ. The constant J must satisfy Liouville's equation

$$\frac{dJ}{dt} \equiv \frac{\partial J}{\partial t} - [J, H] = 0, \tag{3.1}$$

where $[J, H]$ is the Poisson bracket defined by

$$[J, H] \equiv \frac{\partial J}{\partial p_i}\frac{\partial H}{\partial q_i} - \frac{\partial J}{\partial q_i}\frac{\partial H}{\partial p_i}. \tag{3.2}$$

With $\partial J/\partial t = 0$, (3.1) reduces to the equation

$$[J, H] = 0. \tag{3.3}$$

J is expanded as a power series in ϵ:

$$J = \sum_{n=0}^{\infty} \epsilon^n J_n. \tag{3.4}$$

Substituting (1.1) and (3.4) in (3.3) and equating terms in ϵ^n, one obtains

$$\frac{\partial J_0}{\partial q_1} = 0, \tag{3.5}$$

and

$$\frac{\partial J_n}{\partial q_1} = [J_{n-1}, \Omega]. \tag{3.6}$$

Integration of (3.6) gives

$$J_n = \int [J_{n-1}, \Omega]\, dq_1 + G_n, \tag{3.7}$$

where G_n is independent of q_1. Since J_n is required to be periodic in q_1, one obtains the periodicity condition

$$\int_0^{\tau} [J_{n-1}, \Omega]\, dq_1 = 0. \tag{3.8}$$

[8] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publishing Company, London, 1959), p. 280.

To generate the terms in the series for J, it is now convenient to introduce some new terminology. For a periodic function f of q_1 period τ, the functions $\bar{f}$ and $\hat{f}$ are defined as follows:

$$\bar{f} = \frac{1}{\tau}\int_0^\tau f\, dq_1, \tag{3.9}$$

and

$$\hat{f} = \int (f - \bar{f})\, dq_1, \tag{3.10}$$

with the constant of integration chosen so that $\bar{\hat{f}} = 0$. Thus, $\bar{f}$ is the mean value of f and $\hat{f}$ the indefinite integral of the oscillatory part of f. It follows from the definitions that

$$\hat{\bar{f}} = 0, \tag{3.11}$$

$$\frac{\partial}{\partial x}(\bar{f}) = \overline{\left(\frac{\partial f}{\partial x}\right)}, \tag{3.12}$$

and

$$\frac{\partial}{\partial x}(\hat{f}) = \widehat{\left(\frac{\partial f}{\partial x}\right)}, \tag{3.13}$$

where x is any of the variables q_i, p_i. The following results can be proved using integration by parts:

$$\overline{f\bar{g}} + \bar{f}\bar{g} = 0 \tag{3.14}$$

and

$$\widehat{f\bar{g}} + \hat{f}\bar{g} = \hat{f}\hat{g} - \overline{\hat{f}\hat{g}} + \hat{f}\dot{\hat{g}} + \dot{\hat{f}}\hat{\bar{g}}, \tag{3.15}$$

where both f and g are periodic in q_1, period τ. Use is also made of the Poisson bracket relation

$$[f, g] = -[g, f], \tag{3.16}$$

and Jacobi's identity

$$[f, [g, h]] + [h, [f, g]] + [g, [h, f]] = 0. \tag{3.17}$$

Finally for a function K, independent of q_1

$$\overline{[K, f]} = [K, \bar{f}] \tag{3.18}$$

and

$$\widehat{[K, f]} = [K, \hat{f}]. \tag{3.19}$$

The first few terms of J can now be calculated in terms of the operators $^-$, $\hat{}$, and [ ]. Equations (3.7) and (3.8) become

$$J_n = \widehat{[J_{n-1}, \Omega]} + G_n \tag{3.20}$$

and

$$\overline{[J_{n-1}, \Omega]} = 0. \tag{3.21}$$

J_0 is independent of q_1 from (3.5) and the lowest-order periodicity condition is obtained from (3.21):

$$\overline{[J_0, \Omega]} = 0 = [J_0, \bar{\Omega}]. \tag{3.22}$$

In generating a constant of the motion, J, one seeks only a particular solution of this equation and an obvious solution is

$$J_0 = J_0(\bar{\Omega}).$$

Since any function of an invariant is also an invariant, any functional form could be taken for J_0. For simplicity the choice made is that

$$J_0 = \bar{\Omega}. \tag{3.23}$$

Equation (3.20) then gives

$$J_1 = [\bar{\Omega}, \hat{\Omega}] + G_1$$

and the periodicity condition (3.21) gives

$$\overline{[[\bar{\Omega}, \hat{\Omega}]\Omega]} + [G_1, \bar{\Omega}] = 0. \tag{3.24}$$

To solve this equation for G_1 the first term is rearranged as follows:

$$\overline{[[\bar{\Omega}, \hat{\Omega}], \Omega]} = -\overline{[[\Omega, \bar{\Omega}]\hat{\Omega}]} - \overline{[[\hat{\Omega}, \Omega], \bar{\Omega}]},$$

using (3.17), and (3.18),

$$= -\overline{[[\bar{\Omega}, \hat{\bar{\Omega}}], \Omega]} - \overline{[[\hat{\bar{\Omega}}, \Omega], \bar{\Omega}]},$$

using (3.14), (3.16), and (3.19).

$$\therefore \overline{[[\bar{\Omega}, \hat{\Omega}]\Omega]} = -\tfrac{1}{2}\overline{[[\hat{\Omega}, \Omega]\bar{\Omega}]}. \tag{3.25}$$

Equation (3.24) becomes

$$[G_1 - \tfrac{1}{2}\overline{[\hat{\Omega}, \Omega]}, \bar{\Omega}] = 0.$$

With the equation rearranged in this way, a particular integral is easily found:

$$G_1 = \tfrac{1}{2}\overline{[\hat{\Omega}, \Omega]} + F(\bar{\Omega}), \tag{3.26}$$

where F is any function of $\bar{\Omega}$. In generating the invariant series one chooses the particular case $F = 0$. Any other choice merely adds a function of the lower-order invariant J_0 to J_1.

This process can be repeated, a new partial differential equation for G_n having to be solved in each order. The equation for G_2 has been solved in a similar manner to that for G_1. The terms were rearranged to obtain an equation for which a particular integral was easily found. The relations (3.11)–(3.19) were the only ones needed to accomplish this. It has not yet proved possible to demonstrate that such a rearrangement can be made for all values of n.

For the particular case when $\bar{\Omega} = 0$ the lowest-order periodicity condition (3.22) is identically satisfied and J_0 must be determined from the next order periodicity condition. This process can be continued and the invariant is the same as is obtained by putting $\bar{\Omega} = 0$ in the general solution. When $\bar{\Omega} = 0$ it also proved possible to find G_3. To sum up,

therefore,

$$J = \sum_{n=0}^{\infty} \epsilon^n J_n, \tag{3.4}$$

where

$$J_n = \widehat{[J_{n-1}, \Omega]} + G_n \tag{3.7}$$

and

$$\overline{[G_n, \Omega] + \widehat{[[J_{n-1}, \Omega], \Omega]}} = 0.$$

A solution is

$$J_0 = \bar{\Omega},$$

$$G_1 = \tfrac{1}{2}\overline{[\hat{\Omega}, \Omega]},$$

$$G_2 = \tfrac{1}{3}\overline{[\hat{\Omega}, [\hat{\Omega}, \Omega]]} + \tfrac{2}{3}\overline{[\hat{\Omega}, [\hat{\Omega}, \bar{\Omega}]]}, \tag{3.27}$$

and when $\bar{\Omega} = 0$

$$G_3 = \tfrac{1}{8}\overline{[\hat{\Omega}, [\hat{\Omega}, [\hat{\Omega}, \Omega]]]} + \tfrac{1}{8}\overline{\widehat{[[\Omega, \hat{\Omega}], [\Omega, \hat{\Omega}]]}} + \tfrac{1}{4}\overline{[[\overline{[\hat{\Omega}, \Omega]}, \hat{\Omega}], \hat{\Omega}]}.$$

The distinction between Ω being periodic and almost periodic now appears to have been unnecessary. Instead of insisting that J be periodic in q_1, one could have applied the condition that J contain no terms that become indefinitely large as $q_1 \to \infty$. This is assured if

$$\lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} [J_n, \Omega]\, dq_1 = 0. \tag{3.28}$$

With a redefinition of $\bar{f}$ to be

$$\bar{f} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} f\, dq_1 \tag{3.29}$$

[which also includes the previous definition (3.9)], it is apparent that the calculation follows exactly as before. The results obtained in (3.27) therefore apply both for Ω periodic and almost periodic in q_1 with the averaging operator defined by Eq. (3.29).

It must be stressed that this is merely a formal series solution of (3.3). The derivation allows no statement about the validity of the solution and whether the series is convergent or otherwise. However, using Kruskal's method one is able to show that it is in fact an asymptotic series in the limit $\epsilon \to 0$ for the case where Ω is periodic.

4. KRUSKAL'S AVERAGING METHOD

A large part of celestial mechanics is concerned with almost periodic systems and, since one can rarely obtain exact results, a particular aim has been to find approximate descriptions that are valid even for large values of the time. For this latter condition to be realized, it is necessary for the solution to be free from secular terms and one of the achievements of the subject is to have produced perturbation techniques that accomplish this. They are all essentially what is now called "the averaging method" and associated with the names of Poincaré,[9] Bogoliubov,[10] and many others.

A. General Description of the Method

A general formulation of the averaging method has been given by Kruskal[7] and in several important respects he has extended its usefulness. He deals with a set of equations of the form

$$\mathbf{x}_t \equiv d\mathbf{x}/dt = F(\mathbf{x}, \epsilon),$$

such that for $\epsilon = 0$ the point $\mathbf{x}(t)$ traces out closed curves (loops) as t increases. (For this reason his method as it stands is only valid for the case where the lowest order motion is periodic and not for the conditionally periodic case.) It is then possible to transform to new variables $\mathbf{y}$ and an angle-like variable ν such that $\mathbf{y}$ is constant on a loop and ν varies around it. When the motions around the loops all have the same period τ the equations for $\mathbf{y}$ and ν become

$$\mathbf{y}_t = \epsilon \mathbf{g}(\mathbf{y}, \nu), \tag{4.1}$$

$$\nu_t = 1 + \epsilon f(\mathbf{y}, \nu), \tag{4.2}$$

where f and $\mathbf{g}$ are periodic in ν period τ. When ϵ is small, these equations describe a slowly drifting oscillatory motion and the object of the method is to transform to new variables $\mathbf{z}$ and φ which separate the drift motion and the oscillation. It is required that $\mathbf{z}$ be periodic in ν and that φ be an angle-like variable, i.e.,

$$\mathbf{z}(\mathbf{y}, \nu) = \mathbf{z}(\mathbf{y}, \nu + \tau), \tag{4.3}$$

$$\varphi(\mathbf{y}, \nu + \tau) = \varphi(\mathbf{y}, \nu) + \tau, \tag{4.4}$$

and also that the equations of the drift variables should not contain the angle variable, i.e.,

$$\mathbf{z}_t = \epsilon \mathbf{h}(\mathbf{z}) \tag{4.5}$$

$$\varphi_t = 1 + \epsilon \omega(\mathbf{z}). \tag{4.6}$$

Kruskal has shown that it is possible to find systematically $\mathbf{z}$, φ, $\mathbf{h}$, and ω as power series in ϵ to any order required. It is also possible (and this is essential in what follows) to find the inverse transformation as a power series in ϵ, i.e.,

$$\mathbf{y} = \mathbf{y}(\mathbf{z}, \varphi) = \sum_{n=0}^{\infty} \epsilon^n \mathbf{y}_n,$$

$$\nu = \nu(\mathbf{z}, \varphi) = \sum_{n=0}^{\infty} \epsilon^n \nu_n.$$

[9] H. Poincaré, *Les méthodes nouvelles de le mécanique céleste* (Dover Publications, New York, 1957), Vol. 3.

[10] N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Gordon and Breach Science Publishers, New York, 1961).

One can show that the solutions of the transformed equations (4.5) and (4.6) provide asymptotic solutions of Eqs. (4.1) and (4.2) in the following sense. If one writes the partial sums in forms like

$$\mathbf{h}^{[n]} = \sum_{k=0}^{n} \epsilon^k \mathbf{h}_k, \tag{4.7}$$

then $\mathbf{z}'$ and φ' are defined to be solutions of

$$d\mathbf{z}'/dt = \epsilon \mathbf{h}^{[n]}(\mathbf{z}'),$$
$$d\varphi'/dt = 1 + \epsilon\omega^{[n]}(\mathbf{z}').$$

In terms of these, one can write

$$\mathbf{y}' = \mathbf{y}^{[n]}(\mathbf{z}', \varphi'),$$

and

$$\nu' = \nu^{[n]}(\mathbf{z}', \varphi').$$

Kruskal proves that if $\mathbf{y}^*$ and ν^* are solutions of the original equations satisfying the same initial conditions as $\mathbf{y}'$ and ν', then

$$\mathbf{y}^* - \mathbf{y}' = O(\epsilon^{n+1}),$$

and

$$\nu^* - \nu' = O(\epsilon^{n+1})$$

for t within a range of order $1/\epsilon$. That is, solutions of (4.5) and (4.6) give asymptotically correct solutions of (4.1) and (4.2) as $\epsilon \to 0$.

It is apparent that the equations of motion derived from (1.1) are equations of the form (4.1) and (4.2) with

$$(\nu, \mathbf{y}) \sim (\mathbf{q}; \mathbf{p}),$$

and

$$(f, \mathbf{g}) \sim (\Omega\mathbf{p}; -\Omega\mathbf{q}),$$

where

$$\Omega\mathbf{p} = \frac{\partial\Omega}{\partial\mathbf{p}}, \text{ etc.}$$

One could therefore find an asymptotic solution of the equations of motion to any order in ϵ by solving (4.5) and (4.6) for $\mathbf{z}$ and φ as functions of t and then substituting the results to obtain $\mathbf{q}(\mathbf{z}, \varphi)$ and $\mathbf{p}(\mathbf{z}, \varphi)$. However, if one seeks an invariant of the motion rather than the trajectories, it is not necessary even to solve the differential equations for $\mathbf{z}$ and φ.

If C_0 is any closed curve in phase space at time $t = 0$ and if at a later time, t, the points that comprised it now lie on a curve C_t, then it is well known that the action integral taken around C_t:

$$\oint \mathbf{p} \cdot d\mathbf{q}$$

is an invariant of the dynamical system. In order to evaluate the integral, it is necessary to know the general form of C_t which usually entails solving the equations of motion. However, one can see from Eqs. (4.5) and (4.6) that the points comprising a φ-loop (a closed curve on which $\mathbf{z}$ is constant and φ varies) remain a φ-loop as the system develops in time. Choosing C_t to be a φ-loop one can therefore define an invariant K:

$$K(\mathbf{z}) = \oint \mathbf{p} \cdot d\mathbf{q} = \int_0^\tau p_i \frac{\partial q_i}{\partial\varphi}\, d\varphi. \tag{4.8}$$

The time rate of change of K is

$$K_t = \oint (\mathbf{p}_t \cdot d\mathbf{q} - \mathbf{q}_t \cdot d\mathbf{p}),$$

the second term having been integrated by parts, i.e.,

$$K_t = -\oint (H_{\mathbf{q}} \cdot d\mathbf{q} + H_{\mathbf{p}} \cdot d\mathbf{p})$$
$$= -\oint dH.$$

If $\mathbf{p}$ and $\mathbf{q}$ were known exactly as functions of $\mathbf{z}$ and φ and used to evaluate this integral, then K_t would be exactly zero. However, $\mathbf{p}$ and $\mathbf{q}$ as functions of $\mathbf{z}$ and φ are only defined as power series in ϵ. If the invariant is calculated to order n using the partial sums $\mathbf{p}^{[n]}$ and $\mathbf{q}^{[n]}$ instead of $\mathbf{p}$ and $\mathbf{q}$, then the integral does not vanish exactly. One finds

$$K_t^{[n]} = O(\epsilon^{n+1}).$$

The quantity K is therefore only asymptotically invariant.

B. The Coordinate Transformation

As a preliminary to calculating K, the transformations

$$(\mathbf{y}, \nu) \rightleftarrows (\mathbf{z}, \varphi)$$

are obtained. From (4.1) to (4.6) one finds

$$\mathbf{z}_t = \epsilon\mathbf{h}(\mathbf{z}) = \epsilon\mathbf{g} \cdot \mathbf{z}_{\mathbf{y}} + (1 + \epsilon f)\mathbf{z}_\nu \tag{4.9}$$

and

$$\varphi_t = 1 + \epsilon\omega(\mathbf{z}) = \epsilon\mathbf{g} \cdot \varphi_{\mathbf{y}} + (1 + \epsilon f)\varphi_\nu, \tag{4.10}$$

where $\mathbf{z}_{\mathbf{y}} = \partial\mathbf{z}/\partial\mathbf{y}$ is the tensor with ijth component $\partial z_j/\partial y_i$, etc.

On integrating over ν and applying the initial conditions

$$\begin{aligned} \mathbf{z} &= \mathbf{y}, \\ \varphi &= 0, \end{aligned} \tag{4.11}$$

when

$$\nu = 0.$$

Equations (4.9) and (4.10) give

$$\mathbf{z} = \mathbf{y} + \epsilon \int_0^\nu d\nu(\mathbf{h}(\mathbf{z}) - \mathbf{g} \cdot \mathbf{z}_{\mathbf{y}} - f\mathbf{z}_\nu) \tag{4.12}$$

and

$$\varphi = \nu + \epsilon \int_0^\nu d\nu(\omega(\mathbf{z}) - \mathbf{g} \cdot \varphi_{\mathbf{y}} - f\varphi_\nu). \quad (4.13)$$

There is some freedom in choosing the initial conditions of the transformation. In the work of Bogoliubov and Mitropolsky,[10] the integration constants are chosen to eliminate the contribution of the lower limits to the integrals in (4.12) and (4.13). This makes the transformation formulas somewhat simpler, but the only convenient choice, if the inverse transformation is to be obtained also, is that given here.

If one now writes

$$\begin{aligned} &\mathbf{Z} = (\varphi, \mathbf{z}), \quad \mathbf{H}(\mathbf{z}) = (\omega, \mathbf{h}), \\ &\mathbf{Y} = (\nu, \mathbf{y}) \quad \text{and} \quad \mathbf{G}(\mathbf{Y}) = (f, \mathbf{g}), \end{aligned} \quad (4.14)$$

Eqs. (4.12) and (4.13) become

$$\mathbf{Z} = \mathbf{Y} + \epsilon \int_0^\nu d\nu(\mathbf{H}(\mathbf{z}) - \mathbf{G} \cdot \mathbf{Z}_{\mathbf{Y}}). \quad (4.15)$$

The conditions that $\mathbf{z}$ be periodic in ν, and that φ be an angle-like variable with period τ can be written

$$\int_0^\tau d\nu(\mathbf{H}(\mathbf{z}) - \mathbf{G} \cdot \mathbf{Z}_{\mathbf{Y}}) = 0$$

or

$$\begin{aligned} \mathbf{H}(\mathbf{y}) &= \frac{1}{\tau} \int_0^\tau d\nu(\mathbf{G} \cdot \mathbf{Z}_{\mathbf{Y}} + \mathbf{H}(\mathbf{y}) - \mathbf{H}(\mathbf{z})) \\ &= \overline{\mathbf{G} \cdot \mathbf{Z}_{\mathbf{Y}}} + \overline{\mathbf{H}(\mathbf{y})} - \overline{\mathbf{H}(\mathbf{z})}. \end{aligned} \quad (4.16)$$

Equations (4.15) and (4.16) can now be solved iteratively to give a power series expansion of

$$\mathbf{Z} = \sum_0^\infty \epsilon^n \mathbf{Z}_n.$$

$\mathbf{H}$ must also be expanded as follows:

$$\begin{aligned} \mathbf{H}(\mathbf{z}) &= \mathbf{H}_0(\mathbf{z}) + \epsilon\mathbf{H}_1(\mathbf{z}) + \epsilon^2\mathbf{H}_2(\mathbf{z}) + \cdots \\ &= \mathbf{H}_0(\mathbf{z}_0) + \epsilon\left(\mathbf{H}_1(\mathbf{z}_0) + \mathbf{z}_1 \frac{\partial}{\partial \mathbf{z}} \mathbf{H}_0(\mathbf{z}_0)\right) \\ &\quad + \epsilon^2\left(\mathbf{H}_2(\mathbf{z}_0) + \mathbf{z}_1 \frac{\partial}{\partial \mathbf{z}} \mathbf{H}_1(\mathbf{z}_0) \right. \\ &\quad \left. + \frac{1}{2}\left(\mathbf{z}_1 \frac{\partial}{\partial \mathbf{z}}\right)^2 \mathbf{H}_0(\mathbf{z}_0)\right). \end{aligned} \quad (4.17)$$

It is convenient at this stage to introduce the $\tilde{}$ operator defined by

$$\tilde{f} = \int_0^\nu (f - \bar{f})\, d\nu', \quad (4.18)$$

where f is any periodic function of ν. Comparison of Eqs. (3.10) and (4.18) shows that $\tilde{f}$ and $\hat{f}$ differ only by an integration constant. The $\tilde{}$ operator has slightly more complicated properties than the $\hat{}$ operator as follows:

$$\tilde{\bar{f}} = 0, \quad \bar{\tilde{f}} \neq 0 \quad (4.19)$$

and

$$\frac{\partial}{\partial \nu}(\tilde{f}) = \left(\overline{\frac{\partial f}{\partial \nu}}\right) = 0. \quad (4.20)$$

However

$$\frac{\partial}{\partial \nu}(\hat{f}) = f - \bar{f} \quad (4.21)$$

and

$$\left(\widetilde{\frac{\partial f}{\partial \nu}}\right) = f(\nu) - f(0), \quad (4.22)$$

where $f(0)$ is f evaluated at $\nu = 0$. One can also show that

$$\overline{\bar{f}g} + \overline{f\tilde{g}} = \bar{f}\bar{g} + \bar{\bar{f}}\bar{g} \quad (4.23)$$

and

$$\widetilde{\bar{f}g} + \widetilde{f\tilde{g}} = \bar{f}\tilde{g} + \tilde{\bar{f}}\bar{g} + \tilde{\bar{f}}\tilde{g}. \quad (4.24)$$

We can now proceed to calculate the required transformation, using Eqs. (4.15)–(4.24);

$$\begin{aligned} \mathbf{Z}_0 &= \mathbf{Y}, \\ \mathbf{H}_0 &= \overline{\mathbf{G}}, \\ \mathbf{Z}_1 &= \int_0^\nu d\nu(\overline{\mathbf{G}} - \mathbf{G}) = -\tilde{\mathbf{G}}, \\ \mathbf{H}_1 &= \frac{1}{\tau} \int_0^\tau d\nu(-\mathbf{G} \cdot \tilde{\mathbf{G}}_{\mathbf{Y}} + \tilde{\mathbf{g}} \cdot \bar{\mathbf{G}}_{\mathbf{y}}). \end{aligned}$$

Adding $\hat{f}\bar{\mathbf{G}}_\nu$, which is zero by (4.20), to the integrand we can write $\mathbf{H}_1$ as

$$\mathbf{H}_1 = -\overline{\mathbf{G} \cdot \tilde{\mathbf{G}}_{\mathbf{Y}}} + \overline{\tilde{\mathbf{G}} \cdot \bar{\mathbf{G}}_{\mathbf{Y}}} = -\overline{\mathbf{G} \cdot \tilde{\mathbf{G}}_{\mathbf{Y}}} + \tilde{\bar{\mathbf{G}}} \cdot \bar{\mathbf{G}}_{\mathbf{Y}}$$

and similarly

$$\mathbf{Z}_2 = \mathbf{G} \cdot \tilde{\mathbf{G}}_{\mathbf{Y}} - \tilde{\bar{\mathbf{G}}} \cdot \bar{\mathbf{G}}_{\mathbf{Y}}.$$

Evidently one can find $\mathbf{Z}$ and $\mathbf{H}$ to any desired order in ϵ. To find the inverse transformation, we start with the equivalent of Eq. (4.9) for $\mathbf{y}$,

$$\mathbf{y}_t = \epsilon\mathbf{g}(\mathbf{z}, \varphi) = \epsilon\mathbf{h} \cdot \mathbf{y}_{\mathbf{z}} + (1 + \epsilon\omega)\mathbf{y}_\varphi$$

and follow exactly the same procedure. As shown by Kruskal, the series and their inverses are unique and the choice of integration constant shows that the series are equal at $\nu = 0$, $\varphi = 0$. To order ϵ^2, therefore, the transformations are

$$\mathbf{Z} = \mathbf{Y} - \epsilon\tilde{\mathbf{G}} + \epsilon^2\widetilde{(\mathbf{G} \cdot \tilde{\mathbf{G}}_{\mathbf{Y}} - \tilde{\bar{\mathbf{G}}} \cdot \bar{\mathbf{G}}_{\mathbf{Y}})} + O(\epsilon^3) \quad (4.25)$$

$$\mathbf{Y} = \mathbf{Z} + \epsilon\tilde{\mathbf{G}} + \epsilon^2\widetilde{(\tilde{\mathbf{G}} \cdot G_{\mathbf{Z}} - \bar{\mathbf{G}} \cdot \tilde{\bar{\mathbf{G}}}_{\mathbf{Z}})} + O(\epsilon^3). \quad (4.26)$$

The equation of motion for $\mathbf{Z}$, $\mathbf{Z}_t = \epsilon\mathbf{H}$, looks different according to which transformation, $\mathbf{Z} \to \mathbf{Y}$ or $\mathbf{Y} \to \mathbf{Z}$, one calculates, but with the aid of Eq.

(4.24) they can both be shown to be

$$\mathbf{Z}_t = \mathbf{a} + \epsilon\bar{\mathbf{G}}(\mathbf{Z}) - \epsilon^2(\overline{\mathbf{G}\cdot\tilde{\mathbf{G}}_{\mathbf{Z}}} - \bar{\tilde{\mathbf{G}}}\cdot\bar{\mathbf{G}}_{\mathbf{Z}}), \quad (4.27)$$

where $\mathbf{a}$ is the vector $(1, 0, 0\cdots 0)$. For the Hamiltonian system (1.1), $\mathbf{Y}$ and $\mathbf{G}$ can be chosen as

$$\mathbf{Y} = (\mathbf{q}; \mathbf{p}),$$
$$\mathbf{G} = (\Omega_{\mathbf{p}}; -\Omega_{\mathbf{q}}).$$

The averaged coordinates $\mathbf{Z}$ are denoted by

$$\mathbf{Z} = (\mathbf{Q}; \mathbf{P}).$$

For any function $S(\mathbf{q}; \mathbf{p})$ the dyadic products of the transformation formulas can be conveniently written in terms of Poisson brackets,

$$\mathbf{G}\cdot S_{\mathbf{Y}} \equiv -\Omega_{\mathbf{q}}\cdot S_{\mathbf{p}} + \Omega_{\mathbf{p}}\cdot S_{\mathbf{q}} \equiv [\Omega, S].$$

The formulas (4.25) and (4.26) can then be written

$$\mathbf{P} = \mathbf{p} + \epsilon\tilde{\Omega}_{\mathbf{q}} - \epsilon^2\overline{[\widetilde{\Omega, \tilde{\Omega}_{\mathbf{q}}}]} + \epsilon^2[\bar{\tilde{\Omega}}, \bar{\Omega}_{\mathbf{q}}] + O(\epsilon^3), \quad (4.28)$$

$$\mathbf{Q} = \mathbf{q} - \epsilon\tilde{\Omega}_{\mathbf{p}} + \epsilon^2\overline{[\widetilde{\Omega, \tilde{\Omega}_{\mathbf{p}}}]} - \epsilon^2[\bar{\tilde{\Omega}}, \bar{\Omega}_{\mathbf{p}}] + O(\epsilon^3), \quad (4.29)$$

$$\mathbf{p} = \mathbf{P} - \epsilon\tilde{\Omega}_{\mathbf{Q}} - \epsilon^2\overline{[\widetilde{\Omega, \Omega_{\mathbf{Q}}}]} + \epsilon^2[\bar{\Omega}, \bar{\tilde{\Omega}}_{\mathbf{Q}}] + O(\epsilon^3), \quad (4.30)$$

$$\mathbf{q} = \mathbf{Q} + \epsilon\tilde{\Omega}_{\mathbf{P}} + \epsilon^2\overline{[\widetilde{\Omega, \Omega_{\mathbf{P}}}]} - \epsilon^2[\bar{\Omega}, \bar{\tilde{\Omega}}_{\mathbf{P}}] + O(\epsilon^3). \quad (4.31)$$

The equations of motion of the averaged variables $\mathbf{P}$ and $\mathbf{Q}$ are

$$\mathbf{P}_t = -\epsilon\bar{\Omega}_{\mathbf{Q}} - \epsilon^2\overline{[\tilde{\Omega}, \Omega_{\mathbf{Q}}]} + \epsilon^2[\bar{\Omega}, \bar{\tilde{\Omega}}_{\mathbf{Q}}] + O(\epsilon^3), \quad (4.32)$$

$$\mathbf{Q}_t = \mathbf{a} + \epsilon\bar{\Omega}_{\mathbf{P}} + \epsilon^2\overline{[\tilde{\Omega}, \Omega_{\mathbf{P}}]} - \epsilon^2[\bar{\Omega}, \bar{\tilde{\Omega}}_{\mathbf{P}}] + O(\epsilon^3). \quad (4.33)$$

C. The Action Integral

It is now possible to evaluate the action integral

$$K = \int_0^\tau p_i\frac{\partial q_i}{\partial Q_1}\,dQ_1 = \tau\overline{p_i\frac{\partial q_i}{\partial Q_1}}. \quad (4.8)$$

The leading terms comes from

$$\begin{aligned}\overline{p_1\frac{\partial q_1}{\partial Q_1}} &= \frac{1}{\tau}\int_0^\tau dQ_1\{P_1 - \epsilon\tilde{\Omega}_{Q_1} - \epsilon^2\overline{[\widetilde{\tilde{\Omega}, \Omega_{Q_1}}]} \\ &\quad + \epsilon^2[\bar{\Omega}, \bar{\tilde{\Omega}}_{Q_1}] + O(\epsilon^3)\} \\ &\quad \times \{1 + \epsilon(\Omega_{P_1} - \bar{\Omega}_{P_1}) + \epsilon^2(F - \bar{F}) + O(\epsilon^3) \\ &= P_1 - \epsilon\bar{\tilde{\Omega}}_{Q_1} - \epsilon^2\overline{\tilde{\Omega}_{Q_1}(\Omega_{P_1} - \bar{\Omega}_{P_1})} \\ &\quad - \epsilon^2\overline{[\widetilde{\Omega, \Omega_{Q_1}}]} + \epsilon^2[\bar{\Omega}\bar{\tilde{\Omega}}_{Q_1}] + O(\epsilon^3).\end{aligned}$$

The second-order term in $\partial q_1/\partial Q_1$ denoted by $(F - \bar{F})$ has zero average and does not affect the result to this order.

The remainder of $\overline{\mathbf{p}\cdot d\mathbf{q}}$ gives

$$\begin{aligned}\overline{p_i\frac{\partial q_i}{\partial Q_1}} &= \frac{1}{\tau}\int_0^\tau dQ_1\{P_i - \epsilon\tilde{\Omega}_{Q_i} + O(\epsilon^2)\} \\ &\quad \times \{\epsilon(\Omega_{P_i} - \bar{\Omega}_{P_i}) + \epsilon^2(F - \bar{F}) + O(\epsilon^3)\}, \\ &= -\epsilon^2\overline{\tilde{\Omega}_{Q_i}(\Omega_{P_i} - \bar{\Omega}_{P_i})} + O(\epsilon^3).\end{aligned}$$

Now

$$\overline{\tilde{\Omega}_{Q_i}(\Omega_{P_i} - \bar{\Omega}_{P_i})} = \overline{(\tilde{\Omega}_{Q_i} - \bar{\tilde{\Omega}}_{Q_i})(\Omega_{P_i} - \bar{\Omega}_{P_i})},$$

from (4.19)

$$= -\overline{(\Omega_{Q_i} - \bar{\Omega}_{Q_i})(\tilde{\Omega}_{P_i} - \bar{\tilde{\Omega}}_{P_i})},$$

using (4.23)

$$= \tfrac{1}{2}\overline{[\widetilde{(\Omega - \bar{\Omega})}, (\Omega - \bar{\Omega})]},$$
$$= \tfrac{1}{2}[\bar{\Omega}, \Omega] - \tfrac{1}{2}[\bar{\tilde{\Omega}}, \bar{\Omega}].$$

Also,

$$\overline{[\widetilde{\tilde{\Omega}, \Omega_{Q_i}}]} - [\Omega, \bar{\tilde{\Omega}}_{Q_i}] = -\overline{[\widetilde{\Omega, \tilde{\Omega}_{Q_i}}]} - \overline{[\tilde{\Omega}, \tilde{\Omega}_{Q_i}]},$$

using (4.24)

$$= \overline{[\widetilde{\Omega, \Omega - \Omega(0)}]} - \overline{[\tilde{\Omega}, \Omega - \Omega(0)]}$$
$$= [\bar{\Omega}, \Omega(0)] - [\bar{\Omega}, \Omega] + \overline{[\tilde{\Omega}, \Omega(0)]},$$
$$= -\overline{[\tilde{\Omega}, \Omega]}.$$

As a result,

$$\overline{p_i\frac{\partial q_i}{\partial Q_1}} = P_1 - \epsilon\bar{\tilde{\Omega}}_{Q_1} + \frac{\epsilon^2}{2}[\bar{\Omega}, \tilde{\Omega}] + \frac{\epsilon^2}{2}\overline{[\Omega, \tilde{\Omega}]} + O(\epsilon^3).$$

Transforming back to coordinates $\mathbf{p}$ and $\mathbf{q}$, using (4.28) and (4.29), and transforming from $\tilde{\ }$ to $\hat{\ }$ operators, using

$$\tilde{\Omega} = \hat{\Omega} - \hat{\Omega}(0),$$

one finally obtains

$$\overline{p_i\frac{\partial q_i}{\partial Q_1}} = p_1 + \epsilon(\Omega - \bar{\Omega}) - \epsilon^2[\bar{\Omega}, \hat{\Omega}] + \frac{\epsilon^2}{2}\overline{[\Omega, \hat{\Omega}]} + O(\epsilon^3). \quad (4.34)$$

One then has

$$\begin{aligned}K &= \overline{\tau p_i\frac{\partial q_i}{\partial Q_1}} \\ &= \tau H - \epsilon\tau\bar{\Omega} - \epsilon^2\tau([\bar{\Omega}, \hat{\Omega}] - \tfrac{1}{2}\overline{[\Omega, \hat{\Omega}]})\cdots.\end{aligned}$$

This is obviously related to the invariant J calculated from the Poisson bracket method, by

$$K = \tau H - \epsilon\tau J \quad (4.35)$$

at least to order ϵ^2.

The advantages of the Poisson bracket method is that the first few terms of the invariant can easily be found. The advantages of Kruskal's method are that it is systematic (although the task of calculating higher-order terms would be arduous) and that one shows

that the invariant is an asymptotic invariant. The correspondence of the two methods to terms of order ϵ^2 suggests that the Poisson bracket method generates a true asymptotic series. Another advantage of Kruskal's method is that one obtains some physical understanding of what the invariant is. In a periodic dynamical system with an ignorable angle variable, the corresponding action is a constant of the motion. In the present system which is nearly periodic there is an ignorable (in an asymptotic sense) coordinate φ. The invariant K is the action corresponding to this almost ignorable angle variable.

D. Reduction to a System of Lower Dimensions

Having found a constant of the motion, it is well known that new canonical coordinates can be found in which the constant is one of the new momenta. In fact Kruskal shows that new coordinates can be found of which K and Q_1 are a conjugate pair. For the particular systems considered in this paper such a set of coordinates is easily obtained. Although $(\mathbf{Q};\mathbf{P})$ are not canonical coordinates, the set of quantities with K replacing P_1 are. This can be demonstrated directly by transforming the Hamiltonian to obtain

$$H = K + \epsilon\bar{\Omega}(Q_2, \cdots, Q_n; K, P_2, \cdots, P_n) + O(\epsilon^2). \tag{4.36}$$

The equations of motion, (4.32) and (4.33), are of Hamiltonian form with the above Hamiltonian.

Another method of finding new canonical coordinates $(\mathbf{Q}';\mathbf{P}')$ is to solve the Poisson bracket relations

$$\begin{aligned}[P_i', P_j'] &= [Q_i', Q_j'] = 0,\\ [P_i', Q_j] &= \delta_{ij},\end{aligned} \tag{4.37}$$

where P_1' is chosen equal to K. Equations (4.37) have one set of solutions

$$P_i' = p_i + \epsilon\left(\hat{\Omega}_{q_i} + \frac{\partial S}{\partial q_i}\right) + O(\epsilon^2), \quad i \neq 1,$$

$$Q_i' = q_i - \epsilon\left(\hat{\Omega}_{p_i} + \frac{\partial S}{\partial p_i}\right) + O(\epsilon^2),$$

where S is any function independent of q_1. There exists therefore considerable freedom in finding new canonical coordinates. If the particular choice

$$S = -\hat{\Omega}(q_1 = 0)$$

is made, the coordinates are identical with the set mentioned in the previous paragraph since

$$\bar{\Omega} = \hat{\Omega} - \hat{\Omega}(q_1 = 0).$$

As in Sec. 3B, the only one of these transformations that can be readily inverted is the one obtained from Kruskal's averaging method.

In the reduced Hamiltonian (4.37), K appears merely as a parameter and the system has been effectively reduced to $n-1$ dimensions. If this Hamiltonian contains a further small parameter δ, such that for $\delta = 0$ the motion is once more periodic, then another adiabatic invariant exists and can be calculated by the methods of Secs. 3 and 4. It is worth remarking that these other periodicities may not appear in the original Hamiltonian, only being created by the construction of the first invariant and the subsequent transformation of the Hamiltonian. This behavior occurs in plasma physics where firstly the magnetic moment can be found and subsequently the longitudinal and flux invariants. A recent derivation of these from Liouville's equation but not using canonical formalism has been given by Hastie, Haas, and Taylor.[11]

5. WHITTAKER'S ADELPHIC INTEGRAL

A dynamical system of the type under consideration is the motion of a particle near a point of stable equilibrium. The Hamiltonian can be written (Whittaker, Ref. 6, Chap. XVI) in the form

$$H = \sum_{n=2}^{\infty} \epsilon^n H_n, \tag{5.1}$$

where H_n is a homogeneous polynomial of a degree n in the canonical coordinates $(q_i'; p_i')$ with $i = 1, \cdots, N$. In particular, H_2 is of the form (2.5) so that to lowest order the system can be either periodic or conditionally periodic. Invariants of such systems have been studied extensively[1-6] and Contopoulos[12] has shown that the various treatments are equivalent. Whittaker's method is to change to new variables q_i'', p_i'' that reduce the Hamiltonian to

$$H = \sum_{r=1}^{N} (\alpha_r p_r'') + H_3 + H_4 \cdots, \tag{5.2}$$

where the α_r are the periods of the lowest-order simple harmonic motions. H_n is a homogeneous polynomial of degree n in $(p_i)^{\frac{1}{2}}$ and also a function of $\sin q_i''$ and $\cos q_i''$. One looks for a time-independent constant of the motion, Φ, and for simplicity only a two-dimensional system is considered. Φ satisfies the equation

$$[\Phi, H] = 0, \tag{5.3}$$

which when expanded gives successively

$$\alpha_1 \frac{\partial \Phi_2}{\partial q_1''} + \alpha_2 \frac{\partial \Phi_2}{\partial q_2''} = 0, \tag{5.4}$$

$$\alpha_1 \frac{\partial \Phi_3}{\partial q_1''} + \alpha_2 \frac{\partial \Phi_3}{\partial q_2''} = [\Phi_2, H_3], \text{ etc.} \tag{5.5}$$

[11] R. J. Hastie, F. A. Haas, and J. B. Taylor, Ann. Phys. **41**, 2 (1967).

[12] G. Contopoulos, Astron. J. **68**, 1 (1963).

To find a solution Φ different from H, one can take Φ_2 to be any function of p_1'' and p_2'' independent of H_2 (for example Whittaker took $\Phi_2 = \alpha_1 p_1'' - \alpha_2 p_2''$) and the Φ_n can be found as particular integrals of the successive equations. This differs from the method of Sec. 3 because Φ_2 is chosen at this point merely to be different from H_2 rather than being determined by the periodicity condition.

A feature of this solution is that in the series one finds coefficients of the form

$$(\alpha_1 l \pm \alpha_2 m)^{-1},$$

where l and m are integers, the range of possible values of which increases as the order of the term n increases. This means that, with α_1/α_2 irrational and for sufficiently large n, there are in the series coefficients that are arbitrarily large. When α_1/α_2 is rational, the problem is even more acute, since in the formal series for Φ there appear some coefficients with vanishing denominators. Whittaker overcame this difficulty using two techniques. In solving for Φ_n he had selected only a particular integral, but could equally well have added an arbitrary amount of the solution of the homogeneous equation. In some cases it proves possible to avoid zero denominators by a suitable choice of these constants in the lower order terms. For others, however, it is necessary to renormalize the series solution for Φ by, effectively, multiplying by the zero denominator, the new form for Φ now starting with the term that previously had the troublesome coefficient.

The problem with these methods is that it is not clear at any stage whether further difficulties will arise in higher order terms. There is also the curious point that the formal expression for the adelphic integral is different depending on whether α_1/α_2 is rational or irrational and also whether terms of a particular type appear in the Hamiltonian. All of these difficulties arise from the particular choice of Φ_2.

The Hamiltonian (5.2) can be written in standard form (1.1) using the transformation

$$\begin{aligned} p_1 &= \alpha_1 p_1'' + \alpha_2 p_2'', \quad q_1 = \tfrac{1}{2}(q_1''/\alpha_1 + q_2''/\alpha_2), \\ p_2 &= \alpha_1 p_1'' - \alpha_2 p_2'', \quad q_2 = \tfrac{1}{2}(q_1''/\alpha_1 - q_2''/\alpha_2), \end{aligned} \tag{5.6}$$

and the invariant J of Sec. 3 calculated. The question arises as to whether the adelphic integral and J are equivalent or not. The main disparity is that the lowest-order term of the adelphic integral is a function only of p_1 and p_2 (except in some special circumstances associated with rational values of α_1/α_2), whereas the lowest-order term in J is apparently only independent of q_1. However one is able to show that the first term in the series for J is, in general, only a function of p_1 and p_2 (depending on q_2, only when α_1/α_2 is rational). The way in which this occurs is as follows; Ω consists of a number of terms like

$$f_{mn} \begin{matrix} \cos \\ \sin \end{matrix} \{(n\alpha_1 + m\alpha_2)q_1 + (n\alpha_1 - m\alpha_2)q_2\}, \tag{5.7}$$

where f_{mn} is a function of p_1 and p_2 only and m and n take a range of integral values. The lowest-order term in J is $\bar{\Omega}$ which is only nonzero if there are values of m and n such that

$$\alpha_1/\alpha_2 = -m/n$$

and then

$$\bar{\Omega} = f_{mn} \cos \{(n\alpha_1 - m\alpha_2)q_2\}. \tag{5.8}$$

Except under these special circumstances $\bar{\Omega} = 0$ and the lowest order term is $J_1 = \frac{1}{2}\overline{(\Omega, \bar{\Omega})}$. This quantity consists of terms that are averages over q_1 of products of two expressions like (5.7) and leads to integrals of the form

$$I = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau dq_1 f \cos \{((n-k)\alpha_1 + (m-l)\alpha_2)q_1 + ((n-k)\alpha_1 - (m-l)\alpha_2)q_2\}, \tag{5.9}$$

where f is a function of p_1 and p_2 only, and n, m, k, and l take various integral values. When $n = k$ and $m = l$, this integral is equal to f and there always occur some terms of this form in J_1. The only way for I to be nonzero and to depend on q_2 is for

$$(l - m)/(n - k) = \alpha_1/\alpha_2$$

(i.e., for α_1/α_2 to have a particular rational value).

So, apart from these special rational cases, both H and J have lowest-order terms that are functions of p_1 and p_2 only. It is possible therefore to construct another invariant (a function of J and H) that has as its lowest-order term any function of p_1 and p_2 whatsoever. This in fact is what the adelphic integral is. In the three examples quoted by Whittaker (Ref. 6, Chap. XVI), one can show that the adelphic integral is a function of H and J. Examples worked out by Contopoulos and Moutsoulas[5] are also equivalent to J in the same way.

The advantage of the methods of Secs. 3 and 4 is that they overcome all but one of the difficulties encountered by Whittaker. The problem of vanishing denominators does not arise at all. A single expression is obtained for the invariant irrespective of the value of α_1/α_2 and, for the rational case at least, one can show that the series is asymptotically invariant. The problem of small denominators when α_1/α_2 is irrational still occurs however, and it is this fact that prevents one proving that the formal series is asymptotically invariant for all values of α_1/α_2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Results on Certain Non-Hermitian Hamiltonians*

JACK WONG
Lawrence Radiation Laboratory, University of California, Livermore, California

(Received 26 September 1966)

We present a few results on the spectral properties of a class of physically reasonable non-Hermitian Hamiltonians. These theorems relate the spectral properties of a non-self-adjoint operator (of the aforementioned class) in terms of that of a self-adjoint operator. These theorems can be specialized to yield conditions under which the perturbed eigenvalues (of the above class of operators) vary continuously from the eigenvalues of the unperturbed operators. If the Schrödinger equation has to be solved numerically, a knowledge of the spectral properties of the non-Hermitian Hamiltonian would insure when the eigensolutions exist.

1. INTRODUCTION

IN principle, only completely self-contained physical systems are described by Hermitian Hamiltonians, but in practice many physical systems are well approximated as completely isolated. However, one also encounters quantum-mechanical systems that are intimately coupled to the environment. For example, if one is interested in the properties of a hydrogen "atom" in an interacting medium, then the dissipative nature of the medium is appreciable and must be taken into account.[1] One then is immediately confronted with a nonconservative quantum-mechanical system whose Hamiltonian is not self-adjoint. In the literature there exist many examples of non-Hermitian description of physical problems ranging from optical potential model, the theory of α decay of the nuclei and the nuclear reaction theory to nuclear pile theory.

The purpose of this article is to bring attention to a few mathematical results on a fairly broad class of physically reasonable, non-Hermitian Hamiltonians. For convenience, this class of non-Hermitian operators will be called *dissipative* (to be defined) in accord with physical intuition. The mathematical theorems (to be given later) provide important information on the spectrum of a non-Hermitian operator H ($\equiv H_0 + H'$) in terms of the spectrum of a Hermitian subpart H_0; and furthermore the continuous part of the spectrum of H coincides with that of H_0. When the theorems are specialized to perturbation theory, they provide justification (for this class of operators) for the usual perturbation assumption that the eigenvalues of $H \equiv H_0 + gH'$ (H_0 is Hermitian) vary continuously toward that of H_0 as the parameter g goes to zero. Moreover, for operator H satisfying the conditions of the theorems, the sum of the spectral multiplicities of the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ of H ($= H_0 + gH'$) is the same as that of λ_0 of H_0 if $\lambda_i \to \lambda_0$ ($i = 1, 2, \cdots, n$) as $g \to 0$. All this is important mathematical information for physicists who are concerned with the kind of non-Hermitian (or Hermitian) Hamiltonians defined in this article. It is often desirable to solve the full non-Hermitian Hamiltonian Schroedinger equation rather than just the Hermitian Schroedinger equation plus a perturbation part. We list two reasons for this: (1) if the "perturbation" is not small, then clearly, perturbation theory is not even useful, and (2) if the problem can only be solved numerically, then a qualitative knowledge of the eigenspectrum deduced from the mathematical theorems assures one of the existence or nonexistence of the eigensolutions.

It is emphasized that the results of this article are restricted to a special class of non-Hermitian operators, and the proofs are physicists' proofs. The general problem is much too hard and the mathematical facts about non-Hermitian operators are too little known.

Section II contains all the mathematical facts of this article. Section III contains a brief summary.

2. SPECTRAL PROPERTIES OF DISSIPATIVE OPERATORS

An operator H is defined to be linear *dissipative* if the imaginary part of all its eigenvalues λ_n ($n = 1, 2, \cdots$) is less than or equal to zero. If the equality sign holds for all n, then the spectrum is real and H is assumed to be self-adjoint. The adjective "linear dissipative" has its origin in the expression $\exp(-i\lambda_n t)$ which decreases exponentially as the time t becomes large and positive. Actually, what we are going to prove holds true for operators with point spectrum anywhere in the finite complex plane. But with physical systems in mind, and for lack of a more descriptive terminology to designate the class of non-Hermitian operators to be considered in this article, we shall continue to use the term *dissipative* for this class in the sequel. We shall study only those linear dissipative operators H which can be written as a sum of two operators, $H = H_0 + H_1$. It is assumed that (1) the operator H_0 is self-adjoint, (2) the operator H_1

* This work was performed under the auspices of the U.S. Atomic Energy Commission.

[1] T. Nakayama and H. DeWitt, J. Quant. Spectry. Radiative Transfer 4, 623 (1964).

is bounded and (3) the operator $(H_1)^{\frac{1}{2}}R_\lambda\,(H_1)^{\frac{1}{2}}$ is compact[2] (or equivalently completely continuous) where $R_\lambda[=(H_0-\lambda I)^{-1}]$ is the resolvent operator of H_0 for the regular value λ. The operator H_0 is required to be self-adjoint because the spectral properties of self-adjoint operators are well-understood. The operator $(H_1)^{\frac{1}{2}}R_\lambda\,(H_1)^{\frac{1}{2}}$ is required to be compact because a compact operator has only a discrete point spectrum.[2] It turns out that the spectral properties of H bear close resemblance to that of H_0 due to the above restrictions on H_1. The close resemblance of the spectral properties between the operators H and H_0 enables us to deduce useful spectral information on H from that of H_0, in spite of the general lack of understanding concerning non-self-adjoint operators.

In the remainder of this section, we put down the mathematical statements of this article. We only give references to those proofs that are readily available in the published literature. However, we sketch the proofs of those theorems that are not readily accessible.

Lemma 1[2]*:* The spectrum $\sigma(T)$ of a compact (completely continuous) operator T is at most denumerable and has no point of accumulation in the complex plane, except possibly at the origin. Each nonzero point of the spectrum $\sigma(T)$ has finite multiplicity.

Lemma 2[3]*:* Let $T(\lambda)$ be a compact operator and be analytic in the parameter λ in a region G, designated as $\lambda \in G$. Then for all complex $\lambda \in G$, the dimension of the null space of the operator $\mathrm{I} - T(\lambda)$ is a constant α (notice the interesting case $\alpha = 0$) except for *isolated* points at which this dimension may be higher.

First, the null space of an operator T is the collection of all those vectors $\mathbf{x}$ such that $T\mathbf{x} = 0$. Second, for α to be zero for almost all $\lambda \in G$, it suffices to show that $\alpha = 0$ at one point of G.

Theorem A[3,4]*:* We assume that: (1) the self-adjoint operator H_0 is defined on the domain $D(H_0)$, dense everywhere in the Hilbert space L_2. (2) Its resolvent $R = (H_0 - \lambda I)^{-1}$ (λ is said to be in the resolvent set $\rho(H_0)$ of H_0) is given as an integral operator,

$$(R_\lambda f)(x) \equiv \int M(x, y; \lambda) f(y)\, dy. \tag{2.1}$$

(3) The operator S is a bounded operator with domain $D(S) = L_2$; and (4) the operator $SR_\lambda S$ is compact for $\lambda \in \rho(H_0)$. Then, all points of the continuous spectrum $\sigma_c(H_0)$ of the operator H_0 belong to the spectrum of the operator $H \equiv H_0 + S^2[D(H) = D(H_0)]$; and every point outside the spectrum $\sigma(H_0)$ of H_0 is either a regular point of H or a point, in the point spectrum $\sigma_p(H)$ of H, of finite multiplicity.

Remarks: This theorem is essentially due to Martirosyan.[4] The general ideas that such a theorem may be true can be found in Refs. 3, 5, and 6 where, e.g., the operator S^2 is assumed to be compact instead of $SR_\lambda S$ as in Martirosyan's paper. In applying this theorem to the Schroedinger equation, the former assumption falls short of a physicist's needs, because, e.g., a multiplicative function operator such as the potential $V(\mathbf{x})$ ($\equiv S^2$) is *not* compact.

Proof: We will first prove that *if* for some λ, there exists a bounded resolvent operator $B_\lambda \equiv (H - \lambda 1)^{-1}$ of the operator H, then the operator $SB_\lambda S$ is compact. It suffices to show that [assuming $\lambda \in \rho(H_0)$]

$$B_\lambda = R_\lambda - R_\lambda S(I + SR_\lambda S)^{-1} SR_\lambda, \tag{2.2}$$

where the operator $(I + SR_\lambda S)^{-1}$ is defined on all of L_2 and is bounded. Let B_λ exist for some λ, then the equation $(H - \lambda)u = (H_0 + S^2 - \lambda)u = 0$, and consequently the equation $u + R_\lambda S^2 u = 0$ has only a trival solution. Therefore $\phi + (SR_\lambda S)\phi = 0$ has also only a trivial solution. By assumption, the operator $SR_\lambda S$ is compact. Hence, from the Fredholm alternative,[5] it follows that there exists a bounded operator $(I + SR_\lambda S)^{-1}$, defined on all of L_2. The operator S is assumed to be bounded and the norm $\|R_\lambda\|$ of the resolvent operator R_λ is bounded by the inverse of $\mathbf{d}_\lambda$ [the shortest distance from λ to the spectrum $\sigma(H_0)$].[7] Without loss of generality, we can choose $\mathbf{d}_\lambda$ so large such that $\|SR_\lambda S\| \leq \|S\|^2\,\|R_\lambda\| \leq \|S\|^2\,\mathbf{d}_\lambda^{-1} < 1$. Then Eq. (2.2) can be verified as an identity by series expansion of $(I + SR_\lambda S)^{-1}$. This identity equation (2.2) remains valid for all other values of λ reached by analytic continuation. Multiplying both sides of Eq. (2.2) by the operator S, we obtain $SB_\lambda S = SR_\lambda S - (SR_\lambda S)(I + SR_\lambda S)^{-1}(SR_\lambda S)$. Since $(I + SR_\lambda S)^{-1}$ is bounded and $(SR_\lambda S)$ is compact, their product is compact.[8]

To prove the first assertion of the theorem, we assume that λ belongs to the continuous spectrum of the operator $H_0[\lambda \in \sigma_c(H_0)]$, but is a regular point of the operator $H[\lambda \in \rho(H)]$. We arrive at a

[2] N. Dunford and J. T. Schwartz, *Linear Operator Part 1* (Interscience Publishers, Inc., New York and London, 1958), p. 549.

[3] I. C. Gohberg and M. G. Krein, Am. Math. Soc. Transl. Ser. 2 13, 206 (1960).

[4] R. M. Martirosyan, Izv. Akad. Nauk Arm. SSR, Ser. Fiz.-Mat. Nauk, **14**, No. 5, 9 (1961).

[5] F. Riesz and B. Sz-Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955), 2nd ed.

[6] C. L. Dolph and F. Penzlin, Ann. Acad. Sci. Fennicae No. 263 (1959).

[7] See Ref. 3, Eq. (5.5).

[8] See Ref. 5, p. 178.

contradiction. It is evident that the equation

$$(H_0 - \lambda)u = (H - \lambda - S^2)u = 0, \qquad (2.3)$$

is equivalent to the equation

$$(I - B_\lambda S^2)u = 0. \qquad (2.4)$$

The function u cannot be an L_2-solution of Eq. (2.4), otherwise u is an L_2 solution of Eq. (2.3), contrary to the assumption that $\lambda \in \sigma_c(H_0)$. Equation (2.4), $(I - B_\lambda S^2)u = 0$ having only a trivial solution implies that the equation $(I - SB_\lambda S)\phi = 0$ also has only a trivial solution ϕ. As we have just shown, $SB_\lambda S$ is compact. Hence, the Fredholm alternative implies that $(I - SB_\lambda S)^{-1}$ exists and is bounded. But this would imply that the resolvent $R_\lambda = (H_0 - \lambda I)^{-1}$ can be defined by the following identity,

$$R_\lambda = B_\lambda S(I - SB_\lambda S)^{-1}SB_\lambda + B_\lambda;$$

this is contrary to the assumption that $\lambda \in \sigma_c(H_0)$. This proves the first statement of the theorem.

To prove the second statement of the theorem, we start with a point λ not in the spectrum of $H_0[\lambda \in \rho(H_0)]$. The operator $(I + SR_\lambda S)^{-1}$ either exists and is bounded or is not defined. For the first possibility, λ is a regular point of the operator H because the right-hand side of Eq. (2.2) is analytic in λ. Now by analytic continuation other λ values can be reached. If the point λ is such that $(I + SR_\lambda S)^{-1}$ does not exist then by Lemma 2, λ is an isolated point of the spectrum of H. Furthermore, this eigenvalue is of finite multiplicity because of Lemma 1. This proves the second statement of the theorem.

Corollary 1: If λ is an eigenvalue of the operator H, then the complex conjugate of $\lambda(\lambda^*)$ is an eigenvalue of the adjoint of $H(H^+)$.

Proof: Let λ be an eigenvalue of the operator H, then the operator $(I + SR_\lambda S)^{-1}$ is not defined. Therefore the adjoint of this operator $(I + S^+R_{\lambda^*}S^+)^{-1}$ is not defined. But the adjoint of a compact operator (in this case $SR_\lambda S$) is compact.[9] Now retracing the proof of the theorem and taking the adjoint of Eq. (2.2) one can readily show that the resolvent of H^+, $(H_0 + (S^2)^+ - \lambda^*)^{-1}$ is singular at λ^*.

Corollary 2: Let the self-adjoint operator H_0 have the following representation:

$$(H_0u)(\mathbf{x}) = \sum_{i,j=1}^{3} \frac{\partial}{\partial x_i}\left[a_{ij}(\mathbf{x})\frac{\partial u(\mathbf{x})}{\partial x_j}\right] + V(\mathbf{x})u(\mathbf{x}), \quad (a_{ij} = a_{ji}) \qquad (2.5)$$

where $a_{ij}(\mathbf{x})$ are appropriately differentiable functions. Furthermore, let S^2 be represented as a *bounded*, complex-valued function $C(\mathbf{x})$ belonging to L_1. Then the two statements of Theorem A are true.

[9] See Ref. 5, p. 217, footnote 12.

Proof: The resolvent $R_\lambda[\lambda \in \rho(H_0)]$ of a self-adjoint operator of the representation equation (2.5) is an integral operator with a Carlemann type kernel[10] $M(x, y; \lambda)$ [see Eq. (2.1)]. The kernel $M(x, y; \lambda)$ being Carlemann type possesses the following property: If $g(x)$ is defined as

$$g(x) \equiv \int |M(x, y; \lambda)|^2\, dy, \qquad (2.6)$$

then $g(x)$ is not bounded *at most* at a point set of measure zero. But this implies that the operator $(C)^{\frac{1}{2}} \cdot R_\lambda \cdot (C)^{\frac{1}{2}}$ is square-integrable,

$$\int |C(x)^{\frac{1}{2}}M(x, y; \lambda)C(y)^{\frac{1}{2}}|^2\, dx\, dy \leq \left(\max_y |C(y)|\right) \cdot \int |C(x)| \left\{\int |M(x, y; \lambda)|^2\, dy\right\} dx < \infty,$$

since $C(x) \in L_1$. The fact that the operator $(C)^{\frac{1}{2}} \cdot R_\lambda \cdot (C)^{\frac{1}{2}}$ is an L_2 kernel implies that $(C)^{\frac{1}{2}} \cdot R_\lambda \cdot (C)^{\frac{1}{2}}$ is also compact; this follows from the fact that an L_2-kernel can be uniformly approximated in the mean by a kernel of finite rank which ensures compactness.[11] Compactness of the operator $(C)^{\frac{1}{2}} \cdot R_\lambda \cdot (C)^{\frac{1}{2}}$ is just the requirement of Theorem A.

Remarks: This corollary is true for weaker restrictions on the function $C(\mathbf{x})$. Instead of $C(\mathbf{x}) \in L_1$, it is enough that $C(\mathbf{x})$ goes to zero as $|\mathbf{x}|$ goes to infinity. But this refinement entails too much mathematical analysis. Hence, we do not wish to pursue it further here.

Theorem B[12]*:* Let H be a self-adjoint operator, let B be a bounded operator and let λ_0 be some eigenvalue of finite multiplicity ν_0 of the operator H_0 whose distance d from the rest of the spectrum $\sigma(H_0)$ is $d > 2\,\|B\|$. Then in the closure of the circle $|\lambda - \lambda_0| < \|B\|$, the operator $H_0 + B$ has a finite number of eigenvalues and the sum of their respective spectral multiplicities is exactly ν_0.

Remarks: The content of this theorem is very useful to those doing perturbation calculation on a quantum-mechanical system and wishing for information on the perturbed point spectrum. As far as this article is concerned, we only quote it as a plausibility argument for the following restriction on our Hamiltonian H: This article deals only with that class of physical systems whose Hamiltonians $H = H_0 + S^2$ satisfy

[10] M. Sh. Flekser, The spectral function of the operator

$$Bu = -\sum_{ik=1}^{3}(\partial/\partial x_i)[A_{ik}(\mathbf{x})(\partial u/\partial x_k)] + C(\mathbf{x})u,$$

[Mat. Sb. **40** (82),No. 1 (1956)].

[11] See, for e.g., B. Friedman, *Principles and Techniques of Applied Mathematics* (John Wiley & Sons, Inc., London, 1956), p. 39.

[12] Reference 3, Theorem 5.2, p. 29.

Theorem A, and in addition every point outside the continuous spectrum of H_0 is an eigenvalue of *finite* multiplicity of the Hamiltonian H. It should be remembered that Theorem A says: "... every point *outside* $\sigma(H_0)$ is either a regular point of H or is in the point spectrum $\sigma_p(H)$ of H." That is why we need Theorem B to insure that there are only a finite number of points of $\sigma_p(H)$ in a neighborhood of a point of $\sigma_p(H_0)$, if B (of $H = H_0 + B$) is sufficiently small. That the class of Hamiltonians H satisfying Theorems A and B is not empty can be illustrated by the following example. Our example would be an $H = H_0 + S^2$ having the following properties: (1) H satisfies the conditions of Theorem A, (2) H_0 corresponds to a Schrödinger equation with a Yukawa potential; (3) $\|S\|$ is sufficiently small.

Now we turn briefly to the question of completeness of the eigensolutions of any dissipative operator H. In a recent article Fonda *et al.*[13] demonstrated that the set of eigensolutions $\{\psi_n\}$ of H is complete if the associated resolvent operator $B_\lambda(H)$ satisfies the following three assumptions:

(a) There exists a dense set D of functions in the Hilbert space such that if ψ and ϕ belong to D, then $(\phi, B_\lambda\psi)$ is an analytic function of λ with simple poles and a branch cut lying along the *real* λ axis, such that

$$\lim_{\lambda\to\lambda_i}(\lambda - \lambda_i)(\phi, B_\lambda\psi) = -(\phi, P_i\psi), \qquad (i = 1, 2, \cdots, n)$$

$$\lim_{\lambda\to E+i0}(\phi, B_\lambda\psi) - \lim_{\lambda\to E-i0}(\phi B_\lambda\psi) = 2\pi i(\phi, P_E\psi),$$

where λ_i and E belong to the point and the continuous spectrum, respectively. The P_i and P_E are the related projection operators.

(b) There is no accumulation point in the point spectrum $\sigma_p(H)$ and no point in $\sigma_p(H)$ lies in the continuous spectrum $\sigma_c(H)$. Thus, the number of points in $\sigma_p(H)$ is finite.

(c) The function $(\phi, B_\lambda(H)\psi) \to (\phi, I\psi)\lambda^{-1}$ as $|\lambda| \to \infty$ for all λ in the resolvent set $\rho(H)$.

For dissipative operator H of this article, the continuous spectrum is demonstrated to lie within the real λ axis. However, the poles of B_λ corresponding to the point spectrum of H need not be simple. Therefore, we make such an assumption here also. For a finite matrix H, the resolvent has simple poles implies that H can be diagonalized; otherwise H can best be brought into a Jordan canonical form. Under assumption (b), we have to assume that the self-adjoint subpart H_0 (of $H = H_0 + S^2$) has a finite set of points in $\sigma_p(H_0)$. Then by Theorem B, $\sigma_p(H)$ is a finite set.

The asymptotic property of the resolvent $B_\lambda(H)$ of H [assumption (c)] follows from the asymptotic property of $R_\lambda(H_0)$, where H_0 is self-adjoint. This is easily deduced from Eq. (2.2). The above brief discussion indicates what are and what are not assumed in writing down the completeness relation for our class of non-Hermitian operators. Following Fonda *et al.*,[13] we write the following completeness relation:

$$\sum_{\lambda,\mu(\lambda)} |\psi_{\lambda,\mu(\lambda)}\rangle\langle\phi_{\lambda,\mu(\lambda)}| + \int dE\, |\psi_E^+\rangle\langle\phi_E^+| = I, \quad (2.7)$$

where λ sums over the point spectrum $\sigma_p(H)$ and $\mu(\lambda)$ sums over the finite multiplicity of the eigenvalue λ. It should be noticed that ϕ_λ is an eigensolution of the adjoint of H, namely H^+, with eigenvalue λ^* (the complex conjugate of λ),

$$H^+\phi_\lambda = \lambda^*\phi_\lambda. \quad (2.8)$$

In writing down Eq. (2.7) we are aided by Corollary 1 of Theorem A. It also follows from this Corollary that we have biorthonormality between the two complete sets of eigensolutions $\{\phi_\mu\}$ and $\{\psi_\lambda\}$,

$$(\phi_\mu, \psi_\lambda) = \delta_{\mu\lambda}. \quad (2.9)$$

3. SUMMARY

Before concluding, we would like to briefly recapitulate the mathematical results of Sec. II. Theorem A states that the spectrum of the operator $H\,(= H_0 + S^2)$ resembles closely the spectrum of H_0, where S^2 is a bounded operator and $SR_\lambda S$ is compact. Corollary 1 states that if λ_n is an eigenvalue of H, then λ_n^* is an eigenvalue of the adjoint H^+. Corollary 1 results in a relatively simple completeness relation equation (2.7) and in biorthonormality equation (2.9). Corollary 2 specializes the bounded operator S^2 of Theorem A to a multiplicative bounded L_1-function such that direct application can be made to quantum-mechanical problems. Theorem B says that if the perturbation operator B to H_0 is sufficiently small, then the spectral multiplicity of an eigenvalue λ_0 of H_0 is the same as the sum of the spectral multiplicities of all the eigenvalues of $H\,(= H_0 + B)$ in a neighborhood of λ_0.

ACKNOWLEDGMENT

I would like to thank several colleagues for comments on the manuscript. I would also like to acknowledge my particular indebtedness to Dr. J. C. Garrison for earlier collaboration on some aspects of the bound state problem in an interacting medium. This early collaboration is germane to the present work.

[13] L. Fonda, G. C. Ghirardi, T. Weber, and A. Rimini, J. Math. Phys. 7, 1643 (1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Further Exact Solutions of the Dirac Equation*

GEORGE N. STANCIU
University of California, Los Alamos Scientific Laboratory, Los Alamos, New Mexico

(Received 1 February 1967)

The four exactly solvable problems for the Dirac equation previously found by the author are discussed in detail. The presentation is mainly within the two-component relativistic description of a spin $\frac{1}{2}$ particle. Both bound state and scattering solutions are considered.

I. INTRODUCTION

A CAREFUL search of the literature has revealed only six configurations of external electromagnetic fields for which the Dirac equation is exactly solvable. They are: the absence of all fields (free particle), a Coulomb potential,[1] a constant magnetic field,[2] a constant electric field,[3] the field of a plane wave,[4] and the field of a plane wave with a constant magnetic field pointing along the direction of propagation of the plane wave.[5] Recently, the author has found four additional exact solutions of the Dirac equation.[6] The subject of this paper is a more detailed presentation of these solutions.

The first problem encountered in any attempt to solve the Dirac equation is one of algebraic complexity. The Dirac equation is a compact way of writing four coupled differential equations for the four components that comprise the Dirac spinor. The uncoupling of the differential equations is an algebraic problem that can be minimized by adopting a two-component formulism. An electron, $e > 0$, in an external electromagnetic field is described by the relativistically invariant two-component equation,[7,8]

$$[(\mathbf{P} + e\mathbf{A})^2 + m^2 + e\boldsymbol{\sigma} \cdot (\mathbf{H} \dot{+} i\mathbf{E})]\Psi = (W + eV)^2\Psi. \quad (1)$$

The σ's are the ordinary Pauli matrices. The units are $\hbar = c = 1$.

The four-component spinors which are solutions of the Dirac equation are generated from the solutions of the two-component equation by

$$\psi_D = \begin{pmatrix} [\boldsymbol{\sigma} \cdot (\mathbf{P} + e\mathbf{A}) + (W + eV) + m]\Psi \\ [\boldsymbol{\sigma} \cdot (\mathbf{P} + e\mathbf{A}) + (W + eV) - m]\Psi \end{pmatrix}. \quad (2)$$

It can easily be verified that if Ψ is a solution of Eq. (1), then ψ_D is a solution of the Dirac equation

$$[\gamma_j(P + eA)_j + i\gamma_4(W + eV) - im]\psi = 0 \quad (3)$$

in the "standard" or Pauli representation, i.e., $\gamma_j = \rho_2\sigma$, and $\gamma_4 = \rho_3$. The ρ's and σ's are the usual ones.

In order to solve the Dirac equation, it is sufficient to solve the two-component equation and generate the corresponding Dirac solutions by use of Eq. (2). It should be noted that the two-component description of a spin $\frac{1}{2}$ particle is perfectly adequate and one never has to work with the four-component formulism.[7,8]

II. MAGNETIC FIELD SOLUTIONS

Algebraically, the two-component equation is very simple. In fact, when the scalar potential V vanishes, the two-component equation has the canonical form of the Pauli equation. The two-component equation assumes a particularly simple form when V is zero and the magnetic field has only a z component and depends upon only one coordinate, say y. With this choice of the external fields and an asymmetric choice of the gauge, $A_y = A_z = 0$, Eq. (1) becomes

$$[P_y^2 + P_z^2 + (P_x + eA_x(y))^2 + m^2 + esH_z(y)]\Psi_s\chi_s = W^2\Psi_s\chi_s. \quad (4)$$

The spin index s assumes the values $+1$ and -1 corresponding to the spinors $\chi_1 = \begin{vmatrix} 1 \\ 0 \end{vmatrix}$ and $\chi_{-1} = \begin{vmatrix} 0 \\ 1 \end{vmatrix}$, respectively.

The variables P_x and P_z are constants of the motion and can be immediately taken as constants. The trivial plane wave dependence on x and z can be suppressed, and then Ψ_s is only a function of y. After this is done, Eq. (4) is in the canonical form of the one-dimensional Schrödinger equation. Thus one method of constructing exact solutions of Eq. (1) is to choose an $A_x(y)$ such that Eq. (4) becomes equivalent to a Schrödinger equation with a solvable potential. For a constant magnetic field, $A_x(y) = -H_0y$ and Eq. (4) becomes the Schrödinger equation

* Work performed under the auspices of the U.S. Atomic Energy Commission.

[1] P. A. M. Dirac, Proc. Roy. Soc. (London) A117, 610 (1928).
[2] I. I. Rabi, Z. Physik 49, 7 (1928).
[3] F. Sauter, Z. Physik 69, 742 (1931).
[4] D. M. Volkow, Z. Physik 94, 25 (1935).
[5] P. J. Redmond, J. Math. Phys. 6, 1163 (1965).
[6] G. N. Stanciu, Phys. Letters 23, 232 (1966).
[7] R. P. Feynman and M. Gell-Mann, Phys. Rev. 109, 193 (1958).
[8] L. M. Brown, Phys. Rev. 111, 957 (1958) and *Lectures in Theoretical Physics, Boulder* (Interscience Publishers, Inc., N.Y., 1962), Vol. IV, p. 324.

for an harmonic oscillator.[9] Other choices of $A_x(y)$ can be made that lead to solvable potentials.

The vector potential $A_x(y) = -(H_0/a)\tanh(ay)$ and the corresponding magnetic field $H_z(y) = H_0\,\mathrm{sech}^2(ay)$ when substituted into Eq. (4) give, after some re-arrangement,

$$[P_y^2 + \xi^2 - 2P_x\xi\tanh(ay) - (\xi^2 - sa\xi)\,\mathrm{sech}^2(ay)]\Psi_s\chi_s = (W^2 - m^2 - P_x^2 - P_z^2)\Psi_s\chi_s. \quad (5)$$

The quantity ξ is eH_0/a and without loss of generality both of the quantities a and H_0 are taken as positive. Equation (5) is formally equivalent to a one-dimensional Schrödinger equation with a Rosen–Morse potential[10] or an Eckart potential.[11] The Rosen–Morse and the Eckart potentials are essentially the same.

With the introduction of the notation

$$\rho = -2\xi P_x/a^2,$$
$$\gamma = (\xi^2 - s\xi a)/a^2,$$
$$\epsilon = (\xi^2 + m^2 + P_x^2 + P_z^2 - W^2)/a^2;$$

Ψ_s satisfies the differential equation

$$[(d^2/dy^2) - \epsilon a^2 - \rho a^2\tanh(ay) + \gamma a^2\,\mathrm{sech}^2(ay)]\Psi_s = 0. \quad (6)$$

Following Rosen and Morse,[10] we convert the differential equation for Ψ_s into the hypergeometric equation by making the change of variable $\eta = \frac{1}{2}(1 + \tanh(ay))$ and the transformation $\Psi_s = e^{a\sigma y}\cosh^{-\tau}(ay)F_s(y)$. The constants σ and τ are

$$\sigma = -\tfrac{1}{2}[(\epsilon + \rho)^{\frac{1}{2}} - (\epsilon - \rho)^{\frac{1}{2}}],$$
$$\tau = \tfrac{1}{2}[(\epsilon + \rho)^{\frac{1}{2}} + (\epsilon + \rho)^{\frac{1}{2}}],$$

where both square roots are taken as positive quantities. The differential equation satisfied by $F_s(\eta)$ is

$$\eta(1-\eta)F_s''(\eta) + [(\sigma + \tau + 1) - 2\eta(\tau + 1)]F_s'(\eta) + [\gamma - \tau(\tau + 1)]F_s(\eta) = 0.$$

This equation is in the standard form of the hypergeometric equation. The solution regular at $\eta = 0$ which corresponds to $y = -\infty$ is

$$F[\tau + \tfrac{1}{2} - (\gamma + \tfrac{1}{4})^{\frac{1}{2}}, \tau + \tfrac{1}{2} + (\gamma + \tfrac{1}{4})^{\frac{1}{2}}, \sigma + \tau + 1; \eta].$$

This particular hypergeometric function diverges like $e^{2(\tau-\sigma)y}$ as η approaches 1 which corresponds to $y = +\infty$, unless $\tau + \frac{1}{2} + (\gamma + \frac{1}{4})^{\frac{1}{2}}$ or $\tau + \frac{1}{2} - (\gamma + \frac{1}{4})^{\frac{1}{2}}$ equals a negative integer.[10] The first possibility is not allowed since τ and $(\gamma + \frac{1}{4})^{\frac{1}{2}}$ are positive quantities. Therefore, in order to have Ψ_s finite at infinity, we must have

$$\tau = (\gamma + \tfrac{1}{4})^{\frac{1}{2}} - \tfrac{1}{2} - n.$$

In this case the hypergeometric series terminates and is a Jacobi polynomial. It follows directly from the finitions of σ and τ that $\sigma\tau = -\rho/2$, or

$$\sigma = -\tfrac{1}{2}\rho[(\tau + \tfrac{1}{4})^{\frac{1}{2}} - n - \tfrac{1}{2}]^{-1}.$$

The quantum number n is zero or any positive integer which satisfies the relationship

$$n \leq (\gamma + \tfrac{1}{4})^{\frac{1}{2}} - |\tfrac{1}{2}\rho|^{\frac{1}{2}} - \tfrac{1}{2} = |\xi a^{-1} - \tfrac{1}{2}s| - |\xi P_x/a^2|^{\frac{1}{2}} - \tfrac{1}{2}. \quad (7)$$

This result follows from examining the behavior of $\sigma + \tau$ and $\sigma - \tau$. Since the case $\xi a^{-1} < \frac{1}{2}$ and $s = +1$ is clearly impossible, the restriction on n can be written as

$$n \leq \xi a^{-1} - |\xi P_x/a^2|^{\frac{1}{2}} - \tfrac{1}{2}s - \tfrac{1}{2}. \quad (8)$$

If the inequality for n is not satisfied, there are no bound states (quantized orbits).

The energy eigenvalues are calculated from the relationship $\sigma^2 + \tau^2 = \epsilon$. In terms of the original parameters and the quantum number $N = n + \frac{1}{2} + \frac{1}{2}s$, the energy eigenvalues are

$$W_{n,s} = [m^2 + P_x^2 + P_z^2 + \xi^2 - (\xi - Na)^2 - \xi^2P_x^2(\xi - Na)^{-2}]^{\frac{1}{2}}. \quad (9)$$

The limit $a = 0$ corresponds to a constant magnetic field. In this limit n can be zero or any positive integer and the eigenvalues given by Eq. (9) become

$$W_{n,s} = [m^2 + P_z^2 + 2eH_0N]^{\frac{1}{2}}.$$

This is the well-known result for a homogeneous magnetic field.[2,9]

The energy eigenvalues and the restriction on the quantum number n for a positively charged particle are

$$W_{n,s} = [m^2 + P_x^2 + P_z^2 + \xi^2 - (\xi - N'a)^2 - \xi^2P_x^2(\xi - N'a)^{-2}]^{\frac{1}{2}}$$
$$n \leq (\xi/a) - |\xi P_x/a^2|^{\frac{1}{2}} + \tfrac{1}{2}s - \tfrac{1}{2}.$$

The quantum number $N' = n - \frac{1}{2}s + \frac{1}{2}$ and ξ is $|eH_0/a|$.

The scattering solutions are not difficult to construct. Instead of exponentially damped solutions at $y = \pm\infty$, the solutions are oscillatory. This occurs when both $(\epsilon - \rho)^{\frac{1}{2}}$ and $(\epsilon + \rho)^{\frac{1}{2}}$ are pure imaginary. Therefore, let us introduce the notation $\frac{1}{2}(\epsilon - \rho)^{\frac{1}{2}} = i\alpha$ and $\frac{1}{2}(\epsilon + \rho)^{\frac{1}{2}} = i\beta$, where the quantities α and β

[9] M. H. Johnson and B. A. Lippmann, Phys. Rev. 76, 828 (1949).
[10] N. Rosen and P. M. Morse, Phys. Rev. 42, 210 (1932).
[11] C. Eckart, Phys. Rev. 35, 1303 (1930).

are

$$\alpha = [W^2 - (P_x + \xi)^2 - P_z^2 - m^2]^{\frac{1}{2}}/2a$$
$$\beta = [W^2 - (P_x - \xi)^2 - P_z^2 - m^2]^{\frac{1}{2}}/2a.$$

As a particular example of a scattering problem, let us choose the boundary conditions such that at $y = -\infty$ there is an incoming and an outgoing (reflected) wave and at $y = +\infty$ there is only an outgoing (transmitted) wave.

The wavefunction which has the proper asymptotic behavior at $y = -\infty$ is

$$\Psi(y) = e^{-i(\beta-\alpha)ay}\cosh^{-i(\beta+\alpha)}(ay)$$
$$\times \{C_1 F[\tfrac{1}{2} + i(\beta+\alpha) - \Delta, \tfrac{1}{2} + i(\beta+\alpha) + \Delta,$$
$$1 + 2i\alpha; \tfrac{1}{2}(1 + \tanh(ay))]$$
$$+ C_2[\tfrac{1}{2}(1 + \tanh(ay))]^{-2i\alpha} F[\tfrac{1}{2} + i(\beta-\alpha) - \Delta,$$
$$\tfrac{1}{2} + i(\beta-\alpha) + \Delta, 1 - 2i\alpha; \tfrac{1}{2}(1 + \tanh(ay))]\}\chi_s. \tag{10}$$

The notation $\Delta = (\gamma + \frac{1}{4})^{\frac{1}{2}}$ has been introduced and the plane-wave dependence of x and z has been suppressed. The asymptotic behavior of $\Psi(y)$ as y tends towards minus infinity is

$$\Psi(y) \underset{y\to-\infty}{\sim} 2^{i(\beta+\alpha)} e^{2i\alpha ay}[C_1 + C_2 e^{-4i\alpha ay}]\chi_s.$$

In order to calculate the asymptotic behavior of $\Psi(y)$ as y tends towards $+\infty$, one must analytically continue the hypergeometric functions. The analytic continuation of $\Psi(y)$ is[12]

$$\Psi(y) = e^{-i(\beta-\alpha)ay}\cosh^{-i(\beta+\alpha)}(ay)$$
$$\times \{C_1 A_1 F[\tfrac{1}{2} + i(\beta+\alpha) - \Delta, \tfrac{1}{2} + i(\beta+\alpha) + \Delta,$$
$$1 + 2i\beta; \tfrac{1}{2}(1 - \tanh(ay))] + C_1 A_2[\tfrac{1}{2}(1 - \tanh(ay))]^{-2i\beta}$$
$$\times F[\tfrac{1}{2} + i(\alpha-\beta) - \Delta, \tfrac{1}{2} + i(\alpha-\beta) + \Delta,$$
$$1 - 2i\beta; \tfrac{1}{2}(1 - \tanh(ay))] + C_2 A_1'[\tfrac{1}{2}(1 + \tanh(ay))]^{-2i\alpha}$$
$$\times F[\tfrac{1}{2} + i(\beta-\alpha) - \Delta. \tfrac{1}{2} + i(\beta-\alpha) + \Delta,$$
$$1 + 2i\beta; \tfrac{1}{2}(1 - \tanh(ay))] + C_2 A_2'[\tfrac{1}{2}(1 + \tanh(ay))]^{-2i\alpha}$$
$$\times [\tfrac{1}{2}(1 - \tanh(ay))]^{-2i\beta} F[\tfrac{1}{2} - i(\alpha+\beta) - \Delta,$$
$$\tfrac{1}{2} - i(\alpha+\beta) + \Delta, 1 - 2i\beta; \tfrac{1}{2}(1 - \tanh(ay))]\}\chi_s, \tag{11}$$

where

$$A_1 = \frac{\Gamma[1 + 2i\alpha]\Gamma[-2i\beta]}{\Gamma[\frac{1}{2} + i(\alpha-\beta) + \Delta]\Gamma[\frac{1}{2} + i(\alpha-\beta) - \Delta]},$$
$$A_2 = \frac{\Gamma[1 + 2i\alpha]\Gamma[+2i\beta]}{\Gamma[\frac{1}{2} + i(\alpha+\beta) + \Delta]\Gamma[\frac{1}{2} + i(\alpha+\beta) - \Delta]},$$
$$A_1' = \frac{\Gamma[1 - 2i\alpha]\Gamma[-2i\beta]}{\Gamma[\frac{1}{2} - i(\alpha+\beta) + \Delta]\Gamma[\frac{1}{2} - i(\alpha+\beta) - \Delta]},$$
$$A_2' = \frac{\Gamma[1 - 2i\alpha]\Gamma[+2i\beta]}{\Gamma[\frac{1}{2} - i(\alpha-\beta) + \Delta]\Gamma[\frac{1}{2} - i(\alpha-\beta) - \Delta]}. \tag{12}$$

[12] The hypergeometric functions are analytically continued according to Bateman Manuscript Project, *Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, pp. 108–109.

The asymptotic behavior of $\Psi(y)$ as y tends towards $+\infty$ is

$$\Psi(y) \underset{y\to+\infty}{\sim} 2^{i(\beta+\alpha)} e^{-2i\beta ay}$$
$$\times [(C_1 A_1 + C_2 A_1') + (C_1 A_2 + C_2 A_2') e^{4i\beta ay}]\chi_s.$$

There is no spin-flip in the scattering process, since the asymptotic values of $\Psi(y)$ only contain χ_s with the same spin index s. The boundary condition at infinity (only an outgoing wave) is satisfied by choosing $C_1 A_1 + C_2 A_1' = 0$.

The reflection and transmission coefficients are computed from the asymptotic values of the incident, reflected, and transmitted probability currents. These currents are calculated in the four-component formulism by constructing the Dirac spinor by the use of Eq. (2) and then using the definition of the probability current $j_\mu = i(\bar{\psi}_D \gamma_\mu \psi_D)$. The currents can also be calculated strictly within the two component formulism (see Ref. 8). The currents are

$$j_{\text{inc}} = 8a\alpha(P_z + W)\,|C_1|^2,$$
$$j_{\text{ref}} = -8a\alpha(P_z + W)\,|C_2|^2,$$
$$j_{\text{trans}} = 8a\beta(P_z + W)\,|C_1 A_2 + C_2 A_2'|^2.$$

By definition, the reflection and the transmission coefficients are

$$R = \left|\frac{C_2}{C_1}\right|^2 = \left|\frac{A_1}{A_1'}\right|^2$$

and

$$T = \frac{\beta}{\alpha}\left|\frac{C_1 A_1 + C_2 A_2'}{C_1}\right|^2 = \frac{\beta}{\alpha}\left|\frac{A_1' A_2 - A_1 A_2'}{A_1'}\right|^2.$$

By using elementary properties of the gamma function and noting that the quantity $\Delta = (\gamma + \frac{1}{4})^{\frac{1}{2}} = |\xi a^{-1} - \frac{1}{2}s|$ is a real quantity, the expressions for R and T can be written as

$$R = \frac{\cosh 2\pi(\alpha-\beta) + \cos 2\pi\Delta}{\cosh 2\pi(\alpha+\beta) + \cos 2\pi\Delta} \tag{13}$$

and

$$T = \frac{\cosh 2\pi(\alpha+\beta) - \cosh 2\pi(\alpha-\beta)}{\cosh 2\pi(\alpha+\beta) + \cos 2\pi\Delta}. \tag{14}$$

The conservation of probability demands that $R + T = 1$ and is satisfied. Both coefficients depend upon the spin index s via Δ. The reason for this is that the two spin projections have different energies and the incident energy is effectively raised ($s = +1$) and lowered ($s = -1$) depending on the spin projection. The coefficients for scattering with a positively charged particle are obtained from Eqs. (13) and (14) by changing the sign of ξ in α, β, and Δ.

In the above expressions it is assumed that oscillatory solutions exist at $y = \pm\infty$. This occurs when α and β are real or equivalently when α^2 and β^2 are positive. The reality of α and β implies one condition on the energy, namely

$$W^2 > m^2 + P_x^2 + P_z^2 + \xi^2 + 2\,|P_x\xi|\,.$$

There is also the possibility that there are oscillatory solutions at $-\infty$ and an exponentially damped solution at $+\infty$. This corresponds to pure reflection and occurs when α^2 is positive and β^2 is negative. This is only possible when $P_x\xi$ is negative and W satisfies the inequalities

$$m^2 + P_x^2 + P_z^2 + \xi^2 - 2\,|P_x\xi| < W^2 < m^2 + P_x^2 + P_z^2 + \xi^2 + 2\,|P_x\xi|\,.$$

Finally, the case α^2 positive and β equal to zero also gives pure reflection.

Another choice of $A_x(y)$ and $H_z(y)$ that leads to an exactly solvable equation is: $A_x(y) = (H_0/a)(1 - e^{ay})$ and $H_z(y) = H_0e^{ay}$. This choice of the external fields gives the two-component equation

$$[P_y^2 + 2\xi P_x + \xi^2 - (2\xi P_x + 2\xi^2 - s\xi a)e^{ay} + \xi^2e^{2ay}]\Psi_s\chi_s = (W^2 - P_x^2 - P_z^2 - m^2)\Psi_s\chi_s$$

which is formally equivalent to a one-dimensional Schrödinger equation with a Morse potential.[13] Again both a and H_0 are taken as positive without loss of generality and $\xi = eH_0/a$. The spin index s and the spinor χ_s are the same as before.

