A note on the symmetries of the 3*j* and 6*j* coefficients. II

K. Venkatesh
Department of Physics, University of Mysore, Manasagangotri, Mysore 570 006, India

(Received 31 May 1979; accepted for publication 20 August 1979)

The partition of the symmetries of the 3*j* and 6*j* coefficients according to the possible expressions for the number of terms in their series representations are discussed. The canonical parametrization of the 3*j* and 6*j* coefficients introduced by Lockwood are also discussed.

1. INTRODUCTION

In our earlier studies[1,2] we showed that the 72 (144) symmetries of the 3*j* (6*j*) coefficient can be partitioned into nine (twelve) sets of eight (twelve) each depending on the nine (twelve) possible expressions for the number of terms in the series representation for the 3*j* (6*j*) coefficient.

It has been shown[3,4] that the definition of a set of five (six) canonical parameters of the 3*j* (6*j*) coefficient introduces a nine (twelve)-to-one homomorphism of the 72 (144) element group of symmetries on to the eight (twelve)-element group of permutations of canonical parameters. To explicitly obtain the homomorphisms, the intermediate parameters corresponding to each symmetry should be written following the initial order and thus the mapping on to the permutation of canonical parameters can be obtained.

In this paper, we show that the above procedure of obtaining the homomorphisms is exactly equivalent to the following elegant procedure:

(i) We first list the sets of symmetries corresponding to the possible expressions for n starting with the particular n to whose group of permutations of canonical parameters the homomorphisms are due. This n is determined from the minimum of the possible expressions for the number of terms in the series representation for the 3*j* (or 6*j*) coefficient. In the case of the 3*j* coefficient n is given by the minimum of the entries of the Regge square symbol. In our discussion n is taken as $j_1+j_2-j_3$ and $c+d-e$ for the 3*j* and 6*j* coefficients, respectively.

(ii) We list the possible sets of canonical parameters of the 3*j* and 6*j* coefficients.

(iii) The homomorphisms are then obtained by calculating the canonical parameters corresponding to the possible expressions for n.

We discuss these procedures for the 3*j* and 6*j* coefficients in Sec. 2 and Sec. 3, respectively.

2. 3*j* COEFFICIENT

The 72 symmetries of the 3*j* coefficient can be partitioned[1] into nine sets of eight each depending on the nine possible expressions for the number of terms in the series representation for the 3*j* coefficient. We have explicitly calculated these sets of symmetries starting with $n=j_1+j_2-j_3$.

The symmetries of the 3*j* coefficient were discussed in terms of a set of five canonical parameters by Lockwood.[3] The nine sets of canonical parameters[5] corresponding to the possible expressions for n are listed in Appendix A. To explicitly obtain the homomorphisms the intermediate parameters corresponding to each symmetry should be written following the initial order and then the mapping on to the permutation of canonical parameters is obtained. We demonstrate this procedure for a few examples.

Example 2.1: Consider the symmetry

$$\begin{pmatrix} j_2 & j_1 & j_3 \\ -m_2 & -m_1 & -m_3 \end{pmatrix}.$$

Writing the intermediate parameters, viz., β's and α's following the order in which we have defined[1] them, we have

$$\beta_1=j_2+j_1-j_3, \quad \beta_2=j_2+m_2, \quad \beta_3=j_1-m_1,$$
$$\alpha_1=0, \quad \alpha_2=j_1-j_3+m_2, \quad \alpha_3=j_2-j_3-m_1.$$

Since $n=j_1+j_2-j_3, \beta_0=j_1+j_2-j_3$ and $\alpha_0=0$. So, the mapping is obtained as

$$\begin{pmatrix} j_2 & j_1 & j_3 \\ -m_2 & -m_1 & -m_3 \end{pmatrix} \rightarrow b_L a_L; d_L c_L.$$

Example 2.2:

$$(-1)^J \begin{pmatrix} \dfrac{(j_2+m_2)+(j_1+m_1)}{2} & \dfrac{(j_2-m_2)+(j_1-m_1)}{2} & j_3 \\ \dfrac{(j_2+m_2)-(j_1+m_1)}{2} & \dfrac{(j_2-m_2)-(j_1-m_1)}{2} & j_1-j_2 \end{pmatrix}$$

 0022-2488/80/071555-07$1.00

The intermediate parameters corresponding to this symmetry are

$$\begin{aligned} &\beta_1 = j_1 + j_2 - j_3, \quad \beta_2 = j_1 + m_1, \quad \beta_3 = (j_2 - m_2), \\ &\alpha_1 = 0, \quad \alpha_2 = j_1 - j_3 - m_2, \quad \alpha_3 = j_2 - j_3 + m_1,. \\ &n = j_1 + j_2 - j_3 \quad \Rightarrow \beta_0 = j_1 + j_2 - j_3, \quad \text{and} \quad \alpha_0 = 0. \end{aligned}$$

So, the corresponding permutation of the canonical parameters is $d_L c_L; a_L b_L$.

Example 2.3:

$$\begin{pmatrix} j_3 & j_2 & j_1 \\ -m_3 & -m_2 & -m_1 \end{pmatrix}, \quad \begin{aligned} &\beta_1 = j_3 + j_2 - j_1, \quad \beta_2 = j_3 + m_3, \quad \beta_3 = j_2 - m_2, \\ &\alpha_1 = 0, \quad \alpha_2 = j_2 - j_1 + m_3, \quad \alpha_3 = j_3 - j_1 - m_2,. \\ &n = j_1 + j_2 - j_3 \quad \Rightarrow \beta_0 = j_2 - m_2, \quad \alpha_0 = j_3 - j_1 - m_2. \end{aligned}$$

Hence, the corresponding permutation is $b_L a_L; d_L c_L$.

Example 2.4:

$$(-1)^J \begin{pmatrix} \dfrac{j_1 + j_3 + m_2}{2} & \dfrac{j_2 + j_3 + m_1}{2} & \dfrac{j_1 + j_2 + m_3}{2} \\ \dfrac{(j_3 + j_1 - j_2) - (j_2 + m_2)}{2} & \dfrac{(j_2 + j_3 - j_1) - (j_1 + m_1)}{2} & \dfrac{(j_1 + j_2 - j_3) - (j_3 + m_3)}{2} \end{pmatrix},$$

$$\begin{aligned} &\beta_1 = j_3 - m_3, \quad \beta_2 = j_2 + m_2, \quad \beta_3 = j_2 + j_3 - j_1, \\ &\alpha_1 = 0 \quad \alpha_2 = j_2 - j_1 - m_3, \quad \alpha_3 = j_3 - j_1 + m_2, \\ &n = j_1 + j_2 - j_3 \quad \Rightarrow \beta_0 = j_2 + m_2, \quad \alpha_0 = j_3 - j_1 + m_2. \end{aligned}$$

The corresponding permutation is $d_L c_L; a_L b_L$.

Example 2.5:

$$\begin{pmatrix} \dfrac{j_1 + j_2 - m_3}{2} & \dfrac{j_1 + j_3 - m_2}{2} & \dfrac{j_2 + j_3 - m_1}{2} \\ \dfrac{(j_1 + j_2 - j_3) - (j_3 - m_3)}{2} & \dfrac{(j_1 + j_3 - j_2) - (j_2 - m_2)}{2} & \dfrac{(j_2 + j_3 - j_1) - (j_1 - m_1)}{2} \end{pmatrix},$$

$$\begin{aligned} &\beta_1 = j_1 + m_1, \quad \beta_2 = j_3 - m_3, \quad \beta_3 = j_1 + j_3 - j_2, \\ &\alpha_1 = 0, \quad \alpha_2 = j_3 - j_2 + m_1, \quad \alpha_3 = j_1 - j_2 - m_3, \\ &n = j_1 + j_2 - j_3 \quad \Rightarrow \beta_0 = j_1 + m_1, \quad \alpha_0 = j_3 - j_2 + m_1. \end{aligned}$$

The corresponding permutation is $a_L b_L; d_L c_L$.

We have checked this procedure for all the 72 symmetries of the $3j$ coefficient. We find that this procedure is exactly equivalent to calculating the canonical parameters corresponding to the possible expressions for n after listing the nine sets of symmetries.

The eight symmetries described by $n = j_1 + j_2 - j_3$ listed in Table II of Ref. 1 get mapped on to the eight-element group of permutations of canonical parameters corresponding to $n = j_1 + j_2 - j_3$ as follows:

$$\begin{aligned} &a_L b_L; c_L d_L \qquad\qquad b_L a_L; d_L c_L \\ &\qquad a_L b_L; d_L c_L \\ &\qquad b_L a_L; c_L d_L \\ &\qquad c_L d_L; b_L a_L \\ &\qquad d_L c_L; a_L b_L \\ &c_L d_L; a_L b_L \qquad\qquad d_L c_L; b_L a_L \ . \end{aligned}$$

For example the eight symmetries each described by $n = j_2 + j_3 - j_1, n = j_3 + m_3$, and $n = j_1 + m_1$ and their mapping on to the eight-element group of permutations of canonical parameters corresponding to $n = j_1 + j_2 - j_3$ are given in Tables I, II, and III, respectively. The mapping is obtained by calculating the canonical parameters (listed in Appendix A) corresponding to $n = j_2 + j_3 - j_1, n = j_3 + m_3$, and $n = j_1 + m_1$, respectively. These mappings have also been obtained by writing the intermediate parameters for each symmetry and the two procedures are strictly equivalent.

3. 6*j* COEFFICIENT

The 144 symmetries of the $6j$ coefficient can be partitioned[2] into 12 sets of 12 each depending on the 12 possible expressions for the number of terms in the series representation for the $6j$ coefficient. We have explicitly listed these sets of symmetries starting with $n = c + d - e$.

TABLE I. The set of eight symmetries of the $3j$ coefficient described by $n = j_2 + j_3 - j_1$ and their mapping onto the eight-element group of permutations of canonical parameters corresponding to $n = j_1 + j_2 - j_3$.

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \to a_l\, b_l\, ; c_l\, d_l, \quad \begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & -m_2 & -m_3 \end{pmatrix} \to b_l\, a_l\, ; d_l\, c_l$$

$$\begin{pmatrix} j_1 & \dfrac{(j_2 - m_2) + (j_3 - m_3)}{2} & \dfrac{(j_2 + m_2) + (j_3 + m_3)}{2} \\ j_2 - j_3 & \dfrac{(j_3 - m_3) - (j_2 - m_2)}{2} & \dfrac{(j_3 + m_3) - (j_2 + m_2)}{2} \end{pmatrix} \to b_l\, a_l\, ; c_l\, d_l$$

$$\begin{pmatrix} j_2 & \dfrac{(j_3 + m_3) + (j_1 + m_1)}{2} & \dfrac{(j_3 - m_3) + (j_1 - m_1)}{2} \\ j_3 - j_1 & \dfrac{(j_1 + m_1) - (j_3 + m_3)}{2} & \dfrac{(j_1 - m_1) - (j_3 - m_3)}{2} \end{pmatrix} \to a_l\, b_l\, ; d_l\, c_l$$

$$(-1)^{l} \begin{pmatrix} j_3 & \dfrac{(j_1 + m_1) + (j_2 + m_2)}{2} & \dfrac{(j_1 - m_1) + (j_2 - m_2)}{2} \\ j_1 - j_2 & \dfrac{(j_2 + m_2) - (j_1 + m_1)}{2} & \dfrac{(j_2 - m_2) - (j_1 - m_1)}{2} \end{pmatrix} \to d_l\, c_l\, ; a_l\, b_l$$

$$(-1)^{l} \begin{pmatrix} j_3 & \dfrac{(j_2 - m_2) + (j_1 - m_1)}{2} & \dfrac{(j_2 + m_2) + (j_1 + m_1)}{2} \\ j_2 - j_1 & \dfrac{(j_1 - m_1) - (j_2 - m_2)}{2} & \dfrac{(j_1 + m_1) - (j_2 + m_2)}{2} \end{pmatrix} \to c_l\, d_l\, ; b_l\, a_l$$

$$(-1)^{l} \begin{pmatrix} j_3 & j_1 & j_2 \\ -m_3 & -m_1 & -m_2 \end{pmatrix} \to c_l\, d_l\, ; a_l\, b_l, \quad (-1)^{l} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \to d_l\, c_l\, ; b_l\, a_l$$

TABLE II. The set of eight symmetries of the $3j$ coefficient described by $n = j_3 + m_3$ and their mapping on to the eight-element group of permutations of canonical parameters corresponding to $n = j_1 + j_2 - j_3$.

$$\begin{pmatrix} \dfrac{j_1 + j_3 - m_2}{2} & \dfrac{j_2 + j_3 - m_1}{2} & \dfrac{j_1 + j_2 - m_3}{2} \\ \dfrac{(j_1 + j_3 - j_2) - (j_3 - m_3)}{2} & \dfrac{(j_2 + j_3 - j_1) - (j_1 - m_1)}{2} & \dfrac{(j_1 + j_2 - j_3) - (j_3 - m_3)}{2} \end{pmatrix} \to b_l\, a_l\, ; c_l\, d_l$$

$$\begin{pmatrix} \dfrac{j_2 + j_3 + m_1}{2} & \dfrac{j_1 + j_3 + m_2}{2} & \dfrac{j_1 + j_2 + m_3}{2} \\ \dfrac{(j_2 + j_3 - j_1) - (j_1 + m_1)}{2} & \dfrac{(j_1 + j_3 - j_2) - (j_2 + m_2)}{2} & \dfrac{(j_1 + j_2 - j_3) - (j_3 + m_3)}{2} \end{pmatrix} \to a_l\, b_l\, ; d_l\, c_l$$

$$\begin{pmatrix} \dfrac{(j_2 - m_2) + (j_3 - m_3)}{2} & \dfrac{(j_2 + m_2) + (j_3 + m_3)}{2} & j_1 \\ \dfrac{(j_3 - m_3) - (j_2 - m_2)}{2} & \dfrac{(j_3 + m_3) - (j_2 + m_2)}{2} & j_2 - j_3 \end{pmatrix} \to b_l\, a_l\, ; d_l\, c_l$$

$$\begin{pmatrix} \dfrac{(j_1 + m_1) + (j_3 + m_3)}{2} & \dfrac{(j_1 - m_1) + (j_3 - m_3)}{2} & j_2 \\ \dfrac{(j_3 + m_3) - (j_1 + m_1)}{2} & \dfrac{(j_3 - m_3) - (j_1 - m_1)}{2} & j_1 - j_3 \end{pmatrix} \to a_l\, b_l\, ; c_l\, d_l$$

$$(-1)^{l} \begin{pmatrix} \dfrac{(j_1 - m_1) + (j_3 - m_3)}{2} & \dfrac{(j_1 + m_1) + (j_3 + m_3)}{2} & j_2 \\ \dfrac{(j_3 - m_3) - (j_1 - m_1)}{2} & \dfrac{(j_3 + m_3) - (j_1 + m_1)}{2} & j_1 - j_3 \end{pmatrix} \to c_l\, d_l\, ; a_l\, b_l$$

$$(-1)^{l} \begin{pmatrix} \dfrac{(j_2 + m_2) + (j_3 + m_3)}{2} & \dfrac{(j_2 - m_2) + (j_3 - m_3)}{2} & j_1 \\ \dfrac{(j_3 + m_3) - (j_2 + m_2)}{2} & \dfrac{(j_3 - m_3) - (j_2 - m_2)}{2} & j_2 - j_3 \end{pmatrix} \to d_l\, c_l\, ; b_l\, a_l$$

$$(-1)^{l} \begin{pmatrix} \dfrac{j_1 + j_3 + m_2}{2} & \dfrac{j_2 + j_3 + m_1}{2} & \dfrac{j_1 + j_2 + m_3}{2} \\ \dfrac{(j_1 + j_3 - j_2) - (j_2 + m_2)}{2} & \dfrac{(j_2 + j_3 - j_1) - (j_1 + m_1)}{2} & \dfrac{(j_1 + j_2 - j_3) - (j_3 + m_3)}{2} \end{pmatrix} \to d_l\, c_l\, ; a_l\, b_l$$

$$(-1)^{l} \begin{pmatrix} \dfrac{j_2 + j_3 - m_1}{2} & \dfrac{j_1 + j_3 - m_2}{2} & \dfrac{j_1 + j_2 - m_3}{2} \\ \dfrac{(j_2 + j_3 - j_1) - (j_1 - m_1)}{2} & \dfrac{(j_1 + j_3 - j_2) - (j_2 - m_2)}{2} & \dfrac{(j_1 + j_2 - j_3) - (j_3 - m_3)}{2} \end{pmatrix} \to c_l\, d_l\, ; b_l\, a_l$$

TABLE III. The set of eight symmetries of the $3j$ coefficient described by $n = j_1 + m_1$ and their mapping on to the eight-element group of permutations of canonical parameters corresponding to $n = j_1 + j_2 - j_3$.

$$\begin{pmatrix} j_1 & \frac{(j_2-m_2)+(j_3-m_3)}{2} & \frac{(j_2+m_2)+(j_3+m_3)}{2} \\ j_2-j_3 & \frac{(j_3-m_3)-(j_2-m_2)}{2} & \frac{(j_3+m_3)-(j_2+m_2)}{2} \end{pmatrix} \to a_L b_L; c_L d_L$$

$$\begin{pmatrix} j_2 & \frac{(j_1+m_1)+(j_3+m_3)}{2} & \frac{(j_1-m_1)+(j_3-m_3)}{2} \\ j_1-j_3 & \frac{(j_3+m_3)-(j_1+m_1)}{2} & \frac{(j_3-m_3)-(j_1-m_1)}{2} \end{pmatrix} \to b_L a_L; d_L c_L$$

$$\begin{pmatrix} \frac{j_1+j_2+m_3}{2} & \frac{j_2+j_3+m_1}{2} & \frac{j_1+j_3+m_2}{2} \\ \frac{(j_1+j_2-j_3)-(j_3+m_3)}{2} & \frac{(j_2+j_3-j_1)-(j_1+m_1)}{2} & \frac{(j_3+j_1-j_2)-(j_2+m_2)}{2} \end{pmatrix} \to b_L a_L; c_L d_L$$

$$\begin{pmatrix} \frac{j_1+j_2-m_3}{2} & \frac{j_1+j_3-m_2}{2} & \frac{j_2+j_3-m_1}{2} \\ \frac{(j_1+j_2-j_3)-(j_3-m_3)}{2} & \frac{(j_1+j_3-j_2)-(j_2-m_2)}{2} & \frac{(j_2+j_3-j_1)-(j_1-m_1)}{2} \end{pmatrix} \to a_L b_L; d_L c_L$$

$$(-1)^J \begin{pmatrix} \frac{j_1+j_2-m_3}{2} & \frac{j_2+j_3-m_1}{2} & \frac{j_1+j_3-m_2}{2} \\ \frac{(j_1+j_2-j_3)-(j_3-m_3)}{2} & \frac{(j_2+j_3-j_1)-(j_1-m_1)}{2} & \frac{(j_1+j_3-j_2)-(j_2-m_2)}{2} \end{pmatrix} \to d_L c_L; a_L b_L$$

$$(-1)^J \begin{pmatrix} \frac{j_1+j_2+m_3}{2} & \frac{j_1+j_3+m_2}{2} & \frac{j_2+j_3+m_1}{2} \\ \frac{(j_1+j_2-j_3)-(j_3+m_3)}{2} & \frac{(j_1+j_3-j_2)-(j_2+m_2)}{2} & \frac{(j_2+j_3-j_1)-(j_1+m_1)}{2} \end{pmatrix} \to c_L d_L; b_L a_L$$

$$(-1)^J \begin{pmatrix} j_1 & \frac{(j_2+m_2)+(j_3+m_3)}{2} & \frac{(j_2-m_2)+(j_3-m_3)}{2} \\ j_2-j_3 & \frac{(j_3+m_3)-(j_2+m_2)}{2} & \frac{(j_3-m_3)-(j_2-m_2)}{2} \end{pmatrix} \to c_L d_L; a_L b_L$$

$$(-1)^J \begin{pmatrix} j_2 & \frac{(j_1-m_1)+(j_3-m_3)}{2} & \frac{(j_1+m_1)+(j_3+m_3)}{2} \\ j_1-j_3 & \frac{(j_3-m_3)-(j_1-m_1)}{2} & \frac{(j_3+m_3)-(j_1+m_1)}{2} \end{pmatrix} \to d_L c_L; b_L a_L$$

TABLE IV. The set of 12 symmetries (Ref. 6) of the $6j$ coefficient described by $n = c + d - e$ and their mapping on to the 12-element group of permutations of canonical parameters.

$$\begin{Bmatrix} a & b & e \\ d & c & f \end{Bmatrix} \to a_L b_L c_L; d_L e_L, \quad \begin{Bmatrix} b & a & e \\ c & d & f \end{Bmatrix} \to a_L c_L b_L; e_L d_L$$

$$\begin{Bmatrix} a_1 & b_1 & e_1 \\ d_1 & c_1 & f_1 \end{Bmatrix} \to c_L b_L a_L; d_L e_L, \quad \begin{Bmatrix} b_1 & a_1 & e_1 \\ c_1 & d_1 & f_1 \end{Bmatrix} \to b_L c_L a_L; e_L d_L$$

$$\begin{Bmatrix} e_2 & b_2 & a_2 \\ f_2 & c_2 & d_2 \end{Bmatrix} \to b_L a_L c_L; d_L e_L, \quad \begin{Bmatrix} e_2 & a_2 & b_2 \\ f_2 & d_2 & c_2 \end{Bmatrix} \to c_L a_L b_L; e_L d_L$$

$$\begin{Bmatrix} a_3 & b_3 & e_3 \\ d_3 & c_3 & f_3 \end{Bmatrix} \to a_L b_L c_L; e_L d_L, \quad \begin{Bmatrix} b_3 & a_3 & e_3 \\ c_3 & d_3 & f_3 \end{Bmatrix} \to a_L c_L b_L; d_L e_L$$

$$\begin{Bmatrix} e_4 & a_4 & b_4 \\ f_4 & d_4 & c_4 \end{Bmatrix} \to b_L c_L a_L; d_L e_L, \quad \begin{Bmatrix} e_4 & b_4 & a_4 \\ f_4 & c_4 & d_4 \end{Bmatrix} \to c_L b_L a_L; e_L d_L$$

$$\begin{Bmatrix} b_5 & e_5 & a_5 \\ c_5 & f_5 & d_5 \end{Bmatrix} \to c_L a_L b_L; d_L e_L, \quad \begin{Bmatrix} a_5 & e_5 & b_5 \\ d_5 & f_5 & c_5 \end{Bmatrix} \to b_L a_L c_L; e_L d_L$$

TABLE V. The set of 12 symmetries (Ref. 6) of the $6j$ coefficient described by $n = a + f - c$ and their mapping on to the 12-element group of permutations of canonical parameters corresponding to $n = c + d - e$.

$$\begin{Bmatrix} d & f & b \\ a & e & c \end{Bmatrix} \to c_L a_L b_L; d_L e_L, \quad \begin{Bmatrix} c & f & a \\ b & e & d \end{Bmatrix} \to b_L a_L c_L; e_L d_L$$

$$\begin{Bmatrix} d_1 & e_1 & c_1 \\ a_1 & f_1 & b_1 \end{Bmatrix} \to a_L c_L b_L; d_L e_L, \quad \begin{Bmatrix} c_1 & e_1 & d_1 \\ b_1 & f_1 & a_1 \end{Bmatrix} \to a_L b_L c_L; e_L d_L$$

$$\begin{Bmatrix} d_2 & f_2 & b_2 \\ a_2 & e_2 & c_2 \end{Bmatrix} \to c_L a_L b_L; e_L d_L, \quad \begin{Bmatrix} c_2 & f_2 & a_2 \\ b_2 & e_2 & d_2 \end{Bmatrix} \to b_L a_L c_L; d_L e_L$$

$$\begin{Bmatrix} f_3 & d_3 & b_3 \\ e_3 & a_3 & c_3 \end{Bmatrix} \to c_L b_L a_L; d_L e_L, \quad \begin{Bmatrix} f_3 & c_3 & a_3 \\ e_3 & b_3 & d_3 \end{Bmatrix} \to b_L c_L a_L; e_L d_L$$

$$\begin{Bmatrix} d_4 & b_4 & f_4 \\ a_4 & c_4 & e_4 \end{Bmatrix} \to c_L b_L a_L; e_L d_L, \quad \begin{Bmatrix} c_4 & a_4 & f_4 \\ b_4 & d_4 & e_4 \end{Bmatrix} \to b_L c_L a_L; d_L e_L$$

$$\begin{Bmatrix} e_5 & d_5 & c_5 \\ f_5 & a_5 & b_5 \end{Bmatrix} \to a_L c_L b_L; e_L d_L, \quad \begin{Bmatrix} e_5 & c_5 & d_5 \\ f_5 & b_5 & a_5 \end{Bmatrix} \to a_L b_L c_L; d_L e_L$$

TABLE VI. The set of 12 symmetries (Ref. 6) of the 6j coefficient described by $n = b + f - d$ and their mapping on to the 12-element group of permutations of canonical parameters corresponding to $n = c + d - c$.

$$\begin{Bmatrix} f & c & a \\ e & b & d \end{Bmatrix} \to b_L c_L a_L; e_L d_L, \quad \begin{Bmatrix} f & d & b \\ e & a & c \end{Bmatrix} \to c_L b_L a_L; d_L e_L$$

$$\begin{Bmatrix} f_1 & b_1 & d_1 \\ e_1 & c_1 & a_1 \end{Bmatrix} \to c_L b_L a_L; e_L d_L, \quad \begin{Bmatrix} f_1 & a_1 & c_1 \\ e_1 & d_1 & b_1 \end{Bmatrix} \to b_L c_L a_L; d_L e_L$$

$$\begin{Bmatrix} d_2 & c_2 & e_2 \\ a_2 & b_2 & f_2 \end{Bmatrix} \to a_L c_L b_L; e_L d_L, \quad \begin{Bmatrix} c_2 & d_2 & e_2 \\ b_2 & a_2 & f_2 \end{Bmatrix} \to a_L b_L c_L; d_L e_L$$

$$\begin{Bmatrix} d_3 & f_3 & b_3 \\ a_3 & e_3 & c_3 \end{Bmatrix} \to c_L a_L b_L; d_L e_L, \quad \begin{Bmatrix} c_3 & f_3 & a_3 \\ b_3 & e_3 & d_3 \end{Bmatrix} \to b_L a_L c_L; e_L d_L$$

$$\begin{Bmatrix} d_4 & e_4 & c_4 \\ a_4 & f_4 & b_4 \end{Bmatrix} \to a_L b_L c_L; e_L d_L, \quad \begin{Bmatrix} c_4 & e_4 & d_4 \\ b_4 & f_4 & a_4 \end{Bmatrix} \to a_L c_L b_L; d_L e_L$$

$$\begin{Bmatrix} a_5 & c_5 & f_5 \\ d_5 & b_5 & e_5 \end{Bmatrix} \to b_L a_L c_L; d_L e_L, \quad \begin{Bmatrix} b_5 & d_5 & f_5 \\ c_5 & a_5 & e_5 \end{Bmatrix} \to c_L a_L b_L; e_L d_L$$

The symmetries of the 6j coefficient were discussed in terms of a set of six canonical parameters by Lockwood.[4] The 12 sets of canonical parameters corresponding to the 12 possible expressions for n are listed in Appendix A. As described above in the case of 3j coefficient, we give below a few examples of obtaining the mapping on to the permutation of canonical parameters by writing the intermediate parameters corresponding to each symmetry. Since we choose $n = c + d - e$ in our calculation, $\beta_0 = a + b + c + d$ and $\alpha_0 = a + b + e$.

Example 3.1: Consider the symmetry

$$\begin{Bmatrix} b & a & e \\ c & d & f \end{Bmatrix}.$$

Writing the intermediate parameters, viz, β's and α's following the order in which we have defined them,[2] we have

$$\beta_1 = a + b + c + d, \quad \beta_2 = b + c + e + f, \quad \beta_3 = a + d + e + f,$$
$$\alpha_1 = a + b + e, \quad \alpha_2 = c + d + e, \quad \alpha_3 = b + d + f, \quad \alpha_4 = a + c + f.$$

So, the mapping is

$$\begin{Bmatrix} b & a & e \\ c & d & f \end{Bmatrix} \to a_L c_L b_L; e_L d_L.$$

Example 3.2:

$$\begin{Bmatrix} e_4 & b_4 & a_4 \\ f_4 & c_4 & d_4 \end{Bmatrix} = \begin{Bmatrix} \frac{b+a+c-d}{2} & \frac{e+d+a-f}{2} & \frac{e+f+b-c}{2} \\ \frac{b+d+c-a}{2} & \frac{a+f+d-e}{2} & \frac{e+f+c-b}{2} \end{Bmatrix},$$

$$\beta_1 = b + c + a + d, \quad \beta_2 = b + c + e + f, \quad \beta_3 = a + d + e + f,$$
$$\alpha_1 = a + b + e, \quad \alpha_2 = b + d + f, \quad \alpha_3 = a + c + f, \quad \alpha_4 = c + d + e,$$
$$\therefore \begin{Bmatrix} e_4 & b_4 & a_4 \\ f_4 & c_4 & d_4 \end{Bmatrix} \to c_L b_L a_L; e_L d_L.$$

Example 3.3:

$$\begin{Bmatrix} d_1 & e_1 & c_1 \\ a_1 & f_1 & b_1 \end{Bmatrix} = \begin{Bmatrix} d & \frac{c+e+f-b}{2} & \frac{c+e+b-f}{2} \\ a & \frac{e+b+f-c}{2} & \frac{c+b+f-e}{2} \end{Bmatrix},$$

$$\beta_1 = a + d + e + f, \quad \beta_2 = a + d + b + c, \quad \beta_3 = e + f + b + c,$$
$$\alpha_1 = d + c + e, \quad \alpha_2 = a + e + b, \quad \alpha_3 = d + b + f, \quad \alpha_4 = a + c + f,$$
$$\therefore \begin{Bmatrix} d_1 & e_1 & c_1 \\ a_1 & f_1 & b_1 \end{Bmatrix} \to a_L c_L b_L; d_L e_L.$$

Example 3.4:

$$\begin{Bmatrix} e_5 & c_5 & d_5 \\ f_5 & b_5 & a_5 \end{Bmatrix} = \begin{Bmatrix} \frac{c+d+a-b}{2} & \frac{e+d+f-a}{2} & \frac{e+c+b-f}{2} \\ \frac{b+d+a-c}{2} & \frac{e+f+a-d}{2} & \frac{f+b+c-e}{2} \end{Bmatrix},$$

$$\beta_1 = a+d+e+f, \quad \beta_2 = a+d+b+c, \quad \beta_3 = e+f+b+c,$$
$$\alpha_1 = c+d+e, \quad \alpha_2 = a+b+e, \quad \alpha_3 = a+c+f, \quad \alpha_4 = b+d+f,$$

$$\therefore \begin{Bmatrix} e_5 & c_5 & d_5 \\ f_5 & b_5 & a_5 \end{Bmatrix} \to a_L b_L c_L; d_L e_L.$$

Example 3.5:

$$\begin{Bmatrix} d_2 & c_2 & e_2 \\ a_2 & b_2 & f_2 \end{Bmatrix} = \begin{Bmatrix} \frac{d+e+f-a}{2} & c & \frac{d+e+a-f}{2} \\ \frac{e+a+f-d}{2} & b & \frac{a+d+f-e}{2} \end{Bmatrix},$$

$$\beta_1 = e+f+b+c, \quad \beta_2 = e+f+a+d, \quad \beta_3 = b+c+a+d,$$
$$\alpha_1 = c+d+e, \quad \alpha_2 = e+a+b, \quad \alpha_3 = d+f+b, \quad \alpha_4 = c+a+f,$$

$$\therefore \begin{Bmatrix} d_2 & c_2 & e_2 \\ a_2 & b_2 & f_2 \end{Bmatrix} \to a_L c_L b_L; e_L d_L.$$

Example 3.6:

$$\begin{Bmatrix} c_3 & f_3 & a_3 \\ b_3 & e_3 & d_3 \end{Bmatrix} = \begin{Bmatrix} \frac{c+f+e-b}{2} & \frac{c+f+b-e}{2} & a \\ \frac{f+b+e-c}{2} & \frac{c+b+e-f}{2} & d \end{Bmatrix},$$

$$\beta_1 = e+f+b+c, \quad \beta_2 = e+f+a+d, \quad \beta_3 = b+c+a+d,$$
$$\alpha_1 = a+c+f, \quad \alpha_2 = b+e+a, \quad \alpha_3 = c+d+e, \quad \alpha_4 = d+b+f,$$

$$\therefore \begin{Bmatrix} c_3 & f_3 & a_3 \\ b_3 & e_3 & d_3 \end{Bmatrix} \to b_L a_L c_L; e_L d_L.$$

We have checked this procedure for all the 144 symmetries of the 6*j* coefficient. We find that this procedure is exactly equivalent to calculating the canonical parameters corresponding to the possible expressions for n, after listing the 12 sets of symmetries.

The set of 12 symmetries described by $n = c + d - e$ and their mapping on to the 12-element group of permutations of canonical parameters are given in Table IV. As examples, the 12 symmetries each describe by $n = a + f - c$ and $n = b + f - d$, and their mapping on to the 12-element group of permutations of canonical parameters corresponding to $n = c + d - e$ are given in Tables V and VI, respectively. The mapping is obtained by calculating the canonical parameters (listed in Appendx A) corresponding to $n = a + f - c$ and $n = b + f - d$, respectively. This is exactly equivalent to the procedure of obtaining the mapping by writing the intermediate parameters corresponding to each symmetry.[7]

ACKNOWLEDGMENTS

The author is grateful to Dr. G. Ramachandran for his kind encouragement and to Professor B. Sanjeevaiah for providing facilities for research. The author thanks the CSIR (India) for the award of the Senior Research Fellowship. The author is also grateful to the referee for his clarification regarding Lockwood's homomorphisms.

APPENDIX A

The sets of canonical parameters of the 3*j* and 6*j* coefficients.

TABLE VII. The nine possible sets of canonical parameters of the 3j coefficient.

n	a_L	b_L	c_L	d_L
$j_1+j_2-j_3$	$j_3-j_2+m_1$	$j_3-j_1-m_2$	$j_3-j_2-m_1$	$j_3-j_1+m_2$
j_1-m_1	$j_3-j_2+m_1$	$j_3-j_1-m_2$	$j_2-j_3+m_1$	$j_2-j_1-m_3$
j_2+m_2	$j_3-j_2+m_1$	$j_3-j_1-m_2$	$j_1-j_3-m_2$	$j_1-j_2+m_3$
j_1+m_1	$j_2-j_3-m_1$	$j_2-j_1+m_3$	$j_3-j_2-m_1$	$j_3-j_1+m_2$
$j_1+j_3-j_2$	$j_2-j_3-m_1$	$j_2-j_1+m_3$	$j_2-j_3+m_1$	$j_2-j_1-m_3$
j_3-m_3	$j_2-j_3-m_1$	$j_2-j_1+m_3$	$j_1-j_3-m_2$	$j_1-j_2+m_3$
j_2-m_2	$j_1-j_3+m_2$	$j_1-j_2-m_3$	$j_3-j_2-m_1$	$j_3-j_1+m_2$
j_3+m_3	$j_1-j_3+m_2$	$j_1-j_2-m_3$	$j_2-j_3+m_1$	$j_2-j_1-m_3$
$j_2+j_3-j_1$	$j_1-j_3+m_2$	$j_1-j_2-m_3$	$j_1-j_3-m_2$	$j_1-j_2+m_3$

TABLE VIII. The possible sets of canonical parameters of the 6j coefficient.

n	a_L	b_L	c_L	d_L	e_L
$c+d-e$	$a+b-c-d$	$b+e-c-f$	$a+e-d-f$	$e+f-b-c$	$e+f-a-d$
$a+b-e$	$c+d-a-b$	$d+e-a-f$	$c+e-b-f$	$e+f-b-c$	$e+f-a-d$
$b+d-f$	$c+f-b-e$	$a+f-d-e$	$a+c-b-d$	$e+f-b-c$	$e+f-a-d$
$a+c-f$	$d+f-a-e$	$b+f-c-e$	$b+d-a-c$	$e+f-b-c$	$e+f-a-d$
$d+f-b$	$a+b-c-d$	$b+e-c-f$	$a+e-d-f$	$b+c-e-f$	$b+c-a-d$
$a+f-c$	$c+d-a-b$	$d+e-a-f$	$c+e-b-f$	$b+c-e-f$	$b+c-a-d$
$d+e-c$	$c+f-b-e$	$a+f-d-e$	$a+c-b-d$	$b+c-e-f$	$b+c-a-d$
$a+e-b$	$d+f-a-e$	$b+f-c-e$	$b+d-a-c$	$b+c-e-f$	$b+c-a-d$
$c+f-a$	$a+b-c-d$	$b+e-c-f$	$a+e-d-f$	$a+d-e-f$	$a+d-b-c$
$b+f-d$	$c+d-a-b$	$d+e-a-f$	$c+e-b-f$	$a+d-e-f$	$a+d-b-c$
$b+e-a$	$c+f-b-e$	$a+f-d-e$	$a+c-b-d$	$a+d-e-f$	$a+d-b-c$
$c+e-d$	$d+f-a-e$	$b+f-c-e$	$b+d-a-c$	$a+d-e-f$	$a+d-b-c$

[1]K. Venkatesh, J. Math. Phys. **19**, 2060 (1978).
[2]K. Venkatesh, J. Math. Phys. **19**, 1973 (1978).
[3]Loren A. Lockwood, J. Math. Phys. **17**, 1671 (1976).
[4]Loren A. Lockwood, J. Math. Phys. **18**, 45 (1977).

[5]The canonical parameters of Lockwood are indicated by the subscript L.
[6]The subscript in the parameters indicate that the corresponding Regge symmetry is superposed on the tetrahedral symmetry. See T. Regge, Nuovo Cimento **11**, 116 (1959). For example,

$$\begin{Bmatrix} a_5 & e_5 & b_5 \\ d_5 & f_5 & c_5 \end{Bmatrix}$$

means that the Regge symmetry R5 is superposed on the tetrahedral symmetry

$$\begin{Bmatrix} a & e & b \\ d & f & c \end{Bmatrix}.$$

[7]The complete listing of summetry relations and homomorphisms have been given in K. Venkatesh, Ph.D. thesis, Mysore University, 1979, unpublished.

Wigner–Eckart theorem for an arbitrary group or Lie algebra

Vishnu K. Agrawala
Department of Natural Sciences and Technology, Point Park College, Pittsburgh, Pennsylvania 15222

(Received 15 June 1979; accepted for publication 21 August 1979)

Circumstances sufficient for the validity of the Wigner–Eckart theorem are analyzed and the theorem is proved under very mild assumptions. The proof is purely algebraic and shows the theorem to be a direct consequence of Schur's lemma and complete reducibility of tensor product of irreducible representations.

1. INTRODUCTION

The ubiquitous Wigner–Eckart theorem[1,2] of angular momentum theory states that the matrix elements of a tensor operator are proportional to the corresponding Clebsch–Gordan coefficients. There have been several successful attemptsnetic [2,3] to generalize the theorem and extend it to other contexts, such as unitary groups[4–7] higher than SU(2), magnetic symmetry groups,[8] finite groups,[9] compact groups, [10]and unitary representations of locally compact groups. [11]The given proofs depend on the special properties of the operators (e.g., unitarity) or of the groups (e.g., compactness) that are limited to the particular case on hand. As a result the given proofs differ from case to case. Here a simple proof is presented which, perhaps because of its abstract nature, makes transparent which hypotheses are really necessary, and enables one to extend the scope of the theorem as fully as possible.

The statement of the theorem, of course, needs the existence of Clebsch–Gordan coefficients or the operators which they represent. This is insured by assuming that the tensor product of two irreducible modules (at least the two that are being considered) should be completely reducible as a direct sum of irreducible submodules. Of course, this requirement can be proved in a wide variety of particular cases—finite groups, compact groups, finite dimensional representations of semisimple Lie groups, or Lie algebras.[12]

It turns out that just one more assumption is sufficient for proving the Wigner–Eckart theorem—that in an irreducible module over a group, a linear operator which commutes with every element of the group is a multiple of the unit operator. This fact, often known as the second part of Schur's lemma, is easy to prove if we restrict ourselves to finite dimensional modules over algebraically closed fields.

2. MODULES AND TENSOR OPERATORS

The proofs of the Wigner–Eckart theorem are almost identical for groups and Lie algebras, so we shall simultaneously consider representations of a group G or a Lie algebra L. It is not necessary for G and L to be related in any manner. Our treatment will be algebraic for most part and no topology is necessary for G. Let V be a vector space of arbitrary dimension over a field F and let $\mathrm{lin}(V)$ denote the set of all F-linear operators on V and $\mathrm{Aut}(V)$ the set of all invertible ones. For more generality we shall allow our representations to be projective. Thus, by a representation of G in V we mean a function ρ: $G \to \mathrm{Aut}(V)$ that satisfies the property

$$\rho(g_1 g_2) = \alpha(g_1, g_2)\rho(g_1)\rho(g_2), \quad \rho(e) = 1_V \tag{2.1}$$

for all $g_1, g_2 \in G$ and the identity element e of G. Here, α is a function from $G \times G$ into the set F^* of nonzero scalars.

Similarly, a representation of the Lie algebra L is a map ρ: $L \to \mathrm{lin}(V)$ such that

$$\rho([l_1, l_2]) = \beta(l_1, l_2)[\rho(l_1)\rho(l_2) - \rho(l_2)\rho(l_1)]$$

for all $l_1, l_2 \in L$. Here, [,] denotes the Lie product in L and β again is a function from $L \times L$ into F^*.

Given a representation ρ_V of $G(L)$ in V, we can regard V as a module (in the projective sense) over $G(L)$ by defining the action as

$$g \cdot v = \rho_V(g)(v) \quad [l \cdot v = \rho_V(l)(v)] \tag{2.2}$$

for all $g \in G$, $l \in L$, and $v \in V$. Conversely, given a module V over $G(L)$, we obtain a representation of $G(L)$ by reading Eq. (2.2) backwards. Often, for the sake of neat notation, we shall simply write gv instead of $\rho_V(g)v$. This is easily distinguished from the module action because the latter is written $g \cdot v$.

A *submodule* $U \subset V$ is an invariant subspace, i.e., $GU \subset U$ ($LU \subset U$). A module V is *irreducible* if only submodules are $\{0\}$ and V itself. We shall call a module *completely reducible* if it is a direct sum of irreducible submodules.

Given two modules U and V, their vector space tensor product $U \otimes V$ and the space $\mathrm{lin}(U, V)$ of all F-linear maps from U to V can be regarded as modules by means of the equations

$$g \cdot (u \otimes v) = gu \otimes gv \quad [l(u \otimes v) = lu \otimes v + u \otimes lv], \tag{2.3}$$

$$(g \cdot t)u = g[t(g^{-1}u)] \quad [(lt)u = l(tu) - t(lu)] \tag{2.4}$$

for all $u \in U$, $t \in \mathrm{lin}(U, V)$, and $g \in G$ ($l \in L$).

According to the convention which we mentioned before, g, g^{-1}, and l on the right hand side of these equations stand for the operators representing them in the appropriate spaces. Especially, $g^{-1}u$ means $\rho_U(g)^{-1}u$ which may not be the same as $\rho_U(g^{-1})$ for projective representations.

The field F is a module under the trivial action by G or L, i.e., $g \cdot \alpha = \alpha$ ($l \cdot \alpha = 0$) for all $\alpha \in F$ and $g \in G$ ($l \in L$).