With the introduction of the notation

$$g = [W^2 - m^2 - P_z^2 - (P_x + \xi)^2]/a^2,$$
$$f = (s\xi a - 2\xi P_x - 2\xi^2)/a^2,$$

and

$$h = \xi^2/a^2,$$

the differential equation for $\Psi_s(y)$ is

$$[(d^2/dy^2) + ga^2 - fa^2e^{ay} - ha^2e^{2ay}]\Psi_s(y) = 0.$$

This equation is transformed into Laguerre's equation by performing the change of variable $x = 2(h)^{\frac{1}{2}}e^{ay}$ and the transformation $\Psi_s(y) = x^{(-g)^{\frac{1}{2}}}e^{-\frac{1}{2}x}G_s(x)$.[13] The differential equation satisfied by G_s is

$$xG_s''(x) + [(2(-g)^{\frac{1}{2}} + 1) - x]G_s'(x) - \{\tfrac{1}{2} + [f/2(h)^{\frac{1}{2}}] + (-g)^{\frac{1}{2}}\}G_s(x) = 0.$$

The differentiation is with respect to x. Thus G_s is a solution of Laguerre's equation. The regularity conditions[13] imply that bound states only exist if

$$(-g)^{\frac{1}{2}} + [f/2(h)^{\frac{1}{2}}] + \tfrac{1}{2} = -n, \quad (-g)^{\frac{1}{2}} \geq 0. \quad (15)$$

[13] P. M. Morse, Phys. Rev. **34**, 57 (1929).

The function G_s is the associated Laguerre polynomial $L_n^{2(-g)^{\frac{1}{2}}}(x)$ and the two-component wave function in terms of the variable x is $x^{(-g)^{\frac{1}{2}}}e^{-\frac{1}{2}x}L_n^{2(-g)^{\frac{1}{2}}}(x)\chi_s$. The four-component wavefunction is constructed by means of Eq. (2).

The second part of Eq. (15) implies that the quantum number can be zero or any positive integer which satisfies the relationship

$$n \leq [(\xi + P_x)/a] - \tfrac{1}{2}s - \tfrac{1}{2}. \quad (16)$$

The energy eigenvalues are easily calculated from the first part of Eq. (15) and are

$$W_{n,s} = [m^2 + P_z^2 + 2Na(\xi + P_x) - N^2a^2]^{\frac{1}{2}}. \quad (17)$$

The quantum number $N = n + \frac{1}{2}s + \frac{1}{2}$.

The constant field limit is $a = 0$ and in this limit the eigenvalues become the eigenvalues for a constant magnetic field, as they should.

The energy eigenvalues and the restriction on the quantum number n for a positively charged particle are

$$W_{n\,s} = [m^2 + P_z^2 + 2N'a(\xi - P_x) - N'^2a^2]^{\frac{1}{2}},$$
$$n \leq [(\xi - P_x)/a] + \tfrac{1}{2}s - \tfrac{1}{2}.$$

The quantum number $N' = n - \frac{1}{2}s + \frac{1}{2}$ and ξ is $|eH_0/a|$.

The scattering solutions are not of much interest. As y tends towards $-\infty$, oscillatory solutions exist if $(-g)^{\frac{1}{2}}$ is pure imaginary. As y tends towards $+\infty$, only exponentially damped or growing solutions exist and consequently, the only defined scattering problem must have a reflection coefficient of one. Of course, nontrivial scattering problems can be constructed by either letting $A_x(y)$ equal zero past some point λ or matching at λ with the previous field.

III. THE ELECTRIC FIELD SOLUTIONS

The two-component equation for the external field configuration $\mathbf{A} = 0$ and a scalar potential depending upon only one direction, say z, is

$$[\mathbf{P}^2 + m^2 + iseE_z]\Psi_s\chi_s = (W + eV)^2\Psi_s\chi_s. \quad (18)$$

This equation is formally equivalent to Eq. (4), though it describes a different physical situation. Consequently the potentials $V(z) = -(E_0/a)\tanh(az)$ and $V(z) = -(E_0/a)(1 - e^{az})$ lead to equations which are formally equivalent to one-dimensional Schrödinger equations with a Rosen–Morse potential and a Morse potential, respectively. Thus both of these potentials give rise to exactly solvable problems that correspond physically to scattering from one-dimensional repulsive barriers; there are no bound states. (Even nonrelativistically

these potentials have no bound states.) The wave functions satisfying appropriate boundary conditions are easily constructed in terms of the hypergeometric functions for the first potential and the confluent hypergeometric functions for the second potential. The scattering solutions from the second potential are only of the Klein paradox type, since the potential becomes infinitely large and only the "pair production" solutions are possible.[14,15] Of course, a potential that is finite everywhere can be constructed from the second potential by appropriately truncating the potential and matching the truncated part of the potential with a constant potential or any other solvable potential which is finite, for example, the first potential. Only one example of the numerous scattering problems is considered here.

Let us consider the first potential with the boundary conditions such that there are incoming and outgoing waves at $z = -\infty$ and only an outgoing wave at $z = +\infty$. For definiteness, let us take both a and E_0 as positive. Then the function ψ_s obeys the differential equation

$$[(d^2/dz^2) - a^2\epsilon' - \rho' a^2 \tanh(az) + \gamma' a^2 \operatorname{sech}^2(az)]\Psi_s(z) = 0$$

where

$$\epsilon' = -(W^2 - m^2 - P_x^2 - P_z^2 + v^2)/a^2,$$
$$\rho' = 2Wv/a^2,$$
$$\gamma' = -(v^2 + isva)/a^2,$$

and

$$v = eE_0/a.$$

The equation for $\Psi_s(z)$ is identical to Eq. (6) after a trivial relabeling of the variables and constants. Thus the scattering solutions for the present problem can be obtained from the magnetic field scattering field solutions. Performing the trivial relabeling, we have

$$\Psi(z) \underset{z\to-\infty}{\sim} 2^{i(\beta'+\alpha')}e^{2i\alpha' az}[C_1 + C_2 e^{-4i\alpha' az}]\chi_s$$

and

$$\Psi(z) \underset{z\to+\infty}{\sim} 2^{i(\beta'+\alpha')}e^{+2i\beta' az}[C_1A_2 + C_2A_2']\chi_s.$$

The primed constants are

$$\alpha' = [(W + v)^2 - P_x^2 - P_y^2 - m^2]^{\frac{1}{2}}/2a,$$
$$\beta' = [(W - v)^2 - P_x^2 - P_y^2 - m^2]^{\frac{1}{2}}/2a,$$
$$\Delta' = \{-[v^2 + isa)/a^2] + \tfrac{1}{4}\}^{\frac{1}{2}} = -\tfrac{1}{2}s + i(va^{-1}) \equiv -\tfrac{1}{2}s + i\delta.$$

The constants C_1 and C_2 are related by $C_2A_1' + C_1A_1 = 0$. The constants A_1, A_1', A_2, and A_2' are given by Eq. (12) when all of the arguments of the gamma function are replaced by primed quantities. The probability currents are calculated in the same manner as for the magnetic field case and asymptotically are

$$j_{\text{inc}} = 8i\alpha' a(W + v + 2a\alpha' s)\,|C_1|^2,$$
$$j_{\text{ref}} = -8i\alpha' a(W + v - 2a\alpha' s)\,|C_2|^2,$$
$$j_{\text{trans}} = 8i\beta' a(W - v + 2a\beta' s)\,|C_1A_2 + C_2A_2'|^2.$$

The reflection and transmission coefficients are

$$R = \frac{(W + v - 2\alpha' as)}{(W + v + 2\alpha' as)}\left|\frac{A_1}{A_1'}\right|^2,$$

$$T = \frac{(W - v + 2\alpha\beta' s)}{(W + v + 2a\alpha' s)}\left|\frac{A_2A_1' - A_1A_2'}{A_1'}\right|^2.$$

By use of elementary properties of the gamma function and taking into account that $\Delta' = -\frac{1}{2}s + i\delta$ where δ is real, the coefficients can be written as

$$R = \frac{\cosh^2\pi(\alpha' - \beta') - \cosh^2\pi\delta}{\cosh^2\pi(\alpha' + \beta') - \cosh^2\pi\delta}, \tag{19}$$

$$T = \frac{\cosh^2\pi(\alpha' + \beta') - \cosh^2\pi(\alpha' - \beta')}{\cosh^2\pi(\alpha' + \beta') - \cosh^2\pi\delta}. \tag{20}$$

Both of the coefficients are independent of the spin index s. This occurs because both spin projections have the same energy and consequently, the spin splitting of the incident energy that occurred in the magnetic field problem does not happen here. It is apparent that $R + T = 1$, and thus probability is conserved.

In the above expressions it is assumed that both α' and β' are real. If β' is pure imaginary, then the wavefunction at $z = +\infty$ is exponentially damped and pure reflection occurs. β' is pure imaginary when $|W - v| < (P_x^2 + P_y^2 + m^2)^{\frac{1}{2}}$.

The oscillatory solutions occur when both α' and β' are real. The reality of α' and β' implies that $|W - v| > (P_x^2 + P_y^2 + m^2)^{\frac{1}{2}}$. There are two cases to distinguish. If $W > v$, then $W > v + (P_x^2 + P_y^2 + m^2)^{\frac{1}{2}}$. This is the familiar realm of scattering over a potential barrier. If $W < v$, then $W < v - (P_x^2 + P_y^2 + m^2)^{\frac{1}{2}}$ which is only possible if $v > m + (P_x^2 + P_y^2 + m^2)^{\frac{1}{2}}$, since $W \geq m$. This is the region of the Klein paradox, i.e., oscillatory solutions scattering through the potential barrier.[14,15] In this case the reflected current exceeds the incident current and the transmitted current is negative, though $R + T$ is still equal to one. This behavior follows from Eqs. (19) and (20) since $\delta > \alpha' + \beta'$ in the Klein paradox region.

For completeness, we mention that when $(W - v) = (P_x^2 + P_y^2 + m^2)^{\frac{1}{2}}$, which corresponds to $\beta' = 0$, pure reflection occurs.

[14] O. Klein, Z. Physik 53, 157 (1929).

[15] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill Book Company, Inc., N.Y., 1964), pp. 40–42.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Realization of Lie Algebras by Analytic Functions of Generators of a Given Lie Algebra*

P. CHAND,† C. L. MEHTA,** N. MUKUNDA,‡ AND E. C. G. SUDARSHAN
Physics Department, Syracuse University, Syracuse, New York

(Received 30 December 1966)

In this paper we discuss the problem of the Poisson bracket realization of various Lie algebras in terms of analytic functions of the generators of a given Lie algebra. We pose and solve the problem of realizing the general $O(4)$, $O(3, 1)$, and $E(3)$ algebras in terms of analytic functions of the generators of a prescribed realization of an $E(3)$ algebra. A similar problem is solved for the symmetric tensor realizations of $SU(3)$ and $SL(3, R)$. Related questions are discussed for $O(n + 1)$, $O(n, 1)$, $E(n)$, $SU(n)$, and $SL(n, R)$. We study in some detail the finite canonical transformations realized by the generators of the various groups. The relation of these results to the reconstruction problem is briefly discussed.

1. INTRODUCTION

THE success of group-theoretic methods in particle physics has led to a re-examination of these methods as applied to elementary dynamical problems in both quantum and classical mechanics. One finds that sometimes the quantum-mechanical and the classical-mechanical problems can be made to correspond to each other in such a fashion that their invariance groups and their noninvariance groups have the same structure. Thus, for example, the isotropic harmonic oscillator in n dimensions has the invariance group $SU(n)$ and the noninvariance groups $SU(n, 1)$ and $SU(n + 1)$, both in classical and in quantum mechanics.[1]

For quantum-mechanical systems, the Lie algebras of the invariance and noninvariance groups are realized by commutation relations between appropriate dynamical operators, whereas for classical systems the Lie algebras are realized by Poisson brackets between appropriate dynamical variables. That such different realizations exhibit a correspondence between them is quite remarkable. It has been known for quite some time that the infinite-dimensional Lie algebras of polynomials in canonical variables for quantum mechanics (commutator brackets) and for classical mechanics (Poisson brackets) have quite different structures.[2] Yet there exists the same local Lie group in both kinds of dynamics. This, then, suggests that in these two different algebraic systems, there are selected subsets of elements whose Lie algebraic structures are isomorphic.

Given any Lie algebra $\mathcal{A}$ of quantities which possess an associative law of multiplication (in addition to the Lie bracket), we can define an enveloping algebra $\mathcal{E}$ whose elements are polynomials in the elements of the primitive Lie algebra $\mathcal{A}$. The enveloping algebra $\mathcal{E}$ can be given an induced Lie algebra structure by imposing the relation

$$\{X_1 \cdot X_2, Y\} = \{X_1, Y\} \cdot X_2 + X_1 \cdot \{X_2, Y\}, \quad (1.1)$$

where the curly bracket stands for the Lie bracket and the dot (which will usually be omitted) stands for the associative product. We can generalize our definition of the enveloping algebra $\mathcal{E}$ by including among its elements the analytic functions of the elements of the primitive Lie algebra $\mathcal{A}$. This enveloping algebra $\mathcal{E}$ is given an induced Lie-algebra structure by imposing the derivation property. Using this rule, we can identify $\mathcal{E}$ with an infinite-dimensional Lie algebra. In classical mechanics it would be a Lie algebra of Poisson brackets, whereas in quantum mechanics it would be a Lie algebra of commutators. The enveloping algebras in the two cases have, in general, quite different structures. However, in both cases there are certain sets of invariant elements which have vanishing Lie brackets with every element of the primitive Lie algebra and, consequently, with every element of the enveloping algebra. For the quantum-mechanical case, these are the well-known invariant operators, which are expressible as functions of the so-called Casimir invariants. If the N elements of the primitive Lie algebra are denoted by $X_1, X_2, \cdots, X_N$, then the Casimir invariants (for both quantum and classical systems) are homogeneous polynomials of the type

$$C_n = \sum_{\alpha_1=1}^{N} \cdots \sum_{\alpha_n=1}^{N} C^{\alpha_1 \alpha_2 \cdots \alpha_n} X_{\alpha_1} \cdots X_{\alpha_n}, \quad (1.2)$$

* Research supported in part by the U.S. Atomic Energy Commission and by the U.S. Air Force Office of Scientific Research, Office of Aerospace Research.

† Present Address: Agricultural and Mechanical University, College Station, Texas.

** Permanent Address: Department of Physics and Astronomy, University of Rochester, Rochester, New York.

‡ On leave of absence from Tata Institute of Fundamental Research, Bombay, India.

[1] See, for instance, N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965); R. C. Hwa and J. Nuyts, Phys. Rev. **145**, 1188 (1966).

[2] J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949).

where $C^{\alpha_1 \cdots \alpha_n}$ are some numerical coefficients which are symmetric in the indices $\alpha_1, \cdots, \alpha_n$. It is possible that in some particular realizations these Casimir invariants may degenerate into numbers. (In fact, in the case of irreducible realizations, all these invariant elements reduce to numbers.)

In many cases, it is possible to select, out of the infinite-dimensional enveloping algebra, a finite-dimensional subset of elements which constitute another Lie algebra. We show, in this paper, how such construction can be carried out for the realizations of the Lie algebras by Poisson brackets. We begin in Sec. 2 by giving the structures and the Casimir invariants of the Lie algebras of the groups $E(3)$, $O(4)$, $O(3, 1)$, $SU(3)$, and $SL(3, R)$. A simple realization of all these Lie algebras in terms of three pairs of canonical variables q_a, p_a $(a = 1, 2, 3)$ is also given in this section. In Sec. 3, after defining the primitive $E(3)$ Lie algebra and the corresponding generalized enveloping algebra, we give explicit realizations of the $O(4)$, $O(3, 1)$, and other $E(3)$ Lie algebras. The properties of the finite canonical transformations generated by the elements of these Lie algebras are also discussed. A similar problem is discussed for $SU(3)$ and $SL(3, R)$ Lie algebras in Sec. 4. Section 5 deals with a generalization to n dimensions, while in Sec. 6 we discuss briefly the relation of these results to the problem of reconstruction of canonical variables from the generators of a noninvariance group.

2. STRUCTURES AND INVARIANTS OF THE LIE ALGEBRAS OF VARIOUS GROUPS

In this section we briefly outline the structure and invariants of the Lie algebras of $E(3)$, $O(4)$, $O(3, 1)$, $SU(3)$, and $SL(3, R)$.

i. $E(3)$ Lie Algebra

The Lie algebra of the Euclidean group $E(3)$ in three dimensions consists of six elements: J_a, P_a $(a = 1, 2, 3)$, and has the following basic Poisson brackets:

$$\{J_a, J_b\} = \epsilon_{abc} J_c, \tag{2.1}$$

$$\{J_a, P_b\} = \epsilon_{abc} P_c, \tag{2.2}$$

$$\{P_a, P_a\} = 0. \tag{2.3}$$

Here ϵ_{abc} is the completely antisymmetric unit tensor of Levi–Civita. Throughout this paper we employ the usual summation convention according to which a summation is implied over repeated dummy indices.

The two quadratic elements

$$P^2 = P_a P_a \tag{2.4}$$

and

$$\mathbf{J} \cdot \mathbf{P} = J_a P_a \tag{2.5}$$

of the enveloping algebra are invariants. We shall assume that the realization is real and faithful so that P^2 is positive and may be normalized to unity. There are, however, two classes of realizations corresponding to vanishing or nonvanishing of $\mathbf{J} \cdot \mathbf{P}$.

ii. $O(4)$ Lie Algebra

The Lie algebra of the real orthogonal group $O(4)$ in four dimensions also consists of six elements: J_a, K_a, having the following basic Poisson brackets $(1 \leq a, b, c \leq 3)$:

$$\{J_a, J_b\} = \epsilon_{abc} J_c, \tag{2.6}$$

$$\{J_a, K_b\} = \epsilon_{abc} K_c, \tag{2.7}$$

$$\{K_a, K_b\} = \epsilon_{abc} J_c. \tag{2.8}$$

The two quadratic elements

$$J^2 + K^2 = J_a J_a + K_a K_a \tag{2.9}$$

and

$$\mathbf{J} \cdot \mathbf{K} = J_a K_a \tag{2.10}$$

of the enveloping algebra are invariants. For a real faithful realization $J^2 + K^2$ is positive, whereas $\mathbf{J} \cdot \mathbf{K}$ may be positive, negative, or zero.

iii. $O(3, 1)$ Lie Algebra

The Lie algebra of the real pseudo-orthogonal group $O(3, 1)$ in four dimensions again consists of six elements: J_a, K'_a, and has the following basic Poisson brackets:

$$\{J_a, J_b\} = \epsilon_{abc} J_c, \tag{2.11}$$

$$\{J_a, K'_b\} = \epsilon_{abc} K'_c, \tag{2.12}$$

$$\{K'_a, K'_b\} = -\epsilon_{abc} J_c. \tag{2.13}$$

The two quadratic elements

$$J^2 - K'^2 = J_a J_a - K'_a K'_a \tag{2.14}$$

and

$$\mathbf{J} \cdot \mathbf{K}' = J_a K'_a \tag{2.15}$$

of the enveloping algebra are invariants. For a real faithful realization, either invariant may be positive, negative, or zero.

iv. $SU(3)$ Lie Algebra

The Lie algebra of the unimodular unitary group $SU(3)$ in three dimensions consists of eight elements: J_a $(a = 1, 2, 3)$ and five linearly independent elements of a symmetric traceless "quadrupole" tensor Q_{ab}:

$$Q_{ab} = Q_{ba}; \quad Q_{aa} = 0. \tag{2.16}$$

In this case, we have the following basic Poisson brackets:

$$\{J_a, J_b\} = \epsilon_{abc}J_c, \tag{2.17}$$

$$\{J_a, Q_{bc}\} = \epsilon_{abc}Q_{dc} + \epsilon_{acd}Q_{bd}, \tag{2.18}$$

$$\{Q_{ab}, Q_{cd}\} = (\epsilon_{ace}\delta_{bd} + \epsilon_{ade}\delta_{bc} + \epsilon_{bce}\delta_{ad} + \epsilon_{bde}\delta_{ac})J_e. \tag{2.19}$$

There are two basic invariants, one of the second degree,

$$J_aJ_a + \tfrac{1}{2}Q_{ab}Q_{ab}, \tag{2.20}$$

and one of the third degree,

$$(\sqrt{3}/2)(3J_aQ_{ab}J_b - Q_{ab}Q_{bc}Q_{ca}). \tag{2.21}$$

As will be seen in Sec. 4 [see also Eqs. (2.40), (2.41)], there exist some realizations in which the cube of the quadratic invariant equals the square of the cubic invariant.

v. $SL(3, R)$ Lie Algebra

The Lie algebra of the unimodular real linear group $SL(3, R)$ in three dimensions also consists of eight elements: J_a $(a = 1, 2, 3)$ and five linearly independent elements of a symmetric traceless tensor Q'_{ab} $(Q'_{ab} = Q'_{ab}; Q'_{ab} = 0)$. The basic Poisson-bracket relations are

$$\{J_a, J_b\} = \epsilon_{abc}J_c, \tag{2.22}$$

$$\{J_a, Q'_{bc}\} = \epsilon_{abd}Q'_{dc} + \epsilon_{acd}Q'_{bd}, \tag{2.23}$$

$$\{Q'_{ab}, Q'_{cd}\} = -(\epsilon_{ace}\delta_{bd} + \epsilon_{ade}\delta_{bc} + \epsilon_{bce}\delta_{ad} + \epsilon_{bde}\delta_{ac})J_e. \tag{2.24}$$

There are again two basic invariants, one of second degree,

$$J_aJ_a - \tfrac{1}{2}Q'_{ab}Q'_{ab}, \tag{2.25}$$

and one of the third degree,

$$(\sqrt{3}/2)(3J_aQ'_{ab}J_b + Q'_{ab}Q'_{bc}Q'_{ca}). \tag{2.26}$$

It will be seen that, with the particular choice of the over-all coefficient in (2.26), in some realizations the cube of the quadratic invariant equals the negative of the square of the cubic invariant.

For each of these Lie algebras, an explicit construction can be given for the appropriate elements in terms of three pairs of canonical variables q_a, p_a which satisfy

$$\{q_a, p_b\} = \delta_{ab}, \qquad \{q_a, q_b\} = \{p_a, p_b\} = 0. \tag{2.27}$$

The simplest construction for the elements of the $E(3)$ Lie algebra is given by

$$J_a = (\mathbf{q} \times \mathbf{p})_a = \epsilon_{abc}q_bp_c, \qquad P_a = p_a. \tag{2.28}$$

For this realization, the invariant $\mathbf{J} \cdot \mathbf{P}$ vanishes and the other invariant P^2 is given by the dynamical variable p_ap_a. This realization is reducible, since the value of P^2 is unchanged by canonical transformations belonging to $E(3)$. One can now ask whether one could construct other realizations of $E(3)$ in terms of the same canonical variables, such that the $O(3)$ subalgebra generated by J_a is unchanged and that the two invariants $\mathbf{J} \cdot \mathbf{P} = \alpha_0$ and $P^2 > 0$ can be assigned arbitrary values. For this purpose we consider

$$J_a = \epsilon_{abc}q_bp_c \tag{2.29}$$

and

$$P_a = f(J^2)p_a + g(J^2)J_a, \tag{2.30}$$

where f and g are some functions of J^2. If we now impose $\mathbf{J} \cdot \mathbf{P} = \alpha_0$ and $P^2 = 1$, we obtain, from (2.30), the following expression for P_a:

$$P_a = \left\{\frac{1}{p^2}\left(1 - \frac{\alpha_0^2}{J^2}\right)^{\frac{1}{2}}\right\}p_a + \frac{\alpha_0}{J^2}J_a. \tag{2.31}$$

It may easily be verified that (2.29) and (2.31) do furnish a realization of $E(3)$.

For the elements of the $O(4)$ Lie algebra, we have the following simple construction:

$$J_a = \epsilon_{abc}q_bp_c, \tag{2.32}$$

$$K_a = (\beta^2 - J^2)^{\frac{1}{2}}(p^2)^{-\frac{1}{2}}p_a. \tag{2.33}$$

In this particular realization the invariant $\mathbf{J} \cdot \mathbf{K}$ vanishes and the other invariant $J^2 + K^2$ has the value β^2. If we choose J_a given by (2.32) and K_a by an expression similar to the one given on the right-hand side of (2.30) and impose the conditions $\mathbf{J} \cdot \mathbf{K} = \alpha\beta$, $J^2 + K^2 = \alpha^2 + \beta^2$, we obtain a realization with arbitrary values of the two invariants:

$$J_a = \epsilon_{abc}q_bp_c, \tag{2.34}$$

$$K_a = \{(\beta^2 - J^2)(J^2 - \alpha^2)/J^2p^2\}^{\frac{1}{2}}p_a + (\alpha\beta/J^2)J_a. \tag{2.35}$$

It may be checked that this is a solution of (2.8). Without loss of generality we can assume $\beta \geq |\alpha| \geq 0$. The realization given by (2.34), (2.35) is real in the region $\alpha^2 \leq J^2 \leq \beta^2$.

A realization of $O(3, 1)$ is obtained by simply choosing $K'_a = iK_a$ and by analytically continuing α to pure imaginary values, i.e., by putting $\alpha' = i\alpha$, where α' is now real. We thus obtain the following construction for the elements of $O(3, 1)$ Lie algebra:

$$J_a = \epsilon_{abc}q_bp_c, \tag{2.36}$$

$$K'_a = \{(J^2 - \beta^2)(J^2 + \alpha'^2)/J^2p^2\}^{\frac{1}{2}}p_a + (\alpha'\beta/J^2)J_a. \tag{2.37}$$

The two invariants $\mathbf{J} \cdot \mathbf{K}$ and $J^2 - K^2$ have the values $\alpha'\beta$ and $\beta^2 - \alpha'^2$, respectively, and the realization is real in the region $J^2 \geq \beta^2$.

The simplest construction for the elements of the $SU(3)$ Lie algebra is given by

$$J_a = \epsilon_{abc} q_b p_c, \tag{2.38}$$

$$Q_{ab} = q_a q_b + p_a p_b - \tfrac{1}{3}\delta_{ab}(q^2 + p^2). \tag{2.39}$$

In this case the two invariants have the values

$$J_a J_a + \tfrac{1}{2} Q_{ab} Q_{ab} = \tfrac{1}{3}(q^2 + p^2)^2 \tag{2.40}$$

and

$$(\sqrt{3}/2)(3 J_a Q_{ab} J_b - Q_{ab} Q_{bc} Q_{ca}) = -(1/3\sqrt{3})(q^2 + p^2)^3, \tag{2.41}$$

so that for such realizations, the two invariants cannot be assigned independent values [the cube of (2.40) equals the square of (2.41)].

Lastly, a simple construction for the elements of the $SL(3, R)$ Lie algebra in terms of the three pairs of canonical variables q_a, p_a is given by

$$J_a = \epsilon_{abc} q_b p_c, \tag{2.42}$$

$$Q'_{ab} = q_a q_b - p_a p_b - \tfrac{1}{3}\delta_{ab}(q^2 - p^2). \tag{2.43}$$

The two invariants have the values

$$J_a J_a - \tfrac{1}{2} Q'_{ab} Q'_{ab} = -\tfrac{1}{3}(q^2 - p^2)^2 \tag{2.44}$$

and

$$(\sqrt{3}/2)(3 J_a Q'_{ab} J_b + Q'_{ab} Q'_{bc} Q'_{ca}) = (1/3\sqrt{3})(q^2 - p^2)^3. \tag{2.45}$$

For such a realization also, the two invariants cannot be assigned independent values [the cube of (2.44) equals the negative of the square of (2.45)].

All these general constructions are made in terms of one set of three canonical pairs of variables. It is shown in the following sections that it is possible to express the elements of one realization as functions of the elements of another realization. In general, the functional forms involve algebraic functions, rather than polynomials. We therefore have to work with the elements of the generalized enveloping algebra which contains analytic functions (not only polynomials) of the generators. Within such a framework we show, in the following sections, that the generators of the Lie algebras of $O(4)$, $O(3, 1)$, $E(3)$, $SU(3)$, and $SL(3, R)$ can be expressed in terms of the elements of a generalized enveloping algebra of a given $E(3)$ realization.

3. REALIZATION OF $O(4)$, $O(3, 1)$, AND $E(3)$ LIE ALGEBRAS BY ANALYTIC FUNCTIONS OF P AND J

We begin by defining the $E(3)$ representation in terms of which we will construct $O(4)$, $O(3, 1)$, $SU(3)$, $SL(3, R)$, and other $E(3)$ generators. We consider two three-dimensional vectors $\mathbf{P}$ and $\mathbf{J}$ obeying the following basic Poisson-bracket relations:

$$\{J_a, J_b\} = \epsilon_{abc} J_c, \tag{3.1}$$

$$\{J_a, P_b\} = \epsilon_{abc} P_c, \tag{3.2}$$

$$\{P_a, P_b\} = 0. \tag{3.3}$$

Restriction to an irreducible realization of $E(3)$ implies that the two invariants $P^2 > 0$ and $\mathbf{J} \cdot \mathbf{P}$ are given by two preassigned real numbers. Since the multiplication of P_a by a constant does not change the Poisson-bracket relations (3.1)–(3.3), we can normalize P_a such that $P_a P_a = 1$. We then define the "phase space of $E(3)$" as the set of all pairs of real vectors $\mathbf{P}$ and $\mathbf{J}$ obeying the constraints

$$P^2 = P_a P_a = 1; \quad \mathbf{J} \cdot \mathbf{P} = J_a P_a = \alpha_0. \tag{3.4}$$

Because of these two constraints the phase space is now a four-dimensional space. However, we will continue to label the points of this phase space by two vectors $\mathbf{P}$ and $\mathbf{J}$. The finite canonical transformation generated by J_a and P_a are mappings of this phase space on to itself, and provide a realization of the group $E(3)$.

Under a finite canonical transformation generated by J_a, the point (P_a, J_a) is mapped to a new point (P'_a, J'_b) as follows:

$$\begin{aligned} P'_a &= \exp\widetilde{(\mathbf{n} \cdot \mathbf{J})} P_a = R_{ba}(\mathbf{n}) P_b, \\ J'_a &= \exp\widetilde{(\mathbf{n} \cdot \mathbf{J})} J_a = R_{ba}(\mathbf{n}) J_b. \end{aligned} \tag{3.5}$$

By $\exp(\tilde{\phi}) \cdot f$ we mean the following infinite series

$$\exp(\tilde{\phi}) \cdot f = f + \{\phi, f\} + (1/2!)\{\phi, \{\phi, f\}\} + \cdots,$$

where ϕ and f are arbitrary functions of the dynamical variables. Here the vector $\mathbf{n}$ specifies the transformation and $R(\mathbf{n})$ is the real orthogonal matrix corresponding to a rotation by an angle $|\mathbf{n}| = n$ about an axis in the direction of $\mathbf{n}$:

$$R_{ba}(\mathbf{n}) = \cos n \delta_{ba} + \frac{(1 - \cos n)}{n^2} n_b n_a + \frac{\sin n}{n} \epsilon_{abc} n_c. \tag{3.6}$$

It is easy to verify that P'_a, J'_a obey the same Poisson-bracket relations as P_a, J_a, so that the above transformation is in fact a canonical transformation. We denote this transformation by $(\mathbf{0}, R)$. Next consider a finite transformation generated by P_a:

$$\begin{aligned} P'_a &= \exp\widetilde{(\boldsymbol{\lambda} \cdot \mathbf{P})} P_a = P_a, \\ J'_a &= \exp\widetilde{(\boldsymbol{\lambda} \cdot \mathbf{P})} J_a = J_a + \epsilon_{abc} P_b \lambda_c. \end{aligned} \tag{3.7}$$

We denote this transformation by $(\boldsymbol{\lambda}, \mathbf{1})$. It can easily be verified that this transformation is also canonical. A general element of the group $E(3)$ is represented by

the canonical transformation $(\boldsymbol{\lambda}, R)$ obtained by performing $(\mathbf{0}, R)$ first and $(\boldsymbol{\lambda}, \mathbf{1})$ next:

$$\begin{aligned}(\boldsymbol{\lambda}, R) &\equiv (\boldsymbol{\lambda}, \mathbf{1})(\mathbf{0}, R),\\ P_a' &= R_{ba}P_b,\\ J_a' &= R_{ba}(J_b + \epsilon_{bcd}P_c\lambda_d).\end{aligned} \tag{3.8}$$

We have the following composition law:

$$(\boldsymbol{\lambda}', R')(\boldsymbol{\lambda}, R) = (\boldsymbol{\lambda}' + R'\boldsymbol{\lambda}, R'R), \tag{3.9}$$

where $R'\boldsymbol{\lambda}$ is the vector whose components are $R_{ab}\lambda_b$, $(a = 1, 2, 3)$. These finite canonical transformations map the entire phase space of $E(3)$ onto itself, and there is no region in the phase space which is invariant under all the transformations $(\boldsymbol{\lambda}, R)$. Further, we may verify that every transformation $(\boldsymbol{\lambda}, R)$ preserves the restriction $J^2 \geq \alpha_0^2$, which is implied by (3.4) and the reality of P_a and J_a.

We are now interested in constructing generators for $O(4)$ and other Lie algebras as analytic functions of P_a and J_a. We will find that, in general, these generators will be real in some regions of the underlying $E(3)$ phase space and complex in others. When we consider the finite canonical transformations arising from these generators, we can impose the following requirement: There exists some region in the phase space, which is mapped into itself under these canonical transformations; i.e., given any real point (P_a, J_a) in this region, the transformed point (P_a', J_a') also lies in this region, with P_a' and J_a' being real. We then obtain a representation of the group elements by means of finite real canonical transformations operating within this region of phase space. As is seen later, this requirement will, in general, impose further restrictions on the generators.

Let us begin with the construction of the $O(4)$ generators in items of the $E(3)$ generators. We choose to leave the $O(3)$ subalgebra unaltered, so that the first three generators are J_1, J_2, J_3. The most general form of the other three generators K_a $(a = 1, 2, 3)$ is given by

$$K_a = f_1P_a + f_2\epsilon_{abc}J_bP_c + gJ_a, \tag{3.10}$$

where f_1, f_2, and g are functions of J^2, to be determined. Equations (2.6) and (2.7) are automatically satisfied because of (3.1)–(3.3), whereas if we impose (2.8), we get a set of first-order differential equations[3] involving f_1, f_2, and g in their dependence on J^2. For this purpose, we rewrite (2.8) in a slightly different, but equivalent, form:

$$\epsilon_{abc}\{K_b, K_c\} = 2J_a. \tag{3.11}$$

[3] We follow the method of calculation to be found in H. Bacry, Nuovo Cimento **41A**, 221 (1966).

If we substitute (3.10) in (3.11) and also use (1.1) and (3.1)–(3.4), we obtain, after some long but straightforward calculations, the following relation:

$$\begin{aligned}2J_a = 2J_a\{&g^2 - f_2^2 - 2(f_1f_1' + J^2f_2f_2')\\ &- 2\alpha_0(f_1g' - \alpha_0f_2f_2')\}\\ &+ 4P_af_1\{g + J^2g' + \alpha_0f_1'\}\\ &+ 4\epsilon_{abc}J_bP_cf_2\{g + J^2g' + \alpha_0f_1'\},\end{aligned}$$

where primes denote differentiation with respect to J^2. If we multiply this last equation by $(\alpha_0J_a - J^2P_a)$, $(J_a - \alpha_0P_a)$, and $\epsilon_{aef}J_eP_f$ and sum over a in the resulting three equations, we obtain

$$f_1(g + J^2g' + \alpha_0f_1') = 0, \tag{3.12}$$

$$f_2(g + J^2g' + \alpha_0f_1') = 0, \tag{3.13}$$

$$1 = g^2 - f_2^2 - 2(f_1f_1' + J^2f_2f_2') - 2\alpha_0(f_1g' - \alpha_0f_2f_2'), \tag{3.14}$$

where we have assumed $J^2 - \alpha_0^2 \neq 0$. Apart from the trivial solution $f_1 = f_2 = 0$, $g = 1$ (i.e., $\mathbf{K} = \pm\mathbf{J}$), we obtain on solving (3.12)–(3.14) the following functional forms of f_1, f_2, and g:

$$f_1(J^2) = \left\{\frac{(\beta^2 - J^2)(J^2 - \alpha^2)}{(J^2 - \alpha_0^2)}\right\}^{\frac{1}{2}}\cos\{\Theta(J^2)\}, \tag{3.15}$$

$$f_2(J^2) = \left\{\frac{(\beta^2 - J^2)(J^2 - \alpha^2)}{J^2(J^2 - \alpha_0^2)}\right\}^{\frac{1}{2}}\sin\{\Theta(J^2)\}, \tag{3.16}$$

$$g(J^2) = \frac{\alpha\beta}{J^2} - \frac{\alpha_0}{J^2}\left\{\frac{(\beta^2 - J^2)(J^2 - \alpha^2)}{(J^2 - \alpha_0^2)}\right\}^{\frac{1}{2}}\cos\{\Theta(J^2)\}. \tag{3.17}$$

Here $\Theta(J^2)$ is an arbitrary function of J^2, and α and β are two real constants. This is the most general solution of (3.10), since the invariants

$$J^2 + K^2 = \alpha^2 + \beta^2 \tag{3.18}$$

and

$$\mathbf{J}\cdot\mathbf{K} = \alpha\beta \tag{3.19}$$

can be assigned independently. Without loss of generality, we can assume $\beta \geq |\alpha| \geq 0$. In analogy with the quantum-mechanical case, we call the $O(4)$ representation with $\mathbf{J}\cdot\mathbf{K} = 0$, the "symmetric traceless tensor representation."

We now discuss the above solution in some detail. We note first that, according to (3.18) and (3.19), the two invariants are entirely independent of the function Θ. It should therefore be possible to trace the arbitrariness associated with Θ to a freedom in the choice of the form of the generators. This is in fact true. The arbitrary function Θ simply reflects the freedom to make canonical transformations generated

by arbitrary functions $\phi(J^2)$ of J^2:

$$J_a \to J'_a = \exp\,[\widetilde{\phi(J^2)}]J_a = J_a,$$
$$P_a \to P'_a = \exp\,[\widetilde{\phi(J^2)}]P_a,$$
$$= \cos\,[2(J^2)^{\frac{1}{2}}\phi']P_a + \frac{1-\cos\,[2(J^2)^{\frac{1}{2}}\phi']}{J^2}(\mathbf{J}\cdot\mathbf{P})J_a + \frac{\sin\,[2(J^2)^{\frac{1}{2}}\phi']}{(J^2)^{\frac{1}{2}}}\epsilon_{abc}P_bJ_c. \quad (3.20)$$

By a proper choice of the function ϕ, we can eliminate Θ altogether and write the generators K_a in the form

$$K_a = \left\{\frac{(\beta^2-J^2)(J^2-\alpha^2)}{(J^2-\alpha_0^2)}\right\}^{\frac{1}{2}}P_a + \left[\alpha\beta - \alpha_0\left\{\frac{(\beta^2-J^2)(J^2-\alpha^2)}{(J^2-\alpha_0^2)}\right\}^{\frac{1}{2}}\right]\frac{J_a}{J^2}. \quad (3.21)$$

Next we consider the reality properties of K_a and of the canonical transformations generated by K_a. Since $J^2 \geq \alpha_0^2$ in the phase space of $E(3)$, we must choose $\beta^2 > \alpha_0^2$ in order to have some region,

$$\beta^2 \geq J^2 \geq \max\,(\alpha^2, \alpha_0^2),$$

where K_a is real. To discuss the nature of the canonical transformations $(P_a, J_a) \to (P'_a, J'_a)$, we first rewrite (3.21) in the following compact form:

$$\mathbf{K} = \left\{\frac{(\beta^2-J^2)(J^2-\alpha^2)}{(J^2-\alpha_0^2)}\right\}^{\frac{1}{2}}\frac{(\mathbf{J}\times\mathbf{P}\times\mathbf{J})}{J^2} + \frac{\alpha\beta}{J^2}\mathbf{J}, \quad (3.22)$$

from which we can express $\mathbf{P}$ in terms[4] of $\mathbf{K}$ and $\mathbf{J}$:

$$\mathbf{P} = \left\{\frac{(J^2-\alpha_0^2)}{(\beta^2-J^2)(J^2-\alpha^2)}\right\}^{\frac{1}{2}}\frac{(\mathbf{J}\times\mathbf{K}\times\mathbf{J})}{J^2} + \frac{\alpha_0}{J^2}\mathbf{J}. \quad (3.23)$$

Thus, given real vectors $\mathbf{J}$, $\mathbf{P}$ with

$$\beta^2 \geq J^2 \geq \max\,(\alpha^2, \alpha_0^2)$$

(and $P^2 = 1$, $\mathbf{J}\cdot\mathbf{P} = \alpha_0$), $\mathbf{K}$ is real; and conversely, given real vectors $\mathbf{J}$, $\mathbf{K}$ with $\beta^2 \geq J^2 \geq \max\,(\alpha^2, \alpha_0^2)$ (and $J^2 + K^2 = \alpha^2 + \beta^2$, $\mathbf{J}\cdot\mathbf{K} = \alpha\beta$), $\mathbf{P}$ is real. The finite canonical transformations generated by $\mathbf{J}$ and $\mathbf{K}$ can be shown to represent orthogonal rotations on the two vectors $\mathbf{J} \pm \mathbf{K}$:

$$J_a \pm K_a \to J'_a \pm K'_a = \exp\,(\widetilde{\mathbf{n}\cdot\mathbf{J}})(J_a \pm K_a) = R_{ba}(\mathbf{n})(J_b \pm K_b), \quad (3.24)$$

$$J_a \pm K_a \to J'_a \pm K'_a = \exp\,(\widetilde{\boldsymbol{\lambda}\cdot\mathbf{K}})(J_a \pm K_a) = R_{ba}(\pm\boldsymbol{\lambda})(J_b \pm K_b), \quad (3.25)$$

where R_{ba} is the orthogonal matrix given by (3.6). The magnitudes of these vectors are thus fixed:

$$(\mathbf{J} \pm \mathbf{K})^2 = (\beta \pm \alpha)^2. \quad (3.26)$$

Therefore, if we start with a point (P_a, J_a) with $\beta^2 \geq J^2 \geq \max\,(\alpha^2, \alpha_0^2)$, then, by choosing an appropriate canonical transformation in $O(4)$ of the type (3.25), we get an image (P'_a, J'_a) with J'^2 lying anywhere between the values[5] β^2 and α^2 (i.e., $\beta^2 \geq J'^2 \geq \alpha^2$). However, if P_a is real, we also have $J'^2 \geq \max\,(\alpha^2, \alpha_0^2)$, and therefore $\alpha^2 \geq \alpha_0^2$.

We have then the following result: Given the phase space of $E(3)$ with a certain α_0, the generators K_a of (3.21), for $\beta \geq |\alpha| \geq |\alpha_0|$, are real in the region $\beta^2 \geq J^2 \geq \alpha^2$. The finite canonical transformation generated by J_a and K_a map this region into itself and provide a representation of the group $O(4)$. If $|\alpha| < |\alpha_0|$, the K_a are real in the region $\beta^2 \geq J^2 \geq \alpha_0^2$; however, in this case there is no region in the phase space which is invariant under the transformation generated by K_a. [Alternatively, for every finite transformation generated by K_a, there exist some real points (P_a, J_a) which are carried into points (P'_a, J'_a) with complex P'_a.] As a particular case we see that, in order to obtain a real representation of $O(4)$ of the "symmetric traceless tensor" type with $\alpha = 0$, we must start with $\alpha_0 = 0$.

Let us now consider the region of the phase space where the generators (3.21) are not real. The generators K_a become complex outside the region $\beta^2 \geq J^2 \geq \alpha^2$ (with $\beta \geq |\alpha| > |\alpha_0|$), and we do not have a real realization of any Lie algebra. If, however, we make an "analytic continuation" of the parameter α to pure imaginary values, we can generate real realizations of the $O(3, 1)$ Lie algebra in the region $J^2 \geq \beta^2$. We define $K'_a = iK_a$ and simultaneously put $\alpha' = i\alpha$ (with α' real), and obtain, from (3.21), the following expression for K'_a:

$$K'_a = \left\{\frac{(J^2-\beta^2)(J^2+\alpha'^2)}{(J^2-\alpha_0^2)}\right\}^{\frac{1}{2}}P_a + \left[\frac{\alpha'\beta}{J^2} - \frac{\alpha_0}{J^2}\left\{\frac{(J^2-\beta^2)(J^2+\alpha'^2)}{(J^2-\alpha_0^2)}\right\}^{\frac{1}{2}}\right]J_a, \quad (3.27)$$

which is real for $J^2 \geq \beta^2$ (remember also that $\beta^2 \geq \alpha_0^2$). Since J_a and K'_a now satisfy the Poisson-bracket relations (2.11)–(2.13), they generate the

[4] Since we have obtained an expression for $\mathbf{P}$ in terms of $\mathbf{K}$ and $\mathbf{J}$, we can, in all of the present discussion, replace $\mathbf{P}$ by (3.23) and thus obtain realizations of various Lie algebras in terms of analytic functions of $\mathbf{K}$ and $\mathbf{J}$, the generators of $O(4)$.

[5] The fact that β^2 and α^2 are the maximum and minimum values, respectively, attained by J^2 under the finite canonical transformations generated by $\mathbf{K}$ can also be seen from the relation

$$0 = \delta J^2 = \exp\,(\widetilde{\delta\boldsymbol{\lambda}\cdot\mathbf{K}})J^2 - J^2 = (\mathbf{J}\times\mathbf{K})\cdot\delta\boldsymbol{\lambda}$$

and (3.22).

$O(3, 1)$ Lie algebra. The two invariants in this case are

$$J^2 - K'^2 = \beta^2 - \alpha'^2, \tag{3.28}$$

$$\mathbf{J} \cdot \mathbf{K}' = \beta\alpha'. \tag{3.29}$$

We can evaluate the minimum value attained by J^2 under finite canonical transformation generated by $\mathbf{K}'$. Let (P_a, J_a) be the point with a minimum value of J^2. Under an infinitesimal transformation generated by $\mathbf{K}'$, we have

$$\delta(J^2) = 2\mathbf{J} \cdot \delta\mathbf{J} = 2J_a\{\delta\boldsymbol{\lambda} \cdot \mathbf{K}', J_a\} = 2\delta\boldsymbol{\lambda} \cdot (\mathbf{J} \times \mathbf{K}'). \tag{3.30}$$

For J^2 to be minimum, $\delta(J^2) = 0$, and we then find from (3.28)–(3.30) that this minimum value is β^2. Thus, under the finite canonical transformations generated by $\mathbf{J}$ and $\mathbf{K}'$, the region $J^2 \geq \beta^2$ is mapped into itself and we have a real realization of the group $O(3, 1)$.

The above discussion also holds for the case $\alpha = \alpha_0 = 0$, i.e., the generators J_a and K'_a given by

$$K'_a = \{J^2 - \beta^2\}^{\frac{1}{2}} P_a \tag{3.31}$$

provide a real realization of the group $O(3, 1)$ in the region $J^2 \geq \beta^2$. We note, however, that if we replace β^2 by $-\beta^2$ in (3.31), we obtain a realization of the $O(3, 1)$ Lie algebra, real over the entire phase space of $E(3)$. In this case, the entire phase space is mapped into itself under all finite canonical transformations generated by $\mathbf{J}$, or $\mathbf{K}' = \{J^2 + \beta^2\}^{\frac{1}{2}} P_a$.

To summarize, then, starting with an $E(3)$ realization with a given α_0 and with the parameters β, α obeying $\beta \geq |\alpha| \geq |\alpha_0|$, J_a and K_a of (3.21) generate a real realization of the group $O(4)$ in the region $\beta^2 \geq J^2 \geq \alpha^2$. Outside this region, the K_a are complex. One can analytically continue α to $\alpha' = i\alpha$ with α' real, and obtain K'_a of Eq. (3.27), which together with J_a generate a real realization of the group $O(3, 1)$ in the region $J^2 \geq \beta^2$. If $\alpha = \alpha_0 = 0$, we also have a realization of the group $O(3, 1)$, real over the entire phase space.

We conclude this section by considering the "contraction" of the $O(4)$ generators of the equation (3.21) to yield new generators of $E(3)$. For this purpose we set

$$K_a = \beta\bar{P}_a \quad (1 \geq a \geq 3), \tag{3.32}$$

and take the limit $\beta \to \infty$, keeping α fixed. If we substitute (3.32) in the Poisson-bracket relations (2.6)–(2.8) and take the limit $\beta \to \infty$, we find that J_a and $\bar{P}_a$ satisfy (2.1)–(2.3), required of the generators of $E(3)$. Thus J_a and

$$\bar{P}_a = \left[\frac{(J^2 - \alpha^2)}{(J^2 - \alpha_0^2)}\right]^{\frac{1}{2}} P_a + \left[\frac{\alpha}{J^2} - \frac{\alpha_0}{J^2}\left[\frac{(J^2 - \alpha^2)}{(J^2 - \alpha_0^2)}\right]^{\frac{1}{2}}\right] J_a, \tag{3.33}$$

obtained from (3.32) and (3.21) in the limit $\beta \to \infty$, provide a realization of the $E(3)$ Lie algebra. Thus, starting with a realization of $E(3)$ with the generator P_a, J_a and invariants $P^2 = 1$, $\mathbf{J} \cdot \mathbf{P} = \alpha_0$, we have exhibited a realization of $E(3)$ Lie algebra by $\bar{P}_a$, J_a with arbitrary value α for $\mathbf{J} \cdot \bar{\mathbf{P}}$ (and $\bar{P}^2 = 1$). It must be noted that since the value of the invariant $\mathbf{J} \cdot \mathbf{P}$ is changed, (3.33) does not represent a canonical transformation generated by any function of $\mathbf{J}$ or $\mathbf{P}$. It may be seen that the minimum value reached by J^2 under finite canonical transformation generated by $\bar{\mathbf{P}}$ is α^2. Hence, if $|\alpha| \geq |\alpha_0|$, the region $J^2 \geq \alpha^2$, where $\bar{P}_a$ is real, is mapped into itself by the finite canonical transformation generated by $\bar{\mathbf{P}}$ or $\mathbf{J}$, and thus we obtain a real realization of $E(3)$ in this region. On the other hand, if $|\alpha| < |\alpha_0|$, for every finite canonical transformation generated by $\bar{\mathbf{P}}$, there are some real points (P_a, J_a) which are carried into image points (P'_a, J'_a) with complex P'_a.

4. REALIZATION OF $SU(3)$ AND $SL(3, R)$ LIE ALGEBRAS BY ANALYTIC FUNCTIONS OF P AND J

In this section we discuss the realization of $SU(3)$ and $SL(3, R)$ Lie algebras from the generalized enveloping algebra of $E(3)$. The $O(3)$ subalgebra of $SU(3)$, $SL(3, R)$, and $E(3)$ will be taken to be identical, i.e., three of the generators of $SU(3)$ and $SL(3, R)$ are chosen to be J_a $(a = 1, 2, 3)$. The other five generators of $SU(3)$ and $SL(3, R)$ (i.e., the symmetric traceless tensors Q_{ab} and Q'_{ab}, respectively,) are to be determined as functions of $\mathbf{P}$ and $\mathbf{J}$.

It is known that, in general, the unitary irreducible (matrix) representations of the group $SU(3)$ are not only reducible with respect to $O(3)$, but the same $O(3)$ representation may appear more than once.[6] However, for the special class of "completely symmetric tensor" representations, there is no such multiplicity, and only states with the same parity occur. We restrict our considerations to Poisson-bracket "symmetric tensor realizations" of $SU(3)$ and $SL(3, R)$ with even parity [i.e., $Q_{ab}(\mathbf{P}) = Q_{ab}(-\mathbf{P})$ and $Q'_{ab}(\mathbf{P}) = Q'_{ab}(-\mathbf{P})$] from amongst the elements of the generalized enveloping algebra of $E(3)$.

We show below [Eq. (4.26)] that, in any real representation of the group $SU(3)$ with the generator J_a, Q_{ab}, the minimum value of J^2 is always zero. A similar result also holds for $SL(3, R)$. From the examples discussed up to now, we can then conclude that we may restrict ourselves to an underlying $E(3)$ realization with

$$\mathbf{J} \cdot \mathbf{P} = \alpha_0 = 0; \quad P^2 = 1. \tag{4.1}$$

[6] See, for instance, G. Racah, Rev. Mod. Phys. **21**, 494 (1949); V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961).

If the symmetric traceless tensor Q_{ab} is to be constructed from the vectors $\mathbf{P}$ and $\mathbf{J}$ satisfying (4.1), then the most general form for Q_{ab} with even parity $Q_{ab}(\mathbf{P}) = Q_{ab}(-\mathbf{P})$ is given by

$$Q_{ab} = f_0 \cdot \{P_aP_b - \tfrac{1}{3}\delta_{ab}\} + g_0 \cdot \left(\frac{J_aJ_b}{J^2} - \tfrac{1}{3}\delta_{ab}\right) + h_0\{(\mathbf{J}\times\mathbf{P})_aP_b + P_a(\mathbf{J}\times\mathbf{P})_b\}, \quad (4.2)$$

where f_0, g_0, and h_0 are some functions of J^2 to be determined. The absence of a term proportional to

$$(\mathbf{J}\times\mathbf{P})_a(\mathbf{J}\times\mathbf{P})_b - \tfrac{1}{3}J^2\delta_{ab} \equiv \epsilon_{acd}\epsilon_{bef}J_cP_dJ_eP_f - \tfrac{1}{3}J^2\delta_{ab} \quad (4.3)$$

in (4.2) is no loss of generality, since such a term can be re-expressed in terms of δ_{ab}, J_aJ_b, and P_aP_b, which are included in (4.2). This can be seen immediately, if we make use of the identity

$$\epsilon_{acd}\epsilon_{bef} = \delta_{ab}\delta_{ce}\delta_{df} + \delta_{ae}\delta_{cf}\delta_{db} + \delta_{af}\delta_{cb}\delta_{de} - \delta_{ab}\delta_{cf}\delta_{de} - \delta_{ae}\delta_{cb}\delta_{df} - \delta_{af}\delta_{ce}\delta_{db}. \quad (4.4)$$

In order to determine the three functions f_0, g_0, h_0 we must impose the Poisson-bracket relations (2.19). [The relations (2.16)–(2.18) are automatically satisfied with the choice (4.2).] However, it is possible to simplify Q_{ab} first by means of a canonical transformation of the form (3.20), leaving J_a invariant. As we show below, a proper choice of the function $\phi(J^2)$ in (3.20) can eliminate the term proportional to

$$(\mathbf{J}\times\mathbf{P})_aP_b + (\mathbf{J}\times\mathbf{P})_bP_a$$

in (4.2). For this purpose, we rewrite (4.2) with P_a, J_a replaced by P'_a, J'_a:

$$Q_{ab} = f_0(P'_aP'_b - \tfrac{1}{3}\delta_{ab}) + g_0((J'_aJ'_b/J^2) - \tfrac{1}{3}\delta_{ab}) + h_0\{(\mathbf{J}'\times\mathbf{P}')_aP'_b + (\mathbf{J}'\times\mathbf{P}')_bP'_a\}, \quad (4.5)$$

where $\mathbf{P}'$ and $\mathbf{J}'$ are given by (3.20), with $\mathbf{J}\cdot\mathbf{P}$, i.e.,

$$J'_a = \exp\{\widetilde{\phi(J^2)}\}J_a = J_a,$$
$$P'_a = \exp\{\widetilde{\phi(J^2)}\}P_a,$$
$$= \cos\Theta P_a - [\sin\Theta/(J^2)^{\frac{1}{2}}](\mathbf{J}\times\mathbf{P})_a, \quad (4.6)$$

where $\Theta = 2(J^2)^{\frac{1}{2}}\,d\phi(J^2)/d(J^2)$. If we substitute (4.6) in (4.5), we obtain

$$Q_{ab} = f(P_aP_b - \tfrac{1}{3}\delta_{ab}) + g\{(J_aJ_b/J^2) - \tfrac{1}{3}\delta_{ab}\} + h\{(\mathbf{J}\times\mathbf{P})_aP_b + (\mathbf{J}\times P)_bP_a\}, \quad (4.7)$$

where

$$f = f_0\cos 2\Theta + 2h_0(J^2)^{\frac{1}{2}}\sin 2\Theta, \quad (4.8)$$
$$g = g_0 - f_0\sin^2\Theta + h_0(J^2)^{\frac{1}{2}}\sin 2\Theta, \quad (4.9)$$
$$h = h_0\cos 2\Theta - [f_0/2(J^2)^{\frac{1}{2}}]\sin 2\Theta. \quad (4.10)$$

We now choose $\Theta = \frac{1}{2}\tan^{-1}[2h_0(J^2)^{\frac{1}{2}}/f_0]$ so that $h = 0$, and (4.7) then shows that we can restrict our considerations to the case when Q_{ab} is of the form

$$Q_{ab} = f(P_aP_b - \tfrac{1}{3}\delta_{ab}) + g[(J_aJ_b/J^2) - \tfrac{1}{3}\delta_{ab}]. \quad (4.11)$$

The functional dependence of f and g on J^2 is to be determined from the Poisson-bracket relation (2.19). We deduce from (2.19) the essentially equivalent relation

$$\epsilon_{hde}\{Q_{ah}, Q_{cd}\} = 4\delta_{ac}J_e - \delta_{ce}J_a - \delta_{ae}J_c \quad (4.12)$$

and substitute (4.11) in (4.12). If we also use Eqs. (1.1), (3.1)–(3.3), and (4.1), we find, after long but straightforward calculations, that

$$\begin{aligned}4\delta_{ac}J_e - (\delta_{ce}J_a + \delta_{ae}J_c) &= -\tfrac{4}{3}f(f' + g')\delta_{ac}J_e \\ &+ [4/(J^2)^2](J^2fg' - fg + g^2)J_aJ_cJ_e \\ &+ \left\{\tfrac{2}{3}f(f' + g') + \frac{fg}{J^2} - \frac{g^2}{J^2} - 2fg'\right\} \\ &\times (\delta_{ce}J_a + \delta_{ae}J_c) + \tfrac{4}{3}(2fg' - ff') \\ &\times \{P_aP_cP_e + (P_aJ_c + J_aP_c)J_e\}, \quad (4.13)\end{aligned}$$

where primes again denote differentiation with respect to J^2. Equating the coefficients of the various terms, we find that (4.13) is satisfied if and only if

$$f(f' + g') = -3, \quad (4.14)$$
$$f(f' - 2g') = 0, \quad (4.15)$$

and

$$g(g - f) = J^2. \quad (4.16)$$

It is possible to find a solution of the three equations (4.14)–(4.16) for the two functions f, g:

$$f = (\beta^2 - 4J^2)^{\frac{1}{2}}, \quad (4.17)$$
$$g = \tfrac{1}{2}(\beta^2 - 4J^2)^{\frac{1}{2}} \pm \tfrac{1}{2}\beta, \quad (4.18)$$

where β is some arbitrary constant. From (4.11), (4.17), and (4.18), we thus obtain

$$Q_{ab} = (\beta^2 - 4J^2)^{\frac{1}{2}}[P_aP_b + (J_aJ_b/2J^2) - \tfrac{1}{3}\delta_{ab}] \pm \tfrac{1}{2}\beta[(J_aJ_b/J^2) - \tfrac{1}{3}\delta_{ab}]. \quad (4.19)$$

This solution for Q_{ab} obeys the Poisson-bracket relation (2.19). The parameter β and the ambiguity of the sign in (4.19) are related to the quadratic and cubic invariants [cf. (2.25), (2.26)]

$$J_aJ_a + \tfrac{1}{2}Q_{ab}Q_{ab} = \tfrac{1}{3}\beta^2, \quad (4.20)$$
$$\tfrac{1}{2}\sqrt{3}\,\{3J_aQ_{ab}J_b - Q_{ab}Q_{bc}Q_{ca}\} = \pm(\beta^3/3\sqrt{3}). \quad (4.21)$$

As mentioned in Sec. 2, we note that the cube of the quadratic invariant is equal to the square of the cubic invariant.

The generators Q_{ab} in (4.19) are real in the region $J^2 \leq \frac{1}{4}\beta^2$. In order to discuss the finite transformations

generated by J_a and Q_{ab}, we first evaluate the maximum and minimum values attained by J^2 under these transformations.

Let us evaluate the minimum value of J^2. If we just use the Poisson-bracket relations (2.17)–(2.19), we see that, under the finite canonical transformations generated by J_a or Q_{ab}, these quantities go over into certain linear combinations of themselves. In fact, J_a and Q_{ab} transform according to the eight-dimensional adjoint representation of $SU(3)$.[7] To exhibit these transformations we introduce three antisymmetric Hermitian (3×3) matrices Λ^a $(a = 1, 2, 3)$:

$$\Lambda^1 = i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad \Lambda^2 = i\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\Lambda^3 = i\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.22}$$

and the Hermitian symmetric traceless matrices Λ^{ab}:

$$\Lambda^{ab} = \tfrac{1}{2}(\Lambda^a\Lambda^b + \Lambda^b\Lambda^a) - \tfrac{2}{3}\delta_{ab}\mathbf{1}, \tag{4.23}$$

of which only five are linearly independent. Together, Λ^a and Λ^{ab} form a basis for traceless Hermitian (3×3) matrices. Given the variables J_a, Q_{ab}, we form the Hermitian matrix

$$A = J_a\Lambda^a + \tfrac{1}{2}Q_{ab}\Lambda^{ab}. \tag{4.24}$$

Now let U be any unitary unimodular matrix. The matrix $A' = UAU^{-1}$ can also be expanded linearly in terms of Λ^a and Λ^{ab}, i.e.,

$$A' = UAU^{-1} = J'_a\Lambda^a + \tfrac{1}{2}Q'_{ab}\Lambda^{ab}, \tag{4.25}$$

where the coefficients J'_a and Q'_{ab} are linear combinations of J_a and Q_{ab}. By choosing all possible matrices U, we get precisely all those linear combinations J'_a, Q'_{ab} that are obtained by performing all possible finite canonical transformations generated by J_a and Q_{ab} on themselves. Now, given any A, we can choose a U such that A' is diagonal. In that case J_a vanishes, since Λ^a are antisymmetric and Λ^{ab} are symmetric matrices. Thus, starting with any value of J_a, Q_{ab}, there exists a particular transformation generated by J_a, Q_{ab} which takes

$$J_a \to J'_a = 0. \tag{4.26}$$

This proves the statement that in any realization of $SU(3)$, the value $J^2 = 0$ is always attained and this is, of course, the minimum value.

Let us next evaluate the maximum value of J^2. Let (P_a, J_a) be the point where J^2 is maximum. We first perform an orthogonal rotation (3.5), generated by J_a, which leaves J^2 unchanged, such that

$$J_1 = J_2 = 0; \quad J_3 = (J^2)^{\frac{1}{2}}. \tag{4.27}$$

Now under an arbitrary infinitesimal transformation generated by Q_{ab},

$$J_a \to \exp\,\widetilde{(\delta\lambda_{bc} \cdot Q_{bc})}J_a, \tag{4.28}$$

where the $\delta\lambda_{bc}$ are arbitrary, we must have

$$\delta(J^2) = 0. \tag{4.29}$$

Using the Poisson-bracket relation (2.18), we obtain

$$\begin{aligned}\delta(J^2) &= 2J_a(\delta J_a) = 2J_a\{\delta\lambda_{bc}Q_{bc}, J_a\} \\ &= 2J_a[\epsilon_{bad}Q_{dc} + \epsilon_{cad}Q_{bd}]\delta\lambda_{bc},\end{aligned} \tag{4.30}$$

so that (4.28) is satisfied if and only if

$$J_a(\epsilon_{bad}Q_{dc} + \epsilon_{cad}Q_{bd}) = 0. \tag{4.31}$$

If we use (4.27) and the fact that $Q_{ab} = Q_{ba}$, $Q_{cc} = 0$, in (4.31), we obtain the following form for Q_{ab}:

$$Q_{11} = Q_{22} = -\tfrac{1}{2}Q_{33} = q; \quad Q_{ab} = 0, \quad a \neq b, \tag{4.32}$$

where q is some constant. Using the special forms of J_a and Q_{ab}, as given by (4.27) and (4.32) in (4.20) and (4.21), we get the following two equations in J^2 and q:

$$J^2 + 3q^2 = \tfrac{1}{3}\beta^2, \tag{4.33}$$

$$-J^2q + q^3 = \pm\tfrac{1}{27}\beta^3, \tag{4.34}$$

with the solution $J^2 = 0$, $q = \pm\frac{1}{3}\beta$, or $J^2 = \frac{1}{4}\beta^2$, $q = \pm\frac{1}{6}\beta$. The former solution corresponds to the minimum value of J^2 [note that (4.29) is also satisfied when J^2 is minimum], which we already derived earlier, as the latter solution corresponds to the maximum value of J^2.

It is thus seen that the maximum and minimum values of J^2 are $\frac{1}{4}\beta^2$ and 0, respectively, which are just the boundaries of the region where Q_{ab} are real. We therefore conclude that not only are the Q_{ab} real in the region $0 \leq J^2 \leq \frac{1}{4}\beta^2$ of the $E(3)$ phase space, but, in fact, the finite canonical transformation generated by J_a, Q_{ab} carries this region into itself and provides a real realization of the group $SU(3)$.

We conclude this section by a brief discussion of the realization of the $SL(3, R)$ Lie algebra. We see from (4.19) that Q_{ab} become complex in the region $J^2 \gtrsim \frac{1}{4}\beta^2$. We can, however, redefine the generators $Q'_{ab} = iQ_{ab}$ and simultaneously analytically continue β to $\beta' = i\beta$

[7] For any Lie group, the generators transform according to the adjoint representation of the group; for $SU(3)$, this is the octet or eight-dimensional representation. The matrices Λ^a, Λ^{ab} are the Hermitian generators of the three-dimensional representation of $SU(3)$. The generators Λ^a of the $O(3)$ subgroup correspond to the spin-1 representation of $O(3)$, and are here represented in the Cartesian form.

(with β' real) to get

$$Q'_{ab} = (\beta'^2 + 4J^2)^{\frac{1}{2}}\{P_aP_b + \tfrac{1}{2}(J_aJ_b/J^2) - \tfrac{1}{2}\delta_{ab}\} \pm \tfrac{1}{2}\beta((J_aJ_b/J^2) - \tfrac{1}{3}\delta_{ab}). \quad (4.35)$$

J_a and Q_{ab} now obey the Poisson-bracket relations (2.22)–(2.24) corresponding to the unimodular real linear group in three dimensions, $SL(3, R)$. The two invariants of this representation are

$$J_aJ_a - \tfrac{1}{2}Q'_{ab}Q'_{ab} = -\tfrac{1}{3}\beta'^2 \quad (4.36)$$

and

$$\tfrac{1}{2}\sqrt{3}\,(3J_aQ'_{ab}J_b + \tfrac{1}{3}Q'_{ab}Q'_{bc}Q'_{ca}) = \mp\beta^3/3\sqrt{3}. \quad (4.37)$$

The generators Q'_{ab} are real over the entire region $J^2 \geq 0$.

Let us now evaluate the limiting values attained by J^2 under finite canonical transformations generated by J_a and Q'_{ab}. Let (P_a, J_a) be the point where J^2 is stationary and let us first perform an orthogonal rotation generated by J_a (leaving J^2 unchanged) such that J_1, J_2, J_3 are given by (4.27). By following a strictly similar argument as was used to obtain the relation (4.32), we obtain in this case

$$Q'_{11} = Q'_{22} = -\tfrac{1}{2}Q''_{33} = q', \quad Q'_{ab} = 0, \quad a \neq b, \quad (4.38)$$

where q' is some constant. If we use the special forms of J_a and Q'_{ab} as given by (4.27) and (4.38) in (4.36) and (4.37), we obtain the following two equations in J^2 and q':

$$J^2 - 3q'^2 = \tfrac{1}{3}\beta'^2, \quad (4.39)$$

$$J^2q' + q'^3 = \pm\tfrac{1}{27}\beta^3, \quad (4.40)$$

with the only real solution $J^2 = 0$, $q' = \pm\frac{1}{3}\beta'$.

Hence we conclude that J_a and Q'_{ab} generate real finite canonical transformations mapping the entire phase space into itself and providing a realization of the group $SL(3, R)$.

5. GENERALIZATION TO n DIMENSIONS

So far we have restricted our attention to the realization of some Lie algebras in terms of the elements of a generalized enveloping algebra of $E(3)$. In this section, we wish to make some comments about its generalization to n dimensions. There are some special features of the $E(3)$ and $O(4)$ algebras which do not generalize to higher dimensions. However, the symmetric-tensor-type realizations permit an immediate extension to arbitrary dimensions.

We start with the symmetric-tensor realization of $E(n)$, $(n \geq 3)$, with the generators P_a, J_{ab}, $[J_{ab} = -J_{ba}; a, b = 1, 2, \cdots, n]$, which obey the Poisson-bracket relations

$$\{J_{ab}, J_{cd}\} = \delta_{ac}J_{bd} + \delta_{bd}J_{ac} - \delta_{bc}J_{ad} - \delta_{ad}J_{bc}, \quad (5.1)$$

$$\{J_{ab}, P_c\} = \delta_{ac}P_b - \delta_{bc}P_a, \quad (5.2)$$

$$\{P_a, P_b\} = 0. \quad (5.3)$$

We will restrict our discussion to the case when the totally antisymmetric tensors

$$H_{abc} = P_aJ_{bc} + P_bJ_{ca} + P_cJ_{ab} \quad (5.4)$$

and

$$G_{abcd} = J_{ab}J_{cd} + J_{ac}J_{db} + J_{ad}J_{bc} \quad (5.5)$$

identically vanish. For $n = 3$, the constraint (5.5) is empty, whereas the constraint (5.4) reduces to $\mathbf{J} \cdot \mathbf{P} = 0$, which implies that the "helicity" is zero. We also choose the normalization such that

$$P^2 = 1, \quad (5.6)$$

which is always permissible since the multiplication by a constant does not change the Poisson-bracket relations (5.1)–(5.3).

The generators J_{ab} and K_a of $O(n + 1)$ obey the Poisson-bracket relations

$$\{J_{ab}, K_c\} = \delta_{ac}K_b - \delta_{bc}K_a, \quad (5.7)$$

$$\{K_a, K_b\} = J_{ab}, \quad (5.8)$$

and the Poisson bracket of J_{ab} with J_{cd} is given by (5.1).

From (3.21), if we set $\alpha = \alpha_0 = 0$, we can immediately write down the generators K_a in terms of P_a and J_{ab} $(1 \leq a \leq n)$:

$$K_a = (\beta^2 - J^2)^{\frac{1}{2}}P_a, \quad (5.9)$$

where

$$J^2 = \tfrac{1}{2}J_{ab}J_{ab} \quad (5.10)$$

and β is related to the invariant

$$J^2 + K^2 = \tfrac{1}{2}J_{ab}J_{ab} + K_aK_a = \beta^2. \quad (5.11)$$

By direct calculations, it may be verified that the Poisson-bracket relations (5.7) and (5.8) are satisfied and therefore J_{ab} and K_a do serve as generators of $O(n + 1)$. Of course, the form of K_a can be changed, by finite canonical transformations generated by some function of J^2, to

$$K_a = (\beta^2 - J^2)^{\frac{1}{2}}\left\{P_a \cos\Theta + \frac{J_{ab}P_b}{J^2}\sin\Theta\right\}. \quad (5.12)$$

The generator K_a is real in the region $J^2 \leq \beta^2$. For $J^2 > \beta^2$, K_a becomes pure imaginary. In this region

$$K'_a = iK_a = (J^2 - \beta^2)^{\frac{1}{2}}P_a \quad (5.13)$$

and J_{ad} generate a representation of the pseudo-orthogonal group $O(n, 1)$. The generators $K'_a = (J^2 + \beta^2)^{\frac{1}{2}}P_a$ and J_{ab} generate a representation of

$O(n, 1)$ which is real over the entire phase space. [The Poisson-bracket relations of K_a and J_{ab} are similar to (5.7) and (5.8) except for a change of sign in (5.8).]

The $SU(n)$ generators J_{ab}, Q_{ab} obey the Poisson-bracket relations

$$\{J_{ab}, Q_{cd}\} = +\delta_{ac}Q_{bd} + \delta_{ad}Q_{cb} - \delta_{cb}Q_{ad} - \delta_{bd}Q_{ca}, \tag{5.14}$$

$$\{Q_{ab}, Q_{cd}\} = \delta_{bd}J_{ac} + \delta_{ad}J_{bc} + \delta_{bc}J_{ad} + \delta_{ac}J_{bd}. \tag{5.15}$$

In analogy with (4.19) we can again write down Q_{ab} in terms of P_a and J_{ab}:

$$Q_{ab} = (\beta^2 - 4J^2)^{\frac{1}{2}}\left(P_aP_b - \frac{1}{n}\delta_{ab}\right) + \tfrac{1}{2}[(\beta^2 - 4J^2)^{\frac{1}{2}} \pm \beta]\left(\frac{2}{n}\delta_{ab} - \frac{J_{ac}J_{bc}}{J^2}\right), \tag{5.16}$$

where the parameter β is related to the invariant

$$J^2 + \tfrac{1}{2}Q_{ab}Q_{ab} = \tfrac{1}{2}(1 - n^{-1})\beta^2 \tag{5.17}$$

and J^2 is again given by (5.10). From the structure of Q_{ab} in (5.16), it is evident that the Poisson-bracket relation (5.14) is obeyed. It is expected that (5.15) also holds.

The generators Q_{ab} are real in the region $J^2 \leq \frac{1}{4}\beta^2$, and together with J_{ab} provide a real realization of $SU(n)$ in this region. For $J^2 > \frac{1}{4}\beta^2$, Q_{ab} of (5.16) become complex.

Defining $Q'_{ab} = iQ_{ab}$ and choosing $\beta' = i\beta$ (β' real) as before, we obtain from (5.16) the generators of a noncompact group $SL(n, R)$:

$$Q'_{ab} = (\beta'^2 + 4J^2)^{\frac{1}{2}}\left(P_aP_b - \frac{J_{ac}J_{bc}}{2J^2}\right) \pm \tfrac{1}{2}\beta\left(\frac{2}{n}\delta_{ab} - \frac{J_{ac}J_{bc}}{J^2}\right). \tag{5.18}$$

[Q'_{ab} and J_{ab} obey similar Poisson-bracket relations as (5.14), (5.15), except a change of sign in (5.15).] This realization is real over the entire phase space $J^2 > 0$.

It may be noted that for "symmetric tensor" realizations of $O(n + 1)$ and $SU(n)$ and similarly for $O(n, 1)$ and $SL(n, R)$, the quadratic invariants (5.11) and (5.17), respectively, essentially determine the realizations, except for the automorphism

$$J_{ab} \to J_{ab}, \quad K_a \to -K_a, \quad Q_{ab} \to -Q_{ab}, \tag{5.19}$$

which leave all even-degree invariants unaltered, but change the signs of all odd-degree invariants.

6. CONCLUDING REMARKS

We have discussed the realization of certain Lie algebras in terms of the generalized enveloping algebras of certain other prescribed Lie algebras. We now make a few comments on the realizations of Lie algebras in terms of canonical variables. Such constructions are important in connection with explicit realizations of the symmetry groups of Hamiltonian systems as well as in the noninvariance-group description of dynamical systems. The possibility of identifying a dynamical system (with particular emphasis on the quantum-theoretic formulation of particle physics) with the generalized enveloping algebra of a suitable Lie algebra has been discussed elsewhere by one of the authors.[8] The problem of the recovery of the canonical variables for the system is also an essential dynamical problem.

To give an example, consider the special, familiar case of the Kepler problem. The generators of the Euclidean noninvariance group $E(4)$ are, in this case, given by[9]

$$J_{ab} = \epsilon_{abc}J_c \quad (a, b, c = 1, 2, 3), \tag{6.1}$$

$$J_{a4} = B'_a \quad (a = 1, 2, 3), \tag{6.2}$$

$$P_a = K_a \quad (a = 1, 2, 3), \tag{6.3}$$

and

$$P_4 = S', \tag{6.4}$$

where

$$J_a = \epsilon_{abc}q_bp_c, \tag{6.5}$$

$$B'_a = [(-2H)^{\frac{1}{2}}q \cos\{(-2H)^{\frac{1}{2}}(\mathbf{q}\cdot\mathbf{p})\} - (\mathbf{q}\cdot\mathbf{p}) \sin\{(-2H)^{\frac{1}{2}}(\mathbf{q}\cdot\mathbf{p})\}]p_a + [(1/q) \sin\{(-2H)^{\frac{1}{2}}(\mathbf{q}\cdot\mathbf{p})\}]q_a, \tag{6.6}$$

$$K_a = (-2H)^{-\frac{1}{2}}[(\mathbf{q}\cdot\mathbf{p})p_a - p^2q_a + (e/q)q_a], \tag{6.7}$$

$$S' = -(-2H)^{\frac{1}{2}}(\mathbf{q}\cdot\mathbf{p}) \sin[(-2H)^{\frac{1}{2}}(\mathbf{q}\cdot\mathbf{p})] + (1 + 2Hq) \cos[(-2H)^{\frac{1}{2}}(\mathbf{q}\cdot\mathbf{p})], \tag{6.8}$$

$$q = (q_aq_a)^{\frac{1}{2}}; \quad H = \tfrac{1}{2}p_ap_a - e/q. \tag{6.9}$$

Given these ten generators, one could construct the primitive dynamical variables. These points are discussed in more detail in a paper by two of us.[9] We only mention that such a construction yields

[8] E. C. G. Sudarshan, "Currents, Algebras and Dynamical Systems," invited paper at the Eastern Theoretical Physics Conference, Stony Brook, Long Island, New York, 1965.

[9] E. C. G. Sudarshan and N. Mukunda, in *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, Colorado, 1966), Vol. VIII-B, p. 407.

expressions which are undefined (singular) when the Hamiltonian H vanishes. The Hamiltonian H only plays an auxiliary role in the construction of the canonical variables. We could equally well write down some functions of these ten generators which satisfy canonical Poisson-bracket relations. In another publication,[10] one of the authors has discussed the construction of n pairs of canonical variables from the generalized enveloping algebra of the classical groups $SU(n+1)$ and $O(n+2)$.

[10] N. Mukunda, J. Math. Phys. 8, 1069 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Systems of First-Order Linear Ordinary Differential Equations as Canonical or Euler–Lagrange Equations

H. RUND
Department of Mathematics, University of the Witwatersrand, Johannesburg, South Africa

(Received 19 September 1966)

A simple but complete theory is given of the necessary and sufficient conditions which must be satisfied in order that a system of first-order linear ordinary differential equations, whose coefficients are functions of the independent variable, be the canonical or Euler–Lagrange equations of a variational principle.

FOR an n-dimensional single integral problem in the calculus of variations, the Euler–Lagrange equations consist of a system of n second-order ordinary differential equations. These are equivalent to a set of $2n$ first-order equations, the well-known canonical (or Hamiltonian) equations. An obvious problem of both mathematical and physical interest is posed by the converse state of affairs, namely, to find the conditions which must be satisfied in order that a given system of differential equations can be regarded as the canonical or Euler–Lagrange equations of a variational principle.

In the present note this problem is treated for the case when the given system consists of $2n$ first-order linear ordinary differential equations in $2n$ unknown functions whose coefficients are functions solely of the single independent variable t. In this case it is indeed possible to obtain such conditions explicitly, as will be shown below. Two distinct but related approaches suggest themselves. Firstly, one can regard the given system as being possibly a set of $2n$ canonical equations, in which case the required conditions emerge effortlessly as direct consequences of elementary integrability considerations. Secondly, of the $2n$ unknown functions, a set of n functions can be eliminated algebraically. This leads to the replacement of the given system of equations by a set of n second-order differential equations, of which one wishes to know whether or not they can be regarded as the Euler–Lagrange equations of a variational principle. This problem, together with its relation to the first, is also discussed here.

The second question had been considered in a recent paper by Pease.[1] The approach of the latter, however, is based on rather *ad hoc* methods of an algebraic nature (which is quite natural from the point of view of the particular problems treated in the article concerned), whereas the theory presented here depends on a systematic exploitation of integrability conditions. Thus, the conditions derived in the present note are somewhat more general than those found by Pease, which are, in fact, equivalent to those obtained for the case of canonical systems.

It is supposed that the given system of equations for the $2n$ dependent functions $x^A(t)$ is of the form[2]

$$dx^A/dt \equiv \dot{x}^A(t) = S_{AB}(t)x^B, \qquad (A, B = 1, \cdots, 2n), \quad (1)$$

where the entries of the $2n \times 2n$ matrix $\{S_{AB}\}$ defining

[1] M. C. Pease, J. Math. Phys. 6, 1558 (1965). In this paper the general significance of the central problem of the present note is briefly discussed.

[2] Capital Latin indices $A, B, \cdots$, run from 1 to $2n$, lower case indices $i, j, \cdots$ run from 1 to n. The summation convention is applied to both sets of indices. Because of the linearity of the given system (1), the theory presented here is not invariant under arbitrary transformations of the dependent variables. Thus the position of the indices (as sub- or superscripts) is not meant to have any invariant significance (in the sense of co- or contravariance). Both types of indices are used merely in order to facilitate immediate recognition of steps at which the summation convention is operative.

expressions which are undefined (singular) when the Hamiltonian H vanishes. The Hamiltonian H only plays an auxiliary role in the construction of the canonical variables. We could equally well write down some functions of these ten generators which satisfy canonical Poisson-bracket relations. In another publication,[10] one of the authors has discussed the construction of n pairs of canonical variables from the generalized enveloping algebra of the classical groups $SU(n+1)$ and $O(n+2)$.

[10] N. Mukunda, J. Math. Phys. 8, 1069 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Systems of First-Order Linear Ordinary Differential Equations as Canonical or Euler–Lagrange Equations

H. RUND
Department of Mathematics, University of the Witwatersrand, Johannesburg, South Africa

(Received 19 September 1966)

A simple but complete theory is given of the necessary and sufficient conditions which must be satisfied in order that a system of first-order linear ordinary differential equations, whose coefficients are functions of the independent variable, be the canonical or Euler–Lagrange equations of a variational principle.

FOR an n-dimensional single integral problem in the calculus of variations, the Euler–Lagrange equations consist of a system of n second-order ordinary differential equations. These are equivalent to a set of $2n$ first-order equations, the well-known canonical (or Hamiltonian) equations. An obvious problem of both mathematical and physical interest is posed by the converse state of affairs, namely, to find the conditions which must be satisfied in order that a given system of differential equations can be regarded as the canonical or Euler–Lagrange equations of a variational principle.

In the present note this problem is treated for the case when the given system consists of $2n$ first-order linear ordinary differential equations in $2n$ unknown functions whose coefficients are functions solely of the single independent variable t. In this case it is indeed possible to obtain such conditions explicitly, as will be shown below. Two distinct but related approaches suggest themselves. Firstly, one can regard the given system as being possibly a set of $2n$ canonical equations, in which case the required conditions emerge effortlessly as direct consequences of elementary integrability considerations. Secondly, of the $2n$ unknown functions, a set of n functions can be eliminated algebraically. This leads to the replacement of the given system of equations by a set of n second-order differential equations, of which one wishes to know whether or not they can be regarded as the Euler–Lagrange equations of a variational principle. This problem, together with its relation to the first, is also discussed here.

The second question had been considered in a recent paper by Pease.[1] The approach of the latter, however, is based on rather *ad hoc* methods of an algebraic nature (which is quite natural from the point of view of the particular problems treated in the article concerned), whereas the theory presented here depends on a systematic exploitation of integrability conditions. Thus, the conditions derived in the present note are somewhat more general than those found by Pease, which are, in fact, equivalent to those obtained for the case of canonical systems.

It is supposed that the given system of equations for the $2n$ dependent functions $x^A(t)$ is of the form[2]

$$dx^A/dt \equiv \dot{x}^A(t) = S_{AB}(t)x^B, \qquad (A, B = 1, \cdots, 2n), \quad (1)$$

where the entries of the $2n \times 2n$ matrix $\{S_{AB}\}$ defining

[1] M. C. Pease, J. Math. Phys. 6, 1558 (1965). In this paper the general significance of the central problem of the present note is briefly discussed.

[2] Capital Latin indices $A, B, \cdots$, run from 1 to $2n$, lower case indices $i, j, \cdots$ run from 1 to n. The summation convention is applied to both sets of indices. Because of the linearity of the given system (1), the theory presented here is not invariant under arbitrary transformations of the dependent variables. Thus the position of the indices (as sub- or superscripts) is not meant to have any invariant significance (in the sense of co- or contravariance). Both types of indices are used merely in order to facilitate immediate recognition of steps at which the summation convention is operative.

the system are assumed to be of class C^2 in t. The analysis of Eq. (1) according to either of the two points of view described above requires that Eq. (1) be written as a pair of systems each containing n equations. Thus, we have to partition x^A into two sets x^i, y^i, $(i, j = 1, \cdots, n)$, which implies a partition of S_{AB}:

$$S_{AB} = \begin{bmatrix} S_{ij} & T_{ij} \\ Q_{ij} & R_{ij} \end{bmatrix}. \tag{2}$$

Initially no natural prescription for this process suggests itself, but it will be evident almost immediately that the order in which x^i, y^i are selected from x^A is of vital importance. For the moment, however, it is merely assumed that (2) is defined such that T_{ij} is nonsingular. The system (1) is now written as

$$\dot{x}^i = S_{ij}x^j + T_{ij}y^j, \tag{3}$$

$$\dot{y}^i = Q_{ij}x^j + R_{ij}y^j. \tag{4}$$

Beginning with the point of view according to which this is identical with a canonical system,

$$\dot{x}^i = \partial H/\partial y^i, \quad \dot{y}^i = -\partial H/\partial x^i, \tag{5}$$

defined by some Hamiltonian function $H(t, x^j, y^j)$, it is clear that the right-hand sides of Eqs. (3) and (4) must be identically equal to $\partial H/\partial y^i$ and $-\partial H/\partial x^i$, respectively. Since the coefficients in Eqs. (3) and (4) are functions of t only, the relevant integrability conditions can be written down immediately, namely,[3]

$$T_{ij} = T_{ji}, \quad Q_{ij} = Q_{ji}, \tag{6}$$

$$S_{ij} = -R_{ji}, \tag{7}$$

(which obviously depend on the mode of partitioning the matrix S_{AB}). If these conditions are satisfied, the identities involving the derivatives of H can be integrated immediately to yield the following expression for the Hamiltonian:

$$H = \tfrac{1}{2}T_{ij}y^iy^j + S_{ij}y^ix^j - \tfrac{1}{2}Q_{ij}x^ix^j - \psi(t), \tag{8}$$

where $\psi(t)$ is an arbitrary function of t.

Denoting the inverse of the nonsingular matrix $\{T_{ij}\}$ by $\{t_{ij}\}$, we can express the solution of (3) for y^j in the form

$$y^h = t_{hi}\dot{x}^i - t_{hi}S_{ij}x^j. \tag{9}$$

[3] These conditions are essentially equivalent to the relations stipulated by Pease [Eq. (12), Ref. 1], where they are derived in an entirely different manner. In this paper the equation corresponding to Eq. (4) above is multiplied by iP, where $i^2 = -1$ and P is a constant nonsingular $n \times n$ matrix. This construction does not, however, alter the problem in any way since it is merely equivalent to a change of notation in respect of the quantities denoted above by y^i, Q_{ij}, and T_{ij}. Thus, a direct comparison of our results with those of Pease (Ref. 1) is facilitated by the replacement of iP by the unit $n \times n$ matrix I.

The Lagrangian $L(t, x^j, \dot{x}^j)$ corresponding to our Hamiltonian is defined as usual by

$$L(t, x^j, \dot{x}^j) = -H(t, x^j, y^j) + y^j\dot{x}^j, \tag{10}$$

in which y^j is expressed in terms of $(t, x^j, \dot{x}^j)$. The explicit form of L is thus obtained by substituting Eqs. (9) and (8) in Eq. (10), which yields

$$L = \tfrac{1}{2}t_{hi}\dot{x}^h\dot{x}^i - t_{hj}S_{ji}\dot{x}^hx^i + \tfrac{1}{2}(Q_{hi} + t_{jl}S_{li}S_{jh})x^ix^h + \psi(t). \tag{11}$$

Thus: The system (1) represents the canonical equations of a problem in the calculus of variations, if and only if there is a partition (2) of the matrix defining the system for which the integrability conditions (6) and (7) are satisfied. In the latter case the corresponding Lagrangian[4] is given by Eq. (11).

We now turn to the alternative approach to our problem. As before, we solve Eq. (3) for y^h. When the solution (9) is substituted in Eq. (4), the following differential equation of the second order is obtained:

$$(d/dt)(A_{hi}\dot{x}^i) + B_{hi}\dot{x}^i + C_{hi}x^i = 0, \tag{12}$$

in which the coefficients are functions of t, given by

$$A_{hi} = t_{hi}, \tag{13}$$

$$B_{hi} = -(t_{hj}S_{ji} + R_{hj}t_{ji}), \tag{14}$$

$$C_{hi} = -(d/dt)(t_{hj}S_{ji}) - Q_{hi} + R_{hj}t_{jl}S_{li}. \tag{15}$$

We now establish the following:

Lemma: In order that a system of n second-order linear differential equations of the type (12) be the Euler–Lagrange equations corresponding to a Lagrangian $L(t, x^j, \dot{x}^j)$, it is necessary and sufficient that the following conditions are satisfied:

$$A_{ih} = A_{hi}, \tag{16}$$

$$B_{ih} = -B_{hi}, \tag{17}$$

$$\dot{B}_{ih} = (d/dt)B_{ih} = C_{ih} - C_{hi}. \tag{18}$$

Proof: To prove the necessity, we write the Euler–Lagrange equations as

$$E_h(L) = 0, \tag{19}$$

where

$$E_h(L) \equiv \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^h}\right) - \frac{\partial L}{\partial x^h} = \frac{\partial^2 L}{\partial \dot{x}^h \partial \dot{x}^i}\ddot{x}^i + \frac{\partial^2 L}{\partial \dot{x}^h \partial x^i}\dot{x}^i + \frac{\partial^2 L}{\partial t \partial \dot{x}^h} - \frac{\partial L}{\partial x^h}. \tag{20}$$

[4] Except for the arbitrary term $\psi(t)$ and a constant multiplicative factor, this Lagrangian is essentially equivalent to that obtained by Pease (Ref. 1). The latter paper also allows for the possibility of complex x^A. However, because of the linearity of the problem (and the formally similar structures of the integrability conditions for real and complex variables), this possibility is ignored here. Clearly only real-valued Lagrangians can be considered.

Thus, Eqs. (12) and (19) can coincide solely if

$$A_{ih}(t) = \partial^2 L/\partial\dot{x}^i\partial\dot{x}^h, \tag{21}$$

from which Eq. (16) follows directly. Integration of Eq. (21) yields

$$\partial L/\partial\dot{x}^h = A_{ih}\dot{x}^i + f_h(t, x), \tag{22}$$

and furthermore, by virtue of Eq. (16),

$$L = \tfrac{1}{2}A_{ih}\dot{x}^i\dot{x}^h + f_h(t, x)\dot{x}^h + \varphi(t, x), \tag{23}$$

where f_h, φ are functions of (t, x^j) only. In terms of Eq. (23) the relation (20) becomes

$$E_h(L) = \frac{d}{dt}(A_{hi}\dot{x}^i) + \left(\frac{\partial f_h}{\partial x^i} - \frac{\partial f_i}{\partial x^h}\right)\dot{x}^i + \left(\frac{\partial f_h}{\partial t} - \frac{\partial\varphi}{\partial x^h}\right). \tag{24}$$

This is identical with the left-hand side of Eq. (12) if and only if

$$\partial f_h/\partial x^i - \partial f_i/\partial x^h = B_{hi}(t), \tag{25}$$

$$\partial f_h/\partial t - \partial\varphi/\partial x^h = C_{hi}(t)x^i. \tag{26}$$

Clearly Eq. (25) implies Eq. (17), while differentiation of Eq. (26) with respect to x^i gives

$$\partial^2\varphi/\partial x^i\partial x^h = \partial^2 f_h/\partial x^i\partial t - C_{hi},$$

and the requirement that the right-hand side of this relation be symmetric in i, h, taken together with Eq. (25), yields Eq. (18). This establishes the necessity of Eqs. (16)–(18). The sufficiency of these conditions follows from the fact that their validity guarantees the existence of solutions f_h, φ of the system (25) and (26). The form of the required Lagrangian is then given by Eq. (23).

When Eqs. (16)–(18) are applied to the special values (13)–(15), it follows firstly that t_{ij} and hence T_{ij} are symmetric. Similarly, by reversing the roles of x^j, y^j in the above argument, the symmetry of Q_{ij} is established. Thus the conditions (6) are obtained once more. Secondly, Eqs. (17) and (14) give rise to the skew-symmetry condition

$$t_{hj}(S_{ji} + R_{ij}) + t_{ij}(S_{jh} + R_{hj}) = 0, \tag{27}$$

while Eqs. (18) and (15) together with Eq. (27) imply that

$$(d/dt)[t_{hj}(S_{ji} + R_{ij})] = R_{hj}t_{jl}S_{li} - R_{ij}t_{jl}S_{lh}. \tag{28}$$

Thus: In order that the given system (12) of second-order differential equations be a set of Euler–Lagrange equations corresponding to a Lagrangian $L(t, x^j, \dot{x}^j)$, it is necessary and sufficient that the conditions (6), (27), and (28) should hold, in which case L is given by Eq. (23), where in the latter the functions f_h, φ are any solutions of Eqs. (25) and (26).

It is clear that this result describes a state of affairs which is considerably more general than the one which was encountered in the course of the first approach to our problem.

When Eqs. (6), (27), and (28) are satisfied, suitable solutions of Eqs. (25) and (26) are easily found. For instance, with

$$f_h = -\tfrac{1}{2}(t_{hj}S_{ji} + t_{ij}R_{hj})x^i + g_h(t), \tag{29}$$

in which g_h is an arbitrary function of t only, it follows from Eq. (27) that Eq. (25) is satisfied [the right-hand side of the latter being given by (14)], and after substitution of Eq. (29) in Eq. (26), the function φ is found by mere quadratures [the relevant integrability conditions being guaranteed by Eq. (28)].

An important special case arises when the skew-symmetry condition (7) is stipulated. This is stronger than Eq. (27), and it is evident almost immediately that Eq. (7) implies both (27) and (28). Thus, as was to be expected, the conditions (6) and (7) are sufficient to guarantee that the system (12) represents a set of Euler–Lagrange equations. Under these conditions a solution of Eq. (25) is furnished by

$$f_h = -t_{hj}S_{ji}x^i + g_h(t), \tag{30}$$

which follows directly by virtue of (7) and (14); and when this is substituted in Eq. (26), the relation (15) being taken into account, it is found that

$$\partial\varphi/\partial x^h = \dot{g}_h + (Q_{hi} + S_{jh}t_{jl}S_{li})x^i.$$

Since the coefficient of x^i on the right-hand side of this equation is symmetric in i and h, the latter can be integrated directly, which yields φ uniquely up to an arbitrary function $\psi(t)$. When this, together with Eq. (30), is substituted in Eq. (23), we obtain the following special Lagrangian:

$$L^*(t, x^j, \dot{x}^j) = L(t, x^j, \dot{x}^j) + (g_h\dot{x}^h + \dot{g}_h x^h), \tag{31}$$

where L on the right-hand side is given by Eq. (11). Since the additional term in Eq. (31) is an exact differential, we infer that the special solution (30) gives rise to a problem in the calculus of variations which is equivalent[5] to that furnished by the first approach. Also, because of the striking lack of invariance properties of the Lagrangians resulting from this analysis, one cannot expect to find many nontrivial quantities that are conserved as a result of the given system of differential equations.[6]

[5] With regard to the concept of equivalent integrals, reference is made to H. Rund, *The Hamilton–Jacobi Theory in the Calculus of Variations* (D. Van Nostrand Company Ltd., London & New York, 1966), p. 162.

[6] See Ref. 5, p. 73 *et seq.*

It is appropriate that we should briefly mention here an entirely different approach to the problem suggested by Schwartz[7] in a review of the paper of Pease.[1] This approach depends on the following Lagrangian:

$$L = \dot{q}^A\dot{q}^A + (\dot{S}_{AB} + S_{AC}S_{CB})q^Aq^B. \quad (32)$$

Assuming for the sake of simplicity that S_{AB} is symmetric, we may write the Euler–Lagrange equations as

$$\ddot{q}^A - (\dot{S}_{AB} + S_{AC}S_{CB})q^B = 0, \quad (33)$$

which is equivalent to

$$(d/dt)[\dot{q}^A - S_{AB}q^B] + S_{AC}[\dot{q}^C - S_{CB}q^B] = 0. \quad (34)$$

Obviously these equations are satisfied by solutions of the given system (1). But the converse need not hold, as is evident from the consideration of any nonvanishing solution Q^A of the system

$$\dot{Q}^A + S_{AB}Q^B = 0. \quad (35)$$

Thus, the Lagrangian (32) is not really directly concerned with the problem discussed in this note. Its relevance is due, rather, to the fact that its Euler–Lagrange equations can be obtained by differentiation of the given system (1) and the addition of linear combinations of the latter as exemplified by Eq. (34).

In conclusion, it should be remarked that the theory presented here represents a particular aspect of what is commonly referred to as the "inverse problem of the calculus of variations," which does not seem to possess a simple general solution.[8] The special nature of the present analysis is, of course, a direct consequence of the linear form of the system (1) on which the development as a whole is vitally dependent, and which precludes the possibility of a direct generalization to nonlinear systems.

ACKNOWLEDGMENT

The writer gratefully acknowledges valuable discussions with Dr. D. Lovelock of the University of Bristol.

[7] M. J. Schwartz, Math. Rev. **31**, 1152 (1966).

[8] The most far-reaching investigations of the entirely general problem are due to J. Douglas, Trans. Am. Math. Soc. **50**, 71 (1941).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

n-Representability Problem for Reduced Density Matrices

HANS KUMMER*
Department of Mathematics, Queen's University, Kingston, Ontario, Canada

(Received 22 September 1966)

In this paper we prove some theorems about the n-representability problem for reduced density operators. The first theorem (Theorem 6) sharpens a theorem proved by Garrod and Percus. Let $\mathfrak{T}_n^p$ be the set of all n-representable p-density operators. Then a density operator D^p belongs to $\overline{\mathfrak{T}_n^p}$ (the bar indicates the closure with respect to a certain topology) if and only if $\mathrm{Tr}\,(D^pB^p) \geq 0$ for all bounded self-adjoint p-particle operators B^p, such that their n-expansion

$$\binom{n}{p}\Gamma_p^n B^p \equiv \sum_{i_1 < \cdots < i_p} B^p(i_1 \cdots i_p)$$

is a positive operator in n-particle space. Moreover, it is shown that $\overline{\mathfrak{T}_n^p}$ is the closed convex hull of the exposed points of $\mathfrak{T}_n^p$ of finite one-rank (Theorem 9). A more practical version of this theorem may be formulated in the following manner (cf. Theorem 8).

Consider the set γ^p of subspaces of the n-particle space, occurring as an eigenspace to the deepest eigenvalue of a bounded n-particle operator which is the n expansion of some p-particle operator. Choose from every element of γ^p one (and only one) vector (function) and form the corresponding reduced p-particle operator. $\overline{\mathfrak{T}_n^p}$ is the closed convex hull of all these p-density operators (cf. Theorem 9). For $p = 1$, this theorem reduces to Coleman's theorem about the n representability of the 1 matrix.

1. INTRODUCTION

THE purpose of this paper is to discuss the n-representability problem for reduced density matrices. Garrod and Percus[1] have proposed a particular solution of this problem which is appealing due to its reliance on a geometrical argument. However, they formulate their theorem inside an inappropriate mathematical framework so that it is very hard to give a proof for it. Indeed, the proof which can be found in the paper of Garrod and Percus is not quite complete. In this paper we want to reformulate their theorem in such a manner that it becomes possible to establish a complete proof. Second, we give a characterization of the (convex) set of all reduced density matrices by a certain class of its extreme points. Finally, we want to show how the problem in infinite dimensions can be reduced to the analogous problem in finite dimensions.

What is the n-representability problem? In quantum mechanics one associates with every *pure state* a one-dimensional subspace of the Hilbert space $L_2(C, \mu)$ of all square summable functions on the configuration space C of the system. μ is a well-defined measure on C. For an elementary particle of spin S moving all over the physical space R^3, C is the Cartesian product of R^3 with the set $\omega = \{-S, -S+1, \cdots, +S\}$ of integer or half-integer numbers:

$$C_1 = R^3 \times \omega = \bigcup_{j=-S}^{S} R^3(j) \qquad (1.1)$$

wherein $R^3(j) = \{x \in C_1;\ pr_2 x = j\}$ denotes the subset of C_1 for which the spin coordinate has the fixed value j.

The σ-algebra of subsets on which the measure μ is defined consists of sets of the form

$$A = \bigcup_{j=-S}^{+S} A_j,$$

where $A_j \subset R^3(j)$ is Lebesgue measurable and the measure μ is defined by

$$\mu_1(A) = \sum_{j=-S}^{+S} \mu_L(A_j),$$

where μ_L denotes the Lebesgue measure on R^3.

If the system consists of n identical particles of spin S moving all over the physical space R^3, C is equal to the direct sum of n copies of C_1:

$$C_n = C_1(1) + \cdots + C_1(n),$$

and the measure $\mu = \mu_n$ is equal to the n-fold product measure of μ_1.

An element of the Hilbert space $L_2(C_n, \mu_n)$ is an equivalence class with respect to the following equivalence relation:

$$\psi \sim \varphi \Leftrightarrow \psi(\mathbf{x}) = \varphi(\mathbf{x}) \mu_n\text{-(almost everywhere)}$$

of μ_n-square summable complex-valued functions on C_n.

In the case of fermions (bosons) S is a half-integer (integer) and only the antisymmetric (symmetric) subspace $L_2^A(C_n, \mu_n)$ of $L_2(C_n, \mu_n)$ is physically relevant, i.e., all the realizable pure states of the system correspond to one-dimensional subspaces belonging to $L_2^A(C_n, \mu_n)$.

* Post-Doctoral Fellow.
[1] C. Garrod and J. K. Percus, J. Math. Phys. 5, 1756 (1964).

For example, in the case of fermions $L_2^{\wedge}(C_n, \mu_n)$ consists of those elements of $L_2(C_n, \mu_n)$ which have the additional property that they are antisymmetric in the n configuration points:

$$\psi(x_{i_1} \cdots x_{i_n}) = \pm\psi(x_1 \cdots x_n). \tag{1.2}$$

Instead of taking the one-dimensional subspaces as representatives of the states of the systems, we may take as well the corresponding projectors. If we think of such a projector as an integral operator, its kernel is given by

$$P_\psi^n(\mathbf{x}, \mathbf{x}') = \psi(\mathbf{x})\bar{\psi}(\mathbf{x}'), \tag{1.3}$$

where ψ is an element of $L_2(C_n, \mu_n)$ of norm one generating the corresponding one-dimensional subspace.

In the case where p-particle forces $(1 < p \leq n)$ are effective between the particles, the symmetric operators corresponding to the observables of the system are all of the form

$$B = \sum_{i_1 < \cdots < i_p} B^p(i_1 \cdots i_p). \tag{1.4}$$

It is well known that if one is concerned only with computing an expectation value for such a p-particle observable, one does not need all the information contained in the kernel of the projector P_ψ^n, but only the information contained in the p contraction of it, which is defined by

$$D^p(s; s') = \int \psi(s, t)\bar{\psi}(s', t)\, d\mu_{n-p}(t)$$
$$\equiv L_n^p(P_\psi^n)(s; s'), \tag{1.5}$$

where s and t stand for $s \equiv (x_1 \cdots x_p)$; $t \equiv (x_{p+1} \cdots x_n)$ (cf. Refs. 2–4). With Coleman[2] we call the integral operator D^p associated with the kernel (1.5) the *reduced p-density operator* of the system.

The n-representability problem of the first kind consists in the characterization of the set of all positive linear operators in p-particle space $(1 \leq p \leq n - 1)$ which can be derived from n-particle functions, i.e., elements of $L_2^{\wedge}(C_n, \mu_n)$ by the construction indicated by formula (1.5).

However, this paper is concerned with the n-representability problem of the second kind, where the operator P^n is replaced by a general density operator in the space $L_2^{\wedge}(C_n, \mu_n)$, i.e., a linear self-adjoint operator D^n with the additional property $\mathrm{Tr}\,(D^n) = 1$. This is the kind of operator which was introduced into quantum mechanics by von Neumann[5] to describe a mixed state of a system. The set of all these general n-density operators we shall denote by $\mathfrak{T}^n$.

Thus, in turning from the n-representability problem of the first kind to that of the second kind, we drop the property of being idempotent of the n-density operator. Equation (1.5) is replaced by $(1 \leq p \leq n - 1)$:

$$D^p(s; s') = \int D^n(st; s't)\, d\mu_{n-p}(t)$$
$$= L_n^p(D^n)(s; s'). \tag{1.6}$$

The n-representability problem of the second kind, explicitly stated, asks for a characterization of the set of all p-density operators which can be derived from an n-density operator in the way indicated by (1.6). Such a p-density operator we shall call *n representable*. The set of all n-representable p-density operators, i.e., the set $L_n^p(\mathfrak{T}^n)$ will be denoted by $\mathfrak{T}_n^p$.

The physical meaning of the reduced p-density operator lies in the fact that with its help we are able to express the expectation value of an observable of type (1.4) in the state D^n of the system by

$$\langle B \rangle_{D^n} = \binom{n}{p} \mathrm{Tr}\,(B^p D^p). \tag{1.7}$$

Moreover, if we are able to give a characterization of the set $\mathfrak{T}_n^p \equiv L_n^p(\mathfrak{T}^n)$, we can determine the lower and (in an analogous way) the upper edge of the spectrum of B by

$$b_{\min} = \binom{n}{p} \inf_{D^p \in \mathfrak{T}_n^p} \mathrm{Tr}\,(B^p D^p). \tag{1.8}$$

Since the set $\mathfrak{T}^n$ of all n-density operators is convex and the contraction L_n^p defined by Eq. (6) is a linear map, the set $\mathfrak{T}_n^p$ of all n-representable p-density operators is also convex. Since, on the other hand, a compact convex set is completely determined by its extreme points (Krein–Milman theorem), one might guess that the specification of the extreme points would be sufficient to characterize $\mathfrak{T}_n^p$. We see that the present situation is somewhat more complicated, because there is no reasonable topology under which $\mathfrak{T}_n^p$ is compact.

However, we are able to show that there is a topology under which L_n^p and the functionals of type

$$f_{B^p}(D^p) = \mathrm{Tr}\,(B^p D^p) \quad (B^p \text{ bounded!}),$$

defined on a suitable space in which we shall imbed $\mathfrak{T}_n^p$, are both continuous, so that from a physical point of view we are allowed to replace $\mathfrak{T}_n^p$ by its closure

[2] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[3] A. J. Coleman, J. Math. Phys. **6**, 1425 (1965).
[4] T. Ando, Rev. Mod. Phys. **35**, 690 (1963).
[5] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Dover Publications, Inc., New York, 1943).

$\overline{\mathfrak{T}_n^p}$. For instance, the value of $b_{\min}$ given by Eq. (8) will not be affected by this substitution. Moreover, it follows that $\overline{\mathfrak{T}_n^p}$ coincides with the convex closure of certain extreme points of $\mathfrak{T}_n^p$, so that it makes sense to look for these points. (By the convex closure of a subset S of a linear topological space, we mean the smallest closed convex set containing S.)

If the lower edge of the spectrum of B is a discrete nondegenerate eigenvalue, the p-contraction of the projector onto the corresponding eigenvector is certainly an extreme point of $\mathfrak{T}_n^p$. Thus, for example, to every n-particle system in which only p-particle forces occur and whose ground state is nondegenerate, there corresponds such an extreme point. One of the main purposes of this paper is to show that in this way one can get "sufficiently many" extreme points of $\overline{\mathfrak{T}_n^p}$ (cf. Theorems 8 and 9).

Another result of our paper is that the projection cone of $\overline{\mathfrak{T}_n^p}$ from center 0 (cf. Definition A9) in the space of a certain class of integral operators can be obtained as the intersection of an infinity of half-spaces. It is this result which is a precise version of the theorem of Garrod and Percus.[1] The importance of this result is that we are able to approximate a necessary and sufficient condition for the n representability by an infinity of necessary conditions, which will be of great value for constructing a numerical procedure for evaluating $b_{\min}$ using formula (8).

As mentioned at the beginning of the Introduction, the final result of our paper consists in the fact that we were able to reduce the n-representability problem in infinite dimensions in a precise manner to the corresponding problem in finite dimensions, where the algebraic features of the problem appear in their full complexity.

Finally, we should not forget to mention that in the case $p = 1$, the problem has been completely solved by Coleman.[2] His result may be summarized in the following theorem.

Theorem 1: (a) *Bosons:* Every 1-density operator is n representable.

(b) *Fermions:* $\overline{\mathfrak{T}_n^1}$ is the closed convex hull of the set of all density operators of the form $(1/n)P_n$, where P_n projects onto an n-dimensional subspace of $L_2^a(C_1, \mu_1)$; or equivalently, a 1-density operator D^1 belongs to $\overline{\mathfrak{T}_n^1}$ if and only if its norm $\|D^1\|$ satisfies the inequality $\|D^1\| \leq 1/n$.

Proof: The proof of assertion (a) is obvious. Let D^1 be any 1-density operator and $D^1(x_1; x_1')$ its kernel. Then $D^1 = L_n^1(D^n)$, where D^n is the n-density operator with the kernel

$$D^n(x_1 \cdots x_n; x_1' \cdots x_n') = D^1(x_1; x_1') \cdot D^1(x_2; x_2') \cdots D^1(x_n; x_n').$$

The first part of assertion (b) is contained in a more general theorem (Theorem 9) of this paper.

2. A PRECISE MATHEMATICAL DESCRIPTION OF THE SET OF ALL MIXED STATES OF A QUANTUM-MECHANICAL SYSTEM

The mathematical tools we need throughout the paper may be found in Schatten[6] and Robertson and Robertson.[7]

Let H be the (abstract) separable Hilbert space associated with a quantum-mechanical system. As we have already stated in the Introduction, to every pure state of the system there corresponds a projector P onto a one-dimensional subspace of H. We denote the set of all these projectors by $E(H)$. A mixed state in turn is described by an element belonging to "the closure" of the convex hull of $E(H)$. At this point the question immediately arises: in what topology do we take this closure to guarantee that all the operators in the resulting set are of finite trace?

The only way to satisfy this condition is to take the closure inside the so-called trace class $T(H)$ of operators. In order to define this class, let us consider the set $B(H)$ of all bounded linear operators in H. Equipped with the norm

$$\|A\| = \sup_{\|x\|=1} \|A\|, \quad A \in B(H), \tag{2.1}$$

$B(H)$ is a Banach algebra. In $B(H)$ the involution

$$A \to A^*, \tag{2.2}$$

which associates to every A its adjoint, is defined and is continuous. Let $S(H)$ be the set of all self-adjoint bounded linear operators in H, i.e., the set of all fixed points inside $B(H)$ with respect to the involution (2). $S(H)$ is a real closed subspace of $B(H)$. We denote by $S^+(H)$ the set of all positive elements of $S(H)$, i.e., the set

$$S^+(H) = \{A \in S;\ (x, Ax) \geq 0\ \forall\ x \in H\}.$$

$S^+(H)$ is a closed *convex cone* (cf. Definition A7) in $S(H)$, i.e., a closed subset with the additional properties

$$A, B \in S^+ \Rightarrow A + B \in S^+, \tag{2.3a}$$

$$A \in S^+, \quad \alpha \geq 0 \Rightarrow \alpha A \in S^+. \tag{2.3b}$$

[6] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960).

[7] A. P. Robertson and W. J. Robertson, "Topological Vector Spaces," Cambridge Tracts in Mathematics and Mathematical Physics, No. 53 (Cambridge University Press, Cambridge, England, 1964).

We may consider $S^+(H)$ as the positivity domain of a partial order with respect to which $S(H)$ forms a partially ordered vector space.

For every element $A \in S^+(H)$ there is one and only one element $B \in S^+(H)$ such that $B^2 = A$. B is called the square root of A and is denoted by $B = A^{\frac{1}{2}}$. The map $A \to A^{\frac{1}{2}}$ is a continuous map of $S^+(H)$ onto itself. We define an additive positive homogeneous functional on $S^+(H)$ with values on the extended real axis, called the trace.

Definition: The trace of $A \in S^+(H)$ is a nonnegative number defined by

$$\mathrm{Tr}\,(A) = \sum_{i=1}^{\infty} (u_i \mid Au_i), \tag{2.4}$$

where $\{u_i\}$ is any complete orthonormal system (CONS) in H.

The following sequence of equations shows that the trace is independent of the choice of the CONS $\{u_i\}$:

$$\sum_{i=1}^{\infty} (v_i \mid Av_i) = \sum_{i=1}^{\infty} (A^{\frac{1}{2}}v_i \mid A^{\frac{1}{2}}v_i) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |(A^{\frac{1}{2}}v_i \mid u_j)|^2$$
$$= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} |(A^{\frac{1}{2}}u_j \mid v_i)|^2 = \sum_{j=1}^{\infty} (u_j \mid Au_j).$$

Here we have used the fact that the limit of a positive double series does not depend on the order of summation (cf. Ref. 8). Using the trace functional Tr, we are able to define two convex functionals on $B(H)$ with values on the extended real axis by

$$|A| = \mathrm{Tr}\,((A^*A)^{\frac{1}{2}}), \tag{2.5}$$

$$|A| = \{\mathrm{Tr}\,(A^*A)\}^{\frac{1}{2}}. \tag{2.6}$$

It is easy to see that, for all $A \in B(H)$,

$$\|A\| \leq |A| \leq |A|. \tag{2.7}$$

The convexity of the functionals (5) and (6) implies that the classes of operators defined by

$$T(H) = \{A \in B(H);\ |A| < \infty\} \quad \text{(trace class)}, \tag{2.8}$$

$$\mathfrak{K}(H) = \{A \in B(H);\ |A| < \infty\}$$
(Hilbert–Schmidt class) (2.9)

are subspaces of $B(H)$. It is an immediate consequence of (2.7) that $T(H) \subset \mathfrak{K}(H)$. It is even possible to show that $T(H)$ and $\mathfrak{K}(H)$ are two-sided ideals in $B(H)$, which are stable under the (*)-involution (2), and that every element of $T(H)$ can be represented as a product of two elements belonging to $\mathfrak{K}(H)$ (cf. Ref. 6). $|\ |$ defines a norm in $T(H)$, as well as $\|\ \|$ does in $\mathfrak{K}(H)$. Under these norms the sets form normed *-algebras. The question arises: Is each of the spaces $T(H)$ and $\mathfrak{K}(H)$ under its respective norm complete? The answer is in both cases affirmative (cf. Ref. 6).

Let A be an arbitrary element of $B(H)$. Then we denote by R_A the *closure* of its range and by N_A its null space. The dimension of R_A is called the rank of A. By $F(H)$ we denote the set of all elements of $B(H)$ of finite rank. It is easy to verify that $F(H)$ is a two-sided ideal, stable under the (*)-involution inside $B(H)$. Its closure with respect to the uniform topology defined by the norm (2.1) coincides with the set $C(H)$ of all completely continuous linear operators in H. Since $B(H)$ is complete, $C(H)$ is isomorphic to the completion of $F(H)$ equipped with the uniform norm

$$C(H) \cong \mathrm{Compl}\,(F(H), \|\ \|). \tag{2.10}$$

The functionals defined in Eqs. (2.5) and (2.6) restricted to $F(H)$ define on $F(H)$ a norm. It can be shown (cf. Ref. 6) that

$$T(H) \cong \mathrm{Compl}\,(F(H), |\ |), \tag{2.11}$$

$$\mathfrak{K}(H) \cong \mathrm{Compl}\,(F(H), \|\ \|). \tag{2.12}$$

It is not difficult to show that, for every $A \in T(H)$ and any CONS in H, the sequence $|(u_i \mid Au_i)|$ is summable and the sum of the $(u_i \mid Au_i)$'s is independent of the choice of the CONS. We call it the *trace* of A and denote it by

$$\mathrm{Tr}\,(A) = \sum (u_i \mid Au_i), \quad A \in T(H). \tag{2.13}$$

The following lemma summarizes some properties of the trace.[6]

Lemma 1:

(i) $\mathrm{Tr}\,(A^*) = \overline{\mathrm{Tr}\,(A)}$; $A \in T(H)$.
(ii) $\mathrm{Tr}\,(cA) = c\,\mathrm{Tr}\,(A)$; $c = c.$ number; $A \in T(H)$.
(iii) $\mathrm{Tr}\,(A + B) = \mathrm{Tr}\,(A) + \mathrm{Tr}\,(B)$; $A, B \in T(H)$.
(iv) $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$; $A \in T(H)$, $B \in B(H)$.
(v) $\mathrm{Tr}\,(AB) \leq |A|\,\|B\|$; $A \in T(H)$, $B \in B(H)$.

Inside $T(H)$ the self-adjoint elements constitute a closed real subspace $S_T(H)$. The key for our argumentation in the sequel is:

Lemma 2: $S(H)$ and $S_T(H)$ form a dual pair [compare: Definition A1 of Appendix 1 (from Robertson and Robertson, Ref. 7, p. 31)] of real linear spaces with respect to the bilinear form $\mathrm{Tr}\,(AX)$.

Proof: That $\mathrm{Tr}\,(XA)$ is a real bilinear form is a consequence of Lemma 1. For verifying the statements

[8] H. Meschkowski, "Unendliche Reihen," Hochschultaschenbuch Vol. 35, Bibliographisches Institut, Mannheim, 1961.

D2′ and D2″ of Definition A1, one observes that the projectors onto one-dimensional subspaces belong to $S_T(H) \subset S(H)$ and that a self-adjoint operator A, for which $(x, Ax) = 0$ for all $x \in H$, is the zero operator. Q.E.D.

According to Schatten[6] (p. 47, Theorem 2) the trace norm on $S_T(H)$ defines a topology which is compatible with the dual pair $\langle S(H), S_T(H)\rangle$, i.e., under which $S(H)$ is the dual space of $S_T(H)$ (cf. Appendix, Definition A2). Moreover, the uniform norm in $S(H)$ coincides with the norm induced by the trace norm in the dual space $S(H)$ of $S_T(H)$:

$$\sup_{|X|=1} \mathrm{Tr}\,(AX) = \|A\|, \quad A \in S(H); \quad X \in S_T(H). \tag{2.14}$$

The weakest topology on $S_T(H)$ with this property is the *weak topology* generated by the following base at 0:

$$U(0; X_1, \cdots, X_n; \epsilon) = \{Y \in S_T(H); |\mathrm{Tr}\,(X_i Y)| < \epsilon, \quad X_i \in S(H), \quad i = 1, \cdots, n\}. \tag{2.15}$$

What topologies on $S(H)$ are compatible with the dual pair? It can be shown that the strongest (or ultrastrong) topology on $S(H)$ has the property we are asking for. We do not make use of this fact. The weakest topology on $S(H)$ compatible with the dual system is generated by the following base at 0:

$$U(0; Y_1, \cdots, Y_n; \epsilon) = \{X \in S(H); |\mathrm{Tr}\,(XY_i)| < \epsilon, \quad Y_i \in S_T(H), \quad i = 1, \cdots, n\}. \tag{2.16}$$

We refer to it as the *weak* topology on $S(H)$. (Some authors would call it the weak*-topology.) It is perhaps appropriate to remark that the weak topology is also weaker than the uniform topology. This is a consequence of the fact that the strongest topology is weaker than the uniform topology.

If we speak in the sequel of the convex closure of an arbitrary set or the closure of a convex set contained in one of the members of our dual system $\langle S(H), S_T(H)\rangle$, then we understand this operation always with respect to any topology compatible with the dual pair (cf. Theorem A1). The closure will be simply denoted by a bar, while for the convex closure the symbol $\overline{\mathrm{conv}}$ will be used.

Let us now fix our attention to the sets of positive operators belonging to $S(H)$ and $S_T(H)$. We denote them by

$$\mathcal{K}(H) \equiv S^+(H), \quad \mathcal{K}_T(H) \equiv S_T^+(H).$$

Theorem 2 describes some of the features of their geometrical structure.

Theorem 2: (i) The sets $\mathcal{K}(H)$ and $\mathcal{K}_T(H)$ of positive elements in $S(H)$ and $S_T(H)$ are the polar cones of each other (cf. Definition A11) and therefore weakly closed.

(ii) A subset of $\mathcal{K}(H)$ is exposed (cf. Definition A13) if and only if it is of the form $\mathcal{K}(V) = \{X \in \mathcal{K}(H); R_X \subseteq V\}$, where $V \subset H$ is a closed subspace. The exposed rays are therefore of the form $\{\rho P; \rho \geq 0\}$, where P denotes a one-dimensional projector. The statement remains valid after replacing $\mathcal{K}(H)$ by $\mathcal{K}_T(H)$.

(iii) $\mathcal{K}(H)$ is the convex closure of its exposed rays and every extreme ray is exposed. This statement is also true for $\mathcal{K}_T(H)$.

(iv) Neither $\mathcal{K}(H)$ nor $\mathcal{K}_T(H)$ is weakly locally compact.

(v) $\mathcal{K}_T(H)$ does not contain a core point and is therefore nowhere dense in $S_T(H)$ (cf. Definition A8).

Proof: (i) Let $E(H) \subset \mathcal{K}_T(H)$ denote the set of all one-dimensional projectors in H. If we denote the operation of forming the polar cone by $\sim$, we may describe $\mathcal{K}(H)$ by $\mathcal{K}(H) = \tilde{E}(H)$. But since $E(H) \subset \mathcal{K}_T(H)$, it follows that $\tilde{\mathcal{K}}_T(H) \subset \mathcal{K}(H)$. On the other hand, let $A \in \mathcal{K}(H)$, let B be any element of $\mathcal{K}_T(H)$, and let

$$B = \sum_{i=1}^{\infty} \lambda_i(B) P_i$$

be its spectral decomposition. From

$$\lim_{m\to\infty} \left| B - \sum_{i=1}^{m} \lambda_i(B) P_i \right| = \lim_{m\to\infty} \sum_{i=m+1}^{\infty} \lambda_i(B) = 0$$

and the continuity of the functional $\mathrm{Tr}\,(AX)$, it follows that

$$\mathrm{Tr}\,(AB) = \lim_{m\to\infty} \sum_{i=1}^{m} \lambda_i(B)\,\mathrm{Tr}\,(AP_i) \geq 0.$$

Hence $A \in \tilde{\mathcal{K}}_T(H)$ and $\mathcal{K}(H) = \tilde{\mathcal{K}}_T(H)$.

Now we consider $E(H)$ as a subset of $\mathcal{K}(H)$. Then $\mathcal{K}_T(H) = \tilde{E}(H)$. Therefore $\mathcal{K}_T(H)$ is closed and we have, according to the bipolar theorem (Theorem A6),

$$\mathcal{K}_T(H) = \tilde{\tilde{\mathcal{K}}}_T(H) = \mathcal{K}(H).$$

(ii) Let $A \in \tilde{\mathcal{K}}(H) = \mathcal{K}_T(H)$ and let N_A be the null space of A. Then we assert that $\{A\}^\perp \cap \mathcal{K}(H) = \mathcal{K}(N_A)$. The assertion is an immediate consequence of the following lemma:

Lemma 3: Two positive operators A, B, one of them being of trace class, are orthogonal to each other, i.e., $\mathrm{Tr}\,(AB) = 0$ if and only if $R_B \subset N_A$.

Proof: We prove the case where $A \in \mathcal{K}_T(H)$. For the proof in the case $A \in \mathcal{K}(H)$ we have merely to interchange the role of B and A in the subsequent arguments. Let $\{e_i\}$ be a CONS of eigenvectors of A. Then let

$$\mathrm{Tr}\,(AB) = \sum \lambda_i(A)(e_i \mid Be_i) = 0.$$

Since $\lambda_i(A)$ and $(e_i \mid Be_i)$ are nonnegative numbers, it follows that $(e_i \mid Be_i) = 0$ for all i for which $\lambda_i(A) \neq 0$. Thus $Be_i = 0$ for all $e_i \in R_A$ and therefore $N_B \supset R_A$, or equivalently, $R_B \subset N_A$. On the other hand, if $R_B \subset N_A$, then obviously $\mathrm{Tr}\,(AB) = 0$. Q.E.D.

Now let $V \subset H$ be a closed subspace. Then it is not difficult to see that there is an operator $A \in \mathcal{K}_T(H)$ such that $N_A = V$. The representation

$$\mathcal{K}(V) = \{A\}^{\perp} \cap \mathcal{K}(H)$$

shows that $\mathcal{K}(V)$ is an exposed subset of $\mathcal{K}(H)$. The corresponding proof for $\mathcal{K}_T(H)$ is completely analogous.

(iii) As exposed sets, the rays $\mathcal{K}(R_P) = \{\rho P;\ \rho \geq 0,$ $P =$ one-dimensional projector$\}$ are certainly extreme. We show that they are the only extreme rays. For this purpose let A be an element of $\mathcal{K}(H)$ with $\|A\| = 1$, whose range is more than one-dimensional. Let $\{P_\lambda\}$ be the spectral resolution of the identity defined by A, so that $P_\lambda = 0$ for $\lambda < 0$ and $P_1 = I$ (Identity). We distinguish two cases.

(a) The open interval (0, 1) contains a point λ_0 of the spectrum of A. Then the projectors

$$E_2 = P_{\lambda_0} - P_0,$$
$$E_1 = I - P_0 - E_2 = I - P_{\lambda_0},$$

are both different from 0; they are orthogonal and their sum is the projector onto R_A. Thus, if we introduce $A_1 = E_1AE_1$ and $A_2 = E_2AE_2$, we have

$$A_1, A_2 \in \mathcal{K}(H), \quad A_1, A_2 \neq 0$$

and $A = A_1 + A_2$, an equation which shows that the ray generated by A is not extreme in $\mathcal{K}(H)$. An identical proof holds for the $\mathcal{K}_T(H)$ case.

(b) The interval (0, 1) contains no point of the spectrum of A. Then A is a projector onto a closed subspace $V \subset H$ of a dimension higher than one. Let P be a projector onto a one-dimensional subspace of V. Then both P and $A - P$ belong to $\mathcal{K}(H)(\mathcal{K}_T(H))$, are different from zero, and their sum is equal to A: $A = P + (A - P)$. This again shows that the ray generated by A is not extreme. The assertion that the cones $\mathcal{K}(H)$ and $\mathcal{K}_T(H)$ are the convex closures of their extreme rays is intimately connected with assertion (i). Denote again by $E(H)$ the set of all one-dimensional projectors in H. The closed convex cone generated by $E(H)$ coincides, according to the bipolar theorem (cf. Theorem A6), with $\mathcal{K}(H)$ inside $S(H)$ and with $\mathcal{K}_T(H)$ inside $S_T(H)$:

$$E(H) \subset S(H): \quad \tilde{\tilde{E}}(H) = \tilde{\mathcal{K}}_T(H) = \mathcal{K}(H),$$

$$E(H) \subset S_T(H): \quad \tilde{\tilde{E}}(H) = \tilde{\mathcal{K}}(H) = \mathcal{K}_T(H).$$

This completes the proof of assertion (iii).

(iv) That $\mathcal{K}(H)$ is not weakly locally compact can be shown in the following way. Consider $A \in \mathcal{K}_T(H)$ such that $R_A = H$, and let

$$A = \sum_{i=1}^{\infty} a_i P_i \quad 0 < \cdots a_{i+1} \leq a_i \cdots \leq a_1$$

be its spectral decomposition. Let $\{b_i\}$ be a sequence of positive numbers converging monotonically towards zero, such that the series $\sum_{i=1}^{\infty} b_i$ is converging, but at a slower rate than the series $\sum_{i=1}^{n} a_i$, i.e.,

$$\lim_{k\to\infty} (a_k/b_k) = 0.$$

Put

$$B \equiv \sum_{i=1}^{\infty} b_i P_i \in \mathcal{K}_T(H)$$

and

$$\theta = 2 \sup_k (a_k/b_k).$$

We consider the following sequence of operators:

$$C_k \equiv (\epsilon/b_k\theta)P_k \in \mathcal{K}(H) \cap U(0; A; \epsilon), \quad \epsilon > 0,$$

where the weak neighborhood $U(0; A; \epsilon)$ of 0 is defined by Eq. (16). Assume that $\{C_k\}$ contains a certain subsequence (again denoted by $\{C_k\}$) converging weakly to an element $C \in \mathcal{K}(H)$. Then

$$\mathrm{Tr}\,(CA) = \lim_{k\to\infty} \mathrm{Tr}\,(C_kA) = \frac{\epsilon}{\theta} \lim_{k\to\infty} \frac{a_k}{b_k} = 0.$$

Since $R_A = H$, Lemma 3 implies $C = 0$. Therefore we get

$$0 = \mathrm{Tr}\,(CB) = \lim_{k\to\infty} (C_kB) = \epsilon/\theta,$$

i.e., a contradiction: $\mathcal{K}(H) \cap \overline{U(0; A; \epsilon)}$ is not compact. To complete the proof we have to show that the neighborhoods

$$\mathcal{K}(H) \cap U(0; A; \epsilon), \quad A \in \mathcal{K}_T(H), \quad R_A = H$$

form a basis of neighborhoods of the point 0 for the relative topology of $\mathcal{K}(H)$. Let first $A \in \mathcal{K}_T(H)$ and $R_A \neq H$. Choose $B \in \mathcal{K}_T(H)$ such that $R_B = N_A$. Then

$$U(0; A + B; \epsilon) \cap \mathcal{K}(H) \subset U(0; A; \epsilon) \cap \mathcal{K}(H),$$

with $A + B \in \mathcal{K}_T(H)$ and $R_{A+B} = H$. Finally, let $U(0; A_1, \cdots, A_n; \epsilon)$, $A_k \in S_T(H)$ be an arbitrary neighborhood of 0 of type (16) and let

$$A_k = A_k^+ - A_k^-$$

be a representation of A_k as a difference of two elements belonging to $\mathcal{K}_T(H)$. Then

$$U\left(0; \sum_{k=1}^{n}(A_k^+ + A_k^-); \epsilon\right) \cap \mathcal{K}(H) \subset U(0; A_1, \cdots, A_n; \epsilon) \cap \mathcal{K}(H).$$

This completes the proof of the first part of (iv).

In order to see that $\mathcal{K}_T(H)$ also is not locally compact, choose $A \in \mathcal{K}_T(H)$ with $R_A = H$ and let

$$A = \sum_{k=1}^{\infty} a_k P_k$$

be its spectral decomposition. Furthermore, let $B \in (\mathcal{K}H)$, $B \neq 0$, and define

$$P_k' = (\epsilon/2 \, \|B\|)P_k, \quad \epsilon > 0.$$

Then $P_k' \in U(0; B; \epsilon) \cap \mathcal{K}_T(H)$ where $U(0; B; \epsilon)$ is now defined by Eq. (15). Assume $\{P_k'\}$ contains a convergent subsequence again denoted by $\{P_k'\}$ and let P be its limit point. Then

$$\operatorname{Tr}(AP) = \lim_{k \to \infty} \operatorname{Tr}(AP_k') = 0.$$

Now Lemma 3 implies: $P = 0$. But from the weak continuity of the trace we deduce

$$0 = \operatorname{Tr}(P) = \lim_{k \to \infty} \operatorname{Tr}(P_k') = \epsilon/2 \, \|B\|,$$

i.e., a contradiction. The remaining arguments are completely analogous to the ones occurring in the first part of the proof.

(v) Assume that A is a core point of $\mathcal{K}_T(H)$. Then A certainly is not a supporting point and therefore $R_A = H$. Let $A = \sum \lambda_i P_i$ be the spectral representation of A and $\{\mu_i\}$ a sequence of nonnegative numbers converging to zero, such that the series $\sum_{i=1}^{\infty} \mu_i$ is convergent, but at a slower rate than the series

$$\sum_{k=1}^{\infty} \lambda_k.$$

This means that

$$\lim_{k \to \infty} (\lambda_k/\mu_k) = 0.$$

We define now an operator $B \in S_T(H)$ by $B = -\sum \mu_i P_i$. Then $A + \epsilon B \notin \mathcal{K}_T(H)$ for all $\epsilon > 0$. For assume $\epsilon > 0$ and $A + \epsilon B \in \mathcal{K}_T(H)$. It follows that $\lambda_k/\mu_k \geq \epsilon$ for all k and therefore

$$\lim_{k \to \infty} (\lambda_k/\mu_k) \geq \epsilon > 0,$$

in contradiction to the construction of the sequence $\{\mu_k\}$. Q.E.D.

Remarks: The map $V \to \mathcal{K}_T(V)$ is an order isomorphism from the lattice $\mathfrak{L}(H)$ of all closed subspaces of H onto the set of all exposed sets $\mathfrak{L}(\mathcal{K}_T)$ of $\mathcal{K}_T(H)$. The space [0] is mapped onto the exposed set $\{0\}$, consisting of the vertex of $\mathcal{K}_T(H)$. The one-dimensional subspaces of H are mapped onto the set of all extreme rays of $\mathcal{K}_T(H)$. They are the minimal elements of the partially ordered set $\mathfrak{L}(\mathcal{K}_T) - \{0\}$.

Theorem 2 enables us to give a geometrical description of the set of all mixed states of a quantum-mechanical system. We define this set by

$$\mathfrak{S}(H) = \{X \in \mathcal{K}_T(H); \operatorname{Tr}(X) = 1\},$$

i.e., as the intersection of $\mathcal{K}_T(H)$ with the closed linear manifold $\operatorname{Tr}(X) = 1$.

Theorem 2': The set $\mathfrak{S}(H)$ of all mixed states of a quantum-mechanical system with the (separable) Hilbert space H as its space of states is a closed convex set in the space $S_T(H)$ of all self-adjoint trace operators in H with the following properties:

(i) $\mathfrak{S}(H)$ is a subset of the closed linear manifold

$$\{X \in S_T(H); \operatorname{Tr}(X) = 1\}.$$

In fact,

$$\mathfrak{S}(H) = \{X \in \mathcal{K}_T(H); \operatorname{Tr}(X) = 1\}.$$

(ii) The exposed points (cf. Definition A13) correspond to the pure states of the system. $\mathfrak{S}(H)$ coincides with the convex closure of its exposed points. Every extreme point is exposed.

(iii) The supporting points (cf. Definition A13) of $\mathfrak{S}(H)$ are either states which are a mixture of an incomplete set of pure states or are pure states. The exposed sets of $\mathfrak{S}(H)$ coincide with the sets $\mathfrak{S}(V)$ of all mixtures of pure states belonging to a fixed closed subspace V (which may be one-dimensional) of H. Formally,

$$\mathfrak{S}(V) = \{X \in \mathfrak{S}(H); R_A \subseteq V\}.$$

(iv) The projection cone (cf. Definition A9) $\mathcal{K}_T(H)$ of $\mathfrak{S}(H)$ from the center zero is the polar cone of the cone of all bounded positive observables.

(v) The set $\mathfrak{S}(H)$ is nowhere dense in $S_T(H)$. Furthermore, it is not locally compact. (For special notions about convex sets, see Appendix.)

We conclude this section by proving a lemma about the connection between the set $E(H)$ of the one-dimensional projectors and the unitsphere in H.

Lemma 4: Let $x \in H$, $\|x\| = 1$ and P_x the projector onto the one-dimensional subspace spanned by x. Then the map $x \to P_x$ is a continuous map of the unitsphere in H onto $E(H)$.

Proof: Let $x, y \in H$, $\|x\| = \|y\| = 1$. Then $P_x - P_y$ is a self-adjoint operator the range of which is spanned by x and y. We have

$$(P_x - P_y)x = x - (y \mid x) \cdot y,$$
$$(P_x - P_y)y = (x \mid y) \cdot x - y,$$

so that the nonzero eigenvalues of $P_x - P_y$ coincide with the eigenvalues of the matrix

$$\begin{pmatrix} 1 & -(y \mid x) \\ (x \mid y) & -1 \end{pmatrix}.$$

But these eigenvalues are easily calculated:

$$\mu_{1,2} = \pm[1 - |(x \mid y)|^2]^{\frac{1}{2}}.$$

Thus,

$$|P_x - P_y|^2 = (|\mu_1| + |\mu_2|)^2 = 4(1 - |(x \mid y)|^2).$$

Now from $-(x \mid y) = (x \mid x - y) - 1$, it follows that

$$1 - |(x \mid y)|^2 = 2 \operatorname{Re}(x \mid x - y) - |(x \mid x - y)|^2$$
$$\leq 2\,|(x \mid x - y)| \leq 2\,\|x - y\|.$$

Hence we have

$$|P_x - P_y|^2 \leq 8\,\|x - y\|,$$

an inequality which expresses the asserted continuity. Q.E.D.

3. n-REPRESENTABILITY PROBLEM AND ITS DUAL PROBLEM

If H^1 is the Hilbert space associated in nonrelativistic quantum mechanics with an elementary particle, then the Hilbert space associated with n particles is given by the completion of the n-fold tensor product, which we shall denote by H^n:

$$H^n = \overset{n}{\underset{i=1}{\otimes}} H^1(i). \tag{3.1}$$

If $\{e_i\}$ is a CONS in H^1, the same is true for the n-vectors

$$\{e_{i_1} \otimes e_{i_2} \cdots \otimes e_{i_n}\} \tag{3.2}$$

in relation to H^n. We shall refer to the set (3.2) as the product base in H^n associated with $\{e_i\}$. We denote by Π_n the symmetric group of the n particles. Then

$$s \in \Pi_n \to G_s^n t = \sum t^{i_{s(1)} \cdots i_{s(n)}} e_{i_1} \otimes \cdots \otimes e_{i_n}, \quad t \in H^n \tag{3.3}$$

defines a representation of Π_n in H^n.

If we deal with n indistinguishable particles, then it is only a certain subspace $H^{n\wedge}$ of H^n, invariant under the representation (3.3) of Π_n, which is physically relevant. We denote the projector corresponding to this subspace by A_n. A_n can be expressed as

$$A_n = \sum_{s \in \Pi_n} \alpha_s G_s^n, \tag{3.4}$$

where $\alpha_s = 1$ for bosons; $\alpha_s = (-1)^{\sigma(s)}$ for fermions [$\sigma(s)$ = signature of s]. Let i be the map of $S(H^{n\wedge})$ into $S(H)^n$ defined by

$$i(B^n) = A_n B^n A_n \quad B^n \in S(H^{n\wedge}).$$

Then i is an injection. The image $i(S(H^{n\wedge})) \subset S(H^n)$ we denote by S^n. Similarly, we introduce the shorthand notation S_T^n for $i(S_T(H^{n\wedge}))$, $\mathcal{K}^n$ for $i(\mathcal{K}(H^{n\wedge}))$, and so on.

The fundamental dual pair of real linear spaces associated with a system of n indistinguishable particles is given by $\langle S^n, S_T^n \rangle$. The set of observables of such a system can be identified with S^n whereas the set of states coincides with

$$\mathfrak{S}^n \equiv \{D \in S_T^n;\ D \geq 0,\ \operatorname{Tr}(D) = 1\}. \tag{3.5}$$

The set of pure states of the system is represented by the set of all exposed points of $\mathfrak{S}^n$ and will be denoted by

$$E^n = \{P \in \mathfrak{S}^n;\ P^2 = P\}. \tag{3.6}$$

In the sequel, W denotes always a closed subspace of H^1. Let A_n^W be the projector of H^n onto the intersection

$$W^{n\wedge} = \overset{n}{\underset{i=1}{\otimes}} W(i) \cap H^{n\wedge}.$$

Furthermore, let us denote by $S^n[W]$ the closed subspace of S^n, defined by

$$S^n[W] = \{D \in S^n;\ A_n^W D A_n^W = D\} \tag{3.7}$$

and

$$\mathfrak{S}^n[W] = \mathfrak{S}^n \cap S^n[W]. \tag{3.8}$$

Proposition 1: Let $W_1 \subset W_2 \subset \cdots \subset H^1$ be an increasing sequence of *finite*-dimensional subspaces of H^1 with the property that

$$\overline{\bigcup_i W_i} = H^1.$$

Then

$$\mathfrak{S}^n = \overline{\bigcup_i \mathfrak{S}^n[W_i]} \tag{3.9}$$

and

$$S^n = \overline{\bigcup_i S^n[W_i]}, \tag{3.10}$$

where the bar in Eqs. (9) and (10) indicates the closure in any topology compatible with the dual pair $\langle S^n, S_T^n \rangle$.

Proof. It is sufficient to show that

$$E^n \subset \overline{\bigcup_i \mathfrak{F}^n[W_i]}. \tag{3.11}$$

Assume this to be true. Then from Theorem 2 we know that

$$\mathfrak{F}^n = \overline{\text{conv}\,(E^n)} \subseteq \overline{\bigcup_i \mathfrak{F}^n[W_i]}.$$

Since the inverse inclusion is obvious, we have

$$\mathfrak{F}^n = \overline{\bigcup_i \mathfrak{F}^n[W_i]},$$

which proves (3.9). Now let $C \in S^n$. Then there exist elements A, $B \in \mathfrak{F}^n$ and nonnegative numbers α, β such that $C = \alpha A - \beta B$; $\alpha + \beta > 0$.

Because of (3.9), for every $\epsilon > 0$ there exist a natural number $N(\epsilon)$ and elements A', $B' \in \mathfrak{F}^n[W_i]$ for all $i \geq N(\epsilon)$, such that

$$|A - A'| < \epsilon/(\alpha + \beta), \quad |B - B'| < \epsilon/(\alpha + \beta).$$

But $C' \equiv \alpha A' - \beta B' \in S^n[W_i]$ for all $i \geq N(\epsilon)$ and

$$|C - C'| = |\alpha A - \beta B - (\alpha A' - \beta B')|$$
$$\leq \alpha\,|A - A'| + \beta\,|B - B'| < \epsilon.$$

Thus

$$C \in \overline{\bigcup_i S^n[W_i]},$$

and, since the closure of a convex set does not depend on the topology if only it is compatible with the dual pair, the validity of Eq. (3.10) is established.

It remains to prove the inclusion (3.11). With the assumption of the proposition, the one-particle space H^1 can be represented as

$$H^1 = \bigoplus_{i=1}^{\infty} (W_{i+1} \ominus W_i). \tag{3.12}$$

Choose in H^1 a CONS $\{e_i\}$ adapted to the decomposition (3.12) and assume that $\{e_1, \cdots, e_{d_k}\}$ is complete in W_k. Then $\{e_{i_1} \otimes \cdots \otimes e_{i_n}\}$ is a CONS in H^n. It is also not difficult to convince oneself that the set

$$u_{i_1 \cdots i_n} = A_n(e_{i_1} \otimes \cdots \otimes e_{i_n})$$

($i_1 \leq i_2 \leq \cdots \leq i_n$ for bosons, $i_1 < i_2 < \cdots < i_n$ for fermions) is orthogonal and complete in the physically relevant subspace $H^{n\wedge}$.

Now let $P_t \in E^n$ be a one-dimensional projector and t a vector belonging to corresponding subspace such that $\|t\| = 1$. Define

$$k = 1, 2, \cdots; \quad t_k = \sum_{i_1 \cdots i_n \leq d_k} (u_{i_1 \cdots i_n} \mid t) u_{i_1 \cdots i_n}.$$

Then $t = \lim_{k \to \infty} t_k$. Since $\|t\| = 1$, only a finite number of the t_k's can be zero. We denote the subsequence consisting of all nonzero t_k's again with $\{t_k\}$. From

$$\left\| \frac{t_k}{\|t_k\|} - t \right\| \leq (1 - \|t_k\| + \|t_k - t\|) \cdot \|t_k\|^{-1},$$

it follows that

$$\lim_{k \to \infty} (t_k/\|t_k\|) = t. \tag{3.13}$$

If we denote by P_{t_k} the one-dimensional projector corresponding to the subspace generated by t_k, it follows from (3.13) by applying Lemma 4, that

$$\lim_{k \to \infty} |P_{t_k} - P_t| = 0.$$

But obviously $P_{t_k} \in \mathfrak{F}^n[W_k]$. Therefore

$$P_t \in \overline{\bigcup_k \mathfrak{F}^n[W_k]}. \qquad \text{Q.E.D.}$$

In the following theorem we define the contraction operator L_n^p as a mapping from S_T^n into S_T^p. If we represent $D \in S_T^n$ as an integral operator (as it is done in the Schrödinger representation of quantum mechanics), then the action of L_n^p will correspond to the integration over the last $(n - p)$ configuration points.

Theorem 3: Let $D^n \in S_T^n$ and let p be a natural number $0 < p < n$. Let $\{e_i\}$ be a CONS in H^p adapted to the decomposition

$$H^p = H^{p\wedge} \oplus (H^{p\wedge})^{\perp}.$$

Moreover, let P_j^k be the following bounded operator, in H^p:

$$u \in H^p, \quad P_j^k u = (e_j \mid u) e_k.$$

Then the operator D^p defined on $\{e_i\}$ by

$$D^p e_i = \sum_{k=1}^{\infty} \text{Tr}\,([P_k^i \otimes I^{n-p}] D^n) e_k$$

can be extended to a linear bounded operator D^p belonging to S_T^p. The mapping $D^n \to D^p$ is independent of the special choice of the CONS occurring in its definition.

Proof. First of all it follows from $(P_j^k)^* = P_k^j$, using Lemma 1 (i) and (iv), that the matrix $\text{Tr}\,([P_k^i \otimes I^{n-p}] D^n)$ is Hermitian:

$$\text{Tr}\,([P_k^j \otimes I^{n-p}] D^n) = \text{Tr}\,([P_j^k \otimes I^{n-p}]^* D^n)$$
$$= \overline{\text{Tr}\,([P_j^k \otimes I^{n-p}] D^n)}.$$

Let $c_1, \cdots, c_p$ be any finite sequence of complex numbers, such that $\sum_{i=1}^r |c_i|^2 = 1$, and let $x = \sum_{i=1}^r c_i e_i$. Then the projectors onto the subspace spanned by x

can be written as

$$P_x = \sum_{k,j}^{r} c_k \bar{c}_j P_j^k.$$

Therefore, we have for any finite sequence $\{c_i\}$

$$\sum_{k,j} \mathrm{Tr}\,([P_j^k \otimes I^{n-p}]D^n)c_k\bar{c}_j = \mathrm{Tr}\,([P_x \otimes I^{n-p}]D^n) \leq |D^n|.$$

But this is exactly the condition that D^p can be extended to an everywhere-defined bounded and self-adjoint operator in H^p, (cf. Ref. 9, p. 54) which we still denote by D^p. The mapping $D^n \to D^p$ is obviously linear. Let us prove that D^p is positive whenever $D^n \in \mathfrak{K}_T^n$. For this purpose let $x \in H^p$ be arbitrary and $x = \sum_{i=1}^{\infty} c_i e_i$, its Fourier decomposition with respect to the CONS $\{e_i\}$. Then

$$P_x = \sum_{k,j}^{\infty} c_k \bar{c}_j P_j^k.$$

It is not hard to show that the right side converges in the operator norm. Now Lemma 1 (v) implies

$$(x \mid D^p x) = \sum_{k,j} \mathrm{Tr}\,([P_j^k \otimes I^{n-p}]D^n)c_k\bar{c}_j = \mathrm{Tr}\,([P_x \otimes I^{n-p}]D^n) \geq 0.$$

Therefore $D^p \geq 0$, and it makes sense to calculate $|D^p|$:

$$\begin{aligned} |D^p| = \mathrm{Tr}\,(D^p) &= \sum_{j=1}^{\infty} \mathrm{Tr}\,([P_j^j \otimes I^{n-p}]D^n) \\ &= \sum_{j=1}^{\infty} \sum_{i,k} (e_i \otimes u_k \mid D^n(P_j^j e_i) \otimes u_k) \\ &= \sum_{j,k} (e_j \otimes u_k \mid D^n(e_j \otimes u_k)) = \mathrm{Tr}\,(D^n) = |D^n|. \end{aligned} \tag{3.14}$$

Here $\{u_k\}$ denotes any CONS in H^{n-p}. It is clear that $\{e_i \otimes u_k\}$ then represents a CONS in H^n.

Hence, if $D^n \in \mathfrak{K}_T^n$, D^p is also of trace class. If D^n is an arbitrary element of S_T^n, we can write it as $D^n = D^{n+} - D^{n-}$ with $D^{n+}, D^{n-} \in \mathfrak{K}_T^n$. Since $D^n \to D^p$ is linear, it follows that $D^p = D^{p+} - D^{p-}$ is also of trace class.

We show that D^p is zero on the orthogonal complement of $H^{p\wedge}$. It is sufficient to show that

$$\mathrm{Tr}\,([P_k^i \otimes I^{n-p}]D^n) = 0$$

as soon as $e_i \notin H^{p\wedge}$ or $e_k \notin H^{p\wedge}$. For this purpose let us choose $t \in H^{n\wedge}$. Since $H^{n\wedge} \subset H^{p\wedge} \otimes H^{(n-p)\wedge}$, t can be written as

$$t = \sum_{j=1}^{\infty}{}' (e_j \otimes v_j),$$

where the sum $\sum'$ extends over those members of the CONS $\{e_j\}$ which belong to $H^{p\wedge}$. It follows that

$$(A_p \otimes I^{n-p})(P_k^i \otimes I^{n-p})t = \begin{cases} e_i \otimes v_k, & \text{if } e_i, e_k \in H^{p\wedge}, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore we have

$$A_n[P_k^i \otimes I^{n-p}]A_n = A_n(A_p \otimes I^{n-p})(P_k^i \otimes I^{n-p})A_n = 0$$

as soon as $e_i \notin H^{p\wedge}$ or $e_k \notin H^{p\wedge}$. Hence we get in this latter case

$$\begin{aligned} \mathrm{Tr}\,([P_k^i \otimes I^{n-p}]D^n) &= \mathrm{Tr}\,([P_k^i \otimes I^{n-p}]A_n D^n A_n) \\ &= \mathrm{Tr}\,(A_n[P_k^i \otimes I^{n-p}]A_n D^n) = 0. \end{aligned}$$

Thus we have indeed $D^p \in S^p$.

Finally, let us show that the definition of D^p is independent of the choice of the CONS $\{e_i\}$ in H^p. For this purpose let us introduce the group of unitary automorphisms U^p of H^p $(p = 1, \cdots, n)$. Now let $u \in U^p$ and

$$f_k = ue_k = \sum_{j=1}^{\infty} u_k^j e_j.$$

Hence,

$$e_i = u^* f_i = \sum_{j=1}^{\infty} \bar{u}_j^i e_j.$$

If we introduce the operator

$$x \in H^p, \quad Q_j^k x = (f_j \mid x) f_k,$$

we can easily deduce that

$$P_i^j = \sum_{l,k} u_l^i \bar{u}_k^j Q_l^k,$$

where the right side converges in the operator norm. Now let $\hat{D}^p$ be the operator in H^p defined by the linear continuous extension of

$$\hat{D}^p f_l = \sum_{k=1}^{\infty} \mathrm{Tr}\,([Q_k^l \otimes I^{n-p}]D^n) f_k.$$

Then

$$\begin{aligned} \hat{D}^p e_i &= \sum_{l=1}^{\infty} \bar{u}_l^i \hat{D}^p f_l \\ &= \sum_{l,k} \bar{u}_l^i \,\mathrm{Tr}\,([Q_k^l \otimes I^{n-p}]D^n) f_k \\ &= \sum_{l,k,j} \bar{u}_l^i u_k^j \,\mathrm{Tr}\,([Q_k^l \otimes I^{n-p}]D^n) e_j \\ &= \sum_{j=1}^{\infty} \mathrm{Tr}\,([P_j^i \otimes I^{n-p}]D^n) e_j. \\ &= D^p e_i \end{aligned}$$

Hence $\hat{D}^p = D^p$. Q.E.D.

Definition 1: The mapping $D^n \to D^p$ from S_T^n into S_T^p defined in Theorem 3 is called the (n, p)-*contraction operator*. We denote it by L_n^p. Thus $D^p = L_n^p(D^n)$.

[9] N. Akhieser and I. M. Glazman, *Linear Operators in Hilbert Space* (Frederick Ungar Publishing Company, New York, 1961), Vol. I, p. 54.

Theorem 4: The mapping $\Gamma_p^n: S^p \to S^n$, defined by

$$\Gamma_p^n(B^p) = A_n(B^p \otimes I^{n-p})A_n,$$

is the *adjoint mapping* of L_n^p, i.e.,

$$\mathrm{Tr}\,(L_n^p(D^n)B^p) = \mathrm{Tr}\,(D^n\Gamma_p^n(B^p)) \quad D^n \in S_T^n,\ B^p \in S^p.$$

Proof.

Lemma 5: L_n^p is a bounded map (in the trace norm).

Proof: Let $D^n = D^{n+} - D^{n-}$ be the *canonical* representation of D^n as a difference of two elements of $\mathfrak{K}_T^n$ (i.e., $R_{D^{n+}} \perp R_{D^{n-}}$). Then

$$\begin{aligned}|L_n^p(D^n)| &= |L_n^p(D^{n+}) - L_n^p(D^{n-})| \\ &\leq |L_n^p(D^{n+})| + |L_n^p(D^{n-})| \\ &= |D^{n+}| + |D^{n-}| \\ &= |D^n|.\end{aligned}$$

Here we made use of Eq. (3.14).

Now assume that $D^n \in S^n[W]$ (W = finite-dimensional subspace of H^1). Then $L_n^p(D^n) \in S^p[W]$. Choose in $H^{p\wedge}$ a CONS $\{e_i\}$ such that the first n members form a basis in

$$W^{p\wedge} = \bigotimes_{i=1}^{p} W(i) \cap H^{p\wedge}.$$

Then

$$B_W^p = A_p^W B^p A_p^W = \sum_{i,k} (e_i \mid B^p e_k) P_k^i.$$

It follows that

$$\begin{aligned}\mathrm{Tr}\,(L_n^p(D^n)B^p) &= \mathrm{Tr}\,(L_n^p(D^n)B_W^p) \\ &= \sum_{i,k}^{m} \mathrm{Tr}\,([P_k^i \otimes I^{n-p}]D^n)(e_i \mid B^p e_k) \\ &= \mathrm{Tr}\,([B_W^p \otimes I^{n-p}]D^n) \\ &= \mathrm{Tr}\,([B^p \otimes I^{n-p}]D^n) \\ &= \mathrm{Tr}\,(A_n[B^p \otimes I^{n-p}]A_n D^n) \\ &= \mathrm{Tr}\,(\Gamma_p^n(B^p)D^n).\end{aligned}$$

The same equation can be established for arbitrary $D^n \in S^n$ using Proposition 1, the continuity of L_n^p (Lemma 5), and the continuity of the trace.

Definition 2: The adjoint mapping Γ_p^n of the (n, p)-contraction operator is called the (p, n)-*expansion operator.*

Theorem 5: The mappings L_n^p and Γ_p^n have the following properties:

(i) They are linear and order preserving (i.e., they make correspond to a positive element again a positive element).

(ii) L_n^p is bounded with respect to the trace norm $|\ |$ and Γ_p^n is bounded with respect to the operator norm $\|\ \|$.

(iii) They are weakly continuous.

(iv) $L_n^p(S_T^n)$ is dense in S_T^p and Γ_p^n is injective.

(v) L_n^p and Γ_p^n are operator homomorphisms with respect to the group U^1 of unitary automorphisms of H^1.

(vi) Define $L_n^0 \equiv \mathrm{Tr}$ and let p, p' be two natural numbers satisfying $0 \leq p < p' < n'$. Then

$$L_n^p = L_{p'}^p \circ L_n^{p'}; \quad \Gamma_p^n = \Gamma_{p'}^n \circ \Gamma_p^{p'}.$$

Proof: (i) As far as it is not obvious, it has already been proved.

(ii) One-half of this statement is contained in Lemma 5. The other half follows from

$$\|\Gamma_p^n B^p\| = \|A_n(B^p \otimes I^{n-p})A_n\| \leq \|B^p \otimes I^{n-p}\| = \|B^p\|.$$

(iii) This statement is an immediate consequence of Theorem A4.

(iv) To prove that $L_n^p(S^p)$ is dense in S^p, we show that the restriction of L_n^p to $S^n[W]$ is a mapping onto $S^p[W]$ for every finite-dimensional subspace $W \subset H^1$ whose dimension is not smaller than $2n - 1$. The rest is a consequence of Proposition 1.

Assume that $L_n^p(S^n)$ is dense in S^p and let $B^p \neq 0$ be an element of the null space of Γ_p^n. Thus for every $D^n \in S_T^n$

$$\mathrm{Tr}\,(B^p L_n^p(D^n)) = \mathrm{Tr}\,(\Gamma_p^n(B^p)D^n) = 0.$$

It follows that the image $L_n^p(S_T^n)$ is a subset of the closed hyperplane $\{B^p\}^\perp$ of S_T^p, a consequence which contradicts the assumption of $L_n^p(S_T^n)$ being dense in S_T^p. Therefore Γ_p^n is injective.

It remains to show that the restriction of L_n^p to $S^n[W]$ is a mapping onto $S^p[W]$ for all finite-dimensional subspaces $W \subset H$ whose dimension exceeds a certain fixed number. If we denote the restriction of L_n^p to $S^n[W]$ by $\mathring{L}_n^p$ and by $\mathring{\Gamma}_p^n$ its adjoint mapping defined on $S^p[W]$ by $\mathring{\Gamma}_n^p(B^p) = A_n^W(B^p \otimes I^{n-p})A_n^W$, then the following lemma holds:

Lemma 6: Let $\mathfrak{I}^n$ denote the identity operator in $S^n[W]$. Then

$$(n+1)^2 \mathring{L}_{n+1}^n \mathring{\Gamma}_n^{n+1} = (r \pm 2n)\mathfrak{I}^n + \mathring{\Gamma}_{n-1}^n \mathring{L}_{n-1}^n$$
$$(+ \text{ for bosons; } - \text{ for fermions}). \quad (3.15)$$

Proof: In order to simplify the notation we drop in this proof the symbol $\circ$ on the top of the letters $\mathring{L}_n^{n+1}$, $\mathring{\Gamma}_n^{n+1}$, and we write A_n instead of A_n^W. According to a formula due to Sasaki,[2] we have

$$A_{n+1} = (n+1)^{-1}(A_n \pm nA(n, n+1)A_n).$$

Substituting this expression into the formula of definition for Γ_n^{n+1}, we get $D^n \in S^n[W]$

$$\begin{aligned}\Gamma_n^{n+1}(D^n) = (n+1)^{-2} & \\ \times (A_n(D^n \otimes I)A_n \pm nA_n(D^n \otimes I)A_n) & \\ \times (n+1, n)A_n \pm nA_n(n+1, n) & \\ \times A_n(D^n \otimes I)A_n + n^2 A_n(n, n+1) & \\ \times A_n(D^n \otimes I)A_n(n, n+1)A_n .\end{aligned}$$

If we apply on both sides of this equality the operator L_{n+1}^n, we obtain

$$\begin{aligned}&(n+1)^2 L_{n+1}^n \Gamma_n^{n+1}(D^n) \\ &= rD^n \pm 2nD^n + n^2 \Gamma_{n-1}^n L_{n-1}^{n-1}(D^n). \qquad \text{Q.E.D.}\end{aligned}$$

Lemma 7: For $\dim W \geq 2n-1$, the mappings $\mathring{L}_n^p \mathring{\Gamma}_p^n$ $(p = 0, 1, \cdots, n-1)$ are bijective.

Proof: For $p = n-1$ the assertion follows from a reformulation of Eq. (3.15):

$$n^2 \mathring{L}_n^{n-1} \mathring{\Gamma}_{n-1}^n = (r \pm (2n-2))\mathfrak{J}^{n-1} + (n-1)\mathring{\Gamma}_{n-2}^{n-1} \mathring{L}_{n-1}^{n-2}.$$

Since $\mathring{\Gamma}_{n-2}^{n-1} \mathring{L}_{n-1}^{n-2}$ is positive semidefinite on $S^{n-1}[W]$, the right side of this equation is obviously positive definite and therefore bijective, as soon as $r \geq 2n-1$.

Assume the lemma to be true for $p+1$. Since $r \geq 2n-1 \geq 2p+1$, $\mathring{L}_{p+1}^p \mathring{\Gamma}_p^{p+1}$ is bijective by the same argument as before. Therefore $\mathring{\Gamma}_p^{p+1}$ is injective and $\mathring{L}_{p+1}^p$ is surjective. Hence

$$L_n^p \mathring{\Gamma}_p^n = \mathring{L}_{p+1}^p (\mathring{L}_n^{p+1} \mathring{\Gamma}_{p+1}^n) \mathring{\Gamma}_p^{p+1}$$

is bijective. Q.E.D.

The assertion (iv) of Theorem 5 follows from Lemma 7 in a way, we have already indicated.

(v) Let us denote by $\Lambda^p(u), u \in U^1$, the representation of U^1 in $H^{p\wedge}$. Furthermore, let us introduce the notation u^p for the element

$$u^p = \underbrace{u \otimes \cdots \otimes u}_{(p \text{ factors})} \in U^p.$$

Then we have for $B^n \in S^p$

$$\begin{aligned}\Lambda^n(u)\Gamma_p^n(B^p)\Lambda^n(u)^* &= \Lambda^n(u)A_n(B^p \otimes I^{n-p})A_n\Lambda^n(u)^* \\ &= A_n u^n (B^p \otimes I^{n-p}) u^{n^*} A_n \\ &= A_n(u^p B^p u^{p^*} \otimes I^{n-p})A_n \\ &= A_n(\Lambda^p(u)B^p\Lambda^p(u)^* \otimes I^{n-p})A_n \\ &= \Gamma_p^n(\Lambda^p(u)B^p\Lambda^p(u)^*).\end{aligned}$$

By using that L_n^p is adjoint to Γ_p^n, we get the same result for $L_n^p(D^n \in S_T^n)$:

$$\begin{aligned}\operatorname{Tr}[L_n^p(\Lambda^n(u)D^n\Lambda^n(u)^*)B^p] & \\ = \operatorname{Tr}[\Lambda^n(u)D^n\Lambda^n(u)^*\Gamma_p^n(B_p)] & \\ = \operatorname{Tr}[D^n\Gamma_p^n(\Lambda^p(u)^*B^p\Lambda^p(u))] & \\ = \operatorname{Tr}[L_n^p(D^n)\Lambda^p(u)^*B^p\Lambda^p(u)] & \\ = \operatorname{Tr}[\Lambda^p(u)L_n^p(D^n)\Lambda^{p^*}(u)B^p] &\end{aligned}$$

for all $B^p \in S^p$. Hence,

$$L_n^p(\Lambda^n(u)D^n\Lambda^n(u)^*) = \Lambda^p(u)L_n^p(D^n)\Lambda^p(u)^*.$$

(vi) First let us prove this statement for the expansion operator Γ_n^p:

$$\begin{aligned}\Gamma_{p'}^n\Gamma_p^{p'}(B^p) &= A_n[A_{p'}(B^p \otimes I^{p'-p})A_{p'} \otimes I^{n-p'}]A_n \\ &= A_n(A_{p'} \otimes I^{n-p'})(B^p \otimes I^{n-p})(A_{p'} \otimes I^{n-p'})A_n \\ &= A_n(B^p \otimes I^{n-p})A_n \\ &= \Gamma_p^n(B^p)\end{aligned}$$

for all $B^p \in S^p$.

Again it is easy to prove the similar property for the adjoint mapping L_n^p:

$$\begin{aligned}\operatorname{Tr}[L_{p'}^p L_n^{p'}(D^n)B^p] &= \operatorname{Tr}[D^n\Gamma_{p'}^n\Gamma_p^{p'}(B^p)] \\ &= \operatorname{Tr}[D^n\Gamma_p^n(B^p)] \\ &= \operatorname{Tr}[L_n^p(D^n)B^p]\end{aligned}$$

for all $B^p \in S^p$. Q.E.D.

Definition 3: An observable B^n of the n-particle system which has the form $B^n = \Gamma_p^n(B^p)$ is called a *p-particle observable.*

Remark: Let $B^p \in S(H^p)$ such that it commutes with the representation (3.3) of the symmetric group, i.e.,

$$[B^p, G_s^p] = 0, \quad s \in \Pi_p .$$

Let B^n be the operator

$$B^n = \sum_{i_1 < \cdots i_p} B^p(i_1 \cdots i_p),$$

where the notation is understood as usual in literature about quantum mechanics. Then the product B^nA_n can be expressed with the help of the (p, n)-expansion operator as follows:

$$B^nA_n = \Gamma_p^n\left(\binom{n}{p} A_p B^p A_p\right).$$

Thus B^nA_n is a p-particle observable in our sense.

Definition 4: An element D^n of $\mathfrak{J}^n$ is called an *n-density operator.* Its image $D^p = L_n^p(D^p)$ under the (n, p) contraction is called the *reduced n-density operator of order p.*

We denote the set of all reduced n-density operators of order p by $\mathfrak{F}_n^p$: $\mathfrak{F}_n^p \equiv L_n^p(\mathfrak{F}^n)$.

Definition 5: An element of $\mathfrak{F}_n^p$ is called an *n-representable p-density operator.*

Since L_n^p is linear, and $\mathfrak{F}^n$ is convex, $\mathfrak{F}_n^p$ is a convex subset of S^p. Because of Theorem 5 (i) and (vi), we have

$$\mathfrak{F}_n^p \subset \mathfrak{F}^p. \tag{3.16}$$

The projection cone $\mathcal{K}_n^p$ of $\mathfrak{F}_n^p$ from the center 0 (cf. Definition A9) can be represented as

$$\mathcal{K}_n^p = L_n^p(\mathcal{K}_T^n) \subset \mathcal{K}_T^p. \tag{3.17}$$

Since $\mathcal{K}_T^p$ and $\mathfrak{F}^p$ are closed sets, it follows that

$$\overline{\mathcal{K}_n^p} \subset \mathcal{K}_T^p, \tag{3.16'}$$

$$\overline{\mathfrak{F}_n^p} \subset \mathfrak{F}^p. \tag{3.17'}$$

It can be shown that $\overline{\mathcal{K}_n^p}$ and $\overline{\mathfrak{F}_n^p}$ are, in general, proper subsets of $\mathcal{K}_T^p$ and $\mathfrak{F}^p$ (cf. Ref. 2) and the n-representability problem of the second kind, as it was stated in the Introduction, consists in the characterization of $\overline{\mathfrak{F}_n^p}$ as a subset of $\mathfrak{F}^p$ or, equivalently, in the characterization of $\overline{\mathcal{K}_n^p}$ as a subset of $\mathcal{K}_T^p$. Theorem 5(v) implies that $\mathfrak{F}_n^p$ (and $\mathcal{K}_n^p$) are invariant under the representation $\Lambda^p(u)$ of the unitary group U^1:

$$\Lambda^p(u)\mathfrak{F}_n^p\Lambda^p(u)^* = \mathfrak{F}_n^p \quad \text{for all} \quad u \in U^1. \tag{3.18}$$

Since the similarity transformation with a unitary operator is an isometry in S^p, the invariance still holds for the closure:

$$\Lambda^p(u)\overline{\mathfrak{F}_n^p}\Lambda^p(u)^* = \overline{\mathfrak{F}_n^p}. \tag{3.18'}$$

We may summarize the physically relevant content of Theorem 4 by the following Corollary.

Corollary to Theorem 4: (i) The information contained in the reduced n-density operator of order p is sufficient to calculate the expectation value of a p-particle observable

$$\langle \Gamma_p^n(B^p)\rangle_{D^n} = \mathrm{Tr}\,[L_n^p(D^n)B^p] \tag{3.19}$$

[cf. formula (1.7)].

(ii) The maximal and the minimal values α^p and β^p of the spectrum of a p-particle observable $\Gamma_p^n(B^p)$ can be expressed as

$$\alpha^p = \inf_{D^p \in \mathfrak{F}_n{}^p} \mathrm{Tr}\,(D^pB^p), \quad \beta^p = \sup_{D^p \in \mathfrak{F}_n{}^p} \mathrm{Tr}\,(D^pB^p).$$

Since the functional $\mathrm{Tr}\,(D^pB^p)$ is continuous in the first argument D^p, in formula (3.20) $\mathfrak{F}_n^p$ may also be replaced by $\overline{\mathfrak{F}_n^p}$.

An interesting answer to the challenge presented by the n-representability problem was given by Garrod and Percus.[1]

Theorem 6 (Precise Version of the Theorem of Garrod and Percus.): Let D^p be a p-density operator.

(i) If D^p is limit point of n-representable p-density operators and $\Gamma_p^n(B^p)$ is a positive p-particle observable, then $\mathrm{Tr}\,(B^pD^p) \geq 0$.

(ii) If D^p is *not* limit point of n-representable p-density operators, then there is a positive p-particle observable $\Gamma_p^n(B^p)$ which "discovers" this, i.e., for which

$$\langle \Gamma_p^nB^p\rangle_{D^n} = \mathrm{Tr}\,(D^pB^p) < 0.$$

A formal way to summarize these two statements is: Let $D^p \in S_T^p$, such that $\mathrm{Tr}\,(D^p) = 1$; then

$$D^p \in \overline{\mathfrak{F}_n^p} \Leftrightarrow \mathrm{Tr}\,(D^pB^p) \geq 0 \quad \forall \quad B^p \in S^p$$

such that $\Gamma_p^nB^p \geq 0$.

Remark: The formulation and the proof of this theorem given by Garrod and Percus is not quite satisfying because they assume that both operators, the one representing the observable as well as the one representing the state, are of Hilbert–Schmidt class (cf. paragraph 2 of this work). This is especially problematical for the one representing the state, because the trace of an operator of Hilbert–Schmidt class need not to exist. The advantage the authors gain with this assumption is that the proof of their theorem can be performed with the help of mathematical tools offered by the theory of Hilbert spaces, especially the "nearest-point theorem," because the operators of Hilbert–Schmidt class form a Hilbert space with respect to the scalar product: $(A, B) = \mathrm{Tr}\,(AB)$. In the proof of our version, the role of the "nearest-point theorem" is taken over by the "bipolar theorem" (cf. Theorem A6).

Proof: The theorem of Garrod and Percus is an immediate consequence of the bipolar theorem (Theorem A6) applied to the set $\mathfrak{F}_n^p \subset S^p$. According to this theorem we have

$$\overline{\mathfrak{F}_n^p} = \tilde{\mathfrak{F}}_n^p \cap \{D \in S_T^p;\ \mathrm{Tr}\,(D) = 1\}.$$

It remains to show that $\tilde{\mathfrak{F}}_n^p$ consists of all $B^p \in S^p$ such that $\Gamma_p^nB^p$ is a positive p-particle observable. But this is a straightforward consequence of the definition:

$$\begin{aligned} B^p \in \tilde{\mathfrak{F}}_n^p &\Leftrightarrow \mathrm{Tr}\,(D^pB^p) \geq 0 \quad \forall \quad D^p \in \mathfrak{F}_n^p \\ &\Leftrightarrow \mathrm{Tr}\,(L_n^p(D^n)B^p) \geq 0 \quad \forall \quad D^n \in \mathfrak{F}^n \\ &\Leftrightarrow \mathrm{Tr}\,(D^n\Gamma_p^n(B^p)) \geq 0 \quad \forall \quad D^n \in \mathfrak{F}^n \\ &\Leftrightarrow \Gamma_p^n(B^p) \geq 0. \end{aligned}$$

Q.E.D.

The following formulation of the theorem of Garrod and Percus shows that we can approximate a necessary and sufficient condition for n-representability by more and more necessary conditions.

Corollary to Theorem 6: The projection cone $\overline{\mathcal{K}_n^p}$ of $\overline{\mathcal{P}_n^p}$ from the center 0 coincides with the intersection of all half-spaces of type

$$H_{B^p} = \{D \in S_T^p;\ \mathrm{Tr}\,(B^p D) \geq 0\},$$

where $B^p \in S^p$ is such that $\Gamma_p^n(B^p)$ is a positive p-particle observable.

The theorem of Garrod and Percus has some rather remarkable consequences. First of all, it leads to the *dual problem of the n-representability problem*: What are the intrinsic properties of a bounded self-adjoint operator B^p in H^p with the property

$$\Gamma_p^n(B^p) = A_n(B^p \otimes I^{n-p})A_n \geq 0?$$

A more practical modification of this question is: How can we construct operators belonging to $\tilde{\mathcal{P}}_n^p = \tilde{\mathcal{K}}_n^p$? Since Γ_p^n is order preserving [Theorem 5(i)], we have $\mathcal{K}^p \subset \tilde{\mathcal{P}}_n^p$. But the condition $\mathrm{Tr}\,(D^p B^p) \geq 0$ corresponding to an element $B^p \in \mathcal{K}^p$ is trivial: it expresses only that D^p is positive. In this connection the following principle is of some interest:

Principle: Let α_p be the smallest real number contained in the spectrum of an element $B^p \in S^p$ and α_n the smallest real number contained in the spectrum of $\Gamma_p^n(B^p)$. Then

$$\mathrm{Tr}\,(B^p D^p) \geq \alpha_n \tag{3.21}$$

is a necessary condition for n-representability of D^p. It is nontrivial (i.e., it is not a consequence of $D^p \geq 0$) if and only if $\alpha_n > \alpha_p$.

Proof: Since $\Gamma_p^n(B^p - \alpha_n I^p) \geq 0$, $B^p - \alpha_n I^p \in \tilde{\mathcal{P}}_n^p$ and (3.21) is a consequence of the bipolar theorem. From

$$\alpha_p = \inf_{D^p \in \mathcal{P}^p} \mathrm{Tr}\,(B^p D^p) = \inf_{D^p \in \mathcal{P}(H^n)} \mathrm{Tr}\,[(B^p \otimes I^{n-p})D^n]$$
$$\leq \inf_{D^n \in \mathcal{P}^n} \mathrm{Tr}\,[\Gamma_p^n(B^p)D^n] = \alpha_n,$$

it follows that $B^p - \alpha_n I^p \notin \mathcal{K}^p$ if and only if $\alpha_n > \alpha_p$. Q.E.D.

Let us illustrate this principle in the case $p = 1$ and for fermions. Choose a one-dimensional projector $P^1 \in E^1$ and put $B^1 = -P^1$ Then $\alpha_1 = -1$, whereas $\alpha_n = -1/n$. Therefore

$$\mathrm{Tr}\,(B^1 D^1) = -\mathrm{Tr}\,(P^1 D^1) \geq -1/n$$

is a necessary condition for n-representability. It follows that

$$\|D^1\| = \sup_{P^1 \in E^1} \mathrm{Tr}\,(P^1 D^1) \leq 1/n,$$

the well-known necessary (and sufficient) condition for n-representability of the one-density operator (cf. Coleman, Ref. 3).

A certain class of elements belonging to $\tilde{\mathcal{K}}_n^2 = \tilde{\mathcal{P}}_n^2$ was constructed by Garrod and Percus.[1]

Proposition 2: (i) If B is an arbitrary bounded linear operator in H^1, then

$$C = A_2\{[BB^* \otimes I) + (I \otimes BB^*)] + (n-1)[B \otimes B^* + B^* \otimes B]\}$$

belongs to $\tilde{\mathcal{P}}_n^2$.

(ii) A necessary condition that the 2-density operator D^2 is n-representable is that the sesquilinear form

$$G_{D^2}(A, B) = \mathrm{Tr}\,[AB^* L_2^1(D^2)] + (n-1)\,\mathrm{Tr}\,[(A \otimes B^*)D^2],$$

defined on the space $B(H^1)$ of all bounded operators in H^1, is positive semidefinite.

Proof: (ii) This is an immediate conseqeunce of (i) and Theorem 6.

(i) This can be established by a straightforward calculation:

$$\frac{n}{2}\Gamma_2^n(C) = A_n\left\{\sum_{i=1}^n BB^*(i) + \sum_{i<k}[B(i)B^*(k) + B^*(i)B(k)]\right\}$$
$$= A_n\left(\sum_{i=1}^n B(i)\right)\left(\sum_{i=1}^n B(i)\right)^* \geq 0. \qquad \text{Q.E.D.}$$

4. EXPOSED POINTS OF $\mathcal{P}_n^p$; CONNECTION WITH THE FINITE-DIMENSIONAL n-REPRESENTABILITY PROBLEM

In this section we are going to give a description of $\mathcal{P}_n^p$ in terms of its exposed points.