Of particular importance is the *scalar operator submodule* $\mathrm{hom}(U, V)$ of $\mathrm{lin}(U, V)$. This consists of the so called *scalar operators* $s \in \mathrm{lin}(U, V)$ which satisfy $g \cdot s = s$ ($l \cdot s = 0$) for each $g \in G$ ($l \in L$). More explicitly,

$$(g \cdot s)u = g[s(g^{-1}u)] = su \quad [(l \cdot s)u = l(su) - s(lu) = 0],$$

 0022-2488/80/071562-04$1.00

for every $u \in U$. Thus, scalar operators are the ones which intertwine with all elements of G or L; therefore, they are module homomorphisms.

Schur's lemma, part 1, states that if U and V are irreducible modules and $s \in \hom(U,V)$, then either s is invertible or else $s = 0$ (when U and V are nonisomorphic). This is a direct consequence of the definitions because if $s \neq 0$, then $\text{Ker}\, s \neq U$ and $\text{Im}\, s \neq \{0\}$, which by irreducibility of U and V implies $\text{Ker}\, s = \{0\}$ and $\text{Im}\, s = V$.

The second part of Schur's lemma is not so trivial and states that, under certain conditions, equality holds in the obvious inclusion $F 1_W \subset \hom(W,W)$, where W is a irreducible module. The assumptions under which this is valid will be analyzed in Sec. 5.

3. THE WIGNER–ECKART THEOREM

As before, let G denote a group and L a Lie algebra over a field K. Let U and V be G or L modules with V irreducible and let T be a tensor operator submodule of $\text{lin}(U,V)$. We assume that $T \otimes U$ is a direct sum of irreducible modules (some of which may be isomorphic). Also, it is assumed that for every irreducible module W, $\hom(W,W) = F 1_W$. In most applications, T and U will be irreducible but this assumption is not being made here

The nucleus of the proof involves defining a scalar operator $\phi \in \hom(T \otimes U, V)$ by means of the equation

$$\phi(t \otimes u) = tu \tag{3.1}$$

for each $t \in T$, $u \in U$, with extension to all of $U \otimes V$ by linearity. The following algebra shows that ϕ is a scalar operator. For each $t \in T$, $u \in U$, and $g \in G$, we have [from Eqs. (2.3) and (2.4)]

$$\begin{aligned}(g\cdot\phi)(t \otimes u) &= g[\phi g^{-1}(t \otimes u)] \\ &= g[\phi(g^{-1}t \otimes g^{-1}u)] \\ &= g[(g^{-1}t)(g^{-1}u)] \\ &= g\{g^{-1}[tg(g^{-1}u)]\} \\ &= tu = \phi(t \otimes u).\end{aligned}$$

In the case of L modules, we have, for each $l \in L$,

$$\begin{aligned}(l\cdot\phi)(t \otimes u) &= l[\phi(t \otimes u)] - \phi[l(t \otimes u)] \\ &= l(tu) - \phi(lt \otimes u + t \otimes lu) \\ &= l(tu) - \phi(lt \otimes u) - \phi(t \otimes lu) \\ &= l(tu) - (lt)u - t(lu) \\ &= l(tu) - [l(tu) - t(lu)] - t(lu) = 0.\end{aligned}$$

Now let

$$T \otimes U = \bigoplus_{i \in I} W_i$$

be a direct sum decomposition of $T \otimes U$ into irreducible submodules. We assume the indexing to be one to one. Let J be the set of all indices such that each W_j ($j \in J$) is a copy of V. We allow J to be empty. Therefore,

$$T \otimes U = \bigoplus_{j \in J} W_j \oplus \bigoplus_{i \in I - J} W_i. \tag{3.2}$$

Let $p_i: T \otimes U \to W_i$ and $q_i: W_i \to T \otimes U$ be the canonical projections and injections,[13] respectively, associated with the above direct sum decomposition. It is easily verified that these are module homomorphisms. An element $t \otimes u$ in $T \otimes U$ can be written as

$$t \otimes u = \sum_{i \in F} q_i p_i (t \otimes u),$$

where F is a finite subset of I whose cardinality depends on $t \otimes u$. Therefore, we can write

$$\phi(t \otimes u) = \sum_{i \in F} \phi q_i p_i (t \otimes u).$$

Since ϕ and the q_i, $i \in I$, are module homomorphisms, so are $\phi q_i \in \hom(W_i, V)$ for each $i \in I$. By Schur's lemma, ϕq_i is nonzero only when W_i is isomorphic to V, i.e., only when $i \in J$. Letting $f_j: W_j \to V$ denote any isomorphism of modules for each $j \in J$, the above sum can be written as [since the second sum in Eq. (3.2) gives zero]

$$\begin{aligned}\phi(t \otimes u) &= \sum_{j \in J \cap F} \phi q_j p_j (t \otimes u) \\ &= \sum_{j \in J \cap F} (\phi q_j f_j^{-1}) f_j p_j (t \otimes u).\end{aligned}$$

If J is empty, i.e., if $T \otimes U$ has no submodule isomorphic to V, the $\phi q_i = 0$ for all $i \in I$ and hence $\phi = 0$. Therefore, we shall follow the convention that a sum over an empty indexing set is zero.

Now, for each $j \in J$, $\phi q_j f_j^{-1}$ belongs to $\hom(V,V)$ and hence, by Schur's lemma, part 2, it is a multiple γ_j of the unit operator 1_V so that

$$tu = \phi(t \otimes u) = \sum_{j \in J \cap F} \gamma_j f_j p_j (t \otimes u). \tag{3.3}$$

Equation (3.3) is the basis independent version of the Wigner–Eckart theorem stating that the action of $t \in T$ on $u \in U$ is a linear combination of the projection maps $f_j p_j : T \otimes U \to V$. The scalars γ_j are called the *reduced matrix elements* and are independent of the particular vectors t and u. The number of nonzero reduced matrix elements quite naturally depends on t and u because this is the number of terms on the right hand side of Eq. (3.3), namely, the cardinality of $J \cap F$.

The conventional statement of the Wigner–Eckart theorem is written in terms of matrix elements. For that purpose one has to choose bases $\{u_a | a \in A\}$, $\{v_b | b \in B\}$, $\{t_c | c \in C\}$, and $\{w_{jd} | d \in B\}$ in the modules U, V, T, and each W_j ($j \in J$). Then, $\{t_c \otimes u_a | c \in C, a \in A\}$ is a basis of $T \otimes U$. From Eq. (3.3), the (b,a) matrix element of t can be written as

$$\langle b | t | a \rangle = \sum_{j \in J \cap F} \gamma_j \langle b | f_j p_j | c,a \rangle. \tag{3.4}$$

The matrix elements $\langle b | f_j p_j | c,a \rangle$ are simply the Clebsch–Gordan coefficients.

4. APPLICATION TO REDUCIBLE MODULES

Sometimes it may be necessary to find matrix elements of operators between reducible modules (e.g., multiparticle states[14]). The Wigner–Eckart theorem could be used in such cases by first resolving the operator into its operator matrix[15] as shown below. The general formula will be notationally

very cumbersome, so we prefer to present the essential features of the algorithm only. All modules in this section are finite dimensional over an algebraically closed field and completely reducible.

Consider an operator $t \in \text{lin}(U,V)$ where U and V are modules (over G or L). Let

$$U = \underset{m\in M}{\oplus} U_m,$$

$$V = \underset{n\in N}{\oplus} V_n$$

be their direct sum decomposition with $P_m (m\in M)$ and $P'_n (n\in N)$ denoting the natural projections and $Q_m (m\in M)$ and $Q'_n (n\in N)$ denoting the natural injections. The *operator matrix* of t is just an array whose (nm) entry is the linear map t_{nm} from U_m to V_n defined by

$$t_{nm} = P'_n t Q_m.$$

Since the Q_m's are just inclusion maps, to evaluate $t_{nm}(u_m)$ for any $u_m \in U_m$, one simply evaluates $t(u_m)$ and projects it onto its nth component in V_n. Also, it follows easily from the properties of natural projections and injections[13] that

$$t = \sum_{\substack{m\in M \\ n\in N}} Q'_n t_{nm} P_m.$$

Any vector $u\in U$ can be written uniquely as $u = \Sigma_{m\in M} u_m$, where $u_m = P_m(u)$. Thus, tu is known if tu_m is known for each $m\in M$. Moreover, tu_m is known if each of its components $P'_n tu_m$ is known. However, from the above resolution of t, $P'_n tu_m = t_{nm} u_m$. In this sense, (t_{nm}) is the matrix of $t - t_{nm} u_m$ acts on the mth component of u, and gives the nth component of the result, whereas $\Sigma_{m\in M} t_{nm} u_m$ is the nth component of tu. Thus, all information about t is contained in its matrix (t_{nm}). Each "matrix element" t_{nm} belongs to $\text{lin}(U_m, V_n)$. Since we have not made any assumption as to the irreducible nature of t, therefore t_{nm} need not belong to an irreducible submodule of $\text{lin}(U_m, V_n)$ and it is necessary to resolve it as a sum of irreducible tensor operators in $\text{lin}(U_m, V_n)$. This is achieved by taking the direct sum decomposition of $\text{lin}(U_m, V_n) \cong V_n \otimes U_m^* = \oplus_{r\in R} T^r_{mn}$ with the associated projections P''_r and injections Q''_r. The rth projection of t_{nm} is the tensor operator $t^r_{nm} = P''_r(t_{nm})$ belonging to the irreducible submodule T^r_{mn} of $\text{lin}(U_m, V_n)$. The Wigner–Eckart theorem can be applied to t^r_{nm} to find its matrix elements, knowing which one can find matrix elements of t_{nm} and ultimately those of t.

5. ANALYSIS OF ASSUMPTIONS

We now investigate conditions which enable us to prove the two assumptions used earlier for the Wigner–Eckart theorem. The first—Schur's lemma, part 2—can be proved under fairly mild conditions. To discuss them we introduce the following notation and definitions.

Let G denote a group and L a Lie algebra over some field K. Let W be a module over G (or L) which is a vector space over a field F and ρ the corresponding representation of G (or L) in W. It is more convenient to work with the associative algebra A generated by 1_W and $\rho(G)$ [or $\rho(L)$] over F. Clearly, W is an A module and a subspace of W is invariant under G (or L) if and only if it is invariant under A. Thus, W is irreducible for G (or L) if and only if it is irreducible for A. Also, an operator in $\text{lin}(W,W)$ intertwines with elements of $\rho(G)$ [or $\rho(L)$] if and only if it does so with elements of A.

The algebra A is said to be n-fold transitive if given $k \leqslant n$ linearly independent (over F) vectors $x_1,...,x_k$ and any k vectors $y_1,...,y_k$, there exists an element $a\in A$ such that $ax_j = y_j$, $j = 1,...,k$. Thus, A is onefold transitive if and only if W is irreducible.[16] The algebra A is called *dense* if it is n-fold transitive for every positive integer n.

In the course of proving his density theorem, Jacobson[17] has shown that if A is twofold transitive then $\text{hom}(W,W) = F1_W$. Of course, twofold transitivity implies onefold transitivity. Conversely, if A is onefold transitive and $\text{hom}(W,W) = F1_W$, then A is dense and hence certainly twofold transitive. Thus, for an irreducible A module W, the equality $\text{hom}(W,W) = F1_W$ is equivalent to A being twofold transitive.

To provide a simpler criterion for $\text{hom}(W,W) = F1_W$, let us define the spectrum $\sigma(s)$ of $s\in\text{hom}(W,W)$ as the set of scalars $\lambda\in F$ for which $s - \lambda 1_W$ is not invertible. A convenient criterion for $\text{hom}(W,W) = F1_W$ is that $\sigma(s)$ be nonempty for each $s\in\text{hom}(W,W)$; for, if $s = \lambda_s 1_W$, $\lambda_s \in F$, then obviously $\lambda_s \in \sigma(s)$. Conversely, suppose $\lambda_s \in \sigma(s)$. Then, $s - \lambda_s 1_W$ in $\text{hom}(W,W)$ is not invertible and hence, by the first part of Schur's lemma, it must be zero, i.e., $s = \lambda_s 1_W$. Note also that λ_s is the unique member of $\sigma(s)$ in this case.

If W is finite dimensional and F is algebraically closed, the characteristic equation $\det(s - \lambda 1_W) = 0$ has solution in F and so the spectrum of s is nonempty. In the infinite dimensional case it helps to know if s is algebraic because then $\sigma(s)$ coincides with the roots of the minimum polynomial of s which factors into linear factors if F is algebraically closed.

It is easily shown[18] that if the cardinality of F is strictly larger than the F dimension of W, then every $s\in\text{hom}(W,W)$ is algebraic. For instance, if F is the complex field and $\dim_F W$ is countable, then every $s\in\text{hom}(W,W)$ is a scalar multiple of the unit operator. If W is an L module, where L is a finite dimensional Lie algebra over F, then even the cardinality condition on F is not needed for showing that s is algebraic.[19]

Assumptions of a topological nature can sometimes ensure the nonemptiness of the spectrum. For instance, it is a standard result[20] that every bounded linear operator in a normed vector space over the complex field C has a nonempty spectrum. In fact, if W is a complex Hilbert space and the algebra A is self-adjoint ($a\in A \Rightarrow a^* \in A$), then even the weaker type of topological irreducibility (W has no *closed* invariant subspaces except $\{0\}$ and W) is equivalent to $\text{hom}(W,W) = C1_W$.[21]

The second assumption—complete reducibility of modules—needs more stringent conditions. The classical results in this area deal with only finite dimensional modules over real or complex fields. Every finite dimensional real or complex module is completely reducible under any one of the following conditions[12]: (1) G is a compact group, (2) L is the Lie algebra of a compact Lie group, (3) G is a semisimple

Lie group, or (4) L is semisimple.

The above conditions do not guarantee complete reducibility of infinite dimensional modules in general,[22] but Kunze[23] has shown that every square integrable unitary representation of a locally compact group G in a Hilbert space is a Hilbertian direct sum of topologically irreducible square integrable representations. Recall that square integrable representations (known also as discrete series) are those whose matrix elements are all square integrable (with respect to the Haar measure of G).

For the Wigner–Eckart theorem we do not actually need complete reducibility of *every* module but just that the tensor product of two completely reducible (usually irreducible) modules be completely reducible. This holds if the modules in question are L modules but semisimplicity of L is not required.[24]

A fairly satisfactory theory of infinite dimensional representations exists provided one restricts oneself to modules that are complex Hilbert spaces.[16,22,25,26] We shall follow the treatment of Ref. 16 as it uses minimum amount of topology, especially avoiding the use of direct integrals, and does not restrict to unitary representations. No topology is necessary for G but it is necessary that the operators $\rho(g)$ [or $\rho(L)$] in $\mathrm{lin}(W,W)$ representing elements of G (or L) be bounded ($\equiv$ continuous). It is also convenient to require that the algebra A generated by $\rho(G)$ [or $\rho(L)$] or at least its weak closure $\bar{A}$ be self-adjoint. This is not too restrictive because it is fulfilled by most practical representations, e.g., unitary and projective unitary representations.

The set of bounded linear maps from a Hilbert module U to another one V will be denoted by $BL(U,V)$ and its subset consisting of bounded intertwining operators or bounded module homomorphisms by $BH(U,V)$. Isomorphisms of modules now will be required to be isometries, i.e., preserving the inner products. Two modules are said to be *disjoint* if neither contains a submodule isomorphic to the other. The tensor product is taken in the sense of Hilbert space tensor product which itself is a Hilbert space. The direct sum of modules now means the Hilbertian direct sum which allows countable convergent sums instead of finite sums.

Under these circumstances, the following Mackey's[26] version of Schur's lemma can be proved[16]:

(i) Let U and V be G modules. Then U and V are disjoint if and only if $BH(U,V) = \{0\}$.

(ii) A G module W is topologically irreducible if and only if $BH(W,W) = \mathbb{C}1_W$.

To prove the Wigner–Eckart theorem, we consider as before G modules U and V, where V is topologically irreducible, and a tensor operator submodule $T \in BL(U,V)$. We define $\phi \in BL(T \otimes U, V)$ by Eq. (3.1). It is easily verified that ϕ is bounded.

Now, $T \otimes U$ cannot in general be written as a direct sum of irreducibles but according to Theorem IV[Eq. (5.4)] of Ref. 16, it can be decomposed as

$$T \otimes U = H + H^{\perp},$$

where H is a uniquely determined submodule that is a Hilbert space direct sum of topologically irreducible submodules each of which is a copy of V, and $H^{\perp}$ is the orthogonal complement of H, also a submodule. If $T \otimes U$ has no submodule isomorphic to V, then $tu = \phi(t \otimes u) = 0$ and $H = \{0\}$.

Assuming $H \neq \{0\}$, let $P: T \otimes U \to H$, $P^{\perp}: T \otimes U \to H^{\perp}$ be canonical projections and $Q: H \to T \otimes U$, $Q^{\perp}: H^{\perp} \to T \otimes U$ the canonical injections. Then

$$\begin{aligned} tu = \phi(t \otimes u) &= \phi(QP + Q^{\perp}P^{\perp})(t \otimes u) \\ &= \phi Q P(t \otimes u), \end{aligned}$$

because the restriction $\phi Q^{\perp} \in BH(H^{\perp}, V) = 0$ since $H^{\perp}$ is disjoint from V.

Thus, the action of ϕ on $T \otimes U$ is the same as the action of its restriction ϕQ to $H = P(T \otimes U)$. This is similar to the situation of Sec. 3 with ϕQ replacing ϕ and H replacing $T \otimes U$. Letting

$$H = \bigoplus_{j \in J} W_j$$

and proceeding as in Sec. 3 with the same notation as used there, we obtain the following form of the Wigner–Eckart theorem:

$$tu = \sum_{j \in N} \gamma_j f_j p_j [P(t \otimes u)],$$

where N is a countable subset of J. The matrix version easily follows from this as before.

ACKNOWLEDGMENT

The author wishes to thank Professor C.Y. Chao for helpful discussions.

[1] C. Eckart, Rev. Mod. Phys. **2**, 305 (1930).
[2] E. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[3] A. Speiser, *Die Theorie der Gruppen* (Springer, Berlin, 1937).
[4] G.F. Koster, Phys. Rev. **109**, 227 (1958).
[5] B. Diu, Nuovo Cimento **28**, 466 (1963).
[6] J. Ginibre, J. Math. Phys. **4**, 720 (1963).
[7] G.E. Baird and L.C. Biedenharn, J. Math. Phys. **5**, 1730 (1964).
[8] A. Aviron and J. Zak, J. Math. Phys. **9**, 2138 (1968).
[9] P. Rudra, J. Math. Phys. **7**, 935 (1966).
[10] A.U. Klimyk, Rep. Math. Phys. **7**, 153 (1975).
[11] A.U. Klimyk, Theor. Math. Phys. **8**, 668 (1971).
[12] H. Freudenthal and H. de Vries, *Linear Lie groups*, (Academic, New York, 1969), p. 185, Theorem 35.4.
[13] S. MacLane and G. Birkoff, *Algebra* (Macmillan, New York, 1968), p. 210.
[14] J.S. Briggs, J. Math. Phys. **11**, 1198 (1970).
[15] Reference 13, p. 264.
[16] S.A. Gaal, *Linear Analysis and Representation Theory* (Springer, New York, 1973), p. 95.
[17] N. Jacobson, *Structure of Rings* (American Mathematical Society, Providence, 1968), p. 32.
[18] Reference 17, p. 20.
[19] D. Quillen, Proc. Am. Math. Soc. **21**, 171 (1969).
[20] W. Rudin, *Functional Analysis* (McGraw-Hill, New York, 1973), p. 235.
[21] Reference 16, p. 151.
[22] G. Warner, *Harmonic Analysis on Semi-simple Lie groups* (Springer, New York, 1972), p. 458.
[23] R.A. Kunze, J. Funct. Anal. **6**, 454 (1970).
[24] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962), p. 83.
[25] J. Dixmier, *C* Algebras* (North Holland, New York, 1977).
[26] G.W. Mackey, *The Theory of Group Representations* (University of Chicago, Chicago, 1955).

On Gel'fand states of representations of U(n) and the Gel'fand lattice polynomials

Christopher J. Henrich
Perkin–Elmer Computer Systems Division, 106 Apple Street, Tinton Falls, New Jersey 07724

(Received 23 January 1980; accepted for publication 22 February 1980)

Three theorems are proved concerning the Gel'fand lattice polynomials (GLP) introduced by Gazeaux, Dumont-Lepage, and Ronveaux. In these theorems, subdeterminant initial conditions are assumed. (1) The Gel'fand states are linear combinations of GLP corresponding to left Gel'fand lattices. (2) The set of all GLP is a basis for the space of polynomial functions of the elements of a matrix. (3) If the left Gel'fand lattices are arranged in a natural order, then each Gel'fand state is a linear combination of the corresponding GLP and earlier ones in the same order. Finally, a formula is given for computing individual GLP.

1. INTRODUCTION

The irreducible unitary representations of the unitary groups have been studied from many viewpoints. Weyl[1,2] classified them up to equivalence, and constructed a model of each equivalence class, in which he defined a basis that is generally not orthonormal. Using the branching law that Weyl discovered, Gel'fand and Zeitlin[3] defined orthonormal bases; Gel'fand and Graev[4] derived formulas for the matrix elements of the representations with respect to those bases. A general method for constructing irreducible representation spaces for all compact Lie groups was first stated by Godement,[5] was further developed by Zhelobenko,[6] and has since been generalized to the Borel–Weil theorem.[7] Bargmann and Moshinsky,[8,9] and Baird and Biedenharn[10] described the same construction, for the case of U(n), in different notation. An outstanding problem has been the evaluation of the Gel'fand states in these concrete representations, and the further study of the matrix elements of the representations. Louck[11,12] has given a clear exposition of how these problems are related. Several authors[13–15] have found partial results on the explicit form of the Gel'fand states. Nagel and Moshinsky[16] have shown how the states can be constructed by successive applications of certain operators which, in these realizations, are partial differential operators. The results, and the proofs, in all of these works have tended to become quite complicated, and it is hard to see any general pattern amid the details.

Recently, Gazeau, Dumont-Lepage, and Ronveaux,[17] have introduced a series of polynomials which offer some hope of simplification. These are called the Gel'fand Lattice Polynomials (GLP). In their most general form they may be regarded as polynomials on a space of $\binom{2n}{n}-1$ dimensions, related to the exterior algebra over $\mathbf{C}^n$; but a special case, called "subdeteminant initial conditions," defines polynomial functions of the components of an $n\times n$ matrix. The Gel'fand states, and more generally the matrix elements of representations, are such polynomials also; therefore it is reasonable to ask how these sets of polynomials are related. Gazeau, Dumont-Lepage, and Ronveaux made these conjectures:

(1) the Gel'fand states are linear combinations of certain Gel'fand lattice polynomials;

(ii) more generally, all the matrix elements are linear combinations of GLP;

(iii) if the Gel'fand states and the GLP are listed in appropriately chosen orders, then the transformation matrix whose existence is asserted by (i) is triangular. (Thus the Gel'fand states could, in principle, be computed by Gram–Schmidt orthogonalization.).

In Ref. 17 it is proved that (i) holds for $n\leqslant 3$ and (ii) for $n\leqslant 2$; explicit results, for (i) if $n=4$ and (ii) if $n=3$, are stated without complete proof. Conjecture (iii) is consistent with the results of Ref. 17.

The present paper proves these three conjectures. The paper is organized as follows. Sec. 2–5 essentially repeat the definitions of the GLP and related objects from Ref. 17; this is done for ease of reference and also because the conventions chosen in Ref. 17 do not match those of most of the work on Gel'fand states. Section 6 proves the first conjecture and Sec. 7 proves the second conjecture from a special case of the first. In these theorems the essence is the properties of the Gel-'fand lattice polynomials without reference to the Gel'fand states. For the third conjecture, however, we must use some means of constructing the orthogonal Gel'fand states. We choose to adapt the lowering operators of Nagel and Moshinsky to a realization by functions on the group of upper triangular matrices. After discussing in Sec. 8 the convenient orders in which to arrange either Gel'fand states or GLP, in Secs. 9 and 10 we transfer the lowering operators to that group. In Sec. 11 we compute the leading term in an arbitrary Gel'fand state, and thus establish the truth of conjecture (iii).

Finally, we note how individual GLP can be conveniently computed.

2. GEL'FAND LATTICES

A Gel'fand lattice is an arrangement of integers in a certain framework. It is convenient to have some vocabulary for describing the places in the framework. These are an $n\times n$ array of cells; we shall denote a typical cell by $[{}^{j}_{i}]$ where i and j run from 1 to n. We define a partial order on the cells by saying that $[{}^{j}_{i}]\leqslant[{}^{j'}_{i'}]$ if $j\geqslant j'$ and $i\geqslant i'$. We picture the framework as in Fig. 1.

 0022-2488/80/071566-11$1.00

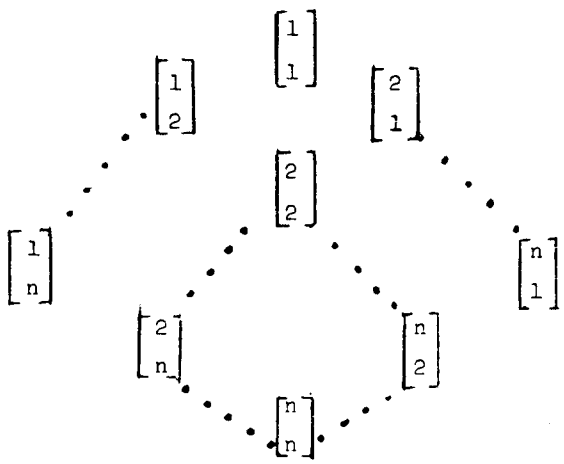


FIG. 1. The framework for Gel'fand lattices.

The *central column* or *median* of the framework is the set of cells $\begin{bmatrix} i \\ i \end{bmatrix}$, $1 \leqslant i \leqslant n$. The k th *left* column, for k between 1 and $n-1$, contains the cells $\begin{bmatrix} j \\ j+k \end{bmatrix}$, $1 \leqslant j \leqslant n-k$; the k th *right* column contains the cells $\begin{bmatrix} i+k \\ i \end{bmatrix}$, $1 \leqslant i \leqslant n-k$. These columns are denoted C_{n-k} and C^{n-k}, respectively; the index is the number of elements in the column. The central column is $C_n = C^n$.

The partial diagonals (in Ref. 17 "quasi diagonals") extend either to the left or to the right, and either up or down, from elements of the central column. A typical upper left partial diagonal contains the elements $\begin{bmatrix} j \\ i \end{bmatrix}$ where $j \leqslant i$ and i is fixed.

A *Gel'fand lattice* m of *dimension* n is an assignment of nonnegative integers $m_i{}^j$ to the cells $\begin{bmatrix} j \\ i \end{bmatrix}$, such that if $\begin{bmatrix} j \\ i \end{bmatrix} \leqslant \begin{bmatrix} j' \\ i' \end{bmatrix}$ then $m_i{}^j \leqslant m_{i'}{}^{j'}$. An equivalent form of this condition is

$$m_i{}^{j+1} \leqslant m_i{}^j \leqslant m_{i-1}{}^j. \tag{2.1}$$

Figure 2 displays a typical Gel'fand lattice. The set of Gel'fand lattices of dimension n is denoted by $\mathscr{G}_n$. The *transpose* ${}^t m$ of a lattice m is defined by

$${}^t m_i{}^j = m_j{}^i.$$

A *binary* Gel'fand lattice is a member $\mathscr{B}$ of $\mathscr{G}_n$ such that every integer $b_i{}^j$ is 0 or 1. The set of binary elements of $\mathscr{G}_n$ is denoted $\mathscr{B}\mathscr{G}_n$. We shall take a special interest in those lattices which have constant values along each lower right partial diagonal; they will be called *left* Gel'fand lattices, and the set of them will be denoted $\mathscr{L}\mathscr{G}_n$.

$\mathscr{G}_n$ is a semigroup, with this addition:

$$(m+m')_i{}^j = m_i{}^j + m'_i{}^j.$$

The zero lattice is the identity element; we shall denote the set of nonzero elements by $\mathscr{G}_n{}^*$.

Among the many possible numbering schemes for the set of binary Gel'fand lattices, there is one that we shall find particularly useful. Let $1 \leqslant l \leqslant n$; let $R = (r_a : 1 \leqslant a \leqslant l)$ and $C = (c_a : 1 \leqslant a \leqslant l)$ be sequences of integers, satisfying

$$1 \leqslant r_a \leqslant n, \quad 1 \leqslant c_a \leqslant n,$$
$$r_1 < r_2 < \cdots < r_l, \quad c_1 < c_2 < \cdots < c_l. \tag{2.2}$$

A binary Gel'fand lattice $b = \beta_C{}^R$ is defined by

$$b_i{}^j = 1, \quad \text{if} \quad j \leqslant l \quad \text{and} \quad i \leqslant n + j - c_j \quad \text{or} \quad i \leqslant l \quad \text{and} \quad j \leqslant n + i - r_i; \quad b_i{}^j = 0 \quad \text{otherwise.} \tag{2.3}$$

The left side of $\beta_C{}^R$ is determined by C, and the right side by R. The condition in the first line of (2.2) may be restated thus: In the lower left partial diagonal starting from $\begin{bmatrix} j \\ j \end{bmatrix}$, $c_j - j$ cells are given the value 0 and the other $n - c_j + 1$ cells are given the value 1.

The transpose of a binary lattice is

$${}^t\beta_C{}^R = \beta_R{}^C. \tag{2.4}$$

3. GEL'FAND LATTICE POLYNOMIALS

Every Gel'fand lattice m can be decomposed into a sum of binary Gel'fand lattices

$$m = b_{(1)} + \cdots + b_{(s)},$$
$$b_{(\alpha)} \in \mathscr{B}\mathscr{G}_n{}^*, \quad s = m_1{}^1, \tag{3.1}$$

usually in more than one way. Let x_b be an indeterminate for every b in $\mathscr{B}\mathscr{G}_n{}^*$. If $m \in \mathscr{G}_n$ let

$$\mathbf{x}^m = \sum x_{b_{(1)}} x_{b_{(2)}} \cdots x_{b_{(s)}}, \tag{3.2}$$

where the sum ranges over all $b_{(1)}, \ldots, b_{(s)}$ in $\mathscr{B}\mathscr{G}_n{}^*$, subject to the constraint (3.1). The *Gel'fand lattice polynomial* (or GLP) associated to m is $\mathbf{x}^m$.

Propositions relating to GLP can sometimes be proved by induction on the number of terms in the sum. For this reason it is convenient to observe [cf. Ref. 17, (17)]

$$\mathbf{x}^m = \sum_{b + \tilde{m} = m} x_b \mathbf{x}^{\tilde{m}},$$

where in the summation $b \in \mathscr{B}\mathscr{G}_n{}^*$ and $\tilde{m} \in \mathscr{G}_n$, $\tilde{m}_1{}^1 = m_1{}^1 - 1$. Equation (3.3) can be regarded as a finite difference equation, the solution of which is completely determined by the values of $\mathbf{x}^b$ for binary lattices b. These values are accordingly called "initial conditions".

If m is a left Gel'fand lattice, then it is easy to see that in any decomposition (3.1) the summands must be left binary Gel'fand lattices.

4. GENERATING FUNCTIONS

The sum in (3.2) ranges over ordered partitions of m. We may describe the same polynomial in terms of unordered

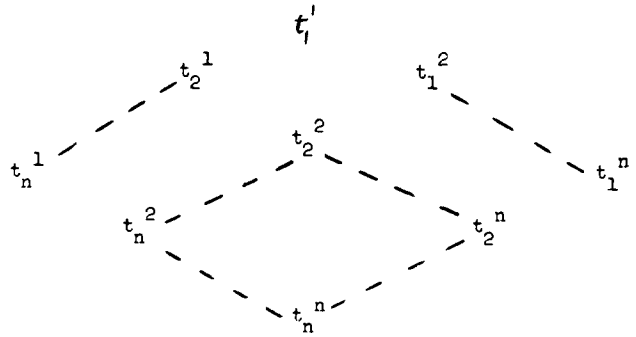


FIG. 2. A Gel'fand Lattice.

partitions of m into distinct terms counted with varying multiplicities. If we have such a partition, say

$$m = \sum_{b \in \mathscr{H}\mathscr{S}_n^*} d_b b, \quad d_b \geqslant 0,$$

then its contribution to (3.2) is

$$(m_1{}^1)! \prod_{b \in \mathscr{H}\mathscr{S}_n^*} (x_b{}^{d_b})/d_b!. \tag{4.1}$$

Now this product has a familiar structure: it is a monomial, $\Pi_b x_b{}^{d_b}$, multiplied by the corresponding multinomial coefficient. One recognizes that $\mathbf{x}^m$ is part of the expansion of $(\Sigma x_b)^{m_1{}^1}$. We shall obtain a generating function for the GLP's if we introduce additional independent variables into this expression, in such a way that it can be separated into individual GLP's.

For this purpose we define a set of $n^2 + n$ indeterminates

$$\lambda_{gh}, \mu_{gh}, \quad 1 \leqslant g \leqslant h \leqslant n. \tag{4.2}$$

If $R = (r_1,...,r_l)$ as in (2.2), we set

$$\boldsymbol{\lambda}_{(R)} = \lambda_{1r_1} \cdots \lambda_{lr_l}, \quad \boldsymbol{\mu}_{(R)} = \mu_{1r_1} \cdots \mu_{lr_l}. \tag{4.3}$$

If $b = \beta_C{}^R$, then the products $\boldsymbol{\lambda}_{(C)}$ and $\boldsymbol{\mu}_{(R)}$ can be characterized in terms of $b_i{}^j$. Indeed, if we make the convention that for any element of $\mathscr{G}_n$, $m_i{}^{n+1} = m_{n+1}{}^j = 0$, then

(i) the exponent of λ_{gh} in $\boldsymbol{\lambda}_{(C)}$ is

$$b^{\,g}_{n-h+g} - b^{\,g}_{n-h+g+1};$$

and

(ii) the exponent of μ_{gh} in $\boldsymbol{\mu}_{(R)}$ is

$$b_g^{\,n-h+g} - b_g^{\,n-h+g+1}.$$

Because the exponents are linear functions of the numbers $b_i{}^j$, we can extend them to all Gel'fand lattices. If $m \in \mathscr{G}_n$, define

$$\langle m, gh \rangle = m^g_{n-h+g} - m^g_{n-h+g+1}, \tag{4.4a}$$

$$\langle gh, m \rangle = m_g^{n-h+g} - m_g^{n-h+g+1}, \tag{4.4b}$$

and

$$\boldsymbol{\lambda}^{\langle m \rangle} = \prod_{g \leqslant h} \lambda_{gh}{}^{\langle m, gh \rangle}, \tag{4.5a}$$

$$\boldsymbol{\mu}^{[m\rangle} = \prod_{g \leqslant h} \mu_{gh}{}^{\langle gh, m \rangle}. \tag{4.5b}$$

Note that

$$\langle m + m', gh \rangle = \langle m, gh \rangle + \langle m', gh \rangle, \tag{4.6a}$$

$$\langle gh, m + m' \rangle = \langle gh, m \rangle + \langle gh, m' \rangle, \tag{4.6b}$$

$$\boldsymbol{\lambda}^{\langle m+m' \rangle} = \boldsymbol{\lambda}^{\langle m \rangle} \boldsymbol{\lambda}^{\langle m' \rangle}, \tag{4.7a}$$

$$\boldsymbol{\mu}^{[m+m'\rangle} = \boldsymbol{\mu}^{[m\rangle} \boldsymbol{\mu}^{[m'\rangle}, \tag{4.7b}$$

and if $b = \beta_C{}^R$, then

$$\boldsymbol{\lambda}^{\langle b \rangle} = \boldsymbol{\lambda}_{(C)}, \tag{4.8a}$$

$$\boldsymbol{\mu}^{[b\rangle} = \boldsymbol{\mu}_{(R)}. \tag{4.8b}$$

Let us set

$$x_b = x_C{}^R, \quad b = \beta_C{}^R;$$

then the generating function is

$$\left[\sum_l \sum_{R,C} \boldsymbol{\lambda}_{(C)} \boldsymbol{\mu}_{(R)} x_C{}^R \right]^S = \sum_{T \in \mathscr{S}\mathscr{S}_n, |T_1| = S} \boldsymbol{\lambda}^{\langle T \rangle} \boldsymbol{\mu}^{[T\rangle} \mathbf{x}^T. \tag{4.9}$$

This generating function is found in Ref. 17, Eq. (54), in slightly different notation, notably, $\lambda_i{}^j$ of that paper is our $\lambda_{j,n-i+1}$ and $\mu_i{}^j = \mu_{i,n-j+1}$. The constraint $\lambda_{ij} = \mu_{ij}$ is imposed in Ref. 17, but it can be omitted without harm.

A generating function for the left GLP can be obtained from (49) by taking special values of the indeterminates μ_{gh}. On the one hand, if T is a left Gel'fand lattice, then the monomial $\boldsymbol{\lambda}^{\langle T \rangle}$ determines T completely. On the other hand, the condition that T be a left Gel'fand lattice is that $\langle gh, T \rangle = 0$ if $g \leqslant h - 1$. We isolate the left GLP in (4.9) if we set $\mu_{gh} = 0$ if $g < h$ and $\mu_{hh} = 1$. We find

$$\left(\sum_l \sum_C \boldsymbol{\lambda}_{(C)} x_C{}^{1\cdots l} \right)^S = \sum_{T \in \mathscr{S}\mathscr{S}_n, |T_1| = S} \boldsymbol{\lambda}^{\langle T \rangle} \mathbf{x}^T. \tag{4.10}$$

4. SUBDETERMINANT INITIAL CONDITIONS

From the definition of the Gel'fand lattice polynomials, it follows that we can assign arbitrary values to the "initial conditions" x^b where b is binary. One interesting assignment is obtained by letting z be an arbitrary $n \times n$ matrix of complex numbers, and letting

$$x_{\beta_C{}^R} = \Delta_C{}^R(\mathbf{z}) = \det \begin{bmatrix} z_{c_1}^{r_1} & \cdots & z_{c_l}^{r_1} \\ \vdots & & \vdots \\ z_{c_1}^{r_l} & \cdots & z_{c_l}^{r_l} \end{bmatrix}. \tag{5.1}$$

We shall then denote

$$\mathbf{x}^{\mathrm{T}} = \Delta^{\,T}(\mathbf{z}). \tag{5.2}$$

In this paper we shall give most of our attention to GLP with either arbitrary independent initial conditions or subdeterminant initial conditions. (Ref. 17 makes interesting use of other choices.) For the sake of brevity we shall sometimes refer to these two cases as "exterior GLP" and "matrix GLP" respectively.

With subdeterminant initial conditions, the left GLP satisfy an important functional equation. Indeed, if $R = (1,...,l)$, then $\Delta_C{}^R$ is a subdeterminant from the first l rows of $\mathbf{z}$; therefore, if $\mathbf{v}$ is a lower triangular matrix with 1's on the diagonal, then

$$\Delta_C{}^R(\mathbf{v}\,\mathbf{z}) = \Delta_C{}^R(\mathbf{z}) \tag{5.3}$$

and if $\boldsymbol{\delta}$ is the diagonal matrix $\mathrm{diag}(\delta_1,...,\delta_n)$, then

$$\Delta_C{}^R(\boldsymbol{\delta}\,\mathbf{z}) = \delta_1 \delta_2 \cdots \delta_l \Delta_C{}^R(\mathbf{z}). \tag{5.4}$$

These equations can be generalized to left Gel'fand lattices that are not binary. It is easily seen that (4.10) and (5.3) imply

$$\Delta^{\,T}(\mathbf{v}\,\mathbf{z}) = \Delta^{\,T}(\mathbf{z}). \tag{5.5}$$

To extend (5.4), we observe that

$$\boldsymbol{\lambda}_C \Delta_C{}^{1\cdots l}(\boldsymbol{\delta}\,\mathbf{z}) = \tilde{\boldsymbol{\lambda}}_C \Delta_C{}^{1\cdots l}(\mathbf{z}),$$

where

$$\tilde{\lambda}_{gh} = \delta_g \lambda_{gh}.$$

It is easily checked that for any Gel'fand latice T,

$$\sum_{h=g}^{n} \langle T, gh \rangle = t_g{}^{g} - t_{g+1}{}^{g+1}$$

and so

$$\tilde{\lambda}^{\langle T \rangle} = \delta^{[t]} \lambda^{\langle T \rangle},$$

where

$$\delta_{[t]} = \delta_1{}^{t_1 - t_2} \ldots \delta_n{}^{t_n}. \tag{5.6}$$

Therefore the generating function implies

$$\Delta^{T}(\delta\, z) = \delta^{[t]} \Delta^{T}(z). \tag{5.7}$$

Now (5.5) and (5.7) are equations that have been studied before; in the work of Baird and Biedenharn,[10] and Bargmann and Moshinsky[8,9] (see also Ref. 15) they are shown to characterize a space $\mathscr{B}^{[t]}$ of polynomial functions of the matrix elements of z. This space is invariant under right multiplication by a matrix u; if we set

$$L_u{}^{[t]} f(z) = f(z\, u), \quad f \in \mathscr{B}^{[t]},$$

then $L^{[t]}$ is a unitary representation of U(n), and a polynomial representation of GL(n). It is the irreducible representation with highest weight $[t] = (t_1{}^1, \ldots, t_n{}^n)$.

At this point the Gel'fand lattice polynomials become relevant to the theory of representations of U(n). The relevance is heightened by the fact that the labeling system for the Gel'fand basis consists of $(n^2 + n)/2$ integers m_{ij}, $1 \leqslant i \leqslant j \leqslant n$, which satisfy

$$m_{i\,j+1} \geqslant m_{ij} \geqslant m_{i+1\,j+1}; \tag{5.8}$$

the representation of U(n) is labelled by $(m_{1n}, \ldots, m_{nn})$. If we set

$$m_{ij} = t^{i}_{n-j+1}$$

then these integers are exactly the left half of a Gel'fand lattice. Therefore the left matrix GLP, which belong to $\mathscr{B}^{[t]}$, enjoy a natural one-to-one correspondence with the Gel'fand basis elements. It is attractive to conjecture that Δ^{T} is proportional to the Gel'fand basis vector; if $n \leqslant 3$, this is true.[10,15,17] For $n \geqslant 4$ it is no longer true, but one can still conjecture that the Gel'fand basis elements are linear combinations of the left matrix GLP. This conjecture amounts to saying that the left matrix GLP belonging to $\mathscr{B}^{[t]}$ are a basis of $\mathscr{B}^{[t]}$. It will be true, if the left matrix GLP are linearly independent.

The relation between Gel'fand lattices and representation theory can be extended. The matrix elements of the irreducible representations with respect to the Gel'fand basis are labeled by two patterns of indices satisfying (5.8), with identical values in the case $j = n$. These patterns are easily unified—as nothing more nor less than a general Gel'fand lattice. The affinity between these two structures is enhanced by their harmonious treatment of a special case: The Gel'fand basis elements are proportional[11,12] to matrix elements in which one pattern of indices have their maximal values; and when the two patterns are put together the maximality of one side results in a left Gel'fand lattice. Therefore it is very plausible that there should be a relation between matrix GLP and representation matrix elements. Such a relation is known[17]; for $n = 2$ the former are proportional to the latter, and for $n = 3$ the matrix elements are a linear combination of matrix GLP.