Definition 6: A subspace V of $H^{n\wedge}$ is called a *p-subspace* if there exists a positive p-particle observable $B^p \in S^p$, such that $V = N_{\Gamma_p{}^n(B^p)}$, i.e., if V occurs as the nullspace of a positive p-particle observable. (It is a consequence of this definition that a p-subspace of $H^{n\wedge}$ is closed.)

Proposition 3: (i) The set $\mathfrak{L}^p(H^{n\wedge})$ of all p-subspaces of $H^{n\wedge}$ forms a lower sublattice (with two elements U, V; their intersection $U \cap V$ is contained in the subclass) of the lattice of all subspaces of $H^{n\wedge}$ containing the subspace $\{0\}$.

(ii)

$$\mathfrak{L}^p(H^{n\wedge}) \subset \mathfrak{L}^{p+1}(H^{n\wedge}),$$

or: a p-subspace is also a $(p+1)$-subspace

$$\mathfrak{L}^n(H^{n\wedge}) = \mathfrak{L}(H^{n\wedge});$$

in particular, every subspace of $H^{n\wedge}$ is an n-subspace.

(iii) Let $W \subset H^1$ be a closed subspace. Then

$$W^{n\wedge} = A_n \bigotimes_{i=1}^{n} W(i)$$

is a 1-subspace of $H^{n\wedge}$.

(iv) Let us denote by $\mathfrak{L}^p(W^{n\wedge})$ the sublattice of all p-subspaces contained in $W^{n\wedge}$. If $U \in \mathfrak{L}^p(W^{n\wedge})$, then U allows the representation

$$U = N_{\mathring{\Gamma}_p^n(B^p)} \cap W^{n\wedge}; \quad B^p \in S^p[W],$$

where $\mathring{\Gamma}_p^n$ stands for the mapping

$$\mathring{\Gamma}_p^n(B^p) = A_n^W(B^p \otimes I^{n-p})A_n^W.$$

Proof: (i) Let $U, V \in \mathfrak{L}^p(H^{n\wedge})$ and $\Gamma_p^n(A^p)$, $\Gamma_p^n(B^p)$ be two positive p-particle observables with U, V as their respective null spaces. Then the null space of

$$\Gamma_p^n(A^p + B^p) = \Gamma_p^n(A^p) + \Gamma_p^n(B^p)$$

is given by $U \cap V$. Therefore $U \cap V \in \mathfrak{L}^p(H^{n\wedge})$.

(ii) is a consequence of Theorem 5(vi).

(iii) Denote by Q the projector onto $W^\perp$. Then $W^{n\wedge}$ is the null space of $\Gamma_p^n(Q)$.

(iv) Let $U = N_{\Gamma_p^n(C^p)}$ and put $B^p = A_p^W C^p A_p^W$. Then

$$\begin{aligned} A_n^W \Gamma_p^n(C^p) A_n^W &= A_n^W (C^p \otimes I^{n-p}) A_n^W \\ &= A_n^W (A_p^W \otimes I^{n-p})(C^p \otimes I^{n-p}) \\ &\qquad \times (A_p^W \otimes I^{n-p}) A_n^W \\ &= A_n^W (A_p^W C^p A_p^W \otimes I^{n-p}) A_n^W \\ &= \mathring{\Gamma}_p^n(B^p). \end{aligned}$$

Therefore the intersection of the null spaces of $\Gamma_p^n(C^p)$ and $\mathring{\Gamma}_p^n(B^p)$ with $W^{n\wedge}$ coincide. Q.E.D.

Definition 7: A p-subspace of $H^{n\wedge}$ is called *minimal* if it is a minimal element of the partially ordered set $\mathfrak{L}^p(H^{n\wedge}) - [0]$.

The importance of the notion of a p-subspace of $H^{n\wedge}$ becomes clear from the following proposition:

Proposition 4: (i) Let e be an exposed subset of $\mathfrak{F}_n^p$, i.e.,

$$e = \{B^p\} \cap \mathfrak{F}_n^p$$

for some $B^p \in \tilde{\mathfrak{F}}_n^p$ (cf. Definition A13). Then the map ϕ from the class $\mathfrak{L}(\mathfrak{F}_n^p)$ of all exposed subsets of $\mathfrak{F}_n^p$ into the lattice $\mathfrak{L}^p(H^{n\wedge})$ of all p-subspaces of $H^{n\wedge}$ defined on e by

$$\phi(e) = N_{\Gamma_p^n(B^p)} \tag{4.1}$$

is an order isomorphism onto $\mathfrak{L}^p(H^{n\wedge}) - [0]$.

(ii) The full preimage of the exposed subset e of $\mathfrak{F}_n^p$ with respect to the (n,p)-contraction is the set

$$L_n^{p(-1)}(e) = \mathfrak{F}(\phi(e)) \equiv \{D^n \in \mathfrak{F}^n; R_{D^n} \subset \phi(e)\}.$$

Proof: (i) The proof is based on the logical equivalence

$$\{B^p\}^\perp \cap \mathfrak{F}_n^p \subset \{C^p\}^\perp \cap \mathfrak{F}_n^p \Leftrightarrow N_{\Gamma_p^n(B^p)} \subset N_{\Gamma_p^n(C^p)}; \quad B^p, C^p \in \tilde{\mathfrak{F}}_n^p. \tag{4.2}$$

It guarantees that ϕ is well defined, injective, and an order isomorphism. That ϕ is surjective onto $\mathfrak{L}^p(H^{n\wedge}) - [0]$ is an immediate consequence of its definition.

Let us then prove equivalence (4.2). Assume $x \in N_{\Gamma_p^n(B^p)}$. It follows that

$$\mathrm{Tr}\,[P_x \Gamma_p^n(B^p)] = \mathrm{Tr}\,[L_n^p(P_x) B^p] = 0,$$

where P_x denotes the one-dimensional projector corresponding to the subspace generated by x. Therefore $L_n^p(P_x) \in \{B^p\}^\perp \cap \mathfrak{F}_n^p$ and hence

$$L_n^p(P_x) \in \{C^p\}^\perp \cap \mathfrak{F}_n^p,$$

or

$$\mathrm{Tr}\,[L_n^p(P_x) C^p] = \mathrm{Tr}\,[P_x \Gamma_p^n(C^p)] = 0.$$

Lemma 3 implies now $x \in N_{\Gamma_p^n(B^p)}$.

On the other hand, suppose $D^p \in \{B^p\}^\perp \cap \mathfrak{F}_n^p$ or $D^p = L_n^p(D^n)$ and $\mathrm{Tr}\,[L_n^p(D^n) B^p] = \mathrm{Tr}\,[D^n \Gamma_p^n(B^p)] = 0$. Then Lemma 3 implies

$$R_{D^n} \subset N_{\Gamma_p^n(B^p)} \subset N_{\Gamma_p^n(C^p)}.$$

Therefore

$$\mathrm{Tr}\,[D^n \Gamma_p^n(C^p)] = \mathrm{Tr}\,(L_n^p(D^n) C^p) = \mathrm{Tr}\,(D^p C^p) = 0,$$

or, differently, expressed, $D^p \in \{C^p\}^\perp \cap \mathfrak{F}_n^p$.

(ii) Let $D^n \in \mathfrak{F}(\phi(e))$. Then $\mathrm{Tr}\,[L_n^p(D^n) B^p] = \mathrm{Tr}\,[D^n \Gamma_p^n(B^p)] = 0$.

Hence $L_n^p(D^n) \in \{B^p\}^\perp \cap \mathfrak{F}_n^p = e$. Conversely, let $D^p \in e$ and let $D^n \in \mathfrak{F}^n$ such that $L_n^p(D^n) = D^p$. It follows that

$$\mathrm{Tr}\,[D^n \Gamma_p^n(B^p)] = \mathrm{Tr}\,[L_n^p(D^n) B^p] = \mathrm{Tr}\,(D^p B^p) = 0.$$

Lemma 3 implies now $D^n \in \mathfrak{F}(\phi(e))$. Q.E.D.

Theorem 8 is an immediate consequence of Proposition 4.

Theorem 8: The preimage of an exposed point of $\mathfrak{F}_n^p$ with respect to the (n,p) contraction L_n^p consists of

the set $\mathfrak{T}(V)$ of all density operators (elements of $\mathfrak{T}^n$) whose range is contained in a fixed *minimal* p-subspace V.

Proof: Since an exposed point $\{D^p\}$ is a minimal element of $\mathfrak{L}(\mathfrak{T}_n^p)$ and ϕ is an order isomorphism, $\phi(\{D^p\})$ must be a minimal p-subspace of $H^{n\wedge}$. The rest is contained in Proposition 4(ii). Q.E.D.

The question arises: Is $\overline{\mathfrak{T}_n^p}$ the convex closure of all exposed points of $\mathfrak{T}_n^p$? This is indeed the case. We shall prove even a somewhat stronger result. Before we formulate it, we should like to draw the attention of the reader to the little fact expressed in Lemma 8.

Lemma 8: Let $W^{n\wedge}$ be as it was in Proposition 3(iii) and (iv). If V is a minimal p-subspace of $H^{n\wedge}$, then $V \cap W^{n\wedge} = [0]$ or $V \subseteq W^{n\wedge}$.

Proof: Since $W^{n\wedge}$ is a 1-subspace [Proposition 3(iii)], it is a p-subspace [Proposition 3(ii)]. Therefore the intersection $U \cap W^{n\wedge}$ is a p-subspace [Proposition 3(i)]. Since U is minimal, $U \cap W^{n\wedge} = [0]$ or

$$U \cap W^{n\wedge} = U. \qquad \text{Q.E.D.}$$

Theorem 9: Let $W_1 \subset W_2 \subset \cdots \subset H^1$ be an increasing sequence of finite-dimensional subspaces of H^1 such that

$$\overline{\bigcup_{i=1}^{\infty} W_i} = H^1,$$

and for any finite-dimensional subspace of H^1, let $E_n^p[W]$ denote the set of all exposed points D^p of $\mathfrak{T}_n^p$ for which one of the two equivalent conditions

(i) $D^p \in S^p[W]$,

(ii) $\phi(\{D^p\}) \subset W^{n\wedge}$

is satisfied.

Then

$$\overline{\mathfrak{T}_n^p} = \overline{\operatorname{conv} \bigcup_i E_n^p[W_i]}. \tag{4.3}$$

Proof: First of all we are going to show that, for any finite-dimensional subspace W of H^1, the set $E_n^p[W]$ coincides with the set of all exposed points of

$$\mathfrak{T}_n^p[W] \equiv L_n^p(\mathfrak{T}^n[W])$$

considered as a convex subset of $S^p[W]$.

It is easy to see that every point of $E_n^p[W]$ is exposed in $\mathfrak{T}_n^p[W]$. Conversely, let us assume that D_0^p is exposed in $\mathfrak{T}_n^p[W]$. Then $\{D_0^p\}$ can be represented as the intersection of a hyperplane $\{B^p\}^\perp \cap S^p[W]$ inside $S^p[W]$ with $\mathfrak{T}_n^p[W]$:

$$\{D_0^p\} = (\{B^p\}^\perp \cap S^p[W]) \cap \mathfrak{T}_n^p[W] = \{B^p\}^\perp \cap \mathfrak{T}_n^p[W].$$

Herein, $B^p \in S^p[W]$ is such that $\overset{\circ}{\Gamma}{}_p^n(B^p) \geq 0$.

Let

$$\alpha \equiv \inf_{D^n \in \mathfrak{T}^n} \operatorname{Tr}(\Gamma_p^n(B^p)D^n)$$

be the lower boundary point of the spectrum of $\Gamma_p^n(B^p)$.

We now define

$$C^p \equiv B^p + (1-\alpha)n\Gamma_1^p(Q), \tag{4.4}$$

wherein Q denotes the projector onto $W^\perp$. Then we have to prove that $C^p \in \tilde{\mathfrak{T}}_n^p$ and

$$\{D_0^p\} = \{C^p\}^\perp \cap \mathfrak{T}_n^p.$$

Indeed, let $D^p \in \mathfrak{T}_n^p$ be arbitrary and assume

$$D^p = L_n^p(D^n) \quad D^n \in \mathfrak{T}^n.$$

Furthermore, let $I^n = P + P^\perp$ be the decomposition of the unity corresponding to the representation

$$H^n = W^{n\wedge} \oplus W^{n\wedge\perp}$$

of H^n and define

$$D_1^n \equiv PD^nP,$$
$$D_2^n \equiv P^\perp D^nP^\perp.$$

It is a fact easy to verify that the operator $\Gamma_p^n(B^p)$ leaves the subspace $W^{n\wedge}$ and, as a consequence, also leaves its orthogonal complement $W^{n\wedge\perp}$ invariant. Therefore we can write

$$\operatorname{Tr}(C^pD^p) = \operatorname{Tr}(\overset{\circ}{\Gamma}{}_p^n(B^p)D_1^n) + \operatorname{Tr}(\Gamma_p^n(B^p)D_2^n) + (1-\alpha)n \operatorname{Tr}(\Gamma_1^n(Q)D_2^n).$$

Now from

$$\operatorname{Tr}(\overset{\circ}{\Gamma}{}_p^n(B^p)D_1^n) \geq 0,$$
$$\operatorname{Tr}(\Gamma_p^n(B^p)D_2^n) \geq \operatorname{Tr}(D_2^n)\alpha,$$
$$(1-\alpha)n \operatorname{Tr}(\Gamma_1^n(Q)D_2^n) \geq (1-\alpha)\operatorname{Tr}(D_2^n),$$

it follows that

$$\operatorname{Tr}(C^pD^p) \geq \operatorname{Tr}(D_2^n) \geq 0,$$

and hence, $C^p \in \tilde{\mathfrak{T}}_n^p$. Moreover, in order that $\operatorname{Tr}(C^pD^p) = 0$, it is necessary that $\operatorname{Tr}(D_2^n) = 0$. Since $D^n \geq 0$, this implies that $D^p \in \mathfrak{T}_n^p[W]$. But then the equation $\operatorname{Tr}(C^pD^p) = 0$ reduces to $\operatorname{Tr}(B^pD^p) = 0$, which, in turn, by assumption, is only satisfied if $D^p = D_0^p$.

The theorem of Klee (Theorem A8) now implies

$$\mathfrak{T}_n^p[W] = \operatorname{conv} E_n^p[W].$$

To prove formula (3) it remains to be shown that

$$\overline{\mathfrak{T}_n^p} = \overline{\bigcup_i \mathfrak{T}_n^p[W_i]}.$$

But from Proposition 1, formula (3.9), we know that

$$\mathfrak{T}^n = \overline{\bigcup_i \mathfrak{T}^n[W_i]}.$$

From the continuity of the contraction operator L_n^p,

it follows that

$$\overline{\mathfrak{F}_n^p} \subset \overline{\bigcup_i \mathfrak{F}_n^p[W_i]}.$$

Since the inverse inclusion is trivial, the proof of formula (3) is complete.

It remains to establish the equivalence of the two conditions (i) and (ii). If $\phi(\{D^p\}) \subset W^{n\wedge}$, Theorem 8 implies that $D^p = L_n^p(D^n)$ with $D^n \in \mathfrak{F}^n[W]$. Therefore, $D^p \in \mathfrak{F}_n^p[W]$. Conversely, assume D^p to be exposed in $\mathfrak{F}_n^p$ and $D^p \in S^p[W]$. If $D^p = L_n^p(D^n)D^n \in \mathfrak{F}^n$, then $\mathrm{Tr}\,(D^n\Gamma_1^n(Q)) = \mathrm{Tr}\,(D^p\Gamma_1^p(Q)) = 0$ and therefore, according to Lemma 3, $R_{D^n} \subseteq W^{n\wedge}$. On the other hand, Theorem 8 tells us that $R_{D^n} \subseteq \phi(\{D^p\})$. Hence $R_{D^n} \subseteq \phi(\{D^p\}) \cap W^{n\wedge}$. Since D^p is exposed in $\mathfrak{F}_n^p$, $\phi(\{D^p\})$ is minimal and Lemma 8 implies $\phi(\{D^p\}) \subset W^{n\wedge}$. Q.E.D.

A density operator $D^p \in \mathfrak{F}^p[W]$, with dim W finite, is said to be of *finite 1-rank*. The motivation for this name lies in the fact that if $D^p \in \mathfrak{F}^p[W]$, then $L_p^1(D^p) \in \mathfrak{F}^1[W] \equiv \mathfrak{F}^1(W)$. Conversely, if the range of the $(p, 1)$ contraction of a p-density operator is a subspace of W, then $\mathrm{Tr}\,(D^p\Gamma_1^p(Q)) = \mathrm{Tr}\,(L_p^1(D^p)Q) = 0$. Hence, by Lemma 3, $D^p \in \mathfrak{F}^p[W]$.

With the help of this notation, we are able to state Theorem 9 briefly:

Corollary to Theorem 9: $\overline{\mathfrak{F}_n^p}$ is the convex closure of the exposed points of $\mathfrak{F}_n^p$ of finite 1-rank.

In the light of this corollary, let us summarize the procedure to construct exposed points whose 1-range is contained in a given finite-dimensional subspace W of H^1.

(1) Choose an operator $B^p \in S^p[W]$ and form

$$\mathring{\Gamma}_p^n(B^p) = A_n^W(B^p \otimes I^{n-p})A_n^W.$$

(2) Look for the eigenspace $V \subseteq W^{n\wedge}$ of $\mathring{\Gamma}_p^n(B^p)$ to the lowest eigenvalue λ. Then V is a p-subspace [since V is the nullspace of $\Gamma_p^n(C^p)$, where $C^p = B^p - \lambda A_n + n(1 - \alpha + \lambda)\Gamma_1^p(Q)$. Herein, α denotes the lower boundary point of the spectrum of $\Gamma_p^n(B^p)$.]

(3) Choose in V a CONS $\{x_i\}$ and let P_{x_i} be the one-dimensional projector corresponding to x_i.

(4) If $L_n^p(P_{x_i}) \neq L_n^p(P_{x_k})$ for some pair $i \neq k$, return to point 1. Otherwise V is a minimal p-subspace and

$$D^p = L_n^p(P_{x_1})$$

is an exposed point of $\mathfrak{F}_n^p$. Sufficiently many exposed points of $\mathfrak{F}_n^p$ can be obtained in this way.

This procedure is a summary of the results contained in the proof of Theorem 9 and Theorem 8.

shows the intimate connection of the n-representability problem of the p-density operator with the eigenvalue problem for a p-particle observable, so that there is no real hope for an analytical solution of the n-representability problem in the case $p > 1$ and for arbitrary dimension of the space W.

In the case $p = 1$ and fermions, the space V obtained by the above procedure is independent of the choice of B^p one-dimensional and spanned by a "Slater determinant":

$$[U] = A_n(e_1 \otimes \cdots \otimes e_n).$$

(U denotes an n-dimensional subspace of V and $e_1, \cdots, e_n$ a CONS in U.)

Therefore, V is minimal and the 1-density operators of type

$$D^1 = L_n^1(P_{[U]}) = (1/n)P_{U,U} \subset W$$

are all exposed points belonging to $E_n^p[W]$. Together with Theorem 9, we get Coleman's Theorem (Theorem 1b).

ACKNOWLEDGMENTS

This paper was written partially when the author was at the Institute for Physical Chemistry of the Swiss Federal Institute of Technology in Zürich, Switzerland, and partially when he was staying as a post-doctoral fellow at the Mathematical Department of Queen's University, Kingston, Ontario, Canada. At both places the author has met people who have supported his work either by passive tolerance or by active intellectual participation. First of all, he wants to express his gratitude to Professor A. J. Coleman from Queen's University for his kind hospitality and the lively interest he has shown to the growing of this work. Secondly, the author should like to thank Professor A. Huber in Zürich and Professor I. Halperin for their stimulating lectures in functional analysis and for helping him personally to overcome some mathematical difficulties. Moreover, he wants to thank Professor H. Primas and Professor H. H. Günthardt in Zürich for allowing the author to spend a good portion of his time thinking about the n-representability problem rather than performing more "useful" MO calculations.

Last but not least, the author is very grateful to his colleague, Robert Erdahl, for his critical discussions and for correcting several grammatical errors in the text.

The author was supported financially in Switzerland by the National Foundations and by the "Paul Karrer Stiftung für Stipendien auf dem Gebiete der Chemie," and in Canada by the National Research Council. The author would like to express his thanks to these three institutions.

APPENDIX: SOME NOTIONS AND THEOREMS ABOUT REAL LINEAR SPACES AND CONVEX SETS (cf. REF. 7)

Definition A1 (Ref. 7, p. 31): A *dual pair* (of real linear spaces) is a pair of real linear spaces $\langle E_1, E_2\rangle$ such that to every pair $(x, u) \in E_1 \times E_2$ there corresponds a real number denoted by xu, such that: (D1) xu is a bilinear form; (D2′) if $xu = 0$ for all $u \in E_2$, then $x = 0$; (D2″) if $xu = 0$ for all $x \in E_1$, then $u = 0$.

Definition A2: A topology in E_1 is called *compatible with the dual pair* $\langle E_1, E_2\rangle$ if it is locally convex and if the topological dual space of E_1 coincides with E_2.

Theorem A1 (Ref. 7, p. 34, Proposition 8): Let $\langle E_1, E_2\rangle$ be a dual pair and let M be a subset of E_1. The operation of taking the convex closure of the set M is independent of the topology as long as it is compatible with the dual pair.

Definition A3 (Ref. 7, p. 32): Let $\langle E_1, E_2\rangle$ be a dual pair. The weakest topology on E_1 compatible with the dual pair coincides with the weakest topology under which the set of all linear functionals $f_u(x) = xu$ $(u \in E_2)$ are continuous. It is called the *weak topology* in E_1.

Definition A4: Let $\langle E_1, E_2\rangle$ and $\langle F_1, F_2\rangle$ be two dual pairs. Let A be a map from E_1 into F_1. Then the expression $(Ax)v$ represents for fixed $v \in F_2$ a linear form in E_1. In other words, there exists an element $v' \in E_1^*$ such that $(Ax)v = xv'$. The map $A' : v \to v'$ is called the *adjoint map of A*. If A is linear, A' is linear.

Theorem A4 (Ref. 7, p. 38): Let $\langle E_1, E_2\rangle$ and $\langle F_1, F_2\rangle$ be two dual pairs. A linear map A from E_1 into F_1 is weakly continuous if and only if $v' = A'v \in E_2$ for every $v \in F_2$.

Definition A5 (Ref. 7, p. 4): Let E be a real linear space. A subset C of E is called *convex* if, whenever $x, y \in C$, the whole segment

$$[xy] = \{z \in E;\ z = \alpha x + (1 - \alpha)y,\ 0 \leq \alpha \leq 1\}$$

is a subset of C.

Definition A6 (Ref. 10, p. 137): A subset C' of a convex set C is called *extreme*, if whenever $[x\,y] \subset C$ shares an inner point with C', then in fact $[x\,y] \subset C'$. An extreme subset of C consisting of a single point is called an *extreme point.*

Definition A7 (Ref. 10, p. 27): A convex subset $\mathfrak{K}$ of a real linear space E is called a *convex cone* if whenever $x \in \mathfrak{K}$, then the half-ray

$$r(x) = \{z \in E;\ z = \alpha x \quad \alpha \geq 0\}$$

is a subset of $\mathfrak{K}$. A convex cone is called *pointed* if it does not contain a one-dimensional subspace of E.

Definition A8 (Ref. 10, p. 11): A point a belonging to a convex cone $\mathfrak{K}$ in a real linear space E is called a *core point of* $\mathfrak{K}$ if for every $x \in E$, there exists an $\epsilon > 0$ such that $a + \epsilon x \in \mathfrak{K}$.

Remark (Cf. Ref. 11, p. 180 "core point" = "algebraisch innerer Punkt."): If a convex cone $\mathfrak{K}$ lying in E does not contain a core point, then it does not contain an inner point in any topology compatible with the linear structure of E.

Definition A9 (Cf. Ref. 11, p. 187): Let C be a convex subset of a real linear space E not containing 0. The *projection cone of C from the center 0* is the smallest convex cone containing C.

Definition A10 (Ref. 7, p. 35): Let $\langle E_1, E_2\rangle$ be a dual pair and let M be a subset of E_1. The subspace of E_2 defined by

$$M^{\perp} = \{u \in E_2;\ xu = 0\ \forall\ x \in M\}$$

is called the *orthogonal complement* of M.

Theorem A5 (Special Case of the Bipolar Theorem): (i) $M^{\perp}$ is a closed subspace of E_2; (ii) $M^{\perp\perp}$ is the smallest closed subspace containing M.

Definition A11 (Ref. 7, p. 34): Let $\langle E_1, E_2\rangle$ be a dual pair and let M be a subset of E_1. Then the subset

$$\tilde{M} = \{u \in E_2;\ xu \geq 0\ \forall\ x \in M\}$$

is called the *polar cone* of M. ($M_1 \subset M_2$ implies $\tilde{M}_1 \supset \tilde{M}_2$.)

Theorem A6 (Bipolar Theorem: Ref. 7, p. 36; cf. also Ref. 12): (i) $\tilde{M}$ is a closed convex cone contained in E_2; (ii) $\tilde{\tilde{M}}$ is the smallest closed convex cone containing M.

Definition A12: Let $\langle E_1, E_2\rangle$ be a dual pair and $a \in E_2$ $(a \neq 0)$. Then the orthogonal complement $\{a\}^{\perp}$

[10] F. A. Valentine, *Convex Sets* (McGraw-Hill Book Company, Inc., New York, 1964).

[11] G. Köthe, *Topologische Lineare Räume* (Springer-Verlag, Berlin, 1960).

[12] I. Halperin, Trans. Roy. Soc. Canada, **47**, Sec. 3, 1 (1953).

is a closed subspace of codimension 1 in E_1 [cf. Theorem A5(i)]. Let $b \in E_1$. Then the set $b + \{a\}^\perp$ is called a *closed hyperplane* in E_1.

Definition A13 (cf. Ref. 11, p. 340): Let $\langle E_1, E_2 \rangle$ be a dual pair and $\mathcal{K}$ a closed convex cone in E_1 and let $a \in \tilde{\mathcal{K}} - \{0\}$. Then $\{a\}^\perp$ is called a *supporting hyperplane* of $\mathcal{K}$. The intersection $\{a\}^\perp \cap \mathcal{K}$ is called an *exposed subset* of $\mathcal{K}$. A point of $\mathcal{K}$ belonging to an exposed subset is called a *supporting point* of $\mathcal{K}$. A ray $r(x) = \{z \in E; z = \alpha x, \alpha \geq 0\}$, being at the same time an exposed subset of $\mathcal{K}$, is called an *exposed ray of* $\mathcal{K}$.

Remark: If $\mathcal{K}$ is the projection cone of a closed convex subset C of a closed hyperplane not containing 0, and if $a \in \mathcal{K}$, then the set $\{a\}^\perp \cap C$ is called an *exposed subset* of C provided it is not empty. An exposed subset of C consisting of a single point is called an *exposed point.*

Theorem A7: Every exposed subset of a closed convex cone $\mathcal{K} \subset E_1$ is closed, convex, and extreme.

Proof: That $\mathcal{K}' = \{a\}^\perp \cap \mathcal{K}$ is closed and convex is a consequence of the same properties of $\{a\}^\perp$ and $\mathcal{K}$ [cf. Theorem A5(i)]. Let $x,y \in \mathcal{K}$ $(x \neq y)$ and $0 < \alpha < 1$. Furthermore, let

$$\alpha x + (1 - y)y \in \mathcal{K}',$$

i.e.,

$$\alpha xa + (1 - \alpha)ya = 0.$$

Since $a \in \tilde{\mathcal{K}}$, $xa \geq 0$ and $ya \geq 0$. It follows that $xa = ya = 0$, so that $x,y \in \mathcal{K}'$. Q.E.D.

Theorem A8 (Theorem of Klee; cf. Ref. 11, p. 340): Let C be a compact convex subset of a normed linear space. C is the closed convex null of the exposed points of C.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Phase Transition of a Bethe Lattice Gas of Hard Molecules†

L. K. RUNNELS*
Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana

(Received 10 March 1967)

A pseudo-lattice, or homogeneous Husimi tree, is simpler statistically than a true lattice in two or three dimensions, since the pseudo-lattice contains only low-order cycles. The prototype is the "Bethe lattice," or Cayley tree, containing no cycles at all. Reported here are the results of studies of the lattice statistics of the Bethe lattice of coordination number three, using the language of the hard molecule lattice gas; the size of an adsorbed molecule prevents simultaneous occupancy of the same site or any of the three nearest-neighbor sites. Using methods related to those used in enumerating graphs, a recursion relation is obtained which must be satisfied by the grand partition function. It is shown that the resulting equation of state is *not* the quasi-chemical equation expected because of the absence of cycles. There is no phase transition of lower than third order and in all likelihood none at all. The quasi-chemical equation of state is obtained only if "surface" effects are eliminated; even then the solution is valid only for activities $z < 4$, at which point the "interior" system undergoes a second-order transition to an ordered state with a finite discontinuity in compressibility.

I. INTRODUCTION

THE intractability of rigorous mathematical equations describing actual physical situations leads naturally to the employment of approximations in the search for partial understanding. These approximations often may be classified as essentially mathematical in nature or as essentially physical simplifications. Into the latter category fall many model systems studied to elucidate the behavior of classical fluids and phase transitions.

One such system is the model of infinitely hard molecules, devoid of mutual attractions. Physically, such a model would be expected to have relevance for temperatures so high that available thermal energy (kT) considerably exceeds in magnitude any intermolecular attractions. In another class of model systems molecular positions are restricted to a discrete set of lattice sites. These calculations may refer directly

† Supported in part by N.S.F. Grant No. GP-6822; computer time supported by N.S.F. Grant No. GP-2812.
* Alfred P. Sloan Foundation Fellow.

is a closed subspace of codimension 1 in E_1 [cf. Theorem A5(i)]. Let $b \in E_1$. Then the set $b + \{a\}^\perp$ is called a *closed hyperplane* in E_1.

Definition A13 (cf. Ref. 11, p. 340): Let $\langle E_1, E_2 \rangle$ be a dual pair and $\mathcal{K}$ a closed convex cone in E_1 and let $a \in \tilde{\mathcal{K}} - \{0\}$. Then $\{a\}^\perp$ is called a *supporting hyperplane* of $\mathcal{K}$. The intersection $\{a\}^\perp \cap \mathcal{K}$ is called an *exposed subset* of $\mathcal{K}$. A point of $\mathcal{K}$ belonging to an exposed subset is called a *supporting point* of $\mathcal{K}$. A ray $r(x) = \{z \in E; z = \alpha x, \alpha \geq 0\}$, being at the same time an exposed subset of $\mathcal{K}$, is called an *exposed ray of* $\mathcal{K}$.

Remark: If $\mathcal{K}$ is the projection cone of a closed convex subset C of a closed hyperplane not containing 0, and if $a \in \mathcal{K}$, then the set $\{a\}^\perp \cap C$ is called an *exposed subset* of C provided it is not empty. An exposed subset of C consisting of a single point is called an *exposed point.*

Theorem A7: Every exposed subset of a closed convex cone $\mathcal{K} \subset E_1$ is closed, convex, and extreme.

Proof: That $\mathcal{K}' = \{a\}^\perp \cap \mathcal{K}$ is closed and convex is a consequence of the same properties of $\{a\}^\perp$ and $\mathcal{K}$ [cf. Theorem A5(i)]. Let $x,y \in \mathcal{K}$ $(x \neq y)$ and $0 < \alpha < 1$. Furthermore, let

$$\alpha x + (1 - y)y \in \mathcal{K}',$$

i.e.,

$$\alpha xa + (1 - \alpha)ya = 0.$$

Since $a \in \tilde{\mathcal{K}}$, $xa \geq 0$ and $ya \geq 0$. It follows that $xa = ya = 0$, so that $x,y \in \mathcal{K}'$. Q.E.D.

Theorem A8 (Theorem of Klee; cf. Ref. 11, p. 340): Let C be a compact convex subset of a normed linear space. C is the closed convex null of the exposed points of C.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Phase Transition of a Bethe Lattice Gas of Hard Molecules†

L. K. RUNNELS*
Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana

(Received 10 March 1967)

A pseudo-lattice, or homogeneous Husimi tree, is simpler statistically than a true lattice in two or three dimensions, since the pseudo-lattice contains only low-order cycles. The prototype is the "Bethe lattice," or Cayley tree, containing no cycles at all. Reported here are the results of studies of the lattice statistics of the Bethe lattice of coordination number three, using the language of the hard molecule lattice gas; the size of an adsorbed molecule prevents simultaneous occupancy of the same site or any of the three nearest-neighbor sites. Using methods related to those used in enumerating graphs, a recursion relation is obtained which must be satisfied by the grand partition function. It is shown that the resulting equation of state is *not* the quasi-chemical equation expected because of the absence of cycles. There is no phase transition of lower than third order and in all likelihood none at all. The quasi-chemical equation of state is obtained only if "surface" effects are eliminated; even then the solution is valid only for activities $z < 4$, at which point the "interior" system undergoes a second-order transition to an ordered state with a finite discontinuity in compressibility.

I. INTRODUCTION

THE intractability of rigorous mathematical equations describing actual physical situations leads naturally to the employment of approximations in the search for partial understanding. These approximations often may be classified as essentially mathematical in nature or as essentially physical simplifications. Into the latter category fall many model systems studied to elucidate the behavior of classical fluids and phase transitions.

One such system is the model of infinitely hard molecules, devoid of mutual attractions. Physically, such a model would be expected to have relevance for temperatures so high that available thermal energy (kT) considerably exceeds in magnitude any intermolecular attractions. In another class of model systems molecular positions are restricted to a discrete set of lattice sites. These calculations may refer directly

† Supported in part by N.S.F. Grant No. GP-6822; computer time supported by N.S.F. Grant No. GP-2812.
* Alfred P. Sloan Foundation Fellow.

to an adsorbed phase or may be regarded as approximations to continuum fluid description—the mesh of the lattice being a measure of the severity of the approximation.

The intersection of these two classes of physical restrictions results in the "hard sphere lattice gas" model, investigated recently by Burley,[1] Temperley,[2] Gaunt and Fisher,[3] Runnels and Combs,[4] Bellemans,[5] Ree and Chesnut,[6] and others. Rather common to these systems is a phase transition between a low-density (disordered) phase and a high-density (ordered) phase. The thermodynamic order of the transition, however, is usually difficult to determine.

Rushbrooke and Scoins[7] discussed the lattice gas with arbitrary nearest-neighbor interactions in the context of the Mayer cluster expansion of an imperfect gas. They established a hierarchy of approximations to the exact equation of state, ordered according to the maximum size of lattice figures included in computing the cluster "integrals" (actually sums): isolated sites, pairs of adjacent sites, higher multiply-connected figures. It was shown that the first approximation amounted to the Langmuir treatment, while the inclusion of pairs of sites was equivalent to the results of Bethe[8] and the "quasi-chemical" approximation of Guggenheim[9]; details were given of the next estimate, which incorporates the smallest multiply-connected figure of the lattice. This "ring" approximation has received considerable use,[10] while systematic inclusion of even larger figures has formed the basis of very accurate expansions.[11]

The work of Rushbrooke and Scoins has led to the assumption that, in some sense, the various approximations are exact for certain "pseudo-lattices." Hence, the quasi-chemical result should describe a Cayley tree (or "Bethe lattice") exactly, since it has no cycles[12]; and the ring approximation has been used by Temperley[13] as an exact treatment of a Husimi tree (or "cactus lattice") bearing hard molecules, since it contains no cycles other than triangles. Two points, however, have not been clearly understood: "surface" effects and the range of densities (activities) describable in this manner. We clarify both of these points for a Cayley tree by treating the entire "lattice," including surface, in a new fashion; the surface-free behavior is obtained from the local intensive variables in regions far removed from the surface. Corresponding investigations can be carried out for more general "star trees," including the cactus lattice, but these will be reserved for a later publication.

II. PARTITION FUNCTION OF A CAYLEY TREE GAS OF HARD MOLECULES

We consider here the model of hard molecules adsorbed on a Cayley tree; the only interaction being an infinite repulsion preventing simultaneous occupancy of adjacent sites (or more than one molecule on the same site). Gaining simplicity at the expense of generality, we consider only the homogeneous tree T_k of coordination number three shown in Fig. 1; all sites have three neighboring sites except those at the "surface" (generation 1) and the one site ("root") at the center (generation k).

Clearly, the "volume" of such a tree is $M_k = 2^k - 1$. We denote the grand partition function for this tree bearing hard molecules by $\Xi_k(z)$, which then is given by $\Xi_k(z) = \sum_n W(k, n)z^n$, where z is the activity and $W(k, n)$ is the number of arrangements of n (indistinguishable) molecules on the M_k sites with no two adjacent. The feature possessed by tree lattices that renders them manageable is a configuration classification scheme closely related to techniques used to enumerate graphs.[14] The configurations of tree T_k can be divided into two classes: those with the root empty and those with the root occupied. The configurations in the former class are exactly those of two independent trees T_{k-1}, and so enumerated by $[\Xi_{k-1}(z)]^2$, while the latter class includes exactly the configurations of four independent trees T_{k-2}, enumerated by $[\Xi_{k-2}(z)]^4$. Thus, the partition function satisfies the recursion

$$\Xi_k(z) = [\Xi_{k-1}(z)]^2 + z[\Xi_{k-2}(z)]^4, \qquad (1)$$

the factor z in the last term accounting for the molecule at the root. Initial conditions are clearly $\Xi_0(z) = 1$, $\Xi_1(z) = 1 + z$.

The three-term, fourth-order relation (1) can be simplified through the introduction of correlation factors c_j defined by

$$c_j(z) = \Xi_j(z)/[\Xi_{j-1}(z)]^2, \qquad (2)$$

which are easily seen to satisfy the two-term recursion

$$c_j(z) = 1 + z/[c_{j-1}(z)]^2. \qquad (3)$$

[1] D. M. Burley, Proc. Phys. Soc. (London) **75**, 262 (1960).
[2] H. N. V. Temperley, Proc. Phys. Soc. (London) **74**, 183, 432 (1959); **77**, 630 (1961); **80**, 813, 823 (1962).
[3] D. S. Gaunt and M. E. Fisher, J. Chem. Phys. **43**, 2840 (1965).
[4] L. K. Runnels and L. L. Combs, J. Chem. Phys. **45**, 2482 (1966).
[5] A. Bellemans and R. K. Nigam, private communication; also, Phys. Rev. Letters **16**, 1038 (1966).
[6] F. H. Ree and D. A. Chesnut, J. Chem. Phys. **45**, 3983 (1966); Phys. Rev. Letters **18**, 5 (1967).
[7] G. S. Rushbrooke and H. I. Scoins, Proc. Roy. Soc. (London) **A230**, 74 (1955).
[8] H. A. Bethe, Proc. Roy. Soc. (London) **A150**, 552 (1935).
[9] E. A. Guggenheim, Proc. Roy. Soc. (London) **A148**, 304 (1935).
[10] D. M. Burley, Proc. Phys. Soc. (London) **77**, 451 (1961); **85**, 1173 (1965); Ref. 5.
[11] Reference 3; D. S. Gaunt, private communication.
[12] C. Domb, Advan. Phys. **9**, 245 (1960).
[13] H. N. V. Temperley, Proc. Phys. Soc. (London) **86**, 185 (1965).
[14] J. Riordan, *An Introduction to Combinational Analysis* (John Wiley & Sons, Inc., New York, 1958), Chap. 6.

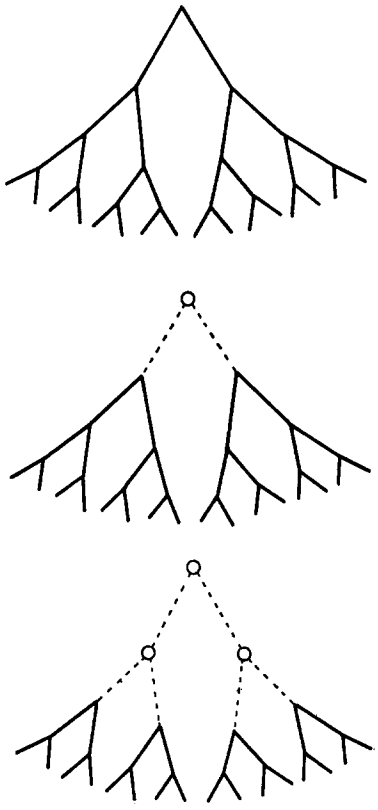


FIG. 1. Homogeneous Cayley tree, or Bethe lattice, of coordination number three. With nearest-neighbor exclusions, lattice configurations are conveniently classified according to the occupancy of the "root": if empty, there remain two independent trees of one generation less (center); but if occupied, there remain four independent trees of two generations less (bottom).

From the definition (2), an indentity can be written for the partition function

$$[\Xi_k(z)]^{2^{-k}} = \prod_{j=1}^{k} [c_j(z)]^{2^{-j}}, \tag{4}$$

and the pressure (in units of energy)

$$p/kT = M_k^{-1} \ln \Xi_k = \sum_j 2^{-j} \ln c_j(z). \tag{5}$$

In the last equation, we have assumed k to be large, so that M_k may be replaced by 2^k.

It should be noticed that the correlation factors c_j and the resulting identities are essentially in the spirit of the successive approximation scheme introduced by Stillinger, Salsburg, and Kornegay[15] in a study of rigid disks at high density, and applied by Salsburg *et al.*[16] to the high-temperature harmonic solid. We shall usually consider the limit of large k, so that the upper limit of the summation in Eq. (5) may be taken as infinity. The question of convergence is an important one and will be taken up shortly.

[15] F. H. Stillinger, Jr., Z. W. Salsburg, and R. L. Kornegay, J. Chem. Phys. **43**, 932 (1965).
[16] Z. W. Salsburg and W. Rudd, Physica **32**, 1601 (1966).

To obtain the fractional density ρ (average number of occupied sites/total number of sites) and further thermodynamic properties, we obtain from Eq. (5) the expansion

$$\rho = z \frac{\partial(p/kT)}{\partial z} = \sum_{j=1}^{\infty} 2^{-j} t_j, \tag{6}$$

where the sequence $t_j = z\partial \ln c_j/\partial z$ satisfies the recursion

$$t_j = b_j(1 - 2t_{j-1}), \quad [t_0 = 0] \tag{7}$$

in terms of the quantities b_j defined by

$$c_j(z) = [1 - b_j(z)]^{-1}. \tag{8}$$

Explicitly, we have

$$t_j = \sum_{i=0}^{j-1} (-2)^i \prod_{n=j-i}^{j} b_n, \tag{9}$$

which may be inserted in Eq. (6) and rearranged to give an alternate expression for the density:

$$\rho = \sum_{j=1}^{\infty} 2^{-j} r_j, \tag{10}$$

where

$$r_j = b_j - b_j b_{j+1} + b_j b_{j+1} b_{j+2} + \cdots,$$
$$= \sum_{i=0}^{\infty} (-1)^i \prod_{n=j}^{j+i} b_n. \tag{11}$$

The rearrangement is justified in the Appendix.

A simple physical interpretation of r_j is possible. From Eq. (1), it is clear that $z[\Xi_{k-2}(z)]^4/\Xi_k(z)$ is the fraction of (activity-weighted) configurations of tree T_k having a molecule at the root—hence the average occupancy of the root at activity z. From the preceding equations, this factor is simply b_k. Then the average occupancy r_j of a site at generation j is given rigorously by

$$r_j = b_j(1 - r_{j+1}); \tag{12}$$

the factor b_j is the weighted fraction of configurations of a tree T_j (i.e., from the selected site out to the surface) with the selected site occupied, while the factor $(1 - r_{j+1})$ accounts for the fraction of configurations of T_k with the inner site adjacent to the selected site vacant to permit occupancy of the selected site. Equation (12) easily iterates to give Eq. (11). The expansion (10) shows that the overall density is a weighted average of the densities of the various generations, the weighting factor 2^{-j} being proportional to the number of sites in generation j.

Exactly the same formulas (5) and (10) may be obtained for the pressure and density by a more tedious process. An explicit expression can be written for the degeneracy factor W in terms of combinatorial factors involving the number of molecules N_j on

generation j. Evaluating the expression, using a maximum term technique, yields the most probable number N_j^* of molecules to be found on generation j; it is found that N_j^* is given by $2^{k-j}r_j$, as required for consistency.

Since the energy is identically zero in the present model, entropy per molecule is given by $s/k = p/\rho kT - \ln z$, and other properties of interest may be obtained by differentiations according to thermodynamics.

III. THERMODYNAMIC BEHAVIOR OF THE ENTIRE CAYLEY TREE

In this section, we discuss the most direct interpretation of the Cayley tree gas; namely, we assume the entire tree (volume) is to be taken into consideration. A numerical approach for finite k is considered first.

We can recursively compute a large number of the correlation coefficients c_j for a fixed value of the activity z, taking $c_0 = 1$. This allows tabulation of the terms t_j also. Through Eqs. (5) and (6), partial sum approximations to p/kT and ρ may be obtained. It is found that these sequences appear to converge rapidly, so that repetition for a series of activities yields the equation of state shown in Fig. 2 upon elimination of z.

Since the partial sum approximants are actually exact descriptions of finite trees (to the extent that $2^k \gg 1$), it could be argued that more than about eighty terms are superfluous, for such a tree contains more than Avogadro's number of sites! In any event, considerably fewer terms are needed for excellent convergence. It is seen that there are no unusual or striking properties. In particular, there seems to be no evidence of phase transitions.

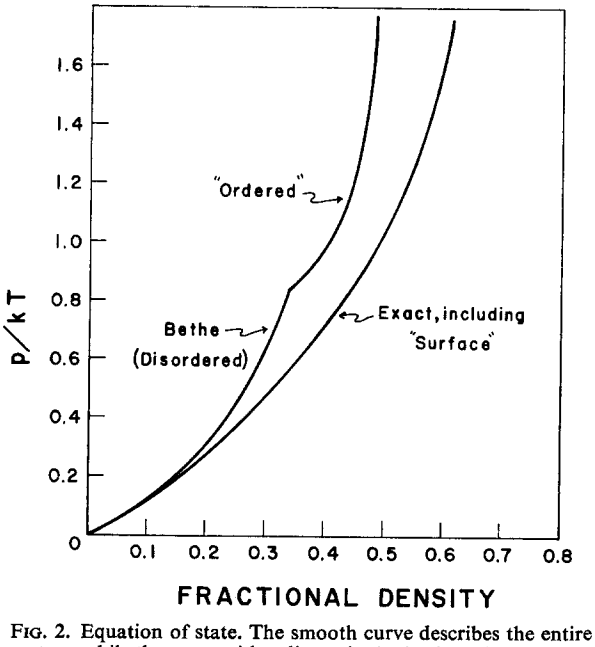


FIG. 2. Equation of state. The smooth curve describes the entire large tree, while the curve with a discontinuity in slope describes the order–disorder transition occurring at the interior sites far from the surface. The quasi-chemical or Bethe "approximation" is exact for the interior sites and densities less than one-third.

The customary procedure in statistical thermodynamics is to pass to the limit $N, V \to \infty$, $N/V = \rho$ remaining finite; clearly this corresponds to the limit $k \to \infty$ and involves the question of the convergence of the various infinite series. It is apparent from Eq. (5) that the convergence properties are determined by the behavior of the terms $c_j(z)$ for large j. These questions are discussed in the Appendix; we here summarize the important results.

The limiting behavior of $c_j(z)$ depends on z. For $z < 4$ the sequence converges to a well-behaved analytic function $c(z)$, while for $z > 4$ the odd terms converge to one function, $c_{2j+1}(z) \to c_+(z)$, and even terms converge to another, $c_{2j}(z) \to c_-(z)$. The three curves meet at $z = 4$, the only candidate for a singularity or phase transition. It is in fact true that at $z = 4$, the various derivatives of $c_j(z)$, become unbounded with increasing j, but not rapidly enough to spoil the convergence of Eqs. (5), (6), and higher derivatives, due to the presence of the factor 2^{-j}. The resulting smooth behavior attributable to this factor is easily understood: most of the sites are near the surface, where they are essentially independent; the highly correlated sites are deep in the interior and there are not many of them.

Another approach to the question of phase transitions, which we only mention, is that of Yang and Lee.[17] Phase transitions may occur, if zeros of the grand partition function $\Xi_k(z)$ close in on the positive real axis (as $k \to \infty$). It appears that for the present problems, zeros do close in on $z = 4$ but with vanishing density, allowing the system to "sneak" through on the real axis with no transition; p/kT is, however, manifestly a nonanalytic function of complex z at $z = 4$, which is probably an essential singularity.

IV. THERMODYNAMIC BEHAVIOR FAR FROM THE SURFACE

In the preceding section, we have discussed the thermodynamics of the entire (large) Cayley tree and found that the behavior is dominated by the large fraction of essentially independent sites near the surface. Another interpretation of this system, possibly more relevant, is a focusing of attention on the local behavior of interior sites far from the surface.

The key to this approach is Eq. (11), which gives

[17] C. N. Yang and T. D. Lee, Phys. Rev. 87, 410 (1952).

the fractional density of the jth generation in terms of the sequence b_j; since the b_j's are given in terms of the c_j's which, in turn, are defined by Eq. (3), in effect, Eq. (11) gives the density of the jth generation in terms of the activity z. An integration then gives the local pressure at generation j in terms of the activity, using the first half of Eq. (6). If the activity can be eliminated from these two expressions, we will have the local equation of state.

The limiting local equation of state would arise in the limit $j \to \infty$. But then, as already discussed, the behavior of the c_j's is very simple; depending on the magnitude of z, either c_j is actually independent of j and given by Eq.(A3), or the c_j's alternate between the two values c_+ and c_- given by Eq.(A4). In either case, the limiting value of b_j is then given by Eq. (8). In the low-density case, where the b_j's are constant, it is readily seen from Eq. (11) that the local density is given by

$$r = \frac{b}{1+b} = \frac{c(z)-1}{2c(z)-1}, \tag{13}$$

where $c^3(z) - c^2(z) = z$. It follows from thermodynamics that the pressure is given by

$$\begin{aligned} p/kT &= \int_0^z z^{-1}\rho(z)\,dz = \int_1^c \frac{3c-2}{c(2c-1)}\,dc \\ &= 2\ln c - \tfrac{1}{2}\ln(2c-1) \\ &= 2\ln(1-\rho) - \tfrac{3}{2}\ln(1-2\rho). \end{aligned} \tag{14}$$

This equation of state, Eq. (14), is precisely the equation of state predicted by the quasi-chemical, or Bethe, approximation. Hence, we see the sense in which the quasi-chemical approximation is exact for pseudo-lattices of the tree variety: the "approximation" is exact for the region far removed from the surface and for low densities, that is, for activities less than 4 or densities less than $\frac{1}{3}$ (for coordination number three), according to Eq. (13). The quasi-chemical equation of state is not exact for higher densities, nor is it exact for any nonzero density for the complete Cayley tree, as shown in Fig. 2.

For activities greater than four, there are two local densities, ρ_+ and ρ_- which are readily found from Eq. (11) to be given by

$$\rho_+ = \frac{b_+(1-b_-)}{1-b_+b_-}, \quad \rho_- = \frac{b_-(1-b_+)}{1-b_+b_-}, \tag{15}$$

where $b_+ = 1 - c_+^{-1}$ and $b_- = 1 - c_-^{-1}$.

In the limit of large activity, b_+ approaches 1 while b_- approaches 0. According to Eq. (15), the local density ρ_+ and ρ_- approach these same limiting values for large activities. This reflects the most efficient, ordered method of filling up the tree at high activities, namely, a filling of alternate layers beginning at the surface, generation 1. To obtain the equation of state in the high activity region, we are faced with the problem of deciding the appropriate meaning of the average density. The entire tree contains $\frac{2}{3}$ of its sites on odd generations; this suggests that the appropriate average density should be $\frac{2}{3}\rho_+ + \frac{1}{3}\rho_-$; on the other hand, beginning with any even generation, the remainder of the tree inward contains $\frac{2}{3}$ of its sites on even generations, leading to an average density expression $\frac{1}{3}\rho_+ + \frac{2}{3}\rho_-$. Viewed strictly from the interior, it is not at all obvious which of these expressions should be used. It seems that the most acceptable choice would be the symmetrical definition $\rho_{av} = \frac{1}{2}(\rho_+ + \rho_-)$ which we adopt. Using Eqs.(A4) and (15), we easily obtain

$$\rho_{av} = \frac{z-2}{2(z-1)}, \quad z > 4. \tag{16}$$

Integrating as before to obtain the local pressure, we find

$$\begin{aligned} p/kT - (p/kT)_{z=4} &= \int_4^z \frac{z-2}{2z(z-1)}\,dz \\ &= \ln z/4 - \tfrac{1}{2}\ln(z-1)/3. \end{aligned} \tag{17}$$

Now, the pressure must be continuous; consequently, using Eq. (14) to evaluate the pressure at $z = 4$, or $\rho = \frac{1}{3}$, we arrive at the high-density equation of state:

$$\begin{aligned} p/kT &= \ln z - \tfrac{1}{2}\ln(z-1) \\ &= \ln(1-\rho_{av}) - \tfrac{1}{2}\ln(1-2\rho_{av}) + \ln 2. \end{aligned} \tag{18}$$

The complete "interior" equation of state is shown in Fig. 2 along with the exact equation of state for the entire Cayley tree.

Thus, it is seen that while the entire Cayley tree behaves continuously, the interior region far from the surface shows a transition at activity $z = 4$. The transition for this interior region is one of the order–disorder type. Clearly, it is not a first-order transition; it can in fact easily be seen from Eqs. (14) and (18) that the transition is a second-order one of the Ehrenfest type, with a finite jump in the compressibility.

V. DISCUSSION

We have discussed the thermodynamic behavior of hard molecules absorbed on a pseudo-lattice of the Cayley tree variety, specifically, a homogeneous Cayley tree of coordination number three. We have found that the quasi-chemical or Bethe equation of state is not exact for the entire tree. There is, however,

a sense in which this equation of state is exact, namely, for low activities and regions far removed from the surface. For activities z greater than four, the disordered Bethe solution is no longer stable and is supplanted by an ordered solution. The "transition" in the interior regions is second order with a finite discontinuity in compressibility, occurring at a density of one third.

The entire tree has been seen to behave continuously, with no transition of lower than third order, and very likely, none at all. This is understandable, when it is realized that the overwhelming majority of the sites of the tree are near the surface, and consequently, essentially independent. Under such conditions, the cooperative behavior necessary for phase transitions is not possible.

APPENDIX. CONVERGENCE QUESTIONS AND THE ABSENCE OF A PHASE TRANSITION

We consider here the question of the existence and continuity of p/kT and its derivatives as functions of activity z. We are inquiring into the possibility of the existence of a phase transition for the complete tree, adopting the criterion[18] that a transition of order n at activity z_0 is characterized by a singularity in $d^n(p/kT)/dz^n$ at z_0 (while all lower derivatives are continuous at z_0). To define completely the mathematical problem, we repeat Eqs. (3) and (5) in the form

$$\varphi(z) = \sum_j 2^{-j} \ln c_j(z), \qquad \text{(A1)}$$

$$c_j(z) = 1 + z/[c_{j-1}(z)]^2, \quad c_0(z) = 1. \qquad \text{(A2)}$$

Much of the work rests on the following three-part *Lemma*, which is easily proved by mathematical induction: (a) $1 \leq c_j \leq 1 + z$; (b) $c_{2j} \leq c_{2j+2}$; (c) $c_{2j+1} \geq c_{2j+3}$. Now, since by (a) there is a uniform bound for the continuous functions $c_j(z)$ on any bounded part of the positive z axis, the series (A1) necessarily converges to a continuous function $\varphi(z)$. (There is thus no zeroth-order transition!)

The even and odd sequence of c_j's are each bounded and monotone, and so each must converge: $c_{2j}(z) \to c_-(z)$, $c_{2j+1}(z) \to c_+(z)$. The two functions c_+ and c_- may be identical (sequence convergence) or distinct (only subsequence convergence). If identical, clearly the unique limit $c(z)$ must satisfy $c(z) = 1 + z/\{c(z)\}^2$, or

$$c^3(z) - c^2(z) = z. \qquad \text{(A3)}$$

The distinct limit functions $c_\pm$ must satisfy $c_\pm(z) = 1 + z/\{1 + z/[c_\pm(z)]^2\}^2$, which is a fifth-order equation that factors into the now extraneous solution (A3) and $c_\pm^2 - zc_\pm + z = 0$, or

$$c_\pm = [z \pm (z^2 - 4z)^{\frac{1}{2}}]/2, \quad z \geq 4. \qquad \text{(A4)}$$

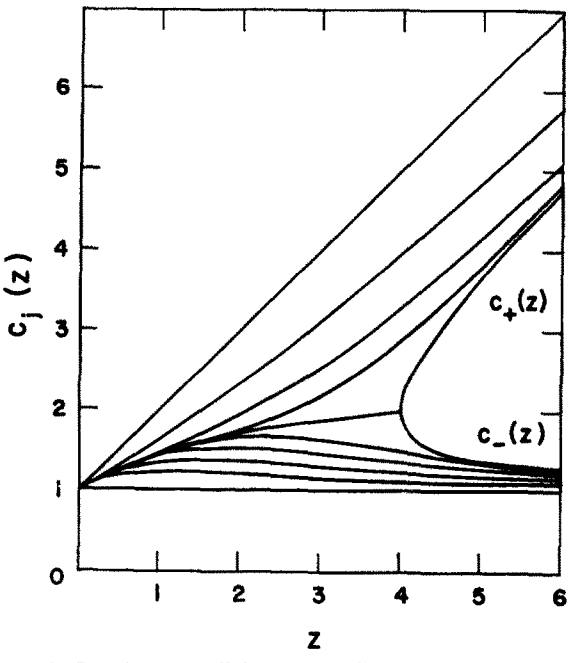


FIG. 3. Correlation coefficients $c_j(z)$. The upper curves are for $j = 1, 3, 7, 15$, and the lower curves are for $j = 0, 2, 4, 8, 16$. Also shown are the limiting curves $c(z)$ and $c_\pm(z)$.

The three curves meet at $z = 4$, each there taking the value 2. Since $c_\pm(z)$ is complex for $z < 4$, while $c_j(z)$ in necessarily real, the unique limit $c(z)$ must obtain for $z < 4$. But it can be shown from Eq. (A2) that this unique limit is not stable for $z > 4$: c_{j+1} is farther from 2 than is c_j. Hence, the alternating limits obtain for $z > 4$. These limiting functions and a few early c_j's are shown in Fig. 3.

It is evident that for any compact (closed and bounded) subset of the positive z axis, *not including* $z = 4$, the infinitely differentiable (rational) functions $c_j(z)$ converge (sequence-wise or subsequence-wise) uniformly to the infinitely differentiable limit function(s). The same statement applies to the sequence $\{\ln c_j(z)\}$, since the c_j's are uniformly bounded away from 0. Due to the presence of the strong convergence factor 2^{-j} in Eq. (A1), $\varphi(z)$ itself is infinitely differentiable everywhere, except possibly at $z = 4$. If there is a phase transition, it must occur at this point.

It is apparent from Fig. 3 that the convergence of $\{c_j(z)\}$ is not uniform in the vicinity of $z = 4$ and that the derivative functions $c_j^{(n)}(4)$ must become unbounded as $j \to \infty$. The crucial question is, how

[18] T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), p. 249.

rapidly (compared to 2^j) they become large, for this determines the differentiability of $\varphi(z)$. We first show that the series given by Eq. (6) converges to a continuous function, even at $z = 4$ (so no first-order transition), and then that the series obtained by formally differentiating (6) also converges to a continuous function $\rho'(z)$ (so no second-order transition).

Now, on any compact set $\mathcal{R}$ containing $z = 4$, the c_j's are bounded above and so the functions $b_j(z)$ are uniformly bounded away from 1:

$$b_j(z) \leq \beta < 1. \tag{A5}$$

(For instance, $\beta = \frac{5}{6}$ works for $\mathcal{R} = [3, 5]$.) The formula (9) is easily established by induction and leads to a bound T_j on $|t_j|$:

$$|t_j| \leq \sum_{i=0}^{j-1} 2^i \beta^{i+1} = \beta \frac{\alpha^j - 1}{\alpha - 1}$$
$$\leq \frac{\alpha^j}{\alpha - 1} = T_j, \tag{A6}$$

where $\alpha = 2\beta < 2$. [We can take $\alpha > 1$, so that $T_j \geq |t_k|$ for $k \leq j$.] The series (6) of continuous functions thus converges absolutely and uniformly on $\mathcal{R}$—necessarily to the continuous function $\rho(z)$, by the comparison test. The absolute convergence also justifies the rearrangement to the alternate form (10).

To investigate the possibility of a second-order transition, we study the continuity of $\partial\rho/\partial z$ near $z = 4$, which is obtained formally by differentiating $\sum 2^{-j} t_j$ in Eq. (6). The derivative $\partial\rho/\partial z$ exists as a continuous function if the series of continuous functions $\sum 2^{-j} \partial t_j/\partial z$ converges uniformly on $\mathcal{R} =$ [3, 5]. Using (8) and the definition $t_j = z\partial \ln c_j/\partial z$, we have $b' = \partial b_j/\partial z = t_j/zc_j$ so that

$$B_j = T_j/3 \geq |t_j/zc_j| = |b_j'| \tag{A7}$$

is a bound on $|b_k'|$ for $k \leq j$, uniform on $\mathcal{R}$. Then, from (9), a bound $T_j^{(1)}$ on $|t_j'|$ is given by

$$|t_j'| \leq \sum_{i=0}^{j-1} 2^i (i+1) \beta^i B_j$$
$$= \frac{\alpha^j[1 - (j+1)\alpha^j + j\alpha^{j+1}]}{3(\alpha - 1)^3}$$
$$= T_j^{(1)}. \tag{A8}$$

Since $\alpha > 1$, for sufficiently large j, say $j \geq J_0$,

$$T_j^{(1)} \leq Kj\alpha^{2j}. \tag{A9}$$

Also since $c_j(4) \to 2$, $b_j(4) \to \frac{1}{2}$, we can for sufficiently large j, $j \geq J_1$, use for β (the bound on b_j) anything greater than $\frac{1}{2}$, say $\beta = \frac{5}{8}$. But then $\alpha = \frac{5}{4}$ and $\omega = \alpha^2 < 2$, so that for $j \geq \max(J_0, J_1)$, we have $|t_j'| \leq Kj\omega^j$. The series $\sum 2^{-j} t_j'$ then converges absolutely and uniformly on $\mathcal{R}$. The limit must then be $\partial\rho/\partial z$ and continuous throughout $\mathcal{R}$. There is thus no second-order transition.

The process could undoubtedly be continued to establish the absence of higher-order transitions. For instance, bounds on $|b_j''|$ could be obtained from $b_j' = t_j/zc_j$ and bounds previously obtained for $|t_j'|$ and $c' = tc_{jj}/z$. Using (9) again would give a bound on $|t_j''|$, and so on. The conclusion would very likely be the absence of a transition of any order, but the required tedium for the general demonstration seems unwarranted.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Electro-Optical Effects. I

M. M. CARROLL
Division of Applied Mechanics, University of California, Berkeley, California
AND
R. S. RIVLIN
Center for the Application of Mathematics, Lehigh University, Bethlehem, Pennsylvania

(Received 14 December 1966)

The theory of Toupin and Rivlin for the propagation of electromagnetic waves in a centro-symmetric isotropic material to which static electric and magnetic fields are applied is specialized to the case when only a static electric field is applied. The reflection–refraction problem for an electromagnetic wave normally incident on the interface between free space and the material is studied for arbitrary direction of the applied electric field. It is found that, in general, there are two refracted rays and that these are not generally normal to the interface.

1. INTRODUCTION

In a previous paper, Toupin and Rivlin[1] have developed a theory for the propagation of plane electromagnetic waves of small amplitude in an initially holohedral isotropic medium (i.e., an isotropic material possessing a center of symmetry) of infinite extent, to which static electric and magnetic fields are applied. The theory is easily specialized to the case when only a static electric field is present and the resulting theory is appropriate for the discussion of electro-optical effects in holohedral isotropic materials.

In this theory, which is linear in the time-dependent electromagnetic field, the constitutive equations involve six constitutive functions of the resultant static field and of the frequency of the electromagnetic wave. The wave velocities are given in terms of these. In the previous paper,[1] this was done explicitly only in the cases when the direction of propagation of the wave (i.e., the direction of the normal to the wavefront) is parallel or perpendicular to the direction of the static field. In Sec. 3 the corresponding result is given for arbitrary inclination of the direction of propagation to the static field. It is seen that there are, in general, two wave velocities corresponding to an arbitrary direction of propagation and that reversal of the direction of propagation changes the wave velocities unless a relation of the Onsager type is satisfied. The nature of the polarization of the electric vectors for each of these waves is discussed in Sec. 4.

In Secs. 5 and 6, we discuss the reflection and refraction of a plane electromagnetic wave incident normally on the surface of an infinite half-space to which a static electric field is applied. The reflected wave is, of course, normal to the surface and it is found that there are, in general, two transmitted waves for which the angles of refraction are nonzero if the tangential component of the static field is nonzero. The ray directions for these lie in the plane of the normal to the interface and the direction of the static electric field. Explicit expressions are obtained for the angles of refraction in terms of the six constitutive functions.

2. CONSTITUTIVE EQUATIONS

The complex electric, magnetic induction, electric displacement, and magnetic intensity fields for an infinite plane electromagnetic wave are given by

$$(\mathbf{E}, \mathbf{B}, \mathbf{D}, \mathbf{H}) = (\mathbf{e}, \mathbf{b}, \mathbf{d}, \mathbf{h})e^{\iota(k\mathbf{n}\cdot\mathbf{x}-\omega t)}, \tag{2.1}$$

where the complex vectors $\mathbf{e}$, $\mathbf{b}$, $\mathbf{d}$, and $\mathbf{h}$ are independent of position and time. ω is the angular frequency of the wave, k is the wavenumber, $\mathbf{n}$ is a unit vector in the direction of propagation, and $\mathbf{x}$ is the position vector of a generic point of space.

We consider the propagation of such a wave in an infinite homogeneous medium to which a strong uniform static electric field $\mathfrak{E}$ is applied. Toupin and Rivlin[1] have shown that, if the material is holohedral isotropic in the absence of any applied field, then the constitutive equations for the wave have the form

$$\begin{aligned} \mathbf{d} &= \boldsymbol{\Phi}\cdot\mathbf{e} + \boldsymbol{\Psi}\cdot\mathbf{b}, \\ \mathbf{h} &= \boldsymbol{\Omega}\cdot\mathbf{e} + \boldsymbol{\Lambda}\cdot\mathbf{b}, \end{aligned} \tag{2.2}$$

where $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Lambda}$ are 3×3 matrices defined by $\boldsymbol{\Phi} = \|\Phi_{ij}\|, \boldsymbol{\Psi} = \|\Psi_{ij}\|, \cdots$, and

$$\begin{aligned} \Phi_{ij} &= \alpha_1\delta_{ij} + \alpha_7\mathfrak{E}_i\mathfrak{E}_j, \quad \Psi_{ij} = -\alpha_3\epsilon_{ijk}\mathfrak{E}_k, \\ \Lambda_{ij} &= \beta_1\delta_{ij} + \beta_7\mathfrak{E}_i\mathfrak{E}_j, \quad \Omega_{ij} = -\beta_2\epsilon_{ijk}\mathfrak{E}_k. \end{aligned} \tag{2.3}$$

Here $\mathfrak{E}_i$ $(i = 1, 2, 3)$ denote the components of $\mathfrak{E}$ in a rectangular Cartesian coordinate system x and the α's and β's are real functions of $\iota\omega$ and $\mathfrak{E}_m\mathfrak{E}_{m'}$ [i.e., the complex conjugates of $\alpha(\iota\omega)$ and $\beta(\iota\omega)$ are $\alpha(-\iota\omega)$ and $\beta(-\iota\omega)$, respectively].[2]

[1] R. A. Toupin and R. S. Rivlin, Arch. Ratl. Mech. Anal. 7, 434 (1961).

[2] The original result had α's and β's as polynomials in $\mathfrak{E}_m\mathfrak{E}_m$. The above extension follows from a theorem of A. S. Wineman and A. C. Pipkin, Arch. Ratl. Mech. Anal. 17, 184 (1964). The notation of the paper by Toupin and Rivlin (Ref. 1) is preserved.

3. THE SECULAR EQUATION

The electromagnetic field equations in the absence of free currents and charges are

$$\begin{aligned} \operatorname{curl}\mathbf{E} + (\partial\mathbf{B}/\partial t) = 0, \quad \operatorname{div}\mathbf{B} = 0, \\ \operatorname{curl}\mathbf{H} - (\partial\mathbf{D}/\partial t) = 0, \quad \operatorname{div}\mathbf{D} = 0. \end{aligned} \tag{3.1}$$

For the infinite plane wave (2.1) we obtain

$$k\mathbf{n}\times\mathbf{e} - \omega\mathbf{b} = 0, \quad k\mathbf{n}\times\mathbf{h} + \omega\mathbf{d} = 0. \tag{3.2}$$

Equations (3.2) together with the constitutive equations (2.2) lead to a system of three homogeneous linear equations in the components of $\mathbf{e}$,

$$\chi_{ij}e_j = 0, \tag{3.3}$$

where

$$\chi_{ij} = \Phi_{ij} + \eta(\epsilon_{jrs}\Psi_{ir}n_s + \epsilon_{ipq}n_p\Omega_{qj}) + \eta^2\epsilon_{ipq}\epsilon_{jrs}n_p\Lambda_{qr}n_s \tag{3.4}$$

and we have used the notation $\eta = k/\omega$. ($1/\eta$ is then the complex velocity.) The secular equation for η is

$$|\chi_{ij}| = 0. \tag{3.5}$$

We choose our rectangular Cartesian coordinate system x such that $n_i = \delta_{i3}$. We then obtain from (3.4)

$$\|\chi_{ij}\| = \begin{Vmatrix} \Phi_{11} + \eta(\Psi_{12} - \Omega_{21}) - \eta^2\Lambda_{22}, & \Phi_{12} - \eta(\Psi_{11} + \Omega_{22}) + \eta^2\Lambda_{21}, & \Phi_{13} - \eta\Omega_{23} \\ \Phi_{21} + \eta(\Psi_{22} + \Omega_{11}) + \eta^2\Lambda_{12}, & \Phi_{22} - \eta(\Psi_{21} - \Omega_{12}) - \eta^2\Lambda_{11}, & \Phi_{23} + \eta\Omega_{13} \\ \Phi_{31} + \eta\Psi_{32}, & \Phi_{32} - \eta\Psi_{31}, & \Phi_{33} \end{Vmatrix}. \tag{3.6}$$

Without loss in generality we may take

$$\mathfrak{E} = (\mathfrak{E}_1, 0, \mathfrak{E}_3), \quad \text{i.e.,} \quad \mathfrak{E}_i = \mathfrak{E}_1\delta_{1i} + \mathfrak{E}_3\delta_{3i}. \tag{3.7}$$

Equations (2.3) then give

$$\begin{aligned} \Phi_{ij} &= \alpha_1\delta_{ij} + \alpha_7\{\mathfrak{E}_1^2\delta_{i1}\delta_{j1} + \mathfrak{E}_1\mathfrak{E}_3(\delta_{i1}\delta_{j3} + \delta_{i3}\delta_{j1}) + \mathfrak{E}_3^2\delta_{i3}\delta_{j3}\}, \\ \Lambda_{ij} &= \beta_1\delta_{ij} + \beta_7\{\mathfrak{E}_1^2\delta_{i1}\delta_{j1} + \mathfrak{E}_1\mathfrak{E}_3(\delta_{i1}\delta_{j3} + \delta_{i3}\delta_{j1}) + \mathfrak{E}_3^2\delta_{i3}\delta_{j3}\}, \\ \Psi_{ij} &= -\alpha_3(\epsilon_{ij1}\mathfrak{E}_1 + \epsilon_{ij3}\mathfrak{E}_3), \\ \Omega_{ij} &= -\beta_2(\epsilon_{ij1}\mathfrak{E}_1 + \epsilon_{ij3}\mathfrak{E}_3), \end{aligned} \tag{3.8}$$

and substituting from Eqs. (3.8) in Eq. (3.6), we obtain

$$\|\chi_{ij}\| = \begin{Vmatrix} \alpha_1 + \alpha_7\mathfrak{E}_1^2 - (\alpha_3 + \beta_2)\mathfrak{E}_3\eta - \beta_1\eta^2, & 0, & (\alpha_7\mathfrak{E}_3 + \beta_2\eta)\mathfrak{E}_1 \\ 0, & \alpha_1 - (\alpha_3 + \beta_2)\mathfrak{E}_3\eta - (\beta_1 + \beta_7\mathfrak{E}_1^2)\eta^2, & 0 \\ (\alpha_7\mathfrak{E}_3 + \alpha_3\eta)\mathfrak{E}_1, & 0, & \alpha_1 + \alpha_7\mathfrak{E}_3^2 \end{Vmatrix}. \tag{3.9}$$

Thus, the secular equation (3.5) becomes

$$(\beta_1 + \beta_7\mathfrak{E}_1^2)\eta^2 + (\alpha_3 + \beta_2)\mathfrak{E}_3\eta - \alpha_1 = 0, \tag{3.10}$$

or

$$(\alpha_1\beta_1 + \alpha_7\beta_1\mathfrak{E}_3^2 + \alpha_3\beta_2\mathfrak{E}_1^2)\eta^2 + \{\alpha_1 + \alpha_7(\mathfrak{E}_1^2 + \mathfrak{E}_3^2)\} \times \{(\alpha_3 + \beta_2)\mathfrak{E}_3\eta - \alpha_1\} = 0. \tag{3.11}$$

From (3.10) and (3.11) we see that there are, in general, four possible values,

$$\eta_1, \eta_2 = \tfrac{1}{2}(\beta_1 + \beta_7\mathfrak{E}_1^2)^{-1}[-(\alpha_3 + \beta_2)\mathfrak{E}_3 \pm \{(\alpha_3 + \beta_2)^2\mathfrak{E}_3^2 + 4\alpha_1(\beta_1 + \beta_7\mathfrak{E}_1^2)\}^{\frac{1}{2}}] \tag{3.12}$$

and

$$\eta_3, \eta_4 = (1/2A)[-(\alpha_3 + \beta_2)\mathfrak{E}_3 \pm \{(\alpha_3 + \beta_2)^2\mathfrak{E}_3^2 + 4\alpha_1 A\}^{\frac{1}{2}}], \tag{3.13}$$

where

$$A = (\alpha_1\beta_1 + \alpha_7\beta_1\mathfrak{E}_3^2 + \alpha_3\beta_2\mathfrak{E}_1^2)/[\alpha_1 + \alpha_7(\mathfrak{E}_1^2 + \mathfrak{E}_3^2)], \tag{3.14}$$

for the inverse complex velocity η of a wave propagating in any direction $\mathbf{n}$. In general these four values are complex. We may expect that η_1 and η_2 correspond to two waves traveling in opposite directions (i.e., that the real parts of η_1 and η_2 are of opposite sign). We see that these waves will have different complex velocities unless

$$(\alpha_3 + \beta_2)\mathfrak{E}_3 = 0. \tag{3.15}$$

Similarly, we may expect that η_3 and η_4 correspond to two waves traveling in opposite directions and these waves also have different complex velocities unless (3.15) holds. We notice that (3.15) is satisfied when $\mathfrak{E}_3 = 0$, i.e., when the static electric field $\mathfrak{E}$ is perpendicular to the propagation direction $\mathbf{n}$ or when a condition of the Onsager type,

$$\alpha_3 + \beta_2 = 0, \tag{3.16}$$

is satisfied.

If we set $\mathfrak{E}_1 = 0$ or $\mathfrak{E}_3 = 0$ in (3.12) and (3.13), we recover the results (10.6), (10.10), and (10.11) of Toupin and Rivlin.[1]

4. POLARIZATION

In this section we consider the form of the waves corresponding to the four values (3.12) and (3.13) of

the inverse complex velocity η. From (3.3) and (3.9) we see that, if

$$\{\alpha_7\beta_7(\mathcal{E}_1^2 + \mathcal{E}_3^2) - \alpha_3\beta_2 + \alpha_7\beta_1 + \alpha_1\beta_7\}\mathcal{E}_1^2 \neq 0, \quad (4.1)$$

the waves corresponding to values of η given by (3.12) have $e_1 = e_3 = 0$. These are transverse waves, linearly polarized in the x_2 direction, i.e., perpendicular to the plane containing $\mathcal{E}$ and $\mathbf{n}$. Also from (3.3) and (3.9) we see that, if the inequality (4.1) holds, then for waves corresponding to values of η given by (3.13), $\mathbf{e}$ has the form

$$\mathbf{e} = (e_1, 0, e_3), \quad (4.2)$$

where

$$e_3/e_1 = -(\alpha_7\mathcal{E}_3 + \alpha_3\eta)\mathcal{E}_1/(\alpha_1 + \alpha_7\mathcal{E}_3^2) = \zeta. \quad (4.3)$$

The corresponding $\mathbf{E}^+$ waves[3] are circularly polarized (see Ref. 1) if and only if $\zeta = \pm\iota$, linearly polarized if and only if ζ is real, and transverse if and only if $\zeta = 0$. For all other values of ζ corresponding to (3.13) the $\mathbf{E}^+$ wave is a skew, elliptically polarized wave. The locus of $\mathbf{E}^+$ at a fixed point in the material is an ellipse in the x_1x_3 plane, i.e., in the plane containing $\mathcal{E}$ and $\mathbf{n}$. The semiaxes L_1 and L_2 of the ellipse are given by[4]

$$L_1^2, L_2^2 = \tfrac{1}{2}e_1e_1^*[1 + \zeta\zeta^* \pm \{(1 + \zeta^2)(1 + \zeta^{*2})\}^{\frac{1}{2}}], \quad (4.4)$$

and $\mathbf{E}^+$ is directed along an axis when

$$\tan 2\phi = \frac{\iota\{e_1^2(1 + \zeta^2) - e_1^{*2}(1 + \zeta^{*2})\}}{e_1^2(1 + \zeta^2) + e_1^{*2}(1 + \zeta^{*2})},$$
$$\phi = k\mathbf{n}\cdot\mathbf{x} - \omega t. \quad (4.5)$$

If the inequality (4.1) does not hold, then the secular equation (3.5) has a repeated root and the corresponding $\mathbf{e}$ vector may take an arbitrary direction in a plane. For example, when $\mathcal{E}_1 = 0$, so that the static field $\mathcal{E}$ is parallel to the propagation direction $\mathbf{n}$, it is easily seen that all waves are linearly polarized in an arbitrary direction normal to the direction of propagation. (cf. Ref. 1.)

5. THE ELECTRIFIED HALF SPACE

We now consider the reflection and transmission of an infinite plane electromagnetic wave incident normally on the interface $x_3 = 0$ (in a rectangular Cartesian coordinate system x) of a half space of holohedral isotropic material occupying the region

$$x_3 > 0. \quad (5.1)$$

The incident wave is assumed to travel in the positive direction of the x_3 axis.

We suppose that a strong static electric field $\mathcal{E}$ is applied and we may, without loss of generality, choose the reference system x so that

$$\begin{aligned}\mathcal{E} &= (\mathcal{E}_1, 0, \mathcal{E}_3) \quad x_3 > 0,\\ &= (\mathcal{E}_1, 0, \bar{\mathcal{E}}_3) \quad x_3 < 0.\end{aligned} \quad (5.2)$$

The constitutive equations for the region $x_3 > 0$ are then given by (2.2) and (3.8) and those for the region $x_3 < 0$ (free space) are

$$\mathbf{d} = \mathbf{e}, \quad \mathbf{h} = \mathbf{b}. \quad (5.3)$$

From the results of Secs. 3 and 4, we see that the complex electric field $\mathbf{E}$ for the wave is given by expressions of the form

$$\begin{aligned}\mathbf{E} &= (e_1, e_2, 0)e^{\iota\omega(x_3 - t)}\\ &\quad + (\bar{e}_1, \bar{e}_2, 0)e^{-\iota\omega(x_3 + t)}, \qquad (x_3 < 0)\\ \mathbf{E} &= (0, {}^{(1)}e_2, 0)e^{\iota\omega(\eta_1x_3 - t)}\\ &\quad + (0, {}^{(2)}e_2, 0)e^{\iota\omega(\eta_2x_3 - t)}\\ &\quad + ({}^{(3)}e_1, 0, {}^{(3)}e_3)e^{\iota\omega(\eta_3x_3 - t)}\\ &\quad + ({}^{(4)}e_1, 0, {}^{(4)}e_3)e^{\iota\omega(\eta_4x_3 - t)}, \quad (x_3 > 0),\end{aligned} \quad (5.4)$$

where η_p $(p = 1, 2, 3, 4)$ are given by (3.12) and (3.13) and [cf. Eq. (4.3)]

$${}^{(p)}e_3/{}^{(p)}e_1 = \zeta_p = -(\alpha_7\mathcal{E}_3 + \alpha_3\eta_p)\mathcal{E}_1/(\alpha_1 + \alpha_7\mathcal{E}_3^2) \quad (p = 3, 4). \quad (5.5)$$

Presumably η_1 and η_2 correspond to waves traveling in opposite directions and similarly η_3 and η_4 correspond to waves traveling in opposite directions. We assume that

$$\eta_1^+ > 0, \quad \eta_2^+ < 0, \quad \eta_3^+ > 0, \quad \eta_4^+ < 0. \quad (5.6)$$

Then, if the material is stable, we have also

$$\eta_1^- > 0, \quad \eta_2^- < 0, \quad \eta_3^- > 0, \quad \eta_4^- < 0. \quad (5.7)$$

From the conditions $\eta_2^- < 0$ and $\eta_4^- < 0$ in (5.7), and the fact that when $x_3 = \infty$ the amplitude of the waves must be zero, it follows that ${}^{(2)}e_i = {}^{(4)}e_i = 0$. Equations (5.4) then become

$$\mathbf{E} = \begin{cases}(e_1, e_2, 0)e^{\iota\omega(x_3 - t)} + (\bar{e}_1, \bar{e}_2, 0)e^{-\iota\omega(x_3 + t)}, & (x_3 < 0) \quad (5.8a)\\ (0, {}^{(1)}e_2, 0)e^{\iota\omega(\eta_1x_3 - t)} + ({}^{(3)}e_1, 0, {}^{(3)}e_3)e^{\iota\omega(\eta_3x_3 - t)}, & (x_3 > 0). \quad (5.8b)\end{cases}$$

We note that in (5.8a) the first term represents the incident wave and the second term the reflected wave. In (5.8b) the terms represent two transmitted waves.

From (5.5) and (5.8), together with the field equations (3.1) and the constitutive equations (2.2)

[3] We denote the real and imaginary parts of a complex quantity by using the superscripts + and −, respectively.

[4] A superscript asterisk is used to denote the complex conjugate.

and (5.3), we see that the magnetic intensity field **H** for the wave is given by

$$\mathbf{H} = \begin{cases} (-e_2, e_1, 0)e^{i\omega(x_3-t)} \\ \quad + (\bar{e}_2, -\bar{e}_1, 0)e^{-i\omega(x_3+t)}, \quad (x_3 < 0) \quad (5.9a) \\ ({}^{(1)}h_1, 0, {}^{(1)}h_3)e^{i\omega(\eta_1 x_3 - t)} \\ \quad + (0, {}^{(3)}h_2, 0)e^{i\omega(\eta_3 x_3 - t)}, \quad (x_3 > 0), \quad (5.9b) \end{cases}$$

where

$${}^{(1)}h_1 = -{}^{(1)}e_2\theta_1, \quad {}^{(1)}h_3 = {}^{(1)}e_2(\beta_2 - \beta_7\mathfrak{E}_3\eta_1)\mathfrak{E}_1, \quad {}^{(3)}h_2 = {}^{(3)}e_1\theta_3, \tag{5.10}$$

and the notations

$$\theta_1 = \beta_2\mathfrak{E}_3 + (\beta_1 + \beta_7\mathfrak{E}_1^2)\eta_1, \quad \theta_3 = \beta_2\mathfrak{E}_3 + \beta_1\eta_3 - \beta_2\mathfrak{E}_1\zeta_3 \tag{5.11}$$

are used.

The boundary conditions at the interface $x_3 = 0$, that the tangential components of **E** and **H** are continuous, give

$$e_1 + \bar{e}_1 = {}^{(3)}e_1, \quad e_2 + \bar{e}_2 = {}^{(1)}e_2, \quad e_1 - \bar{e}_2 = {}^{(1)}e_2\theta_1, \quad e_1 - \bar{e}_1 = {}^{(3)}e_1\theta_3. \tag{5.12}$$

Whence, with (5.5),

$${}^{(3)}e_1 = \frac{2e_1}{1+\theta_3}, \quad {}^{(1)}e_2 = \frac{2e_2}{1+\theta_1}, \quad {}^{(3)}e_3 = \zeta_3\,{}^{(3)}e_1, \quad \bar{e}_1 = \frac{1-\theta_3}{1+\theta_3}e_1, \quad \bar{e}_2 = \frac{1-\theta_1}{1+\theta_1}e_2. \tag{5.13}$$

The remaining boundary conditions, that the normal components of **D** and **B** are continuous at the interface, are automatically satisfied.

We note from (5.13) that if the incident wave is polarized with its electric field in the x_1 direction, then the reflected wave is similarly polarized and there is only one transmitted wave, and this is elliptically polarized in the x_1x_3 plane. On the other hand if the incident wave is polarized with its electric field in the x_2 direction, then the reflected wave is similarly polarized and there is only one transmitted wave, which is linearly polarized in the x_2 direction.

6. RAY DIRECTIONS FOR TRANSMITTED WAVES

The Poynting vector **S**, for a field in which the real electric and magnetic intensity vectors are $\mathfrak{E} + \mathbf{E}^+$ and $\mathbf{H}^+$, respectively, is given by

$$\mathbf{S} = (\mathfrak{E} + \mathbf{E}^+) \times \mathbf{H}^+. \tag{6.1}$$

We denote the Poynting vectors for the transmitted waves with slownesses η_1 and η_3 by $\mathbf{S}_1$ and $\mathbf{S}_3$, respectively. It follows from (5.8) and (5.9) that, for the first of these waves,

$$\mathbf{E}^+ = (0, {}^{(1)}e_2^+\cos\phi_1 - {}^{(1)}e_2^-\sin\phi_1, 0)e^{-\omega\eta_1^- x_3},$$
$$\mathbf{H}^+ = ({}^{(1)}h_1^+\cos\phi_1 - {}^{(1)}h_1^-\sin\phi_1, 0, {}^{(1)}h_3^+\cos\phi_1 - {}^{(1)}h_3^-\sin\phi_1)e^{-\omega\eta_1^- x_3}, \tag{6.2}$$

and, for the second of these waves,

$$\mathbf{E}^+ = ({}^{(3)}e_1^+\cos\phi_3 - {}^{(3)}e_1^-\sin\phi_3, 0, {}^{(3)}e_3^+\cos\phi_3 - {}^{(3)}e_3^-\sin\phi_3)e^{-\omega\eta_3^- x_3}, \tag{6.3}$$
$$\mathbf{H}^+ = (0, {}^{(3)}h_2^+\cos\phi_3 - {}^{(3)}h_2^-\sin\phi_3, 0)e^{-\omega\eta_3^- x_3},$$

where

$$\phi_1 = \omega(\eta_1^+ x_3 - t), \quad \phi_3 = \omega(\eta_3^+ x_3 - t). \tag{6.4}$$

The Poynting vector for each of the transmitted waves varies throughout each cycle. Let $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_3$ be the values of the Poynting vectors for the two transmitted waves averaged over a cycle, then

$$\hat{\mathbf{S}}_p = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} \mathbf{S}_p\,dt \quad (p = 1, 3). \tag{6.5}$$

For the wave given by (6.2) we have

$$\hat{\mathbf{S}}_1 = \tfrac{1}{2}({}^{(1)}e_2^+\,{}^{(1)}h_3^+ + {}^{(1)}e_2^-\,{}^{(1)}h_3^-, 0, -{}^{(1)}e_2^+\,{}^{(1)}h_1^+ - {}^{(1)}e_2^-\,{}^{(1)}h_1^-)e^{-2\omega\eta_1^- x_3}, \tag{6.6}$$

and for the wave given by (6.3) we have

$$\hat{\mathbf{S}}_3 = \tfrac{1}{2}(-{}^{(3)}e_3^+\,{}^{(3)}h_2^+ - {}^{(3)}e_3^-\,{}^{(3)}h_2^-, 0, {}^{(3)}e_1^+\,{}^{(3)}h_2^+ + {}^{(3)}e_1^-\,{}^{(3)}h_2^-)e^{-2\omega\eta_3^- x_3}. \tag{6.7}$$

The vectors lie in the ray directions for the transmitted waves. We note that they both lie in the x_1x_3 plane and are inclined at angles χ_1 and χ_3, respectively, to the x_3 direction (i.e., to the direction of the incident ray), where

$$\tan\chi_1 = -\frac{{}^{(1)}e_2^+\,{}^{(1)}h_3^+ + {}^{(1)}e_2^-\,{}^{(1)}h_3^-}{{}^{(1)}e_2^+\,{}^{(1)}h_1^+ + {}^{(1)}e_2^-\,{}^{(1)}h_1^-}, \quad \tan\chi_3 = -\frac{{}^{(3)}e_3^+\,{}^{(3)}h_2^+ + {}^{(3)}e_3^-\,{}^{(3)}h_2^-}{{}^{(3)}e_1^+\,{}^{(3)}h_2^+ + {}^{(3)}e_1^-\,{}^{(3)}h_2^-}. \tag{6.8}$$

Equations (6.8) may be rewritten as

$$\tan\chi_1 = -\frac{{}^{(1)}e_2^*\,{}^{(1)}h_3 + {}^{(1)}e_2\,{}^{(1)}h_3^*}{{}^{(1)}e_2^*\,{}^{(1)}h_1 + {}^{(1)}e_2\,{}^{(1)}h_1^*}, \quad \tan\chi_3 = -\frac{{}^{(3)}e_3^*\,{}^{(3)}h_2 + {}^{(3)}e_3\,{}^{(3)}h_2^*}{{}^{(3)}e_1^*\,{}^{(3)}h_2 + {}^{(3)}e_1\,{}^{(3)}h_2^*}. \tag{6.9}$$

Introducing (5.10) into (6.9), we obtain, with (5.13) and (5.11),

$$\tan\chi_1 = \frac{[(\beta_2 + \beta_2^*) - \mathfrak{E}_3(\beta_7\eta_1 + \beta_7^*\eta_1^*)]\mathfrak{E}_1}{(\beta_2 + \beta_2^*)\mathfrak{E}_3 + (\beta_1\eta_1 + \beta_1^*\eta_1^*) + (\beta_7\eta_1 + \beta_7^*\eta_1^*)\mathfrak{E}_1^2}, \quad \tan\chi_3 = -\frac{\zeta_3^*\theta_3 + \zeta_3\theta_3^*}{\theta_3 + \theta_3^*}, \tag{6.10}$$

where θ_3 and ζ_3 are given by Eqs. (5.5) and (5.11).

ACKNOWLEDGMENT

One of the authors (R. S. R.) acknowledges the support of a grant from the National Science Foundation to Brown University.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Formal Properties of the Solution to the Radiationless Relativistic Kepler Problem

ROBERT G. CAWLEY*
Clarkson College of Technology, Potsdam, New York†

(Received 24 January 1967)

Formal properties of the solution to the classical radiationless relativistic Kepler problem are discussed with particular attention to the strong coupling limit, a case rarely considered. It is found that fall to the origin can occur for potentials less strongly attractive than required nonrelativistically. One conclusion is that the relativistic centrifugal barrier is less "effective" than the nonrelativistic kind. It is possible to describe these effects in terms of a dimensionless parameter resembling a classical "coupling constant." The quantum-mechanical case is also briefly treated.

I. INTRODUCTION

THE special relativistic "Kepler" problem of a particle moving about an infinitely massive force center without radiation has been solved classically and quantum mechanically with both the Klein–Gordon and Dirac equations, and the solutions are well known.[1,2] A peculiar feature of the solutions to the quantum mechanical cases is their singular behavior when the coupling becomes very strong.[2] It is only for the classical case that the solution for very strongly attractive centers has been written out explicitly, however,[1] and even here little is said about what is happening physically.

In the classical Kepler case, the orbiting particle can fall to the center in a finite time under certain conditions, viz., if the attraction is sufficiently great and the angular momentum about $r = 0$ not too large. While this point is treated in Ref. 1, it does not seem to be generally known to most physicists. Of interest is the question of whether this is a unique feature of the $1/r$ potential or whether it is a peculiarly relativistic effect. Further, what is the relation, if any, to the singular behavior of the quantum-mechanical solutions in the strong coupling case? Our answers to these questions are that these phenomena find their explanation in the same way as the nonrelativistic effect of fall to the origin in a sufficiently strongly attractive $1/r^2$ (or worse) potential in both the classical and quantum-mechanical cases; namely, that the centrifugal barrier is overcome by the attractive center. Anything faster than $1/r^2$ causes collapse nonrelativistically, and relativistically we argue that anything faster than $1/r$ is sufficient.[3a] We conclude that nonrelativistic centrifugal barriers are more effective than their relativistic counterparts, the reason lying essentially in the fact that the latter variety are tied up under a square root in the Hamiltonian.

There are some amusing consequences of these "fall" effects and some of these are pointed out. We also take this opportunity to discuss some of the features of the solutions to the relativistic Kepler problem in particular when the coupling is not much less than one (as it is in most atoms and solar systems). We stress finally that the problem we consider is a *radiationless* problem and that it is a *single-particle problem*, i.e., the mass of the force center must be regarded as infinite—it is the source of an *external* field in other words. It may be noted that the omission of radiation effects is most serious in the strong coupling limit.[3b] Despite this fact, consideration of this problem is of interest because of the qualitative conclusion which we are able to draw, that relativistic centrifugal barriers are weaker than the usual kind. We are able to describe these effects in terms of a dimensionless parameter which resembles a coupling constant. Finally, it may be that the nature of the strong coupling solutions has some sort of a more realistic parallel in strong interactions, though we are unable to say anything definite.

In Sec. II the classical Kepler case is set up and solved in the canonical way, and Sec. III is devoted to a study of the solutions derived in Sec. II and a discussion of the fall to the center in the strong attraction case. Section IV concludes by considering the quantum-mechanical cases.

* Supported in part by the Division of Research, Clarkson College of Technology, Potsdam, New York.

† Present address: U.S. Naval Ordinance Laboratory, Silver Spring, Maryland.

[1] See, e.g., L. Landau and E. Lifshitz, *Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Cambridge, Massachusetts, 1951), pp. 100–102.

[2] See, for example, H. A. Bethe, *Intermediate Quantum Mechanics* (W. A. Benjamin, Inc., New York, 1964).

[3] (a) This statement requires a minor qualification which will be made farther along in the text (see the end of Sec. III). (b) Notice, however that unless the particle is the source of a field, there will, be no radiation at all.

II. SOLUTION OF THE CLASSICAL SPECIAL RELATIVISTIC KEPLER PROBLEM

The Hamiltonian for a particle of mass m moving about another particle of infinite mass is

$$H = (p^2 + m^2c^2)^{\frac{1}{2}}c + V(r). \tag{1}$$

In the Kepler case, $V(r) = -K^2/r$ (attractive).

Here we have formed H in a frame fixed with respect to the force center. Since the mass of the force center is infinite, it cannot be accelerated, so no inertial terms are necessary in Eq. (1). Were the mass of the force center finite, the internal and overall motions would be coupled in a complicated way by retardation effects and the problem would be far more complicated. Relativistically it is not possible to decouple center of mass and relative motions as it is in the nonrelativistic theory. We limit our considerations to the infinite mass case.

(It is for reasons such as these that it is incorrect to apply the Klein–Gordon equation to absorption models of high-energy p–p scattering, as has been done on a few occasions recently.)

In polar coordinates the Hamiltonian is

$$H = \left(p_r^2 + \frac{J^2}{r^2} + m^2c^2\right)^{\frac{1}{2}}c - \frac{K^2}{r}, \tag{2}$$

where p_r is the radial component of the momentum of m and J is its angular momentum about $r = 0$. The Hamilton–Jacobi equation is

$$-\frac{\partial s}{\partial t} = \left[\left(\frac{\partial s}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial s}{\partial \theta}\right)^2 + m^2c^2\right]^{\frac{1}{2}}c - \frac{K^2}{r}. \tag{3}$$

The solution to Eq. (3) is found by standard techniques to be

$$s(r,\theta,t;J,E) = -Et + J\theta + \int^r \left[\frac{1}{c^2}\left(E + \frac{K^2}{r}\right)^2 - \frac{J^2}{r^2} - m^2c^2\right]^{\frac{1}{2}} dr. \tag{4}$$

The orbit comes from

$$\frac{\partial s}{\partial J} = \text{const} \equiv \theta_0 = \theta - \int^r \frac{(J/r^2)\,dr}{\left[\frac{1}{c^2}\left(E + \frac{K^2}{r}\right)^2 - \frac{J^2}{r^2} - m^2c^2\right]^{\frac{1}{2}}}, \tag{5}$$

and the time dependence (the energy integral) from

$$\frac{\partial s}{\partial E} = \text{const} \equiv -t_0 = t + \int^r \frac{\frac{1}{c^2}\left(E + \frac{K^2}{r}\right) dr}{\left[\frac{1}{c^2}\left(E + \frac{K^2}{r}\right)^2 - \frac{J^2}{r^2} - m^2c^2\right]^{\frac{1}{2}}}. \tag{6}$$

There are two entirely different classes of solution, depending on the value of $K^2/Jc \equiv \alpha$. Eq. (5) gives[4]

$$\frac{l}{r} = 1 + \epsilon\cos\beta(\theta - \theta_0), \qquad \alpha < 1, \tag{7a}$$

$$\frac{\tilde{l}}{r} = -1 + \epsilon\cosh\tilde{\beta}(\theta - \theta_0), \quad \alpha > 1, \tag{7b}$$

where

$$\epsilon = \frac{1}{\gamma}\left[1 + \frac{1}{\alpha^2}(\gamma^2 - 1)\right]^{\frac{1}{2}}, \tag{8}$$

$$\gamma = E/mc^2, \tag{9}$$

$$\alpha = K^2/Jc, \tag{10}$$

$$\beta = (1 - \alpha^2)^{\frac{1}{2}}, \quad \tilde{\beta} = (\alpha^2 - 1)^{\frac{1}{2}}, \tag{11}$$

$$l = l_0\beta^2/\gamma, \quad \tilde{l} = l_0\tilde{\beta}^2/\gamma, \tag{12}$$

$$l_0 = J^2/K^2m. \tag{13}$$

When $\alpha = 1$, the solution is the same as the limit of Eqs. (7a) and (7b), the limits being identical. The result is

$$\frac{2l_0}{r} = -1 + \frac{1}{\gamma^2} + (\theta - \theta_0)^2. \tag{14}$$

For comparison with Eq. (7a) the nonrelativistic solution is

$$\frac{l_0}{r} = 1 + \epsilon'\cos(\theta - \theta_0), \tag{15}$$

where

$$\epsilon' = 2\,|E'|\,J^2/K^4m, \tag{16}$$

the conic eccentricity. It is readily ascertained that in the limit that $|E'| = |E - mc^2| \ll mc^2$; Eq. (8) reduces to Eq. (16). Furthermore, when $c \to \infty$, $\beta \to 1$, and $\alpha \to 0$, so Eq. (7a) reduces to Eq. (15). At first it would appear that Eq. (7b) does *not*, but closer inspection reveals that $c \to \infty$, which causes $\alpha \to 0$ and $\beta \to 1$, also results in $\tilde{\beta} \to i$, from which Eq. (7b) $\to$ Eq. (15), apart from a trivial sign which can be absorbed into θ_0.

III. DISCUSSION OF RESULTS

Comparison of Eq. (7a) with Eq. (16) shows that when $\alpha < 1$, the former gives "relativistic conic sections" with ϵ and l playing the role of relativistic "eccentricity" and "*semilatus rectum*," respectively. The novelty is β. When β is very close to unity, the ellipses of the bound case precess forward by about

[4] It is easily seen from (1) that the case of a repulsive potential is handled by replacing K^2 by $-K^2$. This carries through to Eqs. (7a) and (7b) with the result

$$(-|l|/r = 1 + \epsilon\cos\beta(\theta - \theta_0), \qquad \alpha < 1,$$
$$(-|\tilde{l}|/r = -1 + \epsilon\cosh\tilde{\beta}(\theta - \theta_0), \quad \alpha > 1.$$

$\delta\theta \approx \pi\alpha^2$ per revolution. With moderate increase in α and attendant decrease in β, precessing ellipses graduate into the celebrated "open rosettes." As α increases towards one, β drops towards zero, and the appearance of the orbit changes again, starting as a spiral that at first winds on itself with an ever-decreasing radius, shrinking to a minimum value $r_{\min} = l(1+\epsilon)^{-1}$ [see Eq. (7a)], and then unwinding outward again rising back gradually to a maximum radius $r_{\max} = l(1-\epsilon)^{-1}$. The picture is that of a pulsating spiral.

The "hyperbolas" and "parabolas" of Eq. (7a) are much like those of Eq. (15) when $E > mc^2$ (i.e., $E' > 0$). The difference is that they have narrower opening angles and as β falls they even cross themselves.[5] When $\beta \ll 1$, the incoming mass winds several times around the center before re-emerging.

However, at first sight, the most remarkable feature of the results, Eqs. (7a) and (7b), is the fact that there are two entirely different classes of solution, a situation not encountered in the nonrelativistic Kepler problem. Thus, as contrasted to (7a), the bound solution to (7b), where $\alpha > 1$, is a spiral falling from $r_{\max} = \tilde{l}(-1+\epsilon)^{-1}$ to the origin.[6] The absorption time from $r_{\max}$ can be found from the relativistic energy integral (6); it is

$$\tau_{\text{abs}} = \int_0^{\tilde{l}/(\epsilon-1)} \frac{\frac{1}{c^2}\left(E + \frac{K^2}{r}\right) dr}{\left[\frac{1}{c^2}\left(E + \frac{K^2}{r}\right)^2 - \frac{J^2}{r^2} - m^2c^2\right]^{\frac{1}{2}}}, \quad (17)$$

which gives a finite result.[7]

The unbound case when $\alpha \geq 1$ exhibits fall effects also. To see this, let us change the description of the orbit from that of Eq. (7b) by getting rid of θ_0, an angle with a clean physical meaning in the bound case and when $\alpha < 1$, but less so in the present instance. Let the mass m be incident along the $\theta = 0$ direction, then $r(\theta = 0) = \infty$ and this determines θ_0 as

$$\theta_0 = \pm \frac{1}{\tilde{\beta}} \cosh^{-1} \frac{1}{\epsilon}. \quad (18)$$

The lower sign is needed to keep r positive. The result for the orbit is

$$\tilde{l}/r = -1 + \cosh\tilde{\beta}\theta + (1-\epsilon^2)^{\frac{1}{2}} \sinh\tilde{\beta}\theta. \quad (12'')$$

[A similar treatment of Eq. (14) will eliminate θ_0 there, too.] The impact parameter is

$$b = \lim_{r\to\infty} r\sin\theta = \frac{\alpha}{(\gamma^2-1)^{\frac{1}{2}}} l_0 = \frac{J/mc}{(\gamma^2-1)^{\frac{1}{2}}}. \quad (19)$$

Here the picture is that of an incident particle coming in from far out and, rather than scattering and re-emerging, instead being "absorbed" by the central potential.

We can present a centrifugal barrier explanation again as follows: Fix K^2 and vary the impact parameter b from large values down towards zero; K^2/Jc rises from small values, passes through one, and increases beyond. The high J particles scatter in accord with Eq. (7a) but the low angular momentum, low impact-parameter particles are absorbed as dictated by Eq. (19).[8] Crudely,

$$Jc = \frac{mv_0 b}{(1-v_0^2/c^2)^{\frac{1}{2}}} \cdot c = \frac{mc^2}{(1-v_0^2/c^2)^{\frac{1}{2}}} \left(\frac{v_0}{c}\right) b = E\left(\frac{v_0}{c}\right) b, \quad (20)$$

where v_0 is the velocity of m at $r = \infty$; we see that $K^2/Jc > 1$ implies

$$E(v_0/c)b < K^2,$$

or

$$K^2/b > (v_0/c)E,$$

which furnishes a crude confirmation of the statements of the previous paragraph, since K^2/b is a characteristic value of the potential energy.

Finally, it is interesting to note that when $\alpha > 1$ negative energy solutions, but *with the positive square root in* Eq. (1), are possible. Thus if $\gamma < 0$, Eq. (7b) still holds if $\tilde{l}$ and ϵ, which are both proportional to $1/\gamma$, are simply allowed to be negative. The result is just a tighter spiral, starting at a smaller $r_{\max} = |\tilde{l}|/|1+|\epsilon|| = \tilde{l}/(\epsilon-1)$, just the same as the $\gamma > 0$ (and hence $E > 0$) case.

When $\alpha \leq 1$, there are no solutions if $\gamma < 0$ (or if $\gamma^2 < 1-\alpha^2$), as this leads to a negative kinetic energy term.

The reason for these effects is not hard to see. When the coupling is strong, the argument of the square root in Eq. (5) or Eq. (6) is positive for $r < \tilde{l}/(\epsilon-1)$ even for $E < 0$, as the condition $\alpha^2 > 1$ or $K^4/c^2 > J^2$

[5] *Note:* The "parabola" always does!

[6] When $\alpha > i$, ϵ is greater than or less than unity according as E is less than or greater than mc^2, in contradistinction to the case where $\alpha < 1$. Hence $r_{\max}$ as given in the text is positive, as it should be.

[7] Since the reader will make the observation himself, we take this opportunity to point out the amusing fact that the absorption form $r_{\max}$ to $r = 0$ can equally well be played in reverse; viz., emission from $r = 0$ to $r_{\max}$ and then followed by reabsorption, a phenomenon with an obvious, though, as the present effect seems to have nothing to do with range, probably incorrect, quantum analog. There is no trouble with energy or angular momentum conservation because the central mass is infinite—it can serve as a source as well as a sink! It is even more amusing that infinite mass classical particles (such as our force center) can have "intrinsic," though arbitrary, angular momentum.

[8] One is reminded here of the low partial-wave absorption in peripheral production processes at high energies. The "sink" in this case is a probability sink, the large number of open inelastic channels [cf. K. Gottfried and J. D. Jackson, Nuovo Cimento **33**, 309 (1964)].

causes the $1/r^2$ term, which dominates in this region, to have a positive coefficient. When $\alpha^2 < 1$, this is not the case. The effect of the "centrifugal barrier" is summarized in the J^2/r^2 term in the square root of Eq. (5) and we see that it competes at $r = 0$ *not with* $V(r)$ *but with* $[V(r)]^2$. This is the reason for its relative ineffectiveness in preventing collapse. In the nonrelativistic theory, the centrifugal barrier competes with $V(r)$ itself. There is no contradiction here for the nonrelativistic *theory* is obtained from the relativistic theory by expanding the argument of the square root in Eq. (5), getting

$$\left[2m\left(E' + \frac{K^2}{r}\right) - \frac{J^2}{r^2}\right] + \frac{1}{c^2}\left(E' + \frac{K^2}{r}\right)^2,$$

and dropping the last term. This procedure gives an incorrect nonrelativistic *limit* when $r \to 0$, but of course this extra term could never show up in the nonrelativistic *theory*, as the latter can be obtained from the relativistic theory only by making $c \to \infty$, which justifies the dropping of the $[V(r)]^2$ contribution "nonrelativistically."

To look at this in another way, and in the more general case, suppose that near the origin

$$V(r) \sim -k^2 r^n. \tag{21}$$

Then from Eq. (1)

$$\dot{p}_r = -\frac{\partial H}{\partial r} = nk^2 r^{n-1} + \frac{J^2 c^2/r^3}{E + ar^n}. \tag{22}$$

The denominator is the square root appearing in Eq. (1) and hence is positive. Equation (22) can be negative for $r \to 0$ only if $n < 0$. Requiring this with n negative gives

$$|n|\,(k^2/Jc)^2 r^{n+2}(r^n + E/k^2) > 1,$$

or, since $r \to 0$,

$$|n|\,(k^2/Jc)^2 (r^2)^{n+1} > 1. \tag{23}$$

This can be fulfilled for all k^2/Jc only if $n < -1$. If $n = -1$, then the inequality (23) becomes

$$k^2/Jc > 1. \tag{24}$$

In these cases the radial force (22) is sufficiently strongly attractive at the origin that it can cause the particle to fall to the center, as $\dot{p}_r \to -\infty$ there. A Yukawa potential, for example, may be seen to have this property if the value of k^2/Jc is greater than unity in accordance with Eq. (24).

We can conclude that fall occurs in the case of a sufficiently singular potential provided the particle can get close enough to $r = 0$ for the singular character of the potential to take effect. This is certainly the case for any power law attraction if $\gamma < 1$. The unbound case, however, has a trivial "exception," namely, that where the centrifugal barrier dominates over $V(r)$ at larger r and $E - mc^2$ lies low enough that the particle is turned away before it can get in close enough to the origin for the strong attraction there to take hold. Whenever the energy is great enough for the particle to get into the $r = 0$ region, however, it falls to the center as long as $V(r)$ is sufficiently singular.

IV. QUANTUM-MECHANICAL CASES

The nonrelativistic classical problem with

$$V(r) = -k^2/2r^2 \tag{25}$$

gives an orbit equation,

$$\theta - \theta_0 = \int^r \frac{(J/r^2)\,dr}{\left[2mE' + (k^2 m - J^2)\frac{1}{r^2}\right]^{\frac{1}{2}}}, \tag{26}$$

which results in fall if $k^2 m \geq J^2$. The quantum version of this potential gives a wavefunction[9]:

$$R(r) \sim \frac{1}{\sqrt{r}} \cos\left[(\gamma - \tfrac{1}{4})^{\frac{1}{2}} \ln r + c\right], \quad c = \text{const} \tag{27}$$

when $\gamma > \frac{1}{4}$. Here

$$\gamma = mk^2/\hbar^2 - l(l+1). \tag{28}$$

The condition $\gamma > \frac{1}{4}$ is re-expressible as

$$k^2 m > \hbar^2 (l + \tfrac{1}{2})^2, \tag{29}$$

which can be compared with the classical condition [the $\frac{1}{2}$ in Eq. (29) comes from the uncertainty principle]. In Ref. 9 it is argued that the solution (27) reflects a kind of "fall" to the origin because, as $r \to 0$, $R(r)$ has infinitely many nodes for any finite value of E'. The fact that the ground state is supposed to have no nodes, means that it must be at $E' = -\infty$. Since in general a particle is supposed to be in a region of space where $E' > V$, it follows that the ground state is that of the particle confined to $r = 0$. This is quantum-mechanical "fall."

The Dirac and Klein–Gordon equations exhibit similar features when $V(r) = -K^2/r$ near $r = 0$. We look only at the Dirac case. The energy levels are[2]

$$E = mc^2\left[1 + \frac{\alpha_0'^2}{\{n' + [(j+\frac{1}{2})^2 - \alpha_0'^2]^{\frac{1}{2}}\}^2}\right]^{-\frac{1}{2}}, \quad n' = 0, 1, 2, \cdots, \tag{30}$$

[9] L. Landau and E. Lifshitz, *Quantum Mechanics* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1958), p. 118ff.

where

$$\alpha_0'^2 \equiv K^2/\hbar c. \tag{31}$$

These are complex when

$$\alpha' = \frac{K^2}{\hbar(j+\frac{1}{2})c} > 1. \tag{32}$$

Both the "large" and "small" solutions near the origin, independent of E, go as

$$r^{\pm[(j+\frac{1}{2})^2-\alpha_0'^2]^{\frac{1}{2}}} = r^{\pm i[(j+\frac{1}{2})^2(\alpha'^2-1)]^{\frac{1}{2}}}, \tag{33}$$

which gives solutions like (27), and arguments analogous to those given in Ref. 9 may strongly suggest quantum-mechanical "fall" here also.

It is suggestive that the classical "coupling constant" seems to generalize to the quantum-mechanical case and it is hard to resist the temptation to speculate on the possible significance of the effects. Thus, it is probably true that strong interaction phenomena have an entirely different character from electromagnetic interaction phenomena and, perhaps, in a way, are somehow related to "fall" effects. The connection, if any, however, must remain vague until a more realistic model can be set up and solved, but we have no such model.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Complementary Variational Principles and Their Application to Neutron Transport Problems*

G. C. POMRANING
General Dynamics, General Atomic Division
San Diego, California

(Received 19 January 1967)

Several variational principles are developed which give upper and lower bounds for the linear functional (S, ψ), where ψ is the solution of the inhomogeneous equation $H\psi = S$ with H a self-adjoint, positive-definite, linear operator. Some of the principles bound this functional only with respect to small or local variations, whereas others give bounds for arbitrary variations. Several of our results coincide with those of other authors widely scattered throughout the literature, and we show that these principles have a common origin. Other results given are new. Examples of the use of these principles are taken from the field of neutron transport theory, and we use both the linear Boltzmann or transport equation and the diffusion equation. One interesting result is that certain "exact" values of the extrapolated endpoint for the Milne problem which have been reported in the literature fall, due to numerical inaccuracies, outside the bounds computed here.

I. INTRODUCTION

FOR the purposes of this paper, we define a variational principle in the following limited sense: Suppose that a physical situation is entirely described by a given equation (or set of equations). Further suppose that we are only interested in a single gross aspect of the solution of this equation, i.e., some functional of the solution. Finally, suppose that we know an approximate solution to this equation. By definition, we say that this approximate solution differs from the exact solution by first-order terms in some smallness parameters. Then a direct calculation of the functional of interest with this approximate solution will, by definition, contain first-order errors as compared to the exact result. A variational principle is defined to be a definite procedure for using this first-order solution of the equation to obtain an estimate of the functional of interest containing only second-order errors. We further define complementary variational principles as a pair of principles for estimating the same functional, one of which is a minimum principle and the other a maximum principle. By a minimum variational principle we mean that for sufficiently small first-order errors in the approximate solution (trial function), the variational estimate of the functional will necessarily be a second-order overestimate of the true value. Likewise, a maximum principle will always yield a second-order underestimate of the exact result with respect to local errors (variations) in the trial function.

* This work was initiated while the author was a guest at the Brookhaven National Laboratory.

where

$$\alpha_0'^2 \equiv K^2/\hbar c. \tag{31}$$

These are complex when

$$\alpha' = \frac{K^2}{\hbar(j+\frac{1}{2})c} > 1. \tag{32}$$

Both the "large" and "small" solutions near the origin, independent of E, go as

$$r^{\pm[(j+\frac{1}{2})^2-\alpha_0'^2]^{\frac{1}{2}}} = r^{\pm i[(j+\frac{1}{2})^2(\alpha'^2-1)]^{\frac{1}{2}}}, \tag{33}$$

which gives solutions like (27), and arguments analogous to those given in Ref. 9 may strongly suggest quantum-mechanical "fall" here also.

It is suggestive that the classical "coupling constant" seems to generalize to the quantum-mechanical case and it is hard to resist the temptation to speculate on the possible significance of the effects. Thus, it is probably true that strong interaction phenomena have an entirely different character from electromagnetic interaction phenomena and, perhaps, in a way, are somehow related to "fall" effects. The connection, if any, however, must remain vague until a more realistic model can be set up and solved, but we have no such model.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Complementary Variational Principles and Their Application to Neutron Transport Problems*

G. C. POMRANING
General Dynamics, General Atomic Division
San Diego, California

(Received 19 January 1967)

Several variational principles are developed which give upper and lower bounds for the linear functional (S, ψ), where ψ is the solution of the inhomogeneous equation $H\psi = S$ with H a self-adjoint, positive-definite, linear operator. Some of the principles bound this functional only with respect to small or local variations, whereas others give bounds for arbitrary variations. Several of our results coincide with those of other authors widely scattered throughout the literature, and we show that these principles have a common origin. Other results given are new. Examples of the use of these principles are taken from the field of neutron transport theory, and we use both the linear Boltzmann or transport equation and the diffusion equation. One interesting result is that certain "exact" values of the extrapolated endpoint for the Milne problem which have been reported in the literature fall, due to numerical inaccuracies, outside the bounds computed here.

I. INTRODUCTION

FOR the purposes of this paper, we define a variational principle in the following limited sense: Suppose that a physical situation is entirely described by a given equation (or set of equations). Further suppose that we are only interested in a single gross aspect of the solution of this equation, i.e., some functional of the solution. Finally, suppose that we know an approximate solution to this equation. By definition, we say that this approximate solution differs from the exact solution by first-order terms in some smallness parameters. Then a direct calculation of the functional of interest with this approximate solution will, by definition, contain first-order errors as compared to the exact result. A variational principle is defined to be a definite procedure for using this first-order solution of the equation to obtain an estimate of the functional of interest containing only second-order errors. We further define complementary variational principles as a pair of principles for estimating the same functional, one of which is a minimum principle and the other a maximum principle. By a minimum variational principle we mean that for sufficiently small first-order errors in the approximate solution (trial function), the variational estimate of the functional will necessarily be a second-order overestimate of the true value. Likewise, a maximum principle will always yield a second-order underestimate of the exact result with respect to local errors (variations) in the trial function.

* This work was initiated while the author was a guest at the Brookhaven National Laboratory.

In this paper we consider a certain class of inhomogeneous equations and construct complementary variational principles for a particular functional. Specifically, we consider equations of the form

$$H(x)V(x) = S(x), \tag{1}$$

where $H(x)$ is a real, self-adjoint (Hermitian), positive definite, linear operator, x is the independent variable (or represents the set of independent variables), $V(x)$ is the dependent variable, and $S(x)$ is a given inhomogeneous term. The operator H^*, said to be adjoint to H, is defined by

$$(f, Hg) = (g, H^*f), \tag{2}$$

for all real functions f and g in the field of definition of H and H^*, where we have defined the inner product as

$$(f, g) = (g, f) \equiv \int dx f(x)g(x). \tag{3}$$

A self-adjoint operator is one for which $H = H^*$. The restriction throughout this paper to real operators and functions is nonessential and is imposed only for simplicity. A positive definite (or positive bounded below) operator is one which obeys

$$(f, Hf) \geq m(f, f), \tag{4}$$

for all functions f in the domain of H, where m is a real, positive number. We show that one can always construct complementary variational principles which are well-defined from a calculational point of view for the functional I, given by

$$I[V] \equiv (S, V). \tag{5}$$

I is frequently called the weighted average of V. We also show that one can obtain more powerful, though not as well-defined from a computational viewpoint, results. That is, complementary variational principles are defined such that they bound the exact result with respect to small or local variations. We also give methods which necessarily bound the true value of $I[V]$ with respect to all variations, no matter how large. For arbitrary variations, these bounds are also of second order but it is only for small variations, of course, that a second-order result is more than a formal notion.

Some of our results coincide with those of other authors scattered throughout the literature. The treatment given here shows that these principles, which have been discussed in diverse contexts, can be derived from a common starting point. Specifically, we make contact with the work of Roussopoulos,[1] Schwinger[2] and Francis,[3] Selengut,[4] Kato,[5] Slobodyansky,[6] Becker,[7] Yasinsky,[8] and especially Noble.[9] It was the recent work of Noble and subsequent related publications[10,11] which stimulated the present work. Further, several results are given which are believed to be new and useful.

We draw examples of the use of our results from the field of neutron transport theory. In particular, we shall consider the one-velocity transport equation with isotropic scattering[12]

$$\psi(\mathbf{r}) = \int_V d\mathbf{r}' \frac{e^{-\tau}}{4\pi |\mathbf{r} - \mathbf{r}'|^2}\left[\Sigma_s(\mathbf{r}')\psi(\mathbf{r}') + S(\mathbf{r}')\right], \tag{6}$$

where τ, the optical path length between $\mathbf{r}$ and $\mathbf{r}'$, is given by

$$\tau \equiv \int_{\mathbf{r}'}^{\mathbf{r}} ds \Sigma(s), \tag{7}$$

$\psi(\mathbf{r})$ is the scalar flux, $\Sigma(\mathbf{r})$ is the macroscopic collision cross section, $\Sigma_s(\mathbf{r})$ is the macroscopic scattering cross section, and $S(\mathbf{r})$ is the (isotropic) scalar external source. Further, we consider the diffusion theory approximation to Eqs. (6) and (7) given by[12]

$$-\nabla \cdot D(\mathbf{r})\nabla\psi(\mathbf{r}) + \Sigma_a(\mathbf{r})\psi(\mathbf{r}) = S(\mathbf{r}), \tag{8}$$

where $D(\mathbf{r})$, the diffusion coefficient, is defined by

$$D(\mathbf{r}) \equiv 1/3\Sigma(\mathbf{r}), \tag{9}$$

and the macroscopic absorption cross section $\Sigma_a(\mathbf{r})$ is given by

$$\Sigma_a(\mathbf{r}) \equiv \Sigma(\mathbf{r}) - \Sigma_s(\mathbf{r}). \tag{10}$$

II. DEVELOPMENT OF THE COMPLEMENTARY PRINCIPLES

In his work dealing with complementary variational principles, Noble[9] treated the case of two coupled equations of the form

$$TV = \frac{\partial W(x, U, V)}{\partial U}, \tag{11}$$

$$T^*U = \frac{\partial W(x, U, V)}{\partial V}, \tag{12}$$

[1] P. Roussopoulos, Compt. Rend. Acad. Sci. Paris **236**, 1858 (1953).

[2] H. Levine and J. Schwinger, Phys. Rev. **75**, 1423 (1949).

[3] N. Francis, J. Stewart, L. Bohl, and T. Krieger, "Variational Solutions of the Transport Equation," Second Geneva Conference on Peaceful Uses of Atomic Energy, Report 15/P/627 (1958).

[4] D. S. Selengut, Hanford Quarterly Report HW-59126, 89 (1959).

[5] T. Kato, Math. Ann. **126**, 253 (1953).

[6] S. G. Mikhlin, *Variational Methods in Mathematical Physics* (The Macmillan Co., New York, 1964).

[7] M. Becker, *The Principles and Applications of Variational Methods* (M.I.T. Press, Cambridge, Mass., 1964).

[8] J. B. Yasinsky, Trans. Am. Nucl. Soc. **9**, 471 (1966).

[9] B. Noble, *Math. Research Center Report No.* 473, Madison, Wisconsin (1964).

[10] L. R. Rall, J. Math. Anal. Appl. **14**, 174 (1966).

[11] V. Komkov, J. Math. Anal. Appl. **14**, 511 (1966).

[12] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).

where T^* is the operator adjoint to T, and obtained conditions on the function W for complementary variational principles to exist. In order to apply Noble's ideas to Eq. (1), we introduce a second dependent variable in the following way: We define a non-negative operator, say L, as one which satisfies

$$(f, Lf) \geq 0, \tag{13}$$

for all functions f in the field of definition of L. We then write the operator H in Eq. (1) as the sum of two non-negative, self-adjoint operators, a decomposition which we shall later show is always possible. We further represent one of these two operators as the product of an operator T and its adjoint T^*, another decomposition which is always possible.[6] Hence we have the representation

$$H = L + T^*T, \tag{14}$$

where L is non-negative and self-adjoint. To obtain complementary variational principles for $I[V]$ which are useful in a practical sense, we require that L^{-1}, the operator inverse to L, be known. However, we do not require that the operators T and T^* be known. They are introduced only to facilitate the derivation of the principles and do not appear in the final results. We introduce a second dependent variable $U(x)$ according to

$$U = TV. \tag{15}$$

Using Eqs. (14) and (15) in Eq. (1), we have

$$LV + T^*U = S. \tag{16}$$

Equations (15) and (16) are entirely equivalent to the simpler Eq. (1) but have the advantage for our purposes of being closer to the form treated by Noble[9] than is Eq. (1).

Consider the functional [analogous to the primary functional of Noble—his Eq. (11)]

$$F_1[u, v] = 2(S, v) - (v, Lv) + (u, u) - 2(u, Tv), \tag{17}$$

where u and v are arbitrary functions. If we let

$$v = V + \delta v, \tag{18}$$

$$u = U + \delta u, \tag{19}$$

we find, using Eqs. (5), (15), and (16),

$$F_1[u, v] = I[V] - (\delta v, L\delta v) + (\delta u, \delta u) - 2(\delta u, T\delta v). \tag{20}$$

Hence $F_1[u, v]$ is a variational principle for $I[V]$ since it estimates $I[V]$ with second-order errors. However, since δu and δv are independent and arbitrary, $F_1[u, v]$ can either be an overestimate or an underestimate of $I[V]$, depending upon δu and δv.

To derive a functional which always underestimates $I[V]$, let us consider only v to be independent and generate the trial function u from

$$u = Tv, \tag{21}$$

in analogy to Eq. (15). Then we find

$$F_1[Tv, v] \equiv F_2[v] = 2(S, v) - (v, Hv), \tag{22}$$

or

$$F_2[v] = I[V] - (\delta v, H\delta v). \tag{23}$$

Since H is positive definite, Eq. (23) shows that $F_2[v]$ yields a lower bound for $I[V]$ for all v. Note that we have not assumed δv to be small. Of course, it is only for small δv that this lower bound will be close (differing by second-order terms) to the exact result, $I[V]$. Equation (22) is just the Roussopoulos functional[1] in the self-adjoint case and its lower-bound property has previously been noted.[13]

This lower-bound result can be put in a more convenient (and more accurate) form by using a device due to Selengut.[4] We write the trial function $v(x)$ as the product of an amplitude α and a shape function $\bar{v}(x)$, i.e.,

$$v(x) = \alpha\bar{v}(x). \tag{24}$$

Since $F_2[v]$ underestimates $I[V]$ for all functions $v(x)$, we determine the value of α, as a functional of $\bar{v}(x)$, as that value which maximizes $F_2[v]$. That is, we substitute Eq. (24) into Eq. (22) and set the first derivative of the result with respect to α equal to zero. We find, suppressing the bar on $\bar{v}(x)$, for the optimum value of α,

$$\alpha = (S, v)/(v, Hv). \tag{25}$$

Using Eq. (24), with α given by Eq. (25), in $F_2[v]$, we obtain a new functional $G_2[v]$, given by

$$G_2[v] = (S, v)^2/(v, Hv), \tag{26}$$

which is frequently called the Schwinger functional[2] and has been widely used by Francis *et al.*[3] $G_2[v]$ has the advantage over $F_2[v]$ that the normalization of the trial function is irrelevant. From its derivation it is clear that $G_2[v]$, for any trial function $v(x)$, is a lower bound for $I[V]$. A more general result could be obtained by using

$$v(x) = \sum_{n=1}^{N} \alpha_n \bar{v}_n(x) \tag{27}$$

in $F_2[v]$ and maximizing the result with respect to all the α_n. We note that in our lower-bound results, Eqs. (22) and (26), only the operator H and none of the

[13] R. Goldstein, J. Math. Phys. **8**, 473 (1967).

decomposition components, i.e., the operator L, appears. Such is not the case for the upper-bound results which we now derive.

We return to $F_1[u, v]$, Eq. (17), and in this instance treat only u as the independent trial function. Since U and V are related by Eq. (16), we generate the trial function $v(x)$ according to

$$v = L^{-1}(S - T^*u). \tag{28}$$

We then obtain

$$F_1[u, L^{-1}(S - T^*u)] \equiv F_3[u] = (u, u) + ([S - T^*u], L^{-1}[S - T^*u]), \tag{29}$$

or

$$F_3[u] = I[V] + (\delta u, \delta u) + (T^*\delta u, L^{-1}T^*\delta u), \tag{30}$$

where we have used the fact that if L is self-adjoint, so is L^{-1}. Since the operator L was stipulated as non-negative, L^{-1} is non-negative and Eq. (30) shows that $F_3[u]$ is an upper bound for $I[V]$ for all trial functions $u(x)$. Further, the difference between $F_3[u]$ and $I[V]$ is of second order in δu. Again, we have nowhere assumed δu to be small. $F_3[u]$ as given by Eq. (29) is in an inconvenient form, since it contains the operator T^* involved in the decomposition of H. This practical difficulty can be overcome by defining the trial function u as the result of T operating on a function v, i.e.,

$$u = Tv. \tag{31}$$

In view of Eq. (15), for u to be a good approximation to U, v should be a good approximation to V. Using Eq. (31) in $F_3[u]$ and using Eq. (14) to eliminate the T^*T which results, we have an alternate form of $F_3[u]$ which we denote by $F_4[v]$:

$$F_3[Tv] \equiv F_4[v] = 2(S, v) - (v, Hv) + ([Hv - S], L^{-1}[Hv - S]). \tag{32}$$

In this form we see that our upper-bound result is just our lower-bound result $F_2[v]$, plus an additional term which is obviously of second order and positive. Contrary to our lower-bound results, $F_4[v]$ contains the operator L^{-1} and hence is more difficult to use. $F_4[v]$ is substantially the same as a variational principle due to Slobodyansky,[6] which he arrived at from entirely different considerations.

As before, we can obtain a normalization independent result by writing

$$v(x) = \beta\bar{v}(x), \tag{33}$$

and minimizing $F_4[v]$ with respect to β. We find for this value of β, dropping the bar on $\bar{v}(x)$,

$$\beta = (S, v - L^{-1}Hv)/(Hv, v - L^{-1}Hv), \tag{34}$$

which leads to the upper-bound normalization independent result

$$G_4[v] = (S, v - L^{-1}Hv)^2/(Hv, v - L^{-1}Hv) + (S, L^{-1}S). \tag{35}$$

We emphasize that $G_4[v]$ is an overestimate of $I[V]$ for all trial functions $v(x)$, no matter how inaccurate. As with our lower-bound results, one can derive a more general normalization independent upper-bound result by using a trial function of the form

$$v(x) = \sum_{n=1}^{N} \beta_n \bar{v}_n(x), \tag{36}$$

and minimizing $F_4[v]$ with respect to all the β_n.

Let us show the relationship of our results to those of Kato.[5] Let I_U and I_L denote upper and lower bounds, respectively, to the exact result $I[V]$. We have the obvious inequality

$$(I - I_U)(I - I_L) \leq 0. \tag{37}$$

Adding $\frac{1}{4}(I_U - I_L)^2$ to both sides of this inequality, regrouping terms on the left-hand side, and taking the square root of the resulting inequality, we obtain

$$|I - \tfrac{1}{2}(I_L + I_U)| \leq \tfrac{1}{2}(I_U - I_L). \tag{38}$$

To formally obtain Kato's result, we use $F_2[v]$ for I_L and $F_4[v]$ for I_U. We further set $L = H$ in $F_4[v]$, decompose H according to $H = \Gamma^*\Gamma$, and define a function v' as any solution of the equation

$$\Gamma^* v' = S. \tag{39}$$

Then we find

$$I_L = 2(S, v) - (\Gamma v, \Gamma v), \tag{40}$$

$$I_U = (v', v'), \tag{41}$$

and Eq. (38) becomes

$$|I - \tfrac{1}{2}(I_L + I_U)| \leq \tfrac{1}{2}(\Gamma v - v', \Gamma v - v'), \tag{42}$$

which is Kato's result [his Eq. (7)] specialized to the case for which $I[V]$ is the functional of interest. We note that, in general, Eq. (42) is difficult to use since the equation defining v', Eq. (39), involves Γ^* and hence one must explicitly decompose the operator H into the product of an operator and its adjoint. Further, once such a decomposition is known, one must be able to obtain a solution of Eq. (39). However, in the cases for which this is possible, Kato's result is quite powerful since his formulation allows one to bound an arbitrary linear functional of V. (We are considering only a special linear functional, $I[V] \equiv (S, V)$ in this paper.)

Finally, we remark that the introduction of a second dependent variable U into the formalism [see Eqs. (15) and (16)] and the subsequent treatment of either

U or V as the sole dependent variable can be considered as an extension of the Legendre or involutory transformation which Courant and Hilbert[14] introduce in their variational treatment of differential equations. Yasinsky[8] has explicitly shown this connection between $F_2[v]$ and $F_4[v]$ for the Sturm–Liouville equation.

III. DECOMPOSITION OF THE OPERATOR H

In the previous section it was stated that a positive definite operator can always be decomposed into the sum of two non-negative operators. This notion was crucial in the derivation of our complementary variational principles, in particular the upper-bound results. In this section we elaborate on this decomposition and, as a result of this discussion, further develop our upper-bound results. We have seen that our lower-bound results involve only the operator H and hence, insofar as the calculation of lower bounds is concerned, the development is complete.

In certain instances, the required decomposition is obvious. For example, for the diffusion equation given by Eq. (8), we can make the identification

$$L = \Sigma_a(\mathbf{r}), \tag{43}$$

$$T = [D(\mathbf{r})]^{\frac{1}{2}}\nabla; \quad T^* = -\nabla \cdot [D(\mathbf{r})]^{\frac{1}{2}}, \tag{44}$$

since the physics dictates that Σ_a and D are non-negative functions of space. In general, however, such a decomposition by inspection is not possible. The transport equation given by Eqs. (6) and (7) is an example of this more general situation. Let us therefore give a decomposition of a positive definite operator H into the sum of two non-negative operators, say L and L', in the general case. Let $\bar{m}$ be the largest value of m for which Eq. (4) holds and consider the decomposition

$$L = \gamma I, \tag{45}$$

$$L' = H - \gamma I, \tag{46}$$

where $0 \leq \gamma \leq \bar{m}$, and I is the identity operator. Then we have

$$(f, Lf) = \gamma(f,f) \tag{47}$$

and

$$(f, L'f) = (f, Hf) - \gamma(f,f) \geq (\bar{m} - \gamma)(f,f). \tag{48}$$

Clearly $(f, Lf) \geq 0$ since $\gamma \geq 0$ and also $(f, L'f) \geq 0$ since $\gamma \leq \bar{m}$. Thus we have trivially shown that a positive definite operator can always be written as the sum of two non-negative operators.

If the decomposition of H is accomplished according to Eqs. (45) and (46), then $F_4[v]$ and $G_4[v]$ are written

$$F_5[v] = 2(S, v) - (v, Hv) + \gamma^{-1}(Hv - S, Hv - S), \tag{49}$$

$$G_5[v] = \frac{(S, v - Hv/\gamma)^2}{(Hv, v - Hv/\gamma)} + \frac{1}{\gamma}(S, S), \tag{50}$$

where

$$0 \leq \gamma \leq \bar{m}. \tag{51}$$

Although $F_5[v]$ and $G_5[v]$ correspond to special cases of $F_4[v]$ and $G_4[v]$, we have chosen to consider them as separate results (hence the change in subscript from 4 to 5) for purposes of subsequent discussion. In using Eqs. (49) and (50), one would like to choose γ as large as possible, consistent with Eq. (51), since a tighter upper bound is achieved with a larger γ. We note that $F_5[v]$ is just the Roussopoulos functional[1] plus a multiple of the "least squares" principle used by Becker.[7]

A somewhat different restriction on γ can be found by considering Eq. (49) directly. Using Eq. (18), we find

$$F_5[v] = I[V] - (\delta v, H\delta v) + \gamma^{-1}(H\delta v, H\delta v). \tag{52}$$

Hence $F_5[v]$ will be an upper bound for $I[V]$ if we choose γ in the range $0 \leq \gamma \leq \bar{n}$, where $\bar{n}$ is the largest value of n for which the inequality

$$(Hf, Hf) \geq n(f, Hf) \tag{53}$$

holds for all functions f in the field of definition of the operator H. We can easily show that this inequality for γ is no more restrictive than Eq. (51). Consider the inequality

$$(Hf - af, Hf - af) \geq 0, \tag{54}$$

where a is any real number. An expansion of Eq. (54) yields

$$\frac{(Hf, Hf)}{(f, Hf)} \geq 2a - a^2\frac{(f,f)}{(f, Hf)}. \tag{55}$$

Let $a = (f, Hf)/(f,f)$. Then Eq. (55) becomes

$$\frac{(Hf, Hf)}{(f, Hf)} \geq \frac{(f, Hf)}{(f,f)}, \tag{56}$$

or

$$\frac{(Hf, Hf)}{(f, Hf)} \geq \bar{m}. \tag{57}$$

Since $\bar{n}$ is by definition the largest value of n for which Eq. (53) holds, Eq. (57) implies

$$\bar{m} \leq \bar{n}, \tag{58}$$

which is the desired result. In order to use $F_5[v]$ or $G_5[v]$, one needs to know $\bar{m}$ or $\bar{n}$. In practical cases for which one can relate $\bar{n}$ and $\bar{m}$ to a quantity

[14] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I.

amenable to calculation, we show that $\bar{m} = \bar{n}$. This equality may be true in a more general circumstance, but the present generality is sufficient for our purposes.

In order to discuss the upper bound on γ, we introduce the eigenfunctions and eigenvalues of H according to

$$H\Theta_j = \Lambda_j \Theta_j. \tag{59}$$

We now consider the functional $J_m[f]$, defined as

$$J_m[f] \equiv (f, Hf)/(f, f), \tag{60}$$

which, by definition, has a minimum value $\bar{m}$. If there exists a function f in the field of definition of H, say f_0, for which $J_m[f_0] = \bar{m}$, then we shall show that $f_0 = \Theta_0$, the eigenfunction of H corresponding to the lowest eigenvalue, and further $\bar{m} = \Lambda_0$. To show this, we calculate the first variation of $J_m[f]$ and evaluate it at $f = f_0$. Since, by assumption, f_0 minimizes $J_m[f]$, this first variation must be equal to zero. Hence we arrive at

$$(f_0, f_0)(Hf_0, \delta f) - (Hf_0, f_0)(f_0, \delta f) = 0, \tag{61}$$

or, since $J_m[f_0] = \bar{m}$,

$$(Hf_0 - \bar{m}f_0, \delta f) = 0. \tag{62}$$

Since δf is arbitrary, Eq. (62) implies

$$Hf_0 = \bar{m}f_0, \tag{63}$$

i.e., that f_0 and $\bar{m}$ are an eigenfunction and eigenvalue of the operator H. Since the eigenvalues of H are given by

$$\Lambda_j = (\Theta_j, H\Theta_j)/(\Theta_j, \Theta_j), \tag{64}$$

clearly $\bar{m}$ must be the lowest eigenvalue Λ_0. A similar analysis can be applied to the functional

$$J_n[f] \equiv (Hf, Hf)/(f, Hf), \tag{65}$$

which leads to

$$H(Hf_0 - \bar{n}f_0) = 0. \tag{66}$$

Since $H\psi = 0$ implies, by Eq. (4), that $\psi = 0$, Eq. (66) is equivalent to

$$Hf_0 = \bar{n}f_0, \tag{67}$$

and we conclude $\bar{n} = \Lambda_0$. Hence, if a minimizing function exists in the domain of H, we have the important result

$$\bar{m} = \bar{n} = \Lambda_0, \tag{68}$$

and the minimizing function is Θ_0.

However, it may be that no function in the field of definition of H minimizes the functionals J_m and J_n. (This is concerned with the question of whether the limit of a sequence of functions in the domain of H is itself in the domain.) If, however, the eigenfunctions of the operator H are complete in the sense that all functions in the domain of H can be expanded in these eigenfunctions, then we can show quite generally that $\bar{m} = \bar{n} = \Lambda_0$. We expand an arbitrary function f in the domain of H according to

$$f = \sum_{j=0}^{\infty} a_j \Theta_j. \tag{69}$$

Assuming this sum to be sufficiently well-behaved so that the operator H in Eq. (60) can be brought inside the summation, and that the operations of summation and integration can be freely interchanged, we find

$$J_m[f] = \Lambda_0 + \frac{\sum_{j=1}^{\infty} (\Lambda_j - \Lambda_0) a_j^2}{\sum_{j=0}^{\infty} a_j^2}. \tag{70}$$

Since $\Lambda_0 \leq \Lambda_j$, we have $J_m[f] \geq \Lambda_0$. A similar analysis shows that $J_n[f] \geq \Lambda_0$. This proof suffices for the application of our results to the diffusion equation, given by Eq. (8), since it is known[14] that the eigenfunctions of the Sturm–Liouville operator are complete. It is not known, however, whether or not the eigenfunctions of the transport operator, Eqs. (6) and (7), are complete. Nevertheless, a slightly modified version of the above analysis, using the Hilbert–Schmidt theorem for integral operators,[15] allows us to draw the same conclusion, i.e., $0 \leq \gamma \leq \Lambda_0$, for the transport operator corresponding to a one-dimensional homogeneous medium. This proof is given in the Appendix.

If we restrict ourselves to practical cases for which $0 \leq \gamma \leq \Lambda_0$, we need consider the problem of obtaining an estimate of Λ_0 for use in our upper-bound results. Ideally, we would like an accurate underestimate of Λ_0, but as remarked by Mikhlin,[6] methods for achieving this for a general positive definite operator are not yet known. For certain operators, special methods will give an accurate lower bound for Λ_0 (see, e.g., Wing[16]). Let us show that the Rayleigh quotient for the lowest eigenvalue, given by

$$\lambda_0[\theta_0] = (\theta_0, H\theta_0)/(\theta_0, \theta_0), \tag{71}$$

can be of some help in the general case. Here θ_0 is an estimate of Θ_0 and the resulting λ_0 is an estimate of Λ_0. It is well-known[6] and, in fact, our analysis of $J_m[f]$ has shown, that Eq. (71) is a second-order overestimate of Λ_0 if θ_0 differs from Θ_0 by first-order terms. Let us consider $F_5[v]$ with γ replaced by λ_0 according to Eq. (71). Denoting the result by $F_6[v, \theta_0]$, we have

$$F_6[v, \theta_0] = 2(S, v) - (v, Hv) + \lambda_0^{-1}(Hv - S, Hv - S), \tag{72}$$

[15] S. G. Mikhlin, *Integral Equations* (The MacMillan Co., New York, 1964).

[16] G. M. Wing, J. Math. Anal. Appl. **11**, 160 (1965).

with λ_0 given by Eq. (71). Since λ_0 is not an underestimate of Λ_0, we cannot conclude that $F_6[v, \theta_0]$ will be an upper bound for $I[V]$ for all v and θ_0. However, we can show that it will overestimate $I[V]$ with respect to local variations, i.e., it is a minimum variational principle. We have

$$F_6[V, \Theta_0] = (S, V) = I[V], \tag{73}$$

$$\delta F_6[V, \Theta_0] = 0, \tag{74}$$

$$\delta^2 F_6[V, \Theta_0] = (2/\Lambda_0)(H\delta v, H\delta v) - 2(\delta v, H\delta v), \tag{75}$$

$$\delta^3 F_6[V, \Theta_0] = 0, \tag{76}$$

$$\delta^4 F_6[V, \Theta_0] = -(12/\Lambda_0^2)(H\delta v, H\delta v)\delta^2\lambda_0[\Theta_0]. \tag{77}$$

From Eqs. (73) and (74) it is clear that $F_6[v, \theta_0]$ is a variational principle for $I[V]$, i.e., it estimates this quantity with second-order errors. Our previous analysis has shown that the right-hand side of Eq. (75) is non-negative for all δv and hence $\delta^2 F_6 \geq 0$, i.e., $F_6[v, \theta_0]$ is a minimum principle. Since the Rayleigh quotient overestimates Λ_0, we have $\delta^2\lambda_0[\Theta_0] \geq 0$, and thus Eq. (77) shows that $\delta^4 F_6 \leq 0$. Now, $F_5[v]$ has the property that $\delta^n F_5[V] \geq 0$ for all n since $F_5[v]$ overestimates $I[V]$ for all trial functions v. Thus only in fourth and higher-order terms in δv does the character of $F_6[v, \theta_0]$ differ from that of $F_5[v]$. That is, while $F_5[v]$ will overestimate $I[V]$ for any trial function v no matter how inaccurate, $F_6[v, \theta_0]$ will only overestimate $I[V]$ with respect to local variations, i.e., if δv and $\delta\theta_0$ are small enough so that fourth and higher-order terms are negligible compared to second-order terms.

To derive a normalization independent form of $F_6[v, \theta_0]$, we set

$$v(x) = \epsilon\bar{v}(x), \tag{78}$$

where $\bar{v}(x)$ is a shape function and the amplitude ϵ is to be determined as a functional of $\bar{v}(x)$. Analogous to our derivation of $G_5[v]$, we could use Eq. (78) in $F_6[v, \theta_0]$ and obtain ϵ by setting the first derivative of this result with respect to ϵ equal to zero. The result of this would be $G_5[v]$ given by Eq. (50) with γ replaced by λ_0. However, since $F_6[v, \theta_0]$ is not necessarily an overestimate of $I[V]$ for δv and $\delta\theta_0$ large, we cannot interpret this procedure as minimizing $F_6[\epsilon\bar{v}, \theta_0]$ with respect to ϵ. It seems more desirable to determine ϵ from an extremum argument if possible. One such method is to obtain ϵ by minimizing the term $(\delta v, H\delta v)$ occurring in the second variation of $F_6[v, \theta_0]$, Eq. (75). Since this term, and only this term, contributes a negative component to the second variation, we can interpret this procedure as minimizing the magnitude of the negative component of the second variation. It was just this term which was minimized in deriving $G_2[v]$, Eq. (26), and hence we obtain $\epsilon = \alpha$, where α is given by Eq. (25). Our final result is

$$G_6[v, \theta_0] = (S, v)^2/(v, Hv) + \lambda_0^{-1}(\alpha Hv - S, \alpha Hv - S), \tag{79}$$

with α given by Eq. (25). We see that $G_6[v, \theta_0]$, a minimum principle, is just $G_2[v]$, a maximum principle, plus a positive "correction" term. While $G_6[v, \theta_0]$ gives an overestimate of $I[V]$ for δv and $\delta\theta_0$ small (so that fourth-order terms are negligible), it will not necessarily overestimate $I[V]$ if very inaccurate trial functions are used. $G_5[v]$, on the other hand, will yield an overestimate of $I[V]$ for all trial functions v no matter how inaccurate, but its use requires a knowledge of a lower bound on Λ_0, the smallest eigenvalue of the operator H, and this knowledge is not always easy to obtain.

IV. SOME EXAMPLES

As a first example of the use of complementary variational principles, we consider a neutron diffusion theory problem closely related to the example given by Noble.[9] Specifically, we analyze a homogeneous slab of thickness 2τ with a spatially independent external source. Equation (8) can then be written

$$-d^2\psi(x)/dx^2 + \psi(x) = 1/(2\tau)^{\frac{1}{2}}, \tag{80}$$

where the unit of distance has been chosen as the diffusion length $(D/\Sigma_a)^{\frac{1}{2}}$ and, in these units, the source strength has been chosen as $S = 1/(2\tau)^{\frac{1}{2}}$. If the neutron flux is assumed to vanish at the edges of the slab, Eq. (80) is to be solved subject to the boundary conditions

$$\psi(\tau) = \psi(-\tau) = 0. \tag{81}$$

The complementary variational principles will give upper and lower bounds for the functional

$$(S, \psi) = (2\tau)^{-\frac{1}{2}}\int_{-\tau}^{\tau} dx\psi(x). \tag{82}$$

As the thickness of the slab approaches infinity, $\psi(x)$ approaches $(2\tau)^{-\frac{1}{2}}$ (except in the immediate vicinity of the boundaries) and thus (S, ψ) approaches unity. This was the reason for the particular source normalization used in Eq. (80). The exact result for (S, ψ) is easily found to be

$$(S, \psi) = 1 - \tau^{-1}\tanh(\tau). \tag{83}$$

If we choose a parabola satisfying the boundary conditions as the trial function $\varphi(x)$ for use in the variational principles, i.e.,

$$\psi(x) \approx \varphi(x) = \tau^2 - x^2, \tag{84}$$

then Eq. (26) gives the lower bound

$$G_2[\varphi] = \tau^2/[3(1 + \tfrac{2}{5}\tau^2)]. \tag{85}$$

TABLE I. The integrated flux in a slab system of thickness 2τ.

Slab half-thickness	Lower bound	Exact result	Upper bounds		
τ	$G_2[\varphi]$	(S, ψ)	$G_4[\varphi]$	$G_5[\varphi]$	$G_6[\varphi, \theta_0]$
0.	0.	0.	0.	0.	0.
0.5	0.0758	0.0758	0.0769	0.0758	0.0759
1.0	0.2381	0.2384	0.2500	0.2384	0.2420
1.5	0.3947	0.3966	0.4286	0.3967	0.4124
2.0	0.5128	0.5180	0.5714	0.5185	0.5516
3.0	0.6522	0.6683	0.7500	0.6706	0.7321
4.0	0.7207	0.7502	0.8421	0.7558	0.8286
6.0	0.7792	0.8000	0.9231	0.8476	0.9154
10.0	0.8129	0.9000	0.9709	0.9256	0.9676
15.0	0.8242	0.9333	0.9868	0.9622	0.9858
25.0	0.8300	0.9600	0.9952	0.9852	0.9946
∞	0.8333	1.0000	1.0000	1.0000	1.0000

If $G_4[\varphi]$, given by Eq. (35), is used to compute an upper bound, and if we set the operator L equal to the identity operator, we obtain

$$G_4[\varphi] = \tau^2/[3(1 + \tfrac{1}{3}\tau^2)]. \tag{86}$$

Another upper bound can be obtained by making use of $G_5[\varphi]$, Eq. (50). With γ taken as the lowest eigenvalue of the operator in Eq. (80), i.e.,

$$\gamma = 1 + (\pi^2/4\tau^2), \tag{87}$$

the result is

$$G_5[\varphi] = \frac{4\tau^2}{\pi^2 + 4\tau^2} \times \left[1 - \frac{5(12 - \pi^2)^2}{60(12 - \pi^2) + 24(10 - \pi^2)\tau^2}\right]. \tag{88}$$

The use of $G_6[\varphi, \theta_0]$, Eq. (79), with Θ_0 also represented by a parabola, i.e.,

$$\Theta_0 \approx \theta_0 = \tau^2 - x^2, \tag{89}$$

yields yet another upper bound

$$G_6[\varphi, \theta_0] = \frac{\tau^2}{3(1 + \frac{2}{5}\tau^2)} + \tfrac{4}{3}\left(\frac{\tau^2}{5 + 2\tau^2}\right)^3. \tag{90}$$

These results are presented in tabular form as a function of τ, the slab half-thickness, in Table I. We note that $G_5[\varphi]$ is the most accurate estimate of (S, ψ). This is not unexpected since only $G_5[\varphi]$ makes use of an additional piece of information, namely, the lowest eigenvalue of the diffusion theory operator.

As a second diffusion theory example, we consider the problem of Kostin and Brooks[17] of a plane source in an infinite, homogeneous medium. If the source plane is of unit strength and is located at the origin, the appropriate diffusion equation is

$$-D[d^2\psi(x)/dx^2] + \Sigma_a \psi(x) = \delta(x), \tag{91}$$

where $\delta(x)$ is the Dirac delta function. The functional estimated by the complementary variational principles is

$$(S, \psi) = \int_{-\infty}^{\infty} dx \delta(x)\psi(x) = \psi(0), \tag{92}$$

i.e., the flux at the source plane. The exact solution for $\psi(x)$ is given by

$$\psi(x) = (2kD)^{-1}e^{-k|x|}, \tag{93}$$

where $k^2 = \Sigma_a/D$. Hence the exact value of (S, ψ) is just $1/2kD$. For simplicity, we set $kD = \frac{1}{2}$ so that we have $\psi(0) = 1$ as the exact result. As a trial function $\varphi(x)$, we choose

$$\varphi(x) = (2k'D)^{-1}e^{-k'|x|}, \tag{94}$$

which is the exact result for a medium with another absorption cross section Σ_a', with $k'^2 = \Sigma_a'/D$. A direct evaluation of $\psi(0)$ with this trial function gives, using $kD = \frac{1}{2}$,

$$\varphi(0) = (1 + \beta)^{\frac{1}{2}}, \tag{95}$$

where β, defined as

$$\beta = (k^2 - k'^2)/k'^2, \tag{96}$$

is a measure of the accuracy of the trial function. Using $F_2[\varphi]$, Eq. (22), to compute a lower bound, we find

$$F_2[\varphi] = \tfrac{1}{2}(2 - \beta)(1 + \beta)^{\frac{1}{2}}, \tag{97}$$

whereas $G_2[\varphi]$, Eq. (26), gives the lower bound

$$G_2[\varphi] = 2(1 + \beta)^{\frac{1}{2}}/(2 + \beta). \tag{98}$$

If one uses either $F_4[\varphi]$ or $G_4[\varphi]$ to compute an upper bound to $\psi(0)$, one would logically choose the operator L as the absorption cross section Σ_a. On physical grounds, one knows that the appropriate value of γ in $F_5[\varphi]$ or $G_5[\varphi]$ is also Σ_a. That is, the lowest eigenvalue of a homogeneous slab system approaches Σ_a as the slab thickness goes to infinity. Hence, for this infinite medium example, $F_4[\varphi] = F_5[\varphi]$ and $G_4[\varphi] = G_5[\varphi]$. We obtain, using either Eq. (32) or Eq. (49),

$$F_4[\varphi] = F_5[\varphi] = \tfrac{1}{2}(2 + \beta)/(1 + \beta)^{\frac{1}{2}}. \tag{99}$$

Use of the normalization independent upper-bound principles, Eqs. (35) and (50), yields

$$G_4[\varphi] = G_5[\varphi] = \tfrac{1}{2}(2 + \beta)/(1 + \beta)^{\frac{1}{2}}. \tag{100}$$

The unexpected result that Eqs. (99) and (100) are identical must be considered as fortuitous. These results, as a function of β, are given in Table II. We note that all of the variational estimates are more accurate than a direct calculation with the trial function, especially for small values of β. We also see, as expected, that $G_2[\varphi]$ is a better estimate of $\psi(0)$ than is $F_2[\varphi]$.

[17] M. D. Kostin and H. Brooks, J. Math. Phys. 5, 1691 (1964).

TABLE II. The flux at the source plane.[a]

Trial function	Direct calculation	Lower bounds		Upper bound
β	$\varphi(0)$	$F_2[\varphi]$	$G_2[\varphi]$	$F_4[\varphi] = F_5[\varphi] = G_4[\varphi] = G_5[\varphi]$
−0.9	0.316	0.458	0.575	1.741
−0.7	0.548	0.739	0.843	1.186
−0.5	0.707	0.884	0.943	1.061
−0.3	0.837	0.962	0.984	1.016
−0.1	0.949	0.9961	0.9986	1.0014
0.1	1.049	0.9964	0.9989	1.0011
0.3	1.140	0.969	0.991	1.009
0.5	1.225	0.919	0.980	1.020
0.7	1.304	0.848	0.966	1.035
0.9	1.378	0.758	0.951	1.052

[a] Exact value is unity.

More interesting examples result from the consideration of the transport equation given by Eqs. (6) and (7). It is shown in the Appendix that in certain instances the transport operator satisfies all the conditions imposed upon the operator H in the course of our development of complementary variational principles. We first analyze the escape probability problem of Francis *et al.*[3] We consider an arbitrarily shaped homogeneous convex body with a spatially independent source S of neutrons. The appropriate equation is, from Eqs. (6) and (7),

$$\rho = K(c\rho + S), \tag{101}$$

where ρ denotes the scalar flux, $c = \Sigma_s/\Sigma$, and K represents the integral operator

$$Kf(\mathbf{r}) = \frac{1}{4\pi}\int_V d\mathbf{r}' \frac{e^{-|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^2} f(\mathbf{r}'), \tag{102}$$

where the integration extends over the volume V of the body. We have set $\Sigma = 1$ in Eqs. (101) and (102), i.e., the unit of distance is taken as $1/\Sigma$. The escape probability P is defined as the ratio of the rate at which neutrons escape from the body to the rate at which they are born as source neutrons. It is easily shown from neutron conservation considerations that P can be written as

$$P = [SV - (1 - c)(\psi, S)/S]/cSV, \tag{103}$$

where ψ is related to ρ by

$$\psi = c\rho + S, \tag{104}$$

and hence satisfies the equation

$$(I - cK)\psi = S, \tag{105}$$

where I is the identity operator. The complementary variational principles $G_2[\varphi]$ and $G_6[\varphi, \theta_0]$ give us lower and upper bounds, respectively, for (ψ, S), and Eq. (103) then yields upper and lower bounds for the escape probability P. As a trial function for ψ, we use

$$\psi(\mathbf{r}) \approx \varphi(\mathbf{r}) = 1, \tag{106}$$

i.e., we assume that the variable $\psi(\mathbf{r})$ is well represented by a spatially independent function. This is an accurate representation for $c\bar{l}$ much less than one, where $\bar{l}$ is the characteristic dimension of the body, often defined as[18] $\bar{l} = 4V/A$, where A is the surface area of the body and V its volume. We further use

$$\Theta_0(\mathbf{r}) \approx \theta_0(\mathbf{r}) = 1 \tag{107}$$

as the trial function in the Rayleigh quotient for the lowest eigenvalue of the operator $I - cK$. This is a good representation for $\bar{l}$ much less than one. Suppressing the algebra, the use of $G_2[\varphi]$ given by Eq. (26) yields as an upper bound for the escape probability

$$P_U = P_0[1 - c(1 - P_0)]^{-1}, \tag{108}$$

where P_0 is the probability that a neutron from a spatially independent source distribution will escape the body without making any collisions[18] and is given by

$$P_0 = 1 - (1, K1)/V. \tag{109}$$

Equation (108) is just the result of Francis *et al.*[3] A lower bound for the escape probability is found by using $G_6[\varphi, \theta_0]$, Eq. (79). The final result is, after much algebraic manipulation,

$$P_L = P_U(1 - \Delta), \tag{110}$$

where

$$\Delta \equiv \frac{c(1 - c)(1 - P_0)(P_0 - P_1)}{P_0[1 - c(1 - P_0)]^2}, \tag{111}$$

and P_1 is the probability that a neutron from a first collision source distribution will escape the body without making any collisions. The mathematical definition of P_1 is

$$P_1 \equiv 1 - (K1, K1)/(1, K1). \tag{112}$$

It can be shown that $0 \leq P_0 \leq 1$ and hence, since $0 \leq c \leq 1$, Eq. (108) implies $0 \leq P_U \leq 1$, an inequality which the exact escape probability must satisfy. However, the same inequality does not hold for P_L.

To show this explicitly, we consider slab geometry in which case the integrals over y and z in the operator K can be performed to yield[12]

$$Kf(x) = \frac{1}{2}\int_0^\tau dx' E_1(|x - x'|) f(x'), \tag{113}$$

where the integration extends over the slab thickness. P_0 as defined by Eq. (109) becomes[18]

$$P_0 = \tau^{-1}[\tfrac{1}{2} - E_3(\tau)], \tag{114}$$

[18] K. M. Case, F. de Hoffman, and G. Placzek, *Introduction to the Theory of Neutron Diffusion* (Los Alamos Scientific Laboratory, 1953), Vol. I.

TABLE III. The escape probability for $c = 0.8$.

Slab thickness τ	Lower bound P_L	Exact result P	Upper bound P_U
0.2	0.934	0.934	0.934
0.5	0.862	0.862	0.863
1.0	0.757	0.759	0.762
2.0	0.583	0.594	0.606
4.0	0.350	0.382	0.415
8.0	0.159	0.205	0.250
16.0	0.063	0.103	0.139
30.0	0.026	0.055	0.078

TABLE IV. The escape probability for $c = 0.9$.

Slab thickness τ	Lower bound P_L	Exact result P	Upper bound P_U
0.2	0.966	0.966	0.966
0.5	0.926	0.926	0.926
1.0	0.861	0.863	0.865
2.0	0.724	0.743	0.754
4.0	0.495	0.545	0.587
8.0	0.213	0.319	0.400
16.0	0.037	0.163	0.244
30.0	−0.019	0.087	0.145

where $E_3(\tau)$ is the third-order exponential integral defined as[18]

$$E_n(\tau) \equiv \int_1^\infty d\mu \frac{e^{-\tau\mu}}{\mu^n}. \tag{115}$$

P_0 for slabs is tabulated as a function of the slab thickness by Case *et al.*[18] P_1 can be written for the slab case as

$$P_1 = \frac{1 - 2E_3(\tau) - G_{22}(\tau) - G'_{22}(\tau)}{2\tau + 2E_3(\tau) - 1}, \tag{116}$$

where $G_{nm}(\tau)$ and $G'_{nm}(\tau)$ are defined according to

$$G_{nm}(\tau) \equiv \int_0^\tau dt E_n(t)E_m(t), \tag{117}$$

$$G'_{nm}(\tau) \equiv \int_0^\tau dt E_n(t)E_m(\tau - t), \tag{118}$$

as introduced in the astrophysical literature[19] and tabulated by Anthony.[20]

If Δ, defined by Eq. (111), is greater than unity for a particular combination of c and τ, then, according to Eq. (110), P_L will be negative, a nonphysical result. We note that $\Delta = 0$ at $c = 0$ and $c = 1$ and that Δ is non-negative for $0 \leq c \leq 1$ since $(P_0 - P_1) \geq 0$. Hence Δ has a maximum in the range $0 \leq c \leq 1$. We denote by c^* the value of c for which Δ is a maximum. Equating $d\Delta/dc$ to zero for a fixed τ, we find

$$c^* = (1 + P_0)^{-1} \tag{119}$$

and

$$\Delta(c^*) = [(1 - P_0)(P_0 - P_1)]/4P_0^2. \tag{120}$$

Now, $\Delta(c^*)$ increases from zero monotonically as τ increases from zero. Defining τ^* as the value of τ for which $\Delta(c^*)$ is equal to one, we find, using large argument expansions for P_0 and P_1,

$$\tau^* = 3/(1 - \ln 2) \approx 10. \tag{121}$$

Hence P_L gives a physical (non-negative) estimate of the escape probability for any value of c if the slab thickness is less than approximately 10 mean free paths. On the other hand, if c is small enough, P_L will be non-negative for any slab thickness. Letting τ go to infinity in Δ and denoting by $\bar{c}$ the value of c for which Δ is equal to one, we find

$$\bar{c} = 3/(5 - 2 \ln 2) \approx 0.83, \tag{122}$$

and thus we conclude that P_L will be a non-negative estimate of the escape probability if c is less than approximately 0.83 for any thickness slab. Numerical examples of these two cases are given in Tables III and IV. The exact results in these tables were obtained from a numerical integration of the transport equation by Schiff and Stein.[21] We note that, for $\tau < 1$, the variational estimates of the escape probability are very accurate. This is the expected result since in this instance the trial functions used are quite good representations of the true situation. The fact that in certain cases P_L is negative should not be interpreted as a basic shortcoming of the variational method. As we have shown, P_L is negative when c is close to unity and the slab thickness is large. It is in just this case that the trial functions we have used are very poor, and thus P_L, as well as P_U, is a poor estimate of the escape probability. The fact that P_L is negative is just a manifestation of this poorness. The use of better trial functions would, of course, improve the situation.

As a final example, we consider a generalization of Marshak's[22] treatment of the Milne problem, i.e., a semi-infinite, homogeneous half-space bounded by a vacuum with a source of neutrons at infinity. The appropriate transport equation is[12]

$$\rho(x) = \frac{c}{2} \int_0^\infty dx' E_1(|x - x'|)\rho(x'), \tag{123}$$

[19] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).

[20] G. W. Anthony, Hanford Atomic Products Operation Report HW-49188, Richland, Washington (1957).

[21] D. Schiff and S. Stein, Westinghouse Electric Corporation Report WAPD-149, Pittsburgh, Pa. (1956).

[22] R. E. Marshak, Phys. Rev. **71**, 688 (1947).

where the notation is the same as in Eq. (101). We isolate the growing component of $\rho(x)$ by writing

$$\rho(x) = e^{\nu x} - q(x), \tag{124}$$

where ν satisfies

$$2\nu/c = \ln[(1+\nu)/(1-\nu)], \tag{125}$$

and arrive at the equation for $q(x)$:

$$q(x) - \frac{c}{2}\int_0^\infty dx' E_1(|x-x'|)q(x') = S(x), \tag{126}$$

with the source term given by

$$S(x) = (c/2\nu)\{E_1(x) - e^{\nu x}E_1[(1+\nu)x]\}. \tag{127}$$

One knows[18] that, for large x, $q(x)$ decays exponentially according to

$$q(x) \xrightarrow[x\to\infty]{} \Gamma e^{-\nu x}, \tag{128}$$

where Γ is a constant related to the extrapolated endpoint x_0, according to[18]

$$x_0 = -(2\nu)^{-1}\ln\Gamma. \tag{129}$$

We use the complementary variational principles to obtain upper and lower bounds for x_0. By multiplying Eq. (126) by $e^{\nu x}$ and integrating over x from 0 to y, and then letting y increase beyond bound, it can be shown that (S, q) and Γ are related according to[23]

$$\Gamma = \left\{\frac{c}{2\nu}\left[\frac{1}{\nu}\ln\left(\frac{1}{1-\nu}\right) - \frac{1}{c}\right] - (S,q)\right\}\left\{\frac{c}{\nu}\left[\frac{1}{1-\nu} - \frac{1}{c}\right]\right\}^{-1}. \tag{130}$$

The functionals G_2 and G_5 can be used to bound (S, q), Eq. (130) then used to bound Γ, and finally Eq. (129) used to bound x_0.

As a trial function $\bar{q}(x)$ for $q(x)$, we use the large x behavior, i.e.,

$$q(x) \approx \bar{q}(x) = e^{-\nu x}. \tag{131}$$

The application of $G_2[\bar{q}]$, Eq. (26), then gives a lower bound for x_0, which can be written

$$(x_0)_L = -\frac{1}{2\nu}\ln\left(\left\{\left[\frac{\ln(1+\nu)}{\nu} - \frac{1}{1+\nu}\right]^2 - \frac{1}{4\nu^2}\ln^2(1-\nu^2)\right\}\left[\frac{1}{\nu}\left(\frac{1}{1-\nu^2} - \frac{1}{c}\right)\ln(1-\nu^2)\right]^{-1}\right). \tag{132}$$

To use $G_5[\bar{q}]$ to obtain an upper bound for x_0, we require the lowest eigenvalue of the operator in Eq. (126). As argued earlier, this is $\Lambda_0 = 1 - c$. The final result for $(x_0)_U$, while explicit, is algebraically very complex and hence will not be given here. Table V

TABLE V. cx_0 for the Milne problem.

c	Lower bound	Case *et al.*	Mark	Upper bound
0.1	0.853826	0.8539	0.8590	0.853829
0.2	0.78476	0.7851	0.7843	0.78479
0.3	0.74836	0.7491	0.7486	0.74843
0.4	0.7298	0.7305	0.7300	0.7300
0.5	0.7202	0.7207	0.7204	0.7206
0.6	0.7150	0.7155	0.7154	0.7156
0.7	0.7121	0.7127	0.7126	0.7132
0.8	0.7104	0.7113	0.7112	0.7126
0.9	0.7094	0.7106	0.7106	0.7148

gives some numerical results for the bounds on cx_0 as a function of c, as well as the results of Case *et al.*[18] and Mark[24] computed numerically from the exact relation

$$cx_0 = \frac{c^2}{2\nu}\int_0^1 d\mu\left[1 + \frac{c\mu^2}{1-\mu^2}\right] \times \frac{\tanh^{-1}(\nu\mu)}{[(1 - c\mu\tanh^{-1}\mu)^2 + (\pi c\mu/2)^2]}. \tag{133}$$

We note that the bounds are exceedingly tight for small values of c, and somewhat looser as c approaches unity. We also see that the results based on a numerical integration of Eq. (133) fall in certain instances (small c) outside the bounds computed here, indicating an error, albeit small, in the evaluation of Eq. (133). An examination of the integrand of this equation shows that for small values of c the numerical integration is very difficult to perform and this fact no doubt accounts for the discrepancies in Table V. In view of these results, cx_0 was recomputed from Eq. (133) with great care exercised for small values of c. These results, which, together with a discussion of the methods used, are reported elsewhere,[23] were found to fall inside the bounds given in Table V.

APPENDIX

We consider some properties of the transport operator, Eq. (6), appropriate to a nonmultiplying homogeneous medium (i.e., Σ_s and Σ are independent of position) with isotropic scattering. Denoting this operator by H, we have

$$Hf(\mathbf{r}) = f(\mathbf{r}) - \frac{c}{4\pi}\int_V d\mathbf{r}'\frac{e^{-|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^2}f(\mathbf{r}'), \tag{A1}$$

where we have set Σ equal to unity, i.e., we have taken $1/\Sigma$ as the unit of distance and have defined $c = \Sigma_s/\Sigma$. Since the medium is nonmultiplying and all cross sections are non-negative, we have $0 \leq c \leq 1$.

[23] G. C. Pomraning and K. D. Lathrop, Nucl. Sci. Eng. **29**, 305 (1967).

[24] J. C. Mark, National Research Council of Canada Report CRT-338, Montreal (1945).

We immediately note that H is a Hermitian operator and hence has all real eigenvalues. Since both the operator and eigenvalues are real, the eigenfunctions can be normalized such that they are real. We further exclude all but real functions from the domain of H. This restriction is nonessential and is imposed only for simplicity. Our first nontrivial result is that all the eigenvalues of H are non-negative. Consider, for an arbitrary function f in the domain of H,

$$(f, Hf) = \int_V d\mathbf{r} f^2(\mathbf{r}) - \frac{c}{4\pi} \int_V d\mathbf{r} f(\mathbf{r}) \times \int_V d\mathbf{r}' \frac{e^{-|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^2} f(\mathbf{r}'). \quad \text{(A2)}$$

We rewrite this as

$$(f, Hf) = \int_V d\mathbf{r} \int_V d\mathbf{r}' f(\mathbf{r}) \times \left[\delta(\mathbf{r}-\mathbf{r}') - \frac{c}{4\pi} \frac{e^{-|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^2}\right] f(\mathbf{r}'), \quad \text{(A3)}$$

where $\delta(\mathbf{r}) = \delta(x)\delta(y)\delta(z)$, $\delta(\xi)$ being the usual Dirac delta function. Following Wing,[25] we introduce Fourier transforms into Eq. (A3). In particular, we write the term in brackets in Eq. (A3) as the inverse of its Fourier transform. We have

$$\delta(\mathbf{s}) = \frac{1}{(2\pi)^3} \int d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{s}} \quad \text{(A4)}$$

and

$$\frac{e^{-|\mathbf{s}|}}{|\mathbf{s}|^2} = \frac{1}{(2\pi)^3} \int d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{s}} \frac{4\pi}{|\mathbf{k}|} \tan^{-1}|\mathbf{k}|. \quad \text{(A5)}$$

With these results Eq. (A3) can be written

$$(f, Hf) = \frac{1}{(2\pi)^3} \int d\mathbf{k} \left[1 - \frac{c}{|\mathbf{k}|} \tan^{-1}|\mathbf{k}|\right] |\omega(\mathbf{k})|^2, \quad \text{(A6)}$$

where

$$\omega(\mathbf{k}) \equiv \int_V d\mathbf{r} f(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}}. \quad \text{(A7)}$$

Now,

$$1 - (c/|k|) \tan^{-1}|k| \geq 0 \quad \text{for} \quad c \leq 1, \quad \text{(A8)}$$

and thus

$$(f, Hf) \geq 0 \quad \text{for} \quad c \leq 1. \quad \text{(A9)}$$

Since the eigenvalues of H, say Λ_j, are given by

$$\Lambda_j = (\Theta_j, H\Theta_j)/(\Theta_j, \Theta_j), \quad \text{(A10)}$$

where Θ_j is the corresponding eigenfunction, Eq. (A9) implies

$$\Lambda_j \geq 0, \quad \text{(A11)}$$

as was to be shown.

To proceed further, we represent $\mathbf{r}$ by Cartesian coordinates and restrict the domain of H to those functions which depend only upon x, i.e., we consider the slab geometry transport equation. Performing the integrals over y and z in Eq. (A1), we find[12]

$$Hf(x) = f(x) - \frac{c}{2} \int_0^\tau dx' E_1(|x-x'|) f(x'), \quad \text{(A12)}$$

where $E_1(\xi)$ is the first-order exponential integral, defined as[18]

$$E_n(\xi) \equiv \int_1^\infty d\mu \frac{e^{-\xi\mu}}{\mu^n}, \quad \text{(A13)}$$

and the limits of integration in Eq. (A12) correspond to integration over the slab thickness. We now show that the functionals $J_m[f]$ and $J_n[f]$, given by Eqs. (60) and (65), are bounded below by Λ_0, the smallest eigenvalue of H. To this end we introduce the operator K, defined as

$$Kf(x) = \frac{c}{2} \int_0^\tau dx' E_1(|x-x'|) f(x'), \quad \text{(A14)}$$

which has eigenfunctions and eigenvalues according to

$$K\psi_j = \Omega_j \psi_j. \quad \text{(A15)}$$

The kernel of the operator K is symmetric and square integrable

$$\int_0^\tau dx \int_0^\tau dx' \left|\frac{c}{2} E_1(|x-x'|)\right|^2 < \infty, \quad \text{(A16)}$$

and thus K is a Hilbert–Schmidt operator.[15] Wing[25] has shown that K is a non-negative operator and hence has only non-negative eigenvalues. Also, the kernel $(c/2)E_1(|x-x'|)$ is not degenerate and thus the operator K has an infinite number of eigenvalues.[15] If we denote the largest eigenvalue of K by Ω_0, we know on physical grounds (the largest eigenvalue corresponds to the reactivity of the fundamental mode in a critical reactor) that $0 \leq \Omega_0 \leq c$. Ω_0 approaches zero as τ approaches zero, and Ω_0 approaches c as τ approaches infinity.

The theorem of Hilbert–Schmidt[15] states that if $f(x)$ is a square integrable function in the interval $(0, \tau)$, then the function Kf can be expanded in an absolutely and uniformly convergent series with respect to the orthonormal functions $\psi_j(x)$, according to

$$Kf = \sum_{j=0}^\infty a_j \psi_j, \quad \text{(A17)}$$

where

$$a_j = (\psi_j, Kf). \quad \text{(A18)}$$

[25] G. M. Wing, *An Introduction to Transport Theory* (John Wiley & Sons, Inc., New York, 1962).

If we define

$$b_j = (\psi_j, f), \tag{A19}$$

we have, using Eq. (A15),

$$a_j = (K\psi_j, f) = \Omega_j b_j. \tag{A20}$$

Noting that $H = I - K$, where I is the identity operator, we write Eq. (60) as

$$J_m[f] = 1 - (f, Kf)/(f,f), \tag{A21}$$

and using Eqs. (A17) through (A20), we find

$$J_m[f] = 1 - \left(\sum_{j=0}^{\infty} \Omega_j b_j^2\right) \Big/ (f,f). \tag{A22}$$

With the use of Bessel's inequality[15]

$$(f,f) \geq \sum_{j=0}^{\infty} b_j^2, \tag{A23}$$

Eq. (A22) can be written

$$J_m[f] \geq 1 - \Omega_0 + \left[\sum_{j=1}^{\infty} b_j^2(\Omega_0 - \Omega_j)\right] \Big/ \left(\sum_{j=0}^{\infty} b_j^2\right). \tag{A24}$$

Clearly the eigenfunctions of the operators H and K are identical, and the eigenvalues of H, the Λ_j, are related to those of K according to

$$\Lambda_j = 1 - \Omega_j, \tag{A25}$$

with Λ_0 being the smallest eigenvalue of H. We have $1 - c \leq \Lambda_0 \leq 1$ from our earlier physical argument. In terms of the Λ_j, Eq. (A24) becomes

$$J_m[f] \geq \Lambda_0 + \left[\sum_{j=1}^{\infty} b_j^2(\Lambda_j - \Lambda_0)\right] \Big/ \left(\sum_{j=0}^{\infty} b_j^2,\right), \tag{A26}$$

and thus $J_m[f] \geq \Lambda_0$ since $\Lambda_j \geq \Lambda_0$. A similar analysis on Eq. (65) shows that $J_n[f] \geq \Lambda_0$; and hence for the operator H defined by Eq. (A12) we have

$$\bar{m} = \bar{n} = \Lambda_0 \tag{A27}$$

as a rigorous result.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Systems of Observables in Axiomatic Quantum Mechanics*

STANLEY GUDDER
Department of Mathematics, University of Wisconsin, Madison, Wisconsin

(Received 17 April 1967)

Systems of observables are considered in an axiomatic framework for quantum mechanics which generalizes the usual Hilbert space formulation. Familiar concepts such as complete systems of observables and superselection rules are generalized and it is shown that many of the Hilbert space theorems carry over to this abstract formalism. Also functions of observables are considered and some theorems due to John von Neumann are generalized.