It is known[11,12] that every polynomial function of the matrix elements of z is a linear combination of the representation matrix elements. The elements of the representation $(m_{1n}, \ldots, m_{nn})$ are homogeneous polynomials in the $z_j{}^{i}$, of degree $m_{1n} + \cdots + m_{nn}$. It can also be seen that the matrix GLP, $\Delta^{T}(z)$, is homogeneous of degree $t_1{}^1 + \cdots + t_n{}^n$. So in order that every representation matrix element be a combination of matrix GLP, it is necessary and sufficient that every polynomial which is homogeneous of degree S be a combination of matrix GLP Δ^{T}, for which $t_1{}^1 + \cdots + t_n{}^n = S$. We note that exactly the same number of matrix GLP enter this relationship as do matrix elements of the representation. Therefore the former span the same linear space as the latter if and only if they are linearly independent.

In the following sections we shall prove (a) the left matrix GLP for lattices with a given median $[t]$ form a basis of the space $\mathscr{B}^{[t]}$; (b) the matrix GLP form a basis for the space of all polynomials. It is expedient to prove (a) emphasizing linear independence, and (b) emphasizing the fact that the GLP span the space of polynomials.

6. THE FIRST BASIS THEOREM

In this section we shall prove that the polynomials $\Delta^{T}(z)$, for left Gel'fand lattices T, are linearly independent. The proof depends on considering the polynomials expanded as linear combinations of monomials in the matrix elements $z_c{}^{r}$, and arranging these monomials in a certain order. Each monomial

$$\mathbf{z}_\alpha = \prod_{r,c} (z_c{}^{r})^{\alpha_c{}^{r}}$$

is determined by a system of exponents, which we shall write as follows:

$$\alpha = (\alpha_n{}^{n}, \alpha_n{}^{n-1}, \ldots, \alpha_n{}^{1}, \alpha^{n}_{n-1}, \ldots, \alpha^{n}_{n-1}, \ldots, \alpha_1{}^{n}, \ldots, \alpha_1{}^{1}). \tag{6.1}$$

We arrange the monomials in dictionary order, that is, $\alpha \geqslant \beta$ and $\mathbf{z}^{\alpha} > \mathbf{z}^{\beta}$ if in the rows α and β, written as in (6.1), the first unequal pair of exponents satisfy $\alpha_c{}^{r} > \beta_c{}^{r}$.

This order is preserved by multiplication: if $\mathbf{z}^{\alpha} > \mathbf{z}^{\beta}$, then $\mathbf{z}^{\alpha}\mathbf{z}^{\gamma} > \mathbf{z}^{\beta}\mathbf{z}^{\gamma}$. And if also $\mathbf{z}^{\alpha'} \geqslant \mathbf{z}^{\beta'}$, then $\mathbf{z}^{\alpha}\mathbf{z}^{\alpha'} > \mathbf{z}^{\beta}\mathbf{z}^{\beta'}$. In any multiplication of polynomials, therefore, the leading term of the product is formed from the leading terms of the factors.

When we consider the terms in the expansion of a subdeterminant of the matrix $\mathbf{z}$, we see that the leading term is the product of the elements along the main diagonal. In particular, the leading term of $\Delta_C{}^{1 \cdots l}$ is

$$z_{c_1}{}^{1} z_{c_2}{}^{2} \cdots z_{c_l}{}^{l} = \mathbf{z}_{(C)}, \tag{6.2}$$

where we have adapted the notation of (4.3).

Now consider the generating function for the left matrix GLP; it is just a special case of (4.10):

$$\left[\sum_{l} \sum_{C} \lambda_{(C)} \Delta_C{}^{1 \cdots l}(\mathbf{z}) \right]^{S} = \sum_{t_1{}^1 = S} \lambda^{\langle T \rangle} \Delta^{T}(\mathbf{z}). \tag{6.3}$$

On the left side, in each term of the sum the leading monomial in $\mathbf{z}$ has the same exponents as the monomial in λ. We expect to find the same agreement of exponents on the right, that is,

$$\Delta^T(\mathbf{z}) = a_T \mathbf{z}^{\langle T \rangle} + \text{lower terms.} \tag{6.4}$$

The coefficient a_T is a positive integer. A more formal proof of this is not difficult if we proceed by induction on $t_1{}^1 = s$. For $s = 1$, the lattice T is binary, say $T = \beta_c{}^{1\cdots l}$, and $\mathbf{z}^{\langle T \rangle} = \mathbf{z}_{(C)}$ as in (4.8a). Now [cf. (3.3)] if $T_1{}^1 = s > 1$

$$\Delta^T(\mathbf{z}) = \sum_{B + \tilde{T} = T} \Delta^B(\mathbf{z}) \Delta^{\tilde{T}}(\mathbf{z}), \tag{6.5}$$

where $B \in \mathscr{L}\mathscr{B}\mathscr{G}_n{}^*$, $\tilde{t}_1{}^1 = s - 1$. Suppose (6.4) is known for $\tilde{T}$; this is the induction hypothesis. Then in each summand of (6.5), the leading term is

$$a_{\tilde{T}} \mathbf{z}^{\langle B \rangle} \mathbf{z}^{\langle \tilde{T} \rangle} = a_{\tilde{T}} \mathbf{z}^{\langle B \rangle + \langle \tilde{T} \rangle} = a_{\tilde{T}} \mathbf{z}^{\langle T \rangle}.$$

Therefore the leading term in Δ^T is $a_T \mathbf{z}^{\langle T \rangle}$.

This proof also shows that the value of a_T is the number of distinct ordered partitions of T into binary lattices.

Now we observe that distinct polynomials $\Delta^T(\mathbf{z})$ have distinct leading terms, and infer at once that they are linearly independent. This concludes the proof of the main proposition of this section.

The left matrix GLP for lattices with a given median, say $[t]_n$, belong to the representation space $\mathscr{B}^{[t]_n}$. Since the number of polynomials is exactly the dimension of the space, they form a basis.

7. THE SECOND BASIS THEOREM

In this section we prove that the polynomials $\Delta^T(\mathbf{z})$ for all Gel'fand lattices T span the space of polynomial functions over the space of matrices. The proof depends on showing that the matrix GLP $\Delta^T(\mathbf{z})$ is equal, up to a sign, to a *left* matrix GLP $\Delta^{\tilde{T}}(\tilde{\mathbf{z}})$ where $\tilde{T}$ is a certain left Gel'fand lattice of dimension $2n$ and $\tilde{\mathbf{z}}$ is a certain matrix of $2n$ rows and columns. We define $\mathbf{z}$ as in Fig. 3.

The relation between T and $\tilde{T}$ starts with the case of binary lattices. If T is binary, then $\Delta^T(\mathbf{z})$ is a subdeterminant of $\mathbf{z}$; there is a subdeterminant of $\tilde{z}$, formed of the first n rows and a particular selection of n columns, which is proportional to $\Delta^T(\mathbf{z})$. Indeed, let $T = \beta_C{}^R$ where $R = (r_1,\ldots r_l)$ and $C = (c_1,\ldots,c_l)$. Let $S = (s_1,\ldots,s_{n-l})$, where $s_1 > \cdots > s_{n-l}$, be such that

$$(r_1,\ldots,r_l, s_1,\ldots,s_{n-l}) \tag{7.1}$$

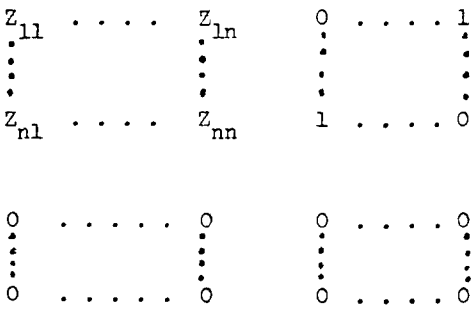


FIG. 3. $\tilde{Z}$ constucted from Z.

is a permutation of $(1,\ldots,n)$, and let $\epsilon(R,S)$ be the sign of that permutation. Let

$$\tilde{C} = (c_1,\ldots,c_l, 2n+1-s_1,\ldots,2n+1-s_{n-l}),$$
$$\tilde{R} = (1\cdots n).$$

Then

$$\Delta_C{}^R(\mathbf{z}) = \epsilon(R,S) \Delta_{\tilde{C}}{}^{\tilde{R}}(\tilde{\mathbf{z}}). \tag{7.2}$$

To prove this equation, consider $\Delta_{\tilde{C}}{}^{\tilde{R}}(\tilde{\mathbf{z}})$. It is the determinant of a matrix whose first l columns came form the left of $\tilde{\mathbf{z}}$, while the others came from the right half. We may permute these columns. We shall preserve the relative order of the former set, and dispose the latter so that each has a 1 on the main diagonal and 0's elsewhere. That is, the column of $\Delta_{\tilde{C}}{}^{\tilde{R}}$ which came from column $n+1-s_a$ of $\tilde{\mathbf{z}}$ is moved into column s_a of the permuted matrix. The effect of this permutation upon the determinant accounts for the factor $\epsilon(R,S)$ in (7.3). Now the permuted matrix is easily seen to have determinant equal to $\Delta_{\tilde{C}}{}^{\tilde{R}}(\mathbf{z})$.

We can restate (7.2) in notation closer to Gel'fand lattices. Let $T = \beta_C{}^R$ and $\tilde{T} = \tilde{\beta}_{\tilde{C}}{}^{\tilde{R}}$, where $\tilde{\beta}$ is analogous to β, in dimension $2n$. Then

$$\Delta^T(\mathbf{z}) = \epsilon(R,S) \Delta^{\tilde{T}}(\tilde{\mathbf{z}}). \tag{7.3}$$

To use this equation we shall need to know how $\tilde{T}$ is related to T. The relationship is simple:

$$\tilde{T} = T \begin{matrix} \mathbf{I} \\ 0 \end{matrix} \mathbf{I}, \tag{7.4}$$

where $\mathbf{I}$ is the $n \times n$ lattice with every cell occupied by 1. The formal proof of (7.5) will be much easier to understand if the reader tries out one or two examples first.

To begin, consider the lower left partial diagonals of T. The first l of these contain, respectively, $n+1-c_1,\ldots,n+1-c_l$ cells occupied by 1; the other cells are occupied by 0. As for the upper right partial diagonals, the first l of them have all cells occupied by 1; after that each one has, respectively, $n+1-s_1,\ldots,n+1-s_{n-l}$ cells occupied by 0. The last assertion is not very intuitive; it can be proved by the following way of describing how the cells of the right side of T are filled. Begin with the cell $\left[{}^n_1\right]$ in the right corner. If $1 \in R$, then fill this cell, and those above it to the left, with 1. Otherwise, if $1 \in S$, then fill this cell, and those obliquely below it, with 0. In either case, move to the rightmost of the remaining unoccupied cells, fill it with 1 or 0 depending on whether $2 \in R$ or $2 \in S$, and fill in those cells whose value is now constrained by (2.1). Proceed similarly for $3,\ldots,n$. Figure 4 shows the result of the process when $n = 7$, $R = (1, 2, 4, 5)$, and $S = (7, 6, 3)$.

Let us take this description of T, and suppose $\tilde{T}$ is defined by (7.4). It is a left binary Gel'fand lattice; therefore $\tilde{T} = \tilde{\beta}_{\tilde{C}}{}^{\tilde{R}}$, where $\tilde{R} = (1,\ldots,n)$ and $\tilde{C}$ is some set of n indices. We determine $\tilde{C}$ by examining the lower left partial diagonals in $\tilde{T}$. The first l of these are the first l lower left partial diagonals of T, each being extended upwards with n cells occupied by 1. So they contain, respectively, $2n+1-c_1,\ldots,$ $2n+1-c_l$ cells occupied by 1. The other lower left partial diagonals of $\tilde{T}$ are formed from the upper right partial dia-

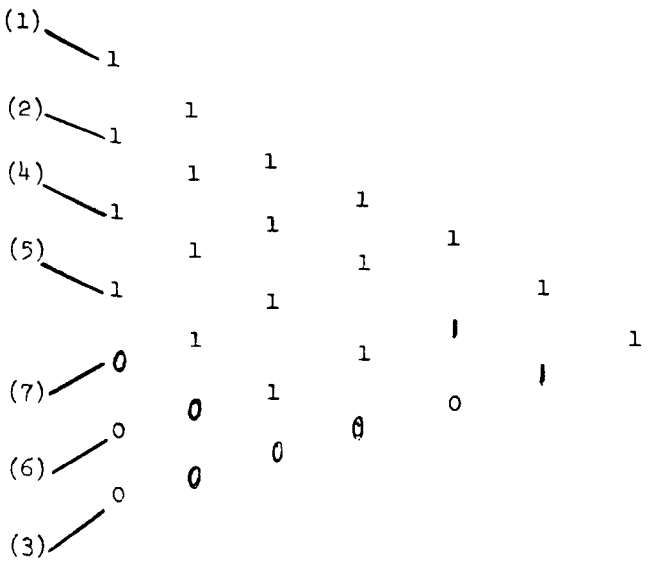


FIG. 4. Example of determination of $\tilde{C}$.

gonals of T by this process: Extend upward with cells filled with 1 until the length of the diagonal is $n+1$, extend downward with cells filled with 0. The number of cells filled with 1 is, respectively, $s_1,\ldots,s_{n-l}$. To summarize, $\tilde{C}=(c_1,\ldots,c_l, 2n+1-s_1,\ldots,2n+1-s_{n-l})$. This definition of $\tilde{C}$ agrees with that which was used in (7.2) and so shows the agreement of (7.2) with (7.4).

Now let us consider the sign factor $\epsilon(R,S)$, and show how it may be obtained from $\tilde{T}$. The permutation (7.1) can be sorted into increasing order by moving s_1 left until it is in correct position relative to $r_1,\ldots,r_l$, then moving s_2 left until it is in correct position relative to the numbers already sorted, and so on. To put s_a in its place requires $n-s_a$ transpositions, so the total is

$$\sum_{a=1}^{n-l}(n-s_a)=n^2-nl-\sum s_a.$$

This number is congruent, modulo 2, to

$$n+nl+\sum s_a \tag{7.5}$$

and (7.5) is the sum

$$\sum_{j=1}^{n}\sum_{i=j}^{j+n}\tilde{T}_i{}^j=\nu(\tilde{T}). \tag{7.6}$$

Let us summarize what we have done in our investigation of (7.2). We assume that $\tilde{\mathbf{z}}$ is related to $\mathbf{z}$ as in Fig. 3, and $\tilde{T}$ is related to T as in (7.4). We define $\nu(\tilde{T})$ by (7.6). Then

$$\Delta^{T}(\mathbf{z})=(-1)^{\nu(\tilde{T})}\Delta^{\tilde{T}}(\tilde{\mathbf{z}}). \tag{7.7}$$

This equation is proved for binary Gel'fand lattices. It clearly extends to general Gel'fand lattices if we generalize (7.5) thus:

$$\tilde{T}=T\begin{matrix} t_1{}^1\mathbf{I} \\ 0 \end{matrix}\quad t_1{}^1\mathbf{I}. \tag{7.8}$$

Indeed, the correspondence of $\tilde{T}$ with T now respects addition of Gel'fand lattices, in that if $T=T'+T''$ then $\tilde{T}=\tilde{T}'+\tilde{T}''$; and $\nu(\tilde{T})=\nu(\tilde{T}')+\nu(\tilde{T}'')$; so (7.7) is easily proved by induction.

By the first independence theorem, the left matrix GLP $\Delta^{\tilde{T}}$ are linearly independent. This does not immediately imply that the matrix GLP, $\Delta^{T}(\mathbf{z})$, are linearly independent, because in (7.7) $\tilde{\mathbf{z}}$ is restricted to be a matrix of the form of Fig. 3. A more indirect argument is needed. As remarked in Sec. 6, every polynomial in one of the Bargmann–Moshinsky spaces is a linear combination of left matrix GLP. We shall use this fact, together with (7.7), to prove that every polynomial is a linear combination of matrix GLP.

Let f be a polynomial function of the matrix elements of an $n\times n$ matrix $\mathbf{z}$. We may separate f into homogeneous parts, so it is no loss of generality to assume that f is in fact homogeneous:

$$f(c\mathbf{z})=c^{t}f(\mathbf{z}). \tag{7.9}$$

Let $\mathbf{w}$ be a second $n\times n$ matrix, and let

$$J=\begin{pmatrix}0 & \cdots & 1\\ \vdots & \iddots & \vdots\\ 1 & \cdots & 0\end{pmatrix}$$

be the reversed-diagonal matrix. Define a function f_1 by

$$f_1(\mathbf{z}\,\mathbf{w})=(\det\mathbf{w})^{t}f(J\,\mathbf{w}^{-1}\mathbf{z}). \tag{7.10}$$

Then f_1 is in fact a polynomial because the matrix elements of $[\det(\mathbf{w})\mathbf{w}^{-1}]$ are polynomials in the matrix elements of $\mathbf{w}$. If $\mathbf{v}$ is a lower triangular matrix with 1's on the diagonal, then

$$f_1(\mathbf{v}\,\mathbf{z},\,\mathbf{v}\,\mathbf{w})=f_1(\mathbf{z},\,\mathbf{w}). \tag{7.11}$$

If δ is a diagonal matrix, then

$$f_1(\boldsymbol{\delta}\,\mathbf{z},\boldsymbol{\delta}\,\mathbf{w})=\boldsymbol{\delta}^{[t\cdots t]}f_1(\mathbf{z},\mathbf{w}). \tag{7.12}$$

Now define a function f_2 of $2n\times 2n$ matrices:

$$f_2\begin{pmatrix}\mathbf{z}_{11} & \mathbf{z}_{12}\\ \mathbf{z}_{21} & \mathbf{z}_{22}\end{pmatrix}=f_1(\mathbf{z}_{11},\mathbf{z}_{12}). \tag{7.13}$$

It follows from (7.11) and (7.12) that

$$f_2\in\mathscr{B}^{[t\cdots t,0\cdots 0]}.$$

By the corollary of the first independence theorem applied to U(2n),

$$f_2=\sum_{\hat{\mathrm{U}}\in\mathscr{T}\mathscr{G}_{2n}}C(\hat{\mathrm{U}})\Delta^{\hat{\mathrm{U}}}. \tag{7.14}$$

Now in (7.13), the lattices $\hat{\mathrm{U}}$ must have their median column equal to $[t,\ldots,t,0,\ldots,0]$; that is

$$\hat{\mathrm{U}}=\mathrm{U}\begin{matrix} t\,\mathbf{I} \\ 0 \end{matrix}\quad t\,\mathbf{I},$$

where $\mathrm{U}\in\mathscr{G}_n$ and $\mathrm{U}_1{}^1\leqslant t$. In analogy with (7.8), we set

$$\tilde{\mathrm{U}}=\mathrm{U}\begin{matrix} \mathrm{U}_1{}^1\mathbf{I} \\ 0 \end{matrix}\quad \mathrm{U}_1{}^1\mathbf{I}.$$

It can be seen that every partition of $\hat{\mathrm{U}}$ into binary lattices contains $t-\mathrm{U}_1{}^1$ terms of the form $\beta^{1\cdots n}_{n+1\cdots 2n}$, while the remaining terms constitute a partition of $\tilde{\mathrm{U}}$. Therefore

$$\Delta^{U}(\mathbf{z}) = \binom{t}{U_1{}^1} \Delta^{1\cdots n}_{n+1\cdots 2n}(\mathbf{z})^{t-U_1{}^1} \Delta^{\bar{U}}(\mathbf{z}). \tag{7.15}$$

When we combine (7.10), (7.13), (7.14), and (7.15) we find

$$f(\mathbf{z}) = \sum \binom{t}{U_1{}^1} c(\hat{U}) \det(J)^{t-U_1{}^1} \Delta^{\bar{U}} \begin{pmatrix} \mathbf{z} & J \\ 0 & 0 \end{pmatrix};$$

by (7.7), this implies that $f(\mathbf{z})$ is a linear combination of the matrix GLP, $\Delta^{U}(\mathbf{z})$. The proof of the second basis theorem is complete.

One corollary of the second basis theorem is a weak form of the "multiplication theorem" (Ref. 17). This is concerned with $\Delta^{T}(\mathbf{z}\,\mathbf{w})$ and its expansion as a sum of products $\Delta^{U}(\mathbf{w})\Delta^{U'}(\mathbf{z})$. Clearly, there is such an expansion, but the theorem just proved offers little information on the lattices U and U' that appear in the sum, or the coefficients. Gazeau, Dumont-Lepage, and Ronveaux have worked out the case $n = 3$, from which it appears that a completely unrestricted sum is much too general. In fact it is reasonable to conjecture from their work that

(a) the lattice elements of U to the *left* of the median coincide with those of T;

(b) the lattice elements of U' to the *right* of the median coincide with those of T;

(c) the elements of U to the right of the median and those of U' to the left of the median are mutual transposes;

(d) the coefficient in the sum depends upon T only through the median elements.

It appears that new ideas are needed to prove these conjectures.

8. THE ORDERING OF GLP AND GEL'FAND STATES

The third result in this paper concerns the linear relation between matrix Gel'fand lattice polynomials and orthonormal Gel'fand states. It depends on arranging each set of polynomials in a certain order, which we shall describe here.

We know that the matrix GLP, $\Delta^{T}(\mathbf{z})$, for left lattices with a given median $[t]$, are contained in $\mathscr{B}^{[t]}$. So it is sufficient to describe the order relation between left lattices with the same median. Essentially, it is the order which we used to prove the first basis theorem in Sec. 6. We shall say that $T > U$, and $\Delta^{T} > \Delta^{U}$, if $\mathbf{z}^{\langle T\rangle} > \mathbf{z}^{\langle U\rangle}$. Now this relation can be defined by comparing differences of corresponding elements of T and U. Given that the medians of T and U are equal, the same effect can be obtained by comparing the elements themselves.

An example will make the logic clearer than a formal argument. Consider Fig. 5, and an analogous figure for the lattice U. The lexical ordering of $\mathbf{z}^{\langle T\rangle}$ and $\mathbf{z}^{\langle U\rangle}$ would compare the differences between elements of T which are linked by solid lines with the corresponding differences for U; if these differences were all equal, then the differences represented by dashed lines would become significant. Now if T and U have a common median, then equality of the differences along the solid lines implies equality of those elements that are linked by the solid lines. Therefore we can describe the lexical order on left Gel'fand lattices by arranging the lattice elements, from the left side of the median, in this order:

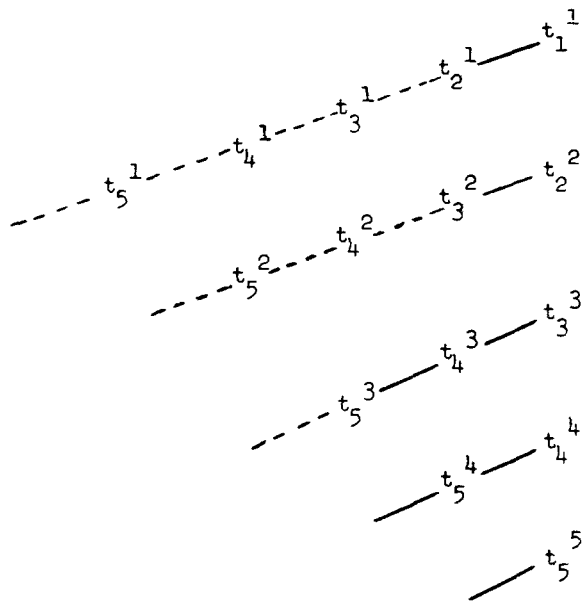


FIG. 5. Ordering of Gel'fand lattices.

$$t_n{}^{n-1},\ldots,t_2{}^1, t_n{}^{n-2},\ldots,t_3{}^1,\ldots,t_n{}^1,$$

and saying that $T > U$ if, in this arrangement, the first place where T and U differ satisfies $t_i{}^j < u_i{}^j$.

9. $\mathscr{B}^{[t]}$ ON A SPACE OF TRIANGULAR MATRICES

The equations which characterize an element F of $\mathscr{B}^{[t]}$ are

$$F(\mathbf{v}\,\mathbf{z}) = F(\mathbf{z}), \tag{9.1}$$

$$F(\boldsymbol{\delta}\,\mathbf{z}) = \boldsymbol{\delta}^{[t]} F(\mathbf{z}), \tag{9.2}$$

where $\mathbf{v}$ is an arbitrary lower triangular matrix with 1's on the diagonal, and $\boldsymbol{\delta}$ is an arbitrary triangular matrix. Equation (9.1) implies that F is determined by its restriction to upper triangular matrix. Equation (9.1) implies that F is determined by its restriction to upper triangular matrices (with arbitrary nonzero diagonal elements), because for almost all matrices $\mathbf{z}$, there is a factorization

$$\mathbf{z} = \mathbf{v}\,\boldsymbol{\zeta}, \tag{9.3}$$

where $\mathbf{v}$ is lower triangular, $\boldsymbol{\zeta}$ is upper triangular, and the diagonal elements of $\mathbf{v}$ are all 1.

We shall find it convenient to make certain computations on members of $\mathscr{B}^{[t]}$, restricted to the upper triangular matrices. To be formal about it, let:

$$M_+ = \{\mathbf{z} \in M(n,\mathbb{C}): z_{ij} = 0 \text{ if } i > j\};$$

if F is a polynomial function on $M(n,\mathbb{C})$, then F_+ is its restriction to M_+. The vector space $\{F_+: F \in \mathscr{B}^{[t]}\}$ will be denoted $\mathscr{B}_+{}^{[t]}$. If $\mathbf{z}$ admits a factorization as in (9.3), then we set

$$\mathbf{v} = \mathbf{P}(\mathbf{z}), \quad \boldsymbol{\zeta} = Q_+(\mathbf{z}). \tag{9.4}$$

The construction of $\mathscr{B}_+{}^{[t]}$ is similar to the definition of the representation spaces as studied by Zhelobenko.[6] He considers the restriction to upper triangular matrices with 1's on the diagonal; (9.1) and (9.2) together imply that a function in $\mathscr{B}^{[t]}$ is determined by that restriction.

The action of $\mathrm{GL}(n,\mathbb{C})$ on $\mathscr{B}_+{}^{[t]}$ can be defined ab-

stractly by

$$L^{[t]}_{(+)u}(F_+) = (L_u{}^{[t]}F)_+.$$

To be a little more concrete, we can say

$$(L^{[t]}_{(+)u}F_+)(\xi) = F_+(Q_+(\xi\,\mathbf{u})), \quad \xi\in M_+, \quad \mathbf{u}\in \mathrm{GL}(n,\mathbb{C}). \tag{9.5}$$

There is a certain "abuse of language" here, because $Q_+(\xi\,\mathbf{u})$ is not defined for all ξ and $\mathbf{u}$. As we shall see, however, its elements are rational functions of the elements of $\xi\,\mathbf{u}$; and the fact that F_+ is the restriction to M_+ of a polynomial satisfying (9.1) on $M(n,\mathbb{C})$ implies that the right side of (9.5) turns out to be a polynomial in ξ and $\mathbf{u}$. (Not every polynomial function on M_+ can be extended to a polynomial on $M(n,\mathbb{C})$ while satisfying (9.1). The requirement for such an extension to exist can, in principle, be formulated in terms of the behavior of F_+ "at infinity"; so far as I know, this has never been worked out.).

Clearly (9.5) will be easier to handle if we know the exact expression of $Q_+(\mathbf{z})$. The usual proof of (9.3) constructs the factors $\mathbf{v}$ and ξ one row or column at a time, by the process known to numerical analysts as "Gaussian elimination." If explict formulas are carried along, then the process eventuates in this result:

$$(Q_+\,\mathbf{z})_{ij} = \Delta^{1\cdots i}_{1\cdots i-1,j}(\mathbf{z})/\Delta^{1\cdots i-1}_{1\cdots i-1}(\mathbf{z}), \quad i\leqslant j, \tag{9.6}$$

$$(\mathbf{P}\,\mathbf{z})_{ij} = \Delta^{1\cdots j-1,i}_{1\cdots j}(\mathbf{z})/\Delta^{1\cdots j}_{1\cdots j}(\mathbf{z}), \quad i\geqslant j. \tag{9.7}$$

These equations, once found or conjectured, can be verified easily. Indeed, (9.6) is left invariant by the transformation $\mathbf{z}\to\mathbf{v}\,\mathbf{z}$ if $\mathbf{v}$ is lower triangular and has 1's on the diagonal; and if $\mathbf{z}\in M_+$, then (9.6) reduces to z_{ij}. A similar argument validates (9.7).

10. LIE ALGEBRA OPERATIONS ON M_+

We intend to use the "lowering" operators of Nagel and Moshinsky[16] to construct Gel'fand states, as polynomials over the space M_+. The lowering operators are built up out of Lie algebra operators, so we shall first describe the action of the Lie algebra operators on $\mathscr{B}_+{}^{[t]}$. In effect, we compute the infinitesimal form of (9.5).

If C is any element of the Lie algebra of $\mathrm{GL}(n,\mathbb{C})$, then there is a one-parameter group $\tau\to\mathbf{u}_\tau$ such that

$$\mathbf{u}_\tau = \exp(\tau C). \tag{10.1}$$

In the representation space $\mathscr{B}_+{}^{[t]}$ we define the action of C by

$$CF_+ = ((d/d\tau)L^{[t]}_{(+)u_\tau}F_+)_{\tau=0}. \tag{10.2}$$

Now F_+ is a polynomial function of the coordinates ξ_{ij}, $i\leqslant j$. We can therefore compute (10.2) if we know the action of C on ξ_{ij}:

$$CF_+(\xi) = \sum_{i<j}(\partial F_+(\xi)/\partial\xi_{ij})(C\xi_{ij}). \tag{10.3}$$

From (9.5) we see that

$$C\xi_{ij} = [(\partial/\partial\tau)Q_+(\xi\mathbf{u}_\tau)_{ij}]_{\tau=0}.$$

To make this equation more explicit, we can refer to (9.6) and the action of C on subdeterminants such as $\Delta_J{}^{1\cdots i}$.

Let us assume that $C = C_b{}^a$ in the notation of Ref. 11. Acting on a polynomial function of a matrix $z = (z_{ij})$,

$$C_b{}^a = \sum_k z_{kb}\frac{\partial}{\partial z_{ka}}.$$

We find $C_b{}^a\Delta_J{}^{1\cdots i} = 0$ unless $a\in J$ and $b\notin J$; if these conditions are satisfied and $J = J'a\,J''$, then

$$C_b{}^a\Delta_J{}^{1\cdots i} = \Delta^{1\cdots i}_{J'bJ''}.$$

One can now apply $C_b{}^a$ to $(Q_+z)_{ij}$ for any choice of a, b, i, and j, using (9.6). The number of different cases is unwieldy, so we shall examine just one, which will be useful in the next section.

The case we consider is that of $C_b{}^a(Q_+z)_{hk}$ where $a<h<b<k$. We compute

$$C_b{}^a\Delta^{1\cdots h}_{1\cdots h-1,k} = (-1)^{h-a-1}\Delta^{1\cdots h}_{1\cdots a-1\cdots h-1,b,k},$$

$$C_b{}^a\Delta^{1\cdots h-1}_{1\cdots h-1} = (-1)^{h-a-1}\Delta^{1\cdots h-1}_{1\cdots a-1,a+1\cdots h-1,b},$$

$$C_b{}^a(Q_+z) = (-1)^{h-a-1}[\Delta^{1\cdots h-1}_{1\cdots h-1}]^{-2}\cdot[\Delta^{1\cdots h-1}_{1\cdots h-1}\Delta^{1\cdots h}_{1\cdots a-1,a+1\cdots h-1,b,k} - \Delta^{1\cdots h-1}_{1\cdots a-1,a+1\cdots h-1,b}\Delta^{1\cdots h}_{1\cdots h-1,k}].$$

We restrict the last expression to z in M_+. There are some simplifications:

$$\Delta^{1\cdots h-1}_{1\cdots h-1}(z) = z_{11}\cdots z_{h-1,h-1},$$

$$\Delta^{1\cdots h}_{1\cdots h-1,k} = z_{11}\cdots z_{h-1,h-1}z_{h,k},$$

$$\Delta^{1\cdots h}_{1\cdots a-1,a+1\cdots h-1,bk} = z_{hk}\Delta^{1\cdots h-1}_{1\cdots a-1,a+1\cdots h-1,b} - z_{hb}\Delta^{1\cdots h-1}_{1\cdots a-1,a+1\cdots h-1,k}$$

The result, for the group action on the upper triangular manifold, is

$$C_b{}^a\xi_{hk} = (-1)^{h-a}(\xi_{11}\cdots\xi_{h-1,h-1})^{-1}\xi_{hb}\times\Delta^{1\cdots h-1}_{1\cdots a-1,a+1\cdots h-1,k}(\xi). \tag{10.4}$$

The determinant in (10.4) has, in general, 2^{h-a-1} terms.

11. COMPUTATION OF THE LEADING TERM OF A GEL'FAND STATE

The main result of this section is that if the orthonormal Gel'fand states $\Gamma(T)$ are restricted to M_+, then the leading term in the polynomial expansion of $\Gamma(T)$ is proportional to $\xi^{\langle T\rangle}$. Together with the first basis theorem, this result will imply that

$$\Gamma(T) = \sum_{T'\leqslant T}\mu(T',T)\Delta^{T'}, \tag{11.1}$$

so that the Gel'fand states can be obtained from the GLP, ordered by increasing leading terms, with a Gram–Schmidt orthogonalization process.

Before starting on the proof, we may look at some cases[17] where $\Gamma(T)(\xi)$ is known. If T is maximal, then it is well known[8,10] that

$$\Gamma(T) = \mathrm{const}\times\prod_{a=1}^{n}(\Delta^{1\cdots a}_{1\cdots a})^{t_a{}^a - t_{a+1}{}^{a+1}}.$$

Now $\Delta^{1\cdots a}_{1\cdots a}(\xi) = \xi_{11}\cdots\xi_{aa}$; when we rearrange factors we find

$$\Gamma(T) = \mathrm{const}\times\prod_{a=1}^{n}\xi_{aa}{}^{t^a_a}.$$

If T is semimaximal, i.e., maximal for U_{n-1}, then

$$\Gamma(T) = \text{const} \times \prod_{a=1}^{n} (\Delta^{1\cdots a}_{1\cdots a-1,n})^{t^a_a - t^a_{a+1}} \times \prod_{a=1}^{n-1} (\Delta^{1\cdots a}_{1\cdots a})^{t_{a+1}^{a+1} - t^a_{a+1}}.$$

Using the fact that $\Delta^{1\cdots a}_{1\cdots a-1,n}(\xi) = \xi_{11}\cdots\xi_{a-1,a-1}\xi_{an}$, we find

$$\Gamma(T) = \text{const} \times \prod_{a=1}^{n} \xi_{an}^{\,t^a_a - t^a_{a+1}} \prod_{a=1}^{n-1} \xi_{aa}^{\,t^a_{a+1}} = \text{const} \times \prod_{a=1}^{n} \xi_{an}^{\langle T,an\rangle} \prod_{a=1}^{n-1} \xi_{aa}^{\langle T,aa\rangle}. \tag{11.2}$$

This formula is worth dwelling on for a moment; it shows that a semimaximal polynomial, on M_+, is a monomial in the diagonal elements of ξ and the elements of the last column. The simplicity of this relation is our motive for computing over M_+ in the present section.

What we shall prove about the general Gel'fand state over M_+ is slightly more than we stated at the beginning of this section. To state it conveniently, let us generalize some of the notation of Sec. 4, in particular (4.5 a, b). If $\tau = (\tau_{gh})$ is a system of integers indexed by (g, h) ranging over the trapezoidal pattern defined by the inequalities $p+1 \leqslant h \leqslant n$, $1 \leqslant g \leqslant h$, then let

$$\xi^\tau = \prod_{h=p+1}^{n} \prod_{g=1}^{h} \xi_{gh}^{\,\tau_{gh}}. \tag{11.3}$$

If $\tau_{gh} = \langle T,gh\rangle$ for some Gel'fand lattice T, then set $\tau = \langle (p)T]$. The notation $\xi^{\langle (p)T]}$ generalizes (4.5a); $\xi^{\langle (p)T]}$ does not depend on the p columns to the far left of T. Let $\xi_{(p)}$ denote the submatrix of ξ containing the first p columns. Every polynomial function F over M_+ can be written

$$F(\xi) = \sum_\tau F_\tau(\xi_{(p)})\xi^\tau.$$

We shall prove that if the left Gel'fand lattice T is maximal for $U(p-1)$ $\Gamma(T)$ has this special form:

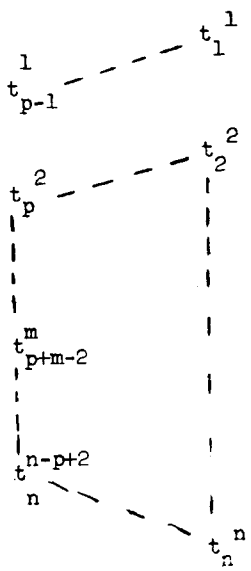


FIG. 6. Indices before application of $L_p{}^m$.

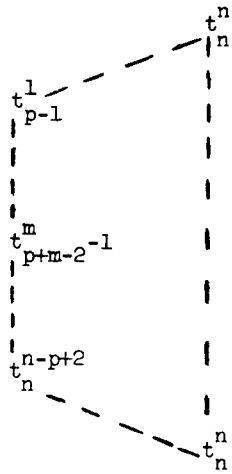


FIG. 7. Indices after application of $L_p{}^m$.

$$\Gamma(T)(\xi) = \text{const} \times \xi^{\langle T|} + \sum_{\tau:\tau < \langle (p)T|} F_\tau(\xi_p)\xi^\tau. \tag{11.4}$$

The proof of (11.4) goes by induction on T, starting from the "least" possible lattice for a given representation space. Because of the way we defined the order on states and lattices (see Sec. 8), this lattice actually belongs to the maximal state. Equation (11.2) shows that the proposition is valid for that state, in fact for all the semimaximal states as well. The step to lower states is made by application of the lowering operators of Nagel and Moshinsky.[16] We need not go into the full details of the construction of these operators; what we need to know is that operators $L_p{}^m$ are defined, where $1 \leqslant m < p \leqslant n$, so that

(i) $L_p{}^m$ is a polynomial in the Lie algebra operators $C_b{}^a$ where $m \leqslant a \leqslant b \leqslant p$;

(ii) if Φ maximal with respect to U_{p-1}, belonging to a representation space of type $(s_1,\ldots,s_{p-1})$, then $L_p{}^m\Phi$ is also maximal with respect to U_{p-1} and belongs to a representation space of type $(s_1,\ldots,s_m{}',\ldots,s_{p-1})$.

If T and T' are the left Gel'fand lattices of Fig. 6 and 7, then

$$\Gamma(T') = \text{const} \times L_p{}^m \Gamma(T). \tag{11.5}$$

The induction step in proving (11.4) is to suppose it is known for T and to prove it for T'. We have

$$\Gamma(T') = \text{const} \times L_p{}^m(\xi^{\langle T|}) + \sum_{\tau:\tau < \langle (p)T|} L_p{}^m(F_\tau \xi^\tau). \tag{11.6}$$

Look at the first term on the right side of (11.6). We can expand it as

$$\text{const} \times \left[L_p{}^m\left(\prod \xi_{a\,p}{}^{\langle T,a\,p\rangle} \prod \xi_{aa}{}^{\langle T,aa\rangle}\right)\right]\xi^{\langle (p)T|} + \text{const} \times \sum_\alpha \left[M_\alpha\left(\prod \xi_{a\,p}{}^{\langle T,a\,p\rangle} \prod \xi_{aa}{}^{\langle Y,aa\rangle}\right)\right] N_\alpha(\xi^{\langle (p)T|}), \tag{11.7}$$

where M_α and N_α are various differential operators; like $L_p{}^m$, they are polynomials in $C_b{}^a$ where $m \leqslant a \leqslant b \leqslant p$. The first part of the expansion is simple, because in it $L_p{}^m$ is acting on a semimaximal state for U_p, and must transform it into another semimaximal state, that is,

$$L_p{}^m\Big(\prod \xi_{a\,p}{}^{\langle T,a\,p\rangle}\prod \xi_{aa}{}^{\langle T,aa\rangle}\Big)$$
$$= \text{const}\times\prod \xi_{a\,p}{}^{\langle T',a\,p\rangle}\prod \xi_{aa}{}^{\langle T',aa\rangle}.$$

Therefore the first part of the expansion is

$$\text{const}\times\xi^{\langle T']}.$$

To complete the proof we must show that the remaining terms in (11.7), and all the other terms in (11.6), can be gathered together in a summation

$$\sum_{\tau:\tau<\langle(p)T']} F'_\tau\,\xi^\tau.$$

Note that $\langle(p)T'] = \langle(p)T]$. For this purpose, we use the following lemma.

Lemma: Let $a\leqslant b\leqslant p$, and let $h\leqslant k$, $k>p$, Let $C_b{}^a$ act on ξ_{hk} as described in Sec. 10. Then the result is either 0 or a sum of terms all lower than ξ_{hk}.

Proof: Given those constraints on the values of a, b, h, and k, there is only one combination in which $C_b{}^a\xi_{hk}\neq 0$, and that is the case that $a<h<b<k$. Equation (10.4) applies. The leading term in a subdeterminant is the main diagonal, so the leading term in $C_b{}^a\xi_{hk}$ is found to be

$$(\xi_{aa}\cdots\xi_{h-1,h-1})^{-1}\xi_{a,a+1}\cdots\xi_{h-2,h-1}\,\xi_{h-1,k}\,\xi_{hb}.$$

Factors from the rightmost columns take precedence in determining the order of monomials, and $\xi_{h-1,k}<\xi_{hk}$, so the whole term is less than ξ_{hk}.

We now apply the Lemma to (11.7) and (11.6). In (11.7), all the parts except the first contain at least one Lie algebra operator acting upon the monomial $\xi^{\langle(p)T']}$. The operators in question are $C_b{}^a$, where $a\leqslant b\leqslant p$, and the variables in $\xi^{\langle(p)T']}$ are ξ_{hk}, where $h\leqslant k$, $k>p$. The Lemma shows that when such an operator meets such a variable, the result must be lower. Thus all the terms which are generated are lower than $\xi^{\langle(p)T']}$. As for the other terms in (11.6), they can be expanded into terms $(L_p{}^mF_\tau)\xi^\tau$ where $\tau<\langle(p)T]$, and other terms where Lie algebra operators act on ξ^τ and lower it even more. Therefore (11.5) implies that (11.4) is true with T' in place of T, and this completes the proof.

12. COMPUTATION OF INDIVIDUAL GLP

In this section we return to the case of arbitrary initial conditions to describe a means of computing individual GLP. We start with the left GLP. The essence of the method is to relate the generating function for left GLP of dimension n to the similar function for dimension $n-1$. Let $X_{(C)} = X_C{}^{1\cdots l}$.