1. INTRODUCTION

IN this paper systems of observables are considered in an axiomatic framework for quantum mechanics similar to that studied by Mackey,[1] Varadarajan,[2] Jauch,[3] Piron,[4] and others.[5] The advantages of considering systems of observables in this abstract setting are at least twofold. First, it is well known that, in general, not all self-adjoint operators on the Hilbert space of state vectors correspond to observables. This class of physically significant self-adjoint operators depends upon the superselection rules present in the quantum-mechanical system.[6] Since all the superselection rules are not as yet known,[7] one cannot specify precisely all the physically significant operators. In the abstract framework one considers only the observables themselves and is not shackled by the presence of physically nonsignificant operators. In the Hilbert space formulation, for example, a set of commuting self-adjoint operators S is said to be complete if it generates a maximal Abelian–von Neumann algebra.[8] But this definition depends upon the interaction of S with nonobservable operators and is thus physically not entirely relevant. In the abstract setting such definitions involve only observable quantities. Secondly, in the abstract framework all extraneous structure is stripped away leaving only the bare physical essentials. In this way the methods of the proofs and their relations to the basic defining axioms become clear and more transparent. The main purpose of this paper is to summarize some of the established work concerning systems of observables and to add to these some results which we feel are interesting and will perhaps be useful. The reader should notice that many of the results we give generalize known theorems concerning systems of observables in a Hilbert space.[9]

2. NOTATION AND KNOWN RESULTS

We now summarize our notation. A *proposition system* $L = \{a, b, c, \cdots\}$ is an orthocomplemented partially ordered set. For this and other definitions the reader is referred to Mackey or Varadarajan.[10] The elements of L are called *propositions* and all sets of propositions will henceforth be assumed to be subsets of one fixed proposition system L. If $a \leq b'$, we say that a and b are *disjoint*. We say a, b *split* and write $a \leftrightarrow b$ if there are mutually disjoint propositions a_1, b_1, c such that $a = a_1 \vee c$, $b = b_1 \vee c$. Splitting propositions describe propositions which are physically compatible. We assume that L satisfies the following:

Axiom 1: For every $a, b, c \in L$ which mutually split, $a \leftrightarrow b \vee c$.

A subset B of L is a *Boolean σ algebra* if B is a proposition system which is a distributive lattice. All the results in this section may be found in Varadarajan[11] (also see Ramsey).[12]

Theorem 2.1: A collection of propositions mutually split if and only if the collection is contained in a Boolean σ algebra.

An *observable* x is a σ homomorphism from the Borel sets $B(R)$ of the real line R to L. We denote the range of x by $R(x)$. If x is an observable and u a Borel function on R, we define the observable $u(x)$ by $u(x)(E) = x[u^{-1}(E)]$ for all $E \in B(R)$. We say that an

* While conducting research for this paper, the author was partially supported by a NSF grant.

[1] G. Mackey, *The Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, Inc., New York, 1963).

[2] V. Varadarajan, Comm. Pure Appl. Math. **15**, 217 (1962).

[3] J. Jauch, Helv. Phys. Acta **33**, 711 (1960).

[4] C. Piron, Helv. Phys. Acta **37**, 439 (1964).

[5] G. Emch and C. Piron, J. Math. Phys. **4**, 469 (1963); J. Jauch and C. Piron, Helv. Phys. Acta **36**, 827 (1963); J. Jauch and B. Misra, *ibid.* **34**, 699 (1961); N. Zierler, Pac. J. Math. **11**, 1151 (1961); S. Gudder, Trans. Am. Math.Soc. **119**, 428 (1965); S. Gudder, Pac. J. Math. **19**, 81 (1966).

[6] G. Wicks, E. Wigner, and A. Wightman, Phys. Rev. **88**, 101 (1952); see also Ref. 3.

[7] E.g., it is not known whether the lepton number is a superselection rule or not.

[8] See Ref. 3.

[9] See Ref. 3.

[10] See Refs. 1 and 2.

[11] See Ref. 2.

[12] A. Ramsey, J. Math. Mech. **15**, 227 (1966).

observable x is *smaller* than an observable y or y is larger than x (denoted by $x \subset y$) if $R(x) \subset R(y)$.

Theorem 2.2: $x \subset y$ if and only if there is a Borel function u such that $x = u(y)$.

A Boolean σ algebra $B \subset L$ is *countably generated* if there is a countable set $C \subset B$ such that the smallest Boolean σ algebra containing C is B itself. A proposition system L is *countably generated* if every Boolean σ algebra in L is countably generated. It follows from a theorem in Riesz and Nagy's book[13] that the proposition system consisting of all closed subspaces of a separable Hilbert space is countably generated. For simplicity, in the sequel we assume that L is countably generated, although the reader should note that many of our results do not require this assumption. In fact, the next two theorems and any results depending upon them are the only results which require L to be countably generated.

Theorem 2.3: A subset of L is the range of an observable if and only if it is a Boolean σ algebra.

Two observables x, y are *simultaneous* (written $x \leftrightarrow y$) if $R(x) \leftrightarrow R(y)$ and a collection of observables $\{x_\alpha : \alpha \in A\}$ is simultaneous if $x_\alpha \leftrightarrow x_\beta$ $\alpha, \beta \in A$. In the special case in which observables are realized by self-adjoint operators, this is equivalent to the operators commuting.

Theorem 2.4: $\{x_\alpha : \alpha \in A\}$ are simultaneous observables if and only if there is an observable x and Borel functions u_α such that $x_\alpha = u_\alpha(x)$, $\alpha \in A$.

This theorem holds for a countable number of observables even if L is not countably generated and it holds for two observables even without Axiom 1.[14] This last statement also applies to two propositions in Theorem 2.1.

3. EQUIVALENT OBSERVABLES AND GENERATORS

In the sequel $x, y, z, \cdots$ denote observables on a fixed proposition system L. If $x \subset y$ and $y \subset x$, then x and y have the same range. In this case we say that x and y are *equivalent* and write $x \sim y$. Of course, equivalent observables need not be equal. Notice that $\sim$ is an equivalence relation. We now give a description of equivalence in terms of one Borel function.

Theorem 3.1: Let x and y be observables. If there is a Borel function u and a set $E \in B(R)$ such that (i) $y(E) = 1$, (ii) u is 1–1 on E, (iii) $x = u(y)$, (iv) $u(\Lambda) \in B(R)$ for every $\Lambda \in B(R)$, $\Lambda \subset E$, then $x \sim y$. Conversely, if $x \sim y$, then there is an $E \in B(R)$ and a Borel function u satisfying (i), (ii), (iii).

Proof: Since $x = u(y)$, $x \subset y$. Let $\Lambda \in B(R)$. Then since u is 1–1, we have $\Lambda \cap E = [u^{-1}(u(\Lambda \cap E))] \cap E$. Since $y(E) = 1$ and $u(\Lambda \cap E) \in B(R)$, we have

$$\begin{aligned} y(\Lambda) &= y(\Lambda \cap E) = y([u^{-1}(u(\Lambda \cap E))] \cap E) \\ &= y[u^{-1}(u(\Lambda \cap E))] = u(y)(u(\Lambda \cap E)) \\ &= x(u(\Lambda \cap E)). \end{aligned}$$

Thus $R(x) = R(y)$ and $x \sim y$. Conversely, if $x \sim y$, there are Borel functions u and v such that $x = u(y)$ and $y = v(x)$. Therefore $y = v(u(y)) = (v \circ u)(y)$. Let $E = \{\lambda : (v \circ u)(\lambda) = \lambda\}$. Then, by Lemma 3.1,[15] $y(E) = 1$. Let $\lambda_1, \lambda_2 \in E$ and suppose $u(\lambda_1) = u(\lambda_2)$. Then $\lambda_1 = (v \circ u)(\lambda_1) = (v \circ u)(\lambda_2) = \lambda_2$ and u is 1–1 on E.

Corollary 3.2: If there is a 1–1 Borel function u whose inverse is a Borel function and if $x = u(y)$, then $x \sim y$.

Let $X = \{x_\alpha : \alpha \in A\}$ be a collection of simultaneous observables. If an observable x satisfies $x_\alpha = u_\alpha(x)$ where u_α are Borel functions for all $\alpha \in A$, then x is called a *generator* of $\{x_\alpha : \alpha \in A\}$. Note by Theorem 2.4 a generator always exists for any family of simultaneous observables. Notice also that if x is a generator of X and $x \subset y$, then y is a generator of X. We say that x is a *minimal generator* of X if for any generator y of X we have $x \subset y$. Notice that if a minimal generator exists, it is unique to within an equivalence. That is, if x_1 and x_2 are minimal generators for X, then $x_1 \sim x_2$. We now prove existence.

Theorem 3.3: If $X = \{x_\alpha : \alpha \in A\}$ is a collection of simultaneous observables, X has a minimal generator.

Proof: By Theorem 2.1 there is a Boolean σ algebra containing $\bigcup \{R(x_\alpha) : \alpha \in A\}$. Let B be the intersection of all Boolean σ algebras containing $\bigcup R(x_\alpha)$. Then B is a Boolean σ algebra and by Theorem 2.3 is the range of an observable x. Applying Theorem 2.2, x is a generator of X and clearly is a minimal generator.

Corollary 3.4: If x is a minimal generator of X, then $R(x)$ is the smallest Boolean σ algebra containing $\bigcup \{R(x_\alpha) : \alpha \in A\}$.

4. COMPLETE SYSTEMS OF OBSERVABLES AND SUPERSELECTION RULES

A system of simultaneous observables

$$X = \{x_\alpha : \alpha \in A\}$$

[13] F. Riesz and Sz. Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955), p. 358.

[14] See Ref. 12.

[15] See S. Gudder, Trans. Am. Math. Soc. **119**, 428 (1965).

is *complete* if $x \leftrightarrow X$ implies $x \in X$. A simple Zorn's lemma argument shows that every observable is contained in a complete system of observables. A Boolean σ algebra is *maximal* if it is not properly contained in a larger Boolean σ algebra. Again using Zorn's lemma, every Boolean σ algebra is contained in a maximal one. An observable x is *maximal* if $R(x)$ is maximal. Clearly every observable is smaller than some maximal observable. Maximal observables are generalizations of self-adjoint operators with simple spectrum (cf. Jauch[16]).

Theorem 4.1: x is maximal if and only if $y \leftrightarrow x$ implies $y \subset x$.

Proof: Suppose x is maximal and $y \leftrightarrow x$. Then $R(x) \cup R(y)$ is contained in a Boolean σ algebra. But since $R(x)$ is maximal, $R(x) \cup R(y) \subset R(x)$ and $R(y) \subset R(x)$. Thus $y \subset x$. Conversely, suppose x is not maximal and is thus strictly smaller than an observable y. Then $y \leftrightarrow x$, and we do not have $y \subset x$.

Corollary 4.2: x is maximal if and only if $y \leftrightarrow x$ implies there is a Borel function u such that $y = u(x)$.

We first obtain some consequences of completeness.

Theorem 4.3: If $X = \{x_\alpha : \alpha \in A\}$ is a system of simultaneous observables, statements 1, 2, 3, 4 are equivalent and so are 6, 7. If X is complete, all the following statements hold:

(1) Every generator of X is maximal.
(2) There is only one Boolean σ algebra B containing $\bigcup \{R(x_\alpha) : \alpha \in A\}$ and hence B is maximal.
(3) All generators of X are equivalent.
(4) Every generator of X is a minimal generator.
(5) If x is a generator of X, then $y \leftrightarrow X$ if and only if $y \leftrightarrow x$.
(6) $\bigcup \{R(x_\alpha) : \alpha \in A\}$ is a maximal Boolean σ algebra.
(7) If x is a generator of X, then

$$R(x) = \bigcup \{R(x_\alpha) : \alpha \in A\}.$$

Proof: We first show that X complete $\Rightarrow$ (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5). Suppose X is complete and x is a generator of X. If $y \leftrightarrow x$, then $y \leftrightarrow X$ and hence $y \in X$. Thus $x \subset y$ and by Theorem 4.1 x is maximal. (1) $\Rightarrow$ (2). Let B be the smallest Boolean σ algebra containing $\bigcup R(x_\alpha)$. By Theorem 2.3 there is an observable x such that $B = R(x)$, and by Theorem 2.2 x is a generator of X. By hypothesis x is maximal, and hence B is the only Boolean σ algebra containing $\bigcup R(x_\alpha)$. (2) $\Rightarrow$ (3) and (3) $\Rightarrow$ (4) are trivial. (4) $\Rightarrow$ (5). Let x be a generator of X. Then by hypothesis x is a minimal generator. If $y \leftrightarrow x$, then clearly $y \leftrightarrow X$. Now suppose $y \leftrightarrow X$. Then $R(y) \cup \bigcup \{R(x_\alpha) : \alpha \in A\}$ is contained in a Boolean σ algebra B, and by Corollary 3.4 $R(x) \subset B_1$ and hence $y \leftrightarrow x$. Now it is trivial that (4) $\Rightarrow$ (1). Now suppose X is complete, and suppose $\bigcup R(x_\alpha) \subset B$ where B is a Boolean σ algebra. If $a \in B$, define the observable x_a by $x_a(\{1\}) = a$, $x_a(\{0\}) = a'$. Then $x_a \leftrightarrow X$ and hence $x_a \in X$. Therefore $a \in \bigcup R(x_\alpha)$ and $\bigcup R(x_\alpha) = B$. Thus (6) holds. That (6) and (7) are equivalent is straightforward.

It can be shown by examples that (1) to (4) are not equivalent to (5), (6), or (7) and that none of these statements imply completeness. We now obtain a characterization of completeness.

Theorem 4.4: $X = \{x_\alpha : \alpha \in A\}$ is a complete system of simultaneous observables if and only if there is a maximal observable x such that $X = \{u(x) : u$ a Borel function$\}$.

Proof: To prove necessity suppose X is complete and let x be a generator of X. By Theorem 4.3 x is maximal. Now by definition $X \subset \{u(x) : u$ a Borel function$\} = Y$. If $y \in Y$, then $y \leftrightarrow X$ and hence $y \in X$. Thus $X = Y$. For sufficiency suppose $y \leftrightarrow X$. Then $y \leftrightarrow x$ and, by Corollary 4.2, $y \in Y = X$.

Corollary 4.5: x is maximal if and only if $\{u(x) : u$ a Borel function$\}$ is complete.

The *center* Z of L is the collection of propositions which split with all propositions. Notice that Z is a Boolean σ algebra. A *superselection rule* (Jauch and Piron call them essential observables)[17] is an observable whose range is in Z. If $R(x) = Z$, then x is a maximal superselection rule in that x is larger than any superselection rule. There exists a maximal superselection rule and it is unique to within equivalence.

Theorem 4.6: Z is the intersection of the collection of maximal Boolean σ algebras.

Proof: Let $\{B_\alpha : \alpha \in A\}$ be the set of maximal Boolean σ algebras. Since $Z \leftrightarrow B_\alpha$, by Theorem 2.1 $Z \cup B_\alpha$ is contained in a Boolean σ algebra and hence $Z \cup B_\alpha = B_\alpha$. Thus $Z \subset B_\alpha$ and $Z \subseteq \bigcap \{B_\alpha : \alpha \in A\}$. Now suppose $a \in \bigcap \{B_\alpha : \alpha \in A\}$. Since every proposition is in a maximal Boolean σ algebra, $a \in Z$ and hence $\bigcap \{B_\alpha : \alpha \in A\} \subseteq Z$.

In a similar way we prove:

Theorem 4.7: The set of superselection rules is the intersection of the collection of all complete sets of simultaneous observables.

[16] See Ref. 3.

[17] See J. Jauch and C. Piron, Helv. Phys. Acta **36**, 827 (1963).

It follows from Theorem 4.4, Corollary 4.5, and Theorem 4.7 that:

Theorem 4.8: x is a superselection rule if and only if x is a Borel function of every maximal observable.

5. STANDARD OBSERVABLES

If $A \subset L$, we define the *splitting set* of A as $A^s = \{b \in L: b \leftrightarrow A\}$. Similarly, if θ is the set of all observables and $X \subset \theta$, we define the *simultaneous set* of X as $X^s = \{y \in \theta: y \leftrightarrow X\}$. Clearly, if $A \subset B (X \subset Y)$, then $B^s \subset A^s (Y^s \subset X^s)$ and $A \subset A^{ss} (X \subset X^{ss})$. It is obvious that the elements of $A(X)$ mutually split (are simultaneous) if and only if $A \subset A^s (X \subset X^s)$. The proof of the following theorem is straightforward.

Theorem 5.1: (1) If $A \subset L$, the following statements are equivalent: (i) The elements of A mutually split; (ii) $A^{ss} \subset A^s$; (iii) A^{ss} is a Boolean σ algebra.
(2) If a Boolean σ algebra A is maximal, then $A = A^{ss}$.
(3) A Boolean σ algebra A is maximal if and only if $A = A^s$.
(4) A set of simultaneous observables X is complete if and only if $X = X^s$.
(5) An observable x is maximal if and only if $\{x\}^s = \{u(x): u \text{ a Borel function}\}$.

An observable y is *affiliated* with an observable x if $z \leftrightarrow x$ implies $z \leftrightarrow y$, i.e., $y \in \{x\}^{ss}$. Notice that y is a superselection rule if and only if it is affiliated with every observable. Notice also that if x is maximal, then every observable affiliated with x is a Borel function of x. We shall see later (self-adjoint operators in a Hilbert space) that the converse need not hold; that is, there are nonmaximal observables whose only affiliated observables are Borel functions of the observable. If an observable x has the property that the only observables affiliated with x are Borel functions of x, we call x a *standard observable* (i.e., x is standard if $\{x\}^{ss} = \{u(x): u \text{ a Borel function}\}$). It is easy to find examples of nonstandard observables. For instance, the inverse mappings of measurable functions on a measure space are always nonstandard observables. We now characterize standard observables.

Theorem 5.2: The following statements are equivalent:
(1) x is a standard observable.
(2) $R(x)$ is the intersection B of the collection of maximal Boolean σ algebras containing $R(x)$.
(3) If $a \notin R(x)$, there is a b such that $b \leftrightarrow R(x)$ and $b \leftrightarrow\!\!\!\!\!\!\rightarrow a$.
(4) $R(x) = R(x)^{ss}$.
(5) The intersection X of the collection of complete systems of simultaneous observables containing x is $\{u(x): u \text{ a Borel function}\}$.

Proof: (1) $\Rightarrow$ (2). Certainly $R(x) \subset B$. Now suppose $a \in B$ and $y \leftrightarrow x$. Then $R(y) \cup R(x)$ is in a maximal Boolean σ algebra B_1 and $a \in B_1$. Therefore $x_a \leftrightarrow y$ where x_a is defined as in Theorem 4.3. By (1) x_a is a function of x and hence $a \in R(x)$. (2) $\Rightarrow$ (3). If $a \notin R(x)$, then there is a maximal Boolean σ algebra B_0 such that $R(x) \subset B_0$ and $a \notin B_0$. Therefore $a \leftrightarrow\!\!\!\!\!\!\rightarrow B_0$ and there is a $b \in B_0$ such that $a \leftrightarrow b$ but $b \leftrightarrow\!\!\!\!\!\!\rightarrow R(x)$. (3) $\Rightarrow$ (4). We know that $R(x) \subset R(x)^{ss}$. Now if $a \notin R(x)$, then by (3) there is a b such that $b \leftrightarrow R(x)$ and $b \leftrightarrow\!\!\!\!\!\!\rightarrow a$. Thus $a \notin R(x)^{ss}$. Hence $R(x)^{ss} \subset R(x)$ and (4) holds. (4) $\Rightarrow$ (5). Again $\{u(x): u \text{ a Borel function}\} \subset X$, so suppose $y \in X$. Let $a \in R(y)$. Now if $b \leftrightarrow R(x)$, then x_b and x are in a complete system of simultaneous observables Y and hence $x_a \in Y$. Therefore $a \leftrightarrow b$ and hence $a \in R(x)^{ss}$. By (4) $a \in R(x)$. Thus $y \subset x$ and by Theorem 2.2 there is a Borel function u such that $y = u(x)$. Hence $X \subset \{u(x): u \text{ a Borel function}\}$ and (5) holds. (5) $\Rightarrow$ (1). Now $\{u(x): u \text{ a Borel function}\} \subset \{x\}^{ss}$. Now suppose $y \in \{x\}^{ss}$ and that X_0 is a complete system of simultaneous observables containing x. Since $X_0 \subset \{x\}^s$, we have $y \leftrightarrow X_0$. Hence $y \in X_0$ and by (5) $y \in \{u(x): u \text{ a Borel function}\}$, so (1) holds.

Notice that in the above proof we did not require that L is countably generated. It follows from the previous theorem that, if x is standard, any observable equivalent to x is standard. Also the identity observable I is standard if and only if there are no superselection rules, other than multiples of I.

Theorem 5.3: Let X be a system of simultaneous observables. Then there is an observable x such that $X^{ss} = \{x\}^{ss}$ and, if x is standard, $X^{ss} = \{u(x): u \text{ a Borel function}\}$.

Proof: Let x be a minimal generator of X. If $y \in X^{ss}$ and $z \in \{x\}^s$, then $y \leftrightarrow z$ so $X^{ss} \subset \{x\}^{ss}$. Conversely, suppose $y \in \{x\}^{ss}$ and $z \in X^s$. We now show $z \in R(x)^s$. If $A = \{a \in R(x): z \leftrightarrow a\}$, then it is easily seen that A is a Boolean σ algebra containing $\bigcup \{R(y): y \in X\}$ and hence $A = R(x)$. Thus $z \in R(x)^s$ and $y \leftrightarrow z$. Therefore $y \in X^{ss}$ and $\{x\}^{ss} \subset X^{ss}$. The last part follows by definition.

A proposition system L is *standard* if every observable on L is standard. We can get necessary and sufficient conditions for L to be standard by replacing $R(x)$ by an arbitrary Boolean σ algebra in Theorem 5.2. It follows from the next theorem due to

John von Neumann[18] that the proposition system of closed subspaces of a separable Hilbert space H is standard.

Theorem 5.4 (von Neumann): If a self-adjoint operator A commutes with every self-adjoint operator which commutes with a self-adjoint operator B on a separable Hilbert space, then A is a Borel function of B.

Von Neumann's theorem is really a little more general in that A may be a closed operator with dense domain; however, this is the form we will find useful. We see that Theorem 5.2 gives necessary and sufficient conditions on an abstract proposition system (which need not be countably generated) for the counterpart of von Neumann's theorem to hold. In fact, Theorem 5.2 can be used to prove Theorem 5.4. To illustrate this we will prove von Neumann's theorem for the special case in which B has countable spectrum. Our theorem does not require that the Hilbert space is separable, so it is not a consequence of von Neumann's theorem, which does not hold in the nonseparable case.[19] We first need two lemmas.

Lemma 5.5: Let ϕ be a vector in a Hilbert space H and let a be a closed subspace of H. If P_ϕ and P_a are the orthogonal projections on ϕ and a, respectively, then $P_\phi \leftrightarrow P_a$ (i.e., P_ϕ and P_a commute) if and only if $\phi \in a$ or $\phi \in a'$ (here $a' = a^\perp$).

Proof: Sufficiency is trivial. To prove necessity suppose $P_\phi \leftrightarrow P_a$ and $\phi \notin a$. Then $\phi = \phi_1 + \phi_2$ where $\phi_1 \in a$, $\phi_2 \in a'$, and $\phi_2 \neq 0$. Now

$$\langle \phi_1, \phi\rangle\phi = P_\phi\phi_1 = P_\phi P_a\phi_1 = P_a P_\phi\phi_1 = P_a\langle\phi_1, \phi\rangle\phi = \langle\phi_1, \phi\rangle\phi_1.$$

Since $\phi \neq \phi_1$, we have $\langle\phi_1, \phi\rangle = 0$, and hence $0 = \langle\phi, \phi_1\rangle = \langle\phi_1, \phi_1\rangle + \langle\phi_2, \phi_1\rangle = \langle\phi_1, \phi_1\rangle$. Thus $\phi_1 = 0$ and $\phi \in a'$.

Lemma 5.6: An orthogonal projection on a (not necessarily separable) Hilbert space H is standard.

Proof: Let P_a be the orthogonal projection on the closed subspace a, and denote the observable corresponding to P_a by P_a also. Then $R(P) = \{0, 1, a, a'\}$. We now show that condition (3) of Theorem 5.2 holds for $R(P)$. Let b be a closed subspace such that $b \notin R(P)$ and let P_b be the corresponding projection. We now show there is a projection P_1 such that $P_1 \leftrightarrow P_a$ but $P_1 \not\leftrightarrow P_b$. If $P_b \not\leftrightarrow P_a$, we are finished, so suppose $P_b \leftrightarrow P_a$. Suppose that for every $\phi \in H$ with $P_\phi \leftrightarrow P_a$ we have $P_\phi \leftrightarrow P_b$. Then, applying Lemma 5.5 if $\phi \in a$, we have $\phi \in b$ or $\phi \in b'$. Now if there is a nonzero $\phi_1 \in a$ such that $\phi_1 \in b$ and a nonzero $\phi_2 \in a$ such that $\phi_2 \in b'$, then $\phi_1 + \phi_2 \in a$ but $\phi_1 + \phi_2 \notin b$ or b', which is a contradiction. Thus $a < b$ or $a < b'$. Similarly, for every $\phi \in a'$ we have $\phi \in b$ or $\phi \in b'$ and again $a' < b$ or $a' < b'$. If $a < b$, then $a' \not< b$, so $a' < b'$ and $b > a$, a contradiction. Similarly, if $a < b'$, then $a' \not< b'$, so $a' < b$ and $b' < a$, a contradiction. Therefore there is a $\phi \in H$ such that $P_\phi \leftrightarrow P_a$ but $P_\phi \not\leftrightarrow P_b$. It thus follows from Theorem 5.2 that P_a is standard.

Corollary 5.7: If a self-adjoint operator A commutes with every self-adjoint operator which commutes with an orthogonal projection P, then A has the form αI, αP, or $\alpha(I - P)$ where α is a real constant.

We now prove our theorem.

Theorem 5.8: Let A be a self-adjoint operator with countable spectrum on a (not necessarily separable) Hilbert space H. If a self-adjoint operator A_1 commutes with every self-adjoint operator which commutes with A, then A_1 is a Borel function of A.

Proof: Let B be the range of the resolution of identity for A. Since A has countable spectrum, there is a countable collection of disjoint nonzero closed subspaces $C = \{a_i\colon\ i = 1, 2, \cdots\}$ such that $\vee a_i = 1 = H$, and every nonzero $a \in B$ is the supremum of a subcollection of C. To show that A is standard, let b be a closed subspace not in B and let P_b be the corresponding orthogonal projection. Suppose for every P_ϕ such that $P_\phi \leftrightarrow P_{a_i}$, $i = 1, 2, \cdots$, we have $P_\phi \leftrightarrow P_b$. Then, as in the proof of Lemma 5.6, we have $a_i < b$ or $a_i < b'$, $i = 1, 2, \cdots$. Let $a_{i(n)}$ be the subcollection of C which satisfies $a_{i(n)} < b$ and $a_{j(n)}$ the subcollection of C which satisfies $a_{j(n)} < b'$. Then $Va_{i(n)} < b$ and $Va_{j(n)} < b'$. Therefore $(Va_{i(n)}) \vee (Va_{j(n)}) < b \vee b' = 1$, which is a contradiction, since every $a_i \in C$ is included in one of the suprema on the left-hand side. As in Lemma 5.6, we conclude that A is standard.

Theorems of this type are of basic importance in considerations of von Neumann algebras and systems of observables in Hilbert space.[20]

[18] See Ref. 13, p. 351.

[19] H. Nakano, Proc. Phys. Math. Soc. Japan **21**, 713 (1939).

[20] See Ref. 3.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Internal Holonomy Groups of Yang-Mills Fields*

HENDRICUS G. LOOS
Douglas Advanced Research Laboratories, Huntington Beach, California and Department of Physics, University of California at Riverside, Riverside, California

(Received 9 March 1967)

Classical Yang–Mills potentials define a Lie group, the internal holonomy group, which is analogous to the ordinary holonomy group defined by the Christoffel symbols in general relativity. The internal holonomy group is an important tool for the dissection and study of the Yang–Mills equations and of local gauge theory in general. A number of theorems on internal holonomy groups is presented, together with some applications.

INTRODUCTION

GROUP-THEORETICAL discussions pertaining to local gauge theories (i.e., gauge theories of the Yang–Mills[1] type) usually involve the gauge group only. However, there exists another group which is defined by the gauge potentials and which is indispensable for most work on the classical Yang–Mills equations. One arrives at this group by recognizing that the gauge potentials can be used to define "parallel displacement" of multiplets to neighboring events; this is possible because of the transformation properties of the gauge potentials. In fact, the gauge potentials serve to define parallelity (or "equivalence") of multiplets at neighboring events in precisely the same way as the Christoffel symbols define parallelity of vectors in Riemann space. When the gauge potentials are known, one can execute a parallel displacement of multiplets around a closed curve in event space, from x^κ back to x^κ; this produces a linear transformation of multiplets at x^κ. Doing this for all closed curves in event space passing through x^κ results in a continuous set of linear multiplet transformations at x^κ. This set turns out to be a Lie group, here called the internal holonomy group $\mathcal{H}(x^\kappa)$, in analogy with the ordinary holonomy group in the Riemann space of general relativity.[2] Just as the ordinary holonomy group can be used to classify solutions of the Einstein field equations, the internal holonomy group $\mathcal{H}$ provides a classification of solutions of the Yang–Mills equations. In fact, whether or not a gauge field has short range depends largely on the internal holonomy group, not on the gauge group. The classical current algebra associated with the gauge field has much more to do with $\mathcal{H}$ than with the gauge group, and $\mathcal{H}$ has an important bearing on conservation laws for the charges which can be constructed from the gauge field.

The event space used here is the Minkowski space of special relativity. x^κ are Cartesian inertial coordinates of events; $\kappa = 0, 1, 2, 3$. The summation convention is used. Partial differentiation with respect to x^κ is written as ∂_κ. □ denotes end of proof.

I. GEOMETRIC IMPLICATIONS OF THE GAUGE POTENTIALS

Yang and Mills[1] considered a multiplet field $\psi(x^\kappa)$, subject to gauge transformations $S(x^\kappa)$ belonging to the gauge group $\mathcal{G}$:

$$\psi' = S^{-1}\psi. \tag{1}$$

Here, $\mathcal{G}$ may be any gauge group, not just $O(3)$, as in the work of Yang and Mills.[1] The gauge potentials[3] of Yang and Mills transform under (1) as

$$B'_\mu = S^{-1}B_\mu S + (i/\epsilon)S^{-1}\partial_\mu S. \tag{2}$$

Writing $\Gamma_\mu = i\epsilon B_\mu$, (2) takes the form

$$\Gamma'_\mu = S^{-1}(\Gamma_\mu S - \partial_\mu S). \tag{3}$$

We call the multiplets $\psi(x^\kappa)$ and $\psi(x^\kappa + dx^\kappa)$ *equivalent* (or parallel), if

$$\psi(x^\kappa + dx^\kappa) - \psi(x^\kappa) = \Gamma_\kappa\psi\, dx^\kappa; \tag{4}$$

the right-hand side is evaluated at x^κ. Such equivalence is invariant under gauge transformations, because of the transformation laws (1) and (3). Hence, *the classical gauge potentials* $\Gamma_\mu(x^\kappa)$ *provide a gauge-invariant definition of equivalence of multiplets at neighboring events.* This fact is the basis for geometric considerations in local gauge theory.[4]

* Work conducted at the Douglas Advanced Research Laboratories under company-sponsored Independent Research and Development funds.

[1] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[2] See J. F. Schell, J. Math. Phys. **2**, 202 (1961); J. M. Goldberg and R. P. Kerr, J. Math. Phys. **2**, 327, 332 (1961).

[3] We use the name "gauge potentials" in a more general sense than Yang and Mills; it may mean the coefficients of expansion of the B_μ in terms of generators of $\mathcal{G}$, or the B_μ or Γ_μ themselves.

[4] That such considerations are possible was already clear from the work of R. Utiyama, Phys. Rev. **101**, 1597 (1956); T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961); S. I. Fickler, Ph.D thesis, Syracuse University (1961); R. L. Arnowitt and S. I. Fickler, Phys. Rev. **127**, 1821 (1962).

Let S in (1) belong to the fundamental representation[5] of $\mathfrak{G}$. Then, we may call the multiplet ψ at x^κ an internal vector, i.e., a vector in some "internal" linear vector space associated with the event x^κ. For clarity, we associate a *separate* internal space with every event x^κ. Denoting internal vector components by indices $a, b, c, \cdots$ ranging from 1 to n, Eqs. (1), (3), and (4) may be written

$$\begin{aligned} \psi_{a'} &= S^a_{a'}\psi_a, \\ \Gamma^{b'}_{\mu a'} &= S^a_{a'}(\Gamma^b_{\mu a}S^{b'}_b - \partial_\mu S^{b'}_a), \\ \psi_a(x^\kappa + dx^\kappa) - \psi_a(x^\kappa) &= \Gamma^b_{\kappa a}\psi_b\, dx^\kappa, \end{aligned} \tag{5}$$

the last equation expressing equivalence of $\psi_a(x^\kappa + dx^\kappa)$ and $\psi_a(x^\kappa)$. This equivalence may be used to define a (gauge-) covariant derivative of an internal vector field $\psi_a(x^\kappa)$:

$$\nabla_\kappa\psi_a = \partial_\kappa\psi_a - \Gamma^b_{\kappa a}\psi_b; \tag{6}$$

$dx^\kappa\nabla_\kappa\psi_a$ is the difference between $\psi_a(x^\kappa + dx^\kappa)$ and an internal vector at $x^\kappa + dx^\kappa$ equivalent to $\psi_a(x^\kappa)$. Under gauge transformations (5), $\nabla_\kappa\psi_a$ transforms as ψ_a, i.e., as an internal vector:

$$\nabla_\kappa\psi_{a'} = S^a_{a'}\nabla_\kappa\psi_a.$$

For a contravariant internal vector field $\psi^a(x^\kappa)$ one has

$$\nabla_\kappa\psi^a = \partial_\kappa\psi^a + \Gamma^a_{\kappa b}\psi^b;$$

this follows from the demand that covariant differentiation follows Leibnitz' rule, and that covariant differentiation of a scalar is the same as ordinary differentiation. For multiplets P^b_a transforming under a tensor representation of $\mathfrak{G}$, we have

$$\nabla_\kappa P^b_a = \partial_\kappa P^b_a - \Gamma^c_{\kappa a}P^b_c + \Gamma^b_{\kappa c}P^c_a; \tag{7}$$

this can be shown by writing P^b_a as the sum of products of co- and contravariant internal vectors. Similar expressions may be found for the covariant derivative of internal tensors of any valence.[6] In the sequel, we mainly discuss multiplets which are covariant internal vectors ψ_a or mixed internal tensors P^b_a (internal operators). Then, it is convenient to suppress the internal indices and write in matrix notation

$$\nabla_\kappa\psi = \partial_\kappa\psi - \Gamma_\kappa\psi \tag{8}$$

for (6), and

$$\nabla_\kappa P = \partial_\kappa P - [\Gamma_\kappa, P] \tag{9}$$

for (7). *Also, from here on, the $n \times n$ matrix fields $\Gamma_\kappa(x^\lambda)$ denote the gauge potentials for the fundamental representation[5] of $\mathfrak{G}$.*

The gauge fields (i.e., the $F_{\mu\nu}$ of Yang and Mills[1]) are

$$\phi_{\kappa\lambda} = \partial_\kappa\Gamma_\lambda - \partial_\lambda\Gamma_\kappa - [\Gamma_\kappa, \Gamma_\lambda]; \tag{10}$$

under (1), the $\phi_{\kappa\lambda}$ transform as internal operators[7]

$$\phi'_{\kappa\lambda} = S^{-1}\phi_{\kappa\lambda}S.$$

By using (8) it is easy to show that[8]

$$\nabla_{[\lambda}\nabla_{\kappa]}\psi = -\tfrac{1}{2}\phi_{\lambda\kappa}\psi, \tag{11}$$

and similarly, by using (9), that

$$\nabla_{[\lambda}\nabla_{\kappa]}P = -\tfrac{1}{2}[\phi_{\lambda\kappa}, P]. \tag{12}$$

From (11) one finds[8]

$$\begin{aligned} \nabla_{[\kappa}\nabla_\lambda\nabla_{\mu]}\psi &= -\tfrac{1}{2}\nabla_{[\kappa}\phi_{\lambda\mu]}\psi \\ &= -\tfrac{1}{2}(\nabla_{[\kappa}\phi_{\lambda\mu]})\psi - \tfrac{1}{2}\phi_{[\lambda\mu}\nabla_{\kappa]}\psi, \end{aligned}$$

but also

$$\nabla_{[\kappa}\nabla_\lambda\nabla_{\mu]}\psi = -\tfrac{1}{2}\phi_{[\kappa\lambda}\nabla_{\mu]}\psi = -\tfrac{1}{2}\phi_{[\lambda\mu}\nabla_{\kappa]}\psi.$$

Hence, it must be that

$$(\nabla_{[\kappa}\phi_{\lambda\mu]})\psi = 0$$

for every covariant internal vector field $\psi(x^\kappa)$, so that we have the "Bianchi identities"

$$\nabla_{[\kappa}\phi_{\lambda\mu]} = 0. \tag{13}$$

Equations (13) are also the integrability conditions for the partial differential equations (10) for $\Gamma_\kappa(x^\kappa)$.

The equivalence of internal vectors at neighboring events, defined by the gauge potentials $\Gamma_\kappa(x^\lambda)$, can be used to execute an equivalence displacement of internal vectors along a curve in event space. The question is whether such equivalence displacement is path dependent. If it is not, we have everywhere in event space

$$\nabla_\kappa\psi = 0, \tag{14}$$

for n linearly independent internal vectors ψ. Covariant differentiation of (14) with respect to x^λ and

[5] The fundamental representation of a Lie group is here meant to be lowest-dimensional faithful representation in a linear vector space over the complex numbers.

[6] Of course, all of this is a simple modification of covariant differentiation of ordinary vectors and tensors in a geometric manifold endowed with an ordinary linear connection. The only difference is that there the $\Gamma^\lambda_{\kappa\mu}$ act on tangent space, while here the $\Gamma^b_{\kappa a}$ act on internal space. In fact, our case has been considered by R. König, J. Deutschen Math. Verein. **28**, 213 (1920); **41**, 169 (1932); T. Sibata, Hiroshima Univ. J. Sci. **5**, 83 (1934); V. Hlavatý, Rend. Cir. Math. Palermo **59**, 1 (1935); J. A. Schouten, Proc. Koninkl. Akad. Amsterdam **27**, 407 (1924); L. Schlesinger, Math. Ann. **99**, 413 (1928); and in modern differential geometry it is a special case of fiber bundles with a linear connection.

[7] The Γ_κ and $\phi_{\kappa\lambda}$ may be called, respectively, the parameters of the internal linear connection, and the internal-curvature tensor operator.

[8] Square brackets around indices denote alternation: $T_{[\kappa\lambda]} = (1/2!)(T_{\kappa\lambda} - T_{\lambda\kappa})$; $V_{[\kappa\lambda\mu]} = (1/3!)(V_{\kappa\lambda\mu} + V_{\mu\kappa\lambda} + V_{\lambda\mu\kappa} - V_{\lambda\kappa\mu} - V_{\kappa\mu\lambda} - V_{\mu\lambda\kappa})$, etc. We generally follow the notation of J. A Schouten, *Ricci Calculus* (Springer-Verlag, Berlin, 1954), 2nd ed.

alternation over λ and κ gives the integrability conditions

$$0 = \nabla_{[\lambda}\nabla_{\kappa]}\psi = -\tfrac{1}{2}\phi_{\lambda\kappa}\psi,$$

using (11). Since this is true everywhere for n linearly independent internal vectors ψ, it follows that

$$\phi_{\lambda\kappa} = 0. \tag{15}$$

This proves—

Theorem 1: In order to have nonvanishing gauge fields $\phi_{\kappa\lambda}(x^\mu)$, the equivalence transport of internal vectors defined by the gauge potentials $\Gamma_\kappa(x^\lambda)$ must be path dependent.

Next we investigate the conditions under which the gauge potentials can be transformed away. We want to consider internal base transformations $S(x^\kappa)$ which are not restricted to belong to $\mathfrak{G}$. Since the internal space is complex n dimensional, the general internal base transformations $S(x^\kappa)$ are nonsingular complex $n \times n$ matrices. We simply extend the transformation laws (1) and (3) to such $S(x^\kappa)$. If $\Gamma'_\kappa = 0$ one must have from (3)

$$\partial_\mu S = \Gamma_\mu S. \tag{16}$$

The integrability conditions for (16) are found by partial differentiation with respect to x^κ, alternation over κ and μ and elimination of first derivatives of S by (16); the result found is (15). This shows—

Theorem 2: If the gauge fields (10) vanish everywhere in event space, then the gauge potentials can be transformed away by an internal base transformation.

All these results are familiar in differential geometry of manifolds with a linear connection.

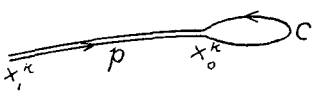
II. INTRODUCTION OF THE INTERNAL HOLONOMY GROUP

The path dependence of equivalence transport of internal vectors [defined by the gauge potentials $\Gamma_\kappa(x^\lambda)$] can be studied best by executing the transport around closed curves in event space. Let C be such a loop through the event x_0^κ; a sense of circumscription of C is provided, and C is, from here on, restricted to be piecewise continuously differentiable. Taking n linearly independent covariant internal vectors around C by equivalence displacement, from x_0^κ back to x_0^κ, results in a linear transformation $H(C)$ of covariant internal vectors at x_0^κ:

$$\psi' = H(C)\psi.$$

Doing this for all closed curves C through the event x_0^κ, one gets a set $\mathcal{H}(x_0^\kappa)$ of linear internal transformations. The inverse of $H(C)$ is produced by equivalence displacement around C in the opposite sense. The composition $H(C_2)H(C_1)$ is the element $H(C_1 + C_2)$, where $C_1 + C_2$ denotes the loop consisting of C_1 and C_2. Hence $\mathcal{H}(x_0^\kappa)$ is a group, and it must be a subgroup of the full complex linear group $GL(n, c)$. From here on, let the matrix elements of the gauge potentials $\Gamma_\kappa(x^\lambda)$ be analytic functions of the event coordinates x^λ. Then, continuous changes of the loop C produce continuous changes in the matrix elements of $H(C)$, and since any loop C through x_0^κ can by continuous modifications be changed into any other loop through x_0^κ, $\mathcal{H}(x_0^\kappa)$ is a connected continuous subgroup of $GL(n, c)$. Hence, $\mathcal{H}(x_0^\kappa)$ is a connected Lie group. We call $\mathcal{H}(x_0^\kappa)$ the "internal holonomy group" (at x_0^κ) of the gauge field.[9] A loop C through x_0^κ can be modified into a loop through x_1^κ by choosing an arbitrary curve p from x_0^κ to x_1^κ, and by forming a spoon-shaped loop with "bowl" C and "handle" p (Fig. 1). If equivalence displacement of internal vectors from x_1^κ to x_0^κ produces a transformation U, the element of $\mathcal{H}(x_1^\kappa)$, belonging to the spoonlike loop is

FIG. 1. Spoonlike loop.



$$U^{-1}H(C, x_0^\kappa)U. \tag{17}$$

If all loops C through x_0^κ are made into spoons with the same handle p, U is the same for all $H(C, x_0^\kappa)$, and one has[10]

$$U^{-1}\mathcal{H}(x_0^\kappa)U \subset \mathcal{H}(x_1^\kappa).$$

By inverting the use of events x_0^κ and x_1^κ, one has

$$U\mathcal{H}(x_1^\kappa)U^{-1} \subset \mathcal{H}(x_0^\kappa).$$

Hence,

$$\mathcal{H}(x_1^\kappa) = U^{-1}\mathcal{H}(x_0^\kappa)U, \tag{18}$$

so that one finds[11]—

Theorem 3: The internal holonomy groups at different events are isomorphic.

[9] The proper differential geometric name for $\mathcal{H}$ is the holonomy group of the linear connection Γ_κ. We use here the name "internal holonomy group," to distinguish it from the "external" holonomy group, belonging to the "external" connection $\Gamma^\mu_{\kappa\lambda}$ (the Christoffel symbols), in case we consider as event space the Riemann space of general relativity. Most of the work presented here holds for that case as well; we then have two linear connections, Γ_κ and $\Gamma^\mu_{\kappa\lambda}$, and the meaning of general covariant differentiation ∇_κ must be extended to involve the $\Gamma^\mu_{\kappa\lambda}$ as well.

[10] The symbol $\subset$ does not exclude the possibility of equality.

[11] This is well known for ordinary holonomy groups; see for instance J. A. Schouten, Ref. 8, p. 362. It is known to hold as well for the holonomy groups of linear connections for fiber bundles, see A. Nyenhuis, Koninkl. Akad. Wetenshap. Amsterdam, Proc. A57, No. 1, 17 (1954), which covers our case.

Theorem 3 implies—

Lemma 1: There exists an internal base (i.e., an internal basis at every event) such that the set of matrices $\mathfrak{L}(x^\kappa)$ constituting the Lie algebra of $\mathfrak{H}(x^\kappa)$ is independent of x^κ.

Proof: The matrix U in (18) depends on the path from x_0^κ to x_1^κ. For fixed x_1^κ one can, for every x^κ, assign a path from x_1^κ to x^κ such that the matrix elements of U are analytic functions of x^κ [because the $\Gamma_\kappa(x^\lambda)$ are analytic]. Then, the internal base transformation $U^{-1}(x^\kappa)$ will result in a set of matrices $\mathfrak{L}(x^\kappa)$ independent of x^κ, and the $\Gamma_\kappa(x^\lambda)$ remain analytic. □

Next we consider the Lie algebra of $\mathfrak{H}$. We need—

Lemma 2: If a matrix representation Ω of a Lie algebra contains the matrices $P(s)$ where s is a continuous parameter, and if $(d/ds)P(s)$ exist as matrices, then they belong to Ω.

Proof[12]*:* Since Ω is a linear vector space, the matrices

$$Q(s, \Delta s) = (1/\Delta s)(P(s + \Delta s) - P(s)), \quad \Delta s \neq 0,$$

belong to Ω. Expand $Q(s, \Delta s)$ in terms of a basis L_i, $i = 1 \cdots m$, of Ω: $Q(s, \Delta s) = q^i(s, \Delta s)L_i$. If

$$\lim_{\Delta s \to 0} Q(s, \Delta s)$$

exists as a matrix, the matrix elements of $Q(s, \Delta s)$ converge in the Cauchy sense. But, by Cramer's rule, the $q^i(s, \Delta s)$ are continuous functions of the matrix elements of $Q(s, \Delta s)$, so that the numbers $q^i(s, \Delta s)$ also converge in the Cauchy sense, as $\Delta s \to 0$. Hence, a limit exists for $q^i(s, \Delta s)$ and it follows that

$$dP/ds = \lim_{\Delta s \to 0} Q(s, \Delta s)$$

belongs to Ω. □

Lemma 3: The covariant derivatives of order p of the gauge fields, alternated over a pair of adjacent indices,

$$\nabla_{\sigma_1} \cdots \nabla_{\sigma_{k-1}} \nabla_{[\sigma_k} \nabla_{\sigma_{k+1}]} \nabla_{\sigma_{k+2}} \cdots \nabla_{\sigma_p} \phi_{\kappa\lambda},$$

can be expressed in terms of derivatives of the gauge fields of order $p - 2$.

Proof: From (12) one has

$$\nabla_{\sigma_1} \cdots \nabla_{\sigma_{k-1}} \nabla_{[\sigma_k} \nabla_{\sigma_{k+1}]} \nabla_{\sigma_{k+2}} \cdots \nabla_{\sigma_p} \phi_{\kappa\lambda} = -\tfrac{1}{2} \nabla_{\sigma_1} \cdots \nabla_{\sigma_{k-1}} [\phi_{\sigma_k \sigma_{k+1}}, \nabla_{\sigma_{k+2}} \cdots \nabla_{\sigma_p} \phi_{\kappa\lambda}],$$

which consists of derivatives of order $p - 2$. □

[12] The author acknowledges a helpful discussion of this point with Dr. D. A. Jacobson.

We now pass to—

Theorem 4: The Lie algebra $\mathfrak{L}(x_1^\kappa)$ of the internal holonomy group $\mathfrak{H}(x_1^\kappa)$ consists of all real linear combinations of the tensor components of the internal operators

$$\phi_{\kappa\lambda}, \nabla_\mu \phi_{\kappa\lambda}, \nabla_\nu \nabla_\mu \phi_{\kappa\lambda}, \cdots, \tag{19}$$

evaluated at x_1^κ.

Proof[13]*:* First, take for the loop C an infinitesimal parallelogram with sides $d_1 x^\kappa$ and $d_2 x^\kappa$. Taking internal vectors around by equivalence transport, one finds the well-known result

$$H(C) = I + \tfrac{1}{2} \phi_{\kappa\lambda} \, df^{\kappa\lambda}, \tag{20}$$

where I is the identity transformation, and $df^{\kappa\lambda} = 2d_1 x^{[\kappa} d_2 x^{\lambda]}$ is the bivector of the infinitesimal loop C. Equation (20) shows that $\phi_{\kappa\lambda}(x_1^\mu)$ belongs to the Lie algebra $\mathfrak{L}(x_1^\mu)$ of $\mathfrak{H}(x_1^\mu)$.

Next, consider a spoonlike loop C_1 as in Fig. 1, taking an infinitesimal bivector $df^{\kappa\lambda}$ for the bowl C. Then, by (17)

$$H(C_1) = I + \tfrac{1}{2} df^{\kappa\lambda} U^{-1} \phi_{\kappa\lambda}(x_0^\mu) U. \tag{21}$$

For $x_1^\kappa = x_0^\kappa + dx^\kappa$ one has

$$U = I - dx^\kappa \Gamma_\kappa, \tag{22}$$

and since the $\Gamma_\kappa(x^\lambda)$ are analytic,

$$\phi_{\kappa\lambda}(x_0^\nu) = \phi_{\kappa\lambda}(x_1^\nu) - dx^\mu \partial_\mu \phi_{\kappa\lambda}(x_1^\nu).$$

Hence, (21) becomes to first order,

$$H(C_1) = I + \tfrac{1}{2} df^{\kappa\lambda} (\phi_{\kappa\lambda}(x_1^\nu) - dx^\mu \nabla_\mu \phi_{\kappa\lambda}(x_1^\nu)). \tag{23}$$

For a spoon with infinitesimal bowl but with a finite handle p we subdivide p into m intervals $\Delta^k x^\kappa$ and apply (23) in successive steps, from x_0^κ to x_1^κ; the result is

$$H = I + \tfrac{1}{2} df^{\kappa\lambda} (1 - \Delta^m x^\nu \nabla_\nu) \cdots (1 - \Delta^1 x^\mu \nabla_\mu) \phi_{\kappa\lambda},$$

which, upon taking the limit $m \to \infty$ becomes

$$H = I + \tfrac{1}{2} df^{\kappa\lambda} (1 - q^\mu \nabla_\mu + q^{\nu\mu} \nabla_\nu \nabla_\mu - \cdots) \phi_{\kappa\lambda}, \tag{24}$$

evaluated at x_1^κ, where

$$\begin{aligned} q^\mu &= x^\mu(s_1) - x^\mu(0), \\ q^{\nu\mu} &= \int_0^{s_1} q'^\nu q^\mu \, ds, \\ q^{\sigma\nu\mu} &= \int_0^{s_1} q'^\sigma q^{\nu\mu} \, ds, \\ &\cdots \\ q'^\mu &= dq^\mu/ds, \end{aligned} \tag{25}$$

[13] We follow A. Nyenhuis, Ref. 11, with some modifications.

and $x^\mu(s)$ is the parametric description of the handle p; $x^\mu(0) = x_0^\mu$, $x^\mu(s_1) = x_1^\mu$. Of course, the handle p must be restricted such that (24) converges. For straight handles p, (24) and (25) become[14]

$$H = I + \tfrac{1}{2} df^{\kappa\lambda}\big(1 - s_1 v^\mu \nabla_\mu + (s_1^2/2!) v^\nu v^\mu \nabla_\nu \nabla_\mu - \cdots\big)\phi_{\kappa\lambda}, \tag{26}$$

where $v^\mu = q'^\mu(0)$. In order to investigate convergence, we calculate U of (21) by applying (22) in succession; the result is the ordered exponential (operator ordering from right to left with increasing s is indicated by T)

$$U(s) = T\exp\left(-\int_0^s \Gamma_\kappa \, dx^\kappa\right) \overset{\text{def}}{=} 1 - \int_0^s \Gamma_\kappa \frac{dx^\kappa}{ds}\, ds + \int_0^s \Gamma_\kappa(s_1) \frac{dx^\kappa}{ds_1}\, ds_1 \int_0^{s_1} \Gamma_\mu(s_2) \frac{dx^\mu}{ds_2}\, ds_2 - \cdots. \tag{27}$$

Since the matrix elements of $\Gamma_\kappa(s)$ are analytic, the matrix elements of $U(s)$ are analytic functions of s as well by an argument similar to the one showing that the matrix elements of $\exp As$ are analytic, when A is a fixed matrix. Equations (26) and (21) give

$$(1 - s v^\mu \nabla_\mu + [s^2/2!] v^\nu v^\mu - \cdots)\phi_{\kappa\lambda}\big|_{x_1} = U^{-1}(s)\phi_{\kappa\lambda}(s)U(s). \tag{28}$$

The matrix elements on the right-hand side of (28) are analytic functions of s, since the matrix elements of $U^{-1}(s)$, $\phi_{\kappa\lambda}(s)$, and $U(s)$ are analytic. Hence, the power-series expansion on the left is absolute convergent for $|s| < s^*$, where s^* is positive.

Taking derivatives of (26) with respect to s_1, putting $s_1 = 0$, and using Lemma 2, one sees that $v^\mu \nabla_\mu \phi_{\kappa\lambda}$, $v^\nu v^\mu \nabla_\nu \nabla_\mu \phi_{\kappa\lambda}$, etc., belong to $\mathfrak{L}$. Taking linear combinations of the operators $v^\nu v^\mu \nabla_\nu \nabla_\mu \phi_{\kappa\lambda}$ with appropriately chosen v^μ, one finds

$$p^{\nu\mu} \nabla_\nu \nabla_\mu \phi_{\kappa\lambda} \in \mathfrak{L},$$

where $p^{\nu\mu}$ is symmetric but otherwise arbitrary. It follows that[15]

$$\nabla_{(\nu} \nabla_{\mu)} \phi_{\kappa\lambda} \in \mathfrak{L}.$$

Application of the same process to higher derivatives shows that

$$\nabla_{(\sigma_1} \cdots \nabla_{\sigma_k)} \phi_{\kappa\lambda} \in \mathfrak{L}\ ; \quad k \text{ any natural number.} \tag{29}$$

Now it will be shown that also the nonsymmetrized covariant derivatives of $\phi_{\kappa\lambda}$ belong to $\mathfrak{L}$. We already know that $\phi_{\kappa\lambda}$, $\nabla_\mu \phi_{\kappa\lambda}$, and $\nabla_{(\nu} \nabla_{\mu)} \phi_{\kappa\lambda}$ belong to $\mathfrak{L}$. By Lemma 3, $\nabla_{[\nu} \nabla_{\mu]} \phi_{\kappa\lambda} \in \mathfrak{L}$. Hence $\nabla_\nu \nabla_\mu \phi_{\kappa\lambda} \in \mathfrak{L}$. For the third derivatives, we already have

$$\nabla_{(\sigma} \nabla_\nu \nabla_{\mu)} \phi_{\kappa\lambda} \in \mathfrak{L}. \tag{30}$$

[14] For the remainder of this proof, absence of an argument x^κ means that the expression is evaluated x_1^κ.

[15] Round brackets around indices denote mixing (symmetrization): $T_{(\kappa\lambda)} = (1/2!)(T_{\kappa\lambda} + T_{\lambda\kappa})$, etc.; see J. A. Schouten, Ref. 8, p. 14.

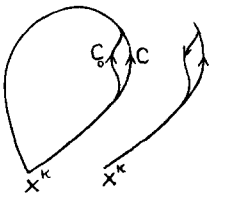


FIG. 2. Loops C, C_0, and $C - C_0$.

Equation (30) contains the sum of all permutations of $\nabla_\sigma \nabla_\nu \nabla_\mu$ operating on $\phi_{\kappa\lambda}$. By Lemma 3, we can interchange any two adjacent indices of covariant differentiations at the expense of an element which we know to belong to $\mathfrak{L}$. In this manner we can change any permutation of $\nabla_\sigma \nabla_\nu \nabla_\mu$ occurring in (30) into the ordering $\nabla_\sigma \nabla_\nu \nabla_\mu$. Doing this for all terms in (30) one finds $\nabla_\sigma \nabla_\nu \nabla_\mu \phi_{\kappa\lambda} \in \mathfrak{L}$. The same procedure works for derivatives of any order. Hence, all elements (19) belong to $\mathfrak{L}$.

It remains to be shown that $\mathfrak{L}$ cannot contain anything else but real linear combinations of the internal operators (19). Suppose a finite loop C gives an infinitesimal transformation H. Because of the continuity of $\Gamma_\kappa(x^\lambda)$ there must exist an infinitesimally different loop C_0 which gives the identity transformation. Then $C - C_0$ (Fig. 2) produces H as well; this loop is a spoon with infinitesimal bowl, or a sum of such spoons. Consider such a spoon (Fig. 3). Then, the element H for the spoon with handle terminated at y_1 on p is given by (24) with $q^\mu, q^{\nu\mu}, \cdots$ evaluated along the section of p from x_0^κ to y_1^κ. Since for y_1^κ close enough to x_0^κ the arc from x_0^κ to y_1^κ is approximated arbitrarily close by a straight line, and since for a straight handle the series (24) becomes a power series like (28), there must be a choice for $y_1^\kappa \neq x_0^\kappa$ such that the series (24) is convergent. For such a choice of y_1^κ the holonomy-group element belonging to the spoon with handle terminated at y_1^κ is $I + df L_1$, where $L_1 \in \mathfrak{L}^*(y_1^\kappa)$, and $\mathfrak{L}^*$ is the set of real linear combinations of the internal operators (19). Since $\mathfrak{K} \subset GL(n, c)$, there are at every event at most $2n^2$ linearly independent operators among (19). Hence, L_1 must be expressible as a linear combination of a *finite* number of operators (19):

$$L_1 = p^{\kappa\lambda}\phi_{\kappa\lambda} + p^{\mu\kappa\lambda}\nabla_\mu \phi_{\kappa\lambda} + \cdots. \tag{31}$$

Extension of the handle from y_1^κ to y_2^κ (along p) gives

$$H(y_2^\kappa) = I + \tfrac{1}{2} df (1 - q^\mu \nabla_\mu + q^{\nu\mu}\nabla_\nu \nabla_\mu - \cdots) L_1(s); \tag{32}$$

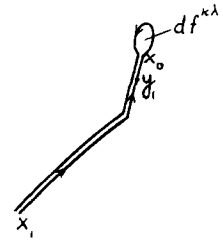


FIG. 3. General spoon-like loop with infinitesimal bowl $df^{\kappa\lambda}$.

where now $q^{\mu}, q^{\nu\mu}, \cdots$ are calculated for the section of p from y_1^{κ} to y_2^{κ} and $L_1(s)$ is calculated from (31), taking the $p^{\kappa\lambda}$, $p^{\mu\kappa\lambda}$, etc., constant. Again, the series (32) converges for y_2^{κ} close enough to y_1^{κ}. The process is repeated by extending the handle to y_3^{κ}, etc. Points of the handle at which the tangent is discontinuous are included in the set of points y_k^{κ}. Because of analyticity of the $\Gamma_{\kappa}(x^{\lambda})$ a finite number of finite steps will be sufficient to reach x_1^{κ}. It follows that for the whole spoon, with handle ending at x_1^{κ}, one has

$$H(x_1^{\kappa}) = I + dfL, \quad L \in \mathfrak{L}^*(x_1^{\kappa}). \quad \square \tag{33}$$

In case every element of $\mathfrak{L}(x_1^{\mu})$ is a linear combination of the $\phi_{\kappa\lambda}(x_1^{\mu})$, the internal holonomy group at x_1^{κ} is called perfect.[16] Theorem 4 implies—

Lemma 4: For the internal base of Lemma 1,

$$[\Gamma_{\kappa}(x^{\lambda}), \mathfrak{L}] \subset \mathfrak{L}. \tag{34}$$

Proof: Choose an internal base such that the matrix set $\mathfrak{L}(x^{\kappa})$ is the same for all x^{κ}; this is possible by Lemma 1. By Theorem 4, $\phi_{\kappa\lambda} \in \mathfrak{L}$ and $\nabla_{\mu}\phi_{\kappa\lambda} = \partial_{\mu}\phi_{\kappa\lambda} - [\Gamma_{\mu}, \phi_{\kappa\lambda}] \in \mathfrak{L}$. By Lemma 2, $\partial_{\mu}\phi_{\kappa\lambda} \in \mathfrak{L}$. Hence, $[\Gamma_{\mu}, \phi_{\kappa\lambda}] \in \mathfrak{L}$. By a similar argument we find $[\Gamma_{\mu}, \nabla_{\nu}\phi_{\kappa\lambda}] \in \mathfrak{L}$, etc., for higher derivatives. Application of Theorem 4 gives (34). $\square$

The considerations in this chapter so far are independent of a relation between the gauge group $\mathfrak{G}$ and the internal holonomy group $\mathfrak{K}$. In fact, we might as well have chosen gauge potentials $\Gamma_{\kappa}(x^{\lambda})$ as four analytic but otherwise arbitrary $n \times n$ matrix fields over event space. The Lie algebra $\mathfrak{L}$ results roughly from closure of the matrices Γ_{κ} under commutation.

Next, we consider the relation between the internal holonomy group $\mathfrak{K}$ and the gauge group $\mathfrak{G}$. Local gauge theory has been developed in order to cope with a situation where, at the onset of the theoretical considerations, we do not know how to choose distinguished internal bases, *to the extent of* $\mathfrak{G}$. In order to have an invariant process of differentiation of internal vectors, a definition of equivalence of internal vectors at neighboring events has been introduced by means of the gauge potentials $\Gamma_{\kappa}(x^{\lambda})$. It would serve no purpose to let this equivalence violate $\mathfrak{G}$, i.e., to admit definitions of equivalence of covariant internal vectors, $v \to v'$, $w \to w'$, such that, when $w = gv$, $g \in \mathfrak{G}$, one would have $w' = g'v'$, and g' not belonging to $\mathfrak{G}$. Hence, we require that $\mathfrak{K}$ *is compatible with* $\mathfrak{G}$, i.e.,

$$g'(x^{\kappa} + dx^{\kappa}) = g(x^{\kappa}) + dx^{\mu}[\Gamma_{\mu}(x^{\lambda}), g(x^{\kappa})] \in \mathfrak{G}(x^{\kappa} + dx^{\mu}), \tag{35}$$

for every $g(x^{\kappa}) \in \mathfrak{G}(x^{\kappa})$. Executing the equivalence transport of the covariant internal vectors w and v, $w = gv$, $g \in \mathfrak{G}$, around a closed curve C, one finds

$$g'(x^{\kappa}) = H(C)g(x^{\kappa})H^{-1}(C), \quad H(C) \in \mathfrak{K}(x^{\kappa}). \tag{36}$$

Since for compatibility, $g'(x^{\kappa}) \in \mathfrak{G}(x^{\kappa})$ for every $g(x^{\kappa}) \in \mathfrak{G}(x^{\kappa})$, one has—

Theorem 5: Compatibility of $\mathfrak{K}$ and $\mathfrak{G}$ requires that either (1) $\mathfrak{K}$ is a subgroup of $\mathfrak{G}$, or (2) $\mathfrak{G}$ is an invariant subgroup of $\mathfrak{K}$.

The second case is unfamiliar. The matter may perhaps be clarified somewhat further by considering the Lie algebras. Without loss of generality we can choose an internal base such that the set $\mathfrak{G}(x^{\kappa})$ of matrices is independent of x^{κ}. Then (35) demands that the Γ_{μ} produce infinitesimal automorphism of $\mathfrak{G}$:

$$[\Gamma_{\mu}(x^{\lambda}), \mathfrak{G}] \subset \mathfrak{G}.$$

If $\mathfrak{G}$ is semisimple, all the infinitesimal automorphisms of $\mathfrak{G}$ are inner automorphisms, so that Γ_{μ} may be chosen in the Lie algebra of $\mathfrak{G}$; the results is Case 1 of Theorem 5. For a nonsemisimple gauge group, one may have Case 2 of Theorem 5.

III. ADJOINT REPRESENTATION

Let $L_i(x^{\kappa})$, $i = 1, \cdots, m$, be a complete set of linearly independent generators of $\mathfrak{K}(x^{\kappa})$. The L_i (L_{ia}^b in index notation) have components subject to base transformations in internal space as well as to base transformations in a real linear vector space of vectors η^i; the Lie algebra $\mathfrak{L}(x^{\kappa})$ of $\mathfrak{K}(x^{\kappa})$ consists of all internal operators $\eta^i L_i$. Hence, with the L_i fixed and known, the η^i specifies an element of the Lie algebra $\mathfrak{L}(x^{\kappa})$; the *real* linear vector space of vectors η^i may be called the (Lie) algebra space of $\mathfrak{K}$. In order to keep the account clear, we introduce a *separate* algebra space for each event. Every event now has a private internal space and a private algebra space. At every event the generators L_{ia}^b form a connecting quantity between these two spaces. In the sequel it will be convenient to have a covariant derivative defined for the algebra elements η^i,

$$\nabla_{\kappa}\eta^i = \partial_{\kappa}\eta^i + \Gamma_{\kappa j}^i\eta^j. \tag{37}$$

The algebra elements $\eta^i(x^{\kappa} + dx^{\kappa})$ and $\eta^i(x^{\kappa})$ are called equivalent if $dx^{\kappa}\nabla_{\kappa}\eta^i = 0$. The question is how

[16] The name is chosen in analogy with ordinary holonomy groups, see V. Hlavatý, J. Math. Mech. **8**, 597 (1959).

to choose the parameters[17] $\Gamma^i_{\kappa j}(x^\lambda)$. The simplest and most convenient choice is such that equivalent η^i's give equivalent internal operators $\eta^i L_i$. Writing from here on ∇_κ for the general covariant derivative, i.e., a derivative covariant under internal base transformations as well as under algebra base transformations, this requirement amounts to $dx^\kappa \nabla_\kappa(\eta^i L_i) = 0$, whenever $dx^\kappa \nabla_\kappa \eta^i = 0$. These equations imply

$$\nabla_\kappa L_i = 0, \tag{38}$$

which, written out with (37) and (9), reads

$$\partial_\kappa L_i - \Gamma^j_{\kappa i} L_j - [\Gamma_\kappa, L_i] = 0. \tag{39}$$

There is—

Lemma 5: There exist an internal base and an algebra base such that the $L_i(x^\kappa)$ are constant matrices.

Proof: By Lemma 1, there exists an internal base such that the matrix set $\mathfrak{L}(x^\kappa)$ is the same for all x^κ. Using this internal base, algebra bases over event space can be chosen such that $\partial_\kappa L_i = 0$. □

Theorem 6: For the base system of Lemma 5, the $\Gamma^j_{\kappa i}$ are uniquely determined by the gauge potentials $\Gamma_\kappa(x^\lambda)$ and the L_i.

Proof: For the base system of Lemma 5, (39) becomes

$$\Gamma^j_{\kappa i} L_j = -[\Gamma_\kappa, L_i]. \tag{40}$$

By Lemma 4, the commutator on the right-hand side lies in $\mathfrak{L}$, so that (40) may be written

$$\Gamma^j_{\kappa i} L_j = c^j_{\kappa i} L_j, \tag{41}$$

where the $c^j_{\kappa i}$ can be calculated from the Γ_κ and the L_i. Since the L_j are linearly independent, (41) amounts to $\Gamma^j_{\kappa i} = c^j_{\kappa i}$. □

The structure constants $c_{ij}{}^k$ of $\mathfrak{K}$ are defined by

$$[L_i, L_j] = c_{ij}{}^k L_k. \tag{42}$$

Covariant differentiation, use of (38), and the linear independence of L_k gives

$$\nabla_\kappa c_{ij}{}^k = 0. \tag{43}$$

With (43), all the algebraic concomitants of the structure constants are covariant constant as well:

$$\nabla_\kappa g_{ij} = 0, \quad \nabla_\kappa g_{ijk} = 0, \quad \cdots, \tag{44}$$

where[18] $g_{ij} = c_{ki}{}^l c_{lj}{}^k$ (group metric),

$$g_{ijk} = c_{mi}{}^l c_{lj}{}^n c_{nk}{}^m, \quad \cdots.$$

The covariant constancy of the L_i has another important consequence. Since equivalence transport is defined for internal vectors as well as for algebra elements η^i, we can take L_i around a loop C through x^κ by equivalence transport. Since $\nabla_\kappa L_i = 0$, the $L_i(x^\kappa)$ defined by equivalence transport always coincides with the original $L_i(x^\kappa)$. Now, also take a covariant internal vector v_a and an algebra element η^i around C by equivalence transport. Let the resulting transformations be

$$\begin{aligned} v'_a &= H^b_a v_b, \\ \eta'^i &= T^{(-1)i}_j \eta^j. \end{aligned} \tag{45}$$

Since $\eta^i v^a L^b_{ia} v_b$ is a scalar, it does not change under equivalence transport, so that

$$\eta'^i v'^a L'^b_{ia} v'_b = \eta^i v^a L^b_{ia} v_b. \tag{46}$$

Since $L'^b_{ia} = L^b_{ia}$, and (46) is true for arbitrary η^i and v_b, it follows with (45) that

$$T^{(-1)i}_j H^{-1a}_d L^b_{ia} H^c_b = L^c_{jd},$$

or, suppressing internal indices,

$$T^{(-1)i}_j H^{-1} L_i H = L_j. \tag{47}$$

For infinitesimal transformations

$$\begin{aligned} H &= I + t^k L_k, \\ T^i_j &= \delta^i_j + a^i_j, \end{aligned} \tag{48}$$

(47) gives

$$a^i_j L_i = -t^k c_{kj}{}^i L_i,$$

or, since the L_i are linearly independent,

$$a^i_j = -t^k c_{kj}{}^i. \tag{49}$$

Equations (49) and (48) show that the Lie algebra of the Lie group of matrices T^i_j is the adjoint representation of $\mathfrak{L}$. Hence, we have—

Theorem 7: The map which (38) defines between the internal holonomy group $\mathfrak{K}(x^\kappa)$ and the holonomy group acting on the algebra elements η^i is the adjoint representation.

The same result may be derived from the integrability conditions of (38), seen as partial differential equations for L_i. Covariant differentiation of (38) with respect to x^λ, and alternation over λ and κ gives

$$0 = \nabla_{[\lambda} \nabla_{\kappa]} L_i = -\tfrac{1}{2} \phi_{\lambda\kappa i}{}^j L_j - \tfrac{1}{2}[\phi_{\lambda\kappa}, L_i], \tag{50}$$

[17] η^i defines a Lie algebra element $\eta^i L_i$ as well as a Lie group element $\exp(\eta^i L_i)$. Hence, we could call the space of elements η_i the "group space of $\mathfrak{K}$," and the algebra space is then the tangent space to the group space at the identity. The linear connection Γ^j_{ki} of (37) should not be confused with the well-known linear connections $\Gamma^{(+)j}_{ki}$ and $\Gamma^{(-)j}_{ki}$ in group space, which are set by the first and second parameter groups, and which define equivalence of group vectors at neighboring points of group space. Our Γ^j_{ki} relate algebra spaces at neighboring events.

[18] See J. A. Schouten, Ref. 8, p. 220.

where

$$\phi_{\lambda\kappa i}{}^{j} = \partial_\lambda \Gamma^j_{\kappa i} - \partial_\kappa \Gamma^j_{\lambda i} - \Gamma^k_{\lambda i}\Gamma^j_{\kappa k} + \Gamma^k_{\kappa i}\Gamma^j_{\lambda k}.$$

On account of Theorem 4, $\phi_{\kappa\lambda}$ can be expanded in terms of the L_i:

$$\phi_{\kappa\lambda} = B^i_{\kappa\lambda} L_i. \tag{51}$$

Equation (49) then gives

$$\phi_{\kappa\lambda i}{}^{j} = -B^k_{\kappa\lambda} c_{ki}{}^{j}. \tag{52}$$

Differentiation of (51) and (52) gives

$$\nabla_\mu \phi_{\kappa\lambda} = (\nabla_\mu B^i_{\kappa\lambda}) L_i, \tag{53}$$

and

$$\nabla_\mu \phi_{\kappa\lambda i}{}^{j} = -(\nabla_\mu B^k_{\kappa\lambda}) c_{ki}{}^{j}, \tag{54}$$

on account of (38) and (43). A theorem can be proved for the Lie algebra of the holonomy group acting on algebra elements η^i, which reads just like Theorem 4, but with $\phi_{\kappa\lambda}$, $\nabla_\mu \phi_{\kappa\lambda}$, etc., replaced by $\phi_{\kappa\lambda i}{}^{j}$, $\nabla_\mu \phi_{\kappa\lambda i}{}^{j}$, etc. Equations (51), (52), (53), (54), and similar relations produced by further differentiation then results in (49).

Similar considerations apply to the mapping between holonomy groups for different representations of $\mathfrak{G}$; the role of the L_i is then played by another connecting quantity. For example, the map between the holonomy groups for the spinor and vector representations of $O(3)$ is set by the integrability conditions for the partial differential equations expressing the covariant constancy of the Pauli spin matrices. Another example, somewhat outside the context of the present paper, is the case of the Riemann space of general relativity where one may consider the map between the ordinary holonomy group and the holonomy group operating on Dirac spinors. Then the covariant constancy of the Dirac matrices sets the map, and Schrödinger's relation[19] between the Riemann curvature and the spin curvature follows from the integrability conditions for these equations.

We turn to the question whether the Γ_κ can be expanded in terms of the L_i.

Theorem 8 (R. P. Treat): For $\mathfrak{K}$ semisimple, there exists an internal base and an algebra base such that $\partial_\kappa c_{ki}{}^{j} = 0$ and $\Gamma^j_{\kappa i} = -b^k_\kappa c_{ki}{}^{j}$ for some b^k_κ.

Proof[20]*:* Choose the base system of Lemma 5. Then (42) shows that $\partial_\kappa c_{ij}{}^{k} = 0$. Hence, (43) gives

$$-\Gamma^l_{\kappa i} c_{lj}{}^{k} - \Gamma^l_{\kappa j} c_{il}{}^{k} + \Gamma^k_{\kappa l} c_{ij}{}^{l} = 0.$$

Transvecting with $c_{mk}{}^{j}$, and using (44) and the Jacobi identity

$$c_{[il}{}^{k} c_{m]k}{}^{j} = 0 \tag{55}$$

gives

$$\Gamma^l_{\kappa i} g_{lm} = -\Gamma_{\kappa j}{}^{l} c_{im}{}^{k} c_{lk}{}^{j}. \tag{56}$$

For a semisimple group, the group metric g_{lm} is nonsingular. Then, (56) implies

$$\Gamma^m_{\kappa i} = -\Gamma^l_{\kappa j} c_{lk}{}^{j}\, c_i{}^{mk}. \tag{57}$$

The total antisymmetry of the c_{ijk} can be used to write $c_i{}^{mk} = c^k{}_i{}^{m}$, and (57) becomes

$$\Gamma^m_{\kappa i} = -(\Gamma_{\kappa j}{}^{l}\, c_l{}^{kj}) c_{ki}{}^{m},$$

which shows that

$$\Gamma^m_{\kappa i} = -b^k_\kappa c_{ki}{}^{m}, \tag{58}$$

with

$$b^k_\kappa = \Gamma^l_{\kappa j} c_l{}^{kj}. \qquad \square \tag{59}$$

Using Treat's theorem, we can show—

Theorem 9: If $\mathfrak{K}$ is semisimple, there exists a base system such that $\Gamma_\kappa = b^i_\kappa L_i$, $\partial_\kappa L_i = 0$.

Proof: For the base system of Lemma 5, (38) becomes (40). By Theorem 8, (58) holds so that (40) becomes

$$[\Gamma_\kappa, L_i] = b^k_\kappa c_{ki}{}^{j} L_j. \tag{60}$$

Equation (60) is solved in general by

$$\Gamma_\kappa = b^k_\kappa L_k + M_\kappa = L_\kappa + M_\kappa, \tag{61}$$

where M_κ does not belong to $\mathfrak{L}$ and

$$[M_\kappa, L_i] = 0. \tag{62}$$

We now calculate the gauge field belonging to the gauge potentials (61) and find, using (62), that

$$\phi_{\kappa\lambda} = L_{\kappa\lambda} + M_{\kappa\lambda}, \tag{63}$$

where $L_{\kappa\lambda}$ and $M_{\kappa\lambda}$ are, respectively, the gauge fields belonging to L_κ alone and to M_κ alone. Since $L_{\kappa\lambda} \in \mathfrak{L}$ and Theorem 4, $\phi_{\kappa\lambda} \in \mathfrak{L}$, (62) implies

$$M_{\kappa\lambda} \in \mathfrak{L}. \tag{64}$$

Denote by $\mathcal{M}(x^\lambda)$ the set of all $n \times n$ matrices which commute with all elements of $\mathfrak{L}(x^\lambda)$. Clearly, $M_\kappa \in \mathcal{M}$. Since $\mathfrak{L}(x^\kappa)$ does not depend on x^κ, $\mathcal{M}(x^\kappa)$ does not depend on x^κ either. Then by a lemma which can be proved in much the same way as Lemma 2, one finds $\partial_\kappa M_\lambda \in \mathcal{M}$. Also, with the Jacobi identity and (62), one can show

$$[M_\kappa, M_\lambda] \in \mathcal{M}.$$

Hence,

$$M_{\kappa\lambda} = \partial_\kappa M_\lambda - \partial_\lambda M_\kappa - [M_\kappa, M_\lambda] \in \mathcal{M}. \tag{65}$$

[19] See E. Schrödinger, Sitzber. Preuss. Akad. Wiss. Physik-Math. Kl. XI, 105 (1932); H. G. Loos, Ann. Phys. (N.Y.) 25, 91 (1963).

[20] We show the proof given by R. P. Treat (private communication).

But since $\mathfrak{L}$ is semisimple, $\mathfrak{L}$ contains no elements commuting with all elements of $\mathfrak{L}$; i.e., $\mathfrak{M}$ and $\mathfrak{L}$ have no elements in common. Hence, (64) and (65) can only be true if

$$M_{\kappa\lambda} = 0. \tag{66}$$

Now, subject the gauge potentials (61) to an internal base transformation S; in the new base, one has

$$\Gamma'_\kappa = S^{-1}(L_\kappa S + M_\kappa S - \partial_\kappa S).$$

We try to find $S(x^\kappa)$ such that

$$M_\kappa S - \partial_\kappa S = 0; \tag{67}$$

if this could be done we would have

$$\Gamma'_\kappa = S^{-1}L_\kappa S \in S^{-1}\mathfrak{L}S = \mathfrak{L}',$$

so that then

$$\Gamma'_\kappa \in \mathfrak{L}'. \tag{68}$$

The integrability conditions for (67) are obtained by partial differentiation with respect to x^λ, alternation over κ and λ and elimination of first derivatives of S by means of (67); the result found is (66). Hence, the integrability conditions for Eqs. (67) are satisfied, and we can find solutions of (67) subject to initial conditions, such as $S(x_0^\kappa) = I$. At an event $x_0^\kappa + dx^\kappa$, one of them has from (67)

$$S(x_0^\kappa + dx^\kappa) = I + dx^\kappa M_\kappa(x_0^\lambda).$$

Calculation of $L' = S^{-1}LS$ for any $L \in \mathfrak{L}$ at $x_0^\kappa + dx^\kappa$ gives $L' = L - dx^\kappa[M_\kappa(x_0^\lambda), L] = L$, because of (62). Hence, in the new base we have the same $L_i(x^\kappa)$ as in the old base. □

Note that Theorem 9 holds for compact as well as noncompact $\mathfrak{K}$ and regardless of whether or not $\mathfrak{K}$ acts irreducibly in internal space.

IV. YANG–MILLS EQUATIONS

The Yang–Mills equations for the gauge potentials are

$$\tilde{J}^\kappa = \tilde{g}\nabla_\lambda \phi^{\kappa\lambda}; \tag{69}$$

$\tilde{J}_\kappa$ is the generalized current density operator (isospin current density in the original Yang–Mills paper[1]), $\phi_{\kappa\lambda}$ are the gauge fields (10) constructed from the gauge potentials $\Gamma_\kappa(x^\lambda)$, and $\tilde{g} = |\mathrm{Det}\, g_{\kappa\lambda}|^{\frac{1}{2}}$, where $g_{\kappa\lambda}$ is the metric tensor in event space. Several exact solutions of (69) for a point charge (i.e., $\tilde{J}^\kappa = 0$, except on the world line $x^1 = x^2 = x^3 = 0$) are known.[21] Among these, only Treat's solution has short range components; it has a non-Abelian internal holonomy group. All known solutions with Abelian internal holonomy groups are Coulomb-like[22] (i.e., long range). In fact, we have—

Theorem 10 (C. A. Uzes): For gauge fields with Abelian internal holonomy group, the Yang–Mills equations and the Bianchi identities reduce to Maxwell's equations.

Proof[23]*:* For $\mathfrak{K}$ Abelian, the structure constants vanish. Then, by (52)

$$\phi_{\kappa\lambda i}{}^j = 0. \tag{70}$$

Now consider the transformation of the $\Gamma^j_{\kappa i}$ under a change $A^{i'}_i$ of algebra base[24]:

$$\Gamma^{j'}_{\kappa i'} = A^i_{i'}(\Gamma^j_{\kappa i}A^{j'}_j - \partial_\kappa A^{j'}_i).$$

We try to find an algebra base transformation $A^{i'}_i$ such that $\Gamma^{j'}_{\kappa i'} = 0$. This requires

$$\Gamma^j_{\kappa i}A^{j'}_j - \partial_\kappa A^{j'}_i = 0. \tag{71}$$

The integrability conditions for these partial differential equations are just (70). Hence, (71) can be integrated so that there exists an algebra base with

$$\Gamma^j_{\kappa i} = 0 \tag{72}$$

everywhere in event space. Taking such a base and using the expansion (51) in (69) and (13) gives

$$\begin{aligned} \tilde{J}^\kappa &= \tilde{g}\nabla_\lambda B^{\kappa\lambda i}L_i, \\ 0 &= \nabla_{[\kappa}B^i_{\lambda\mu]}L_i. \end{aligned} \tag{73}$$

By Leibnitz' rule, (38) and the linear independence of the L_i, (73) implies

$$\begin{aligned} \tilde{j}^{\kappa i} &= \tilde{g}\nabla_\lambda B^{\kappa\lambda i}, \\ 0 &= \nabla_{[\kappa}B^i_{\lambda\mu]}, \end{aligned} \tag{74}$$

where

$$\tilde{J}^\kappa = \tilde{j}^{\kappa i}L_i,$$

the latter expansion is possible by Theorem 4. Using (72) and remembering that all along we have been using Cartesian inertial event coordinates and a Minkowskian event space, Eq. (74) may be written

$$\begin{aligned} \tilde{j}^{\kappa i} &= \tilde{g}\partial_\kappa B^{\kappa\lambda i}, \\ 0 &= \partial_{[\kappa}B^i_{\lambda\mu]}, \end{aligned}$$

which are Maxwell's equations. □

[21] (a) M. Ikeda and Y. Miyachi, Progr. Theoret. Phys. (Kyoto) **27**, 474 (1962); (b) H. G. Loos, Nucl. Phys. **72**, 677 (1965); (c) Hendricus G. Loos, J. Math. Phys. **8**, 1870 (1967); (d) R. P. Treat, "A Short Range Gauge Field," Nuovo Cimento (to be published).

[22] For the monopole field, of course.

[23] We follow, with some modifications, a proof given by C. A. Uzes (private communication).

[24] This transformation is determined by the requirement that equivalence of algebra elements η^i at neighboring events is invariant under algebra base transformations.

Corollary 1: In order that gauge fields can have nonelectrodynamic features, their internal holonomy group must be non-Abelian.

Note that $\mathfrak{K}$ may be Abelian for a non-Abelian gauge group $\mathfrak{G}$. An example is provided by Treat's solution[21] for the case of vanishing f; this makes $\mathfrak{K}$ Abelian keeping the non-Abelian gauge group with generators A, B, C, and commutators $[A, B] = C$, $[A, C] = B$, $[B, C] = 0$.

Theorem 11: When a free gauge field (i.e., a gauge field satisfying the Yang–Mills equations for zero current density) has an internal holonomy group which is the direct product of two semisimple groups $\mathfrak{K}_1$ and $\mathfrak{K}_2$, then it can be decomposed into two noninteracting fields.

Proof: Since a direct product of semisimple groups is semisimple, Theorem 9 applies. Hence, there is a base system such that

$$\Gamma_\kappa = \Gamma_{1\kappa} + \Gamma_{2\kappa}, \tag{75}$$

where $\Gamma_{1\kappa} \in \mathfrak{L}_1$, $\Gamma_{2\kappa} \in \mathfrak{L}_2$, and $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are the Lie algebras of $\mathfrak{K}_1$ and $\mathfrak{K}_2$, respectively. From (10) and $[\Gamma_{1\kappa}, \Gamma_{2\lambda}] = 0$, one has

$$\phi_{\kappa\lambda} = \phi_{1\kappa\lambda} + \phi_{2\kappa\lambda}, \tag{76}$$

where $\phi_{1\kappa\lambda}$ and $\phi_{2\kappa\lambda}$ are, respectively, the gauge fields constructed from $\Gamma_{1\kappa}$ alone, and from $\Gamma_{2\kappa}$ alone. Putting (75) and (76) in the free Yang–Mills equations [(69) with $\mathfrak{J}^\kappa = 0$] and in the Bianchi identities (13) gives two uncoupled sets of equations:

$$\nabla_\lambda^{(1)}\phi_1^{\kappa\lambda} = 0, \quad \nabla_{[\mu}^{(1)}\phi_{1\kappa\lambda]} = 0, \tag{77}$$

$$\nabla_\lambda^{(2)}\phi_2^{\kappa\lambda} = 0, \quad \nabla_{[\mu}^{(2)}\phi_{2\kappa\lambda]} = 0, \tag{78}$$

where $\nabla_\lambda^{(1)}$ means that $\Gamma_{1\lambda}$ only is used in the calculations of the covariant derivative. Equations (77) and (78) show that the fields $\phi_{1\kappa\lambda}$ and $\phi_{2\kappa\lambda}$ propagate independently. □

The gauge potentials $\Gamma_\kappa(x^\lambda)$ define a large number of covariant divergence-free current densities.[25] The simplest string of such current densities is

$$\begin{aligned}
&\mathfrak{J}_1^\kappa = \tilde{g}\nabla_\lambda\phi^{\kappa\lambda} \quad \text{(Yang–Mills current density)},\\
&\mathfrak{J}_2^\kappa = \tilde{g}\nabla_\lambda\phi^{[\kappa|\mu|}\phi_\mu^{\lambda]},\\
&\mathfrak{J}_3^\kappa = \tilde{g}\nabla_\lambda\phi^{[\kappa|\mu}\phi_\mu^{\nu|}\phi_\nu^{\lambda]},\\
&\quad\cdot \qquad\cdot\\
&\quad\cdot \qquad\cdot\\
&\quad\cdot \qquad\cdot
\end{aligned} \tag{79}$$

For all these current densities one has $\nabla_\kappa\mathfrak{J}^\kappa = 0$; this can be verified by using (12). One has—

Theorem 12: If equivalence displacement preserves the norm of internal vectors, the norm being defined by means of an internal metric tensor, then the current densities (79) lie in the Lie algebra of the unitary group belonging to the internal metric.

Proof: If the internal metric tensor g has a signature consisting of $m - p$ plus signs and p minus signs, then $\mathfrak{K} \subset U(m - p, p)$. Hence, all elements of $\mathfrak{L}$ are antiself-adjoint with respect to g. Taking the adjoint of each of the internal operators (79) and using the fact that in this case covariant differentiation commutes with the process of taking the adjoint (since the internal metric is covariant constant) shows that all the currents in (79) are antiself-adjoint with respect to g. □

A similar theorem holds for the derivative currents discussed in Ref. 25. We briefly mention some further results involving internal holonomy groups, which have been proved elsewhere. The proofs will not be repeated here.

Theorem 13[26]*:* Spherically symmetric analytic[27] solutions of the point-charge Yang–Mills equations have Abelian internal holonomy groups.

This theorem together with Corollary 1 clarifies why Ikeda and Miyachi's spherically symmetric solutions[21] are electromagnetic.

Theorem 14[28]*:* Free gauge fields (i.e., gauge fields satisfying the Yang–Mills equations for zero current density) have, in the classical theory, a positive-definite energy density, if their internal holonomy group is semisimple and compact.

A covariant divergence-free current density $\mathfrak{J}^\kappa$, for instance one of the string (79), does not always give rise to as many conserved charges as there are linearly independent algebraic components of $\mathfrak{J}^\kappa$. This situation is entirely governed by the internal holonomy group, and is discussed in Ref. 25. The principal argument is that the calculation of the total charge carried by $\mathfrak{J}^\kappa$,

$$Q = \int_{t_1} \mathfrak{J}^\kappa \, df_\kappa, \tag{80}$$

[25] H. G. Loos, Ann. Phys. (N.Y.) **36**, 486 (1966).

[26] H. G. Loos, Ref. 21(b).

[27] Except, of course, on the world line of the point charge.

[28] R. P. Treat, Ref. 21(d).

where $d\tilde{f}_\kappa$ is an element of the hypersurface $x^0 = t$, over which the integration (80) is carried out, involves summation of internal operators $\tilde{J}^\kappa d\tilde{f}_\kappa$ at different events, a process which generally is not gauge invariant.

V. DISCUSSION

The properties of internal holonomy groups discussed here show how rich a geometric structure is defined by the Yang–Mills potentials, and how much bearing this structure has on local gauge theory. Of particular importance is the relation between the range of a solution of the Yang–Mills equations and its internal holonomy group.

The demonstration of existence of a classical short-range Yang–Mills field by Treat[21] should help to rejuvenate the theory of Yang–Mills fields [not restricted to $O(3)$, of course] as a possible theory of strong interactions. The internal holonomy group and associated geometric concepts form an indispensable tool for further investigation of Yang–Mills fields.

Quantization has not been considered in the present paper, because so much clarification is already needed on the classical level.

We started out with the gauge potentials and their transformation properties under gauge transformations, and showed that the gauge potentials define gauge-invariantly an equivalence of multiplets at neighboring events. The gauge potentials usually are introduced as compensating fields, starting with gauge-invariance requirements on the Lagrangian. But another point of view is possible which, although presently not fashionable, is really simpler. Regardless of the Lagrangian, the mere fact that field equations for a multiplet field occur in the theory demands definition of a gauge-invariant concept of differentiation, if we want the field equations to have an invariant meaning under event-dependent gauge transformations. In other words, a gauge-covariant derivative must be defined for the same reason that a covariant derivative must be defined in general relativity. It would then be most appropriate to call the gauge potentials Γ_κ parameters of the internal linear connection, and to call the gauge fields $\phi_{\kappa\lambda}$ the internal-curvature tensor operator; this has been done in previous papers.[29] In the present paper a compromise has been made between simplicity and readability; we call Γ_κ and $\phi_{\kappa\lambda}$ by their "physical" names: gauge potentials and gauge fields, and we do not start with the internal connection Γ_κ introduced as a geometric necessity.

Theorem 9 is of great importance for practical work with the Yang–Mills equations. For instance, if one wants to prove nonexistence of localized nonsingular solutions of the free Yang–Mills equations with a certain internal holonomy group $\mathcal{H}$, one had better know what restriction to place on the gauge potentials Γ_κ, such that the internal holonomy group is indeed $\mathcal{H}$, without excluding any such solutions. Theorem 9 is a step in this direction, but a stronger theorem valid for any $\mathcal{H}$ is sorely needed.

ACKNOWLEDGMENT

The author wishes to thank Richard P. Treat for several helpful and stimulating discussions.

[29] See papers by H. G. Loos, Ref. 21(b), (c) and Ref. 25, and R. P. Treat, Ref. 21(d).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

A Class of Null Spectroscopic Coefficients*

B. R. JUDD AND H. T. WADZINSKI
Department of Physics, Johns Hopkins University, Baltimore, Maryland

(Received 16 January 1967)

The irreducible representations occurring in the decomposition of the Kronecker products $U \times U$ and $W \times W$ for irreducible representations U and W of the groups G_2 and R_7 are separated into symmetrical and antisymmetrical parts. Application of the Wigner–Eckart theorem yields new selection rules and explains many of the apparently accidental zeros in the tabulation by Nielson and Koster of the reduced matrix elements of various single-particle tensor operators for the electronic configurations f^n. The method is applied in detail to the magnetic spin–spin interaction between electrons in the f shell, and the new selection rules are presented in tables.

1. INTRODUCTION

A FEW years ago, Nielson and Koster[1] tabulated a large number of matrix elements for sums of single-particle tensor operators taken between states belonging to configurations of equivalent f electrons. In spite of the fact that they excluded all matrix elements for which the familiar triangular conditions for angular momentum quantum numbers were not satisfied, one is immediately struck by the large number of matrix elements whose values are zero. Many can be accounted for by generalizing the concept of triangular conditions. The states that enter the matrix elements are defined by the representations W and U of the respective groups R_7 and G_2;[2] and in some cases the identity representation is not contained in the triple Kronecker products $W \times W' \times W''$ or $U \times U' \times U''$ that describe the transformation properties of the bra, the operator, and the ket of a given matrix element. For the group R_3, the analogous statement is identical to the triangular condition. In a few other cases, the tensor operators play the role of the infinitesimal operators of a group, and thus cannot connect states belonging to different irreducible representations. Considerations of these kinds explain many of the zero matrix elements, but a surprisingly large residue remains. For example, of the 368 entries for the reduced matrix elements of the sixth-rank tensor $\mathbf{U}^{(6)}$ for the configuration f^4, no fewer than 21 vanish; and of these 21, only two can be readily understood. We are thus left with 19 apparently accidental zeros.

It is the purpose of this article to point out the existence of a general class of null matrix elements, for which many of the apparently accidental zeros of Nielson and Koster are special cases. As a result of the analysis, new selection rules are obtained, not only for the operators studied by Nielson and Koster, but also for operators of physical interest (such as the spin–spin interaction) whose matrix elements are not included in these authors' tables. The method makes use of the symmetry properties possessed by various product functions, and it is to these that we first turn.

2. SYMMETRY

It is well known that the states of the configuration j^2, comprising two equivalent particles each of angular momentum j, are symmetric or antisymmetric with respect to the interchange of the particles according as J, the quantum number of the total angular momentum, is such that $2j - J$ is even or odd, respectively. This can be expressed in the language of group theory. The $2j + 1$ states for a single particle can be regarded as forming bases for either the representation [1] of the unitary group U_{2j+1}, or the representation $\mathfrak{D}_j$ of the subgroup R_3 of U_{2j+1}. The decomposition

$$[1] \times [1] = [2] + [11] \tag{1}$$

breaks up the product functions into their symmetrical and antisymmetrical parts; the representations $\mathfrak{D}_j$ of R_3 that occur in the associated decomposition

$$\mathfrak{D}_j \times \mathfrak{D}_j = \sum_{J=0}^{2j} \mathfrak{D}_J \tag{2}$$

belong to [2] if $2j - J$ is even and to [11] if $2j - J$ is odd.

It is clear from this form of the statement that the separation into symmetrical and antisymmetrical parts is not necessarily confined to product functions that form bases for R_3. In fact, the Kronecker product $\Gamma \times \Gamma$ for any N-dimensional representation Γ of a group G may be so decomposed. Since we know that the dimensions of [2] and [11] are $\frac{1}{2}N(N+1)$ and $\frac{1}{2}N(N-1)$, respectively, it is often possible to carry out the decomposition in simple cases purely from dimensional considerations. This procedure

* This work was partially supported by the U.S. Atomic Energy Commission.

[1] C. W. Nielson and G. F. Koster, *Spectroscopic Coefficients for the p^n, d^n, and f^n Configurations* (The MIT Press, Cambridge, Massachusetts, 1963).

[2] G. Racah, Phys. Rev. 76, 1352 (1949).

TABLE I. Separation of $U \times U$ into symmetrical and antisymmetrical parts.

$U \times U$	Symmetric	Antisymmetric
(00) × (00)	(00)	
(10) × (10)	(00) (20)	(10) (11)
(11) × (11)	(00) (20) (22)	(11) (30)
(20) × (20)	(00) $(20)^2$ (21) (22) (40)	(10) (11) (21) (30) (31)
(21) × (21)	(00) $(20)^2$ (21) $(22)^2$ (30) (31) $(40)^2$ (41) (42)	(10) $(11)^2$ (21) $(30)^2$ $(31)^2$ (32) (33) (41) (50)
(22) × (22)	(00) (20) $(22)^2$ (40) (41) (42) (44) (60)	(11) (30) (31) (33) (41) (50) (52)
(30) × (30)	(00) $(20)^2$ (21) $(22)^2$ (31) $(40)^2$ (32) (41) (42) (60)	(10) (11) (21) $(30)^2$ $(31)^2$ (32) (33) (41) (50) (51)
(31) × (31)	(00) $(20)^3$ $(21)^2$ $(22)^2$ (30) $(31)^2$ $(40)^4$ $(32)^2$ $(41)^3$ $(42)^4$ (43) (44) (50) $(51)^2$ (52) $(60)^2$ (61) (62)	(10) $(11)^2$ $(21)^2$ $(30)^3$ $(31)^4$ (40) $(32)^3$ $(33)^2$ $(41)^3$ (42) (43) $(50)^2$ $(51)^3$ $(52)^2$ (53) (61) (70)
(40) × (40)	(00) $(20)^2$ (21) $(22)^2$ (31) $(40)^3$ (32) $(41)^2$ $(42)^3$ (43) (44) (51) (52) $(60)^2$ (61) (62) (80)	(10) (11) (21) $(30)^2$ $(31)^2$ $(32)^2$ $(33)^2$ $(41)^2$ (42) (43) $(50)^2$ $(51)^2$ (52) (53) (61) (70) (71)

TABLE II. Separation of $W \times W$ into symmetrical and antisymmetrical parts.

$W \times W$	Symmetric	Antisymmetric
(000) × (000)	(000)	
(100) × (100)	(000) (200)	(110)
(110) × (110)	(000) (111) (200) (220)	(110) (211)
(111) × (111)	(000) (111) (200) (210) (220) (222)	(100) (110) (211) (221)
(200) × (200)	(000) (200) (220) (400)	(110) (310)
(210) × (210)	(000) (111) $(200)^2$ (211) $(220)^2$ (222) (310) (311) (321) (400) (420)	$(110)^2$ $(211)^2$ (221) $(310)^2$ (321) (330) (411)
(211) × (211)	(000) $(111)^2$ $(200)^2$ $(210)^2$ (211) $(220)^2$ (221) $(222)^2$ (310) $(311)^2$ (320) $(321)^2$ (322) (331) (400) (410) (420) (422)	(100) $(110)^2$ (210) $(211)^3$ $(221)^3$ (300) $(310)^2$ (311) (320) $(321)^2$ (322) (330) (332) (411) (421)
(220) × (220)	(000) (111) (200) $(220)^2$ (222) (311) (321) (331) (400) (420) (422) (440)	(110) (211) (221) (310) (321) (322) (330) (411) (431)
(221) × (221)	(000) $(111)^2$ $(200)^2$ $(210)^2$ (211) $(220)^2$ (221) $(222)^2$ (310) $(311)^3$ (320) $(321)^3$ (322) $(331)^3$ (332) (333) (400) (410) $(420)^2$ (421) $(422)^2$ (430) (431) (432) (440) (442)	(100) $(110)^2$ (210) $(211)^3$ $(221)^3$ (300) $(310)^2$ (311) $(320)^2$ $(321)^3$ $(322)^2$ $(330)^2$ (331) $(332)^2$ $(411)^2$ $(421)^2$ $(431)^2$ (432) (433) (441)
(222) × (222)	(000) (111) (200) (210) (220) (222) (311) (321) (331) (333) (400) (410) (420) (422) (430) (432) (440) (442) (444)	(100) (110) (211) (221) (300) (310) (320) (322) (330) (332) (411) (421) (431) (433) (441) (443)

becomes unsatisfactory for N larger than 10 or so, since the solutions to the dimensional equations are not always unambiguous. It is then more convenient to combine our knowledge of the procedure for R_3 with the branching rules for the reductions $G_2 \rightarrow R_3$ and $R_7 \rightarrow G_2$, extending the available tables of branching rules[2] and Kronecker products[3] where necessary.

[3] P. B. Nutter, Raytheon Tech. Memor. T-544 (1964).

For example, the decomposition of the Kronecker product (11) × (11), where (11) is a representation of G_2, is made by first noting that for the reduction $G_2 \rightarrow R_3$, the branching rule

$$(11) \rightarrow \mathfrak{D}_1 + \mathfrak{D}_5$$

is valid. Now $\mathfrak{D}_J \times \mathfrak{D}_J$ is readily separable into symmetrical and antisymmetrical parts by Eq. (2)

above; and the cross term

$$\mathfrak{D}_1 \times \mathfrak{D}_5 + \mathfrak{D}_5 \times \mathfrak{D}_1$$

in the product gives rise to pairs of representations $\mathfrak{D}_{J'}$, one of which is symmetrical, the other antisymmetrical. We thus find that

$$(\mathfrak{D}_1 + \mathfrak{D}_5) \times (\mathfrak{D}_1 + \mathfrak{D}_5)$$

decomposes into the symmetrical part

$$S^2D^2G^2HI^2LN$$

and the antisymmetrical part

$$P^2FGH^2IKM,$$

where, for brevity, the spectroscopic symbol L is used for the representation $\mathfrak{D}_L$. From the tables of Nutter,[3]

$$(11) \times (11) = (00) + (11) + (20) + (22) + (30).$$

Racah's branching rules[2] give

$$(00) \rightarrow S,$$
$$(11) \rightarrow PH,$$
$$(20) \rightarrow DGI,$$
$$(22) \rightarrow SDGHILN,$$
$$(30) \rightarrow PFGHIKM.$$

We see at once from the presence of the M and N terms that (22) and (30) must correspond to the symmetrical and antisymmetrical parts, respectively. Only one I term, a symmetrical one, is unaccounted for; it follows that (20), to which this term belongs, must also be symmetrical. Proceeding in this way, we find that the structure

$$(00) + (20) + (22)$$

corresponds to the symmetrical representation [2], while

$$(11) + (30)$$

corresponds to the antisymmetrical representation [11].

Once this kind of analysis has been performed for a few of the simpler representations, it is possible to set up a chain calculation, working solely within G_2. Suppose, for example, the separation of $U_1 \times U_1$ is required, where U_1 is a representation of G_2. A product $U' \times U''$ is sought that possesses the following properties: (1) The irreducible representations U_i into which $U' \times U''$ decomposes contain U_1, preferably once; (2) the separation of all products $U_i \times U_i$ into symmetrical and antisymmetrical parts is known for all i except $i = 1$; (3) the separations of $U' \times U'$ and $U'' \times U''$ are also known. We now write

$$(U' \times U'')^2 = \left(\sum_i U_i\right)^2.$$

The separation of $(U' \times U'')^2$ into symmetrical and antisymmetrical parts can be carried out by expanding it as $(U' \times U') \times (U'' \times U'')$; the cross terms between U_1 and U_i $(i \neq 1)$ give rise to equal symmetrical and antisymmetrical parts. It follows that $U_1 \times U_1$ can be found by a process of subtraction.

The separation of $U \times U$ into symmetrical and antisymmetrical parts has been carried out for every representation U of G_2 that occurs in the classification of the states of f^n. The results are assembled in Table I. Analogous methods can be used for the products $W \times W$ of irreducible representation W of R_7; the results of Table I can be used with the branching rules for $R_7 \rightarrow G_2$ to facilitate the analysis. The results are assembled in Table II for all representations W used in classifying the states of f^n.

3. WIGNER–ECKART THEOREM

The usefulness of Tables I and II becomes apparent when a few special cases are examined. Consider, for example, the matrix element

$$\langle\!\langle (30)LM_L + q | \, U_q^{(6)} \, | (20)IM_L \rangle\!\rangle,$$

where the representations (30) and (20) belong to the group G_2. The tensors $(2k + 1)^{\frac{1}{2}}\mathbf{U}^{(k)}$ for $k = 2, 4$, and 6 transform according to the representations (20) of G_2 and (200) of R_7.[2] This means that the above matrix element is proportional to the coupling coefficient

$$\langle\!\langle (30)LM_L + q \mid (20)Iq, (20)IM_L \rangle\!\rangle. \qquad (3)$$

According to the Wigner–Eckart theorem, this coefficient can be broken up into an isoscalar factor and a vector-coupling (VC) coefficient:

$$\langle\!\langle (30)LM_L + q \mid (20)Iq, (20)IM_L \rangle\!\rangle$$
$$= \langle\!\langle (30)L \mid (20)I + (20)I \rangle\!\rangle (LM_L + q \mid Iq, IM_L). \qquad (4)$$

Now we see from Table I that (30) occurs in the antisymmetrical part of (20) × (20), and hence the coefficient (3) must reverse its sign if the two parts $(20)Iq$ and $(20)IM_L$ are interchanged. Equation (4) shows that this is guaranteed if L is odd, since the VC coefficient itself changes sign. But if L is even, the VC coefficient is invariant, as is also the isoscalar factor. Hence, the coefficient (3) is zero for even L. Not all the VC coefficients for a given L can be zero, and hence the only remaining possibility is that the isoscalar factor vanishes. In other words,

$$\langle\!\langle (30)L \mid (20)I + (20)I \rangle\!\rangle = 0 \quad (L \text{ even}).$$

This implies

$$\langle\!\langle (30)L \| \, U^{(6)} \, \| (20)I \rangle\!\rangle = 0 \quad (L \text{ even}), \qquad (5)$$

since reduced matrix elements are equivalent (apart

from a nonvanishing factor) to the corresponding isoscalar factors. Equation (5) accounts for two of the 19 unexplained zeros mentioned in Sec. 1.

If (30) had occurred in the symmetrical part of the reduction of (20) × (20)—as (40) does, for example—then Eq. (4) would imply the vanishing of the isoscalar factor in all those cases where the VC coefficient changes sign; that is, when L is odd. As examples of this we have

$$(\!(40)L\| U^{(k)} \|(20)k\!) = 0 \quad (L \text{ odd}),$$
$$(\!(22)H\| U^{(k)} \|(20)k\!) = 0.$$

On the other hand,

$$(\!(31)L\| U^{(k)} \|(20)k\!) = 0 \quad (L \text{ even}),$$

since (31), like (30), occurs in the antisymmetrical part of the decomposition of (20) × (20).

These methods can be equally well applied to the group R_7. From Table II, it can be seen that (220) occurs in the symmetrical part of the decomposition of (200) × (200), from which it follows that

$$(\!(220)L\| U^{(k)} \|(200)k\!) = 0 \quad (L \text{ odd}).$$

Five of the remaining 17 apparently accidental zeros are examples of this equation. It is clear that the arguments that have been used to obtain these results are in no way peculiar to G_2 and R_7. In fact, the conditions that have been given by de Swart[4] for the vanishing of certain isoscalar factors for SU_3 are the analogs of our results.

It should be added here that Tables I and II can be used to gain additional information about certain nonvanishing matrix elements. For example, the representation (21) occurs twice in the reduction of (20) × (20)—once as a symmetrical representation, and once as an antisymmetrical one. From this we may deduce, for example, that the parameter A occurring in the equation

$$(\!(f^3(20)^2k\| U^{(k)} \| f^3(21)^2L\!) = A(\!(f^4(20)^3k\| U^{(k)} \| f^4(21)^3L\!) \quad (6)$$

assumes only one value when L is even (namely, $-1/\sqrt{2}$), and another one (namely, $3/\sqrt{2}$) when L is odd. This goes beyond a naive application of the Wigner–Eckart theorem, which would not predict any proportionality.

4. INTRA-ATOMIC INTERACTIONS

The discussion in the preceding section is limited to the tensors $\mathbf{U}^{(k)}$. These enter naturally in crystal-field calculations of the type where the electric potential of the crystal lattice is expanded in spherical harmonics, taking the nucleus of the ion under study as the origin. When we consider other types of operators, such as those corresponding to the spin–orbit, spin–spin, or spin–other-orbit interactions in an atom, a number of new features arise.

In the first place, we can sometimes use the factorization of the isoscalar factors themselves to derive more general results. For example, the spin–orbit interaction with a configuration of the type f^n can be represented as the scalar part of a double tensor such as $\mathbf{V}^{(11)}$, and this belongs to the representations (110) of R_7 and (11) of G_2.[5] The matrix element

$$(\!(f^n(211)(20)SLM_S + \pi M_L + q| V^{(11)}_{\pi q} | f^n(110)(11)S'PM_SM_L\!)$$

involves the isoscalar factor

$$(\!(211)(20)L \,|\, (110)(11)P + (110)(11)P\!),$$

which factorizes as follows:

$$(\!(211)(20)L \,|\, (110)(11)P + (110)(11)P\!) = (\!(211)(20) \,|\, (110)(11) + (110)(11)\!) \times (\!(20)L \,|\, (11)P + (11)P\!). \quad (7)$$

Now (20) occurs in the symmetrical part of the decomposition of (11) × (11), whereas (211) occurs in the antisymmetrical part of the decomposition of (110) × (110). Since there exist for the configuration f^2 nonvanishing matrix elements of $\mathbf{V}^{(11)}$ that involve the factor $(\!(20)L \,|\, (11)P + (11)P\!)$, it follows that

$$(\!(211)(20) \,|\, (110)(11) + (110)(11)\!) = 0. \quad (8)$$

Hence, all matrix elements of the spin–orbit interaction in the f shell vanish between states labeled (211)(20) and (110)(11). Results of this kind, though more powerful than those exemplified by Eq. (6), are correspondingly rarer. Equation (8) is in fact the only one of its kind that can be obtained for the spin–orbit interaction in f^n.

Interactions more complex than the spin–orbit interaction involve irreducible representations U and W of higher dimensionality. The frequent occurrence of like irreducible representations in the decomposition of $U \times U$ complicates the derivation of results analogous to Eq. (8), since the isoscalar factors $(\Gamma U'L'' \,|\, UL + UL')$ involve an additional classificatory symbol Γ. One way to be sure that such an isoscalar factor is nonvanishing for at least one pair (L, L') is to note that, on forming their squares and summing over L and L', the result must be unity.

[4] J. J. de Swart, Rev. Mod. Phys. 35, 916 (1963).

[5] A. G. McLellan, Proc. Phys. Soc. (London) 76, 419 (1960).

TABLE III. Decomposition of matrix elements of the spin–spin interaction for f^2.

Φ	Φ'	$(\Phi\|\| z_1 \|\|\Phi')$	$(\Phi\|\| z_2 \|\|\Phi')$	$(\Phi\|\| z_3 \|\|\Phi')$	$(\Phi\|\| z_4 \|\|\Phi')$
^3P	^3P	-9	63	0	1
^3P	^3F	$6(3)^{\frac{1}{2}}$	$48(3)^{\frac{1}{2}}$	$(3)^{-\frac{1}{2}}$	0
^3F	^3F	$-(14)^{\frac{1}{2}}$	$112(14)^{\frac{1}{2}}$	0	0
^3F	^3H	$2(22)^{\frac{1}{2}}$	$16(22)^{\frac{1}{2}}$	$-3(22)^{-\frac{1}{2}}$	0
^3H	^3H	$(143)^{\frac{1}{2}}$	$-7(143)^{\frac{1}{2}}$	0	$9(143)^{-\frac{1}{2}}$

Confronted, then, with a coefficient

$$(W'\Gamma U'L'' \,|\, WUL + WUL')$$

for which W' and U' occur in opposite symmetry parts of the decompositions of $W \times W$ and $U \times U$, we may deduce that the isoscalar factor

$$(W'\Gamma U' \,|\, WU + WU)$$

vanishes. In this way we can obtain analogs of Eq. (8).

5. SPIN–SPIN INTERACTION

As an example of an interaction whose matrix elements have not yet been tabulated for f^n, we consider the magnetic spin–spin interaction H_{ss} between pairs of electrons:

$$H_{ss} = 4\beta^2 \sum_{i>j} \left[\frac{\mathbf{s}_i \cdot \mathbf{s}_j}{r_{ij}^3} - \frac{3(\mathbf{r}_{ij} \cdot \mathbf{s}_i)(\mathbf{r}_{ij} \cdot \mathbf{s}_j)}{r_{ij}^5} \right].$$

The symbol β is the Bohr magneton. The decomposition of H_{ss} into parts corresponding to unique pairs of representations WU is carried out in a way precisely analogous to the decomposition already effected for the d shell.[6] For f electrons, we find that we can write

$$H_{ss} = h_1 + h_2 + h_3 + h_4, \tag{9}$$

where the operators h_r ($r = 1, 2, 3, 4$) correspond to $WU = (200)(20)$, $(220)(20)$, $(220)(21)$, and $(220)(22)$, respectively. The spin–spin interaction is the scalar part of a double tensor possessing ranks of 2 in the orbital and spin spaces; and so, for any of the operators h_r, the J dependence of a matrix element can be separated out as follows:

$$(\gamma SLJ|\, h_r \,|\gamma' S'L'J') = \delta(J, J')(-1)^{S'+L+J} \begin{Bmatrix} S' & L' & J \\ L & S & 2 \end{Bmatrix} (\gamma SL\|\, t_r^{(22)} \,\|\gamma' S'L').$$

All the matrix elements of H_{ss} can be rapidly calculated once the reduced matrix elements $(\Psi\|\, t_r^{(22)} \,\|\Psi')$ are known. Those for f^n are related to those for f^{n-1} through the equation

$$(\Psi\|\, t_r^{(22)} \,\|\Psi') = [n/(n-2)] \times \sum (\Psi\{\,|\,\bar{\Psi})(\bar{\Psi}\|\, t_r^{(22)} \,\|\bar{\Psi}')(\bar{\Psi}'\,|\,\}\Psi'), \tag{10}$$

where the sum runs over the states $\bar{\Psi}$ and $\bar{\Psi}'$ of f^{n-1}, and where $(\Psi\{\,|\,\bar{\Psi})$ and $(\bar{\Psi}'\,|\,\}\Psi')$ are coefficients of fractional parentage. These last have been tabulated by Nielson and Koster. The procedure is thus to set up a chain calculation, starting from f^2. This configuration is the simplest for which a two-particle operator has nonvanishing matrix elements. The matrix elements in question are conveniently defined by

$$(\Phi\|\, t_r^{(22)} \,\|\Phi') = a_r(\Phi\|\, z_r \,\|\Phi'),$$

where Φ and Φ' denote states of f^2. The quantities $(\Phi\|\, z_r \,\|\Phi')$ are set out in Table III, and the coefficients a_r, defined in terms of Marvin's radial integrals M^k,[7] are given by

$$a_1 = 4(55M^0 - 44M^2 - 50M^4)/165,$$
$$a_2 = (66M^2 - 175M^4)/770,$$
$$a_3 = 8(143M^2 - 175M^4)/77,$$
$$a_4 = -3(286M^2 + 175M^4)/22.$$

The matrix elements of H_{ss} for any configuration f^n can thus be calculated by repeated use of Eq. (10), taking the results for f^2 as a starting point. Of course, full use is made of the group-theoretical properties of the various component operators h_r; indeed, it is to take advantage of these properties that the decomposition (9) is made in the first place. If an operator h_r (corresponding to the irreducible representations U_r and W_r) is taken between states characterized by $WUSL$ and $W'U'S'L'$, then the matrix element will vanish unless the various identity representations occur at least once in the decompositions of all four triple Kronecker products $W \times W_r \times W'$, $U \times U_r \times U'$, $\mathfrak{D}_S \times \mathfrak{D}_2 \times \mathfrak{D}_{S'}$, and $\mathfrak{D}_L \times \mathfrak{D}_2 \times \mathfrak{D}_{L'}$. The last two products lead to the well-known selection rules ΔS, $\Delta L = 0, \pm 1, \pm 2$. Such applications of group theory do not concern us here, however. Our central theme is the usefulness of Tables I and II in deriving new selection rules. These can be obtained by the methods of Secs. 3 and 4, and are listed in Table IV. As in Sec. 3, all inessential quantum numbers are suppressed. The matrix elements in which L does

[6] B. R. Judd, Physica 33, 174 (1967).

[7] H. H. Marvin, Phys. Rev. 71, 102 (1947).

TABLE IV. Vanishing matrix elements of the operators h_r of the spin–spin interaction.

$((30)G\|\, h_r \,\|(20)D)$	$(r = 1, 2)$
$((31)L\|\, h_r \,\|(20)D)$	$(r = 1, 2;\ L$ even$)$
$((40)F\|\, h_r \,\|(20)D)$	$(r = 1, 2)$
$((40)F\|\, h_3 \,\|(21)D)$	
$((40)F\|\, h_4 \,\|(22)D)$	
$((30)G\|\, h_4 \,\|(22)D)$	
$((31)L\|\, h_4 \,\|(22)D)$	(L even)
$((220)L\|\, h_1 \,\|(200)D)$	(L odd)
$(WL\|\, h_2 \,\|(220)(20)D)$	[if $W = (222)$, L odd;
$(WL\|\, h_3 \,\|(220)(21)D)$	or $W = (211)$, L even;
$(WL\|\, h_4 \,\|(220)(22)D)$	or $W = (221)$, L even]
$(W(10)\|\, h_2 \,\|(220)(20))$	$[W = (111), (222)]$
$(W(20)\|\, h_2 \,\|(220)(20))$	$[W = (211), (221)]$
$((222)(30)\|\, h_2 \,\|(220)(20))$	
$(W(10)\|\, h_3 \,\|(220)(21))$	$[W = (111), (222)]$
$(W(20)\|\, h_3 \,\|(220)(21))$	$[W = (211), (221)]$
$(W(20)\|\, h_4 \,\|(220)(22))$	$[W = (211), (221)]$
$((222)(30)\|\, h_4 \,\|(220)(22))$	

not appear are particularly useful, since they include many special cases. They are the analogs of Eq. (7).

6. CONCLUSION

From the extensive listing in Table IV, it is apparent that the method based on the symmetry or antisymmetry of product functions is a valuable tool for deriving new selection rules. It could evidently be applied to other types of interaction, such as the spin–other-orbit interaction. Since, for the spin–spin interaction, the rules are obtained for the component operator h_r rather than H_{ss} itself, it is apparent that these rules would appear in a direct physical way only if the radial integrals M^k are of such a magnitude that one of the coefficients a_r dominates all the others. Calculations of the M^k for rare-earth ions[8] indicate that a_1 comprises positive and negative parts that nearly cancel, but this is the only obvious simplification that can be made. This means that we must be content to apply the selection rules to the operators h_r rather than to H_{ss}.

It is perhaps worth mentioning that a few of the selection rules that have been discussed in this paper can be obtained by alternative methods. Equation (7), for example, was obtained a few years ago[9] by studying the commutation relation between two tensors of the type $\mathbf{V}^{(11)}$ and $\mathbf{U}^{(3)}$. It appears that this method, which is completely different from the approach developed above, cannot be generalized to reproduce more than a few isolated zeros. Although the method based on the separation of $U \times U$ and $W \times W$ into symmetrical and antisymmetrical parts is much more powerful, yet there remains a significant number of unexplained zeros in the tables of Nielson and Koster. Whether some other symmetry properties have yet to be discovered is a matter for speculation.

[8] M. Blume, A. J. Freeman, and R. E. Watson, Phys. Rev. **134**, A320 (1964).

[9] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill Book Company, Inc., New York, 1963), p. 221.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Timelike Cluster Properties in Nonrelativistic Scattering

JOHN R. TAYLOR
Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado

(Received 16 February 1967)

It is generally assumed that experiments which are sufficiently separated in space or time should be mutually independent. In scattering theory this independence is expressed by the cluster decomposition of the S matrix. It has been proved by Hunziker that for experiments which are separated in space the cluster property does hold in nonrelativistic scattering. The purpose of this paper is to prove the same result for experiments which are separated in time.

I. INTRODUCTION

THE cluster decomposition of the S matrix[1,2] expresses the independence of experiments which are sufficiently separated in space or time. It asserts that the probability of any specified outcomes for two well-separated experiments should be the product of the two separate probabilities. A precise statement is as follows: Let $(g' \leftarrow f')$ and $(g'' \leftarrow f'')$ denote any two physical processes leading from initial states f' and f'' to final states g' and g''. Let $(g'_\xi \leftarrow f'_\xi)$ denote the process obtained by rigidly translating the process $(g' \leftarrow f')$ through the 4-vector ξ, and regard the two processes $(g'_\xi \leftarrow f'_\xi)$ and $(g'' \leftarrow f'')$ as a single composite process $(g'_\xi \,\&\, g'' \leftarrow f'_\xi \,\&\, f'')$. The cluster property simply asserts that, as $\xi \to \infty$, the probability for this composite process should satisfy

$$P(g'_\xi \,\&\, g'' \leftarrow f'_\xi \,\&\, f'') \xrightarrow[\xi\to\infty]{} P(g' \leftarrow f')P(g'' \leftarrow f''). \tag{1.1}$$

This must be formulated in the language of quantum scattering theory, where the asymptotic states are in one-to-one correspondence with the rays in a Hilbert space $\mathcal{H}_{\text{asym}}$. The probability amplitude for a process $(g \leftarrow f)$ is the matrix element of the scattering operator S between the corresponding vectors g and f; i.e.,

$$P(g \leftarrow f) = |\langle g, Sf\rangle|^2. \tag{1.2}$$

A composite state of the type $f' \,\&\, f''$ is represented by the tensor product $f' \otimes f''$, and so the quantum-mechanical form of the cluster property (1.1) is just

$$|\langle (g'_\xi \otimes g''), S(f'_\xi \otimes f'')\rangle| \xrightarrow[\xi\to\infty]{} |\langle g', Sf'\rangle\langle g'', Sf''\rangle|. \tag{1.3}$$

The cluster property (1.3) can be considered from two distinct points of view. If one is trying to *construct* a scattering theory, then one may wish to adopt the cluster property as a basic assumption. This is the procedure in analytic S-matrix theory, where it is shown[2] that the limit (1.3) for the moduli of S-matrix elements leads to the corresponding limit for the matrix elements themselves[3]:

$$\langle (g'_\xi \otimes g''), S(f'_\xi \otimes f'')\rangle \xrightarrow[\xi\to\infty]{} \langle g', Sf'\rangle\langle g'', Sf''\rangle. \tag{1.4}$$

This limit in turn leads to the well-known momentum-space cluster equations which are the starting point of the postulate of analyticity.

On the other hand, if one already has a complete dynamical theory (such as quantum field theory or nonrelativistic Schrödinger theory), then, having recognized the significance of the limit (1.3), one must obviously verify that the limit does hold in the given theory. This has been done for quantum field theory by Hepp, who has shown that as $\xi \to \infty$, in either a spacelike[4] or a timelike[5] direction, not only the limit (1.3) but also (1.4) are always valid. (The essential assumptions are that the field theory be almost local and that all fields have mass greater than zero.) In nonrelativistic Schrödinger theory Hunziker has shown[6] that the limit (1.4) holds when $\xi \to \infty$ in a purely spacelike direction. (In this case the essential condition is that all potentials be short range.)

It is the purpose of this paper to fill the obvious remaining gap in this second line of research; namely, to prove for nonrelativistic Schrödinger theory that the limit (1.4) holds when $\xi \to \infty$ in a timelike direction. This is done under approximately the same conditions used by Hunziker; specifically, that all particles interact by two-body potentials which are square integrable in the relative coordinate.[7]

[1] E. H. Wichmann and J. H. Crichton, Phys. Rev. **132**, 2788 (1963).
[2] J. R. Taylor, Phys. Rev. **142**, 1236 (1966).
[3] In order to deduce Eq. (1.4), it is necessary to strengthen the limit (1.3) either to include some statement of the rate at which the limit is approached or to include some uniformity for separations into three distinct processes. (See Ref. 2.) This strengthening of the limit presumably eliminates long-range interactions such as the Coulomb force.
[4] K. Hepp, Helv. Phys. Acta **37**, 659 (1964).
[5] K. Hepp, J. Math. Phys. **6**, 1762 (1965).
[6] W. Hunziker, J. Math. Phys. **6**, 6 (1965).
[7] Hunziker (Ref. 6) uses the slightly weaker condition of local square integrability together with the requirement that all potentials be $O(r^{-s})$, $s > 1$, at large distances. The question of how weak the conditions for cluster decomposition can be made is somewhat academic since, as discussed below, stronger conditions are needed to guarantee asymptotic completeness.

In Sec. II, I outline all the necessary notations. An N-particle system is defined, with states represented by vectors in the appropriate Hilbert space:

$$\mathcal{H} = \mathfrak{L}_2(R^{3N}).$$

Following the notation of Jauch,[8] the Møller wave operators $\Omega_{\pm}^{\alpha}$ are defined as limits of the channel operators $\Omega^{\alpha}(t)$ and are used to construct a scattering operator S. In order that S be unitary and have the property (1.2), it is necessary to introduce an auxiliary space $\mathcal{H}_{\text{asym}}$. This new space labels the asymptotic initial and final states of the system[9] and is the space on which S acts as a unitary operator. Finally, as a preliminary to the proof of the cluster property, the set of N particles is split into two subsets, or clusters, C' and C'' [corresponding to the states f' and f'' in Eq. (1.4)] and the operator $T'(\tau)$, which delays the motion of the subset C', is defined. This operator is characterized by the relation

$$T'(\tau)(f' \otimes f'') = f'_{\tau} \otimes f'',$$

where, as in Eq. (1.4), f'_{τ} represents the asymptotic free state f' delayed by time τ (τ positive or negative).

Section III contains the proof of the cluster property. This is given in three stages. The first and principal step, stated as a theorem, establishes the cluster property, as a strong operator limit, for the operators $\Omega^{\alpha}(t)$. This has a corollary the same result for the Møller wave operators $\Omega_{\pm}^{\alpha}$, and this in turn leads to the cluster property of the S matrix.

Since the mathematical details of the proof are quite complicated, it is appropriate to describe physically what happens and, indeed, why one expects the cluster property always to be valid. In most cases this is easily understood. If, for example, we do an experiment on radioactive decay today and a scattering experiment in the same place tomorrow, then tomorrow's experiment is unaffected by today's since all the decay products have dispersed before the incident scattering particles arrive. The independence of these two experiments is as true in classical mechanics as in quantum mechanics and may be called the *classical cluster property*. However, in special cases one of the decay products of today's experiment may be left at rest, in which case it will still be in our laboratory tomorrow and may be expected to interfere with tomorrow's experiment. In classical mechanics the cluster property does not hold for this special case. In quantum mechanics it is saved by the phenomenon of wave-packet spreading. If we delay tomorrow's experiment long enough, the remaining decay product has spread out too much to influence appreciably the second experiment. In this case the independence of the two experiments is a purely quantum effect, and may be called the *quantum cluster property*.

The proof of the cluster property given in this paper depends only on the latter effect—the spreading of wave packets—which is, of course, present in all cases. Rather than prove the result directly for any set of states, it is convenient first to consider Gaussian wave packets, whose familiar $t^{-\frac{3}{2}}$ spreading is explicitly known. [See Eq. (3.10) below.] Once established for Gaussians, the result is easily extended to arbitrary wavefunctions by approximating the latter with a sum of the former.

II. NOTATION

The system to be considered consists of N distinguishable spinless particles which are labeled by $i = 1, 2, \cdots, N$ and have Hamiltonian

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + \sum_{i<j} V_{ij}(x_{ij}), \tag{2.1}$$

where p_i, x_i, m_i are the momentum, position, and mass of the ith particle, and $x_{ij} = (x_i - x_j)$. I assume throughout that the potentials $V_{ij}(x_{ij})$ are square integrable, which is sufficient to insure that H is self-adjoint[10] on the Hilbert space $\mathcal{H} = \mathfrak{L}_2(R^{3N})$. A simplifying feature of nonrelativistic scattering is that transitions between states with different particles are impossible. This allows one to confine attention to a single N-particle system as just defined.

The central idea in scattering theory is that any collision state, if followed far enough into the past or future, should resemble a free state, on which the interparticle potentials have no effect. The well-known difficulties which occur with more than two particles ($N > 2$) arise from the possibility that some subset of $n < N$ particles may be able to form a bound state, in which case this subset may enter a scattering experiment either as n free particles or as one freely moving bound state. Both possibilities must be allowed for, and in the latter case the interactions V_{ij} within the subset never become ineffective, however large $|t|$. One is led to introduce the concept of an *asymptotic channel*. A channel α is defined by its $n_{\alpha} \leq N$ freely moving fragments, each fragment being either one of the original N particles or a

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

[9] As is well known, the asymptotic states of an N-particle system ($N > 2$) are not in (1–1) correspondence with the rays of the N-particle Hilbert space $\mathcal{H}$. It is to remedy this defect that the space $\mathcal{H}_{\text{asym}}$ has been introduced by, for example, L. D. Fadeev in "Mathematical Problems of the Quantum Theory of Scattering for a Three-Particle System," Stoklow Mathematical Institute, Publication No. 69 (1963), and F. Coester, Helv. Phys. Acta **38**, 7 (1965).

[10] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).

definite bound state of some of them. Corresponding to each channel α, there is a channel Hamiltonian

$$H_\alpha = \sum_{k=1}^{n_\alpha} \left(\frac{P_{\alpha k}^2}{2M_{\alpha k}} + E_{\alpha k} \right), \tag{2.2}$$

where $P_{\alpha k}$, $M_{\alpha k}$, $E_{\alpha k}$ denote the momentum of the kth fragment in the channel α, its mass, and minus its binding energy (which is zero if the fragment happens to be a single particle).

In the Schrödinger picture, any state of the system has time dependence $\exp(-iHt)F$, with F in $\mathcal{H}$. One expects that certain of these states should behave, as $t \to \pm\infty$, like free states $\exp(-iH_\alpha t)f_\pm$ in the channel α; and it can in fact be shown[11] under the stated assumptions that there are domains $\mathcal{D}^\alpha$ of $\mathcal{H}$ such that to each f in $\mathcal{D}^\alpha$ there corresponds states $F_\pm$ in $\mathcal{H}$ for which

$$e^{-iHt}F_\pm \xrightarrow[t\to\pm\infty]{} e^{-iH_\alpha t}f. \tag{2.3}$$

This means that the states f in $\mathcal{D}^\alpha$ label the possible initial and final free states in the channel α. If we define the unitary operator

$$\Omega^\alpha(t) = e^{iHt}e^{-iH_\alpha t},$$

the result (2.3) implies that $\Omega^\alpha(t)$ has strong limits $\Omega_\pm^\alpha$ on $\mathcal{D}^\alpha$ as $t \to \pm\infty$. The significance of the Møller operators $\Omega_\pm^\alpha$ is that if the system is initially in the free state f of channel α (or finally in the state g of channel β), then at $t = 0$ the actual state of the system is

$$F = \Omega_-^\alpha f \quad (\text{or } G = \Omega_+^\beta g).$$

The transition amplitude $S_{\beta\alpha}(g,f)$ from an initial state f in channel α to a final state g in channel β is, therefore,

$$\begin{aligned} S_{\beta\alpha}(g,f) &= \langle G, F\rangle \\ &= \langle \Omega_+^\beta g, \Omega_-^\alpha f\rangle \quad (f\in\mathcal{D}^\alpha,\ g\in\mathcal{D}^\beta). \end{aligned} \tag{2.4}$$

We now wish to define a single operator S whose matrix elements $\langle g, Sf\rangle$ are the transition amplitudes (2.4). The difficulty is that the various domains $\mathcal{D}^\alpha$, $\alpha = 0, 1, 2\cdots$, can overlap, and any vector f lying in both $\mathcal{D}^\alpha$ and $\mathcal{D}^\beta$ ($\alpha \neq \beta$) represents two distinct asymptotic states when regarded as a member first of $\mathcal{D}^\alpha$ and then of $\mathcal{D}^\beta$. To remove this ambiguity, it is natural to define a new space $\mathcal{H}_{\text{asym}}$ as follows: Firstly, for each channel α we introduce a distinct space $\mathcal{D}'^\alpha$ isomorphic to $\mathcal{D}^\alpha$. Corresponding vectors of $\mathcal{D}^\alpha$ and $\mathcal{D}'^\alpha$ label the same asymptotic state in channel α; but since the new spaces $\mathcal{D}'^\alpha$, $\alpha = 0, 1, 2\cdots$, are all distinct, each new vector labels a unique asymptotic state. We now define

$$\mathcal{H}_{\text{asym}} = \oplus_\alpha \mathcal{D}'^\alpha,$$

which, to the extent that the domains $\mathcal{D}^\alpha$ overlap (as subspaces of $\mathcal{H}$), is a space larger than $\mathcal{H}$. Each vector of $\mathcal{H}_{\text{asym}}$ labels a unique asymptotic state; those which lie in the subspaces $\mathcal{D}'^\alpha$ enjoy a privileged role since they belong to a definite channel, but, as we shall see directly, every vector f in $\mathcal{H}_{\text{asym}}$ has a unique image $\Omega_\pm f$ in the space $\mathcal{H}$ of the actual system. Before showing this, it is convenient to agree to omit the prime in $\mathcal{D}'^\alpha$; this leads to no serious ambiguities.

Linear Møller operators $\Omega_\pm$ are defined on $\mathcal{H}_{\text{asym}}$ in the natural way; that is, for $f\in\mathcal{D}^\alpha$,

$$\Omega_\pm f = \Omega_\pm^\alpha f.$$

Since the $\Omega_\pm^\alpha$ are isometric (and hence uniformally bounded), this defines unique maps of $\mathcal{H}_{\text{asym}}$ into $\mathcal{H}$. Further, since the orthogonality theorem[8] establishes that the ranges $\mathcal{R}_\pm^\alpha$ of the $\Omega_\pm^\alpha$ are orthogonal,

$$\mathcal{R}_+^\alpha \perp \mathcal{R}_+^\beta \quad \text{and} \quad \mathcal{R}_-^\alpha \perp \mathcal{R}_-^\beta \quad (\alpha \neq \beta),$$

the maps $\Omega_\pm$ of $\mathcal{H}_{\text{asym}}$ into $\mathcal{H}$ are actually isometric maps of $\mathcal{H}_{\text{asym}}$ *onto* their ranges $\mathcal{R}_\pm$:

$$\Omega_\pm : \mathcal{H}_{\text{asym}} = \oplus_\alpha \mathcal{D}^\alpha \xrightarrow{(1-1)} \mathcal{R}_\pm = \oplus_\alpha \mathcal{R}_\pm^\alpha \subset \mathcal{H}.$$

The operators $\Omega_\pm$ have adjoints defined on $\mathcal{H}$, and so we can define

$$S = \Omega_+^*\Omega_-,$$

which maps $\mathcal{H}_{\text{asym}}$ into itself. According to Eq. (2.4), the transition amplitude $S_{\beta\alpha}(g,f)$ is just the matrix element

$$S_{\beta\alpha}(g,f) = \langle g, Sf\rangle \quad (f\in\mathcal{D}^\alpha,\ g\in\mathcal{D}^\beta).$$

It should be emphasized that, although the condition that the potentials V_{ij} be $\mathfrak{L}_2$ is sufficient for the existence of the S operator and for the proof of the cluster property in Sec. III, we do not have a physically meaningful scattering theory unless we can also guarantee asymptotic completeness; i.e., that the ranges $\mathcal{R}_\pm$ of $\Omega_\pm$ satisfy

$$\mathcal{R}_+ = \mathcal{R}_- = \mathcal{M},$$

where $\mathcal{M}$ denotes the continuum subspace of H in $\mathcal{H}$. This condition is needed to ensure that the asymptotic states of $\mathcal{H}_{\text{asym}}$ are in fact *all* the possible asymptotic scattering states of the system. Unfortunately, no general proof of asymptotic completeness has been found and in the cases $N = 2$ and 3, where proofs are known, the conditions on the potentials are stronger than those used here.[12]

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

[12] For $N = 2$, S. T. Kuroda [Nuovo Cimento 12, 431 (1959)] has established asymptotic completeness when V is both $\mathfrak{L}_2$ and $\mathfrak{L}_1$. T. Ikebe [Arch. Ratl. Mech. Anal. 5, 1 (1960)] has established the same result using slightly weaker conditions. The case $N = 3$ has been treated by Fadeev. (See Footnote 9.)

If we choose to *assume* that the theory is asymptotically complete, then, since both maps

$$\Omega_-: \mathcal{K}_{\text{asym}} \xrightarrow{(1-1)} \mathcal{R}_- \qquad (\mathcal{R}_- = \mathcal{R}_+)$$
$$\Omega_+: \mathcal{K}_{\text{asym}} \xrightarrow{(1-1)} \mathcal{R}_+$$

are isometric, the operator $S = \Omega_+^*\Omega_-$ is unitary on $\mathcal{K}_{\text{asym}}$.

Having set up the necessary scattering formalism, we must introduce two more concepts essential for the proof of the cluster property. The first of these is the operator $T(\tau)$, which delays an asymptotic free state, and the second is the notion of a cluster. It seems reasonably obvious that the generator of time delays for an asymptotic free state in channel α should be the corresponding free Hamiltonian $\mathcal{K}_\alpha$. To see that this is really so, let us consider two asymptotic states f in $\mathfrak{D}^\alpha$ and

$$f_\tau = e^{iH_\alpha\tau}f \equiv T_\alpha(t)f. \tag{2.5}$$

It is easily seen that f_τ is also in $\mathfrak{D}^\alpha$, and so the corresponding actual states at $t = 0$ are $F = \Omega^\alpha f$ and $F_\tau = \Omega^\alpha f_\tau$. Using the intertwining relation[8]

$$\Omega^\alpha e^{iH_\alpha\tau} = e^{iH\tau}\Omega^\alpha, \tag{2.6}$$

we can rewrite the latter as

$$F_\tau = \Omega^\alpha e^{iH_\alpha\tau}f = e^{iH\tau}F;$$

that is, the actual state defined by F_τ is the same at $t = 0$ as was that defined by F at $t = -\tau$. This establishes that f_τ, as given by Eq. (2.5), is the correctly delayed asymptotic free state.

We next consider a partition of the N particles into two disjoint subsets, or clusters, C' and C''. Since we have to prove that, when the two clusters are well separated, each behaves as if the other were absent, it is natural to introduce cluster Hamiltonians

$$H' = \sum_i{}' \frac{p_i^2}{2m_i} + \sum_{i<j}{}' V_{ij}(x_{ij})$$

and similarly H'', where $\sum'$ denotes summation over those i and j contained in C'. The Hamiltonian H' determines the behavior of the cluster C' in the absence of C'' and vice versa. Clearly

$$H = H' + H'' + V'^{,''},$$

where $V'^{,''}$ is made up of all potentials which link the two clusters.

We shall say that the clusters C' and C'' are *consistent* with a channel α if each fragment of α is entirely contained in either C' or C''. In this case the channel Hamiltonian (2.2) can be split up as $H_\alpha = H'_\alpha + H''_\alpha$, where H'_α contains the energy of those fragments belonging to C', and similarly H''_α. In terms of these we define Møller wave operators for each separate cluster:

$$\Omega_\pm^{\alpha'} = \lim_{t\to\pm\infty} \Omega^{\alpha'}(\tau),$$

where

$$\Omega^{\alpha'}(\tau) = e^{iH'\tau}e^{-iH'_{\alpha'}\tau},$$

and similarly for C''. Since these operators can be defined for any channel α which is consistent with the clusters C' and C'', they define wave operators $\Omega'_\pm$ on the whole subspace of $\mathcal{K}_{\text{asym}}$ spanned by these channels.

If α is a channel consistent with C' and C'' and $f' \otimes f''$ is one of its asymptotic states, then, according to Eq. (2.5), the operator which delays f' (while leaving f'' alone) is

$$T'_\alpha(\tau) = e^{iH'_\alpha\tau}.$$

Just as the operators $\Omega^{\alpha'}$ led to Ω', so the operators $T'_\alpha(\tau)$ define a delay operator $T'(\tau)$ on the whole subspace of $\mathcal{K}_{\text{asym}}$ consistent with the clusters[13] C' and C''.

Finally, it is convenient for the proofs of Sec. III to note that the Hamiltonians H' and H'_α can be written as

$$H' = P'^2/2M' + h'$$

and

$$H'_\alpha = P'^2/2M' + h'_\alpha, \tag{2.7}$$

and similarly H'' and H''_α, where P' and M' are the total momentum and mass of the cluster C' and h' and h'_α are self-adjoint "internal" Hamiltonians which commute with P' (although not with each other).

III. PROOF OF THE CLUSTER PROPERTY

The principal step in the proof of the timelike cluster property is the following theorem, whose proof is given at the end of this section:

Theorem:

$$[\Omega^\alpha(t) - \Omega^{\alpha'}(t)\Omega^{\alpha''}(t)]T'_\alpha(\tau) \xrightarrow[\tau\to\pm\infty]{} 0 \tag{3.1}$$

in the strong sense on $\mathcal{K}$ and uniformly in $-\infty < t < \infty$.

This theorem has the following corollary, whose proof is also deferred to the end of the section.

Corollary:

$$[\Omega_\pm^\alpha - \Omega_\pm^{\alpha'}\Omega_\pm^{\alpha'}]T'_\alpha(\tau) \xrightarrow[\tau\to\pm\infty]{} 0 \tag{3.2}$$

strongly on $\mathfrak{D}^\alpha$.

[13] This is the largest subspace on which one wants to define $T'(\tau)$, since it is physically meaningless to separate states in the channel α into clusters C', C'' with which α is inconsistent. (For example, if particles 1 and 2 belong to the same bound state of channel α, it is senseless to delay the arrival of 1 but not 2.)

Since this holds for every channel α which is consistent with clusters C' and C'', it is in fact true on the whole subspace of $\mathcal{K}_{\text{asym}}$ spanned by these channels; i.e.,

$$[\Omega_\pm - \Omega'_\pm\Omega''_\pm]T'(\tau) \xrightarrow[\tau\to\pm\infty]{} 0 \tag{3.3}$$

strongly on the subspace of $\mathcal{K}_{\text{asym}}$ consistent with C' and C''.

The result (3.3) is the timelike cluster decomposition of the Møller operators. From it we can now prove the cluster property of the S matrix (1.4):

$$\langle (g'''_\tau \otimes g''''), S(f'_\tau \otimes f'')\rangle \xrightarrow[\tau\to\pm\infty]{} \langle g''', Sf'\rangle\langle g'''', Sf''\rangle, \tag{3.4}$$

where

$$f'_\tau \otimes f'' = T'(\tau)f' \otimes f'',$$

as defined in Sec. II. The extra primes which appear in Eq. (3.4) are supposed to emphasize that the clusters C', C'' into which the initial state is split are not necessarily the same as C''', C'''' of the final state. The case $C' \neq C'''$ is relatively trivial. In this case f' and g''' contain different particles and the matrix element $\langle g''', Sf'\rangle$ vanishes, making the limit (3.4) zero. [This is just as it should be. The probability for the processes $(2N \leftarrow N)$ today and $(3N \leftarrow 4N)$ tomorrow must be zero, even though $(5N \leftarrow 5N)$ is perfectly possible.] When $C' = C'''$ (and hence $C'' = C''''$), we have split initial and final states into the same clusters and the separated processes $(g' \leftarrow f')$ and $(g'' \leftarrow f'')$ are both allowed. In this case the limit (3.4) is, in general, nonzero and is physically more interesting. We consider this latter case first.

A. Separation of Initial and Final States into Identical Clusters

$$\begin{aligned}
&\langle (g'_\tau \otimes g''), S(f'_\tau \otimes f'')\rangle \\
&\quad = \langle \Omega_+T'(\tau)(g' \otimes g''), \Omega_-T'(\tau)(f' \otimes f'')\rangle \\
&\quad \xrightarrow[\tau\to\pm\infty]{} \langle (\Omega'_+T'g' \otimes \Omega''_+g''), (\Omega'_-T'f' \otimes \Omega''_-f'')\rangle \\
&\qquad\qquad\qquad\qquad \text{[by Eq. (3.3)]} \\
&\quad = \langle \Omega'_+T'g', \Omega'_-T'f'\rangle\langle \Omega''_+g'', \Omega''_-f''\rangle \\
&\quad = \langle g', Sf'\rangle\langle g'', Sf''\rangle,
\end{aligned}$$

as required. [In passing from the second-to-last line to the last, I have used the intertwining relation (2.6), which implies that

$$\begin{aligned}
\langle \Omega'_+T'g', \Omega'_-T'f'\rangle &= \langle e^{iH'\tau}\Omega'_+g', e^{iH'\tau}\Omega'_-f'\rangle \\
&= \langle g', Sf'\rangle.]
\end{aligned}$$

B. Separation of Initial and Final States into Different Clusters

We have next to consider matrix elements of the form $\langle T'''g, ST'f\rangle$ where the clusters C' and C''' are different. As before,

$$\begin{aligned}
\langle T'''(\tau)g, ST'(\tau)f\rangle &= \langle \Omega_+T'''g, \Omega_-T'f\rangle \\
&\xrightarrow[\tau\to\pm\infty]{} \langle \Omega'''_\pm\Omega''''_+T'''g, \Omega'_-\Omega''_-T'f\rangle,
\end{aligned} \tag{3.5}$$

by Eq. (3.3). Since Ω'' and T' commute, we can now use the intertwining property for Ω' and T', and similarly for Ω''' and T'''. This gives for (3.5)

$$\langle e^{iH'''\tau}\Omega'''_+\Omega''''_+g, e^{iH'\tau}\Omega'_-\Omega''_-f\rangle. \tag{3.6}$$

Finally, we use the fact (whose proof is given at the end of the section) that, if the clusters C' and C''' are different, then the operator

$$e^{-iH'''\tau}e^{iH'\tau}$$

converges weakly to zero (i.e., its matrix elements go to zero) on $\mathcal{K}$ as $\tau \to \pm\infty$. This immediately shows that (3.6) tends to zero, which is the required result.

This completes the derivation of the timelike cluster properties, and it remains only to give the three proofs which have so far been omitted.

C. Proof of Theorem

The result (3.1) which we wish to prove can be rewritten as

$$\|[1 - e^{-iHt}e^{i(H'+H'')t}]e^{-i(H_\alpha'+H_\alpha'')t}e^{iH_\alpha'\tau}f\| \xrightarrow[\tau\to\pm\infty]{} 0 \tag{3.7}$$

for all f in $\mathcal{K}$ and uniformly in $-\infty < t < \infty$. It is sufficient to prove this on any dense set and we accordingly consider those f with wavefunctions

$$f(x_1, \cdots, x_N) = g'(X')g''(X'')\phi'(y')\phi''(y'').$$

Here X' and X'' denote the centers of mass of the two clusters and g' and g'' are Gaussians of the form

$$g(x) = e^{-\frac{1}{2}(x-a)^2} \tag{3.8}$$

with a any vector. The functions ϕ' and ϕ'' are arbitrary $\mathfrak{L}_2$ functions of the internal coordinates y' and y'' of the clusters C' and C''.

The first step is to rewrite the operator in square brackets $[\cdots]$ of Eq. (3.7) as the integral of its derivative:

$$[\cdots] = i\int_0^t e^{-iHt'}V'^{,''}e^{i(H'+H'')t'}\,dt'.$$

Using the fact that the norm of an integral is less than the integral of the norm, we find the following bound for the left-hand side of Eq. (3.7):

$$\begin{aligned}
\|(3.7)\| &< \int_{-\infty}^{\infty} \|V'^{,''}e^{i(H'+H'')t'}e^{-iH_\alpha'(t-\tau)}e^{-iH_\alpha''t}f\|\,dt' \\
&= \int_{-\infty}^{\infty} N(t', t, \tau)\,dt'.
\end{aligned} \tag{3.9}$$

The well-known property of Gaussian wave packets that

$$|(e^{-ip^2t/2m}g)(x)|^2 = \mu^{\frac{3}{2}}(t)e^{-\mu(t)(x-a)^2} = \mu^{\frac{3}{2}}g^2(\mu, x), \tag{3.10}$$

where

$$\mu(t) = (1 + t^2/m^2)^{-1}, \tag{3.11}$$

together with the expansions (2.7) allow us to evaluate the time dependence of the center-of-mass part of the wavefunction in Eq. (3.9):

$$|(e^{i(H'+H'')t'}e^{-iH'_\alpha(t-\tau)}e^{-iH'_\alpha t}f)(x_1 \cdots x_N)|^2 = [\mu'(t-\tau-t')\mu''(t-t')]^{\frac{3}{2}}\,|g'(\mu', X')g''(\mu'', X'') \times \phi'(y', t', t-\tau)\phi''(y'', t', t)|^2, \tag{3.12}$$

where the functions $\phi(y, t', t)$ are related to $\phi(y)$ by the unitary transformation,

$$\phi'(\cdot, t', t) = e^{ih't'}\,e^{-ih'_\alpha t}\phi(\cdot) \tag{3.13}$$

and similarly ϕ''.

Returning to Eq. (3.9), we note that $V'^{,''}$ is a finite sum of the potentials V_{ij} which link the two clusters. By the triangle inequality it is sufficient to consider the contribution of each V_{ij} separately. Using the result (3.12), we find for such a typical contribution

$$N_{ij}^2 = (\mu'\mu'')^{\frac{3}{2}}\int dX'\,dX''\,dy'\,dy'' \times |V_{ij}(x_{ij})g'(\mu', X')g''(\mu'', X'')\phi'\phi''|^2.$$

To evaluate this integral we replace the variable X'' by x_{ij}, in which case (because x_i and x_j belong to different clusters) $X'' = X' + \xi$, where ξ is some linear combination of the remaining variables. The integration over X' (i.e., the integration of the two Gaussians) can now be performed explicitly, giving an answer which is bounded by

$$\text{const}\,(\mu' + \mu'')^{-\frac{3}{2}}.$$

The integration over x_{ij} now gives a constant since $V_{ij}(x_{ij})$ is $\mathfrak{L}_2$, and those over y', y'' give some other constant since $\|\phi'\|$ and $\|\phi''\|$ are time-independent [Eq. (3.13)]. The result is therefore

$$N^2(t', t, \tau) < \text{const}\,[\mu'\mu''/(\mu' + \mu'')]^{\frac{3}{2}} = \text{const}\left[2 + \left(\frac{t-\tau-t'}{M'}\right)^2 + \left(\frac{t-t'}{M''}\right)^2\right]^{-\frac{3}{2}}.$$

Substitution of this bound into Eq. (3.9) and the simple change of variable from t' to $t' - t$ shows that the norm (3.7) satisfies

$$\|(3.7)\| < \text{const}\int_{-\infty}^{\infty} dt'\left[2 + \left(\frac{t'+\tau^2}{M'}\right) + \left(\frac{t'}{M''}\right)^2\right]^{-\frac{3}{4}}.$$

This bound is independent of t and tends to zero as $\tau \to \pm\infty$. Q.E.D.

D. Proof of Corollary

To prove the limit (3.2) we use the following three inequalities. Firstly, from the main theorem (3.1) it is clear that, given $\epsilon > 0$ and f in $\mathcal{H}$, we can find a τ_ϵ such that

$$\|[\Omega^\alpha(t) - \Omega^{\alpha'}(t)\Omega^{\alpha''}(t)]T'_\alpha(\tau)f\| < \tfrac{1}{3}\epsilon$$

for all $|\tau| > \tau_\epsilon$ and all t. If we now fix $|\tau| > \tau_\epsilon$ and $f \in \mathcal{D}^\alpha$, then the usual proof[11] of the convergence of $\Omega^\alpha(t)$ shows that[14]

$$\|[\Omega^\alpha(t) - \Omega^\alpha_+]T'_\alpha(\tau)f\| < \tfrac{1}{3}\epsilon \tag{3.14}$$

for sufficiently large positive t (and the same condition with Ω^α_- for large negative t). Similarly,[15]

$$\|[\Omega^{\alpha'}(t)\Omega^{\alpha''}(t) - \Omega^{\alpha'}_+\Omega^{\alpha''}_+]T'_\alpha(\tau)f\| < \tfrac{1}{3}\epsilon \tag{3.15}$$

for sufficiently large t. Combining these three inequalities and using the triangle inequality in the form

$$\|a - d\| < \|a - b\| + \|b - c\| + \|c - d\|,$$

we find that

$$\|[\Omega^\alpha_+ - \Omega^{\alpha'}_+\Omega^{\alpha''}_+]T'_\alpha(\tau)f\| < \epsilon$$

for any $|\tau| > \tau_\epsilon$. Q.E.D.

E. Proof of the Weak Convergence of $\exp(-iH'''\tau)\exp(iH'\tau)$

If the clusters C' and C''' are different, then there is at least one particle contained in C' but not in C''', or vice versa. We assume the former possibility, and then label this particle as number 1 and the remaining particles of C' as $2, \cdots, n$.

We wish to prove that

$$\langle e^{iH'''\tau}f_1, e^{iH'\tau}f_2\rangle \xrightarrow[\tau\to\pm\infty]{} 0$$

for any f_1 and f_2 in $\mathcal{H}$. Since it is sufficient to prove the result for any dense sets in $\mathcal{H}$, we consider wavefunctions of the form

$$f_1(x_1, \cdots, x_N) = g_1(x_1)\phi_1(x_2, \cdots, x_N)$$

and

$$f_2(x_1, \cdots, x_N) = g_2(X')\phi_2(y', x_{n+1}, \cdots, x_N),$$

where g_1 and g_2 are Gaussians of the form (3.8) and ϕ_1 and ϕ_2 are arbitrary $\mathfrak{L}_2$ functions of the remaining $N - 1$ variables. Just as in Eq. (3.12), the τ dependence of these wavefunctions can be evaluated. Because

[14] One must check that $T'_\alpha f$ is in $\mathcal{D}^\alpha$. This will be done directly. (See Footnote 15.)

[15] By setting $\tau = 0$ in the proof of the main theorem, one can see that the domains of Ω^α and $\Omega^{\alpha'}\Omega^{\alpha''}$ coincide. The latter domain is obviously invariant under T'_α and so, therefore, is the former. This justifies Eqs. (3.14) and (3.15). (See Footnote 14.)

x_1 is in C' but not in C''', the results are

$$|(e^{iH'''\tau}f_1)(x_1, \cdots, x_N)| = g_1(x_1)\,|\phi_1(x_2, \cdots, x_N, \tau)|$$

and

$$|(e^{iH'\tau}f_2)(x_1, \cdots, x_N)| = \mu^{\frac{1}{2}}(\tau)g_2(\mu, X')\,|\phi_2(y', \cdots, x_N, \tau)|,$$

where $\mu(\tau)$ and $g_2(\mu, X')$ are the same as in Eqs. (3.10) and (3.11) and $\phi_1(\cdot, \tau)$ and $\phi_2(\cdot, \tau)$ are related by unitary transformations to $\phi_1(\cdot)$ and $\phi_2(\cdot)$ [cf. Eq. (3.13)]. We note that

$$g_2(\mu X') \leq 1,$$

whence

$$|\langle e^{iH''\tau}f_1, e^{iH'\tau}f_2\rangle| < \mu^{\frac{1}{2}}(\tau)\int g_1(x_1)\,dx_1 \times \int |\phi_1(\cdots, \tau)\phi_2(\cdots, \tau)|\,dx_2 \cdots dx_N.$$

By the Schwartz inequality, this is

$$\leq (2\pi)^{\frac{3}{2}}\,\|\phi_1\|\,\|\phi_2\|\,(1 + \tau^2/M'^2)^{-\frac{3}{4}}.$$

Since $\|\phi_1\|$ and $\|\phi_2\|$ are independent of τ, this tends to zero as $\tau \to \pm\infty$. Q.E.D.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Invariant Imbedding Equations for the Dissipation Functions of an Inhomogeneous Finite Slab with Anisotropic Scattering*

R. BELLMAN, H. KAGIWADA, R. KALABA, AND S. UENO
The RAND Corporation, Santa Monica, California

(Received 30 March 1967)

Two complete sets of differential-integral equations are derived for the dissipation functions of an inhomogeneous finite slab which scatters anisotropically. The first set is for the case of monodirectional illumination of the upper boundary, and the second is for the monodirectional illumination of the lower boundary. Conservation equations link the reflection, transmission, and dissipation functions.

THE dissipation function is important to the study of scattering problems. Equations are derived for the dissipation functions of an inhomogeneous finite slab with anisotropic scattering. This paper may be of interest to those engaged in research in radiative transfer, neutron transport, and mathematical physics.

1. INTRODUCTION

In earlier studies[1-3] we have shown the importance of the dissipation function in various analytical studies of transport in a rod. Further, in our most recent study[4] we derived, by means of the particle-counting procedure, an equation for the dissipation function of a homogeneous slab with isotropic scattering, and wrote the conservation relation that relates the reflection, transmission, and dissipation functions.

Here, with the aid of the invariant imbedding technique, we derive a complete set of integro-differential equations for the dissipation functions of an inhomogeneous finite slab with anisotropic scattering, using the equation of transfer and taking into account the polarity of the optical properties of the medium. Furthermore, we obtain the conservation relation for the reflection, transmission, and dissipation functions.

* This research is sponsored by the U.S. Air Force under Project RAND—Contract No. F44620-67-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq. USAF. Views or conclusions contained in this paper should not be interpreted as representing the official opinion or policy of the U.S. Air Force.

[1] R. E. Bellman, K. L. Cooke, R. E. Kalaba, and G. M. Wing, "Existence and Uniqueness Theorems in Invariant Imbedding—I: Conservation Principles," The RAND Corporation, RM-3611-ARPA, May 1963.

[2] R. E. Bellman, H. H. Kagiwada, R. E. Kalaba, and M. C. Prestrud, *Invariant Imbedding and Time-dependent Transport Processes* (American Elsevier Publishing Company, Inc., New York, 1964).

[3] R. E. Bellman, R. E. Kalaba, and G. M. Wing, Proc. Natl. Acad. Sci. 46, 1258 (1960).

[4] R. E. Bellman, H. H. Kagiwada, R. E. Kalaba, and S. Ueno, "The Invariant Imbedding Equations for the Dissipation Function of a Homogeneous Finite Slab," Quart. Appl. Math. (to be published).

x_1 is in C' but not in C''', the results are

$$|(e^{iH'''\tau}f_1)(x_1, \cdots, x_N)| = g_1(x_1)\,|\phi_1(x_2, \cdots, x_N, \tau)|$$

and

$$|(e^{iH'\tau}f_2)(x_1, \cdots, x_N)| = \mu^{\frac{1}{2}}(\tau)g_2(\mu, X')\,|\phi_2(y', \cdots, x_N, \tau)|,$$

where $\mu(\tau)$ and $g_2(\mu, X')$ are the same as in Eqs. (3.10) and (3.11) and $\phi_1(\cdot, \tau)$ and $\phi_2(\cdot, \tau)$ are related by unitary transformations to $\phi_1(\cdot)$ and $\phi_2(\cdot)$ [cf. Eq. (3.13)]. We note that

$$g_2(\mu X') \leq 1,$$

whence

$$|\langle e^{iH''\tau}f_1, e^{iH'\tau}f_2\rangle| < \mu^{\frac{1}{2}}(\tau)\int g_1(x_1)\,dx_1 \times \int |\phi_1(\cdots, \tau)\phi_2(\cdots, \tau)|\,dx_2 \cdots dx_N.$$

By the Schwartz inequality, this is

$$\leq (2\pi)^{\frac{3}{2}}\,\|\phi_1\|\,\|\phi_2\|\,(1 + \tau^2/M'^2)^{-\frac{3}{4}}.$$

Since $\|\phi_1\|$ and $\|\phi_2\|$ are independent of τ, this tends to zero as $\tau \to \pm\infty$. Q.E.D.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Invariant Imbedding Equations for the Dissipation Functions of an Inhomogeneous Finite Slab with Anisotropic Scattering*

R. BELLMAN, H. KAGIWADA, R. KALABA, AND S. UENO
The RAND Corporation, Santa Monica, California

(Received 30 March 1967)

Two complete sets of differential-integral equations are derived for the dissipation functions of an inhomogeneous finite slab which scatters anisotropically. The first set is for the case of monodirectional illumination of the upper boundary, and the second is for the monodirectional illumination of the lower boundary. Conservation equations link the reflection, transmission, and dissipation functions.

THE dissipation function is important to the study of scattering problems. Equations are derived for the dissipation functions of an inhomogeneous finite slab with anisotropic scattering. This paper may be of interest to those engaged in research in radiative transfer, neutron transport, and mathematical physics.

1. INTRODUCTION

In earlier studies[1-3] we have shown the importance of the dissipation function in various analytical studies of transport in a rod. Further, in our most recent study[4] we derived, by means of the particle-counting procedure, an equation for the dissipation function of a homogeneous slab with isotropic scattering, and wrote the conservation relation that relates the reflection, transmission, and dissipation functions.

Here, with the aid of the invariant imbedding technique, we derive a complete set of integro-differential equations for the dissipation functions of an inhomogeneous finite slab with anisotropic scattering, using the equation of transfer and taking into account the polarity of the optical properties of the medium. Furthermore, we obtain the conservation relation for the reflection, transmission, and dissipation functions.

* This research is sponsored by the U.S. Air Force under Project RAND—Contract No. F44620-67-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq. USAF. Views or conclusions contained in this paper should not be interpreted as representing the official opinion or policy of the U.S. Air Force.

[1] R. E. Bellman, K. L. Cooke, R. E. Kalaba, and G. M. Wing, "Existence and Uniqueness Theorems in Invariant Imbedding—I: Conservation Principles," The RAND Corporation, RM-3611-ARPA, May 1963.

[2] R. E. Bellman, H. H. Kagiwada, R. E. Kalaba, and M. C. Prestrud, *Invariant Imbedding and Time-dependent Transport Processes* (American Elsevier Publishing Company, Inc., New York, 1964).

[3] R. E. Bellman, R. E. Kalaba, and G. M. Wing, Proc. Natl. Acad. Sci. 46, 1258 (1960).

[4] R. E. Bellman, H. H. Kagiwada, R. E. Kalaba, and S. Ueno, "The Invariant Imbedding Equations for the Dissipation Function of a Homogeneous Finite Slab," Quart. Appl. Math. (to be published).

2. MONODIRECTIONAL ILLUMINATION OF THE UPPER BOUNDARY

2.1. Equation of Transfer

Consider a plane-parallel, inhomogeneous, and anisotropically scattering atmosphere of finite optical thickness $(\tau_1 - \tau_0)$. Suppose that parallel rays of constant net flux π per unit area normal to the direction of the propagation are falling on the upper surface $\tau = \tau_0$ in the direction $(-\mu_0, \varphi_0)$, where μ_0 $(0 < \mu_0 \leq 1)$ is the cosine of the polar angle measured from the inward directed normal and the parameter φ_0 is the azimuth angle $(0 \leq \varphi_0 \leq 2\pi)$. Directions are measured with respect to the upward normal.

We shall use the notation $I(\tau, +\mu, \varphi)$ $(0 < \mu \leq 1, 0 \leq \varphi \leq 2\pi)$ to denote an upwelling intensity, and $I(\tau, -\mu, \varphi)$ $(0 < \mu \leq 1, 0 \leq \varphi \leq 2\pi)$ to denote a downwelling intensity at level τ. In this case, the polar angle is measured from the upward-directed vertical at the upper boundary. We shall also write $I_+(\tau, \Omega) = I(\tau, +\mu, \varphi)$ and $I_-(\tau, \Omega) = I(\tau, -\mu, \varphi)$.

The equation of transfer appropriate to the nonclassical radiation field takes the form

$$\mu[dI(\tau, \Omega)/d\tau] = I(\tau, \Omega) - J(\tau, \Omega), \tag{1}$$

where Ω stands for (μ, φ) $(-1 \leq \mu \leq 1, 0 \leq \varphi \leq 2\pi)$, and J represents the source function

$$J(\tau, \Omega) = \frac{\lambda(\tau)}{4\pi}\int \gamma(\tau, \Omega, \Omega')I(\tau, \Omega')\, d\Omega', \tag{2}$$

where $d\Omega' = d\mu'\, d\varphi'$ and the integration intervals for μ' and φ' are $(-1, +1)$ and $(0, 2\pi)$, respectively. In Eq. (2), λ represents the albedo for single scattering and γ is the phase function normalized to 4π on the unit sphere. The nonclassical radiation field consists of the diffuse radiation field and the reduced incident radiation.

Equation (1) should be solved subject to the boundary conditions

$$I_-(\tau_0, \Omega) = \pi\delta(\mu - \mu_0)\delta(\varphi - \varphi_0), \tag{3}$$

$$I_+(\tau_1, \Omega) = 0. \tag{4}$$

2.2. Scattering and Transmission Functions

On extending the invariant imbedding procedure given in preceding papers[5,6] to the case of anisotropic scattering, and writing

$$I_+(\tau_0, \Omega) = (S/4\mu)(\tau_0, \tau_1; \Omega, \Omega_0), \tag{5}$$

$$I_-(\tau_1, \Omega) = (T/4\mu)(\tau_0, \tau_1; \Omega, \Omega_0) + \pi\delta(\mu - \mu_0)\delta(\varphi - \varphi_0)e^{-(\tau_1-\tau_0)/\mu}, \tag{6}$$

where Ω_0 stands for (μ_0, φ_0) $(0 < \mu_0 \leq 1, 0 \leq \varphi_0 \leq 2\pi)$, the invariant imbedding equations for the scattering and transmission functions take the forms

$$\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)S(\tau_0, \tau_1; \Omega, \Omega_0) - \frac{\partial S}{\partial \tau_0}(\tau_0, \tau_1; \Omega, \Omega_0)$$
$$= \lambda(\tau_0)\Big\{\gamma(\tau_0, \Omega, -\Omega_0)$$
$$+ \frac{1}{4\pi}\int_- S(\tau_0, \tau_1; \Omega, \Omega')\gamma(\tau_0, \Omega', -\Omega_0)\frac{d\Omega'}{\mu'}$$
$$+ \frac{1}{4\pi}\int_+ \gamma(\tau_0, \Omega, \Omega'')S(\tau_0, \tau_1; \Omega'', \Omega_0)\frac{d\Omega''}{\mu''}$$
$$+ \frac{1}{16\pi^2}\int_-\int_+ S(\tau_0, \tau_1; \Omega, \Omega')\gamma(\tau_0, \Omega', \Omega'')$$
$$\times S(\tau_0, \tau_1; \Omega'', \Omega_0)\frac{d\Omega'}{\mu'}\frac{d\Omega''}{\mu''}\Big\} \tag{7}$$

and

$$\frac{1}{\mu_0}T(\tau_0, \tau_1; \Omega, \Omega_0) - \frac{\partial T}{\partial \tau_0}(\tau_0, \tau_1; \Omega, \Omega_0)$$
$$= \lambda(\tau_0)\Big\{\gamma(\tau_0, -\Omega, -\Omega_0)e^{-(\tau_1-\tau_0)/\mu}$$
$$+ \frac{1}{4\pi}\int_- T(\tau_0, \tau_1; \Omega, \Omega')\gamma(\tau_0, \Omega', -\Omega_0)\frac{d\Omega'}{\mu'}$$
$$+ \frac{1}{4\pi}e^{-(\tau_1-\tau_0)/\mu}\int_+ \gamma(\tau_0, \Omega, \Omega'')S(\tau_0, \tau_1; \Omega'', \Omega_0)\frac{d\Omega''}{\mu''}$$
$$+ \frac{1}{16\pi^2}\int_-\int_+ T(\tau_0, \tau_1; \Omega, \Omega')\gamma(\tau_0, \Omega', \Omega'')$$
$$\times S(\tau_0, \tau_1; \Omega'', \Omega_0)\frac{d\Omega'}{\mu'}\frac{d\Omega''}{\mu''}\Big\}, \tag{8}$$

where $-\Omega$ (or $-\Omega_0$) stands for $(-\mu, \varphi)$ (or $-\mu_0, \varphi_0$), and subscripts $+$ and $-$ on the integrals indicate that the integration in the half-range is over positive or negative values of μ' only in the phase function. In other words, the integration intervals for μ' and φ' are $(0, +1)$ and $(0, 2\pi)$, respectively, irrespective of the sign of the subscript. Equations (7) and (8) should be solved subject to the boundary conditions

$$[S(\tau_0, \tau_1; \Omega, \Omega_0)]_{\tau_0=\tau_1} = 0,$$
$$[T(\tau_0, \tau_1; \Omega, \Omega_0)]_{\tau_0=\tau_1} = 0. \tag{9}$$

The principle of reciprocity provides us with

$$S(\tau_0, \tau_1; \Omega, \Omega_0) = S(\tau_0, \tau_1; \Omega_0, \Omega), \tag{10}$$

$$T(\tau_0, \tau_1; \Omega, \Omega_0) = T(\tau_1, \tau_0; \Omega_0, \Omega), \tag{11}$$

where $T(\tau_1, \tau_0; \Omega_0, \Omega)$ is the transmission function conjugate to $T(\tau_0, \tau_1; \Omega, \Omega_0)$ when the lower boundary τ_1 is illuminated monodirectionally. Equation (7)

[5] R. E. Bellman and R. E. Kalaba, Proc. Natl. Acad. Sci. **42**, 629 (1956).
[6] S. Ueno, Astrophys. J. **132**, 729 (1960).

agrees with that given by Goldstein.[7] In the isotropic case, Eq. (7) reduces to that given by Busbridge.[8] Furthermore, in the homogeneous medium the scattering and transmission functions reduce to those given by Chandrasekhar.[9]

2.3. Invariant Imbedding Equations for the Absorption Function

We define the absorption function L in the following manner. Let

$$\frac{L(\tau_0, \tau_1, \Omega_0)}{\mu_0} = \text{[the probability of ultimate absorption of a photon with direction } -\Omega_0 \text{ which is incident on the upper boundary } \tau_0 \text{ of the atmosphere of optical thickness } (\tau_1 - \tau_0), \text{ where } -\Omega_0 \text{ stands for } (-\mu_0, \varphi_0)\text{]}.$$

In other words, πL represents the rate of production of truly absorbed photons in a cylinder of unit base area extending from $\tau = \tau_0$ to $\tau = \tau_1$, the input having direction $(-\mu_0, \varphi_0)$ and the net incident flux being π. The quantity πL corresponds to the total flux of absorbed radiation.

In a manner similar to the principle of invariance approach for diffuse reflection and transmission by a homogeneous slab (cf. Chandrasekhar[9]), we formulate the principle of invariant imbedding with the aid of the absorption integral operator $\mathfrak{L}$. In this case we consider the balance of radiation flux in the layer between τ and τ_1.

The downward normal flux of radiation $I_-(\tau, \Omega)$, incident on the level τ, consists of the flux of upgoing radiation $I_+(\tau, \Omega)$ at any level τ, the flux of downgoing radiation $I_-(\tau_1, \Omega')$ through the bottom τ_1, and the flux of downgoing radiation absorbed by the atmosphere of optical thickness $(\tau_1 - \tau)$. Subscripts + and − refer to upgoing and downgoing radiation, respectively. Therefore, we have

$$\int_0^1 \int_0^{2\pi} \{I_-(\tau, \Omega') - I_+(\tau, \Omega') - I_-(\tau_1, \Omega')\}\mu' \, d\mu' \, d\varphi' = \mathfrak{L}_\tau\{I_-(\tau, \Omega')\}, \quad (12)$$

where the $\mathfrak{L}_\tau$ operator is defined by

$$\mathfrak{L}_\tau\{f_\pm(\Omega')\} = \int_0^1 \int_0^{2\pi} L(\tau_i, \tau_j, \Omega') f_\pm(\Omega') \, d\mu' \, d\varphi' \quad (13)$$

for range (τ_i, τ_j), where $\tau_0 \leq \tau_i < \tau_j \leq \tau_1$. In Eq. (12) we have $L(\tau, \tau_1, \Omega')$ for range (τ, Ω_1).

On differentiating Eq. (12) with respect to τ, passing to the limit $\tau = \tau_0$, and making use of boundary conditions (3) and (4), we get

$$\int_0^1 \int_0^{2\pi} \left\{\left[\frac{dI_-(\tau, \Omega')}{d\tau}\right]_{\tau=\tau_0} - \left[\frac{dI_+(\tau, \Omega')}{d\tau}\right]_{\tau=\tau_0}\right\}\mu' \, d\mu' \, d\varphi' = \int_0^1 \int_0^{2\pi} \left\{\frac{\partial L(\tau_0, \tau_1, \Omega')}{\partial \tau_0} I_-(\tau_0, \Omega') + L(\tau_0, \tau_1, \Omega')\left[\frac{dI_-(\tau, \Omega')}{d\tau}\right]_{\tau=\tau_0}\right\} d\mu' \, d\varphi'. \quad (14)$$

On recalling Eqs. (1)–(6), we obtain (after some minor rearrangement of the terms) one of the desired equations:

$$\begin{aligned} -\frac{\partial L}{\partial \tau_0}(\tau_0, \tau_1, \Omega_0) + \frac{L}{\mu_0}(\tau_0, \tau_1, \Omega_0) &= \frac{\lambda(\tau_0)}{4\pi} \int_- L(\tau_0, \tau_1, \Omega') \gamma(\tau_0, \Omega', -\Omega_0) \frac{d\Omega'}{\mu'} \\ &+ \frac{\lambda(\tau_0)}{16\pi^2} \int_- L(\tau_0, \tau_1, \Omega') \frac{d\Omega'}{\mu'} \\ &\times \int_+ \gamma(\tau_0, \Omega', \Omega'') S(\tau_0, \tau_1; \Omega'', \Omega_0) \frac{d\Omega''}{\mu''} \\ &+ 1 - \frac{\lambda(\tau_0)}{4\pi} \int \gamma(\tau_0, \Omega', -\Omega_0) \, d\Omega' \\ &+ \frac{1}{4\pi} \int_+ S(\tau_0, \tau_1; \Omega', \Omega_0) \frac{d\Omega'}{\mu'} - \frac{\lambda(\tau_0)}{16\pi^2} \\ &\times \int \int_+ \gamma(\tau_0, \Omega', \Omega'') S(\tau_0, \tau_1; \Omega'', \Omega_0) \, d\Omega' \frac{d\Omega''}{\mu''}, \end{aligned} \quad (15)$$

where the lack of a subscript on an integral indicates that the integration is over the whole range of μ' and φ'.

Similarly, consider another balance equation of radiation flux in the atmosphere between τ_0 and τ as below: The sum of the downward normal flux of radiation $I_-(\tau_0, \Omega)$ incident on the upper boundary τ_0 and the upward flux of the radiation $I_+(\tau, \Omega)$ at any level τ results in the flux of the diffusely reflected radiation $I_+(\tau_0, \Omega)$, the flux of downward radiation $I_-(\tau, \Omega)$ incident on the surface τ, and the flux of the radiation that comes from the absorption of the radiation $\{I_-(\tau_0, \Omega) + I_+(\tau, \Omega)\}$, incident on the upper boundary and level τ, by the atmosphere of optical thickness $(\tau - \tau_0)$. Then

$$\begin{aligned} &\int_0^1 \int_0^{2\pi} \{I_-(\tau_0, \Omega') + I_+(\tau, \Omega') - I_+(\tau_0, \Omega') - I_-(\tau, \Omega')\}\mu' \, d\mu' \, d\varphi' \\ &= \mathfrak{L}_\tau\{I_-(\tau_0, \Omega') + I_+(\tau, \Omega')\} \\ &= \int_0^1 \int_0^{2\pi} L(\tau_0, \tau, \Omega')\{I_-(\tau_0, \Omega') + I_+(\tau, \Omega')\} \, d\mu' \, d\varphi'. \end{aligned} \quad (16)$$

[7] J. W. Goldstein, Astrophys. J. 132, 473 (1960).
[8] I. W. Busbridge, Astrophys. J. 133, 198 (1961).
[9] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).