It is helpful to introduce some notation for a certain map from left Gel'fand lattices of dimension n to those of dimension $n-1$. If $T\in\mathscr{L}\mathscr{G}_n$, we characterize $T(n-1)=T'$ in $\mathscr{L}\mathscr{G}_{n-1}$ by

$$\langle T',ab\rangle = \langle T,ab\rangle,\quad 1\leqslant a\leqslant b\leqslant n-1.$$

Equivalently,

$$(t')_i{}^j = t^j_{i+1},\quad i\geqslant j,$$
$$(t')_i{}^j = t^i_{i+1},\quad i<j.$$

With the notation $\langle(n-1)T]$ of Sec. 11, we have

$$\lambda^{\langle T]} = \lambda^{\langle T']}\lambda^{\langle(n-1)T]}. \tag{12.1}$$

Consider the generating function for left GLP of dimension n:

$$\Big(\sum_C \lambda_{(C)}\mathbf{X}_{(C)}\Big)^s = \sum_{T:t_1{}^1=s}\lambda^{\langle T]}\mathbf{X}^T.$$

On the left, we may partition the summands into those for which $n\in C$ and those for which $n\notin C$; there are 2^{n-1} of the former and $2^{n-1}-1$ of the latter. Let $A(l,n-1)$ denote the set of multi-indices $D=(d_1,\ldots,d_l)$ where $1\leqslant d_1<\cdots<d_l\leqslant n-1$. Then

$$\sum_C \lambda_{(C)}\mathbf{X}_{(C)} = \lambda_{1n}\mathbf{X}_{(n)} + \sum_{l=1}^{n-1}\sum_{D\in A(l,n-1)}\lambda_D(\mathbf{X}_{(D)} + \lambda_{l+1,n}\mathbf{X}_{(Dn)}).$$

Let

$$\mathbf{Y}_D = \mathbf{X}_{(D)} + \lambda_{l+1,n}\mathbf{X}_{(Dn)}, D\in A(l,n-1).$$

Then

$$\sum_C \lambda_{(C)}\mathbf{X}_{(C)} = \lambda_{1n}\mathbf{X}_{(n)} + \sum_D \lambda_{(D)}\mathbf{Y}_{(D)}.$$

Therefore, from (12.2),

$$\sum_{T:t_1{}^1=s}\lambda^{\langle T]}\mathbf{X}^T = \Big(\lambda_{1n}\mathbf{X}_{(n)} + \sum_D \lambda_{(D)}\mathbf{Y}_{(D)}\Big)^s$$
$$= \sum_r \binom{s}{r}(\lambda_{1n}\mathbf{X}_{(n)})^{s-r}\Big(\sum_D \lambda_{(D)}\mathbf{Y}_{(D)}\Big)^r$$
$$= \sum_{r=t_2{}^1}\sum_{T'}\binom{s}{r}(\lambda_{1n}\mathbf{X}_{(n)})^{s-r}\lambda^{\langle T']}\mathbf{Y}^{T'}.$$

We use (12.1) to find, for a fixed choice of $T(n-1)$,

$$\sum_{(n-1)T:t_1{}^1=s}\lambda^{\langle(n-1)T]}\mathbf{X}^T = \binom{t_1{}^1}{t_2{}^1}(\lambda_{1n}\mathbf{X}_{(n)})^{\langle T,1n\rangle}\mathbf{Y}^{T(n-1)}.$$

From this generating function, individual terms can be extracted:

$$\mathbf{X}^T = \binom{t_1{}^1}{t_2{}^1}X_n{}^{\langle T,1n\rangle}\prod_{l=2} \times[(\partial/\partial\lambda_{ln})^{\langle T,ln\rangle}/\langle T,ln\rangle!]\cdot\mathbf{Y}^{T(n-1)}|_{\lambda=0}.$$

If we introduce these differential operators:

$$\mathbb{D}_{ln} = \sum_{D\in A(l-1,n-1)} X_{(Dn)}\partial/\partial X_{(D)},\quad l\geqslant 2,$$

then

$$\mathbf{X}^T = \binom{t_1{}^1}{t_2{}^1}X_n{}^{\langle T,1n\rangle}\prod_{l=2}^n(\mathbb{D}_{ln}{}^{\langle T,ln\rangle}/\langle T,ln\rangle!)\mathbf{X}^{T'}. \tag{1.24}$$

We can make this equation a little more convenient by introducing another indeterminate $X_{(\,)}$, corresponding to the empty set, and

$$\mathbb{D}_{1n} = X_{(n)}\,\partial/\partial X_{(\,)}.$$

Then, because $\langle T,1n\rangle = t_1{}^1 - t_2{}^1$,

$$X_{(\,)}{}^{t_2{}^1}X^T$$

$$= \prod_{l=1}^{n} (\mathbb{D}_{ln}{}^{\langle T,ln\rangle}/\langle T,ln\rangle!) X_{(\ )}{}^{t_1'} X^{T'}. \tag{12.5}$$

Let us define

$$\mathbb{D}^{\langle (n\ \ 1)T\ \rangle} = \prod_{l=1}^{n} (\mathbb{D}_{ln}{}^{\langle T,ln\rangle}/\langle T,ln\rangle!). \tag{12.6}$$

Then (12.5) can be restated

$$X_{(\ )}{}^{t_1'} X^{T} = X_{(\ )}{}^{\langle T,1n\rangle} \mathbb{D}^{\langle (n-1)T\ \rangle} X_{(\ )}{}^{\langle T,1n\rangle} (X_{(\ )}{}^{t_2'} X^{T'}).$$

We have arranged the powers of $X_{(\ )}$ in this equation to show how the process of constructing X^T from $X^{T'}$ may be iterated. If we start from

$$X^0 = 1,$$

then we have

$$X_{(\ )}{}^{t_1'} X^{T} = \prod_{m=1}^{n} \{X_{(\ )}{}^{\langle T,1m\rangle} \mathbb{D}^{\langle (m-1)\ T\ \rangle} X_{(\ )}{}^{\langle T,1m\rangle}\} 1. \tag{12.7}$$

To extend this result from left GLP to arbitrary GLP, we shall use the relations between Gel'fand lattices of different dimensions that we developed in Sec. 7. In fact, the map $\rho: T \to \bar{T}$, defined by (7.8), is a homomorphism from the semigroup $\mathscr{L}_n$ to the semigroup $\mathscr{L}\mathscr{G}_{2n}$. It maps the generators of $\mathscr{G}_n$, i.e., the binary lattices, onto certain generators of $\mathscr{L}\mathscr{G}_n$; specifically, $\rho(\underline{\beta}_S{}^R) = \beta_{\bar{S}}{}^{\bar{R}}$ with the notation of (7.2). Every partition of $\bar{T}$ is the image of a partition of T; therefore if we set

$$\rho(X_T) = X_{\bar{T}}$$

and extend ρ to homomorphism of polynomials in the indeterminates X_T, then

$$\rho(\mathbf{X}^T) = \mathbf{X}^{\bar{T}}. \tag{1.28}$$

Equation (12.8) may be compared with (7.7). In (7.7) the sign factor was necessary because we were concerned with the values of matrix GLP for certain matrixes. Equation (12.8) implies that we can construct (12.7); then apply ρ^{-1}, replacing each factor $X_{\bar{C}}{}^{(1\cdots n)}$ by $X_C{}^R$ according to the rule given in Sec. 7.

A program, written in the programming language Pascal, for computing individual left GLP is available from the author.

[1]H. Weyl, Math. Z. **23**, 271, 328 (1925); *ibid.* **24**, 377, 798 (1926).
[2]H. Weyl, *The structure and representations of continuous groups*, Institute of Advanced Studies, Princeton, 1935 (unpublished lecture notes); *The Classical Groups*, (Princeton U.P., Princeton, New Jersey, 1946).
[3]I.M. Gel'fand and M.L. Zeitlin, Dok. Akad. Nauk. SSSR **71**, 825 (1950).
[4]I.M. Gel'fand and M.I. Graev, Izv. Akad. Nauk. SSSR Ser. Mat. **29**, 1329 (1965) [Amer. Math. Soc. Transl. Ser. 2 **64**, 116 (1967)].
[5]R. Godement, Trans. Am. Math. Soc. **73**, 496 (1952).
[6]D.P. Zhelobenko, Russ. Math. Surv. **17**, 1 (1962).
[7]S. Warner, *Harmonic Analysis on Semi-Simple Lie Groups*, Vol. I, (Springer, New York, 1972); esp. Chap. 3.1.
[8]V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961).
[9]M. Moshinsky, Nucl. Phys. **31**, 384 (1962); J. Math. Phys. **4**, 1128 (1963).
[10]G.E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[11]J.D. Louck, J. Math. Phys. **6**, 1786 (1965).
[12]J.D. Louck, Am. J. Phys. **38**, 3 (1970).
[13]M. Ciftan and L.C. Biedenharn, J. Math. Phys. **10**, 221 (1969); M. Ciftan, J. Math. Phys. **10**, 1635 (1969).
[14]A.C.T. Wu, J. Math. Phys. **12**, 437 (1971); C.-S. Huang and A.C.T. Wu, J. Math. Phys. **14**, 263 (1971).
[15]C.J. Henrich, J. Math. Phys. **16**, 2271 (1975).
[16]J.G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); Rev. Mex. Phys. **14**, 29 (1965).
[17]J.P. Gazeau, M.-Cl. Dumont-Lepage, and A. Ronveaux, J. Math. Phys. **19**, 734 (1978).

Micu-type invariants of exceptional simple Lie algebras

Vishnu K. Agrawala
Department of Natural Sciences and Technology, Point Park College, Pittsburgh, Pennsylvania 15222

(Received 17 July 1979; accepted for publication 22 August 1979)

A formula simplifying the quartic trace in exceptional simple Lie algebras is derived and used for expressing higher order Micu-type invariants as polynomials in the second order Casimir invariant.

1. INTRODUCTION

In an earlier paper,[1] an algorithm was described for computing the eigenvalues of Micu-type invariants for the classical simple Lie algebras A_r, B_r, C_r, and D_r. However, the techniques used there could not be used for the exceptional Lie algebras because the expressions for the projection operators are much more complicated for these algebras. An independent method is presented here which takes advantage of the fact that for the exceptional simple Lie algebras E_r $(r = 6,7,8)$,F_4, and G_2, the tensor product of the adjoint module with itself decomposes as a direct sum of only five irreducible modules, no two of which are isomorphic. This fact turns out to be so restrictive that all Micu-type invariants are found to be simple polynomials of the second order Casimir invariant. In Micu's paper,[2] the question of algebraic dependence of his series of invariants was left moot for the exceptional case. Our Eq. (3.1) explicitly shows them to be algebraically dependent.

The notation of Ref. 1 will be maintained here without any change, except that L will now stand for any of the simple exceptional Lie algebras E_r $(r = 6,7,8)$,F_4, or G_2. Greek indices will run from 1 through dim$L$ and summation over repeated indices is implied. In any L-module M, we write $M_\alpha = G_\alpha(M)$ for the operators representing the basis elements G_α of L. The module L itself in the adjoint representation will be denoted by A. For our application in Sec. 3, F will be the fundamental (lowest dimensional) module, but all results of Sec. 2 hold even when F is an arbitrary faithful irreducible L-module.

2. QUARTIC TRACE

The main feature of the present method of computing invariants is the representation (2.2) of the quartic trace $\mathrm{Tr}_F(F_\alpha F_\beta F_\gamma F_\delta)$ as a linear combination of simpler tensors. In order to analyze the mathematical structure behind it, we first show that dim hom $(L \otimes L \otimes L \otimes L,\mathbb{C}) = 5$ by exploiting the isomorphisms hom$(L \otimes L \otimes L \otimes L,\mathbb{C})$ $\cong$ hom$(A \otimes A, A^* \otimes A^*) \cong$ hom$(A \otimes A, A \otimes A)$. Now for the exceptional simple Lie algebras, $A \otimes A$ is a direct sum

$$A \otimes A = \overset{5}{\underset{j=1}{\oplus}} M_j$$

of five irreducible submodules, no two of which are isomorphic (Table II in Appendix). If $\pi_j : A \otimes A \to M_j$ and $i_j : M_j \to A \otimes A$ denote the canonical projections and injections associated with this direct sum decomposition, then any operator σ in hom $(A \otimes A, A \otimes A)$ may be written as

$$\sigma = \sum_{j=1}^{5} \sum_{k=1}^{5} i_j \pi_j \sigma i_k \pi_k .$$

By Schur's lemma, $\pi_j \sigma i_k \in \mathrm{hom}(M_k, M_j)$ is zero if $j \neq k$ and is a multiple s_j of the unit operator if $j = k$; hence, the above sum reduces to

$$\sigma = \sum_{j=1}^{5} s_j i_j \pi_j . \tag{2.1}$$

The five orthogonal projectors $i_j \pi_j$ $(j = 1,\ldots,5)$ are obviously linearly independent while Eq. (2.1) shows that they span hom $(A \otimes A, A \otimes A)$. These five projectors thus form a

TABLE I. Basic data about exceptional simple Lie algebras.

	E_6	E_7	E_8	F_4	G_2
dim L	78	133	248	52	14
Adjoint module, A	ω_6	ω_6	ω_1	ω_4	ω_2
Fundamental module, F	ω_1	ω_1	ω_1	ω_1	ω_1
dim F	27	56	248	26	7
$\gamma(A)$	1	1	1	1	1
$\gamma(F)$	13/18	19/24	1	2/3	1/2
$\mathrm{Tr}_F(F_\alpha F_\beta)/g_{\alpha\beta}$	1/4	1/3	1	1/3	1/4
a_F	1/576	1/648	1/300	1/324	1/192
b_F	1/24	1/18	1/6	1/18	1/24

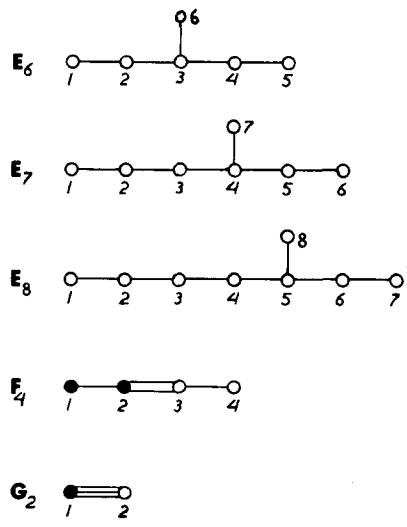


FIG. 1. Dynkin diagrams showing the numbering of simple roots. Black dots represent the shorter roots. The ratio (length of larger root)/(length of smaller root) is 2 for F_4 and 3 for G_2.

 0022-2488/80/071577-02$1.00

TABLE II. Clebsch–Gordan series for $A \otimes A$, where A is adjoint module. $M(\lambda)$ denotes the module with highest weight λ.

Algebra type	Clebsch–Gordan series for $A \otimes A$
E_6	$M(2\omega_6) \oplus M(\omega_6) \oplus M(\omega_1+\omega_5) \oplus M(\omega_3) \oplus M(0)$
E_7	$M(2\omega_6) \oplus M(\omega_6) \oplus M(\omega_5) \oplus M(\omega_2) \oplus M(0)$
E_8	$M(2\omega_1) \oplus M(\omega_7) \oplus M(\omega_2) \oplus M(\omega_1) \oplus M(0)$
F_4	$M(2\omega_4) \oplus M(2\omega_1) \oplus M(\omega_4) \oplus M(\omega_3) \oplus M(0)$
G_2	$M(2\omega_2) \oplus M(2\omega_1) \oplus M(3\omega_1) \oplus M(\omega_2) \oplus M(0)$

basis of hom $(A \otimes A, A \otimes A)$ which must have dimension five.

Now the operation of taking the trace defines a tensor $t \in L^* \otimes L^* \otimes L^* \otimes L^*$ by

$$t(G_\alpha \otimes G_\beta \otimes G_\gamma \otimes G_\delta) = \mathrm{Tr}_F(F_\alpha F_\beta F_\gamma F_\delta),$$

with extension to the entire space by linearity. This tensor is invariant, i.e., $t \in \hom(L \otimes L \otimes L \otimes L, \mathbb{C})$ because the trace form is always invariant. To see this, let G denote the Lie group corresponding to L. Then for every $g \in G$ we have

$$\begin{aligned}(g\cdot t)(G_\alpha \otimes G_\beta \otimes G_\gamma \otimes G_\delta) &= (g\cdot \mathrm{Tr}_F)(F_\alpha \otimes F_\beta \otimes F_\gamma \otimes F_\delta)\\ &= g\cdot(\mathrm{Tr}_F g^{-1}\cdot(F_\alpha F_\beta F_\gamma F_\delta))\\ &= \mathrm{Tr}_F(g^{-1}(F_\alpha F_\beta F_\gamma F_\delta)g)\\ &= \mathrm{Tr}_F(F_\alpha F_\beta F_\gamma F_\delta)\\ &= t(G_\alpha \otimes G_\beta \otimes G_\gamma \otimes G_\delta).\end{aligned}$$

We shall expand the fourth order invariant tensor t as a linear combination of five linearly independent fourth order tensors made out of second and third order trace forms. More explicitly, in terms of components, we write

$$\begin{aligned}&\mathrm{Tr}_F(F_\alpha F_\beta F_\gamma F_\delta)\\ &\quad = a_1 g_{\alpha\beta}g_{\gamma\delta} + a_2 g_{\alpha\gamma}g_{\beta\delta} + a_3 g_{\alpha\delta}g_{\beta\gamma}\\ &\qquad + a_4 f_{\mu\alpha\beta} f^{\mu}{}_{\gamma\delta} + a_5 f_{\mu\beta\gamma} f^{\mu}{}_{\delta\alpha}.\end{aligned} \tag{2.2}$$

The tensors on the right-hand side are invariant because they are obtained by tensoring and contracting

$$g_{\alpha\beta} = \mathrm{Tr}_A(A_\alpha A_\beta),$$

$$f_{\alpha\beta\gamma} = \mathrm{Tr}_A(A_\alpha A_\beta A_\gamma - A_\alpha A_\gamma A_\beta),$$

which are themselves invariant, being linear combinations of trace forms.

To find the coefficients a_1, a_2, a_3, a_4, a_5 in Eq. (2.2), we can contract it with $g^{\alpha\beta}, g^{\beta\gamma}, g^{\gamma\alpha}, f^{\alpha\beta}_{\lambda\cdot\cdot}$ and $f^{\beta\gamma}_{\lambda\cdot\cdot}$ thereby getting a system of five equations. The matrix of the system is nonsingular showing that the tensors used on the right-hand side of Eq. (2.2) form a basis of hom $(L \otimes L \otimes L \otimes L, \mathbb{C})$. The solution is $a_1 = a_2 = a_3 = a_F$ and $a_4 = a_5 = b_F$ where

$$a_F = \frac{\dim F\,\gamma(F)}{6\dim L}\,\frac{(6\gamma(F)-1)}{(\dim L+2)},$$

$$b_F = \dim F\,\gamma(F)/(6\dim L).$$

We wish to remark that Eq. (2.2) provides a quick and direct proof of Okubo's quartic trace identity.[3] If X is any nonzero member of L, it can be extended to a basis $\{X_1 = X, X_2, \ldots, X_{\dim L}\}$ of L. Let the corresponding representatives in the module F be denoted by F_α $(\alpha = 1,\ldots \dim L)$. Then setting $\alpha = \beta = \gamma = \delta = 1$ in Eq. (2.2), using Eq. (5.1) of Ref. 1, and the fact that $f_{\mu 11} = 0$, we get the quartic trace identity

$$\mathrm{Tr}_F(F_1^4) = 2a_F g_{11} = \frac{6\gamma(F)-1}{2(\dim L+2)}\mathrm{Tr}_F(F_1^2).$$

3. EIGENVALUES OF INVARIANTS

In view of their definition, Eq. (3.8) of Ref. 1, the eigenvalue of a Micu-type invariant U_{2k} in any module M is obtained by contracting Eq. (2.2) with $V^{\alpha}_{(2k-3)}(M) M^{\beta} M^{\gamma} \times M^{\delta}$, where $M^{\alpha} = G^{\alpha}(M)$. The result of contraction, simplified with the help of Eqs. (4.3), (4.7), and (5.1) of Ref. 1, reduces to

$$u_{2k}(M) = C_F(M) u_{2k-2}(M),$$

where

$$C_F(M) = \frac{\dim F\,\gamma(F)}{12\dim L}\left(\frac{[6\gamma(F)-1][6\gamma(M)-1]}{2+\dim L}+1\right).$$

Iteration of this finally yields

$$u_{2k}(M) = [C_F(M)]^{k-1}\gamma(M). \tag{3.1}$$

APPENDIX

Some useful information about exceptional simple Lie algebras is collected here for easy reference. Figure 1 shows the numbering of simple roots. We have chosen to number them so that for each Lie algebra, the fundamental representation has highest weight $\omega_1 = (1,0,\ldots,0)$. The data in Table I were computed by a program written by Agrawala and Belinfante,[4] while the Clebsch–Gordan series given in Table II was taken from Seligman's book.[5]

[1] V.K. Agrawala, "Micu-type invariants of simple Lie algebras," J. Math. Phys. **20**, 2178 (1980).

[2] M. Micu, Nucl. Phys. **60**, 353 (1964).

[3] S. Okubo, J. Math. Phys. **20**, 586 (1979).

[4] V.K. Agrawala and J.G. Belinfante, BIT **11**, 1 (1971).

[5] G.B. Seligman, *Rational Methods in Lie Algebras* (Dekker, New York, 1976), Appendix A.

Addition of complex angular momentum operators

K. N. Joshi
Department of Mathematics, Garhwal University, Srinagar 246174 (Garhwal), U. P., India

B. S. Rajput[a)]
Department of Physics, Garhwal University, Srinagar 246174 (Garwhal), U. P., India

(Received 18 January 1979; accepted for publication 6 August 1979)

Representations of the components of complex angular momentum operator have been carried out in three different subspaces of the space in which the generators of proper, orthochronous homogeneous Lorentz group operate. Addition of complex angular momentum operators have been derived in the canonical basis of their eigenvectors.

1. INTRODUCTION

Some noncompact groups have played a major role in the understanding and development of important physical ideas. Among these the most striking one is the use of SL(2,C) for the group theoretical treatment of Virasoro–Shapiro model.[1] Group SL(2,C) is the universal covering group of the homogeneous Lorentz group SO(3,1), the correspondence of the elements from the former to the latter being one to one. Developments in the last ten years in physics have made important expansions with respect to SL(2,C) representations which include infinitesimal nonunitary representations[2] as well. Unitary representations of the homogeneous Lorentz group have been obtained and classified by Bargmann,[3] Gel'fand and Naimark,[4] and Barut and Wilson.[5] Though the representation theory of the homogeneous Lorentz group SO(3,1) [or the covering group SL(2,C)] is of great importance in physics and the knowledge of its Clebsch–Gordan (CG) coefficients is of fundamental importance in various areas of elementary particle physics because, in addition to its extensive use in scattering theory and the theory of wave equations,[6] it serves as the group of motion of space of independent kinematical variables in one approach,[7] as the invariance group of the zero angle elastic scattering amplitude in the second approach[8,9] and as the group generating a complete set of basis functions in the third,[10] it is surprising that up to 1970 the attention had hardly been paid to reconstructing and investigating the representation of SL(2,C) groups in an explicitly analytically continuable form while most of the work had been devoted to[11–14] construct its representations explicitly.

The question of obtaining the unitary CG coefficients by analytically continuing the known spinor CG coefficients was first considered by Anderson *et al.*[15] who obtained the recursion and symmetry relations and obtained the CG coefficients associated with the coupling two SL(2,C) principal series representations. It has recently been shown by Wong and Yeh[16] that the boost matrix elements of SO(3,1) obtained by Smorodinskii and Shepelev[17] are connected to those of SO(4) by analytic continuation and that the CG coefficients of SO(3,1) are analytic continuation of X functions from SO(4) but their results are very complicated due to the summation of two Fourier series. In view of such considerations we developed in our earlier paper[18] a compact operator formulation to reformulate the Gel'fand–Naimark theory of representations of SL(2,C) by combining the generators of the homogeneous Lorentz group to undertake the studies of operators associated with the collective relativistic motion of a body rotating about one axis and moving with relativistic velocity along the other. Diagonal elements of this operator have been interpreted as the components Z_1, Z_2, and Z_3 of complex angular momentum operator Z in complex space and its representations have been derived in the basis of the eigenvectors of the third component of the usual angular momentum operator.

In the present paper the representations of the components of complex angular momentum operators have been derived in three different subspaces R_{l-1}, R_l, and R_{l+1} of the space R in which the generators of the homogeneous Lorentz group operate and the eigenvalues of the complex angular momentum operator and its third component have been calculate in the canonical basis of its eigenvectors. It has been shown that the eigenvector of the components of complex angular momentum operator forms the complete set in the subspace R_l while in the subspaces R_{l-1} and R_{l+1} separately the sets of eigenvectors are not complete. Addition of complex angular momentum operators has been carried out in the subspace R_l and the corresponding value of CG coefficients[19,20] have been derived. The operator method introduced here for the representation of SO(3,1) has already been used in generalized[21] form to treat the projective representation of SL(3,C). Similarly, it can find applications in the representation theory of other noncompact groups relevant to physics such as the de Sitter groups or the conformal groups.

2. REPRESENTATIONS OF COMPLEX ANGULAR MOMENTUM OPERATORS

Denoting by J_j and K_j ($j = 1,2,3$) the generators of rotations about and boosts along the jth axis, respectively, we have shown in our earlier paper[18] that the linear combinations

$$Z_j = \tfrac{1}{2}(J_j + iK_j) \tag{2.1a}$$

and

[a)]On leave from Physics Department, Kurukshetra University, Kurukshetra, India.

 0022-2488/80/071579-04$1.00

$$X_j = \tfrac{1}{2}(J_j - iK_j) \tag{2.1b}$$

satisfy the Lie algebra of two independent angular momenta in complex spaces (J,K) and $(J,-K)$ and they have therefore been defined as the components of complex angular momentum operators in these spaces. These operators have also been introduced by Bars and Gursey[22] as left handed and right handed vector operators to treat the Gel'fand–Naimark Z basis for the homogeneous Lorentz group.

Since the operators Z_j and X_j given in Eqs. (2.1) commute, the representation of the homogeneous Lorentz group can be considered as a direct product of the two groups generated by Z_j and X_j as has already been shown by Smorodinskii and Huszar[23] by considering the two parameter subgroup $H = \mathrm{SO}(2) \otimes \mathrm{SO}(1,1)$. Combining the components Z_1 and Z_2 in the following manner:

$$\begin{aligned} Z_+ &= Z_1 + iZ_2, \\ Z_- &= Z_1 - iZ_2, \end{aligned} \tag{2.2}$$

we have shown[18] that the space R in which the generators of the homogeneous Lorentz group operate may be analyzed into a linear sum of invariant subspaces R_l, in each of which an irreducible representation of weight "l" of the group of ordinary rotation is obtained. Choosing the eigenvectors of the operator J_3 as the canonical basis $|l,m\rangle$ in each subspace R, the following representations of the operators Z_+, Z_-, and Z_3 have been obtained:

$$\begin{aligned} Z_3|l,m\rangle &= mA_1|l,m\rangle - A_2(l^2-m^2)^{1/2}|l-1,m\rangle \\ &\quad + A_3[(l+1)^2-m^2]^{1/2}|l+1,m\rangle, \end{aligned} \tag{2.3a}$$

$$\begin{aligned} Z_+|l,m\rangle &= A_1[(l-m)(l+m+1)]^{1/2}|l,m+1\rangle \\ &\quad - A_2[(l-m)(l-m-1)]^{1/2}|l-1,m+1\rangle \\ &\quad - A_3[(l+m+1)(l+m+2)]^{1/2} \\ &\quad \times|l+1,m+1\rangle, \end{aligned} \tag{2.3b}$$

$$\begin{aligned} Z_-|l,m\rangle &= A_1[(l+m)(l-m+1)]^{1/2}|l,m-1\rangle \\ &\quad + A_2[(l+m)(l+m-1)]^{1/2}|l-1,m-1\rangle \\ &\quad + A_3[(l-m+1)(l-m+2)]^{1/2} \\ &\quad \times|l+1,m-1\rangle, \end{aligned} \tag{2.3c}$$

where

$$\begin{aligned} A_1 &= \left(1 + \frac{l_0 l_1}{l(l+1)}\right), \\ A_2 &= \frac{1}{l}\left(\frac{(l^2-l_0^2)(l^2-l_1^2)}{4l^2-1}\right)^{1/2}, \\ A_3 &= \frac{1}{l+1}\left\{\frac{[(l+1)^2-l_0^2]\,[(l+1)^2-l_1^2]}{4(l+1)^2-1}\right\}^{1/2}, \end{aligned} \tag{2.4}$$

where $m = -l, -l+1, \ldots, 0, \ldots, l-1, l$, $l = l_0, l_0+1, \cdots$ and l_1 is some complex number. It is obvious from Eqs. (2.3) that each representation is uniquely defined by the pair of numbers l_0 and l_1, where l_0 is the smallest weight participating in the irreducible representation and the number l_1 is arbitrary.

Let us now consider the following particular cases of representations of these operators.

Case I: When only the weight l participates in the representation, the general equations (2.3a)–(2.3c) reduce to the following form:

$$Z_3|l,m\rangle = mA_1|l,m\rangle, \tag{2.5a}$$

$$Z_+|l,m\rangle = A_1[(l-m)(l+m+1)]^{1/2}|l,m+1\rangle, \tag{2.5b}$$

$$Z_-|l,m\rangle = A_1[(l+m)(l-m+1)]^{1/2}|l,m-1\rangle. \tag{2.5c}$$

In this basis of representations we obviously get the following matrix elements of the operators Z_+, Z_-, and Z_3:

$$\langle l,m'|Z_3|l,m\rangle = mA_1\delta_{m,m'}, \tag{2.6a}$$

$$\langle l,m'|Z_+|l,m\rangle = A_1[(l-m)(l+m+1)]^{1/2}\delta_{m,m+1}, \tag{2.6b}$$

$$\langle l,m'|Z_-|l,m\rangle = A_1[(l+m)(l-m+1)]^{1/2}\delta_{m,m-1}. \tag{2.6c}$$

Let us now introduce the operator Z^2 as

$$Z^2 = Z_+Z_- - Z_3 + Z_3^2, \tag{2.7}$$

for which the following commutation rules may be readily derived by using the commutation rules[18] of the components of the complex angular momentum operator:

$$[Z^2,Z_1] = [Z^2,Z_2] = [Z^2,Z_3] = 0. \tag{2.8}$$

Using Eqs. (2.8) and (2.5a)–(2.5c), we get

$$Z^2|l,m\rangle = A_1\{A_1[l(l+1)+m] - m\}|l,m\rangle$$

and

$$Z_3|l,m\rangle = mA_1|l,m\rangle. \tag{2.9}$$

Similar results may be derived for the operators $X^2 = X_+X_- - X_3 + X_3^2$ and X_3 and it may be shown that in the space R_l the basis $|l,m\rangle$ give's the simultaneous eigenstates of the four commuting operators Z^2, X^2, Z_3, and X_3, with mA_1 and mB being the complex eigenvalues of the last two operators where A_1 is given by Eq. (2.4) and B may be derived to have the following value:

$$B = \left(1 - \frac{l_0 l_1}{l(l+1)}\right). \tag{2.10}$$

Changing the canonical basis from $|l,m\rangle$ to $|L,M\rangle$, where M represents the eigenvalues of Z_3 and $L(L+1)$ are the eigenvalues of Z^2, we get

$$Z^2|L,M\rangle = L(L+1)|L,M\rangle, \tag{2.11a}$$

$$Z_3|L,M\rangle = M|L,M\rangle, \tag{2.11b}$$

where

$$L = \frac{-1 \pm [1 + 4A_1^2 l(l+1) + 4A_1 M - 4M]^{1/2}}{2} \tag{2.12a}$$

and

$$M = mA_1 \tag{2.12b}$$

are complex angular momenta and magnetic quantum numbers, respectively. From these relations we may get the different allowed values of M in terms of L corresponding to various allowed values of m in terms of l. For instance, if $m = -l$, we get

$$L = A_1 l \quad \text{or} \quad -A_1(l+1), \tag{2.13}$$

with the corresponding values of M given by

$$M = -L \quad \text{or} \quad L+1, \tag{2.14}$$

where we choose only first value in accordance with its cor-

respondence with the value of m. Similarly, corresponding to the values of $m = -l+1, -l+2,...,0,...,l-1,l$, we get $M = -L+A_1, -L+2A_1,...,0,...,L-A_1,L$. In this way, corresponding to $(2l+1)$ values of m, we get $[(2L+A_1)/A_1]$ values of M. In the special case when $A_1 = 1$, i.e., $l_1 = 0$, we get the components of the operators K corresponding to the pure Lorentz transformations vanishing. Then, $L = l$ and consequently the components of complex angular momentum operator reduce to those of ordinary angular momentum operator.

Case II: When only the weight "$l-1$" participates in the representations of the complex angular momentum operator, the general equations (2.3) reduce to the following form:

$$Z_3|l-1,m\rangle = -A_2(l^2-m^2)^{1/2}|l-1,m\rangle, \tag{2.15a}$$

$$Z_+|l-1,m\rangle = -A_2[(l-m)(l-m-1)]^{1/2}|l-1,m+1\rangle, \tag{2.15b}$$

$$Z_-|l-1,m\rangle = A_2[(l+m)(l+m-1)]^{1/2}|l-1,m-1\rangle. \tag{2.15c}$$

Then, for the operator Z^2 given by Eq. (2.7), we get

$$Z^2|l-1,m\rangle = A_2(l^2-m^2)^{1/2}\{A_2(l^2-m^2)^{1/2} - A_2[l^2-(m-1)^2]^{1/2}+1\}|l-1,m\rangle. \tag{2.16}$$

Similar results may be derived for the operator X^2 and X_3 and it can be shown that, in the space R_{l-1}, the basis $|l-1,m\rangle$ gives the simultaneous eigenstates of the operators Z^2, X^2, Z_3, and X_3 with $\mp A_2(l^2-m^2)$, respectively, being the complex eigenvalues of last two operators. Changing the canonical basis from $|l-1,m\rangle$ to $|L,M\rangle$, Eqs. (2.16) and (2.15a) reduce to

$$Z^2|L,M\rangle = L(L+1)|L,M\rangle \text{ and } Z_3|L,M\rangle = M|L,M\rangle, \tag{2.17}$$

respectively, where

$$L = -\tfrac{1}{2} + \{1+4M^2-4M+4A_2M \times[l^2-(m-1)^2]^{1/2}\}^{1/2}/2 \tag{2.18a}$$

and

$$M = -A_2(l^2-m^2)^{1/2}. \tag{2.18b}$$

From these relations we may readily get the following values of M corresponding to different allowed values of m:

$$M = 0, -A_2(2l-1)^{1/2}, -A_2[2(2l-2)]^{1/2},..., -lA_2,..., -A_2(2l-1)^{1/2}, 0, \tag{2.19}$$

which are all negative and each value is repeated twice.

Case III: When only the weight "$l+1$" participates in the representations of the complex angular momentum operator, the general representation equations (2.3) reduce to the following form:

$$Z_3|l+1,m\rangle = A_3[(l+1)^2-m^2]^{1/2}|l+1,m\rangle, \tag{2.20a}$$

$$Z_+|l+1,m\rangle = -A_3[(l+m+1)(l+m+2)]^{1/2}|l+1,m+1\rangle, \tag{2.20b}$$

$$Z_-|l+1,m\rangle = A_3[(l-m-1)(l-m+2)]^{1/2}|l+1,m-1\rangle. \tag{2.20c}$$

Then, for the operator Z^2, given in Eq. (2.7), we get the following representation:

$$Z^2|l+1,m\rangle = A_3[(l+1)^2-m^2]^{1/2} \times\{A_3[(l+1)^2-m^2]^{1/2} - A_3[(l+1)^2-(m-1)^2]^{1/2}-1\}|l+1,m\rangle. \tag{2.21}$$

In the canonical basis $|L,M\rangle$, Eqs. (2.20) and (2.21) may be written as Eq. (2.17), where M and L have the following values:

$$L = -\tfrac{1}{2} + \{1+4M^2-4M-4A_3M \times[(l+1)^2-(m-1)^2]^{1/2}\}^{1/2}/2 \tag{2.22}$$

and

$$M = A_3[(l+1)^2-m^2]^{1/2}, \tag{2.23}$$

which give the following values for M corresponding to the values of $m = -l, -l+1,...,0,...,l-1,l$:

$$M = A_3(2l+1)^{1/2}, A_3(2.2l)^{1/2}, A_3[3(2l-1)]^{1/2},..., A_3(l+1),...,A_3(2.2l)^{1/2}, A_3(2l+1)^{1/2}, \tag{2.24}$$

which are all positive and each value is repeated twice.

3. ADDITION OF COMPLEX ANGULAR MOMENTUM OPERATORS

In the space when only the weight l participates in the representations of the components of the complex angular momentum operator, we get the similar set of values of M in terms of L in the basis $|L,M\rangle$ as that of m in terms of l in the basis $|l,m\rangle$. Therefore, the addition of complex angular momentum operators in the canonical basis $|L,M\rangle$ may be derived in the way similar to that of the addition of ordinary angular momentum operators in the basis $|l,m\rangle$. As such, the eigenstates $|L,M\rangle$ of the operators Z^2 and Z_3, when $Z = Z_1 + Z_2$, may be given as follows in terms of eigenstates $|L_1,M_1\rangle$ and $|L_2,M_2\rangle$ of Z_1^2 and Z_2^2, respectively:

$$|L,M\rangle = \sum_{M_1,M_2} C(L_1M_1,L_2M_2,LM)|L_1,M_1,L_2,M_2\rangle, \tag{3.1}$$

where

$$|L_1,M_1,L_2,M_2\rangle = |L_1,M_1\rangle\,|L_2,M_2\rangle, \tag{3.2}$$

and the coefficients $C(L_1M_1,L_2M_2,LM)$, which are nonzero only when $M = M_1 + M_2$, are given by

$$C(L_1M_1,L_2M_2,LM) = \left(\frac{2L+A_1}{A_1}\right)^{1/2} V(L_1L_2L;M_1M_2M), \tag{3.3}$$

where

$$V(L_1L_2L;M_1M_2M) = \left[\frac{(L_1+L_2-L)!(L_1-L_2+L)!(-L_1+L_2+L)!}{(L_1+L_2+L+1)!}\right]^{1/2} [(L_1+M_1)!(L_1-M_1)!(L_2+M_2)!$$

$$\times(L_2 - M_2)!(L + M)!(L - M)!]^{1/2} \sum_z \{(-1)^z/z!(L_1 + L_2 - L + z)!(L_1 - M_1 - z)!$$
$$\times(L_2 + M_2 - z)!(-L_2 + L + M_1 + z)!(-L_1 + L - M_2 + z)!\}, \tag{3.4}$$

where z runs through zero and integers. This value of $C(L_1 M_1, L_2 M_2, LM)$ in the basis $|L,M\rangle$ is similar to that of the Clebsch–Gordan coefficients for addition of ordinary angular momentum, operators in the basis $|l,m\rangle$ except that we get a factor $[(2L + A_1)/A_1]^{1/4}$ in place of $(2l + 1)^{1/2}$ and angular momenta and magnetic quantum numbers given by Eqs. (2.12a) and (2.12b), respectively, are complex. We may therefore consider these coefficients as CGC for complex angular momentum operators in the basis $|L,M\rangle$. The existence of these Clebsch–Gordan coefficients with complex angular momenta and magnetic quantum numbers has already been utilized by Wong and Yeh[16] to demonstrate that CG coefficients of SO(3,1) are the analytic continuation of well known functions from SO(4). Actually, the boost matrix elements of SO(3,1) should be connected to those of SO(4) by analytic continuation.

We conclude that in the subspace R_{l-1} the allowed values of M given by Eq. (2.19) are only negative values and therefore the basis vectors $|L,M\rangle$ do not form the complete set in this subspace. Similar is the case in subspace R_{l+1} where all the allowed values of M are only positive. It is therefore not possible to derive the addition of complex angular momentum operators in these subspaces separately in the simple way discussed above for the subspace R_l.

[1] I. Bars and F. Gursey, Phys. Rev. D **4**, 1769 (1971).
[2] K. Szego and K. Toth, J. Math. Phys. **12**, 846 (1971).
[3] V. Bargmann, Ann. Math. **48**, 568 (1947).
[4] I.M. Gel'fand and M.A. Naimark, Izv. Akad. Nauk. SSR **11**, 411 (1947).
[5] A.O. Barut and R. Wilson, J. Math. Phys. **17**, 900 (1976).
[6] N. Mukunda, J. Math. Phys. **9**, 50 (1968).
[7] E.G. Stoyanov, D.T. Stoyanov, and I.T. Todorov, ICTP, Trieste, Preprint IC/67/58; C. Fronsdal, Phys. Rev. **156**, 1653 (1967).
[8] J.A. Smorodinsky and N.J. Vilenkin, Zh. Eksp. Teor. Fiz. **46**, 1793 (1964) [Sov. Phys. JETP **19**, 1209 (1964)]; M. Sheftel, J.A. Smorodinsky, and P. Winternitz, Yad. Fiz. **7**, 1325 (1968) [Sov. J. Nucl. Phys. **7**, 785 (1968)].
[9] A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).
[10] R. Delbourgo, A. Salam, and J. Strathdee, Phys. Lett. B **25**, 230 (1967).
[11] J.F. Boyce, R. Delbourgo, A. Salam, and J. Strathdee, ICTP, Trieste, Preprint IC/67/69.
[12] M. Toller, Nuovo Cimento A **53**, 671 (1968).
[13] Y. Nambu, Phys. Rev. **160**, 1171 (1967).
[14] G. Domokos and E. Sckonberg, Phys. Rev. D **2**, 1026 (1970).
[15] R.L. Anderson, R. Raczka, M.A. Rashid, and P. Winternitz, J. Math. Phys. **11**, 1050, 1059 (1970).
[16] M.K.F. Wong and H.-Y. Yeh, J. Math. Phys. **18**, 1768 (1977).
[17] Ya.A. Smorodinskii and G.I. Shebelev, Sov. J. Nucl. Phys. **13**, 248 (1971).
[18] K.N. Joshi and B.S. Rajput, "Representations of Complex Angular Momentum Operators," J. Math. Phys. (to be published).
[19] A. Messiah, *Quantum Mechanics*, Vol. II (Wiley, New York, 1962).
[20] I.M. Gel'fand, R.A. Minlos and Z.Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963).
[21] I. Bars, J. Math. Phys. **14**, 759 (1973).
[22] I. Bars and F. Gursey, J. Math. Phys. **13** 131 (1972).
[23] Ya.A. Smorodinskii and M. Huszar, Theor. Math. Phys. **4**, 867 (1970).

Ermakov systems, nonlinear superposition, and solutions of nonlinear equations of motion

James L. Reid and John R. Ray
Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29631

(Received 7 February 1980; accepted for publication 20 March 1980)

We report several important additions to our original discussion of Ermakov systems. First, we show how to derive the Ermakov system from more general equations of motion. Second, we show that there is a general nonlinear superposition law for Ermakov systems. Also, we give explicit examples of the nonlinear superposition law. Finally, we point out that any ordinary differential equation can be included in many Ermakov systems.