On differentiating Eq. (16) with respect to τ, passing to the limit τ_1, and making use of boundary conditions (3) and (4), we have

$$\int_0^1\int_0^{2\pi}\left\{\left[\frac{dI_+}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_1}-\left[\frac{dI_-}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_1}\right\}\mu'\,d\mu'\,d\varphi'$$
$$=\int_0^1\int_0^{2\pi}\left\{\frac{\partial L}{\partial\tau_1}(\tau_0,\tau_1,\Omega')I_-(\tau_0,\Omega')\right.$$
$$\left.+L(\tau_0,\tau_1,\Omega')\left[\frac{dI_+}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_1}\right\}d\mu'\,d\varphi'. \quad (17)$$

On using Eqs. (1)–(6), we find (after some rearrangement of the terms) another important equation:

$$\frac{\partial L}{\partial\tau_1}(\tau_0,\tau_1,\Omega_0)$$
$$=e^{-(\tau_1-\tau_0)/\mu_0}\left(1-\frac{\lambda(\tau_1)}{4\pi}\int\gamma(\tau_1,\Omega',-\Omega_0)\,d\Omega'\right)$$
$$+\frac{\lambda(\tau_1)}{4\pi}e^{-(\tau_1-\tau_0)/\mu_0}\int_-L(\tau_0,\tau_1,\Omega')\gamma(\tau_1,\Omega',-\Omega_0)$$
$$\times\frac{d\Omega'}{\mu'}+\frac{\lambda(\tau_1)}{16\pi^2}\int_-L(\tau_0,\tau_1,\Omega')\frac{d\Omega'}{\mu'}$$
$$\times\int_-\gamma(\tau_1,\Omega',\Omega'')T(\tau_0,\tau_1;\Omega'',\Omega_0)\frac{d\Omega''}{\mu''}$$
$$-\frac{\lambda(\tau_1)}{16\pi^2}\int\int_-\gamma(\tau_1,\Omega',\Omega'')T(\tau_0,\tau_1;\Omega'',\Omega_0)\,d\Omega'$$
$$\times\frac{d\Omega''}{\mu''}+\frac{1}{4\pi}\int_-T(\tau_0,\tau_1;\Omega',\Omega_0)\frac{d\Omega'}{\mu'}. \quad (18)$$

Equations (15) and (18) are the required invariant imbedding equations, together with boundary condition

$$[L(\tau_0,\tau_1,\Omega_0)]_{\tau_0=\tau_1}=0. \quad (19)$$

When no re-emission of interacting photons takes place (i.e., in the case of pure absorption with $\lambda=0$), Eq. (18) becomes

$$L(\tau_0,\tau_1,\Omega_0)=\mu_0(1-e^{-(\tau_1-\tau_0)/\mu_0}). \quad (20)$$

The above equation is readily derived for the rate of production of absorbing photons from the equation of transfer.

Furthermore, passing to the limit $\tau=\tau_0$ in Eq. (12) and using boundary conditions (3), (4), (5), and (6), we get

$$\frac{L(\tau_0,\tau_1,\Omega_0)}{\mu_0}+\frac{1}{4\pi\mu_0}\int_+S(\tau_0,\tau_1;\Omega',\Omega_0)\,d\Omega'$$
$$+\frac{1}{4\pi\mu_0}\int_-T(\tau_0,\tau_1;\Omega',\Omega_0)\,d\Omega'+e^{-(\tau_1-\tau_0)/\mu_0}=1. \quad (21)$$

This represents the principle of energy conservation for the standard diffuse reflection and transmission problem in an inhomogeneous medium. Probabilistically, given a photon incident in a given direction $(-\mu_0,\varphi_0)$ on the upper boundary, the sum of the probability of absorption in the finite atmosphere, the probability of diffuse reflection, the probability of diffuse transmission, and the probability of direct transmission results in unity.

In the homogeneous medium with isotropic scattering [i.e., $\lambda(\tau)=\lambda$, $\gamma(\tau,\Omega,\Omega_0)=1$, and $\tau_0=0$], Eqs. (15) and (18) reduce, respectively, to

$$\frac{\partial L}{\partial\tau_1}(\tau_1,\mu_0)+\frac{L}{\mu_0}(\tau_1,\mu_0)$$
$$=\left(1+\frac{1}{2}\int_0^1S(\tau_1;\mu',\mu_0)\frac{d\mu'}{\mu'}\right)$$
$$\times\left(1-\lambda+\frac{\lambda}{2}\int_0^1L(\tau_1,\mu')\frac{d\mu'}{\mu'}\right) \quad (22)$$

and

$$\frac{\partial L}{\partial\tau_1}(\tau_1,\mu_0)=\left\{e^{-\tau_1/\mu_0}+\frac{1}{2}\int_0^1T(\tau_1;\mu',\mu_0)\frac{d\mu'}{\mu'}\right\}$$
$$\times\left\{1-\lambda+\frac{\lambda}{2}\int_0^1L(\tau_1,\mu')\frac{d\mu'}{\mu'}\right\}. \quad (23)$$

In Eqs. (22) and (23), $S(\tau_1;\mu,\mu_0)$ and $T(\tau_1;\mu,\mu_0)$ are, respectively, the Chandrasekhar scattering and transmission functions.

Then, combining Eqs. (22) and (23), we obtain

$$L(\tau_1,\mu_0)=\frac{\lambda}{2}\mu_0(1-e^{-\tau_1/\mu_0})\int_0^1L(\tau_1,\mu')\frac{d\mu'}{\mu'}$$
$$+(1-\lambda)\mu_0\left[1-e^{-\tau_1/\mu_0}+\frac{1}{2}\int_0^1S(\tau_1;\mu',\mu_0)\frac{d\mu'}{\mu'}\right.$$
$$\left.-\frac{1}{2}\int_0^1T(\tau_1;\mu',\mu_0)\frac{d\mu'}{\mu'}\right]+\frac{\lambda}{4}\mu_0\int_0^1\int_0^1L(\tau_1,\mu')$$
$$\times\{S(\tau_1;\mu'',\mu_0)-T(\tau_1;\mu'',\mu_0)\}\frac{d\mu'}{\mu'}\frac{d\mu''}{\mu''}. \quad (24)$$

Furthermore, Eq. (20) becomes

$$1=e^{-\tau_1/\mu_0}+\frac{L(\tau_1,\mu_0)}{\mu_0}+\frac{1}{2\mu_0}\int_0^1S(\tau_1;\mu',\mu_0)\,d\mu'$$
$$+\frac{1}{2\mu_0}\int_0^1T(\tau_1;\mu',\mu_0)\,d\mu'. \quad (25)$$

Equations (23) and (25) coincide, respectively, with Eqs. (2) and (6) of Ref. 4.

3. MONODIRECTIONAL ILLUMINATION OF THE LOWER BOUNDARY

3.1. Equation of Transfer

Suppose that a parallel beam of radiation of constant flux π per unit area normal to the direction of the propagation is incident on the lower boundary $\tau = \tau_1$ of the atmosphere considered in the previous section at a fixed polar angle $\cos^{-1}\mu_0$ $(0 < \mu_0 \leq 1)$ with respect to the upward normal at the bottom and at azimuth φ_0 $(0 \leq \varphi_0 \leq 2\pi)$.

The equation of transfer appropriate to this case is given by Eq. (1), together with boundary conditions

$$I_-(\tau_0, \Omega) \equiv 0, \tag{26}$$

$$I_+(\tau_1, \Omega) = \pi\delta(\mu - \mu_0)\delta(\varphi - \varphi_0). \tag{27}$$

3.2. Scattering and Transmission Functions

In a manner similar to that given in Sec. 2.2, putting

$$I_+(\tau_0, \Omega) = (T/4\mu)(\tau_1, \tau_0; \Omega, \Omega_0) + \pi\delta(\mu - \mu_0)\delta(\varphi - \varphi_0)e^{-(\tau_1-\tau_0)/\mu_0}, \tag{28}$$

$$I_-(\tau_1, \Omega) = (S/4\mu)(\tau_1, \tau_0; \Omega, \Omega_0), \tag{29}$$

the invariant imbedding equations for the $S(\tau_1, \tau_0; \Omega, \Omega_0)$ and $T(\tau_1, \tau_0; \Omega, \Omega_0)$ functions take the forms

$$\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)S(\tau_1, \tau_0; \Omega, \Omega_0) + \frac{\partial S}{\partial \tau_1}(\tau_1, \tau_0; \Omega, \Omega_0) = \lambda(\tau_1)\Big\{\gamma(\tau_1, -\Omega, \Omega_0) + \frac{1}{4\pi}\int_- \gamma(\tau_1, \Omega, \Omega'')S(\tau_1, \tau_0; \Omega'', \Omega_0)\frac{d\Omega''}{\mu''} + \frac{1}{4\pi}\int_+ S(\tau_1, \tau_0; \Omega, \Omega')\gamma(\tau_1, \Omega', \Omega_0)\frac{d\Omega'}{\mu'} + \frac{1}{16\pi^2}\int_+\int_- S(\tau_1, \tau_0; \Omega, \Omega')\gamma(\tau_1, \Omega', \Omega'') \times S(\tau_1, \tau_0; \Omega'', \Omega_0)\frac{d\Omega'}{\mu'}\frac{d\Omega''}{\mu''}\Big\}, \tag{30}$$

$$\frac{1}{\mu_0}T(\tau_1, \tau_0; \Omega, \Omega_0) + \frac{\partial T}{\partial \tau_1}(\tau_1, \tau_0; \Omega, \Omega_0) = \lambda(\tau_1)\Big\{\gamma(\tau_1, \Omega, \Omega_0)e^{-(\tau_1-\tau_0)/\mu} + \frac{1}{4\pi}\int_+ T(\tau_1, \tau_0; \Omega, \Omega')\gamma(\tau_1, \Omega', \Omega_0)\frac{d\Omega'}{\mu'} + \frac{1}{4\pi}e^{-(\tau_1-\tau_0)/\mu}\int_- \gamma(\tau_1, \Omega, \Omega'')S(\tau_1, \tau_0; \Omega'', \Omega_0) \times \frac{d\Omega''}{\mu''} + \frac{1}{16\pi^2}\int_+\int_- T(\tau_1, \tau_0; \Omega, \Omega')\gamma(\tau_1, \Omega', \Omega'') \times S(\tau_1, \tau_0; \Omega'', \Omega_0)\frac{d\Omega'}{\mu'}\frac{d\Omega''}{\mu''}\Big\}, \tag{31}$$

together with boundary conditions

$$[S(\tau_1, \tau_0; \Omega, \Omega_0)]_{\tau_0=\tau_1} \equiv 0, \quad [T(\tau_1, \tau_0; \Omega, \Omega_0)]_{\tau_0=\tau_1} \equiv 0. \tag{32}$$

The principle of reciprocity is written in the forms

$$S(\tau_1, \tau_0; \Omega, \Omega_0) = S(\tau_1, \tau_0; \Omega_0, \Omega), \tag{33}$$

$$T(\tau_1, \tau_0; \Omega, \Omega_0) = T(\tau_0, \tau_1; \Omega_0, \Omega). \tag{34}$$

Apart from matters of notation, Eqs. (30) and (31) reduce to those given by Yanovitskii.[10] Furthermore, in the homogeneous case the scattering and the transmission functions reduce to those given by Chandrasekhar.[9]

3.3. Invariant Imbedding Equations for the Absorption Function

Let

$\frac{L(\tau_1, \tau_0, \Omega_0)}{\mu_0}$ = [the probability of ultimate absorption of photon with direction Ω_0 which is incident on the lower boundary τ_1 of the atmosphere of optical thickness $(\tau_1 - \tau_0)$, where Ω_0 stands for (μ_0, φ_0)].

In a manner similar to that used in a previous section, we shall formulate the principle of invariant imbedding with the aid of the absorption operator integral $\mathfrak{L}^*$, defined by

$$\mathfrak{L}_\tau^*\{f_\pm(\Omega')\} = \int_0^1\int_0^{2\pi} L(\tau_i, \tau_j, \Omega')f_\pm(\Omega')\,d\mu'\,d\varphi', \tag{35}$$

for the range (τ_j, τ_i), where $\tau_0 \leq \tau_j < \tau_i \leq \tau_1$.

Consider the balance equation of radiation flux in the atmospheric layer between τ and τ_1 as follows: The sum of the downward normal flux of radiation $I_-(\tau, \Omega)$ incident on the level τ, and the upward flux of the external radiation $I_+(\tau_1, \Omega)$ incident on the bottom, results in the flux of upward radiation $I_+(\tau, \Omega)$ at level τ, the flux of the diffusely reflected radiation $I_-(\tau_1, \Omega)$ at the bottom, and the flux of the radiation that comes from the absorption of radiation

$$\{I_-(\tau, \Omega) + I_+(\tau_1, \Omega)\},$$

incident on the level τ and on the bottom τ_1, by the atmosphere of optical thickness $(\tau_1 - \tau)$. Then

$$\int_0^1\int_0^{2\pi}\{I_-(\tau, \Omega') + I_+(\tau_1, \Omega') - I_+(\tau, \Omega') - I_-(\tau_1, \Omega')\}\mu'\,d\mu'\,d\varphi' = \mathfrak{L}_\tau^*\{I_-(\tau, \Omega') + I_+(\tau_1, \Omega')\}, \tag{36}$$

where $L(\tau_i, \tau_j, \Omega') \equiv L(\tau_1, \tau, \Omega')$. Equation (36) is similar in form to Eq. (16) for monodirectional illumination of the upper boundary.

[10] E. G. Yanovitskii, Russ. Astron. J. 38, 912 (1961).

On differentiating Eq. (36) with respect to τ, passing to the limit τ_0, and making use of boundary conditions (26) and (27), we have

$$\int_0^1\int_0^{2\pi}\left\{\left[\frac{dI_-}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_0}-\left[\frac{dI_+}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_0}\right\}\mu'\,d\mu'\,d\varphi'$$
$$=\int_0^1\int_0^{2\pi}\left\{\frac{\partial L}{\partial\tau_0}(\tau_1,\tau_0,\Omega')I_+(\tau_1,\Omega')\right.$$
$$\left.+L(\tau_1,\tau_0,\Omega')\left[\frac{dI_-}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_0}\right\}d\mu'\,d\varphi'. \quad (37)$$

On recalling Eqs. (1), (2), and (26)–(29), from Eq. (37) we get

$$-\frac{\partial L}{\partial\tau_0}(\tau_1,\tau_0,\Omega_0)$$
$$=e^{-(\tau_1-\tau_0)/\mu_0}\left[1-\frac{\lambda(\tau_0)}{4\pi}\int\gamma(\tau_0,\Omega',\Omega_0)\,d\Omega'\right]$$
$$+\frac{\lambda(\tau_0)}{4\pi}e^{-(\tau_1-\tau_0)/\mu_0}\int_+L(\tau_1,\tau_0,\Omega')\gamma(\tau_0,\Omega',\Omega_0)$$
$$\times\frac{d\Omega'}{\mu}+\frac{\lambda(\tau_0)}{16\pi^2}\int_+\int_+L(\tau_1,\tau_0,\Omega')\gamma(\tau_0,\Omega',\Omega'')$$
$$\times T(\tau_1,\tau_0;\Omega'',\Omega_0)\frac{d\Omega'}{\mu'}\frac{d\Omega''}{\mu''}$$
$$-\frac{\lambda(\tau_0)}{16\pi^2}\int\int_+\gamma(\tau_0,\Omega',\Omega'')T(\tau_1,\tau_0;\Omega'',\Omega_0)\,d\Omega'$$
$$\times\frac{d\Omega''}{\mu''}+\frac{1}{4\pi}\int_+T(\tau_1,\tau_0;\Omega',\Omega_0)\frac{d\Omega'}{\mu'}. \quad (38)$$

Similarly, consider another balance equation of radiation flux in the atmosphere between τ_0 and τ as follows: The upward normal flux of radiation $I_+(\tau,\Omega)$ at the level τ results in the flux of transmitted radiation at the upper boundary $I_+(\tau_0,\Omega)$, the flux of downward radiation $I_-(\tau,\Omega)$ at the level τ, and the flux of radiation that comes from the absorption of radiation $I_-(\tau,\Omega)$ by the atmosphere between τ_0 and τ. Therefore, we have

$$\int_0^1\int_0^{2\pi}\{I_+(\tau,\Omega')-I_+(\tau_0,\Omega')-I_-(\tau,\Omega')\}\mu'\,d\mu'\,d\varphi'$$
$$=\mathfrak{L}_\tau^*\{I_+(\tau,\Omega')\}, \quad (39)$$

where $L(\tau_i,\tau_j,\Omega')=L(\tau,\tau_0,\Omega')$.

On differentiating Eq. (39) with respect to τ, passing to the limit τ_1, and making use of boundary conditions (26) and (27), we obtain

$$\int_0^1\int_0^{2\pi}\left\{\left[\frac{dI_+}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_1}-\left[\frac{dI_-}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_1}\right\}\mu'\,d\mu'\,d\varphi'$$
$$=\int_0^1\int_0^{2\pi}\left\{\frac{\partial L}{\partial\tau_1}(\tau_1,\tau_0,\Omega')I_+(\tau_1,\Omega')\right.$$
$$\left.+L(\tau_1,\tau_0,\Omega')\left[\frac{dI_+}{d\tau}(\tau,\Omega')\right]_{\tau=\tau_1}\right\}d\mu'\,d\varphi'. \quad (40)$$

On making use of Eqs. (1), (2), and (26)–(29), from Eq. (40) we get

$$\frac{\partial L}{\partial\tau_1}(\tau_1,\tau_0,\Omega_0)+\frac{L(\tau_1,\tau_0,\Omega_0)}{\mu_0}$$
$$=\frac{\lambda(\tau_1)}{4\pi}\int_+L(\tau_1,\tau_0,\Omega')\gamma(\tau_1,\Omega',\Omega_0)\frac{d\Omega'}{\mu'}$$
$$+\frac{\lambda(\tau_1)}{16\pi^2}\int_+\int_-L(\tau_1,\tau_0,\Omega')\gamma(\tau_1,\Omega',\Omega'')$$
$$\times S(\tau_1,\tau_0;\Omega'',\Omega_0)\frac{d\Omega'}{\mu'}\frac{d\Omega''}{\mu''}$$
$$+1-\frac{\lambda(\tau_1)}{4\pi}\int\gamma(\tau_1,\Omega',\Omega_0)\,d\Omega'$$
$$+\frac{1}{4\pi}\int_-S(\tau_1,\tau_0;\Omega',\Omega_0)\frac{d\Omega'}{\mu'}-\frac{\lambda(\tau_1)}{16\pi^2}$$
$$\times\int\int_-\gamma(\tau_1,\Omega',\Omega'')S(\tau_1,\tau_0;\Omega'',\Omega_0)\frac{d\Omega'\,d\Omega''}{\mu''}. \quad (41)$$

In this case a complete set of the invariant imbedding equations consists of Eqs. (38) and (41), together with boundary condition

$$[L(\tau_1,\tau_0,\Omega_0)]_{\tau_0=\tau_1}=0. \quad (42)$$

In the case of pure absorption, from Eq. (38) we have Eq. (20).

Furthermore, passing to the limit $\tau=\tau_1$ in Eq. (39) and using boundary conditions (28) and (29), we obtain the principle of energy conservation:

$$\frac{L(\tau_1,\tau_0,\Omega_0)}{\mu_0}+\frac{1}{4\pi\mu_0}\int_-S(\tau_1,\tau_0;\Omega,\Omega_0)\,d\Omega$$
$$+\frac{1}{4\pi\mu_0}\int_+T(\tau_1,\tau_0;\Omega,\Omega_0)\,d\Omega+e^{-(\tau_1-\tau_0)/\mu_0}=1, \quad (43)$$

which is similar in form to Eq. (21).

When $\lambda(\tau)=\lambda$, $\gamma(\tau,\Omega,\Omega_0)=1$, and $\tau_0=0$, Eqs. (38) and (41) become, respectively, Eqs. (23) and (22).

ACKNOWLEDGMENTS

One of the authors (S. Ueno) is indebted to the National Academy of Sciences–National Research Council for the tenure of a Senior Postdoctoral Resident Research Associateship under the National Aeronautics and Space Administration during the summer of 1966; he also expresses his gratitude to Dr. Robert Jastrow of the Institute for Space Studies for his hospitality and kind interest in this work. In addition, it is a pleasure to thank Dr. A. Arking and Dr. J. Potter for their helpful discussions concerning this subject.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Potential-Correlation Function Duality in Statistical Mechanics

C. DE DOMINICIS
D.Ph.T.–Centre d'Etudes Nucléaires de Saclay, France
AND
F. ENGLERT*
Université Libre de Bruxelles, Belgium

(Received 12 April 1967)

The Feynman path-integral formulation is used to analyze the grand-partition function Z and the n-body Green's function G_n or rather their cumulants (or connected parts) $\tilde{G}_n$ for a system governed by p-body interaction potentials v_p. It is shown that the functional relationship expressing $\tilde{G}_n$ in terms of v_p is invariant under the transformation exchanging $-(-)^m\tilde{G}_m$ and v_m everywhere. Under the same transformation $\log Z$ undergoes a change of sign. The content of these results is discussed in conclusion.

INTRODUCTION

THERE have been many attempts in the past[1,2] to formulate statistical mechanics in such a way as to replace the potentials appearing in a perturbation theory by more physical or "observable" quantities such as densities or Green's functions. These efforts toward "renormalized" theories were motivated either by the search for more convergent expansions or by the necessity of new, "anomalous" solutions. We would like to show that there actually exists an extremely simple duality relationship between potentials and Green's functions for quantum systems. More precisely, we are going to show[3] that the functional forms (as given, e.g., by perturbation theory) for the grand-partition function Z and the successive n-body Green's functions G_n in terms of the original p-body potentials v_p governing the system remain essentially *invariant*, if, within phase factors, we exchange everywhere v_m and G_m (or rather the cumulant or "connected" part of G_m). In the end, we comment upon some related results of classical statistical mechanics, and discuss the meaning and content of these dual relationships.

The relationships obtained are entirely formal in the sense that no attempt is made to discuss the conditions of existence for the initial, intermediate, or final steps of the proof.

I. DEFINITIONS

To begin with, consider a system of identical particles governed by the Hamiltonian

$$\hat{H}(\Phi^\dagger,\Phi)=\int\hat{v}_1(1,1')\Phi^\dagger(1)\Phi(1')\,d1\,d1' + \frac{1}{(2!)^2}\int\hat{v}_2(12,1'2')\Phi^\dagger(1)\Phi^\dagger(2) \times\Phi(1')\Phi(2')\,d1\,d2\,d1'\,d2'+\cdots, \quad (1)$$

where variable j stands for position x_j (or momentum k_j) and internal degrees of freedom, where $\hat{v}_1$ is the kinetic part minus the chemical potential, $\hat{v}_2$ a two-body interaction potential, etc.

Those potentials may also be time-dependent, or rather "temperature"-dependent, for the purpose of application to equilibrium statistical mechanics. In that case, the variable j also includes[4] a "temperature" variable $\tilde{u}_j$ varying in the interval $(0,\beta=(kT)^{-1})$, where T is the temperature, or the Fourier conjugate of $\tilde{u}_j$, $\tilde{\omega}_j$ taking the values $2n_j i\pi/\beta$ for a bose system where n_j is any integer. The quantity defined by the right-hand side of Eq. (1) is, in that case, β times the Hamiltonian. It is more convenient to use dimensionless variables by introducing $0<u_j<1$ and $\omega_j=2n_j i\pi$, and by always working with the combinations $v=\beta\hat{v}$ instead of $\hat{v}$. $\Phi^\dagger$ and Φ are creation and annihilation operators properly normalized. In the following, we consider only bose systems to avoid signs and we stick to the momentum-"energy" variables to have discrete summations.

* Supported in part by the U.S. Air Force under grant No. AFE OAR 63-51 and monitored by the European Office of Aerospace Research.

[1] J. Yvon, Act. Sci. Ind. 203 (1935), to quote the earliest attempt known to the authors, where the one-body potential is eliminated in favor of the "observable" one-body density, yielding, in particular, the so-called virial expansion.

[2] See, for example, the review article of C. Bloch, *Studies in Statistical Mechanics* (North-Holland Publishing Company, Amsterdam, 1965), Vol. 3, p. 1.

[3] A perturbation expansion approach to this result viewed in the framework of field theory together with comments on possible applications to "bootstrap" theory of interaction is being published elsewhere; F. Englert and C. De Dominicis, Nuovo Cimento (to be published).

[4] Section 20 of Ref. 2 discusses the pseudoperiodicity conditions imposed on many-"time" potentials.

II. FEYNMAN PATH-INTEGRAL FORMULATION

The equilibrium statistical mechanics of such a system[5] is contained in the "action"

$$S(\Phi^\dagger, \Phi) = \sum_{j,j'} \delta_{j,j'}\omega_j \Phi^\dagger(j)\Phi(j') - H(\Phi^\dagger, \Phi). \quad (2)$$

For reasons of symmetry and more generality, we write the action in the form

$$S_v(\Phi^\dagger, \Phi) = -\sum \frac{1}{p!\,q!} v_{p,q}(1, \cdots p; 1', \cdots q') \times \Phi^\dagger(1)\cdots\Phi^\dagger(p) \times \Phi(1')\cdots\Phi(q'),$$

the summation being over the momentum-"energy" variables k_j, ω_j, and over the subscripts p, q characterizing now the various "potentials" $v_{p,q}$. Here we have incorporated the ω_j term of Eq. (2) in the one-body part of the action

$$v_{1,1}(j, j') = \beta\hat{v}_1(j, j') + \omega_j\delta_{jj'}.$$

Also, we now have introduced anomalous potentials $v_{p,q}$ $(q \neq p)$ such as, for example, source or sink potentials $v_{0,1}$, $v_{1,0}$.

As is well known,[5] the grand-partition function is given by the functional integral

$$Z(v) = \int \prod_j \frac{d\varphi(j)\, d\varphi^*(j)}{(2\pi)} \exp\{S_v(\varphi^*, \varphi)\} = \int D(\varphi^*, \varphi) \exp\{S_v(\varphi^*, \varphi)\}, \quad (3)$$

where the integration is over the *c-number variables* $\varphi(j)$ and $\varphi^*(j)$, that is, variables indexed by k_j, ω_j. The domain of integration is the whole space for the real and imaginary part of each $\varphi(j)$. Likewise, the average T product of any number of operators ("thermal" Green's function) is given in Fourier transform by the dimensionless functional integral form

$$G_{n,m}(1, 2\cdots n; 1', 2'\cdots m') = \frac{1}{Z(v)} \int D(\varphi^*, \varphi)\varphi(1)\cdots\varphi(n)\varphi^*(1')\cdots\varphi^*(m') \times \exp\{S_v(\varphi^*, \varphi)\}, \quad (4)$$

or in functional differential form, by

$$G_{n,m}(1, 2\cdots n; 1', 2'\cdots m') = \frac{1}{Z(v)} \prod_{j=1}^{n}\left[-\frac{\delta}{\delta v_{0,1}(j)}\right]\prod_{j'=1'}^{m'}\left[-\frac{\delta}{\delta v_{1,0}(j')}\right] \times Z\{v\}. \quad (5)$$

[5] See, for example, N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), and J. S. Bell in *The Many-Body Problem*, E. Caianiello, Ed. (Academic Press, Inc., New York, 1962), for its quantum statistical mechanics aspect.

That is, the quantity $\rho = e^{S_v}/Z$ can be thought of as a density matrix in a functional phase space, the observables (the average value of which are calculated with the weight ρ) being products of φ, φ^* corresponding to T products of operators Φ, $\Phi^\dagger$. To emphasize the functional dependence of $G_{n,m}$ upon $v_{p,q}$ [whether in its closed form as in the right-hand side of Eq. (4) or in its perturbation expansion form in terms of Feynman diagrams], we write

$$G_{n,m}(1, 2, \cdots n; 1', 2', \cdots m') = \mathcal{F}_{n,m}\{1, \cdots n; 1', \cdots m' \mid v_{p,q}\}.$$

III. DUAL RELATIONSHIPS

Let us consider now the generating functional $\Xi(\eta^*, \eta)$, defined by

$$\Xi(\eta^*, \eta) = \int D(\varphi^*, \varphi) \times \exp\left\{S_v(\varphi^*, \varphi) + i\sum_i (\eta(1)\varphi^*(1) + \eta^*(1)\varphi(1))\right\}, \quad (6)$$

and expand $\Xi(\eta^*, \eta)$ in powers of η, η^*. According to the definition (4), we have

$$\Xi(\eta^*, \eta) = Z(v)\left[\sum_{p=0,q=0}^{\infty} \frac{(i)^{p+q}}{p!\,q!} G_{p,q}(1\cdots p; 1'\cdots q') \times \eta^*(1)\cdots\eta^*(p)\eta(1')\cdots\eta(q')\right].$$

If we introduce the cumulants $\tilde{G}_{n,m}$, defined by

$$\tilde{G}_{n,m}(1, 2, \cdots n; 1', 2', \cdots m') = \prod_{j=1}^{n}\left[-\frac{\delta}{\delta v_{0,1}(j)}\right]\prod_{j'=1'}^{m'}\left[-\frac{\delta}{\delta v_{1,0}(j')}\right] \log Z\{v\}, \quad (7)$$

that is, the "connected" parts of $G_{n,m}$, we have

$$\Xi(\eta^*, \eta) = Z(v)\exp\left\{\sum_{\substack{p,q \\ p^2+q^2\neq 0}} \frac{(i)^{p+q}}{p!\,q!} \tilde{G}_{p,q}(1\cdots p; 1'\cdots q') \times \eta^*(1)\cdots\eta^*(p)\eta(1')\cdots\eta(q')\right\}. \quad (8)$$

Again, to exhibit the functional dependence of $\tilde{G}_{n,m}$ upon the "potentials" $v_{p,q}$, we write

$$\tilde{G}_{n,m}(1, 2, \cdots n; 1', 2', \cdots m') = \tilde{\mathcal{F}}_{n,m}\{1\cdots n; 1'\cdots m' \mid v_{p,q}\},$$

where $\tilde{\mathcal{F}}_{n,m}$ may be thought as the sum of all connected Feynman diagrams with m incoming and n outgoing external lines, constructed with vertices $-v_{p,q}$ and propagator $[v_{1,1}]^{-1}$.

TABLE I. Correspondence between original and dual systems.

	Original system	Dual system
Potentials	$v_{p,q}$	$-(i)^{p+q}\tilde{G}_{pq}$
Grand partition function	$Z\{v_{pq}\}$	$Z\{-(i)^{p+q}\tilde{G}_{p,q}\} = Z^{-1}\{v_{pq}\}$
Cumulants	$\tilde{G}_{n,m}(1\cdots n; 1'\cdots m')$ $= \tilde{\mathcal{F}}_{n,m}\{1\cdots n; 1'\cdots m' \mid v_{p,q}\}$	$-(i)^{n+m}v_{n,m}(1\cdots n; 1'\cdots m')$ $= \tilde{\mathcal{F}}_{n,m}\{1\cdots n, 1'\cdots m' \mid -(i)^{p+q}\tilde{G}_{p,q}\}$

A. Dual Relationships for Z

We now integrate both sides of Eq. (8) over all variables η^*, η. The left-hand side, as given by Eq. (6), yields a product of δ functions in φ^*, φ for each corresponding variable in η^*, η space. Hence, after integrations over φ^*, φ, we obtain an exponential with null argument. We thus have

$$1 = Z\{v_{p,q}\}\cdot Z\{-(i)^{p+q}\tilde{G}_{p,q}\}, \tag{9}$$

where $Z\{-(i)^{p+q}\tilde{G}_{p,q}\}$ is the grand-partition function where the potentials $v_{p,q}$ occurring in the "action" are replaced by the cumulants $\tilde{G}_{p,q}$ multiplied by $-(i)^{p+q}$.

B. Dual Relationships for $G_{n,m}$

We integrate both sides of Eq. (8) over η, η^* after multiplication by

$$\exp\left(-i\sum_1(\eta(1)\varphi_0^*(1) + \eta^*(1)\varphi_0(1))\right).$$

We obtain

$$\exp\{S_v(\varphi_0^*, \varphi_0)\} = Z(v)\int D(\eta^*, \eta)$$
$$\times \exp\left\{S_{\tilde{G}}(\eta^*, \eta) - i\sum_1(\eta(1)\varphi_0^*(1) + \eta^*(1)\varphi_0(1))\right\},$$

where $S_{\tilde{G}}$ is obtained from S_v by substituting $-(i)^{p+q}\tilde{G}_{p,q}$ for $v_{p,q}$.

Again, expansion of the right-hand side in powers of φ_0^*, φ_0 gives

$$\exp\{S_v(\varphi_0, \varphi_0^*)\} = Z\{v_{p,q}\}Z\{-(i)^{p+q}\tilde{G}_{p,q}\}$$
$$\times\left[\sum_{p,q}(-)^{p+q}\frac{1}{p!\,q!}\langle\eta(1)\cdots\eta(p)\eta^*(1')\cdots\eta^*(q')\rangle_{\tilde{G}}\right.$$
$$\left.\times\,\varphi(1')\cdots\varphi(q')\varphi^*(1)\cdots\varphi^*(p)\right],$$

where

$$\langle\eta(1)\cdots\eta(n)\eta^*(1')\cdots\eta^*(m')\rangle_{\tilde{G}}$$
$$= \mathcal{F}_{n,m}\{1\cdots n; 1'\cdots m' \mid -(i)^{p+q}\tilde{G}_{p,q}\}$$

is the Green's function in a *dual* system where the potentials $v_{p,q}$ are replaced by $-(i)^{p+q}\tilde{G}_{p,q}$. Using Eq. (9) and the definitions of the cumulants, we obtain

$$-(i)^{n+m}v_{n,m}(1\cdots n; 1'\cdots m')$$
$$= \tilde{\mathcal{F}}_{n,m}\{1\cdots n; 1'\cdots m' \mid -(i)^{p+q}\tilde{G}_{p,q}\}. \tag{10}$$

Thus, in this dual system, the cumulants [that is, in perturbation theory, the sum of all connected diagrams constructed with vertices $(i)^{p+q}\tilde{G}_{p,q}$, propagator $[\tilde{G}_{1,1}]^{-1}$ and having m incoming, n outgoing lines] are, within a phase factor, numerically equal to the original potentials $v_{n,m}$. We have the correspondence.[6] See Table I.

C. Alternate Form

Suppose, for simplicity, the system to be "normal," that is

$$G_{n,m} = 0, \quad n \neq m.$$

Call

$$G_{n,n} = G_n.$$

We also have

$$G_1 = \tilde{G}_1$$

diagonal in k, ω space, and we recall that the unperturbed propagator in the original system is

$$G_1^{(0)} \equiv [v_{1,1}]^{-1}.$$

Define the "correlation functions" C_n by

$$\tilde{G}_n(1,\cdots n; 1',\cdots n') = [G_1(1)\cdots G_1(n)G_1(1')\cdots G_1(n')]$$
$$\times\, C_n(1\cdots n; 1'\cdots n'). \tag{11}$$

By changing variables of integration in Eq. (8) from $\eta(j)$ to $\bar{\eta}(j)$ with

$$\eta(j) = \bar{\eta}(j)[G_1(j)]^{-1},$$

we obtain the duality relationship in the form

$$\log Z\{[G_1^{(0)}]^{-1}; v_n\} = +\sum_j \log G_1(j)$$
$$-\log Z\{[G_1]^{-1}; -(-)^nC_n\}. \tag{12}$$

In the original system, we had

$$\begin{aligned}\text{propagator:}\quad & G_1^{(0)} \equiv [v_{1,1}]^{-1},\\ \text{vertices:}\quad & -v_n \equiv -v_{n,n},\quad n > 1.\end{aligned} \tag{13}$$

In the dual system, we have

$$\begin{aligned}\text{propagator:}\quad & G_1 \equiv [C_1]^{-1},\\ \text{vertices:}\quad & (-)^nC_n,\quad n > 1,\end{aligned} \tag{14}$$

[6] Results of Table I remain valid if the factors $(i)^{p+q}$ or $(i)^{n+m}$ are suppressed everywhere (this immediately follows from the structure of the diagram expansions, for example). In that form, however, the propagator has an opposite sign.

Results of Ref. 3 for field theory are recovered by the substitutions

$$\beta \to i \text{ and } \sum_{n_j=-\infty}^{+\infty} \to \int_{-\infty}^{+\infty}\frac{d\omega}{2\pi},$$

where ω is the usual energy variable.

and we have the correspondence ($n \geq 1$):

C_n = sum of all distinct Feynman diagrams with their $2n$ external lines removed, built with propagator $G_1^{(0)}$, vertices $-v_p, p > 1$; (15)

$-(-)^n v_n$ = sum of all distinct Feynman diagrams with their $2n$ external lines removed, built with propagator G_1, and vertices $(-)^p C_p, p > 1$. (16)

D. Dual Equations of Motion

The hierarchy of equations of motion for the original system can be viewed as expressing that, by partial integration, one has

$$\int D(\varphi^*\varphi) \frac{\delta}{\delta\varphi(j)} \exp(S_v) = 0, \tag{17}$$

$$\int D(\varphi^*\varphi)\varphi(1) \frac{\delta}{\delta\varphi(1')} \exp(S_v) = -\delta_{1,1'} \int D(\varphi^*\varphi) \exp(S_v); \tag{18}$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

Likewise, dual equations of motion follow by partial integration of the integrand $\exp(S_{\bar{G}})$ upon $\eta^*(j)$ or $\eta(j)$. However, they are equations which couple each "correlation function" C_n to *all* higher-order C_{n+r} with $r = 0, 1, 2, \cdots$.

IV. EXTENSIONS AND DISCUSSION

The procedure used here for a quantum system could also be applied to the functional integral form of spin, Ising or classical systems.[7] The results would be slightly more complicated, because the unperturbed part of the effective "action" is in those cases nonlinear in the field variables (φ^*, φ).

It is interesting to notice that in the case of classical systems, whether or not at equilibrium, one has the following "dual" relationships[8]:

$$\mu_n(1, 2, \cdots n) = \sum_{p \geq n} \frac{p!}{(p-n)!} \int D_p(1, 2, \cdots p) \times d(n+1)\, d(n+2) \cdots dp, \tag{19}$$

$$D_n(1, 2, \cdots n) = \frac{1}{n!} \sum_{p=0}^{\infty} \frac{(-)^p}{p!} \int \mu_{n+p}(1, 2, \cdots, n+p) \times d(n+1)\, d(n+2) \cdots d(n+p). \tag{20}$$

Here each variable j stands for position andmo mentum, $\mu(1, 2, \cdots n)$ is the average n-point distribution function, and $D(1, 2, \cdots n)$ is the density in phase space of a system of n particles. Equation (19) is a definition of μ_n in terms of D_p; Eq. (20) is the result of inverting Eq. (19) to obtain D_n in terms of μ_p. J. Yvon[9] has also written relationships nearly identical to (19) and (20) for quantum statistical mechanics. Clearly, Eqs. (19) and (20) contain only the combinatorics of statistical mechanics. In contradistinction, the dual relationships established in Sec. III are also containing quantum dynamics through the starting form of the "action" and, in particular, the linearity in the fields φ^*, φ of its unperturbed part.

From the point of view of "renormalization," the dual formulation of quantum statistical mechanics is not very satisfactory for at least two reasons. First, the conjugate variables v_n and C_n are not on an equal footing. For example, a system having $v_2 \neq 0$, $v_{2+r} = 0$, $r > 0$, has, in general, *all* its C_n's nonvanishing. If the potentials v_n are indeed the "natural" variables of $\log Z$, the correlation functions C_n are usually understood as "natural" variables for the *entropy*, that is, for the quantity

$$F^{(\infty)} = \log Z\{v\} + \sum_j [[G_1^{(0)}(j)]^{-1} + \omega_j] G_1(j) + \mathrm{Tr} \sum_{p=2}^{\infty} \frac{1}{(p!)^2} v_p G_p. \tag{21}$$

Here the trace is over v_p and G_p considered as matrices with one index $1, 2, \cdots, p$ and the other $1', 2', \cdots p'$.

Using Eqs. (12) and (16), the entropy can then be written as a functional of the C_p's alone:

$$F^{(\infty)}\{C_1 \cdots C_n \cdots\} = \sum_j [\log G_1(j) + \omega_j G_1(j)] + \mathrm{Tr}\left[\sum_{p=1}^{\infty} \frac{1}{(p!)^2} v_p\{C_1 \cdots C_n \cdots\} G_p\right] - \log Z\{[G_1]^{-1}; -(-)^n C_n\}, \tag{22}$$

where $v_p\{C_1 \cdots C_p \cdots\}$ is given by Eq. (16).

There is actually a considerable amount of cancellation between the last two terms of Eq. (22). The functional $\log Z$ contains all connected diagrams, whereas $F^{(\infty)}$ will turn out to contain only a much restricted class of diagrams (irreducible diagrams and "ladders"). The analysis of these cancellations together with the introduction of new independent "observables" conjugate to v_n is the object of a separate paper.[10]

ACKNOWLEDGMENT

We wish to thank Dr. R. Balian for interesting discussions. One of us (F. E.) takes this opportunity to thank the Service de Physique Théorique du Centre d'Etudes Nucléaires de Saclay for the hospitality extended to him.

[7] See, for example, S. F. Edwards, Phil. Mag. 4, 1171 (1959) for classical systems; B. Mühlschlegel and H. Zittartz, Z. Physik 175, 553 (1963) for Ising systems.

[8] J. Yvon, *Les corrélations et l'entropie en mécanique statistique classique* (Cie Dunod, Paris, 1966). Equations (19) and (20) of the text are his equations (4) and (5), Chap. 2, Sec. 2.

[9] J. Yvon, Cours de Mécanique Statistique à la Faculté des Sciences de Paris (1966).

[10] C. De Dominicis and F. Englert (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

Group Formulation of Potential Scattering*

R. FINKELSTEIN AND D. LEVY
Department of Physics, University of California, Los Angeles, California

(Received 29 March 1967)

Potential scattering is formulated in terms of an integral equation on the group space of O_3. The potential need not be spherically symmetric except at large distances from the scattering center.

1. INTRODUCTION

IN particle physics, the physical assumptions are usually expressed either in group or in dispersion language. By establishing a correspondence between group space and momentum space, it is perhaps possible to relate these two approaches in an interesting way. We have attempted to explore this idea by formulating the general scattering problem in terms of an integral equation on the space of the symmetry group.

Here we present some results for potential scattering. The method is not restricted to spherically symmetric potentials, although we work with O_3. However, we do assume that space is isotropic far from the scatterer.

2. INTEGRAL EQUATIONS IN MOMENTUM SPACE

We consider the integral equations for the Green's function and the scattering amplitude:

$$(E - \mathbf{p}^2)g(\mathbf{p}, \mathbf{p}') - \int \tilde{V}(\mathbf{p} - \mathbf{p}'')g(\mathbf{p}'', \mathbf{p}')\, d\mathbf{p}'' = \delta(\mathbf{p} - \mathbf{p}'), \quad (2.1)$$

$$f(\mathbf{p}, \mathbf{p}') = \mathring{f}(\mathbf{p}, \mathbf{p}') - \int \frac{\tilde{V}(\mathbf{p} - \mathbf{p}'')f(\mathbf{p}'', \mathbf{p}')\, d\mathbf{p}''}{\mathbf{p}''^2 - E - i\epsilon}, \quad (2.2)$$

where

$$V(\mathbf{p} - \mathbf{p}') = \left(\frac{1}{2\pi}\right)^3 \int e^{i(\mathbf{p}-\mathbf{p}')\mathbf{x}} V(\mathbf{x})\, d\mathbf{x} \quad (2.3a)$$

$$= (-2\pi^2)^{-1}\mathring{f}(\mathbf{p}, \mathbf{p}'), \quad (2.3b)$$

and we have taken $2m = \hbar = 1$.

Let us put $z = E + i\epsilon$ and regard these equations as functions of the complex parameter z:

$$(p^2 - z)g(p, p', z) + \int \tilde{V}(\mathbf{p} - \mathbf{p}'')g(\mathbf{p}'', \mathbf{p}', z)\, d\mathbf{p}'' = -\delta(\mathbf{p} - \mathbf{p}''), \quad (2.1')$$

$$f(\mathbf{p}, \mathbf{p}', z) + \int \tilde{V}(\mathbf{p} - \mathbf{p}'') \frac{1}{(\mathbf{p}'')^2 - z} f(\mathbf{p}'', \mathbf{p}', z)\, d\mathbf{p}'' = \mathring{f}(\mathbf{p}, \mathbf{p}'). \quad (2.2')$$

We may first solve these equations for z real and negative; the scattering solution may then be obtained by continuation to the positive real axis, since $g(\mathbf{p}, \mathbf{p}', z)$ and $f(\mathbf{p}, \mathbf{p}', z)$ are regular everywhere in the complex z-plane with the exception of the real axis.

The reason for proceeding in this way is to connect with a compact, rather than a noncompact, group.

3. GROUP SPACE

We transform these equations into integral equations on the group space of O_3 by introducing the following correspondence between a point (**p**) in momentum space and a point (**w**) in group space:

$$e^{(i/2)\boldsymbol{\sigma}\mathbf{w}} = \frac{p_0 + i\mathbf{p}\boldsymbol{\sigma}}{p_0 - i\mathbf{p}\boldsymbol{\sigma}}. \quad (3.1)$$

The components of **p** provide a stereographic coordinate system for group space. In this coordinate system the group metric[1] is

$$g_{\alpha\beta} = G^2\delta_{\alpha\beta}, \quad (3.2)$$

where

$$G = \frac{P_0^2}{p^2 + p_0^2}. \quad (3.2a)$$

By Eq. (3.2) the group metric is determined by the free propagator.

Alternatively one assigns a non-Euclidean metric to momentum space in a way which depends on the local propagator. With this metric momentum, space *is* the group space, and its volume is, of course, finite.

The irreducible representations of the rotation group $D^j_{mn}(p)$ satisfy the orthogonality relations

$$\int D^j_{mn}(p)\bar{D}^{j'}_{m'n'}(p)g^{\frac{1}{2}}\, d\mathbf{p} = \delta^{jj'}\delta_{mm'}\delta_{nn'}\Omega/d_j, \quad (3.3)$$

where

$$d_j = 2j + 1, \quad (3.3a)$$

$$g^{\frac{1}{2}} = G^3, \quad (3.3b)$$

and Ω is the volume of the group space

$$\Omega = \tfrac{1}{4}\pi^2 p_0^3. \quad (3.3c)$$

Let us introduce the orthonormal set

$$D_\mu(\mathbf{p}) = (d_j/\Omega)^{\frac{1}{2}} D^j_{mn}(\mathbf{p}), \quad \mu = (jmn). \quad (3.4)$$

Then,

$$\int D_\mu(\mathbf{p})\bar{D}_\nu(\mathbf{p})\, d\tau = \delta_{\mu\nu} \quad d\tau = g^{\frac{1}{2}}\, d\mathbf{p}. \quad (3.5)$$

* Work supported in part by the National Science Foundation.

[1] R. Finkelstein, J. Math. Phys. 8, 443 (1967).

The δ function on the group space is

$$D(\mathbf{p}-\mathbf{p}')=\sum_{\mu} D_{\mu}(\mathbf{p})\bar{D}_{\mu}(\mathbf{p}'). \tag{3.6}$$

This function is related to the usual δ function on momentum space in the following way:

$$\delta(\mathbf{p}-\mathbf{p}')=g^{\frac{1}{2}}D(\mathbf{p}-\mathbf{p}') \tag{3.7}$$

$$=g^{\frac{1}{2}}\sum_{\mu} D_{\mu}(\mathbf{p})\bar{D}_{\mu}(\mathbf{p}'). \tag{3.7a}$$

4. INTEGRAL EQUATIONS ON GROUP SPACE

Let us now put $z=-p_0^2$, where p_0 is real, into Eqs. (2.1′) and (2.2′), since we are going to obtain solution for z real and negative.

At the same time, we introduce the new (amputated) functions

$$G(\mathbf{p},\mathbf{p}')=G^{-2}(p/p_0)g(\mathbf{p},\mathbf{p}')G^{-2}(p'/p_0), \tag{4.1}$$

$$F(\mathbf{p},\mathbf{p}')=G^{-1}(p/p_0)f(\mathbf{p},\mathbf{p}')G^{-1}(p'/p_0). \tag{4.2}$$

Then,

$$G(\mathbf{p},\mathbf{p}')-(2\pi^2p_0^2)^{-1}\int \mathring{F}(\mathbf{p},\mathbf{p}'')G(\mathbf{p}'',\mathbf{p}')\,d\tau''$$

$$=-D(\mathbf{p}-\mathbf{p}')/p_0^2, \tag{4.3}$$

$$F(\mathbf{p},\mathbf{p}')-(2\pi^2p_0^2)^{-1}\int \mathring{F}(\mathbf{p},\mathbf{p}'')F(\mathbf{p}'',\mathbf{p}')\,d\tau''$$

$$=\mathring{F}(\mathbf{p},\mathbf{p}'). \tag{4.4}$$

The amputated Green's function and scattering amplitude satisfy integral equations which differ only in the inhomogeneous term.

It is now natural to expand the two point functions on the group space in terms of the irreducible representations:

$$\mathring{F}(\mathbf{p},\mathbf{p}')=\sum D_{\mu}(\mathbf{p})\mathring{c}_{\mu\nu}\bar{D}_{\nu}(\mathbf{p}'), \tag{4.5}$$

$$F(\mathbf{p},\mathbf{p}')=\sum D_{\mu}(\mathbf{p})c_{\mu\nu}\bar{D}_{\nu}(\mathbf{p}'), \tag{4.6}$$

$$G(\mathbf{p},\mathbf{p}')=\sum D_{\mu}(\mathbf{p})\gamma_{\mu\nu}\bar{D}_{\nu}(\mathbf{p}'), \tag{4.7}$$

and, of course,

$$D(\mathbf{p},\mathbf{p}')=\sum_{\mu} D_{\mu}(\mathbf{p})\bar{D}_{\mu}(\mathbf{p}'). \tag{4.8}$$

Substituting in the integral equations, one obtains the algebraic equations

$$\gamma_{\mu\nu}-(2\pi^2p_0^2)^{-1}\sum_{\lambda}\mathring{c}_{\mu\lambda}\gamma_{\lambda\nu}=-\delta_{\mu\nu}/p_0^2, \tag{4.9}$$

$$c_{\mu\nu}-(2\pi^2p_0^2)^{-1}\sum_{\lambda}\mathring{c}_{\mu\lambda}c_{\lambda\nu}=\mathring{c}_{\mu\nu}, \tag{4.10}$$

with the solutions

$$-\gamma=\frac{1}{p_0^2}\frac{1}{1-(2\pi^2p_0^2)^{-1}\mathring{c}}, \tag{4.11}$$

$$c=\frac{\mathring{c}}{1-(2\pi^2p_0^2)^{-1}\mathring{c}}$$

$$=-p_0^2\mathring{c}\gamma, \tag{4.12}$$

and

$$p_0^2\gamma+(2\pi^2p_0^2)^{-1}c=-1. \tag{4.13}$$

Therefore, the Green's function and scattering amplitude are given explicitly by continuation of the following expressions:

$$-G(\mathbf{p},\mathbf{p}')=\frac{1}{p_0^2}\sum_{\mu,\nu} D_{\mu}(\mathbf{p})[[1-(2\pi^2p_0^2)^{-1}\mathring{c}]^{-1}]_{\mu\nu}\bar{D}_{\nu}(\mathbf{p}'), \tag{4.14}$$

$$F(\mathbf{p},\mathbf{p}')=\sum_{\mu\nu} D_{\mu}(\mathbf{p})[\mathring{c}[1-(2\pi^2p_0^2)^{-1}\mathring{c}]^{-1}]_{\mu\nu}\bar{D}_{\nu}(\mathbf{p}'). \tag{4.15}$$

If the matrix $\mathring{c}$ happens to be diagonal, these results are particularly simple, but in any case, one has a formal solution of the scattering problem. It follows from (4.13) that

$$p_0^2G(\mathbf{p},\mathbf{p}')+(2\pi^2p_0^2)^{-1}F(\mathbf{p},\mathbf{p}')=-\sum D_{\mu}(\mathbf{p})\bar{D}_{\mu}(\mathbf{p}')$$

$$=-\delta(\mathbf{p}-\mathbf{p}')G^{-3}(\mathbf{p}), \tag{4.16}$$

or

$$(p_0^2+p^2)g(\mathbf{p},\mathbf{p}')(p_0^2+p'^2)+(2\pi^2)^{-1}f(\mathbf{p},\mathbf{p}')$$

$$=-\delta(p-p')(p^2+p_0^2). \tag{4.17}$$

In particular, if $\mathbf{p}\neq\mathbf{p}'$,

$$-2\pi^2g(\mathbf{p},\mathbf{p}')=(p_0^2+p^2)^{-1}f(\mathbf{p},\mathbf{p}')(p_0^2+p'^2)^{-1}. \tag{4.17a}$$

5. BOUND STATES

The condition for bound states is

$$\Phi(\mathbf{p})=-\frac{1}{2\pi^2E}\int \mathring{F}(\mathbf{p},\mathbf{p}')\Phi(\mathbf{p}')\,d\tau', \tag{5.1}$$

where

$$\Phi(\mathbf{p})=G^{-2}(p/p_0)\Phi(\mathbf{p}),$$

and $\phi(\mathbf{p})$ is the probability amplitude in momentum space. Putting

$$\Phi(\mathbf{p})=\sum V_{\mu}D_{\mu}(\mathbf{p}), \tag{5.2}$$

we obtain

$$\sum_{\lambda}(\mathring{c}_{\mu\lambda}+2\pi^2E\delta_{\mu\lambda})V_{\lambda}=0. \tag{5.3}$$

The condition for an eigensolution is

$$|\mathring{c}_{\mu\nu}+2\pi^2E\delta_{\mu\nu}|=0, \tag{5.4}$$

which, of course, is the condition that c becomes singular or that the inversion of (4.9) and (4.10) is not possible.

If $\mathring{c}$ happens to be diagonal, then

$$\mathring{c}_{\mu\lambda}=\mathring{c}(N,p_0)\delta_{\mu\lambda} \tag{5.5a}$$

or

$$\mathring{c}(N,p_0)=-2\pi^2E(N), \tag{5.5b}$$

where N labels the bound state. That is, when p_0 corresponds to a bound state, the following condition is satisfied:

$$\mathring{c}(N,p_0)-2\pi^2p_0^2=0. \tag{5.5c}$$

6. DISCUSSION OF THE MATRIX $\mathring{c}$

In order to investigate this method, it is necessary to learn something about $\mathring{c}$, which determines the

representation of the Born approximation on the group space. We have, by inverting (4.5),

$$\mathring{c}_{\mu\nu} = \iint \mathring{F}(\mathbf{p}, \mathbf{p}')\bar{D}_\mu(\mathbf{p})D_\nu(\mathbf{p}')\, d\tau\, d\tau', \tag{6.1}$$

where

$$\begin{aligned}\mathring{F}(\mathbf{p}, \mathbf{p}') &= -2\pi^2 G^{-1}(p/p_0)G^{-1}(p'/p_0)\tilde{V}(\mathbf{p} - \mathbf{p}') \\ &= -2\pi^2 G^{-1}(p/p_0)G^{-1}(p'/p_0)\int e^{i(\mathbf{p}-\mathbf{p}')\mathbf{x}}V(\mathbf{x})\, d\mathbf{x}.\end{aligned} \tag{6.2}$$

Therefore,

$$\begin{aligned}\mathring{c}_{\mu\nu} &= -2\pi^2 \int d\mathbf{x}\, V(\mathbf{x}) \int d\tau G^{-1}(p)e^{ip\mathbf{x}}\bar{D}_\mu(\mathbf{p}) \\ &\quad \times \int d\tau' G^{-1}(p')e^{-ip'\mathbf{x}}D_\nu(\mathbf{p}') \\ &= -2\pi^2 \int d\mathbf{x}\, V(\mathbf{x})\bar{\Psi}_\mu(\mathbf{x})\Psi_\nu(\mathbf{x}),\end{aligned} \tag{6.3}$$

where

$$\Psi_\mu(x) = \int e^{-ip\mathbf{x}}G^{-1}(p)D_\mu(\mathbf{p})\, d\tau \tag{6.4a}$$

$$= \int e^{-ip\mathbf{x}}G^2(p)D_\mu(\mathbf{p})\, d\mathbf{p}. \tag{6.4b}$$

The $G^2(\mathbf{p}/p_0)D_\mu(\mathbf{p}/p_0)$ are known to be eigensolutions of the Coulomb problem in the momentum representation,[1] provided that $p_0^2 = -E$, where E belongs to the Balmer spectrum. They are, however, not eigenfunctions of the angular momentum. The $\Psi_\mu(x)$ are therefore eigenfunctions of the Coulomb problem in configuration space under the same restrictions. They are, of course, not eigenfunctions of the angular momentum either.

According to (6.3), the matrix $\mathring{c}$ is obtained by calculating matrix elements of the arbitrary potential $V(x)$ with respect to the basis $\Psi_\mu(x)$—which we shall call the Coulomb basis, even when p_0 is not determined by the Balmer formula.

Another way to characterize the functions Ψ^j_{mn} is as follows:

$$\frac{e^{-ip\mathbf{x}}}{G(p/p_0)} = \sum_\mu \Psi_\mu(\mathbf{x})\bar{D}_\mu(\mathbf{p}). \tag{6.5}$$

It is important to notice that these functions are not orthonormal, and therefore,

$$(\mathring{c}^2)_{\mu\lambda} \neq (2\pi^2)^2 \int \bar{\Psi}_\mu(\mathbf{x})[V(\mathbf{x})]^2\Psi_\lambda(\mathbf{x})\, d\mathbf{x}.$$

In fact,

$$\begin{aligned}(\mathring{c}^2)_{\mu\lambda} &= \sum_\sigma \mathring{c}_{\mu\sigma}\mathring{c}_{\sigma\lambda} \\ &= (2\pi^2)^2 \sum \left(\int \bar{\Psi}_\mu V\Psi_\sigma\, d\mathbf{x}\right)\left(\int \bar{\Psi}_\sigma V\Psi_\lambda\, d\mathbf{x}\right) \\ &= (2\pi^2)^2 \iint d\mathbf{x}\, d\mathbf{x}' V(\mathbf{x})V(\mathbf{x}')\bar{\Psi}_\mu(\mathbf{x})\Psi_\lambda(\mathbf{x}') \\ &\quad \times \left(\sum_\sigma \Psi_\sigma(\mathbf{x})\bar{\Psi}_\sigma(\mathbf{x}')\right),\end{aligned} \tag{6.6}$$

where

$$\sum_\sigma \Psi_\sigma(\mathbf{x})\bar{\Psi}_\sigma(\mathbf{x}') = \sum_\sigma \iint d\mathbf{p}\, d\mathbf{p}' e^{-ip\mathbf{x}}e^{ip\mathbf{x}'}G^2(p)G^2(p')D_\sigma(\mathbf{p})\bar{D}_\sigma(\mathbf{p}').$$

But

$$\sum_\sigma D_\sigma(\mathbf{p})\bar{D}_\sigma(\mathbf{p}') = \delta(\mathbf{p} - \mathbf{p}')G^{-3},$$

and therefore,

$$\begin{aligned}\sum_\sigma \Psi_\sigma(\mathbf{x})\bar{\Psi}_\sigma(\mathbf{x}') &= \int d\mathbf{p}e^{ip(\mathbf{x}'-\mathbf{x})}G(p) \\ &= 2\pi^2 p_0^2 \mathring{G}(\mathbf{x}' - \mathbf{x}),\end{aligned} \tag{6.7}$$

where

$$\begin{aligned}\mathring{G}(\mathbf{x}) &= \frac{1}{2\pi^2}\int \frac{1}{p_0^2 + p^2}e^{ip\mathbf{x}}\, d\mathbf{p} \\ &= \frac{1}{2\pi^2}\int \frac{1}{p^2 - E}e^{ip\mathbf{x}}\, d\mathbf{p},\end{aligned} \tag{6.8}$$

which is, after continuation, the usual Green's function of the free particle. It follows from (6.6) that

$$(\mathring{c}^2)_{\mu\lambda} = (2\pi^2)^3 p_0^2 \iint \bar{\Psi}_\mu(\mathbf{x})V(\mathbf{x})\mathring{G}(\mathbf{x} - \mathbf{x}') \times V(\mathbf{x}')\Psi_\lambda(\mathbf{x}')\, d\mathbf{x}\, d\mathbf{x}'$$

and the complete perturbation expansion of the amputated amplitude follows from (4.12),

$$c = \mathring{c}[1 + (2\pi^2 p_0^2)^{-1}\mathring{c} + \cdots],$$

and therefore is

$$\begin{aligned}F(\mathbf{p}, \mathbf{p}') &= \sum D_\mu(\mathbf{p}) \left\{\int \bar{\Psi}_\mu(\mathbf{x})\left[(-2\pi^2)V(\mathbf{x}) + (-2\pi^2)^2\right.\right. \\ &\quad \left.\left. \times \int V(\mathbf{x})\mathring{G}(\mathbf{x} - \mathbf{x}')V(\mathbf{x}')\, d\mathbf{x}' + \cdots\right]\Psi_\lambda(\mathbf{x})\, d\mathbf{x}\right\}\bar{D}_\lambda(\mathbf{p}') \\ &= G^{-1}(p)G^{-1}(p')\left\{(-2\pi^2)\int e^{i(\mathbf{p}-\mathbf{p}')\mathbf{x}}V(\mathbf{x})\, d\mathbf{x}\right. \\ &\quad + (-2\pi^2)^2 \iint e^{i(\mathbf{p}-\mathbf{p}')\mathbf{x}} \\ &\quad \left. \times V(\mathbf{x})\mathring{G}(\mathbf{x} - \mathbf{x}')V(\mathbf{x}')\, d\mathbf{x}\, d\mathbf{x}' + \cdots\right\}\end{aligned}$$

by Eq. (6.5). The scattering amplitude itself is

$$f(p, p') = \langle p|\mathring{f} + \mathring{f}\mathring{G}\mathring{f} + \cdots|p'\rangle, \tag{6.9}$$

which is a more familiar form of the solution to the Lippmann–Schwinger equation.

7. CASIMIR POTENTIALS

The simplest potentials in the present approach lead to matrices $\mathring{c}$ (and therefore c) which are diagonal and will be called diagonal potentials. In addition, let the diagonal entries depend only on j and not on m and n; that is, the potential is a Casimir operator with respect to the group and will be called a Casimir potential. Then,

$$c_{\mu\nu} = c(N)\,\delta_{\mu\nu}, \tag{7.1}$$

$$\gamma_{\mu\nu} = \gamma(N)\,\delta_{\mu\nu}. \tag{7.2}$$

The Coulomb potential is the simplest such example, but all Casimir potentials may be treated in the same way as we now show. We now have

$$F(\mathbf{p}, \mathbf{p}') = \frac{1}{\Omega} \sum_{jmn} Nc(N) D^j_{mn}(\mathbf{p}) \bar{D}^j_{mn}(\mathbf{p}'),$$

where $N = d_j = 2j + 1$ by (3.4). Then,

$$F(\mathbf{p}, \mathbf{p}') = \frac{1}{\Omega} \sum_{N=1}^{\infty} Nc(N) \chi^{(N)}(pp'^{-1}) = \frac{1}{\Omega} \sum_{N=1}^{\infty} Nc(N) \chi^N(\phi), \tag{7.3}$$

where

$$\chi^{(N)}(\phi) = \frac{\sin N\phi}{\sin \phi} \tag{7.4}$$

is the character, and

$$\sin^2 \tfrac{1}{2}\phi = G(p)(\mathbf{p} - \mathbf{p}')^2 G(p')/p_0^2. \tag{7.5}$$

The connection between momentum space and group space is made explicit in (7.5).

For the Green's function, we have similarly

$$G(\mathbf{p}, \mathbf{p}') = \frac{1}{\Omega} \sum_{N=0}^{\infty} N\gamma(N) \chi^{(N)}(\phi). \tag{7.6}$$

8. COULOMB POTENTIAL

In this case, we know[1]

$$\mathring{F}(\mathbf{p}, \mathbf{p}') = e^2\pi^2 p_0 \left[\frac{1}{\Omega} \sum_{jmn} D^j_{mn}(\mathbf{p}) \bar{D}^j_{mn}(\mathbf{p}') \right] = e^2\pi^2 p_0 \sum_{\mu} D_\mu(\mathbf{p}) \bar{D}_\mu(\mathbf{p}') \frac{1}{d_\mu} \tag{8.1}$$

or

$$\mathring{c}_{\mu\nu} = \frac{e^2\pi^2 p_0}{N} \delta_{\mu\nu}. \tag{8.2}$$

Then,

$$\mathring{c}_{\mu\nu} = \mathring{c}\delta_{\mu\nu}, \quad \text{where} \quad \mathring{c} = \frac{A}{N}, \tag{8.3}$$

$$c_{\mu\nu} = c\delta_{\mu\nu}, \quad \text{where} \quad c = \frac{A}{N - a}, \tag{8.4}$$

$$\gamma_{\mu\nu} = \gamma\delta_{\mu\nu}, \quad \text{where} \quad \gamma = \frac{N}{N - a}, \tag{8.5}$$

and

$$A = e^2\pi^2 p_0, \tag{8.6}$$

$$a = e^2/2p_0. \tag{8.7}$$

To evaluate $F(\mathbf{p}, \mathbf{p}')$ and $G(\mathbf{p}, \mathbf{p}')$, consider

$$H(\mu, \phi, a) = \sum_{N=1}^{\infty} \frac{N\chi^N(\phi)\mu^N}{N - a} = \mu \frac{d}{d\mu} \sum_{N=1}^{\infty} \frac{\chi^N(\phi)\mu^N}{N - a}. \tag{8.8}$$

Let

$$x_\pm = \mu e^{\pm i\phi}. \tag{8.9}$$

Then

$$\mu^N \chi^N(\phi) = \mu^N \frac{\sin N\phi}{\sin \phi} = \frac{x_+^N - x_-^N}{2i \sin \phi}$$

and

$$H(\mu, \phi, a) = \mu \frac{d}{d\mu} S(\mu, \phi, a), \tag{8.10}$$

where

$$S(\mu, \phi, a) = \frac{1}{2i \sin \phi} \sum_1^{\infty} \frac{x_+^N - x_-^N}{N - a}. \tag{8.11}$$

We choose an integral representation of the sum as follows:

$$\sum_1^{\infty} \frac{x^N}{N - a} = \sum_1^{\infty} x^N \int_0^1 y^{N-a-1}\, dy = \int_0^1 y^{-a-1} \frac{xy}{1 - xy}\, dy,$$

and

$$S(\mu, \phi, a) = \mu \int_0^1 y^{-a} [1 - 2\mu y \cos \phi + \mu^2 y^2]^{-1}\, dy. \tag{8.12}$$

Therefore,

$$H(\mu, \phi, a) = \mu \frac{d}{d\mu} \int_0^1 \frac{\mu y^{-a}\, dy}{1 - 2\mu y \cos \phi + \mu^2 y^2} = \mu \int_0^1 dy y^{-a-1} \frac{d}{d\mu} \frac{\mu y}{1 - 2\mu \cos \phi + \mu^2 y^2} = \mu \int_0^1 dy y^{-a} \frac{d}{d(\mu y)} \frac{\mu y}{1 - 2\mu \cos \phi + (\mu y)^2}. \tag{8.13}$$

Similarly,

$$\frac{dH}{d\mu} = \int_0^1 dy y^{-a} \frac{d}{d(\mu y)} (\mu y) \frac{d}{d(\mu y)} \frac{\mu y}{1 - 2\mu \cos \phi + \mu^2 y^2}. \tag{8.14}$$

From Eqs. (7.3) and (8.4), we have the scattering amplitude:

$$F(\mathbf{p}, \mathbf{p}') = \frac{A}{\Omega} \sum \frac{N}{N - a} \chi^N(\phi) \tag{8.15}$$

$$= \frac{A}{\Omega} H(1, \phi, a). \tag{8.16}$$

The corresponding expression for the Green's function follows from (7.6) and (8.5):

$$-G(\mathbf{p}, \mathbf{p}') = \frac{1}{\Omega p_0^2} \sum \frac{N^2}{N - a} \chi^N(\phi) \tag{8.17}$$

$$= \frac{1}{\Omega p_0^2} \left(\frac{d}{d\mu} H(\mu, \phi, a) \right)_{\mu=1}. \tag{8.18}$$

Therefore,

$$F(\mathbf{p}, \mathbf{p}') = \frac{A}{\Omega} \int_0^1 dy y^{-a} \frac{d}{dy} \left(\frac{y}{1 - 2y \cos \phi + y^2} \right), \tag{8.19}$$

$$-G(\mathbf{p},\mathbf{p}') = \frac{1}{\Omega p_0^2}\int_0^1 dy y^{-a}\frac{d}{dy}\,y\,\frac{d}{dy}\,y \times\left(\frac{1}{1-2y\cos\phi+y^2}\right). \quad (8.20)$$

This representation of the Coulomb Green's function has been given by Schwinger.[2]

9. GENERAL CASIMIR POTENTIAL

The general diagonal potential may be expressed in terms of the Coulomb potential by introducing the Stieltjes transform:

$$c(N) = \int_0^\infty \frac{\tilde{c}(\lambda)}{N+\lambda}\,d\lambda, \quad (9.1)$$

where

$$\tilde{c}(\lambda) = \frac{i}{2\pi}\left[c(\lambda e^{i\pi}) - c(\lambda e^{-i\pi})\right] \quad (9.1a)$$

is the discontinuity of $c(\lambda)$ on the negative real axis. Then, by (7.3) and (9.1),

$$F(p,p') = \frac{1}{\Omega}\int_0^\infty d\lambda\tilde{c}(\lambda)\sum_{N=1}^\infty \frac{N}{N+\lambda}\chi^N(\phi)$$
$$= \frac{1}{\Omega}\int_0^\infty d\lambda\tilde{c}(\lambda)H(1,\phi,-\lambda)$$
$$= \frac{1}{\Omega}\int_0^\infty d\lambda\tilde{c}(\lambda)\int_0^1 dy y^\lambda\frac{d}{dy}\left(\frac{y}{1-2y\cos\phi+y^2}\right)$$

or

$$F(\mathbf{p},\mathbf{p}') = \frac{1}{\Omega}\int_0^1 \hat{c}(y)\frac{d}{dy}\left(\frac{y}{1-2y\cos\phi+y^2}\right)dy, \quad (9.2)$$

where

$$\hat{c}(y)/y = \int_0^\infty \tilde{c}(\lambda)y^{\lambda-1}\,d\lambda \quad (9.2a)$$

is the Mellin transform.

Starting from (7.6), we similarly obtain,

$$G(\mathbf{p},\mathbf{p}') = \frac{1}{\Omega}\int_0^1 \hat{\gamma}(y)\frac{d}{dy}\left(\frac{y}{1-2y\cos\phi+y^2}\right)dy, \quad (9.3)$$

where $\hat{\gamma}(y)$ is the Mellin transform of

$$\tilde{\gamma}(\lambda) = \frac{i}{2\pi}\left[\gamma(\lambda e^{i\pi}) - \gamma(\lambda e^{-i\pi})\right]. \quad (9.3a)$$

10. WATSON TRANSFORM

We again start with

$$F(\mathbf{p},\mathbf{p}') = \frac{1}{\Omega}\sum_1^\infty Nc(N)\chi^N(\phi)\mu^N. \quad (10.1)$$

Since N and $\chi^N(\phi)$ both vanish at $N=0$, the sum may be extended to include $N=0$, unless $c(N)$ becomes correspondingly infinite. We consider only those potentials for which $c(N)$ does not increase so fast.

[2] J. Schwinger, J. Math. Phys. 5, 1606 (1964).

Let us next make a Watson–Sommerfeld transformation by choosing a contour which includes the entire real axis. Then,

$$F(\mathbf{p},\mathbf{p}') = \frac{1}{2i\Omega}\oint\frac{\lambda(-\mu)^\lambda c(\lambda)\chi^{(\lambda)}(\phi)}{\sin\pi\lambda}\,d\lambda. \quad (10.2)$$

When the contour is opened up, one obtains poles from only $c(\lambda)$. Then, in the usual way,

$$F(\mathbf{p},\mathbf{p}') = -\frac{1}{2i\Omega}\int_{L-i\infty}^{L+i\infty}\frac{\lambda c(\lambda)(-\mu)^\lambda\chi^\lambda(\phi)}{\sin\pi\lambda}\,d\lambda + \frac{\pi}{\Omega}\sum_k\frac{r_k a_k(-\mu)^{a_k}\chi^{a_k}(\phi)}{\sin\pi a_k}, \quad (10.3)$$

where the sum is over the poles of $c(\lambda)$ and r_k is the residue at a_k.

In the Coulomb case, we have, by Eq. (8.4),

$$F(\mathbf{p},\mathbf{p}') = \frac{iA}{2\Omega}\int_{L-i\infty}^{L+i\infty}\frac{\lambda(-)^\lambda\chi^\lambda}{(\lambda-a)\sin\pi\lambda}\,d\lambda + \frac{A}{\Omega}\left[\frac{\pi a(-)^a}{\sin\pi a}\right]\chi^{(a)}, \quad (10.4)$$

and

$$G(\mathbf{p},\mathbf{p}') = -\frac{1}{\Omega}\left(\frac{dH}{d\mu}\right)_{\mu=1}, \quad (10.5)$$

where $H(\mu)$ is by Eq. (8.8) the same as Eq. (10.1) and therefore, also the same as Eq. (10.3) with

$$c(N) = \frac{\Omega}{N-a}.$$

Therefore,

$$H(\mu) = -\frac{1}{2i}\int_{L-i\infty}^{L+i\infty}\frac{\lambda}{\lambda-a}(-\mu)^\lambda\chi^{(\lambda)}\,d\lambda + \frac{\pi a}{\sin\pi a}(-\mu)^a\chi^{(a)}, \quad (10.6)$$

$$G(\mathbf{p},\mathbf{p}') = \frac{1}{2i\Omega}\int_{L-i\infty}^{L+i\infty}\frac{e^{\pi i\lambda}\lambda^2\chi^{(\lambda)}}{\lambda-a}\,d\lambda + \frac{\pi a^2}{\sin\pi a}e^{\pi i a}\chi^{(a)}. \quad (10.7)$$

11. CONTINUATION TO POSITIVE ENERGY

To obtain the physical limit, we continue from $z=-p_0^2$ to $z=s+i\epsilon$, where s is the scattering energy. Then,

$$G(p/p_0)^{-1} = 1 + p^2/p_0^2 = 1 - p^2/z.$$

After continuation to the scattering region, we have

$$G^{-1} \sim 1 - \frac{s}{s+i\epsilon} \sim \frac{i\epsilon}{s}. \quad (11.1)$$

Then,

$$\sin^2\frac{\phi}{2} \sim -\frac{1}{s}(\mathbf{p}-\mathbf{p}')^2 G(p)G(p'),$$

where $G(p)$ and $G(p')$ both stand for $s/i\epsilon$. Then also

$$\cos\phi \sim \frac{2}{s} G(p)G(p')(\mathbf{p}-\mathbf{p}')^2,$$

$$\sin\phi \sim \frac{2i}{s} G(p)G(p')(\mathbf{p}-\mathbf{p}')^2,$$

$$e^{i\phi} \sim 0,$$

$$e^{-i\phi} \sim 2\cos\phi.$$

Now let

$$x = e^{-i\phi}$$

$$= \frac{4}{s} G(p)G(p')(\mathbf{p}-\mathbf{p}')^2. \tag{11.2}$$

The character then goes to the following limit:

$$\chi^\lambda(\phi) = \frac{e^{i\lambda\phi} - e^{-i\lambda\phi}}{e^{i\phi} - e^{-i\phi}} \sim x^{\lambda-1}, \tag{11.3}$$

and (10.3) gives us

$$\Omega F(\mathbf{p}, \mathbf{p}') = \frac{i}{2}\int_{L-i\infty}^{L+i\infty} \frac{\lambda c(\lambda)(-\mu)^\lambda x^{\lambda-1}}{\sin\pi\lambda}\, d\lambda + \pi\sum_k \frac{r_k a_k (-\mu)^{a_k} x^{a_k-1}}{\sin\pi a_k}. \tag{11.4}$$

As we approach the physical limit, $x \to \infty$ and the integral goes as x^{L-1}.

Now assume that there are no poles of $c(\lambda)$ to the left of the imaginary axis (the situation for the Coulomb potential). Then the terms in the sum do not vanish faster than x^{-1}. On the other hand, since $L < 0$, the integral disappears in the physical limit and the exact solution is simply

$$F(\mathbf{p}, \mathbf{p}') = \frac{\pi}{\Omega x}\sum_k \frac{r_k a_k (-\mu)^{a_k} x^{a_k}}{\sin\pi a_k}. \tag{11.5}$$

In the Coulomb case,

$$c(\lambda) = \frac{A}{\lambda - a}, \quad \mu = 1.$$

Then there is just a single pole at a, and therefore,

$$F(\mathbf{p}, \mathbf{p}') = \frac{\pi}{\Omega}\frac{Aa(-1)^{-a}x^{-a-1}}{\sin\pi a}, \tag{11.6a}$$

which in the limit is the same as

$$F(\mathbf{p}, \mathbf{p}') = -\frac{A}{\Omega}\frac{\pi a}{\sin\pi a} e^{-\pi i a}\chi^{(a)}, \tag{11.6b}$$

where $\chi^{(a)}$ is the character, and

$$A = e^2\pi^2 p_0,$$

$$a = e^2/2p_0,$$

$$\Omega = \pi^2 p_0^3/4.$$

This expression for the scattering amplitude agrees with that found by Okubo and Feldman.[3]

It is remarkable that this simple expression is the exact solution of the Coulomb problem. To express it in a more familiar form, return to the variable x and use (11.2). Then,

$$f(p, p') = -\frac{e^2}{(p-p')^2}\left[-\frac{4(p-p')^2}{s}\right]^{-a} \times \Gamma(1-a)\Gamma(1+a)[G(p)G(p')]^a, \tag{11.7}$$

$$a = \frac{e^2}{2p} i.$$

The dependence on $t = -(p-p')^2$ is correct. However, the factor $[G(p)G(p')]^a$ becomes ∞^a as one goes onto the energy shell. This feature of the result stems from the infinite range of the Coulomb potential and requires a renormalization which distinguishes between the assumed asymptotic plane waves and the actual asymptotic states which are distorted plane waves. This point has been discussed by several authors.[4] In the Appendix, it is shown that in the physical limit

$$[G(p)G(p')]^a \to \left[\frac{1}{\Gamma(1-a)}\right]^2 (-1)^{1+a}. \tag{11.8}$$

One then gets the usual result for the complete amplitude:

$$f(p, p') = \frac{e^2}{t}\left(-\frac{t}{4s}\right)^{-(i/2)(e^2/k)} \frac{\Gamma[1-(ie^2/2k)]}{\Gamma[1+(ie^2/2k)]}. \tag{11.9}$$

In the general case of a Casimir amplitude, one finds from Eq. (11.5) and (11.8)

$$f(\mathbf{p}, \mathbf{p}') = \frac{1}{t}\sum_k r_k\left(-\frac{t}{s}\right)^{a_k} \frac{\Gamma(1+a_k)}{\Gamma(1-a_k)}. \tag{11.10}$$

This general result may be described as a superposition of Coulomb-like amplitudes. It is not the same as the amplitude produced by a superposition of Coulomb potentials, which, of course, would still be a Coulomb potential and would still be described by only a single pole.

12. REGGE POLES

The basis functions appearing in the expansion (10.1) are the characters of the rotation group and are related to the Tschebyscheff polynomials as follows: $U_{2j}(x) = \chi^N = (\sin N\phi/\sin\phi)$, where $x = \cos\phi$. In the Coulomb bound state problem, these basis functions are labeled by the principal quantum number. Therefore, we have an expansion in partial waves which are Tschebyscheff instead of Legendre functions and these are labeled by the principal instead of the angular momentum quantum number. The operator corresponding to the principal quantum number may be expressed in terms of the Runge–Lenz

[3] S. Okubo and D. Feldman, Phys. Rev. 117, 292 (1960).

[4] See, for example, W. Ford, Phys. Rev. 133, 1616 (1964).

vector. In the general non-Coulombic case, this vector is not conserved, but it may still be used to define the partial waves of Eq. (10.1). That is, of course, analogous to using an expansion in angular momentum states when the potential is not spherically symmetric.

We have made a Watson–Sommerfeld transformation of the character expansion instead of the Legendre expansion, and have thereby obtained a representation of Casimir scattering amplitudes as a sum over poles in the N plane, instead of the L plane.

According to Regge, the Watson–Sommerfeld transformation of the Legendre series for a Yukawian amplitude leads to the result that the scattering amplitude at large, unphysical (t/s) approaches the sum

$$f(p, p') = \sum_k \frac{\beta_k}{\sin \pi\alpha_k}\left(-\frac{t}{2s}\right)^{\alpha_k}, \qquad (12.1)$$

where the α_k are poles of the partial waves of the Legendre expansion when continued into the complex L plane. The corresponding transformation of the character series for Casimir amplitudes leads to the exact representation, valid at all physical t/s:

$$f(p, p) = \sum_k \frac{r_k}{\sin \pi a_k} a_k\left(-\frac{t}{s}\right)^{a_k}, \qquad (12.2)$$

where the a_k are the poles of the partial waves of the character expansion when continued into the complex N plane. It follows that the amplitudes produced by Yukawian potentials behave at large t/s like Casimir amplitudes insofar as they may be represented by Eqs. (12.1) and (12.2). Of course, the a_k trajectories are not the same as the α_k (Regge) trajectories, although they both give the bound states (for positive integral L and N, respectively). Finally, the Yukawian matrix does in fact approach diagonal form at large s/μ.

13. REMARKS

(a) This last result suggests that an approximate method based on the smallness of the off-diagonal elements might be useful.

(b) The general method indicated here can be applied to potentials which are not isotropic, and avoids the non-separability problem of the Schrodinger equation. In particular, it appears natural to investigate multipole potentials since they are simply derivatives of the Coulomb potential and simpler in the context of the rotation group than Yukawians (and also less singular than r^{-n}).

(c) The general potential corresponding to the matrix $\mathring{c}$ is not local. Although all local potentials may be represented in this way, it would be interesting to know how locality limits $\mathring{c}$.

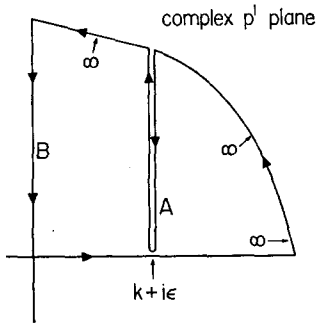


FIG. 1. Contour used in integration.

APPENDIX

We are interested in the limit of

$$L(p, p') = [G(p)G(p')]^a$$

as we go onto the energy shell. As already remarked, $L(p, p')$ blows up if taken between states of precisely defined momentum, but these are not correct asymptotic states if the potential has infinite range. Therefore, we assume initial and final wave packets $\omega(p - k)$ and calculate the following amplitude:

$$L(\mathbf{k}) = \iint L(\mathbf{p}, \mathbf{p}')\omega(\mathbf{p} - \mathbf{k})\omega(\mathbf{p}' - \mathbf{k})\, d\mathbf{p}\, d\mathbf{p}'.$$

Then, $L(\mathbf{k}) = I(\mathbf{k})^2$, where

$$I(\mathbf{k}) = \int [G(\mathbf{p})]^a \omega(\mathbf{p} - \mathbf{k})\, d\mathbf{p}.$$

To calculate this, choose wave packets of the following form:

$$\omega_\Delta(p - k) = \delta(\Omega - \Omega_0) \times \frac{\exp\left(i\frac{(p-k)}{\Delta}\right) - \exp\left(-i\frac{(p-k)}{\Delta}\right)}{2\pi i(p - k - i\epsilon)}.$$

This packet has a spread Δ and, as $\Delta \to 0$, becomes a δ function. We have explicitly (after angular integrations)

$$I(k) = \int_0^\infty \frac{dp'[e^{i(p'-k)/\Delta} - e^{-i(p'-k)/\Delta}]}{(2\pi i)(p' + k)^{-a}(p' - k - i\epsilon)^{1-a}}.$$

We evaluate this integral by utilizing the contour shown in Fig. 1.

In the limit $\Delta \to 0$, only the contribution from the cut A is significant. The path B gives a part which goes to zero like Δ. We find that

$$I(k) = \frac{(i)^a}{\Gamma(1 - a)i}(2k\Delta)^a = \frac{i^a}{\Gamma(1 - a)i}\sigma.$$

We renormalize by dropping the infinite phase factor (σ). Then we find that

$$L(k) = \left(\frac{1}{\Gamma(1 - a)}\right)^2(-1)^{1+a},$$

as claimed.

A more careful discussion of this question will be found in a forthcoming paper by one of the authors (D. L.).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 10 OCTOBER 1967

A Necessary Restriction on Correlation Functions Arising from 3- (or More) Particle Systems in One Dimension

H. R. Post
Chelsea College of Science and Technology, London, England

(Received 15 March 1967)

A necessary condition on any pair-correlation function, arising from the symmetry of a system of three or more identical particles, is derived.

THE purpose of this paper is to point out a necessary restriction on the pair-correlation function corresponding to a three- (or more) particle system.

By a pair-correlation function, we mean that function $C(\xi)$ which denotes the probability of encountering *any* two particles at a distance ξ from each other (we consider here a one-dimensional laboratory space). This is the pair-correlation function we are *forced* to use if the system under consideration consists of identical bosons or fermions (i.e., particles without individuality), and is the function we may *choose* to use *if* we are interested in the distance between *any* two particles, regardless of identity, even in the case of distinguishable particles.

We are concerned here with a system of a finite number of particles $N \geq 3$ distributed over the space $-\infty \leq x \leq +\infty$.

$$\int_{-\infty}^{+\infty} C(\xi)\, d\xi = (N-1)N$$

is the normalizing condition.

$$C(\xi) = C(-\xi)$$

and

$$C(\xi) \geq 0 \quad \text{everywhere.}$$

We now establish a further condition:

$$\int_a^{2a} C(\xi)\, d\xi \leq \tfrac{2}{3} \int_0^{+\infty} C(\xi)\, d\xi \quad \text{for all } a.$$

The proof is as follows.

Consider any configuration of labeled particles 1, 2, 3, $\cdots$ in the locations $x_1, x_2, x_3, \cdots$.

This corresponds to a system of the following values:

$$\xi = \xi_3 = +(x_1 - x_2), \quad \xi = \xi_3' = (x_2 - x_1) = -\xi_3,$$
$$\xi = \xi_2 = +(x_1 - x_3), \quad \xi = \xi_2' = (x_3 - x_1) = -\xi_2,$$

and

$$\xi = \xi_1 = +(x_2 - x_3) = \xi_2 - \xi_3,$$
$$\xi = \xi_1' = (x_3 - x_2) = \xi_2' - \xi_3' = -\xi_1.$$

Thus, we have a "Ritz spectrum" (a triplet in the positive region of ξ) such that the location of any "line" (δ function) is given by a ξ value which is the sum or difference of the ξ values of the other two lines.

A consequence of this relationship within the triplet is that at most two of the δ functions may be in the interval $a \leq \xi \leq 2a$, since, for any two δ-function members of a Ritz triplet within the interval $a \leq \xi \leq 2a$, there must be a third δ-function member of the triplet outside that interval, either in the interval $0 \leq \xi \leq a$ or in the region $2a \leq \xi$. Thus, for this Ritz spectrum's contribution $R(\xi)$ to $C(\xi)$ we have

$$\int_a^{2a} R(\xi)\, d\xi \leq \tfrac{2}{3} \int_0^{\infty} R(\xi)\, d\xi \quad \text{for all } a. \qquad (1)$$

Any $C(\xi)$ function can be regarded as a superposition of such Ritz triplets. Now the relationship (1) is "inherited" in the sense that any superposition of R's again fulfills the condition (1). Hence,

$$\int_a^{2a} C(\xi)\, d\xi \leq \tfrac{2}{3} \int_0^{\infty} C(\xi)\, d\xi \quad \text{for all } a. \quad \text{Q.E.D.}$$

We had to choose the somewhat awkward integral condition (1), since the Ritz principle as such is not "inherited" in the sense that it does not hold in general for a superposition of "alien" triplets, and thus does not yield a necessary condition on the correlation function.

The restriction is significant for systems of particles whose interactions include short-range repulsion, and for fermions in general. It holds *a fortiori* for $N > 3$.