I. INTRODUCTION

In an earlier paper[1] we have proven that the quantity

$$I = \tfrac{1}{2}(\rho\dot{x} - x\dot{\rho})^2 + \int^{x/\rho} f(\eta)d\eta + \int^{\rho/x} g(\eta)d\eta, \quad (1.1)$$

is constant if ρ satisfies the equation

$$\ddot{\rho} + \omega^2(t)\rho = f(x/\rho)/(\rho^2 x), \quad (1.2)$$

and x satisfies the auxiliary equation

$$\ddot{x} + \omega^2(t)x = g(\rho/x)/(x^2\rho), \quad (1.3)$$

where ω^2, f, and g are arbitrary functions and the overdots indicate differentiation with respect to t. The proof of this result, following Ermakov,[2] can be obtained by eliminating ω^2 between (1.2) and (1.3) and then carrying out a few simple manipulations to arrive at the result that (1.1) is constant. We refer to (1.2) and (1.3) as an Ermakov pair of equations and call I the Ermakov invariant associated with this pair. Several cases of these equations have proven useful for describing various physical systems with explicit time dependence. For example, the case $f = x/\rho$, $g = 0$ leads to the Lewis invariant[3]

$$I = \tfrac{1}{2}[(\rho\dot{x} - x\dot{\rho})^2 + (x/\rho)^2], \quad (1.4)$$

which can be used to give an exact quantum treatment of the time-dependent harmonic oscillator.[4]

Katzin and Levine[5] and Lutzky[6] derived the invariant (1.4) and the auxiliary equation for the time-dependent harmonic oscillator starting from the Lagrangian

$$L = \tfrac{1}{2}[\dot{\rho}^2 - \omega^2(t)\rho^2], \quad (1.5)$$

and employing Noether's theorem. In a later paper[7] we extended Lutzky's procedure and derived the Ermakov invariant and auxiliary equation for a Lagrangian of the form

$$L = \tfrac{1}{2}\left[\dot{\rho}^2 - \omega^2(t)\rho^2 + 2\sum_i G_i(t)F_i(\rho)\right]. \quad (1.6)$$

Applying Noether's theorem to this Lagrangian leads to

$$f(x/\rho) = \sum_i c_i (x/\rho)^{2m_i - 1}, \quad (1.7)$$

$$g(\rho/x) = k\rho/x, \quad (1.8)$$

where c_i, k, and $m_i \neq 0, -1/2, -1$ are constants (see Ref. 7 for details). The Ermakov invariant follows by employing (1.1), with the forms (1.7) and (1.8). An advantage of the Noether theorem approach is that one obtains both the Ermakov invariant and auxiliary equation starting from a given equation of motion. For example, starting from the Lagrangian for the equation of motion

$$\ddot{\rho} + \omega^2(t)\rho = \sum_i G_i dF_i/d\rho, \quad (1.9)$$

we obtain both the auxiliary equation

$$\ddot{x} + \omega^2(t)x = k/x^3, \quad (1.10)$$

and the Ermakov invariant for this pair. In the Ermakov-type calculation, e.g. the elimination of ω^2 between (1.1) and (1.2), one must know both equations before one can derive the Ermakov invariant. For a general equation of motion one does not always know the auxiliary equation (or equations, since there may be an infinite number of different auxiliary equations). A disadvantage of the Noether theorem approach is that one obtains only special cases of the general results (1.2) and (1.3).

In view of the foregoing comments one would like a constructive procedure for starting with a given nonlinear equation of motion and deriving the auxiliary equation and Ermakov invariant. We present such a procedure for the Ermakov pair (1.2) and (1.3) in Sec. II of this paper. The hints for developing this procedure are contained in a paper by Bandić.[8]

The techniques used in the proofs in Sec. II suggest a general method of relating the solutions of (1.2) and (1.3) using the Ermakov invariant. We shall develop this general relationship in Sec. III. In Sec. IV we present explicit examples of the use of the theory of Sec. III to obtain solutions of one of the equations (1.2) or (1.3) in terms of solutions of the other equation. These results are sometimes referred to as nonlinear superpostion laws in the literature.[9] In Sec. V we present our conclusions and suggestions for further work.

Before introducing the analysis of Sec. II we present a simple generalization of the Ermakov system (1.1), (1.2), and (1.3). This system can easily be converted into a form having linear friction. To prove this assertion we change to a new independent variable s defined by

$$\frac{ds}{\theta(s)} = dt, \quad (1.11)$$

where $\theta(s)$ is an arbitrary function. Equations (1.2) and (1.3)

 0022-2488/80/071583-05$1.00

are converted into the equations

$$\frac{d^2\rho}{ds^2} + p(s)\frac{d\rho}{ds} + \Omega^2(s)\rho = \frac{f(x/\rho)}{(\rho^2 x\theta^2)}, \tag{1.12}$$

and

$$\frac{d^2x}{ds^2} + p(s)\frac{dx}{ds} + \Omega^2(s)x = \frac{g(\rho/x)}{(x^2\rho\theta^2)}, \tag{1.13}$$

where

$$p(s) = \frac{1}{\theta}\frac{d\theta}{ds}, \tag{1.14}$$

$$\Omega^2(s) = \frac{\omega^2}{\theta^2}. \tag{1.15}$$

The Ermakov invariant (1.1) is converted into the form

$$I = \frac{1}{2}\theta^2\left(\rho\frac{dx}{ds} - x\frac{d\rho}{ds}\right)^2 + \int^{x/\rho} f(\eta)d\eta + \int^{\rho/x} g(\eta)d\eta. \tag{1.16}$$

Thus the Ermakov set with linear friction, (1.12), (1.13), and (1.16), can be derived directly from the Ermakov set without friction. For this reason we shall work with the friction-free case for simplicity and without loss of generality. Lutzky in Ref. (10) has noted the case with friction.

II. Derivation of the Ermakov System[8]

Consider the differential equation

$$\ddot{\rho} + \omega^2(t)\rho = q_1(t)f(q_2(t)\rho), \tag{2.1}$$

where q_1, q_2, and f are arbitrary functions along with the related equation

$$\ddot{x} + \omega^2(t)x = q_3(t)g(q_4(t)x), \tag{2.2}$$

where q_3, q_4, and g are arbitrary functions. We seek conditions on q_1, q_2, q_3, and q_4 which allow the solutions of (2.1) and (2.2) to be related by the equation

$$\rho = xr, \tag{2.3}$$

where r is required to satisfy an ordinary differential equation of the type

$$\frac{d^2r}{d\tau^2} = F(r), \tag{2.4}$$

where the variable τ and the function F will be defined below. Thus, the function r, which relates the solutions of (2.1) and (2.2), satisfies Newton's law for a one-dimensional system.

In order to derive the form of F we substitute (2.3) in (2.1) to obtain the equation

$$r(\ddot{x} + \omega^2 x) + x\ddot{r} + 2\dot{x}\dot{r} = q_1 f(q_2\rho). \tag{2.5}$$

We next use (2.2) in the first term in (2.5) which yields

$$rq_3 g(q_4 x) + x\ddot{r} + 2\dot{x}\dot{r} = q_1 f(q_2\rho). \tag{2.6}$$

Now by changing to a new independent variable τ defined by

$$d\tau = dt/x^2, \tag{2.7}$$

we convert (2.6) to the form

$$rq_3 x^3 g(q_4 x) + r'' = x^3 q_1 f(q_2 rx), \tag{2.8}$$

where the prime denotes differentiation with respect to τ. We now require that equation (2.8) be of the form (2.4). This implies the functions q_1, q_2, q_3, and q_4 have the form

$$q_1 = 1/(x^3 r^3), \tag{2.9a}$$

$$q_2 = 1/(xr^2), \tag{2.9b}$$

$$q_3 = 1/(x^3 r), \tag{2.9c}$$

and

$$q_4 = r^2/\rho. \tag{2.9d}$$

These expressions for the q's are not unique, since we could multiply any of them by any function of r and (2.8) would still have to form (2.4). However, since the functions f and g in (2.8) are arbitrary the choices (2.9a–d) do not introduce any loss of generality; we have made these choices for simplicity in later results. With these choices for the q's, (2.8) becomes

$$r'' = f(1/r)/r^2 - g(r). \tag{2.10}$$

This equation has the immediate energy integral

$$\tfrac{1}{2}r'^2 + V(r) = I = \text{const}, \tag{2.11}$$

where the potential energy $V(r)$ is

$$V(r) = \int^{1/r} f(\eta)d\eta + \int^{r} g(\eta)d\eta. \tag{2.12}$$

Next, using the above expressions for the q's in the original equations (2.1) and (2.2), and using (2.3) to eliminate r, we obtain

$$\ddot{\rho} + \omega^2\rho = f(x/\rho)/(\rho^2 x), \tag{2.13}$$

$$\ddot{x} + \omega^2 x = g(\rho/x)/(x^2\rho). \tag{2.14}$$

These last equations are precisely the Ermakov pair (1.2) and (1.3). The energy integral, (2.11), when expressed in terms of ρ/x and t, becomes the Ermakov invariant (1.1). Thus we have given a constructive derivation of the Ermakov system. *The assumption that $r = \rho/x$ satisfy a Newton's law equation,* (2.4), *implies that the pair of equations* (2.1) *and* (2.2) *are an Ermakov pair.* The energy associated with the equation satisfied by r is exactly the Ermakov invariant for the related pair of equations.

III. NONLINEAR SUPERPOSITION—GENERAL THEORY

We now assume that we have the Ermakov system (1.1), (1.2), and (1.3). Using the results of the previous section, we can write the Ermakov invariant (1.1) in the form (2.11)

$$\tfrac{1}{2}r'^2 + V(r) = I, \tag{3.1}$$

where $r = \rho/x$ and $V(r)$ is given by (2.12). The derivative in (3.1) is with respect to τ where t and τ are related by (2.7). Equation (3.1) can be integrated immediately to yield

$$\tau + c = \frac{1}{\sqrt{2}}\int \frac{dr}{[1 - V(r)]^{1/2}}, \tag{3.2}$$

where c is a constant of integration. If we can perform the integral in (3.2) then we obtain the general solution for r, the two integration constants being I and c. Thus, if we know any particular solution to the x equation, say x_1, then we obtain the general solution to the ρ equation as

$$\rho = x_1(t)r\left(\int \frac{dt}{x_1^2} + c, I\right), \tag{3.3}$$

after eliminating τ through (2.7). Clearly if we know a particular solution to the ρ equation and can solve for r via (3.2) then we obtain the general solution to the x equation as

$$x = \rho/r. \tag{3.4}$$

Thus, the nonlinear superposition works either way in x and ρ. Equation (3.3) and (3.4) represents a nonlinear superposition law for the Ermakov pair (1.2) and (1.3). Lutzky, in Ref. 6, used the procedure discussed above to obtain nonlinear superposition with the Lewis invariant (1.4). He also used the same general procedure to derive nonlinear superposition for a particular nonlinear equation discussed in Ref. 7.[10] We shall discuss these results in the next section. The general nonlinear superposition law represented by (3.3) does not seem to have been previously mentioned.

The Ermakov invariant provides a link between the two equations. It also allows one to derive the nonlinear superposition law (3.3). Although the Ermakov pair (1.2) and (1.3) have a special form, they still represent many equations of motion since f and g are arbitrary functions. Thus, we have a general nonlinear superposition law for a large class of nonlinear equations of motion.[9]

IV. NONLINEAR SUPERPOSITION—EXAMPLES

As a first example of the general theory developed in the previous section we choose the Ermakov–Pinney equation, $f = x/\rho$, and the time-dependent harmonic oscillator equation for the auxiliary equation $g = 0$, i.e.,

$$\ddot{\rho} + \omega^2(t)\rho = 1/\rho^3, \tag{4.1}$$

$$\ddot{x} + \omega^2(t)x = 0. \tag{4.2}$$

The Ermakov invariant for this case is just the Lewis invariant. The nonlinear superposition law for this Ermakov pair has been discussed many times starting with Ermakov in Ref. 2. The result is that if x_1 and x_2 are linearly independent solutions of (4.2) with Wronskian $W = x_1\dot{x}_2 - \dot{x}_1x_2$, then the general solution to (4.1) can be written

$$\rho = (Ax_1^2 + Bx_2^2 + 2Cx_1x_2)^{1/2}, \tag{4.3}$$

where A, B, and C are constants related by

$$AB - C^2 = 1/W^2. \tag{4.4}$$

Here we shall derive (4.3) using the general theory of the previous section. For the potential $V(r)$ for the pair (4.1) and (4.2) we obtain

$$V(r) = \int^{1/r} f(\eta)d\eta = 1/2r^2. \tag{4.5}$$

Equation (3.2) can be integrated and solved for r to yield

$$r = (1/2I + 2I(\tau + c)^2)^{1/2}. \tag{4.6}$$

The nonlinear superposition law then yields

$$\rho = xr = x[1/2I + 2I(\tau + c)^2]^{1/2}. \tag{4.7}$$

We may pick x to be a particular solution x_1, and write (4.7) in the form

$$\rho = \left[\left(\frac{1}{2I} + 2Ic^2\right)x_1^2 + \frac{2I}{W^2}x_2^2 + \frac{4Ic}{W}x_1x_2\right]^{1/2}, \tag{4.8}$$

where x_2 is a solution of (4.2) which has Wronskian W with x_1. The solution x_2 is introduced into (4.7) with the Wronskian formula,

$$x_2 = Wx_1\int \frac{dt}{x_1^2} = Wx_1\tau, \tag{4.9}$$

which defines a second linearly independent solution of (4.2) from a given solution x_1. Solution (4.8) is just the known nonlinear superposition law (4.3). Although the result (4.8) is not new we believe the simple proof we have given of this result is new. The other proofs of (4.8) use an algebraic process first discussed by Ermakov. In this procedure one writes two Ermakov invariants for x_1 and x_2, keeping the ρ solution the same. Then one eliminates $\dot{\rho}$ between the resulting invariants. This procedure leads, after considerable algebra, to (4.8). This algebraic method works when the Ermakov pair are uncoupled, i.e., essentially only for the pair (4.1) and (4.2).

As a simple variation of the previous example consider the Ermakov pair

$$\ddot{\rho} + \omega^2(t)\rho = 0, \tag{4.10}$$

$$\ddot{x} + \omega^2(t)x = 1/x^3, \tag{4.11}$$

or $f = 0$, $g = \rho/x$. This pair leads to a different form of the superposition law. Here the potential is

$$V(r) = \int^{(r)} g(\eta)d\eta = r^2/2. \tag{4.12}$$

The solution for r yields

$$r = \sqrt{2I}\sin(\tau + c) = \rho/x. \tag{4.13}$$

The result (4.13) was obtained previously by Lutzky.[6] Equation (4.13) is not as useful as (4.7) in our previous treatment because here the variable τ involves the solution to the nonlinear equation (4.11).

As another example we mention the Ermakov pair discussed in Ref. 7. These equations are a special case of (1.7) and (1.8) derived from Noether's theorem, and have the form

$$\ddot{\rho} + \omega^2(t)\rho = c_1x^{m-4}\rho^{1-m} + c_2x^{2m-4}\rho^{1-2m}, \tag{4.14}$$

$$\ddot{x} + \omega^2(t)x = k/x^3, \tag{4.15}$$

or

$$f(x/\rho) = c_1(x/\rho)^{m-3} + c_2(x/\rho)^{2m-3}, \tag{4.16}$$

$$g(\rho/x) = k\rho/x. \tag{4.17}$$

Here the potential V is

$$V(r) = \tfrac{1}{2}kr^2 + \frac{c_1}{(m-2)r^{m-2}} + \frac{c_2}{2(m-1)r^{2(m-1)}}. \tag{4.18}$$

The integral (3.2) for r can be performed in simple form only for $I = 0$ for $V(r)$ given by (4.18). Lutzky[10] has considered this case with $k \neq 0$ and using the associated Ermakov invariant and the theory of Sec. III, has derived the particular solution of (4.14) found earlier by Reid [see Ref. 7, Eq. (3.7)]. The resulting solution, (3.7) of Ref. 7, is only a particular solution since one must set $I = 0$ to do the r integral. In carrying out his derivation Lutzky made use of the following particular solution to the auxiliary equation (4.15)

$$x = \left(\frac{-4k}{W^2}\right)^{1/4}(x_1x_2)^{1/2}, \tag{4.19}$$

where x_1 and x_2 are linearly independent solutions of the time-dependent harmonic oscillator equation with Wronskian W. We shall not present further details of this example but shall refer the reader to Refs. 7,10.

As another example of the system (4.14), (4.15), we consider the case $k = 0$ for the Ermakov pair (4.14) and (4.15). The potential energy is (4.18) with $k = 0$. As in the previous example the r integral (3.2) can be performed in simple form only for the case $I = 0$. Performing the integral and solving for r yields

$$r = \rho/x = (a\tau^2 + 2ac\tau + ac^2 + b)^{1/m}, \tag{4.20}$$

here the constants a and b are defined by

$$a = m^2c_1/(2(2-m)), \tag{4.21}$$

$$b = c_2(m-2)/(c_1(1-m)). \tag{4.22}$$

By picking any particular solution x_1 we obtain a particular solution to the ρ equation in the form

$$\rho = \left(\frac{a}{W^2}x_1^{m-2}x_2^2 + \frac{2ac}{W}x_1^{m-1}x_2 + (ac^2+b)x_1^m\right)^{1/m}, \tag{4.23}$$

where τ has again been eliminated in favor of x_2 through the Wronskian theorem (4.9). In summary, (4.23) is a particular solution to

$$\ddot{\rho} + \omega^2(t) = c_1{x_1}^{m-4}\rho^{1-m} + c_2{x_1}^{2m-4}\rho^{1-2m}, \tag{4.24}$$

if x_1 and x_2 are linearly independent solutions of

$$\ddot{x} + \omega^2(t)x = 0, \tag{4.25}$$

with Wronskian W. Thus (4.23) is a nonlinear superposition law for the Ermakov pair (4.24) and (4.25).

We shall not work out further examples in this paper but shall let the above examples serve to illustrate the nonlinear superposition law (3.3).

It seems clear from the form of (3.3) that explicit superposition rules such as (4.8) or (4.23) will obtain only when the x-equation does not contain ρ. From (1.3) we see that the only uncoupled case occurs when $g = k\rho/x$, the resulting x-equation being

$$\ddot{x} + \omega^2(t)x = k/x^3. \tag{4.26}$$

The only viable auxiliary equation is, therefore, the Ermakov–Pinney equation: the linear auxiliary equation follows with $k = 0$. However, even when both equations of an Ermakov pair are coupled with each other, the functions $\rho = xr$ or $x = \rho/r$, along with integral (3.2) for r, still provide a general solution to one member of the Ermakov pair in terms of a particular solution to the other member. We may say in this coupled case that the nonlinear superposition is of an implicit form.

V. CONCLUSIONS

In this paper we have extended our study of the Ermakov system,

$$\ddot{\rho} + \omega^2(t)\rho = f(x/\rho)/(\rho^2x), \tag{5.1}$$

$$\ddot{x} + \omega^2(t)x = g(\rho/x)/(x^2\rho), \tag{5.2}$$

$$I = \tfrac{1}{2}(\rho\dot{x} - x\dot{\rho})^2 + \int^{x/\rho} f(\eta)d\eta + \int^{\rho/x} g(\eta)d\eta, \tag{5.3}$$

in two important ways. First we presented a constructive procedure for arriving at the Ermakov system starting from equations of motion having the general form (2.1) and (2.2). The central idea in this procedure is to require the ratio $r = \rho/x$ to satisfy an equation of the form of (2.4), or Newton's second law. This requirement forces the general equations (2.1) and (2.2) to be the Ermakov pair (5.1) and (5.2). The equation for r has a simple energy integral and the energy is exactly the Ermakov invariant (5.3). The second major result is that since the energy equation for r (2.11) can be solved, the Ermakov invariant provides a general connection between the solutions of the form

$$\rho = xr\left(\int \frac{dt}{x^2} + c, I\right). \tag{5.4}$$

We refer to (5.4) as a nonlinar superposition law for the Ermakov pair (5.1) and (5.2). We have given several examples showing how this formal nonlinear superposition law gives explicit nonlinear superposition laws in special cases. Lutzky was the first to exploit this procedure, in Ref. 10. Although this method probably only gives explicit nonlinear superposition laws when the x equation has the Ermakov–Pinney form (4.26) there are still a large number of possible cases because the ρ equation contains the arbitrary function f.

As a final point we mention the following interesting fact concerning the Ermakov set (5.1)–(5.3). In the Ermakov derivation of the invariant (5.3) one eliminates $\omega^2(t)$ between the two equations (5.1) and (5.2) and carries out a few simple manipulations to arrive at invariant (5.3). We usually think of $\omega^2(t)$ as a given function of the independent variable t. However, a little reflection on this derivation of (5.3) shows that in fact ω^2 can depend on the dependent variables x,ρ in an arbitrary way, i.e., in the form $\omega^2(\rho,\dot{\rho},\ddot{\rho},\ldots;x,\dot{x},\ddot{x},\ldots;t)$ and the Ermakov elimination procedure is still valid. This means that formally, any differential equation can be written as a member of an Ermakov pair

$$\ddot{\rho} + \omega^2(\rho,\dot{\rho},\ddot{\rho},\cdots;x,\dot{x},\ddot{x},\cdots;t)\rho = f(x/\rho)/(\rho^2x), \tag{5.5}$$

$$\ddot{x} + \omega^2(\rho,\dot{\rho},\ddot{\rho},\cdots;x,\dot{x},\ddot{x},\cdots;t)x = g(\rho/x)/(x^2\rho), \tag{5.6}$$

with I given by (5.3) being constant. This may not lead to important results in all cases, but it is at least interesting that every ordinary differential equation belongs to an infinite number of Ermakov pairs. It also follows that the nonlinear superposition law (5.4) also holds for an arbitrary ordinary differential equation. Again we emphasize that the elimination procedure for a generalized ω^2 will probably not lead to practical results for every differential equation, however, we find it a somewhat remarkable result.

We believe, based on the results of this paper, along with those in the references, that Ermakov-type systems are a fruitful field for study.

ACKNOWLEDGMENTS

We would like to thank our friends John C. Zimmerman for his translation of Ermakov's paper and Morton Lutzky for sending us copies of his papers prior to publication.

[1]J. R. Ray and J. L. Reid, Phys. Lett. A **71**, 317 (1979).
[2]V. P. Ermakov, Univ. Izv. Kiev **20**, No. 9, 1 (1880).
[3]H. R. Lewis, Jr., Phys. Rev. Lett. **18**, 510 (1967).
[4]H. R. Lewis, Jr. and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[5]G. H. Katzin and J. Levine, J. Math. Phys. **18**, 1267 (1977).
[6]M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[7]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979).
[8]I. Bandić, Bull. Math. Soc. Math. R. S. Roumanie **11**, (59), 6 (1967).
[9]W. F. Ames, *Nonlinear Equations in Abstract Spaces*, (Academic, New York, 1978), pp. 43–66.
[10]M. Lutzky, "An Application of Ray–Reid Invariants," preprint.

On the intrinsic geometry of certain nonlinear equations: The sine-Gordon equation

F. J. Chinea
Dept. de Métodos Matemáticos, Facultad de Ciencias Físicas, Universidad de Madrid, Madrid-3, Spain
Dept. de Física Fundamental, UNED, Madrid-3, Spain

(Received 27 November 1979; accepted for publication 22 February 1980)

The classical relation between two-dimensional spaces of constant curvature and certain nonlinear partial differential equations is formulated in group-theoretic terms by means of the underlying semisimple isometry group. Rather than working with the metric and curvature in the given constant curvature space, it is then possible to consider the equivalent system consisting of a pair of first-order partial differential equations in flat space for two so(2,1) vectors satisfying a pair of SO(2,1)-invariant algebraic constraints. Such equations determine a SL(2,R) principal bundle with flat connection. The construction is carried out in detail for the case of the sine-Gordon equation. The connection is shown to give rise to a spectral problem of the inverse scattering type by means of a gauge transformation. Bäcklund transformations are shown to be automorphisms of the connection characterized by a certain so(2,1) null vector. The generation of solutions of the nonlinear equation from known ones is seen to be determined by gauge transformations leaving invariant a fixed set of gauge conditions.

1. INTRODUCTION

Some geometric ideas developed during the last century are at the root of many concepts now widely used in the study of certain nonlinear partial differential equations of interest in physics. In recent years, comparatively more emphasis has been put in the development and use of very powerful analytical techniques bearing on the functional analysis aspects of the subject, among which the method of the inverse scattering is especially important. Quite recently, however, there has been a simultaneous and renewed interest in the geometric approach.[1-6]

In the present work some of these geometric aspects are developed. The point of view that the central feature is the existence of an underlying isometry group is adopted.

The more prominent role given here to the group structure, as opposed to the purely differential-geometric concepts, may be supported by the following reasons:

(i) The generalization to higher dimensions of constructions based on spaces of constant curvature seems unlikely, as a consequence of the additional differential constraints forced on the metric by the constant curvature condition in such dimensions.

(ii) Such a structure is very powerful in terms of finding exact solutions of a given equation. Its absence renders such a task almost impossible.

(iii) A natural generalization of the quantization procedure for linear equations seems to require a certain invariance group in order to provide a particle interpretation for quantized fields obeying nonlinear equations.

In this spirit, the classical relation among certain nonlinear equations in $1+1$ dimensions and two-dimensional Riemannian spaces of nonvanishing constant curvature[7] is formulated in terms of a maximal (three-parametric) isometry group, whose existence is assured by the condition of constant curvature.

Section 2 is concerned with the geometric setting. Rather than working with the two-dimensional metric and the Killing equation it satisfies, one deals with the equivalent problem of considering a pair of Lie algebra structure equations satisfied by two 3-vectors p and q and the algebraic constraints among such vectors. The equivalence of the two approaches is made possible by the fact that the isometry group is semisimple. Section 3 deals with the relation among the said constraints and the type of nonlinear equation obtained. It is also found that the integrability condition for the structure equations is precisely the desired nonlinear equation.

The consequences of this approach are developed in detail in the following sections for the sine-Gordon system. In Sec. 4 the basic equations are written down. They take the form of first-order evolution equations for an orthonormal frame of so(2,1) vectors. Interestingly enough, one of the vectors in the frame is seen to obey the equations for a O(2,1)-invariant chiral field.

Representing the mentioned equations for the so(2,1) frame in SL(2,R) spinor language, a SL(2,R) principal bundle with connection is constructed in Sec. 5. It has been previously recognized that the spectral problem corresponding to the inverse-scattering method for the sine-Gordon equation and certain others may be put in the language of bundle connections.[5] In the present approach, the bundle and connection, as well as the vanishing curvature condition for such a connection, are *derived* as a consequence of the equations in Sec. 4. The resulting connection is shown to give rise to the spectral problem by means of a gauge transformation.

In agreement with the basic aim of emphasizing the group-theoretic aspects of the problem, all geometric constructions are intrinsic and make no use of embeddings of the original two-dimensional manifold in a higher-dimensional space. In particular, the Bianchi–Bäcklund transformations,

 0022-2488/80/071588-05$1.00

which were originally derived by a rather complicated and clever construction using a three-dimensional embedding space,[7] are here obtained in an intrinsic fashion. In Sec. 6, the Bäcklund transformations for the sine-Gordon equation are shown to be a particular case of invariance transformations (gauge transformations leaving the connection invariant).

Finally, the matter of generation of solutions from known ones is shown in Sec. 7 to be equivalent to the study of gauge transformations that leave a given set of gauge conditions invariant. Such gauge conditions, which turn out to consist of both linear and quadratic equations, are written down explicitly.

2. GEOMETRIC DESCRIPTION: SEMISIMPLE ISOMETRY GROUPS

Let $\mathscr{M}$ be a complete, connected, simply connected two-dimensional (pseudo-) Riemannian manifold with metric g and constant scalar curvature R. The latter condition implies the existence of a maximal transitive isometry group on $\mathscr{M}$.[8,9] In particular, in the case of constant negative curvature, the connected isometry group is isomorphic to SO(2,1). As a consequence, there will exist three independent Killing fields of infinitesimal isometries. Let them be denoted by $\xi^{i}_{(\mu)}$, where the index μ describes a given field ($\mu = 1, 2, 3$) and the index i a given contravariant component ($i = 1, 2$). In the case $R \neq 0$, the metric g may be expressed in terms of the Killing fields $\xi_{(\mu)}$ by means of the following relation[10]:

$$g^{ij} = \sigma K^{\mu\nu}\xi^{i}_{(\mu)}\xi^{j}_{(\nu)}\,, \tag{2.1}$$

where σ is a constant and $K^{\mu\nu}$ the Killing–Cartan metric for the corresponding isometry group. The possibility of writing such an expression clearly depends on the fact that the isometry group is semisimple for $R \neq 0$; it has no analog in the $R = 0$ case, which corresponds to the (pseudo-) Euclidean group in two dimensions. Equation (2.1) may be shown to hold by checking its validity for a particular model of a two-dimensional space of constant curvature, such as a sphere (hyperboloid) in $\mathbb{R}^3$ with Euclidean (Minkowski) metric, and then using the fact that all spaces of equal constant curvature and dimension are isometric.[8]

The Killing equations

$$\mathscr{L}_{\xi} g = 0$$

are trivially satisfied by g as given by (2.1), taking into account the structure equations for the Lie algebra of the isometry group and the expression for the Killing–Cartan form in terms of the structure coefficients. In particular, in the case of a SO(2,1) isometry group, the following holds

$$\left(\xi_{(\mu)},\xi_{(\nu)}\right) = \epsilon_{\mu\nu}{}^{\alpha}\xi_{(\alpha)}\,, \tag{2.2}$$

where $\epsilon_{\mu\nu\alpha}$ is the Levi–Civita tensor in three dimensions, greek indices being raised and lowered by means of $K^{\mu\nu}$ and its inverse and $K^{\mu\nu} = \mathrm{diag}\,(+\ +\ -)$.

Equation (2.2) may be rewritten in three-dimensional vector notation by defining

$$p_{\mu} = \xi^{1}_{(\mu)}, \quad q_{\mu} = \xi^{2}_{(\mu)}\,. \tag{2.3}$$

Equation (2.2) now reads

$$p = p_u \times p + p_v \times q, \quad q = q_u \times p + q_v \times q\,, \tag{2.4}$$

where the cross product of two vectors a and b is defined by

$$(a \times b)^{\alpha} = \epsilon_{\mu\nu}{}^{\alpha} a^{\mu} b^{\nu}$$

and the subindices u and v respectively denote partial differentiation with respect to coordinates x^1 and x^2 in $\mathscr{M}$.

3. NONLINEAR WAVE EQUATIONS AND CONSTRAINTS

Let us now suppose that the metric g on $\mathscr{M}$ depends on a certain field $\varphi(u,v)$ defined on $\mathscr{M}$ (and possibly on its derivatives). This will be denoted by

$$ds^2 = g_{ij}[\varphi]\,dx^i\,dx^j\,.$$

The scalar curvature R will then be given as a functional expression in terms of φ and its derivatives. The condition of constant curvature,

$$R[\varphi] = \text{constant},$$

is then a partial differential equation for φ. Such an equation is obviously of second order if g depends on the field φ only, and not on its derivatives. In particular cases, the resulting expression takes the form of a nonlinear wave equation

$$\Box\varphi = \mathscr{F}(\varphi),$$

where $\mathscr{F}(\varphi)$ is a certain functional of φ and $\Box$ the two-dimensional d'Alembertian operator in *flat* space. It is a classical fact[7] that the sine-Gordon equation

$$\varphi_{uv} = \sin\varphi \tag{3.1}$$

may be obtained in this fashion from the metric

$$ds^2 = du^2 + 2\cos\varphi\, du\, dv + dv^2, \tag{3.2}$$

where u and v play the role of light-cone coordinates in the corresponding equation (3.1).

Similarly, it is easy to see that the Liouville equation, characterized by $\mathscr{F}(\varphi) = -e^{\varphi}$, corresponds to the metric

$$ds^2 = e^{\varphi}\, du\, dv\,. \tag{3.3}$$

A number of other equations, such as the Korteweg–de Vries equation, may be associated with spaces of constant curvature.[6]

The preceeding correspondence among spaces of constant curvature and nonlinear partial differential equations may be translated into the language of Killing fields, which will play a central role in the sequel due to the possibility of expressing g in terms of such vector fields by means of (2.1). From (2.1) and (2.3), one gets

$$g^{11} = \sigma p^2, \quad g^{12} = \sigma\, p\cdot q, \quad g^{22} = \sigma q^2 \tag{3.4}$$

(where the scalar product of two 3-vectors a and b is defined by $a\cdot b = K_{\mu\nu}a^{\mu}b^{\nu}$).

It is clear that the vectors p and q must satisfy certain constraints in order for g to give rise to a particular nonlinear equation. For equations of the second order, such constraints are algebraic. For instance, inspection of (3.2) and (3.3) shows that in these examples the constraints are[11]

$$p^2 = 0, \quad q^2 = 0 \quad \text{(Liouville)}, \tag{3.5}$$

$$q^2(1-p^2) + (p\cdot q)^2 = 0, \quad p^2(1-q^2) + (p\cdot q)^2 = 0 \quad \text{(sine-Gordon)}. \tag{3.6}$$

The complete information about a particular nonlinear equation and its symmetries is thus given by Eqs. (2.4) together with an appropriate pair of invariant constraints satisfied by p and q. As p and q depend on the field φ through (3.4), Eqs. (2.4) will contain expressions in the field φ and its first derivatives. Their integrability condition is precisely the requirement that φ satisfy the desired equation. It is easy to see that this is the case by means of the following consideration: The integrability conditions for (2.4) are equivalent to the requirement of the existence of three Killing fields that satisfy an so(2,1) algebra. But we have seen that the existence of a corresponding maximal isometry group is equivalent to the condition that $\mathscr{M}$ has constant negative curvature. By construction, the latter is satisfied if and only if φ is a solution of the given nonlinear equation.

As a consequence, we shall deal in the following with Eqs. (2.4) and the appropriate constraints rather than with metrical quantities such as g and R. Notice that it is crucial in this respect that the derivatives involved in (2.2) may be considered as *ordinary* ones, due to the antisymmetry of (2.2) in the covariant derivatives of the $\xi_{(\mu)}$.

4. THE SINE-GORDON EQUATION

In what follows, the preceeding approach is developed in the specific case of the sine-Gordon equation. For convenience, variables $t = u - v$ and $x = u + v$ will be used. The sine-Gordon equation

$$\Box\varphi = -\sin\varphi \tag{4.1}$$

then corresponds to the metric

$$ds^2 = \sin^2\tfrac{1}{2}\varphi\, dt^2 + \cos^2\tfrac{1}{2}\varphi\, dx^2. \tag{4.2}$$

The algebraic constraints satisfied by the transformed variables $\tilde{p}$ and $\tilde{q}$ (denoted in the following by the same symbols p and q) are now

$$1/p^2 + 1/q^2 = 1, \quad p\cdot q = 0. \tag{4.3}$$

Introducing the vectors ζ, η, ω, defined by

$$\zeta = (p^2)^{-1/2}p, \quad \eta = (q^2)^{-1/2}q, \quad \omega = \zeta\times\eta \tag{4.4}$$

[where $(p^2)^{-1/2} = \sin\frac{1}{2}\varphi$ and $(q^2)^{-1/2} = \cos\frac{1}{2}\varphi$, due to (4.2) and (4.3)], the following relations hold:

$$\zeta^2 = \eta^2 = 1, \quad \omega^2 = -1, \quad \zeta\cdot\eta = \eta\cdot\omega = \omega\cdot\zeta = 0. \tag{4.5}$$

The vectors ζ, η, and ω thus form a trihedral, orthonormal with respect to the Killing–Cartan metric.

Equations (2.4) may be written in terms of the new variables ζ, η, and ω. Using (4.5), the resulting equations may be solved explicitly for the first derivatives of these variables and the following expressions are obtained:

$$\zeta_t = -\tfrac{1}{2}\varphi_x\eta, \quad \zeta_x = -\tfrac{1}{2}\varphi_t\eta + \cos\tfrac{1}{2}\varphi\,\omega,$$
$$\eta_x = \tfrac{1}{2}\varphi_t\zeta, \quad \eta_t = \tfrac{1}{2}\varphi_x\zeta - \sin\tfrac{1}{2}\varphi\,\omega. \tag{4.6}$$

The evolution of the trihedral is completely specified by (4.6). The remaining equations

$$\omega_t = -\sin\tfrac{1}{2}\varphi\,\eta, \quad \omega_x = \cos\tfrac{1}{2}\varphi\,\zeta \tag{4.7}$$

are just a consequence of the definition of ω and of conditions (4.5). Incidentally, (4.4) and (4.7) imply

$$p = (\omega_t^2)^{-\frac{1}{2}}(\omega_x^2)^{-\frac{1}{2}}\omega_x, \quad q = -(\omega_t^2)^{-\frac{1}{2}}(\omega_x^2)^{-\frac{1}{2}}\omega_t$$

and hence Eqs. (2.4) may be written as second-order equations for ω, where φ and its derivatives do not appear explicitly.

The so(2,1) vector ω admits a definite geometric interpretation as a vector in three-dimensional space. First notice that ω satisfies the following equations:

$$\omega^2 = -1,$$
$$\omega_t^2 + \omega_x^2 = 1, \tag{4.8}$$
$$\omega_t\cdot\omega_x = 0,$$

as implied by (4.5) and (4.7).

Suppose now that a two-surface is defined in three-dimensional space, with a Minkowski metric given by $K_{\mu\nu}$, by the condition

$$\omega^2(t,x) = -1. \tag{4.9}$$

The metric induced by $K_{\mu\nu}$ on the embedded surface (4.9) is

$$ds^2 = \omega_i\omega_j\, dx^i\, dx^j = \omega_t^2\, dt^2 + \omega_x^2\, dx^2$$
$$= \sin^2\tfrac{1}{2}\varphi\, dt^2 + \cos^2\tfrac{1}{2}\varphi\, dx^2,$$

where (4.7) and (4.8) have been taken into account. The original metric g on $\mathscr{M}$ is thus obtained. The vector ω, which was constructed by an intrinsic procedure in terms of the isometries of $\mathscr{M}$, turns out to represent the radius vector in $\mathbb{R}^3$ whose locus, when t and x take all possible values, determines an (at least local) embedding of $\mathscr{M}$ in $\mathbb{R}^3$ with a Minkowski metric.

Finally, it should be noticed that (4.8) are just the constraint equations characterizing a O(2,1)-invariant chiral field. In a very interesting article,[12] it has been shown how to derive the sine-Gordon equation from the O(3)-invariant chiral model. In the present context, the converse construction is obtained, namely, the O(2,1)-invariant field ω is constructed from the isometries of the manifold $\mathscr{M}$ associated with the sine-Gordon equation. A similar construction may be carried out for a SO(3)-invariant sine-Gordon surface $\mathscr{M}'$ with positive curvature, thus showing the complete equivalence of the sine-Gordon and chiral O(3) theories. The reason for our preference for working with a SO(2,1) invariance group rather than with the corresponding compact group SO(3) is merely technical and will become apparent in Sec. 6.

5. SL(2,R) BUNDLE ASSOCIATED WITH THE SINE-GORDON EQUATION

Equations (4.6) and (4.7) may be written as

$$A_t = MA, \quad A_x = NA, \tag{5.1}$$

with

$$M = \begin{pmatrix} 0 & -\frac{1}{2}\varphi_x & 0 \\ \frac{1}{2}\varphi_x & 0 & -\sin\frac{1}{2}\varphi \\ 0 & -\sin\frac{1}{2}\varphi & 0 \end{pmatrix},$$

$$N=\begin{pmatrix} 0 & -\tfrac{1}{2}\varphi_t & \cos\tfrac{1}{2}\varphi \\ \tfrac{1}{2}\varphi_t & 0 & 0 \\ \cos\tfrac{1}{2}\varphi & 0 & 0 \end{pmatrix}, \tag{5.2}$$

$$A=\begin{pmatrix} \zeta \\ \eta \\ \omega \end{pmatrix}. \tag{5.3}$$

The integrability condition for (5.1) is

$$(M_x - N_t + [M,N])A = 0$$

or, taking into account the linear independence of ζ, η, and ω, simply

$$M_x - N_t + [M,N] = 0\,. \tag{5.4}$$

The matrices M and N are elements of the Lie algebra so(2,1); they may be expressed as

$$M^{\mu}{}_{\nu} = 2\epsilon^{\mu}{}_{\alpha\nu}m^{\alpha}, \quad N^{\mu}{}_{\nu} = 2\epsilon^{\mu}{}_{\alpha\nu}n^{\alpha}, \tag{5.5}$$

with appropriate vectors m^{α}, n^{α}.

Equations (5.1)–(5.4) may be expressed in the corresponding SL(2,R) spinor representation for convenience. The correspondence is (the same symbols are used for the original and transformed variables):

$$A \to A = A^{\mu}\rho_{\mu}, \quad M \to M = m^{\mu}\rho_{\mu}, \quad N \to N = n^{\mu}\rho_{\mu}\,,$$

where the sl(2,R) matrices ρ_{μ} are defined by

$$\rho_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \rho_3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and satisfy the relation

$$\rho_{\mu}\rho_{\nu} = \epsilon_{\mu\nu}{}^{\alpha}\rho_{\alpha} + K_{\mu\nu}I\,,$$

where I is the 2×2 identity matrix. Equations (5.1)–(5.4) now read

$$A_t = [M,A], \quad A_x = [N,A], \tag{5.6}$$

$$M=\tfrac{1}{2}\begin{pmatrix} \sin\tfrac{1}{2}\varphi & -\tfrac{1}{2}\varphi_x \\ \tfrac{1}{2}\varphi_x & -\sin\tfrac{1}{2}\varphi \end{pmatrix},$$

$$N=\tfrac{1}{2}\begin{pmatrix} 0 & \cos\tfrac{1}{2}\varphi - \tfrac{1}{2}\varphi_t \\ \cos\tfrac{1}{2}\varphi + \tfrac{1}{2}\varphi_t & 0 \end{pmatrix}, \tag{5.7}$$

$$A=\begin{pmatrix} \zeta & \eta-\omega \\ \eta+\omega & -\zeta \end{pmatrix}, \tag{5.8}$$

$$M_x - N_t + [M,N] = 0\,. \tag{5.9}$$

Let us now look at the transformation properties of the matrices A, M, and N. The natural requirement that the transformed quantities again describe the evolution of a certain orthonormal frame will be imposed. In particular, the trihedral ζ, η, ω should transform in such a way that the orthogonality conditions (4.5) hold for the transformed vectors $\tilde{\zeta}$, $\tilde{\eta}$, and $\tilde{\omega}$. That amounts to transforming ζ, η, ω by means of a rotation in SO(2,1), or, equivalently, to transform A as given by (5.8) in the following way:

$$\tilde{A} = S^{-1}AS\,, \tag{5.10}$$

where S is a SL(2,R) matrix depending in general on the coordinates t and x.

The transformation rules for M and N are then derived from the required invariance of equations (5.6), giving as a result

$$\tilde{M} = S^{-1}MS - S^{-1}S_t, \quad \tilde{N} = S^{-1}NS - S^{-1}S_x\,. \tag{5.11}$$

It is an immediate consequence of (5.11) that the null trace condition which M and N should satisfy as elements of sl(2,R), as well as the integrability condition (5.9), are automatically satisfied by the transformed quantities $\tilde{M}$ and $\tilde{N}$. Equations (5.11) are obviously the transformation rules for the gauge fields in a non-Abelian gauge theory. In other words, M and N are the components of a connection in a principal SL(2,R) bundle with base $\mathcal{M}$. Condition (5.9) just asserts that the curvature of such a connection vanishes.

A given connection (M, N) gives rise to a spectral problem by means of a gauge transformation in the following manner: A scattering problem may be set up for the sine-Gordon equation by means of a certain pair of 2×2 matrices,[13] which in coordinates t-x and using the present notation read:

$$\tilde{M}=\tfrac{1}{2}\begin{pmatrix} \lambda - \dfrac{1}{4\lambda}\cos\varphi & -\tfrac{1}{2}(\varphi_t+\varphi_x) - \dfrac{1}{4\lambda}\sin\varphi \\ \tfrac{1}{2}(\varphi_t+\varphi_x) - \dfrac{1}{4\lambda}\sin\varphi & -\lambda + \dfrac{1}{4\lambda}\cos\varphi \end{pmatrix},$$

$$\tilde{N}=\tfrac{1}{2}\begin{pmatrix} \lambda + \dfrac{1}{4\lambda}\cos\varphi & -\tfrac{1}{2}(\varphi_t+\varphi_x) + \dfrac{1}{4\lambda}\sin\varphi \\ \tfrac{1}{2}(\varphi_t+\varphi_x) + \dfrac{1}{4\lambda}\sin\varphi & -\lambda - \dfrac{1}{4\lambda}\cos\varphi \end{pmatrix},$$

where λ is a spectral parameter. It is easy to see that these particular $(\tilde{M}, \tilde{N})$ and an arbitrary pair (M, N) given by (5.7) are in fact related by a gauge transformation (5.11). Notice that $\tilde{M}$ and $\tilde{N}$ are traceless and satisfy the vanishing curvature condition.[5] As a consequence, they must be of the form

$$\tilde{M} = -Q^{-1}Q_t, \quad \tilde{N} = -Q^{-1}Q_x\,,$$

for a certain nonsingular matrix Q with determinant equal to a constant (which may be obviously considered to be unity). Similarly,

$$M = -R^{-1}R_t, \quad N = -R^{-1}R_x$$

for a certain matrix R with $\det R = 1$. $(\tilde{M}, \tilde{N})$ is then obtained from (M, N) by a transformation of the type (5.11) with $S = R^{-1}Q$.

6. BÄCKLUND TRANSFORMATIONS AS INVARIANCE TRANSFORMATIONS OF THE CONNECTION

The connection (M, N) transforms as a one-form under coordinate transformations, as may be easily seen from (5.6). Going back to u-v coordinates, (5.7) and (5.9) read

$$M=\tfrac{1}{2}\begin{pmatrix} \sin\tfrac{1}{2}\varphi & \cos\tfrac{1}{2}\varphi - \tfrac{1}{2}\varphi_u \\ \cos\tfrac{1}{2}\varphi + \tfrac{1}{2}\varphi_u & -\sin\tfrac{1}{2}\varphi \end{pmatrix},$$

$$N=\tfrac{1}{2}\begin{pmatrix} -\sin\tfrac{1}{2}\varphi & \cos\tfrac{1}{2}\varphi + \tfrac{1}{2}\varphi_v \\ \cos\tfrac{1}{2}\varphi - \tfrac{1}{2}\varphi_v & \sin\tfrac{1}{2}\varphi \end{pmatrix}, \tag{6.1}$$

$$M_v - N_u + [M,N] = 0. \tag{6.2}$$

Consider now infinitesimal gauge transformations that leave M, N *invariant*. Putting

$$S \simeq I + \epsilon\xi$$

in Eq. (5.11) (where ξ is a traceless matrix), the following

equations are obtained to first order in ϵ:

$$\xi_u = [M,\xi\,], \quad \xi_v = [N,\xi\,]\,. \tag{6.3}$$

We shall consider in detail the case where ξ is a nilpotent matrix, $\xi^2 = 0$ (that is to say, ξ is the spinor representative of a null vector). The matrix ξ may be then represented as

$$\xi = a(\cos\tfrac{1}{2}\theta\rho_1 + \sin\tfrac{1}{2}\theta\,\rho_2 + \rho_3)\,,$$

where a and θ are functions of u and v. Equations (6.3) take the following form:

$$\tfrac{1}{2}(\varphi-\theta)_u = \sin\tfrac{1}{2}(\varphi+\theta), \quad \tfrac{1}{2}(\varphi+\theta)_v = \sin\tfrac{1}{2}(\varphi-\theta)\,, \tag{6.4}$$

$$(\ln a)_u = \cos\tfrac{1}{2}(\varphi+\theta), \quad (\ln a)_v = \cos\tfrac{1}{2}(\varphi-\theta)\,. \tag{6.5}$$

Equations (6.4) are just the Bäcklund transformations for the sine-Gordon equation (or, more precisely, the Bianchi transformations[7]—in order to get a one-parameter family of Bäcklund transformations, it is sufficient to use the invariance of the resulting sine-Gordon equation under the transformation $u \to ku$, $v \to k^{-1}v$). It is well known that the integrability conditions for (6.4) imply that both φ and θ satisfy the sine-Gordon equation. The integrability condition for (6.5) is satisfied by virtue of (6.4).

It is a rather remarkable fact that in the case $\xi^2 = 0$, S satisfies $S = I + \epsilon\xi$ *exactly*, with ϵ no longer restricted to being infinitesimal, and equations (6.3) become finite, exact equations. This explains the rather puzzling fact that Bäcklund transformations were classically obtained by a *finite* geometric construction, rather than as an infinitesimal deformation, the latter being the usual way to derive similar results in the differential geometry of surfaces.

Infinitesimal transformations generated by ξ with $\xi^2 \neq 0$ may also be considered. They also depend on two functions,[14] combinations of which satisfy the sine-Gordon equation, but their separation is not so clean as that of θ and a in (6.4) and (6.5). It is for this reason that we prefer to consider a SO(2,1)-invariant rather than a SO(3)-invariant theory. In the latter, due to the positive definite metric on SO(3), it is not possible to have a nontrivial matrix representing a vector with $\xi^2 = 0$. Again, this is related to the classical fact that surfaces of positive curvature do not admit Bäcklund transformations.[7]

7. GENERATION OF SOLUTIONS AND GAUGE CONDITIONS

The gauge transformations (5.11) play an important role in the theory. It is through them that the bundle connection found by considering the isometries of $\mathscr{M}$ leads to a problem of the inverse-scattering type. On the other hand, they include the Bäcklund transformations as a special case of invariance transformations.

Another point of great interest is connected with Eqs. (5.11): The generation of new solutions of the given nonlinear partial differential equation from known ones.[15] Given matrices M, N satisfying the trace condition and the integrability condition (5.9), new matrices $\tilde{M}$, $\tilde{N}$ satisfying the same properties may be obtained by means of an arbitrary S in (5.11). However, for the purposes of solution generation, one important requirement has to be added: $\tilde{M}$, $\tilde{N}$ have to be of the same from as M and N. This guarantees that it is possible to extract from $\tilde{M}$, $\tilde{N}$ a new field $\tilde{\varphi}$ satisfying the desired equation.[17] In order to formulate this requirement, the allowed class of matrices must be restricted by some additional conditions, which may be considered as "gauge conditions" on the connection. For the sine-Gordon connection, such conditions are

$$\begin{aligned} &(M+M^T)_u = (M+M^T)(M-M^T)\,,\\ &(N+N^T)_v = (N+N^T)(N-N^T)\,, \qquad (7.1)\\ &\rho_2(M+M^T)\rho_2 = N+N^T\,,\\ &(M+M^T)^2 = \tfrac{1}{4}\,, \qquad (7.2) \end{aligned}$$

where the symbol T denotes the transpose of a given matrix. Equations (7.1) and (7.2) are easily seen to be equivalent to requiring M, N to be of the form (6.1). Equations (7.1) lead to

$$\Box\varphi = -\gamma\sin\varphi\,,$$

where γ is a constant, while Eq. (7.2) normalizes the value of γ to unity.

ACKNOWLEDGMENTS

The author wishes to thank L. Abellanas, F. Guil, and L. Martínez Alonso for conversations.

[1] H.C. Wahlquist and F.B. Estabrook, J. Math. Phys. **16**, 1 (1975).
[2] R. Hermann, Phys. Rev. Lett. **36**, 835 (1976).
[3] G.L. Lamb, Jr., Phys. Rev. Lett. **37**, 235 (1976).
[4] F. Lund and T.D. Regge, Phys. Rev. D **14**, 1524 (1976).
[5] M. Crampin, F.A.E. Pirani, and D.C. Robinson, Lett. Math. Phys. **2**, 15 (1977).
[6] R. Sasaki, Nucl. Phys. B **154**, 343 (1979).
[7] L.P. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Dover, New York, 1960).
[8] L.P. Eisenhart, *Riemannian geometry* (Princeton U., Princeton, New Jersey, 1926).
[9] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vol. I (Interscience, New York, 1963).
[10] F.J. Chinea, in *Proceedings of the* VI *International School of Cosmology and Gravitation*, edited by P.G. Bergmann and V. de Sabbata (Plenum, New York, 1980).
[11] The value of the constant σ may be made equal to unity by appropriately choosing a particular constant value for R.
[12] K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[13] M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[14] Notice that Eqs. (6.3) imply ξ^2 = constant (using the fact that ξ^2 is a multiple of the identity for two-dimensional traceless matrices, and as a consequence ξ^2 commutes with M and N). Such a condition cuts down the number of independent functions in ξ from three to two.
[15] Cf. Ref. 16 for an alternative approach to the subject of solution generation by means of the isometries of $\mathscr{M}$.
[16] F.J. Chinea, Phys. Lett. A **72**, 281 (1979).
[17] Related systems, characterized by matrices satisfying Eq. (5.9), are dealt with in Refs. 18–20. The additional requirement on M and N is not one of form invariance, but of a certain analytic structure of these matrices.
[18] A.A. Belavin and V.E. Zakharov, Phys. Lett. B **73**, 53 (1978).
[19] V.E. Zakharov and A.V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1978).
[20] V.A. Belinskii and V.E. Zakharov, Sov. Phys. JETP **48**, 985 (1978).

Bäcklund transformations in several variables

Dean A. Payne
Applied Physics Laboratory, Johns Hopkins University, Laurel, Maryland 20810

(Received 18 November 1979; accepted for publication 29 February 1980)

A differential geometric method of constructing Bäcklund transformations for a second-order partial differential equation in n independent variables (n arbitrary) is developed and several examples worked out.

INTRODUCTION

Bäcklund transformations have appeared in the recent literature on those nonlinear partial differential equations which admit special solutions called solitons. Until now these transformations have been limited to differential equations in two independent variables. This paper describes a way to construct Bäcklund transformations for second-order partial differential equations in any number of independent variables. The technique is based on a differential geometric approach which views a partial differential equation as a submanifold of an appropriate jet space. In this approach a Bäcklund transformation between equations in n independent variables is a $3n$-dimensional submanifold M (with certain regularity conditions) of $J^1(R^n,R)\times J^1(R^n,R)$ where $J^1(R^n,R)$ is the space of 1-jets of maps from R^n to R. The differential equations related by the transformation arise from integrability conditions for the two differential systems on M inherited from the two contact system on the respective factors of $J^1(R^n,R)\times J^1(R^n,R)$.

1. BACKGROUND

For the sake of simplicity all manifolds and all functions will be assumed to be C^∞. Moreover, all arguments concerning the existence of solutions to partial differential equations will be local.

Let M be a real C^∞-differential manifold. Then $C(M)$ denotes the ring of C^∞-functions on M, $T(M)$ and $T^*(M)$ denote the tangent and cotangent bundles, and $\Gamma(M)$ and $\Gamma^*(M)$ denote the $C(M)$-modules of global C^∞-sections of $T(M)$ and $T^*(M)$. The r-jet of a map $f:M\to N$ between differential manifolds M and N at a point $p\in M$ will be denoted by $j^r_p(f)$. The set of all r-jets of maps of M into N forms a differential manifold which will be denoted $J^r(M,N)$. The r-graph of a map $f:M\to N$ is the map

$$j^r(f):M\to J^r(M,N)$$

defined by

$$j^r(f)(p)=j^r_p(f).$$

There are natural projections

$$\pi^{r,k}:J^r(M,N)\to J^k(M,N),\quad r\geqslant k \tag{1.1}$$

and

$$\alpha:J^r(M,N)\to M, \tag{1.2}$$

defined by

$$\pi^{r,k}(J^r_p(f))=j^k_p(f) \text{ and } \alpha(j^r_p(f))=p.$$

Definition[1]: A system of rth-order partial differential equations for maps of M into N is a differential manifold Σ together with a one–one differentiable map

$$i:\Sigma\to J^r(M,N). \tag{1.3}$$

A solution is a map

$$f:M\to N$$

such that the image of the r-graph $J^r(f)$ lies on the image of Σ under i.

$J^r(M,N)$ admits a globally defined Pfaffian system $\Omega^r(M,N)$ called the rth order contact system. Its importance is due to the following proposition.[1]

Proposition: Let $\eta:M'\to J^r(M,N)$ be a map such that

$$\lambda=\alpha\circ\eta:M'\to M$$

is diffeomorphism. Then a necessary and sufficient condition for

$$\eta\circ\lambda^{-1}:M\to J^r(M,N)$$

to be the r-graph of a map $\zeta:M\to N$ is

$$\eta^*\Omega^r(M,N)=0. \tag{1.4}$$

We shall only need the contact systems $\Omega^1(R^n,R)$ and $\Omega^2(R^n,R)$, defined on $J^1(R^n,R)$ and $J^2(R^n,R)$, respectively, by their generators as

$$\Omega^1(R^n,R)=\left\{dz-\sum_{i=1}^n p_i dx_i\right\}, \tag{1.5}$$

$$\Omega^2(R^n,R)=\begin{cases} dz-\sum_{i=1}^n p_i dx_i, \\ dp_i-\sum_{j=1}^n p_{ij}dx_j, \quad i=1,\ldots,n, \end{cases} \tag{1.6}$$

where x_i, z, and p_{ij} are global coordinates on $J^2(R^n,R)$, and x_i, z, p_i are global coordinates on $J^1(R^n,R)$. The coordinates are defined by

$$\begin{aligned} &x_i(j^r_p(f))=x_i(p), \quad \text{the } i\text{th coordinate of } p\in R^n, \\ &z(j^r_p(f))=f(p), \\ &p_i(j^r_p(f))=\frac{\partial f}{\partial x_i}(p), \\ &p_{ij}(j^r_p(f))=\frac{\partial^2 f}{\partial x_j\partial x_i}(p). \end{aligned} \tag{1.7}$$

In the development of Sec. 3 we shall need the following

 0022-2488/80/071593-10$1.00

results from the theory of Pfaffian systems. For the standard concepts of Pfaffian systems refer to Gardner.[1,2]

Definition: Let $\omega\in\Gamma^*(M)$ be a 1-form on a manifold M. We shall say that ω is regular at a point $p\in M$ if there exists some neighborhood U of p and some integer $r\geqslant 1$ such that

$$d\omega^r\wedge\omega=0 \text{ for all } q\in U$$

and (1.8)

$$d\omega^{r-1}\wedge\omega\neq 0 \text{ for all } q\in U.$$

We say that ω is regular in an open set V if it is regular at each point of V.

Theorem (*The Classical Pfaff Problem*[3]): Let $\omega\in\Gamma^*(M)$ be a regular 1-form at a point $p\in M$ so that there exists a neighborhood U of p such that

$$d\omega^r\wedge\omega=0 \text{ for all } q\in U$$

and

$$d\omega^{r-1}\wedge\omega\neq 0 \text{ for all } q\in U.$$

Then there exists a neighborhood V of p and functions $f_1,...,f_r,g_1,...,g_r\in C(V)$ such that on V

$$\omega=\sum_{i=1}^{r} f_i dg_i$$

and $g_1,...,g_r$ are functionally independent. Moreover, as special cases we have (i) If $d\omega^r\neq 0$ for $q\in U$ then the functions $f_1,...,f_r,g_1,...g_r\in C(V)$ are all functionally independent. (ii) If $d\omega^r=0$ for all $q\in U$ then there exists a neighborhood V of p and functionally independent functions $z,f_1,...,f_{r-1},g_1,...,g_{r-1}\in C(V)$ such that on V

$$\omega=dz+\sum_{i=1}^{r-1} f_i dg_i. \tag{1.9}$$

(Note that a 1-form ω may be regular without falling into either of the above cases, e.g., if $d\omega^r=0$ at p but is not identically zero in any neighborhood of p.)

We would like to draw a corollary from this result but must first introduce the concept of the Cartan system of a Pfaffian system. Let I be a Pfaffian system on a manifold M. Then there is an intrinsically defined vector field system

$$\text{Char}I=\{X\in\Gamma(M):\langle X,\omega\rangle = 0 \text{ and } X\lrcorner d\omega\in I \text{ for all } \omega\in I\}$$

called the characteristic system of I. Here $\langle\ ,\ \rangle$ is the bilinear pairing of the exterior algebra $\Lambda\Gamma(M)$, of $\Gamma(M)$, with the exterior algebra $\Lambda\Gamma^*(M)$, of $\Gamma^*(M)$, and the symbol $X\lrcorner$ denotes the so-called "interior product by X" defined as follows. Let $X\in\Gamma(M)$, $\Omega\in\Lambda^r\Gamma^*(M)$, and $Y\in\Lambda^{r-1}\Gamma(M)$ $(r\geqslant 1)$, i.e., X is a vector field on M, Ω is an element of the rth exterior power of $\Gamma^*(M)$ (an r-form), and Y is an element of the $(r-1)$th exterior power of $\Gamma(M)$. Then $X\lrcorner$ is a mapping $X\lrcorner:\Lambda^r\Gamma^*(M)\to\Lambda^{r-1}\Gamma^*(M)$ defined by the relation

$$\langle Y,X\lrcorner\Omega\rangle=\langle X\wedge Y,\Omega\rangle.$$

The annihilator

$$C(I)=\text{Char}(I)^{\perp}=\{\phi\in\Gamma^*(M):\langle X,\phi\rangle = 0 \text{ for all } X\in\text{Char}(I)\}$$

is called the Cartan system of I. The basic property of this system is given by the following local theorem of Cartan.[4]

Theorem: Let I be a nonsingular Pfaffian system with a nonsingular Cartan system $C(I)$. Then $C(I)$ is completely integrable. Moreover, if $(x^1,...,x^p)$ is a local system of first intergals then locally there exist generators for I which only depend on $(x^1,...,x^p)$ and their differentials.

Corollary: Let $\omega\in\Gamma^*(M)$ be a 1-form on an m-dimensional manifold M satisfying

$$d\omega^r\wedge\omega=0 \quad \text{for all } p\in V \tag{1.10}$$

and

$$d\omega^{r-1}\wedge\omega\neq 0 \quad \text{for all } p\in V, \tag{1.11}$$

for some connected open set V and some integer $r\geqslant 1$. Then the maximum dimension (actually attained) of an integral manifold of ω in V is $m-r$.

Proof: Since $d\omega^r\wedge\omega=0$ and $d\omega^{r-1}\wedge\omega\neq 0$ for all $p\in V$, then for any $q\in V$ there is a neighborhood U of q on which

$$d\omega^{r-1}\wedge\omega$$

is monomial and

$$C(\omega)=\{\pi\in\Gamma^*(U):\pi\wedge d\omega^{r-1}\wedge\omega=0\}. \tag{1.12}$$

(See Ref. 5). Clearly then the Pfaffian systems $\{\omega\}$ and $C(\omega)$ are nonsingular so that Theorem (1.10) is applicable. From Eq. (1.12) we see that there are $2r-1$ first integrals of $C(\omega)$ and hence there exists a coordinate system $(u^1,...,u^m)$ on a neighborhood of q which we can take to be U, with

$$C(\omega)=\{du^1,...,du^{2r-1}\}.$$

Theorem (1.10) then implies that there is a generator ϕ of the Pfaffian system $\{\omega\}$ which depends only on $(u^1,...,u^{2r-1})$ and their differentials. Since ϕ and ω both generate the same Pfaffian system on U they must have the same integral manifolds in U. However, since ϕ depends only on $(u^1,...u^{2r-1})$ and their differentials we must have

$$d\phi^r=0 \text{ for all } p\in U,$$

while, of course,

$$d\phi^{r-1}\wedge\phi\neq 0 \text{ for all } p\in U.$$

The latter follows since ϕ and ω generate the same Pfaffian system and therefore $\omega=\rho\phi$ on U for some nonzero function $\rho\in C(U)$. It is the presence of the function ρ which leads us to use ϕ rather then ω in our argument since $d\omega^r=\rho^{r-1}d\phi^{r-1}\wedge\phi\wedge d\rho$ may not fall into either of the two cases of Theorem (1.9) while ϕ necessarily falls into case (ii). Hence by Theorem (1.9) there is a neighboorhood of the arbitrary point q of V, which again we may without loss of generality take to be U, and functionally independent functions $z,f_1,...f_{r-1},g_1,...g_{r-1}\in C(U)$ such that on U

$$\phi=dz+\sum_{i=1}^{r-1} f_i dg_i.$$

Clearly $z=z(q)$, $g_i=g_i(q)$, $i=1,...,r-1$ define an integral manifold of ω through q with dimension $m-r$ as desired.

To show $m-r$ is the maximum dimension of an integral manifold of ω in V let $\sigma:N\to M$ be an integral manifold of ω through $q\in V$. Then, because in the neighborhood U of q

$$\phi=dz+\sum_{i=1}^{r-1} f_i dg_i,$$

and

$$d\phi = \sum_{i=1}^{r-1} df_i \wedge dg_i,$$

we must have

$$\sigma^*\phi = \sigma^* dz + \sigma^* \sum_{i=1}^{r-1} f_i dg_i = 0, \tag{1.13}$$

$$\sigma^* d\phi = \sum_{i=1}^{r-1} \sigma^* df_i \wedge \sigma^* dg_i = 0. \tag{1.14}$$

By the lemma below, Eq. (1.14) implies that

$$\dim\{\sigma^* df_1,\ldots,\sigma^* df_{r-1},\sigma^* dg_1,\ldots,\sigma^* dg_{r-1}\} \leqslant r-1.$$

But if we extend $f_1,\ldots,f_{r-1},g_1,\ldots,g_{r-1},z$ to a local coordinate system about q we see that $T^*_{\sigma^{-1}(q)}N$ is spanned by $\sigma^* df_1,\ldots,\sigma^* df_{r-1},\sigma^* dg_1,\ldots,\sigma^* dg_{r-1},\sigma^* dz$ plus σ^* of $n-(2r-1)$ other 1-forms. By Eq. (1.13) $\sigma^* dz$ is a linear combination of $\sigma^* dg_i$, $i=1,\ldots,r-1$ and hence the dimension of $T^*_{\sigma^{-1}(q)}N$ is $\leqslant n-(2r-1)+r-1=n-r$. Thus the dimension of N is $\leqslant n-r$ and the proof is complete. |||

Lemma: Let $x_1,\ldots,x_n,y_1,\ldots,y_n$ be elements of a vector space V. Then $\Sigma_{i=1}^n x_i \wedge y_i = 0$ implies that the span of $x_1,\ldots,x_n,y_1,\ldots,y_n$ has dimension $\leqslant n$.

Proof: The proof is by induction on n. The result is clearly true for $n=1$. Now assume it is true for $n=k$. We must prove that $\Sigma_{i=1}^{k+1} x_i \wedge y_i = 0$ implies that the span of $x_1,\ldots,x_{k+1},y_1,\ldots,y_{k+1}$ has dimension $\leqslant k+1$. If any one of the terms $x_j \wedge y_j$ in the sum is zero the result follows immediately from the inductive hypothesis. If none of the terms $x_j \wedge y_j$ is zero than in particular $x_{k+1} \wedge y_{k+1} \neq 0$ so that x_{k+1}, y_{k+1} are part of a basis for V. Relative to this basis let $\xi_{k+1} \in V^*$ be dual to x_{k+1} so that $\xi_{k+1} \lrcorner x_{k+1} = 1$ but $\xi_{k+1} \lrcorner y_{k+1} = 0$ [the interior product by ξ_{k+1}, $\xi_{k+1}\lrcorner$, is defined in the paragraph preceding Theorem (1.10)]. Then

$$\begin{aligned} 0 &= \xi_{k+1} \lrcorner \sum_{i=1}^{k+1} x_i \wedge y_i \\ &= (\xi_{k+1} \lrcorner x_1) y_1 - x_1(\xi_{k+1} \lrcorner y_1) + \cdots \\ &\quad + (\xi_{k+1} \lrcorner x_k) y_k - x_k(\xi_{k+1} \lrcorner y_k) + y_{k+1} \end{aligned} \tag{1.16}$$

expresses y_{k+1} as a linear combination of $x_1,\ldots,x_k,y_1,\ldots,y_k$.

Now let $E(V)$ be the exterior algebra of V and define ϕ: $E(V) \to E(V)$ by

$$\phi(\alpha) = \alpha - x_{k+1} \wedge (\xi_{k+1} \lrcorner \alpha).$$

ϕ is easily seen to be an algebra morphism and clearly $\phi(x_{k+1} \wedge y_{k+1}) = 0$. Then

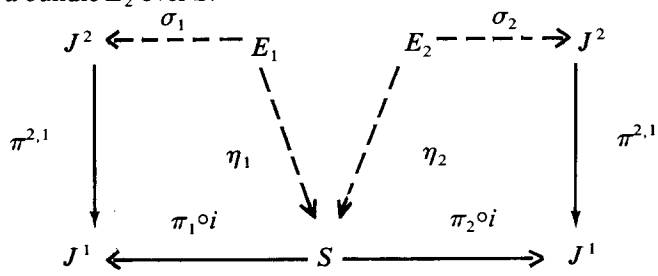
$$0 = \phi\left(\sum_{i=1}^{k+1} x_i \wedge y_i\right) = \phi\left(\sum_{i=1}^{k} x_i \wedge y_i\right) = \sum_{i=1}^{k} \phi(x_i) \wedge \phi(y_i)$$

implies by the induction hypothesis that the span of $\phi(x_1),\ldots,\phi(x_k),\phi(y_1),\ldots,\phi(y_k)$ has dimension $\leqslant k$. But for any vector $x \in V$ $\phi(x) = x - x_{k+1}(\xi_{k+1} \lrcorner x)$ and thus the span of $\phi(x_1),\ldots,\phi(x_k),\phi(y_1),\ldots,\phi(y_k),x_{k+1}$ equals the span of $x_1,\ldots,x_k,y_1,\ldots,y_k,x_{k+1}$, which therefore has dimension $\leqslant k+1$. Finally, the proof is completed by the observation that Eq. (1.16) implies that the span of $x_1,\ldots,x_{k+1},y_1,\ldots,y_{k+1}$ equals the span of $x_1,\ldots,x_{k+1},y_1,\ldots,y_k$. |||

2. BÄCKLUND TRANSFORMATIONS: THE DIFFERENTIAL EQUATIONS

Let π_1 and π_2 be the canonical projections from $J^1(R^n,R) \times J^1(R^n,R)$ onto the first and second factors respectively, and let $i: S \to J^1(R^n,R) \times J^1(R^n,R)$ be a $3n$-dimensional submanifold of the product such that $\pi_1 \circ i$: $S \to J^1(R^n,R)$ and $\pi_2 \circ i: S \to J^1(R^n,R)$ are submersions, (i.e., their differentials $(\pi_j \circ i)_*$ are of maximal rank). We show that with the addition of one more condition S leads to two single second order partial differential equations for functions from R^n to R. It is these two differential equations which will be related by the Bäcklund transformation S.

$J^2(R^n,R)$ is a bundle over $J^1(R^n,R)$ with projection $\pi^{2,1}$. The map $\pi_1 \circ i: S \to J^1(R^n,R)$ induces a bundle E_1 over S constructed by letting the fiber over $x \in S$ be the fiber over $(\pi_1 \circ i)(x)$ in $J^1(R^n,R)$. Similarly $\pi_2 \circ i: S \to J^1(R^n,R)$ induces a bundle E_2 over S.



$\eta_j: E_j \to S$ projections, $(\pi_j \circ i)$ and σ_j submersions, $j=1,2$. (2.1)

Let ω denote the contact form on $J^1(R^n,R)$

$$\omega = dz - \sum_{i=1}^{n} p_i dx_i.$$

Then $d\omega^n \wedge \omega = (\Sigma_1^n dx_i \wedge dp_i)^n \wedge (dz - \Sigma_1^n p_i dx_i)$ $= n! dx_1 \wedge dp_1 \wedge \cdots \wedge dx_n \wedge dp_n \wedge dz \neq 0$. Since $\pi_1 \circ i$ and $\pi_2 \circ i$ are submersions, $\omega_1 = (\pi_1 \circ i)^* \omega$ and $\omega_2 = (\pi_2 \circ i)^* \omega$ are both nonzero 1-forms on S which also satisfy

$$(d\omega_i)^n \wedge \omega_i \neq 0, \text{ for all } p \in S,\ i = 1,2.$$

In turn, since $\eta_1: E_1 \to S$ and $\eta_2: E_2 \to S$ are projections,

$$\eta_1^*(d\omega^n{}_2 \wedge \omega_2) \neq 0 \quad \text{for all } p \in E_1,$$

$$\eta_2^*(d\omega^n{}_1 \wedge \omega_1) \neq 0 \quad \text{for all } p \in E_2.$$

(Note that η_1^* acts on ω_2 and η_2^* on ω_1.)

Recall that the contact system $\Omega^2(R^n,R)$ on $J^2(R^n,R)$ is generated by the $(n+1)$ 1-forms

$$\Omega^2(R^n,R) = \begin{cases} dz - \sum_{i=1}^{n} p_i dx_i, & \\ dp_i - \sum_{j=1}^{n} p_{ij} dx_j, & i = 1,\ldots,n. \end{cases}$$

Let μ be the $(n+1)$-form obtained by wedging together the $(n+1)$ generators of $\Omega^2(R^n,R)$. By the construction of the bundles E_1 and E_2 over S the maps $\sigma_1: E_1 \to J^2(R^n,R)$ and $\sigma_2: E_2 \to J^2(R^n,R)$ are submersions, since they cover the submersions $\pi_1 \circ i$ and $\pi_2 \circ i$, respectively. Hence $\sigma_1^* \mu$ and $\sigma_2^* \mu$ are nowhere vanishing $(n+1)$-forms on E^1 and E^2, respectively.

Now consider the $3n$-forms

$$\nu_2 = \eta_2^*(d\omega_1^{n-1} \wedge \omega_1) \wedge \sigma_2^* \mu \quad \text{on } E_2 \tag{2.2}$$

and

$$\nu_1 = \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \wedge \sigma_1^*\mu \quad \text{on } E_1. \tag{2.3}$$

Recall that a Pfaffian system I defined on a bundle $\pi: B \to M$ is said to be semibasic if

$$\ker \pi_* \subset I^1 = \{x \in TM : \langle x, w \rangle = 0 \text{ for all } w \in I\}.$$

The Pfaffian systems

$$\{\sigma_2^*\Omega^2(R^n, R), \eta_2^*\omega_1\} \quad \text{on } E_2$$

and

$$\{\sigma_1^*\Omega^2(R^n, R), \eta_1^*\omega_2\} \quad \text{on } E_1$$

are easily seen to be semibasic, and as such we have that locally

$$\begin{aligned} \nu_2 &= \eta_2^*(d\omega_1^{n-1} \wedge \omega_1) \wedge \sigma_2^*\mu = h_2\eta_2^*(dV_s), \\ \nu_1 &= \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \wedge \sigma_1^*\mu = h_1\eta_1^*(dV_s), \end{aligned} \tag{2.4}$$

for some functions $h_1 \in C(E_1)$ and $h_2 \in C(E_2)$, where dV_s is a local volume form on S. (This follows since both are semibasic $3n$-forms over a $3n$-dimensional manifold.) Thus, if the map $\nu_i: E_i \to \Lambda^{3n}\Gamma^*(E_i)$ is nontrivial and vanishes independently of the fiber of σ_i $(i = 1,2)$, the image under σ_i of the zero set of ν_i will define a single second-order partial differential equation in $J^2(R^n, R)$ for each $i = 1,2$. That they are single equations follows since the vanishing of ν_i is equivalent to the vanishing of the function $h_i \in C(E_i), i = 1,2$. We have proved the following theorem.

Theorem: Let $i: S \to J^1(R^n, R) \times J^1(R^n, R)$ be a $3n$-dimensional submanifold of $J^1(R^n, R) \times J^1(R^n, R)$. Then S determines two single second-order partial differential equations for functions from R^n to R if the following conditions are satisfied.

(i) $\pi_1 \circ i: S \to J^1(R^n, R)$ and $\pi_2 \circ i: S \to J^1(R^n, R)$ are submersions.

(ii) The $3n$-forms

$$\nu_1 = \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \wedge \sigma_1^*\mu \in \Lambda^{3n}\Gamma^*(E_1)$$

and

$$\nu_2 = \eta_2^*(d\omega_1^{n-1} \wedge \omega_1) \wedge \sigma_2^*\mu \in \Lambda^{3n}\Gamma^*(E_2)$$

are not identically zero and vanish independently of the fiber of σ_1 and σ_2, respectively. That is,

$$\nu_i(x) = 0 \text{ implies } \nu_i\{\sigma_i^{-1}[\sigma_i(x)]\} = 0, \quad i = 1,2. \tag{2.5}$$

The differential equations are

$$\sigma_1[Z(\nu_1)] \subset J^2(R^n, R) \text{ and } \sigma_2[Z(\nu_2)] \subset J^2(R^n, R),$$

where $Z(\nu_i)$ = zero set of ν_i, $i = 1,2$. Notice that one begins with the submanifold S and proceeds to the differential equations. One cannot, in general, construct a submanifold S which will give rise to two given differential equations in this way. The reasons are similar to those for the classical case of two independent variables (see Ref. 6).[6]

3. BÄCKLUND TRANSFORMATIONS: THE SOLUTIONS

To better understand Bäcklund transformations and the way that they lead from a solution of one differential equation to a solution or family of solutions of the related differential equations, let us compare them with the well-known contact transformations. A contact transformation is a diffeomorphism $u: J^1(R^n, R) \times J^1(R^n, R)$ such that $u^*\omega = \rho\omega, \rho \neq 0$, where ω is the contact form. Note that u^{-1} is also a contact transformation. If $i: \Sigma \to J^1(R^n, R)$ is a differential equation then $u \circ i: \Sigma \to J^1(R^n, R)$ is another differential equation. Moreover, if $f: R^n \to R$ is a solution to the first equation so that $\mathrm{im}[j^1(f)]$ lies on the image of i: $\Sigma \to J^1(R^n, R)$ and $j^1(f)^*\omega = 0$, then clearly $u \circ j^1(f): R^n \to J^1(R^n, R)$ has its image on the image of $u \circ i: \Sigma \to J^1(R^n, R)$, and $[u \circ j^1(f)]^*\omega = j^1(f)^* \circ u^*\omega = j^1(f)^*\rho\omega = 0$. Hence $u \circ j^1(f): R^n \to J^1(R^n, R)$ is a local solution to the differential equation $u \circ i: \Sigma \to J^1(R^n, R)$ if only $\alpha \circ u \circ j^1(f): R^n \to R^n$ is a local diffeomorphism, i.e., if only the image of $u \circ j^1(f): R^n \to J^1(R^n, R)$ is transverse to the fiber of $\alpha: J^1(R^n, R) \to R^n$.

As an example consider the Legendre transformation $u: J^1(R^2, R) \to J^1(R^2, R)$ defined by

$$u: \quad \begin{aligned} X &= p, \\ Y &= q, \\ Z &= xp + yq - z, \\ P &= x, \\ Q &= y, \end{aligned}$$

and the so called Clairaut differential equation[7]

$$z - xp - yq - f(p,q) = 0.$$

Under the transformation $u: J^1(R^2, R) \to J^1(R^2, R)$ this becomes

$$-Z - f(X,Y) = 0,$$

which is not a differential equation at all. In fact, if a, b are arbitrary constants the map $\theta: R^2 \to J^1(R^2, R)$ defined by

$$\theta(\xi, \eta) = [a, b, -f(a,b), \xi, \eta]$$

has its image in the set $\{-Z - f(X,Y) = 0\}$. Moreover,

$$\theta^*\omega = 0,$$

and therefore $u^{-1} \circ \theta: R^2 \to J^1(R^2, R)$ has its image in the set $\{z - xp - yq - f(p,q) = 0\}$ and satisfies

$$(u^{-1} \circ \theta)^*\omega = \theta^* \circ (u^{-1})^*\omega = \theta^*(-\omega) = 0.$$

$\theta: R^2 \to J^1(R^2, R)$ is, however, not transverse to the fiber of α and so is not the 1-graph of a function. It is merely a two-dimensional integral manifold of ω. $u^{-1} \circ \theta: R^2 \to J^1(R^2, R)$ will be a solution of the Clairaut equation if only its image is transverse to the fiber of $\alpha: J^1(R^2, R) \to R^2$. Since

$$u^{-1} \circ \theta(\xi, \eta) = [\xi, \eta, a\xi + b\eta + f(a,b), a, b],$$

we see that, indeed, the image of $u^{-1} \circ \theta$ is transverse to the fiber of α, for $\alpha \circ u^{-1}(\xi, \eta) = (\xi, \eta)$. Hence

$$z = ax + by + f(a,b)$$

is a two-parameter solution of the Clairaut equation. It is in just this way that a contact transformation is used to find either a simpler differential equation or, as in this case, an obvious integral manifold of ω which, when transformed back, yields a solution to the original equation.

It is important to observe that a solution to a given differential equation when transformed under a contact transformation may not lead to a solution to the transformed dif-

ential equation. For under a contact transformation the base and fiber coordinates may well be intermixed as in the Legendre transformation. Thus, although one is guaranteed that if $f: R^n \to R$ is a solution to the original differential equation then $u \circ j^1(f): R^n \to J^1(R^n, R)$ will have its image on the image of the transformed differential equation and will be an integral manifold of ω, one is not guaranteed that the image of $u \circ j^1(f)$ will be transverse to the fiber of α—a necessary condition if $u \circ j^1(f)$ is to be a 1-graph of a function from $R^n \to R$ and hence a solution of the transformed equation. There is, then, a transversality condition on a given solution f: $R^n \to R$ if it is to give rise to a solution to the transformed differential equation. In the case of the Legendre transformation, since $\alpha \circ u \circ j^1(f) = (p,q)$, it is clear that the transversality condition is

$$\frac{\partial(p,q)}{\partial(x,y)} = f_{xx} f_{yy} - (f_{xy})^2 \neq 0.$$

Notice that given a contact transformation some solutions will give rise to a transformed solution and some will not.

Since a contact transformation is a diffeomorphism its graph is a $(2n+1)$-dimensional submanifold

$$i: S' \to J^1(R^n, R) \times J^1(R^n, R)$$

of $J^1(R^n, R) \times J^1(R^n, R)$ and satisfies the condition that $\pi_1 \circ i$ and $\pi_2 \circ i$ be submersions. Generalizing, we might consider relations on $J^1(R^n, R)$ rather than functions,—in this case submanifolds

$$i: S \to J^1(R^n, R) \times J^1(R^n, R)$$

of $J^1(R^n, R) \times J^1(R^n, R)$ which still satisfy the condition that $\pi_1 \circ i$ and $\pi_2 \circ i$ be submersions. This is exactly what we have done to define a Bäcklund transformation, letting S be a $3n$-dimensional submanifold of $J^1(R^n, R) \times J^1(R^n, R)$. Because a Bäcklund transformation is really a relation on $J^1(R^n, R)$ and not a function, we should expect a solution of one of the related equations to give rise to possibly a whole family of solutions to the other. However, just as for contact transformations, a given solution of one differential equation may lead to no solution of the other if the appropriate transversality condition fails to be satisfied. Of course, a Bäcklund transformation relates second-order partial differential equations while a contact transformation relates first-order partial differential equations.

With this background let us begin to investigate how a Bäcklund transformation, which we have just seen should more properly be called a Bäcklund relation, uses a solution of one of its related differential equations to create possibly a whole family of solutions to the other, provided that the appropriate transversality condition is satisfied. Recall that we are dealing only with local solutions.

For any function $f: R^n \to R$ {whether or not $\mathrm{im}(j^2(f)) \subset \sigma_1(Z(\nu_1))\}\, \sigma_1^{-1}[\mathrm{im}(j^2(f))]$ is an integral manifold of $\sigma_1^* \Omega^2(R^n, R)$ of dimension $2n-1$. Let $M_1 = \sigma_1^{-1}[\mathrm{im}(j^2(f))]$ and $\delta_1: M_1 \to E_1$ be the canonical injection.

Lemma: If $\nu_1 = \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \wedge \sigma_1^* \mu = 0$ at $\delta_1(p) \in E_1$ then

$$\delta_1^* \circ \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) = 0 \text{ at } p \in M_1. \tag{3.1}$$

The same result holds if we exchange the subscripts 1 and 2 so that we are in the bundle E_2.

Proof: To prove that $\nu_1 = 0$ at $\delta_1(p)$ implies that $\delta_1^* \circ \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) = 0$ at p we prove that $\delta_1^* \circ \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \neq 0$ at p implies that $\nu_1 \neq 0$ at $\delta_1(p)$. $\delta_1^* \circ \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \neq 0$ at p implies that there exist vectors $x_1, \ldots, x_{2n-1} \in T_p M_1$ such that

$$\langle \delta_{1*} x_1 \wedge \ldots \wedge \delta_{1*} x_{2n-1}, \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \rangle \neq 0$$

at $\delta_1(p) \in E_1$. Since $\sigma_1^* \mu$ is nowhere zero on E_1 and since $\delta_1^* \circ \sigma_1^* \Omega^2(R^n, R) = 0$, there must exist vectors

$$y_1, \ldots, y_{n+1} \in T_{\delta_1(p)} E_1$$

not tangent to $\delta_1(M_1)$ (i.e., not in the image of δ_{1*} at p) for which

$$\langle y_1 \wedge \cdots \wedge y_{n+1}, \sigma_1^* \mu \rangle \neq 0$$

at $\delta_1(p) \in E_1$. Now at $\delta_1(p)$,

$$\begin{aligned}
&\langle \delta_{1*} x_1 \wedge \cdots \wedge \delta_{1*} x^{2n-1} \wedge y_1 \wedge \cdots \wedge y_{n+1}, \\
&\quad \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \wedge \sigma_1^* \mu \rangle \\
&\qquad = \langle \delta_{1*} x_2 \wedge \cdots \wedge \delta_{1*} x_{2n-1} \wedge y_1 \cdots \wedge y_{n+1}, \\
&\qquad\quad [\delta_{1*} x_1 \lrcorner \eta_1^*(d\omega_2^{n-1} \wedge \omega_2)] \wedge \sigma_1^* \mu \rangle \\
&\qquad = \langle y_1 \wedge \cdots \wedge y_{n+1}, [\delta_{1*} x_{2n-1} \lrcorner \cdots \\
&\quad \delta_{1*} x_1 \lrcorner \eta_1^*(d\omega_2^{n-1} \wedge \omega_2)] \wedge \sigma_1^* \mu \rangle,
\end{aligned}$$

since $\delta_{1*} x_i \lrcorner \sigma_1^* \mu = 0$ $(i = 1, \ldots, 2n-1)$ by virtue of the fact that $\delta_1^* \circ \sigma_1^* \Omega^2(R^n, R) = 0$. Then the above becomes

$$\begin{aligned}
&\langle \delta_{1*} x_1 \wedge \cdots \wedge \delta_{1*} x_{2n-1}, \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \rangle \\
&\quad \cdot \langle y_1 \wedge \cdots \wedge y_{n+1}, \sigma_1^* \mu \rangle,
\end{aligned}$$

which being nonzero, proves that $\nu_1 = \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) \wedge \sigma_1^* \mu$ is in fact nonzero in $\Lambda^{3n}(T^*_{\delta_1(p)} E_1)$, as desired. |||

Lemma: If M_1 contains an integral manifold N_1 of $\delta_1^* \circ \eta_1^* \omega_2$ of dimension $\geqslant n$ then

$$\nu_1 = 0 \text{ at all points } p \in \delta_1(N_1). \tag{3.2}$$

The same result holds if we exchange the subscripts 1 and 2 so that we are in E_2.

Proof: $T_{\delta_1(p)} E_1 / \ker \eta_{1*}$ is a $3n$-dimensional vector space and since ν_1, $\eta_1^* \omega_2$, $\eta_1^* d\omega_2$, and $\sigma_1^* \Omega^2(R^n, R)$ are all semibasic, each can be thought of as belonging to the appropriate exterior power of

$$(T_{\delta_1(p)} E_1 / \ker \eta_{1*})^*.$$

Let $W \subset T_{\delta_1(p)} E_1 / \ker \eta_{1*}$ be the $(2n-2)$-dimensional vector space defined by

$$\begin{aligned}
W = \{ &w \in T_{\delta_1(p)} E_1 / \ker \eta_{1*}: \\
&\langle w, \eta_1^* \omega_2 \rangle = \langle w, \sigma_1^* \Omega^2(R^n, R) \rangle = 0 \}.
\end{aligned}$$

Then, since $\nu_1 = (\eta_1^* d\omega_2)^{n-1} \wedge \eta_1^* \omega_2 \wedge \sigma_1^* \mu$, where μ is the product of the $n+1$ generators of $\Omega^2(R^n, R)$, we have from the theory of exterior algebras[5] that $\nu_1 \neq 0$ at $\delta_1(p)$ is equivalent to the nondegeneracy of the skew-symmetric bilinear form $\eta_1^* d\omega_2$ on W. We thus must show that $\eta_1^* d\omega_2$ is degenerate on W.

Let N_1 be an at least n-dimensional integral manifold of $\delta_1^* \circ \eta_1^* \omega_2$ passing through $p \in M_1$. Then

$$V = \delta_{1*}(T_p N_1) / \ker \eta_{1*}$$

is an at least n-dimensional subspace of W. Let

$$V^{\perp} = \{x \in W: \langle x, y \lrcorner \eta_1^* d\omega_2 \rangle = 0 \text{ for all } y \in V\}$$

be the orthogonal complement relative to $\eta_1^* d\omega_2$ of V in W. (Note the different use of the notation $V^{\perp}$ here from elsewhere in this paper.) Since N_1 is an integral manifold of $\delta_1^* \circ \eta_1^* \omega_2$, for any $x, y \in \delta_{1*}(T_p N_1)$,

$$\langle x, y \lrcorner \eta_1^* d\omega_2 \rangle = 0$$

and therefore $V \subset V^{\perp}$.

Now

$$\dim(V \cap V^{\perp}) + \dim(V + V^{\perp}) = \dim V + \dim V^{\perp}$$

or, since $V \subset V^{\perp}$,

$$\dim V + \dim(V + V^{\perp}) = \dim V + \dim V^{\perp},$$

which implies

$$\dim(V + V^{\perp}) = \dim V^{\perp}.$$

Thus $\dim V^{\perp} \geqslant n$ since $\dim V \geqslant n$. But this is impossible for a nondegenerate bilinear form since for such a form $\dim V = \dim V^0$, where

$$V^0 = \{w \in W^*: \langle x, w \rangle = 0 \text{ for all } x \in V\}$$

has dimension $\leqslant (2n-2) - n = n-2$. Thus $\eta_1^* d\omega_2$ is degenerate on W and so $\nu_1 = 0$ at $\delta_1(p)$ as desired. |||

Now let f: $U_1 \subset R^n \to R$ be a local solution about the point $x_0 \in R^n$ of the differential equation

$$\sigma_1[Z(\nu_1)] \subset J^2(R^n, R) \tag{3.3}$$

determined in Sec. 2. Subject to a transversality condition, to be introduced, we must show how this leads to a solution or family of solutions of the differential equation

$$\sigma_2[Z(\nu_2)] \subset J^2(R^n, R). \tag{3.4}$$

Again let $M_1 = \sigma_1^{-1}[\mathrm{im}(j^2(f))]$, and let $\delta_1: M_1 \to E_1$ be the canonical injection. Since f is a solution of (3.3), $\nu_1 = 0$ along the image of δ_1 in E_1 and hence by lemma 3.1,

$$\delta_1^* \circ \eta_1^*(d\omega_2^{n-1} \wedge \omega_2) = 0$$

everywhere in M_1. $\delta_1^* \circ \eta_1^* \omega_2$ may be regular at some points of the fiber $(\sigma_1 \circ \delta_1)^{-1}(j^2_{x_0}(f))$ and fail to be regular at other points of the fiber (see Ex. 2 in Sec. 4). Assuming that $\delta_1^* \circ \eta_1^* \omega_2$ is regular at some point $p_0 \in (\sigma_1 \circ \delta_1)^{-1}(j^2_{x_0}(f))$, there exists a neighborhood U of p_0 such that

$$\delta_1^* \circ \eta_1^*(d\omega_2^r \wedge \omega_2) = 0 \text{ for all } p \in U$$

and

$$\delta_1^* \circ \eta_1^*(d\omega_2^{r-1} \wedge \omega_2) \neq 0 \text{ for all } p \in U,$$

for some $r \leqslant n-1$. (In most cases of interest it seems the regularity condition (1.8) is satisfied with $r = n-1$.)

By the proof of Corollary (1.11) we know that in some neighborhood V about p_0, $\delta_1^* \circ \eta_1^* \omega_2$ can be expressed as $\delta_1^* \circ \eta_1^* \omega_2 = \rho_2 \phi_2$ for some nonzero function $\rho \in C(V)$ and some 1-form $\phi_2 \in \Gamma^* V$, where ϕ_2 depends only on the $2r-1$ first integrals of the Cartan system of $\delta_1^* \circ \eta_1^* \omega_2$, and hence satisfies

$$d\phi_2^r = 0 \text{ for all } p \in V$$

and

$$d\phi_2^{r-1} \wedge \phi_2 = 0 \text{ for all } p \in V.$$

This implies that in some neighborhood about p_0, which we can take to be V again, ϕ_2 may be expressed as

$$\phi_2 = dg_1 + \sum_{i=2}^{r} f_i dg_i \quad 1 \leqslant r \leqslant n-1, \tag{3.5}$$

where all the functions $g_1, \ldots, g_r, f_2, \ldots, f_r$ are functionally independent, and we may assume $g_i(p_0) = 0$ $(i = 1, \ldots, r)$.

Since we are interested only in integral manifolds of $\delta_1^* \circ \eta_1^* \omega_2$, we will work with ϕ_2 rather than $\delta_1^* \circ \eta_1^* \omega_2$.

For each $c = (c_1, \ldots, c_r)$ in some neighborhood C of zero in R^r

$$L_c = \{p \in M_1: g_i(p) = c_i, \quad i = 1, \ldots, r\}$$

is an integral manifold of $\delta_1^* \circ \eta_1^* \omega_2$ of dimension $2n-1-r \geqslant n$. Thus there is a foliation $F = \{L_c\}_{c \in C}$ of V by integral manifolds of $\delta_1^* \circ \eta_1^* \omega_2$, each of dimension $2n-1-r \geqslant n$. L_0 is the leaf through p_0.

Recall that $\alpha: J^1(R^n, R) \to R^n$ takes $j^1_x(f)$ to $x \in R^n$. Let $\alpha \circ (\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c$ denote the restriction of $\alpha \circ (\pi \circ i) \circ (\eta_1 \circ \delta_1)$ to L_c, $c \in C$. Let $q_0 = \alpha \circ (\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)(p_0)$. Then if

$$(\alpha \circ (\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_0)_{*_{p_0}}: T_{p_0} L_0 \to T_{q_0} R^n$$

has maximal rank the image of $(\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_0$ in $J^1(R^n, R)$ must be locally transverse to the fiber of α: $J^1(R^n, R) \to R^n$. Moreover, since maximal rank is an open condition, there exist a neighborhood about p_0 which we will again call V and a neighborhood about zero in R^r which we will again call C, such that

$$(\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c \cap V: L_c \cap V \to J^1(R^n, R)$$

has image transverse to the fiber of α for every $c \in C$; and in fact,

$$\alpha \circ (\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c \cap V: L_c \cap V \to R^n$$

will be a submersion or, if $r = n-1$, a diffeomorphism onto its image $U_c \subset R^n$, $c \in C$.

This implies that $(\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c \cap V: L_c \cap V \to J^1(R^n, R)$ has rank at least n. But, since $L_c \cap V$ is an integral manifold of $\delta_1^* \circ \eta_1^* \omega_2 = \delta_1^* \circ \eta_1^*(\pi_2 \circ i)^* \omega$, we see that if k is the maximum of the rank of $((\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c \cap V)_*$ on $L_c \cap V$ then there exists some neighborhood W in $L_c \cap V$ for which $(\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|W: W \to J^1(R^n, R)$ has an image that is a k-dimensional integral manifold of ω in $J^1(R^n, R)$. But $d\omega^n \wedge \omega \neq 0$ and $d\omega^{n+1} \wedge \omega = 0$ everywhere in $J^1(R^n, R)$ and hence, since $J^1(R^n, R)$ has dimension $2n+1$, Corollary (1.11) implies that the maximum dimension of an integral manifold of ω is $2n+1-(n+1) = n$. Hence k must be $\leqslant n$ while at the same time k must be $\geqslant n$. Then $k = n$ and we have that

$$(\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c \cap V: L_c \cap V \to J^1(R^n, R)$$

has an image which is an n-dimensional integral manifold of ω transverse to the fiber of α for each $c \in C$. Thus the image of $(\pi_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c \cap V$ is the image of a 1-graph of some function $f_c: U_c \subset R^n \to R$ for each $c \in C$, where U_c is the open set of R^n equal to the image of $\alpha \circ (\eta_2 \circ i) \circ (\eta_1 \circ \delta_1)|L_c \cap V$.

Observe that the family of integral manifolds $L_c = \{p \in M_1: g_i(p) = c_i, i = 1, \ldots, r\}$ is by no means unique and hence that the family of functions f_c is also not unique for a

given solution f of the original equation (3.3). For from (3.5) we see that ϕ_2 can be expressed as

$$\phi_2 = dg_1 + \sum_{i=2}^{r} f_i dg_i \quad 1 \leqslant r \leqslant n-1,$$

and so ϕ_2 can also be expressed as

$$\phi_2 = d(g_1 + f_r g_r) + \sum_{i=2}^{r-1} f_i dg_i - g_r df_r,$$

etc., so that another family of integral manifolds might be

$$\{ p \in M_1 : (g_1 + f_r g_r)(p) = c_1, f_r = c_r, g_i = c_i, \quad i = 2,\ldots,r-1 \},$$

etc. (See Examples 1 and 3 in Sec. 4.)

To show that $f_c : U_c \to R$ is a local solution of the second-order partial differential equation $\sigma_2(Z(\nu_2)) \subset J^2(R^n, R)$ we must show that the image of $j^2(f_c) : U_c \to J^2(R^n, R)$ lies in $\sigma_2(Z(\nu_2))$. To do this we use Lemma (3.2).

First observe that, since $L_c \subset M_1$ and M_1 is a $(2n-1)$-dimensional integral manifold of $\eta_1^* \omega_1$, $\eta_1 \circ \delta_1 : L_c \to S$ is an at least n-dimensional integral manifold of ω_1 as well as of ω_2. Next observe that if δ_2 is the canonical injection of $\sigma_2^{-1}[\mathrm{im}(j^2(f_c))]$ into E_2 then, by the nature of the bundle $E_2 \to S$, the map $\eta_2 \circ \delta_2 : \sigma_2^{-1}[\mathrm{im}(j^2(f_c))] \to S$ contains the set $\eta_1 \circ \delta_1(L_c)$ in its range. That is, the range of $\eta_2 \circ \delta_2$ contains an at least n-dimensional integral manifold of ω_1. Hence through each point of $\sigma_2^{-1}[\mathrm{im}(j^2(f_c))]$ passes an at least n-dimensional integral manifold of $\eta_2^* \omega_1$ which, by Lemma (3.2), implies that $\nu_2 = 0$ along $\sigma_2^{-1}[\mathrm{im}(j^2(f_c))]$. So the image of $j^2(f_c) : U_c \to J^2(R^n, R)$ does, in fact, lie in $\sigma_2(Z(\nu_2))$ as desired, and each $f_c : U_c \to R$, $c \in C$ is a local solution of $\sigma_2(Z(\nu_2))$.

We summarize this section with the following theorem:

Theorem: Let f be a local solution about $x_0 \in R^n$ of the second-order partial differential equation $\sigma_1(Z(\nu_1)) \subset J^2(R^n, R)$, let $M_1 = \sigma_1^{-1}[\mathrm{im}(j^2(f))]$, and let $\delta_1 : M_1 \to E_1$ be the canonical injection.

(a) Assume that $\delta_1^* \circ \eta_1^* \omega_2$ is regular at some point p_0 of the fiber $(\sigma_1 \circ \delta_1)^{-1}(j^2_{x_0}(f))$ in M_1 so that on some neighborhood U of p_0 one can write

$$\delta_1^* \circ \eta_1^* \omega_2 = \rho\left(dg_1 + \sum_{i=2}^{r} f_i dg_i \right), \quad 1 \leqslant r \leqslant n-1, \tag{3.7}$$

where the functions $g_1,\ldots,g_r, f_2,\ldots,f_r$ are functionally independent on U and ρ is a nonzero function in $C(U)$. Then if the integral manifold $L_0 = \{ p \in M_1 : g_i(p) = g_i(p_0), i = 1,\ldots,r \}$ is transverse to the fiber of $\alpha \circ \pi_2 \circ i \circ \eta_1 \circ \delta_1 : M_1 \to R^n$, there is an r-parameter family of local solutions to the differential equation $\sigma_2(Z(\nu_2))$. Note that since the expression (3.7) for $\delta_1^* \circ \eta_1^* \omega_2$ is not unique it may happen that L_0 is not transverse to the fiber of $\alpha \circ \pi_2 \circ i \circ \eta_1 \circ \delta_1$, but that for a different expression of $\delta_1^* \circ \eta_1^* \omega_2$ the corresponding integral manifold would be. See examples 1 and 3 in Sec. 4.

(b) If $\delta_1^* \circ \eta_1^* \omega_2$ is not regular at some point p_0 of the fiber $(\sigma_1 \circ \delta_1)^{-1}(j^2_{x_0}(f))$ it is still possible that there may exist an integral manifold through p_0 yielding a solution to $\sigma_2(Z(\nu_2))$ (see Example 2, Sec. 4) but very little else else is known.

4. EXAMPLES

Example 1: A Bäcklund transformation for second-order partial differential equations on R^3 will be a nine-dimensional submanifold of the 14-dimensional product manifold $J^1(R^3, R) \times J^1(R^3, R)$. Let $i : S \to J^1(R^3, R) \times J^1(R^3, R)$ be such a submanifold defined implicitly by the five equations $X_1 = x_1$, $X_2 = x_2$, $X_3 = x_3$, $P_1 = \exp(z - p_2 - p_3)$, $P_2 = p_1$, where the lower case letters are the coordinates on the first factor of $J^1(R^3, R) \times J^1(R^3, R)$ and the upper case letters the coordinates on the second factor. Then S may be coordinatized by $x_1, x_2, x_3, z, p_1, p_2, p_3, P_2, P_3$ or by $X_1, X_2, X_3, Z, P_1, P_2, P_3, z, p_3$, as in

$$\begin{aligned}
&\text{(a)} \quad x_i = x_i, \quad z = z, \quad p_i = p_i, \quad X_i = x_i, \quad Z = \ln p_1 + P_2 + P_3, \quad P_1 = P_2, \quad P_2 = P_2, \quad P_3 = P_3, \\
&\text{or} \\
&\text{(b)} \quad x_i = x_i, \quad z = z, \quad p_1 = \exp(Z - P_2 - P_3), \quad p_2 = P_1, \quad p_3 = p_3, \quad X_i = X_i, \quad Z = Z, \quad P_i = P_i,
\end{aligned} \qquad (i = 1,2,3; p_1 > 0), \tag{4.1}$$

respectively. (4.1a) demonstrates that $\pi_1 \circ i : S \to J^1(R^3, R)$ is a submersion, and (4.1b) demonstrates that $\pi_2 \circ i : S \to J^1(R^3, R)$ is a submersion.

When the contact form $\omega = dz - \Sigma_{i=1}^{3} p_i dx_i$ on $J^1(R^3, R)$ is pulled back to S via $\pi_1 \circ i$ we obtain

$$\omega_1 = (\pi_1 \circ i)^* \omega = dz - \sum_{i=1}^{3} p_i dx_i \tag{4.2a}$$

or

$$\omega_1 = (\pi_1 \circ i)^* \omega = dz - \exp(Z - P_2 - P_3) dX_1 - P_1 dX_2 - p_3 dX_3, \tag{4.2b}$$

depending on the coordinate system used for S. When the contact form ω is pulled back to S via $\pi_2 \circ i$ we obtain

$$\omega_2 = (\pi_2 \circ i)^* \omega = dZ - \sum_{i=1}^{3} P_i dX_i \tag{4.3a}$$

or

$$\omega_2 = (\pi_2 \circ i)^* \omega = d(\ln p_1 + P_2 + P_3) - p_2 dx_1 - P_2 dx_2 - P_3 dx_3, \tag{4.3b}$$

depending again on the coordinate system used for S.

The bundles E_1 and E_2 can be coordinatized, respective-

ly, by x_i,z,p_i,P_2,P_3,p_{ij} and $X_i,Z,P_i,z,p_3,p_{ij}(1\leqslant i\leqslant j\leqslant 3)$. With respect to these coordinate systems one can show with a little calculation that

$$\begin{aligned}\nu_1 &= \eta_1^*(d\omega_2^2\wedge\omega_2)\wedge\sigma_1^*\mu\\ &= (p_2 - p_{11}/p_1 - p_{22} - p_{23})\\ &\quad\times dx_1dx_2dx_3dzdp_1dp_2dp_3dP_2dP_3\end{aligned}$$

and

$$\begin{aligned}\nu_2 &= \eta_2^*(d\omega_1^2\wedge\omega_1)\wedge\sigma_2^*\mu\\ &= (\exp(Z - P_2 - P_3)(P_2 - P_{22} - P_{23}) - P_{11})\\ &\quad\times dX_1dX_2dX_3dZdP_1dP_2dP_3dp_3dz.\end{aligned}$$

[The calculations are made easy if one observes that the factor of $\sigma_i^*\mu$ in ν_i means that one is essentially doing algebra modulo the ideal generated by $\sigma_i{}^*\Omega^2(R,{}^3R)$.] Thus this Bäcklund transformation relates the two differential equations

$$p_2 - p_{11}/p_1 - p_{22} - p_{33} = 0 \tag{4.4a}$$

and

$$\exp(Z - P_2 - P_3)(P_2 - P_{22} - P_{23}) - P_{11} = 0, \tag{4.4b}$$

where, if we hope to transform a solution of (4.4a) into a solution of (4.4b), we must asssume that the solution of (4.4a) satisfies $p_1 > 0$.

Clearly $f(x_1,x_2,x_3) = ae^{x_2} + bx_1$, $b>0$, a arbitrary, satisfies (4.4a) with $p_1>0$, and we will use it to arrive at a family of solutions to (4.4b). Recall that $M_1 = \sigma_1^{-1}[\mathrm{im}(j^2(f))]$ and $\delta_1\colon M_1\to E_1$ is the canonical injection. M_1 has coordinates x_1,x_2,x_3,P_2,P_3 and (4.3b) implies that

$$\delta_1^*\circ\eta_1^*\omega_2 = dP_2 + dP_3 - ae^{x_2}dx_1 - P_2dx_2 - P_3dx_3.$$

This may be written in the form (3.7) as

$$\begin{aligned}\delta_1^*\circ\eta_1^*\omega_2 = {}& e^{x_2}[d((P_2+P_3)e^{-x_2} - ax_1)\\ &- P_3e^{-x_2}d(x_3 - x_2)].\end{aligned}$$

So we can obtain integral manifolds of $\delta_1^*\circ\eta_1^*\omega_2$ by setting $x_3 - x_2 = C_1$ and $(P_2+P_3)e^{-x_2} - ax_1 = C_2$, C_1 and C_2 constants. However, since $x_3 - x_2 = X_3 - X_2$, this integral manifold is not transverse to the fiber of $\alpha\circ\pi_2{}^i\circ\eta_1\circ\delta_1$ and so does not yield a solution to (4.4b). We therefore rewrite $\delta_1^*\circ\eta_1^*\omega_2$ as

$$\begin{aligned}&\delta_1^*\circ\eta_1^*\omega_2\\ &= e^{x_2}\{d[(P_2+P_3)e^{-x_2} - ax_1 - P_3e^{-x_2}(x_3 - x_2)]\\ &\quad + (x_3 - x_2)d(P_3e^{-x_2})\}\end{aligned}$$

and obtain integral manifolds by setting $P_3e^{-x_2} = C_1$ and $(P_2+P_3)e^{-x_2} - ax_1 - P_3e^{-x_2}(x_3 - x_2) = C_2$, C_1 and C_2 constants. This implies that $P_3 = C_1e^{x_2}$,

$$P_2 = C_1(x_3 - x_2 - 1)e^{x_2} + ax_1e^{x_2} + C_2e^{x_2}.$$

This is an integral manifold transverse to the fibers of $\alpha\circ\pi_2\circ i\circ\eta_1\circ\delta_1$ and (4.1a) implies that $Z = \ln p_1 + P_2 + P_3 = \ln b + [C_1(x_3 - x_2) + ax_1 + C_2]e^{x_2}$. We therefore have obtained a two-parameter family of solutions,

$$f_{c_1,c_2}(X_1,X_2,X_3) = \ln b + [C_1(X_3 - X_2) + aX_1 + C_2]e^{X_2}, \tag{4.5}$$

of Eq. (4.4b).

To demonstrate the lack of uniqueness in the family of solutions to (4.4b), observe that one may also express $\delta_1^*\circ\eta_1^*\omega_2$ as

$$\begin{aligned}\delta_1^*\circ\eta_1^*\omega_2 = {}& (P_2 - ax_1e^{x_2})d(\ln|P_2 - ax_1e^{x_2}| - x_2)\\ &+ P_3d(\ln|P_3| - x_3),\end{aligned}$$

so that we can obtain integral manifolds by setting $\ln|P_3| - x_3 = C_1$ and $\ln|P_2 - ax_1e^{x_2}| - x_2 = C_2$. This implies that

$$P_3 = k_1e^{x_2},$$

$$P_2 = k_2e^{x_2} + ax_1e^{x_2},$$

where k_1,k_2 are arbitrary constants (k_1 may $=0$ if we let $P_3 = 0$ and k_2 may $=0$ if we let $P_2 - ax_1e^{x_2} = 0$ in the expression for $\delta_1^*\circ\eta_1^*\omega_2$). Again (4.1a) implies that $Z = \ln b + (k_2 + ax_1)e^{x_2} + k_1e^{x_2}$, yielding the family of solutions

$$f_{k_1,k_2}(X_1,X_2,X_3) = \ln b + (k_2 + aX_1)e^{X_2} + k_1e^{X_2} \tag{4.6}$$

to (4.4b). Clearly (4.5) and (4.6) are distinct families of solutions.

In an exactly similar manner one can begin with a solution of (4.4b), say $f(X_1,X_2,X_3) = \ln b(k_2 + aX_1)e^{X_2} + k_1e^{X_2}$, and obtain a family of solutions to (4.4a). For this f we find from (4.2b) that

$$\begin{aligned}\delta_2^*\circ\eta_2^*\omega_1 &= dz - bdX_1 - ae^{X_2}dX_2 - p_3dX_3\\ &= d(z - bX_1 - ae^{X_2}) - p_3dX_3.\end{aligned}$$

Since $x_3 = X_3$ we cannot have $X_3 =$ constant and expect a solution. Hence we rewrite

$$\delta_2^*\circ\eta_2^*\omega_1 = d(z - bX_1 - ae^{X_2} - p_3X_3) + X_3dp_3,$$

which yields the family of solutions

$$f_{c_1,c_2}(x_1,x_2,x_3) = bx_1 + ae^{x_2} + c_1x_3 + c_2.$$

It is easy to see that if we started with the solution $f(x_1,x_2,x_3) = bx_1 + ae^{x_2} + cx_3 + d$ to (4.4a) we would obtain the same families of solutions, (4.5) and (4.6), to (4.4b). If we then used either of these as the original solution to (4.4b) and worked back to a family of solutions to (4.4a) we would arrive at the same solution $f(x_1,x_2,x_3) = bx_1 + ae^{x_2} + cx_3 + d$. It is therefore not always possible to generate a new solution to one of the related equations by transforming to a solution of the related equation and then transforming that back to a solution of the original equation. One may well arrive at nothing new. In the case of the sine-Gordon equation on R^2, however, this procedure does work and the hope for such cases is one of the motivations for Bäcklund transformations.

Example 2: A Bäcklund transformation for second-order differential equations on R^2 will be a six-dimensional submanifold of the ten-dimensional product manifold $J^1(R^2,R)\times J^1(R^2,R)$. For such an example consider Ex. 1, p. 441, in Ref. 6. $i\colon S\to J^1(R^2,R)\times J^1(R^2,R)$ is defined implicitly by the four equations

$$\begin{aligned}x_1 &= X_1,\\ x_2 &= X_2,\\ p_1 &= \exp(z/(X_1+X_2) - (X_1+X_2)P_1)\\ &\quad + z/(X_1+X_2) - (X_1+X_2)P_1 + 1,\end{aligned} \tag{4.7}$$

$$p_2 = \exp(Z + z/(X_1 + X_2)) + Z + z/(X_1 + X_2) + 1,$$

where we assume throughout this example that $X_1 + X_2 = x_1 + x_2 \neq 0$.

Let $\alpha = \exp[z(X_1 + X_2) - (X_1 + X_2)P_1]$, $\beta = \exp[Z + z/(X_1 + X_2)]$, and let $P = \alpha + 1$, $Q = \beta + 1$. Then

$$e^{p_1 - P} = P - 1, \tag{4.8a}$$

$$e^{p_2 - Q} = Q - 1, \tag{4.8b}$$

which defines P as an everywhere increasing function of p_1 and Q as an everywhere increasing function of p_2, both functions > 1. This allows us to solve (4.7) for X_1, X_2, Z, P_1 as

$$\begin{aligned} X_1 &= x_1, \\ X_2 &= x_2, \\ Z &= \ln(Q(p_2) - 1) - z/(x_1 + x_2), \\ P_1 &= z/(x_1 + x_2)^2 - \ln(P(p_1) - 1)/(x_1 + x_2). \end{aligned} \tag{4.9}$$

(4.7) enables one to coordinatize S by X_1, X_2, Z, P_1, P_2, z, which shows that $\pi_2 \circ i$ is a submersion, while (4.9) enables one to coordinatize S by $x_1, x_2, z, p_1, p_2, P_2$, which shows that $\pi_1 \circ i$ is a submersion.

We have then that

$$\omega_1 = (\pi_1 \circ i)^* \omega = dz - \sum_{i=1}^{2} p_i dx_i, \tag{4.10a}$$

or

$$\begin{aligned} \omega_1 &= (\pi_1 \circ i)^* \omega \\ &= dz - [\exp(z/(X_1 + X_2) - (X_1 + X_2)P_1) \\ &\quad + z/(X_1 + X_2) - (X_1 + X_2)P_1 + 1]dX_1 \\ &\quad - [\exp(Z + z/(X_1 + X_2)) + Z \\ &\quad + z/(X_1 + X_2) + 1]dX_2 \end{aligned} \tag{4.10b}$$

and

$$\omega_2 = (\pi_2 \circ i)^* \omega = dZ - \sum_{i=1}^{2} P_i dX_i, \tag{4.11a}$$

or

$$\begin{aligned} \omega_2 &= (\pi_2 \circ i)^* \omega \\ &= d[\ln(Q(P_2) - 1) - z/(x_1 + x_2)] - [z/(x_1 + x_2)^2 \\ &\quad - \ln(P(p_1) - 1)/(x_1 + x_2)]dx_1 - P_2 dx_2, \end{aligned} \tag{4.11b}$$

depending on the coordinate system used for S.

The bundles E_1 and E_2 can be coordinatized respectively by x_i, z, p_i, P_2, p_{ij} and X_i, Z, P_i, z, P_{ij} $(1 \leqslant i \leqslant j \leqslant 2)$. From (4.8a) and (4.8b) one has that

$$p_1 - \ln[P(p_1) - 1] = P(p_1) \tag{4.12}$$

and

$$Q'(p_2) = \frac{Q(p_2) - 1}{Q(p_2)}, \tag{4.13}$$

where $Q'(p_2)$ denotes the derivative of Q with respect to p_2. Using these results and the above coordinate systems on E_1 and E_2 it is not hard to show that $\nu_1 = \eta_1^*(d\omega_2 \wedge \omega_2) \wedge \sigma_1^* \mu$ vanishes if and only if $p_{12} = P(p_1)Q(p_2)/(x_1 + x_2)$ and that $\nu_2 = \eta_2^*(d\omega_1 \wedge \omega_1) \wedge \sigma_2^* \mu$ vanishes if and only if $P_{12} + P_1/(X_1 + X_2) = Z/(X_1 + X_2)^2$. Hence this Bäcklund transformation relates the differential equations

$$p_{12} = P(p_1)Q(p_2)/(x_1 + x_2) \tag{4.14a}$$

and

$$P_{12} + P_1/(X_1 + X_2) = Z/(X_1 + X_2)^2. \tag{4.14b}$$

Let $f = f(x_1, x_2)$ be a local solution of (4.14a), let $M_1 = \sigma_1^{-1}[\mathrm{im}(j^2(f))]$, and let $\delta_1 : M_1 \to E_1$ be the canonical injection as usual. Then from (4.11b) we have

$$\begin{aligned} \delta_1^* \circ \eta_1^* \omega_2 = \frac{Q'(p_2)}{Q(p_2) - 1}(p_{21}dx_1 + p_{22}dx_2) \\ - \frac{p_1 dx_1 + p_2 dx_2}{x_1 + x_2} + \frac{z}{(x_1 + x_2)^2} d(x_1 + x_2) \\ - \left[\frac{z}{(x_1 + x_2)^2} - \frac{\ln(P(p_1) - 1)}{x_1 + x_2}\right] dx_1 \\ - P_2 dx_2. \end{aligned}$$

Using (4.12) and (4.13) this simplifies to

$$\begin{aligned} &\delta_1^* \circ \eta_1^* \omega_2 \\ &= \left[\frac{p_{21}}{Q(p_2)} - \frac{P(p_1)}{x_1 + x_2}\right] dx_1 \\ &\quad + \left[\frac{p_{22}}{Q(p_2)} - \frac{p_2}{x_1 + x_2} + \frac{z}{(x_1 + x_2)^2} - P_2\right] dx_2. \end{aligned}$$

Since f is a solution of (4.14a) the first term is zero and we are left with

$$\begin{aligned} &\delta_1^* \circ \eta_1^* \omega_2 \\ &= \left[\frac{p_{22}}{Q(p_2)} - \frac{p_2}{x_1 + x_2} + \frac{z}{(x_1 + x_2)^2} - P_2\right] dx_2. \end{aligned}$$

Because we are dealing with differential equations on R^2 (so that $d\omega_2^{n-1} \wedge \omega_2 = d\omega_2 \wedge \omega_2$) we know that

$$\delta_1^* \circ \eta_1^*(d\omega_2 \wedge \omega_2) = 0$$

and hence that regularity of $\delta_1^* \circ \eta_1^* \omega_2$ as defined in (1.8) is equivalent to nonsingularity of $\delta_1^* \circ \eta_1^* \omega_2$. But if $\delta_1^* \circ \eta_1^* \omega_2$ is nonsingular at p_0 then its only integral manifold through p_0 is obtained by setting x_2 constant, which means that it cannot be transverse to the fiber of $\alpha \circ \pi_2 \circ i \circ \eta_1 \circ \delta_1$ and hence yields no solution of (4.14b). However, $\delta_1^* \circ \eta_1^* \omega_2$ does have an integral manifold transverse to the fiber of $\alpha \circ \pi_2 \circ i \circ \eta_1 \circ \delta_1$, obtained by letting

$$P_2 = \frac{p_{22}}{Q(p_2)} - \frac{p_2}{x_1 + x_2} + \frac{z}{(x_1 + x_2)^2}.$$

In this case there are no parameters in the solution—it is unique and determined by (4.9) to be $Z = \ln(Q(p_2) - 1) - z/(x_1 + x_2)$.

Observe, however, that if one begins with a solution to (4.14b) he may obtain a one-parameter family of solutions to (4.14a). In particular, one can show that the zero solution of (4.14b) leads to the one-parameter family of solutions to (4.14a) defined implicitly by

$$\exp(z/(x + y))/(x + y)[\exp(z/(x + y)) + 1] = C,$$

where C is a positive constant if $x + y > 0$, and a negative constant if $x + y < 0$.

Example 3: Until now our examples have identified the base coordinates x_i and X_i, $i = 1, \ldots, n$. In this example they are not identified and moreover the differential equations related by the transformation are identical.

Let $i: S \to J^1(R^3,R) \times J^1(R^3,R)$ be defined implicitly by the five equations $X_1 = x_2 + p_3, X_2 = x_1 - P_3, X_3 = x_1 + p_2, P_1 = -p_1, P_2 = x_3 - x_2 - p_3$. Thus S may be coordinatized by x_i, z, p_i, Z, P_3 or by X_i, Z, P_i, z, p_3 $(i = 1,2,3)$ as

$$
\begin{array}{lll}
\text{(a)} & x_i = x_i, & \text{(b)} \quad x_1 = X_2 + P_3, \\
& z = z, & x_2 = X_1 - p_3, \\
& p_i = p_i, & x_3 = X_1 + P_2, \\
& X_1 = x_2 + p_3, & z = z, \\
& X_2 = x_1 - P_3, & p_1 = -P_1, \\
& \text{or} & (i = 1,2,3), \qquad (4.15) \\
& X_3 = x_1 + p_2, & p_2 = X_3 - X_2 - P_3, \\
& Z = Z, & p_3 = p_3, \\
& P_1 = -p_1, & X_i = X_i, \\
& P_2 = x_3 - x_2 - p_3, & Z = Z, \\
& P_3 = P_3, & P_i = P_i,
\end{array}
$$

respectively. From the symmetry of (4.15a) and (4.15b) it is clear that the differential equations related by the transformation S must be identical. Calculating as in the previous examples one finds that the differential equation related to itself by S is

$$
\begin{aligned}
0 = {} & p_{11} - p_{13} - p_{22} - p_{23} + 2p_{11}p_{23} - 2p_{12}p_{13} \\
& - p_{23}^2 + p_{22}p_{33} + p_{12}p_{33} - p_{13}p_{23} + p_{11}p_{23}^2 \\
& - 2p_{12}p_{13}p_{23} - p_{11}p_{22}p_{23} + p_{22}p_{13}^2 + p_{33}p_{12}^2. \qquad (4.16)
\end{aligned}
$$

The function

$$
f(x_1,x_2,x_3) = \tfrac{1}{2}(x_1^2 + x_2^2 + x_3^2) \qquad (4.17)
$$

satisfies (4.16) and, with the same notation as before, we find that

$$
\begin{aligned}
\delta_1^* \circ \eta_1^* \omega_2 &= dZ + x_1 d(x_2 + x_3) - (x_3 - x_2 - x_3) d(x_1 - P_3) \\
&\quad - P_3 d(x_1 + x_2) \\
&= d(Z + x_1x_2 - x_2P_3 + x_1x_3) - (x_3 + P_3)dx_1.
\end{aligned}
$$

So $x_1 = X_2 + P_3 = C_1$ and $Z + x_1x_2 - x_2P_3 + x_1x_3 = C_2$, C_1 and C_2 constants, determine an integral manifold of $\delta_1^* \circ \eta_1^* \omega_2$ which we can write, after appropriate substitutions from (4.15), as

$$
\begin{aligned}
& P_3 = C_1 - X_2, \\
& Z = C_1P_2 - (C_1 - X_2)P_2 - C_1(X_1 + P_2) + C_2.
\end{aligned}
$$

This is paramaterized by X_1, X_2, P_2 which, since P_2 is one of the coordinates of the fiber of $\alpha \circ \pi_2 \circ \eta_1 \circ \delta_1$, might lead one to believe that this integral manifold is not transverse to the fiber of $\alpha \circ \pi_2 \circ i \circ \eta_1 \circ \delta_1$ and hence that it yields no solution of (4.16). However, for the solution (4.17) one has $p_i = x_i$ $(i = 1,2,3)$ and hence from (4.15a) $P_2 = x_3 - x_2 - p_3 = -x_2 = -p_2$ and $X_3 = x_1 + p_2$. Together these imply that $P_2 = x_1 - X_3$. Then, since $x_1 = C_1$ on our integral manifold, we have $P_2 = C_1 - X_3$ which, after appropriate substitution, yields

$$
\begin{aligned}
& P_2 = C_1 - X_3, \\
& Z = C_1(X_3 + X_2 - X_1) - X_2X_3 + (C_2 - C_1^2).
\end{aligned}
$$

Thus we have the two-parameter family of solutions

$$
f_{c_1,c_2}(X_1,X_2,X_3) = C_1(X_3 + X_2 - X_1) - X_2X_3 + (C_2 - C_1^2) \qquad (4.18)
$$

to (4.16).

The nonuniqueness of the transformed family of solutions can again be demonstrated if we write $\delta_1^* \circ \eta_1^* \omega_2$ in the form

$$
\begin{aligned}
\delta_1^* \circ \eta_1^* \omega_2 = {} & d(Z + x_1x_2 - x_2P_3 + x_1x_3 \\
& - (x_3 - P_3)x_1) + x_1 d(x_3 + P_3)
\end{aligned}
$$

and obtain an integral manifold of $\delta_1^* \circ \eta_1^* \omega_2$ by setting

$$
\begin{aligned}
& Z + x_1x_2 - x_2P_3 + x_1x_3 - (x_3 + P_3)x_1 = C_2, \\
& P_3 + x_3 = C_1.
\end{aligned}
$$

Again, since $p_i = x_i$ $(i = 1,2,3)$ for the solution (4.17), (4.15a) implies that $X_1 = x_2 + x_3$, $X_2 = x_1 - P_3$, $X_3 = x_1 + x_2$, from which one can deduce that $x_3 = X_1 - X_3 + X_2 + P_3$. Hence $P_3 + x_3 = C_1$ implies that

$$
P_3 = \tfrac{1}{2}(X_3 - X_1 - X_2 + C_1)
$$

and, upon appropriate substitution, that

$$
\begin{aligned}
Z = {} & -\tfrac{1}{2}C_1(X_1 - X_2 - X_3) - \tfrac{1}{2}(X_1X_2 + X_1X_3 + X_2X_3) \\
& + \tfrac{1}{4}(X_1^2 + X_2^2 + X_3^2) + C_2.
\end{aligned}
$$

The latter determines a two-parameter family of solutions to (4.16) distinct from that of (4.18).

[1] R. Gardner, Commun. Pure Appl. Math. **22**, 597 (1969).
[2] R. Gardner, Trans. Am. Math. Soc. **126**, 514 (1967).
[3] A.R. Forsyth, *Theory of Differential Equations*, Part 1 (Dover, New York, 1959).
[4] E. Cartan, *Les Systems Differentiels Exterieurs et Leurs Applications Geometriques* (Hermann, Paris, 1945).
[5] R. Gardner, J. Differential Geometry **2**, 25 (1968).
[6] A.R. Forsyth, *Theory of Differential Equations*, Part 6 (Dover, New York, 1959).
[7] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. 2 (Interscience, New York, 1962).

Elementary solutions of the linear transport equation for continuously varying spatial media

G. C. Pomraning
School of Engineering and Applied Science, University of California, Los Angeles, California 90024

E. W. Larsen
Theoretical Division, University of California Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87544

(Received 2 October 1979; accepted for publication 7 March 1980)

We demonstrate that it is possible to construct elementary solutions (eigenfunctions) of the linear transport equation for certain types of continuously varying spatial media. In general, both discrete and continuum modes result, which appear to be complete on the half-range. A detailed analysis is given for an "exponential" medium, including numerical results and a half-range completeness proof. A "linear" medium is also considered. A general method is presented for constructing, jointly, the spatial variation of a medium and the corresponding functional forms of the eigenfunctions. Our results represent a partial generalization of the singular eigenfunction technique to media with continuous spatial variation.

I. INTRODUCTION

One of the outstanding achievements in linear transport theory in the last twenty years has been the development of the singular eigenfunction method of solution. The use of these eigenfunctions was first suggested by Davidson,[1] and independently proposed by Wigner,[2] and Van Kampen, both in the plasma[3] and neutron transport[4] context. The classic paper in this area is due to Case.[5] It was Case's paper, which appeared in 1960, that led to the full development of the method for neutron transport problems. Many extensions of Case's early work have been made, including extensions to multigroup problems, anisotropic scattering, multiple regions, finite problems, criticality (eigenvalue) problems, etc. Excellent reviews of the progress in this area since 1960 are given by Case and Zweifel,[6] McCormick and Kuscer,[7] and Greenberg and Zweifel.[8] Although developed in the neutron transport arena, the singular eigenfunction method has found application is plasma physics,[9] radiative transfer,[10] and kinetic theory.[11] The papers sired by the original Case paper are too numerous to reference here, but number in the hundreds.

However, except for one recent article,[12] the explicit construction of elementary solutions (eigenfunctions) of the transport equation for media whose cross sections vary continuously in space (nonhomogeneous media) is a generalization of the singular eigenfunction method which has eluded researchers. In the above-mentioned article,[12] Mullikin and Siewert explicitly construct, for a special nonhomogenous-medium transport equation (containing one free parameter,s), a continuum of elementary solutions which reduce to the standard continuum solutions as the spatial variations tend to zero ($s\to\infty$). (Mullikin and Siewert do not use these elementary solutions to generate solutions of the full transport equation but, instead, to develop a singular integral equation which the angular flux exiting from a halfspace must satisfy. This equation was derived earlier, and in a different way, by Martin.[13] Also, in Ref. 12, Mullikin and Siewert generalize Martin's method to derive a singular integral equation which the flux exiting from a very general nonhomogeneous one-dimensional medium must satisfy. However, this method does not explicitly involve elementary solutions of the appropriate transport equation, and we shall not discuss it further here.)

In part of what follows we shall extend some of the ideas originated by Mullikin and Siewert.[12] However, the broader goal of this article is to demonstrate that the construction of elementary solutions of the transport equation for continuously variable spatial media is explicitly possible for a rich variety of such media. For the spatial variations considered, we analytically and explicitly construct both "continuum" and "discrete" eigensolutions. That is, these solutions converge to the standard homogeneous-medium continuum and discrete eigensolutions in the limit as the spatial variations in the media tend to zero.

Our results are obtained for the time-independent, one-speed, slab geometry transport equation with isotropic scattering, which can be written[6]

$$\mu\frac{\partial\psi(z,\mu)}{\partial z}+\psi(z,\mu)=\tfrac{1}{2}c(z)\int_{-1}^{1}d\mu'\,\psi(z,\mu'). \tag{1.1}$$

For the case $c(z)=\text{const}$, separation of variables in Eq. (1.1) leads to a "continuum plus two discrete" eigensolutions.[5] These solutions have been shown to be complete for both full-range ($-1\leqslant\mu\leqslant 1$) and half-range ($0<\mu\leqslant 1$) problems. Moreover, full- and half-range formulas for the expansion coefficients in an eigenfunction expansion can be written explicitly.

For $c(z)\neq\text{const}$, our results as reported here are not as complete as we would like in two respects. First, for the spatial variations considered, we generally obtain only one discrete mode and what seems to be only a portion of the continuum. Nevertheless, in all cases there *appears*, to be enough solutions to solve the problem of a source-free half-space with an incident flux on the boundary (the albedo problem). That is, the eigenfunctions we obtain appear to be complete on the half-range. The second difficulty is that the

 0022-2488/80/071603-10$1.00

singular integral equations which arise in determining the appropriate half-range expansion coefficients do not seem to be explicitly solvable. In addition, the methods needed to prove existence of a solution, and thus guarantee half-range completeness, are dependent upon the particular $c(z)$ under consideration. However, for the one functional form of $c(z)$ which we have investigated in the most detail, we have been able to prove half-range completeness for sufficiently weak spatial variations. Also, for this $c(z)$ we have constructed numerical solutions of the singular integral equations for the half-range coefficients which converge nicely as the numerical approximation is refined. Thus, for this particular $c(z)$ we have strong evidence that the corresponding eigenfunctions always satisfy a standard half-range completeness criterion.

The outline of this paper is as follows. In the next section we consider media whose spatial variations are characterized by

$$c(z) = \frac{c + kbe^{-z/s}}{1 + be^{-z/s}}, \quad s > 0, \tag{1.2}$$

where b, c, k and s are constants. Our analysis of Eq. (1.1), with $c(z)$ given by Eq. (1.2), leads to a "continuum plus one discrete" solution in general. However, for sufficiently small s, a "continuum plus two discrete" solution is extant. In Sec. III a half-range completeness proof is given for $|b|$ sufficiently small. Section IV presents numerical results which show that the singular integral equation for the expansion coefficients, arising from attempting to solve Eq. (1.1) with $c(z)$ given by Eq. (1.2) and boundary conditions

$$\psi(0,\mu) = f(\mu), \quad 0 < \mu \leqslant 1, \tag{1.3}$$

$$\lim_{z\to\infty} \psi(z,\mu) = 0, \tag{1.4}$$

has a well-behaved solution for various (not necessarily small) values of b. Taken together, Secs. III and IV constitute strong evidence that the eigenfunctions constructed in Sec. II are complete on the half-range for any value of b.

The remainder of the paper is more preliminary in nature, suggesting other functional forms for $c(z)$ which might lead to eigenfunctions complete on the half-range. In Sec. V we describe a method for obtaining a self-consistent functional form for $c(z)$ and one (or more) corresponding discrete eigensolutions of Eq. (1.1). The form of this discrete eigensolution suggests the proper form of the continuum modes for this $c(z)$. One of the possibilities considered is the bilinear form

$$c(z) = \left(\frac{a+z}{b+z}\right) c, \tag{1.5}$$

and in Sec. VI we explicitly construct for this case a "continuum plus one discrete" solution. The continuum eigenfunctions is this case involve higher order distributions (e.g., $1/x^2$) than encountered in the $c(z) = \text{const}$ or $c(z)$ given by Eq. (1.2) cases. We conclude the paper with a discussion in Sec. VII.

II. AN "EXPONENTIAL" MEDIUM

In this section we treat Eq. (1.1) with $c(z)$ defined by Eq. (1.2). The transport equation can then be written as

$$(1 + be^{-z/s})\left(\mu \frac{\partial\psi}{\partial z} + \psi\right) = (c + kbe^{-z/s}) \tfrac{1}{2} \int_{-1}^{1} \psi \, d\mu'. \tag{2.1}$$

We seek solutions of this equation having the form

$$\psi_\nu(z,\mu) = f_\nu(\mu) e^{-z/\nu} + b g_\omega(\mu) e^{-z/\omega}, \tag{2.2}$$

where ω is defined in terms of ν by

$$\frac{1}{\omega} = \frac{1}{\nu} + \frac{1}{s}. \tag{2.3}$$

(In the work of Mullikin and Siewert on elementary solutions,[12] $c(z) = c_0 \exp(-z/s)$. This can be obtained here by taking the limits $b\to 0$, $c\to 0$, and $bk\to c_0$ in Eq. (2.1). Also, our assumed solution form given by Eqs. (2.2) and (2.3) is identical to that in Ref. 12. Therefore, the results in this section are an extension of those of Mullikin and Siewert.)

Introducing Eq. (2.2) into Eq. (2.1) and equating the coefficients of $\exp[-(1/\nu + (n/s))z]$ for $n = 0,1$, and 2 yields the following three equations:

$$\left(1 - \frac{\mu}{\nu}\right) f_\nu(\mu) = \frac{c}{2} \int_{-1}^{1} f_\nu(\mu') \, d\mu', \tag{2.4}$$

$$\left(1 - \frac{\mu}{\nu}\right) f_\nu(\mu) + \left(1 - \frac{\mu}{\omega}\right) g_\omega(\mu) = \frac{k}{2} \int_{-1}^{1} f_\nu(\mu') \, d\mu' + \frac{c}{2} \int_{-1}^{1} g_\omega(\mu') \, d\mu', \tag{2.5}$$

$$\left(1 - \frac{\mu}{\omega}\right) g_\omega(\mu) = \frac{k}{2} \int_{-1}^{1} g_\omega(\mu') \, d\mu'. \tag{2.6}$$

Equation (2.4) has the usual solutions corresponding to $c(z) = \text{const}$, namely

$$f_\nu(\mu) = \lambda(c,\nu)\delta(\nu - \mu) + \frac{c\nu}{2} P \frac{1}{\nu - \mu} \equiv \phi(c,\nu,\mu), \tag{2.7}$$

$$\lambda(c,\mu) = 1 - \frac{c\nu}{2} P \int_{-1}^{1} \frac{d\mu'}{\nu - \mu'}, \tag{2.8}$$

which satisfy

$$\int_{-1}^{1} f_\nu(\mu) d\mu = 1, \tag{2.9}$$

$$(1 - (\mu/\nu)) f_\nu(\mu) = c/2. \tag{2.10}$$

Here the eigenvalue ν can be any of the usual continuum ($-1 < \nu < 1$) or discrete ($\nu = \pm \nu_0$) modes defined by the dispersion law $\lambda(c,\nu_0) = 0$.

We now introduce Eqs. (2.9) and (2.10) into Eqs. (2.5) and (2.6) and simplify, to obtain the following two remarkably simple equations:

$$(1 - (\mu/\omega)) g_\omega(\mu) = k/2, \tag{2.11}$$

$$\int_{-1}^{1} g_\omega(\mu') \, d\mu' = 1. \tag{2.12}$$

If ω were a completely free parameter, these equations would have the solution

$$g_\omega(\mu) = \phi(k,\omega,\mu), \tag{2.13}$$

with $-1 \leqslant \omega \leqslant 1$ and $\omega = \pm \omega_0$, where ω_0 and k are related by the usual dispersion formula

$$0 = \lambda(k,\omega_0) = 1 - \tfrac{1}{2} k\omega_0 \int_{-1}^{1} \frac{d\mu'}{\omega_0 - \mu'}. \tag{2.14}$$

However, ω is defined in terms of ν by Eq. (2.3); thus the above conditions on the values of ν and ω leads to a restricted set of values of ν for which solutions of the form given by (2.3) are possible.

In particular, the set of continuum ($-1\leqslant\nu\leqslant 1$) values of ν which are consistent with the condition $-1\leqslant\omega\leqslant 1$ are

$$-\left(\frac{s}{s+1}\right)\leqslant\nu\leqslant 1. \tag{2.15}$$

These values of ν correspond to values of ω in the interval

$$-1\leqslant\omega\leqslant\frac{s}{s+1}. \tag{2.16}$$

Thus, continuum solutions of the form derived above exist for ν satisfying the inequalities Eq. (2.15).

Next, if $\pm\nu_0$ are the two discrete roots appropriate to Eq. (2.7), then the corresponding discrete values of ω_0 are defined by

$$1/\omega_0{}^{\pm} = \pm(1/\nu_0 + 1/s). \tag{2.17}$$

Since $\omega_0^- \neq -\omega_0^+$, Eq. (2.14) cannot be satisfied by *both* ω_0^- and ω_0^+. However Eq. (2.14) can be satisfied provided only one of the discrete solutions is desired, by selecting k appropriately. It is convenient here to choose the solution which decays at $z = +\infty$; and thus we take $\nu = +\nu_0$; then $\omega = \omega_0 = \omega_0^+$ is defined by Eq. (2.3) and k is defined by Eq. (2.14). (Thus, "k is the value of $c(z) = \text{const}$ which corresponds to the positive discrete eigenvalue ω_0.") The resulting solution is real and bounded provided (i) $c < 1$ (then ν_0, ω_0, and k are all real), and (ii) $s > \nu_0/(\nu_0 - 1)$ (then $\omega_0 > 1$, and so $g_{\omega_0}(\mu)$ is bounded). Alternately, if k is given and $0 < k < c$, then Eq. (2.14) with $\omega = \omega_0^+$ defined by Eq. (2.17) provides an equation for s.

To summarize, we have constructed solutions of Eq. (2.1) of the form

$$\psi_\nu(z,\mu) = \phi(c,\nu,\mu)e^{-z/\nu} + b\phi(k,\omega,\mu)e^{-z/\omega}, \tag{2.18}$$

where $\phi(c,\nu,\mu)$ are the usual eigenfunctions [defined by Eq. (2.7)], ω is defined in terms of ν by Eq. (2.3), and c, s, and b are any constants satisfying:

$$\text{(i) } 0 < c < 1, \tag{2.19}$$

$$\text{(ii) } s > \nu_0/(\nu_0 - 1), \tag{2.20}$$

$$\text{(iii) } -\infty < b < +\infty. \tag{2.21}$$

Equation (2.18) holds for $-s/(s+1) < \nu < 1$ and for $\nu = \nu_0$. The discrete root $\nu = \nu_0$ defines a discrete value $\omega = \omega_0$ by Eq. (2.3), and then k and s are related by Eq. (2.14).

We conclude this section with two remarks. First, by setting $b = 0$ in Eqs. (2.1) and (2.18), it is clear that $c(z) = c = \text{const}$ and the $\psi_\nu(z,\mu)$ reduce to the usual eigensolutions for $-s/(s+1) < \nu \leqslant 1$ and $\nu = \nu_0$. Since half-range completeness for $0 < \mu \leqslant 1$ has been proved for this ($b = 0$) case, the possibility exists that half-range completeness for $0 < \mu \leqslant 1$ can be proved for $|b|$ sufficiently small. We show in the next section that such a proof is indeed possible. (In addition, we can obtain the usual homogeneous media eigenfunctions by letting $s \to +\infty$.)

Second, we have imposed the condition Eq. (2.20) to guarantee that the solution Eq. (2.18) corresponding to $\nu = \nu_0$ is bounded. If we consider

$$0 < s < \nu_0/(\nu_0 - 1), \tag{2.22}$$

then $0 < \omega_0^+ < 1$ and one can show that ψ_ν, defined by (2.18) is still a solution for any k. However, the term $\phi(k,\omega_0^+,\mu)$ is now a distribution and is not bounded. For this situation, one can construct a second discrete eigensolution which corresponds to $-\nu_0$. This solution has the form of Eq. (2.18), with $\nu = -\nu_0$ and $\omega = \omega_0^-$. Now there are two cases. If $-1 \leqslant \omega_0^- \leqslant 1$, then $\phi(k,\omega_0^-,\mu)$ will be unbounded and k will be arbitrary; otherwise, $\phi(k,\omega_0^-,\mu)$ will be bounded and k will be defined by the dispersion law

$$0 = \lambda(k,\omega_0^-). \tag{2.23}$$

Thus, for s satisfying Eq. (2.22), one can always construct two "discrete" modes; at least one, and possibly both, of these modes contain a distributional part which decays spatially like a continuum mode. Presumably, in the solution of a half-space problem, the expansion coefficients for the continuum modes would have to be unbounded to cancel the unboundedness in the discrete mode. In fact, it appears that both continuum and discrete modes can be found for arbitrary parameters in Eq. (1.2) if the ansatz given by Eq. (2.2) is extended from two to N terms according to

$$\psi_\nu(z,\mu) = \sum_{n=0}^{N-1} f_n(\mu)e^{-z/\nu_n}, \tag{2.24}$$

where

$$1/\nu_n = (1/\nu) + (n/s), \tag{2.25}$$

and the value of N depends upon the parameters. The resulting discrete mode(s) may or may not be unbounded per the discussion above, depending upon the precise values of the parameters b, c, k, and s. A detailed analysis of this more general problem will not be presented here, but will be considered in a future article.

III. HALF-RANGE COMPLETENESS AND FULL-RANGE ORTHOGONALITY FOR THE "EXPONENTIAL" MEDIUM SOLUTIONS

We consider here the boundary value problem

$$\mu\frac{\partial\psi}{\partial z} + \psi = \tfrac{1}{2}\left(\frac{c + kbe^{-z/s}}{1 + be^{-z/s}}\right)\int_{-1}^{1}\psi\,d\mu', \quad 0 < z < +\infty, \tag{3.1}$$

$$\psi(0,\mu) = f(\mu), \quad 0 < \mu \leqslant 1, \tag{3.2}$$

$$\psi(+\infty,\mu) = 0, \tag{3.3}$$

where c, b, and s satisfy Eqs. (2.19)–(2.21) and k is defined by Eq. (2.14). The general solution, which decays as $z \to \infty$, and which is constructible from the elementary solutions found in Sec. II, is

$$\psi(z,\mu) = a(\nu_0)\psi_{\nu_0}(z,\mu) + \int_0^1 a(\nu)\psi_\nu(z,\mu)\,d\nu, \tag{3.4}$$

where ψ_ν is defined by Eq. (2.18) and $a(\nu)$ is the expansion coefficient. This is a solution of the problem given by Eqs. (3.1)–(3.3) if and only if the boundary condition Eq. (3.2) is satisfied

$$f(\mu) = a(\nu_0)\phi(c,\nu_0,\mu) + \int_0^1 a(\nu)\phi(c,\nu,\mu)\,d\nu$$

$$+ b\left[a(\nu_0)\phi(k,\omega_0,\mu) + \int_0^1 a(\nu)\phi(k,\omega,\mu)\,d\nu\right],$$

$$0<\mu\leqslant 1. \tag{3.5}$$

To prove that this equation has a unique solution $a(\nu)$ for $|b|$ sufficiently small, we need to define two Banach spaces. For any $p>1$, let X_p be the space of functions $\xi(\mu)$ defined for $0<\mu\leqslant 1$ which satisfy

$$\|\xi\|_{X_p} \equiv \left(\int_0^1 |\mu\xi(\mu)|^p\,d\mu\right)^{1/p} < \infty. \tag{3.6}$$

Next, for any $0<c<1$, let $Y_{p,c}$ be the space of functions $\eta(\nu)$ defined for $0<\nu\leqslant 1$ and $\nu=\nu_0$, which satisfy

$$\|\eta\|_{Y_{p,c}} = \left(|\eta(\nu_0)|^p + \int_0^1 |\eta(\nu)\ln(1-\nu)|^p\,d\nu\right)^{1/p}. \tag{3.7}$$

For any $0<c<1$ and any function $\eta\in Y_{p,c}$, Larsen, Sancaktar, and Zweifel[14] have shown that the function ξ defined by

$$\xi(\mu) = (L_c\eta)(\mu)$$

$$= \eta(\nu_0)\phi(c,\nu_0,\mu) + \int_0^1 \eta(\nu)\phi(c,\nu,\mu)\,d\nu \tag{3.8}$$

is a function in X_p. Moreover, the operator $L_c: Y_{p,c}\to X_p$ is bounded, and its inverse exists and is also bounded.

Now we use Eq. (2.3) to obtain, by a change of variables,

$$\int_0^1 a(\nu)\phi(k,\omega,\mu)\,d\nu = \int_0^{s/(s+1)} a\left(\frac{\omega s}{s-\omega}\right)\left(\frac{s}{s-\omega}\right)^2 \times\phi(k,\omega,\mu)\,d\omega. \tag{3.9}$$

Let us define the operator T: $Y_{p,c}\to Y_{p,k}$ by

$$(Ta)(\omega) = \begin{cases} \left(\dfrac{s}{s-\omega}\right)^2 a\left(\dfrac{\omega s}{s-\omega}\right), & 0<\omega<\dfrac{s}{s+1}, \\ 0, & \dfrac{s}{s+1}\leqslant\omega\leqslant 1, \\ a(\nu_0), & \omega=\omega_0. \end{cases} \tag{3.10}$$

It is a trivial calculation to show that T is a bounded operator satisfying

$$\|T\|\leqslant\max\left[1,\left(\frac{1+s}{s}\right)^{(1-2/p)}\right]. \tag{3.11}$$

Equations (3.9) and (3.10) imply

$$a(\nu_0)\phi(k,\omega_0,\mu) + \int_0^1 a(\nu)\phi(k,\omega,\mu)\,d\nu$$

$$= (Ta)(\omega_0)\phi(k,\omega_0,\mu) + \int_0^1 (Ta)(\omega)\phi(k,\omega,\mu)\,d\omega$$

$$= (L_k Ta)(\mu). \tag{3.12}$$

Now Eq. (3.5) can be written as

$$f(\mu) = (L_c a)(\mu) + b(L_k Ta)(\mu), \tag{3.13}$$

or

$$[I + b(L_c^{-1}L_k T)]a(\nu) = (L_c^{-1}f)(\nu). \tag{3.14}$$

If b satisfies

$$|b| < (\|L_c^{-1}\|\|L_k\|\|T\|)^{-1}, \tag{3.15}$$

then

$$\|bL_c^{-1}L_k T\|\leqslant b\,\|L_c^{-1}\|\|L_k\|\|T\| < 1,$$

and thus the integral equation (3.14) has a unique solution, which can be expressed as a Neumann Series. This solution is exactly the half-range coefficients necessary to solve the problem given by Eqs. (3.1)–(3.3). This completes the proof that the "$\nu>0$" solutions $\psi_\nu(0,\mu)$ satisfy the usual half-range completeness criterion for the interval $0<\mu\leqslant 1$.

Next we shall derive a full-range orthogonality relationship which will allow a solution to the half-space albedo problem via Siewert's integral equation method.[15] Let $\psi(z,\mu)$ and $\theta(z,\mu)$ be any solutions of Eq. (1.1) which vanish as $z\to+\infty$. Then $\psi(z',\mu)$ and $\theta(z',-\mu)$ satisfy

$$\mu\frac{\partial}{\partial z'}\psi(z',\mu) + \psi(z',\mu) = \frac{c(z')}{2}\int_{-1}^1 \psi(z',\mu')\,d\mu', \tag{3.16}$$

$$-\mu\frac{\partial}{\partial z'}\theta(z',-\mu) + \theta(z',-\mu) = \frac{c(z')}{2}\int_{-1}^1 \theta(z',-\mu')\,d\mu'. \tag{3.17}$$

Multiplying Eq. (3.16) by $\theta(z',-\mu)$, Eq. (3.17) by $\psi(z',\mu)$, subtracting the resulting equations, and integrating over $-1\leqslant\mu\leqslant 1$ and $z\leqslant z'<\infty$ gives

$$\int_{-1}^1 \mu\psi(z,\mu)\theta(z,-\mu)\,d\mu = 0. \tag{3.18}$$

Now, if we specialize this result to the function $c(z)$ considered in Sec. II and take $\psi(z,\mu)=\psi_\nu(z,\mu)$ and $\theta(z,\mu)=\psi_{\nu'}(z,\mu)$ for $\nu>0$ and $\nu'>0$, we obtain

$$\int_{-1}^1 \mu\psi_\nu(z,\mu)\psi_{\nu'}(z,-\mu)\,d\mu = 0, \tag{3.19}$$

or

$$\int_{-1}^1 \mu[f_\nu(\mu)+Kg_\omega(\mu)][f_{\nu'}(-\mu)+Kg_{\omega'}(-\mu)]\,d\mu = 0, \tag{3.20}$$

where K is any constant. Equation (3.20) [or Eq. (3.19)] is the desired orthogonality condition.

IV. NUMERICAL RESULTS

The half-range completeness proof given in the last section is valid for $|b|$ sufficiently small. To strengthen our belief that these eigenfunctions are complete for any value of b, we solve the singular integral equation (3.5) for the expansion coefficient $a(\nu)$, using a straightforward numerical method. The purpose here is not to suggest an efficient numerical scheme, but rather to demonstrate that a solution can be obtained, thus giving experimental evidence of half-range completeness.

We proceed as follows. We form the first $N+1$ angular moments of Eq. (3.5) by multiplying this equation by μ^n and integrating over $0<\mu\leqslant 1$. This gives

$$f_n = a(\nu_0)[\phi_n(c,\nu_0) + b\phi_n(k,\omega_0)] + \int_0^1 d\nu\; a(\nu)[\phi_n(c,\nu) + b\phi_n(k,\omega)], \quad 0\leqslant n\leqslant N, \tag{4.1}$$

TABLE I. The half-space Albedo.

$f(\mu)$	$b \backslash N$	2	4	8	16
1	0.	0.630684	0.630583	0.630581	0.630581
	0.1	0.579865	0.579721	0.579718	0.579718
	1.0	0.399516	0.399083	0.399078	0.399078
	10.0	0.271905	0.271099	0.271058	0.271058
μ	0.	0.613733	0.614110	0.614117	0.614117
	0.1	0.562881	0.563300	0.563308	0.563308
	1.0	0.382885	0.383297	0.383308	0.383308
	10.0	0.256150	0.256363	0.256351	0.256351
μ^2	0.	0.604532	0.604750	0.604754	0.604754
	0.1	0.553782	0.554047	0.554052	0.554052
	1.0	0.374310	0.374655	0.374660	0.374660
	10.0	0.248169	0.248449	0.248446	0.248446

where we have defined

$$f_n \equiv \int_0^1 d\mu\, \mu^n f(\mu), \tag{4.2}$$

$$\phi_n(c,\nu) \equiv \int_0^1 d\mu\, \mu^n \phi(c,\nu,\mu). \tag{4.3}$$

We perform the integration in Eq. (4.1) by using N-point Gauss quadrature on the interval $0 \leqslant \nu \leqslant 1$. This yields the matrix equation

$$f_n = \sum_{m=0}^{N} A_{nm} a_m, \tag{4.4}$$

where we have defined the matrix elements

$$A_{nm} = w_m [\phi_n(c,\nu_m) + b\phi_n(k,\omega_m)], \quad m \neq 0, \tag{4.5}$$

$$A_{n0} = \phi_n(c,\nu_0) + b\phi_n(k,\omega_0). \tag{4.6}$$

Here we have also defined

$$a_m \equiv a(\nu_m), \quad a_0 \equiv a(\nu_0), \tag{4.7}$$

where the ν_m, $m \leqslant 1 \leqslant N$, are the Gauss quadrature points with corresponding weights w_m. The points ω_m are given by

$$\omega_m = s\nu_m/(s + \nu_m). \tag{4.8}$$

The $\phi_n(c,\nu)$ are conveniently computed from a recurrence relationship. In both the continuum and discrete cases, we have

$$(\nu - \mu)\phi(c,\nu,\mu) = c\nu/2. \tag{4.9}$$

Multiplying Eq. (4.9) by μ^n and integrating gives

$$\phi_{n+1}(c,\nu) = \nu[\phi_n(c,\nu) - c/2(n+1)], \quad n \geqslant 0. \tag{4.10}$$

The results for $n = 0$ needed to initialize Eq. (4.10) are computed by direct integration. We obtain

$$\phi_0(c,\nu_0) = \frac{c\nu_0}{2} \ln\left(\frac{\nu_0}{\nu_0 - 1}\right), \tag{4.11}$$

$$\phi(c,\nu) = 1 - \frac{c\nu}{2} \ln\left(\frac{1+\nu}{\nu}\right), \quad 0 < \nu \leqslant 1. \tag{4.12}$$

Once the expansion coefficient $a(\nu)$, or the discrete form a_m, is known, we can compute the half-space albedo R (reflection probability) as follows. The albedo is defined as

$$R \equiv \frac{J_{\text{out}}}{J_{\text{in}}} = \frac{1}{J_{\text{in}}} \int_{-1}^{0} d\mu\, |\mu| \psi(0,\mu), \tag{4.13}$$

where

$$J_{\text{in}} \equiv \int_0^1 d\mu\, \mu f(\mu) = f_1. \tag{4.14}$$

Equation (4.13) can alternatively be written as

$$R = 1 - \frac{1}{f_1} \int_{-1}^{1} d\mu\, \mu \psi(0,\mu). \tag{4.15}$$

From Eq. (3.4) we have

$$\int_{-1}^{1} d\mu\, \mu \psi(0,\mu) = \int_{-1}^{1} d\mu\, \mu a(\nu_0)[\phi(c,\nu_0,\mu) + b\phi(k,\omega_0,\mu)] + \int_{-1}^{1} d\mu\, \mu \int_0^1 d\nu\, a(\nu)[\phi(c,\nu,\mu) + b\phi(k,\omega,\mu)]. \tag{4.16}$$

Integration of Eq. (4.9) over $-1 \leqslant \mu \leqslant 1$, using the fact that the integral of ϕ over μ is unity [see Eqs. (2.12) and (2.13)], gives

$$\int_{-1}^{1} d\mu\, \mu\phi(c,\nu,\mu) = \nu(1-c). \tag{4.17}$$

Hence Eq. (4.16) becomes

$$\int_0^1 d\mu\, \mu \psi(0,\mu) = a(\nu_0)[\nu_0(1-c) + b\omega_0(1-k)] + \int_0^1 d\nu\, a(\nu)[(\nu(1-c) + b\omega(1-k)]. \tag{4.18}$$

If we use N-point Gauss quadrature to perform the integration in Eq. (4.18), we obtain

$$R = 1 - \frac{1}{f_1} \sum_{m=0}^{N} B_m a_m, \tag{4.19}$$

where

$$B_m \equiv w_m [\nu_m(1-c) + b\omega_m(1-k)], \quad m \neq 0, \tag{4.20}$$

$$B_0 \equiv \nu_0(1-c) + b\omega_0(1-k). \tag{4.21}$$

As an application of the foregoing numerical method, we obtained explicit results for $c = 0.96$ and $s = 2$, which yield $k = 0.686954$ from Eq. (2.17). We employed three different incoming angular distributions, namely μ^i, $i = 0,1,2$,

and four different values of b, namely $b = 0, 0.1, 1.0$, and 10. The matrix equation (4.4) was solved numerically for the a_m on a computer, using a standard linear equation solution package. The case $b = 0$ corresponds to $c(z) = \text{const}$, in which instance the eigenfunctions are known to be complete and hence Eq. (3.5) is guaranteed to have a solution. This case serves as a check on the numerical method. In particular, for $b = 0$ and $f(\mu) = 1$, the albedo is known to be $R = 0.631$,[16] which agrees with the converged value shown in Table I. For $b \neq 0$ this table shows similar convergence of the computed albedo as N, the order of the numerical approximation, is increased. The expansion coefficient $a(\nu)$ converges in a similar manner. We note from these results that the computed albedo follows the expected trends with i and b. That is, as the parameter i increases, the incoming angular distribution is more peaked, which implies more absorption in the half-space and hence a smaller albedo. Also, as b increases, the value of $c(z)$, particularly for small z, decreases, which implies a more highly absorbing half-space and thus a smaller albedo. An extreme example which we computed used $b = 10^6$. In this case, $c(z)$ is essentially a constant, equal to $c(0) = k$, for those values of z which experience a significant neutron flux. Accordingly, we should find an albedo corresponding to $c(z) = \text{const} = k$. This value is $R = 0.248$, and this is the result we obtained. Thus we have generated numerically additional evidence that the eigenfunctions are complete on the half-range $0 < \mu \leqslant 1$. That is, we have actually computed, albeit numerically, a solution to the integral equation (3.5) for the expansion coefficient $a(\nu)$.

In closing this section, we mention that the numerical scheme used is very sensitive to numerical roundoff errors. In particular, oscillations in the continuum coefficients a_m became evident for large values of N ($N \gtrsim 16$) for all values of b, including $b = 0$. The onset of these oscillations occurs for lower values of N as either b and/or the power i on the incident flux μ^i increases. This is the expected trend with b and i, since increasing these parameters increases the anisotropy of the angular flux, thus making the continuum contribution more important. This numerical roundoff problem was mitigated by using double precision arithmetic (16 significant digits). We also found it necessary to use direct numerical integration to compute $\phi_n(c,\nu_0)$ and $\phi_n(k,\omega_0)$ rather than the recurrence formula Eq. (4.10). Since $\nu_0, \omega_0 > 1$, Eq. (4.10) is an unstable recurrence relationship for the discrete terms. The fact that using more accurate arithmetic delayed the onset of the oscillations is a clear indication that numerical roundoff is the cause of the instability.

Several other simple numerical schemes were also tried and they, too, were sensitive to roundoff errors. This numerical work indicates that a computer solution of a singular integral equation of the type given by Eq. (3.5), even for $b = 0$, is extremely difficult. A much more clever scheme than outlined here is required for an efficient numerical solution. However as we stated earlier, our purpose was not to develop such a scheme, but merely to demonstrate numerically that a solution for $a(\nu)$ exists.

V. USE OF THE INTEGRAL TRANSPORT EQUATION

Given a $c(z)$, one would like to be able to obtain in a systematic way the corresponding functional form (ansatz) for $\psi(z,\mu)$ which will lead to discrete and/or continuum eigensolutions. We do not know how to do this for a general function $c(z)$. However, we present in this section a method for constructing a consistent pair consisting of (i) $c(z)$, and (ii) a functional form for $\psi(z,\mu)$ which is guaranteed to produce at least one eigenfunction. In the cases we have investigated, this same functional form also leads to other eigenfunctions (continuum and/or discrete).

A formal integration of Eq. (1.1) leads to

$$\psi(z,\mu) = \frac{1}{4\pi\mu}\int_{-\infty}^{z} dz'\, Q(z') e^{-(z-z')/\mu}, \quad \mu > 0, \tag{5.1}$$

$$\psi(z,\mu) = \frac{1}{4\pi|\mu|}\int_{z}^{\infty} dz'\, Q(z') e^{-(z'-z)/|\mu|}, \quad \mu < 0, \tag{5.2}$$

where we have defined the scalar flux as

$$\phi(z) \equiv 2\pi \int_{-1}^{1} d\mu\, \psi(z,\mu), \tag{5.3}$$

and introduced the product

$$Q(z) \equiv c(z)\phi(z). \tag{5.4}$$

Integration of Eqs. (5.1) and (5.2) over μ gives

$$Q(z) = \frac{c(z)}{2}\int_{-\infty}^{\infty} dz'\, E_1(|z-z'|)Q(z'), \tag{5.5}$$

where $E_1(z)$ is the first-order exponential integral. Equation (5.5) is the well-known integral (Peierls) transport equation in planar geometry, written here in terms of $Q(z)$ rather than the more usual $\phi(z)$.

Our procedure to generate a $c(z)$ and a corresponding form for $\psi(z,\mu)$ is as follows. We arbitrarily choose a function $Q(z)$ with certain free parameters, say N in number. Equation (5.5) is used to compute $c(z)$, and Eqs. (5.1) and (5.2) are used to compute the corresponding $\psi(z,\mu)$. This solution for $\psi(z,\mu)$ is generalized to a functional form with the same spatial dependence, but with angular coefficients to be determined. Use of this functional form as the assumed solution for Eq. (1.1) with the corresponding $c(z)$ just determined (which contains N free parameters) is then guaranteed to produce equations for the angular coefficients which will yield at least one (in general, discrete) solution, and hopefully additional solutions (discrete and/or continuum). The reason that a discrete mode (as opposed to a continuum mode) is guaranteed is that for the integral in Eq. (5.5) to converge, $Q(z)$ cannot grow faster than $\exp(|z|)$. Since the constructed $c(z)$ in general will not vanish at $z = \pm\infty$, this means $\phi(z)$ cannot grow faster than $\exp(|z|)$, implying a discrete mode. In the special case that the constructed $c(z)$ does vanish at $z = \pm\infty$, it may be possible for the corresponding solution to represent a continuum mode.

As an example of this procedure, we consider

$$Q(z) = e^{-z/\nu}(1 + \epsilon e^{-z/s}), \quad s > 0, \tag{5.6}$$

where ϵ, ν, and s are free parameters. Use of Eq. (5.6) in Eqs. (5.1), (5.2), and (5.5) yields

$$\psi(z,\mu) = \frac{1}{4\pi}\left[\frac{\nu e^{-z/\nu}}{\nu-\mu} + \epsilon\frac{\omega e^{-z/\omega}}{\omega-\nu}\right], \quad -1 \leqslant \mu \leqslant 1, \tag{5.7}$$

and

$$c(z) = \frac{c + \epsilon c e^{-z/s}}{1 + (\epsilon c/k)e^{-z/s}}. \tag{5.8}$$

Here we have defined

$$1/\omega \equiv (1/\nu) + (1/s), \tag{5.9}$$

and the pairs (c,ν) and (k,ω) are related by the usual dispersion relationship [see Eq. (2.8)]

$$\lambda(c,\nu) = 1 - \frac{c\nu}{2}\ln\left(\frac{\nu+1}{\nu-1}\right) = 0, \tag{5.10}$$

$$\lambda(k,\omega) = 1 - \frac{k\omega}{2}\ln\left(\frac{\omega+1}{\omega-1}\right) = 0. \tag{5.11}$$

Equation (5.7) suggests looking for solutions of Eq. (1.1) of the form

$$\psi_\nu(z,\mu) = \phi_{0\nu}(\mu)e^{-z/\nu} + \phi_{1\nu}(\mu)e^{-z/\omega}. \tag{5.12}$$

Such a form will reproduce Eq. (5.7) and, hopefully, lead to additional eigenfunctions. Equation (5.12) is, in fact, the solution form used in Sec. II, where it was shown that this form also generates a continuum of solutions, as well as a second discrete mode for s sufficiently small. If we replace the free parameter ϵ with the free parameter $b \equiv \epsilon c/k$, then it is clear that the $c(z)$ given by Eq. (5.8) is identical to the $c(z)$ analyzed in Sec. II. The $Q(z)$ given by Eq. (5.6) is easily generalized to the sum of N, rather than two, exponential terms. One finds a correspondingly more complex (but with more free parameters) form for $c(z)$, and the appropriate functional form for $\psi_\nu(z,\mu)$ is the sum of N exponential terms.

As a second example, we set

$$Q(z) = (a + z)e^{-z/\nu}, \tag{5.13}$$

where the constants a and ν are free parameters. The results in this case are

$$\psi(z,\mu) = \frac{\nu e^{-z/\nu}}{4\pi(\nu-\mu)}\left[(a+z) - \frac{\mu\nu}{\nu-\mu}\right], \quad -1 \leqslant \mu \leqslant 1, \tag{5.14}$$

and

$$c(z) = \left(\frac{a+z}{b+z}\right)c. \tag{5.15}$$

Here the constant c is related to ν through the usual dispersion relationship given by Eq. (5.10), and the constant b is given by

$$b = a - \nu_0\left[\frac{1 - \nu_0^2(1-c)}{\nu_0^2 - 1}\right]. \tag{5.16}$$

For a $c(z)$ given by Eqs. (5.15) and (5.16), Eq. (5.14) suggests looking for eigensolutions of Eq. (1.1) of the form

$$\psi_\nu(z,\mu) = \phi_{0\nu}(\mu)e^{-z/\nu} + \phi_{1\nu}(\mu)ze^{-z/\nu}. \tag{5.17}$$

We take up this case in detail in the next section and find, of course, that the discrete mode given by Eq. (5.14) is reproduced. In addition, however, we find continuum modes involving higher order distributions than those previously encountered in transport theory problems.

Just as with our last example, the $Q(z)$ given by Eq. (5.13) can be generalized to an Nth order polynomial with $N+1$ free parameters. One is led to a $c(z)$ which is a const c, determined from Eq. (5.10), multiplying a ratio of two Nth order polynomials. This ratio contains N free parameters. The corresponding functional form which suggests itself is just the extension of Eq. (5.17) to an Nth order polynomial in z. We have not worked out the details for a general N, but are confident that continuum modes involving even higher order distributions will result.

As a final example, we consider a damped periodic function given by

$$Q(z) = e^{-z/\nu}(a + \cos\omega z). \tag{5.18}$$

This leads to a $c(z)$ which is periodic in z and given by

$$c(z) = \left\{\frac{a + \cos\omega z}{a + g(\nu)\cos\omega[z - z_0(\nu)]}\right\}c, \tag{5.19}$$

where ν and c are again related by the usual dispersion relationship Eq. (5.10), and $g(\nu)$ and $z_0(\nu)$ are simple explicit functions of the discrete root ν. The corresponding functional form suggested is

$$\psi_\nu(z,\mu) = e^{-z/\nu}[\phi_{0\nu}(\mu) + \phi_{1\nu}(\mu)\cos\omega z + \phi_{2\nu}(\mu)\sin\omega z]. \tag{5.20}$$

No further details of this case have been worked out.

There appears to be virtually no limit to the functions $Q(z)$ which can be considered to construct a consistent $c(z)$ and a functional form for the corresponding eigenfunctions. Clearly, the more complex $Q(z)$ is, the more complex will be the resulting $c(z)$, the functional form for $\psi(z,\mu)$, and the ensuing analysis to look for continuum modes and other discrete modes. On the other hand, certain functions $Q(z)$ could very well lead to unexpected and interesting solutions of the transport equation.

VI. A "LINEAR" MEDIUM

In this section, we treat Eq. (1.1) with $c(z)$ defined by Eq. (1.5). The transport equation can be written as

$$(b+z)\left(\mu\frac{\partial\psi}{\partial z} + \psi\right) = (a+z)\frac{c}{2}\int_{-1}^{1}\psi\,d\mu'. \tag{6.1}$$

The results derived in Sec. V suggest that we seek solutions having the form

$$\psi_\nu(z,\mu) = f_\nu(\mu)e^{-z/\nu} + g_\nu(\mu)ze^{-z/\nu}. \tag{6.2}$$

Introducing Eq. (6.2) into Eq. (6.1) and equating the coefficients of 1, z, and z^2 gives the following three equations:

$$b\left[\left(1 - \frac{\mu}{\nu}\right)f_\nu(\mu) + \mu g_\nu(\mu)\right] = \frac{ca}{2}\int_{-1}^{1} f_\nu(\mu')\,d\mu', \tag{6.3}$$

$$\left[\left(1 - \frac{\mu}{\nu}\right)f_\nu(\mu) + \mu g_\nu(\mu)\right] + b\left(1 - \frac{\mu}{\nu}\right)g_\nu(\mu) = \frac{c}{2}\int_{-1}^{1} f_\nu(\mu')\,d\mu' + \frac{ca}{2}\int_{-1}^{1} g_\nu(\mu')\,d\mu', \tag{6.4}$$

$$\left(1 - \frac{\mu}{\nu}\right)g_\nu(\mu) = \frac{c}{2}\int_{-1}^{1} g_\nu(\mu')\,d\mu'. \tag{6.5}$$

Straightforward algebraic manipulation of these equations yields

$$\left(1 - \frac{\mu}{\nu}\right)g_\nu(\mu) = \frac{c}{2}, \tag{6.6}$$

$$\left(1-\frac{\mu}{\nu}\right)f_\nu(\mu)=\frac{ca}{2}-\mu g_\nu(\mu), \tag{6.7}$$

$$\frac{1}{b}\int_{-1}^{1} f_\nu(\mu')\,d\mu'=\int_{-1}^{1} g_\nu(\mu')\,d\mu'=1. \tag{6.8}$$

In deriving the normalization condition Eq. (6.8), we have assumed $a\neq b$ so that $c(z)\neq$ constant.

Equation (6.6) with the normalization (6.8) has the usual solutions

$$g_\nu(\mu)=\phi(c,\nu,\mu), \tag{6.9}$$

where we use the definitions given by Eqs. (2.7) and (2.8). At this point, the parameter ν can be any of the continuum ($-1<\nu<1$) or discrete ($\nu=\pm\nu_0$) modes.

To solve Eq. (6.7) with the normalization given by Eq. (6.8), it is appropriate to treat the discrete and continuum values of ν separately. Thus, we first set $\nu=+\nu_0$ in Eq. (6.7) and solve for f_{ν_0} to get

$$f_{\nu_0}(\mu)=a\left(\frac{c\nu_0}{2}\frac{1}{\nu_0-\mu}\right)-\frac{c\nu_0^2}{2}\frac{\mu}{(\nu_0-\mu)^2}. \tag{6.10}$$

Imposing the normalization Eq. (6.8) and using

$$1=1-\lambda(c,\nu_0)=\frac{c\nu_0}{2}\int_{-1}^{1}\frac{d\mu}{\nu_0-\mu},$$

we obtain

$$b=a-\gamma, \tag{6.11}$$

where

$$\gamma=\frac{c\nu_0^2}{2}\int_{-1}^{1}\frac{\mu}{(\nu_0-\mu)^2}\,d\mu=\nu_0\left[\frac{1-\nu_0^2(1-c)}{\nu_0^2-1}\right]. \tag{6.12}$$

(In the integral definition of γ, the integrand for positive values of μ dominate those for negative values, and therefore $\gamma>0$.) We note that Eq. (6.11) is consistent with our previous result, Eq. (5.16). Equation (6.11) is a constraint on the values of a and b. If we now set $\nu=-\nu_0$ in Eqs. (6.7) and (6.8), we obtain the form (6.10) for $f_{-\nu_0}(\mu)$ (with ν_0 replaced by $-\nu_0$), and the condition

$$b=a+\gamma. \tag{6.13}$$

Since $\gamma>0$, Eqs. (6.11) and (6.13) cannot simultaneously be satisfied, and therefore only one discrete mode of the form (6.2) can exist. The description of this mode is given above, and will be summarized below.

Now we shall consider the continuum modes $-1\leqslant\nu\leqslant1$. Equation (6.7) becomes

$$(\nu-\mu)f_\nu(\mu)=\frac{ca\nu}{2}-\mu\nu\lambda(c,\nu)\delta(\nu-\mu)-\mu\frac{c\nu^2}{2}\times P\frac{1}{\nu-\mu}. \tag{6.14}$$

To solve this equation, we shall use the identities

$$x[-\delta'(x)]=\delta(x) \tag{6.15}$$

and

$$x\,P(1/x^2)=P(1/x), \tag{6.16}$$

where $\delta'(x)$ and $P(1/x^2)$ are the distributions defined by Ref. 17

$$\int_{-\infty}^{+\infty}u(x)\delta'(x)dx=-u'(0), \tag{6.17}$$

$$\int_{-\infty}^{+\infty}u(x)\,P\left(\frac{1}{x^2}\right)dx=P\int_{-\infty}^{+\infty}\frac{u(x)-u(0)}{x^2}dx. \tag{6.18}$$

[Equations (6.15) and (6.16) follow from

$$\int_{-\infty}^{+\infty}u(x)x[-\delta'(x)]dx=-\int_{-\infty}^{+\infty}xu(x)\delta'(x)dx=u(0)=\int_{-\infty}^{+\infty}u(x)\delta(x)dx$$

and

$$\int_{-\infty}^{+\infty}u(x)x\,P\left(\frac{1}{x^2}\right)dx=P\int_{-\infty}^{+\infty}\frac{u(x)x-0}{x^2}dx=\int_{-\infty}^{+\infty}u(x)\,P\left(\frac{1}{x}\right)dx,$$

where $u(x)$ is any test function.]

Now we can write the solution of Eq. (6.14) as

$$f_\nu(\mu)=\rho(\nu)\delta(\nu-\mu)+\frac{ca\nu}{2}P\frac{1}{\nu-\mu}+\mu\nu\lambda(c,\nu)\delta'(\nu-\mu)-\mu\frac{c\nu^2}{2}P\frac{1}{(\nu-\mu)^2}, \tag{6.19}$$

where $\rho(\nu)$ is determined by the normalization condition given by Eq. (6.8):

$$b=\rho(\nu)+a[1-\lambda(c,\nu)]+\nu\lambda(c,\nu)\times\int_{-1}^{1}\mu\delta'(\nu-\mu)\,d\mu-\frac{c\nu^2}{2}\int_{-1}^{1}\mu\,P\frac{1}{(\nu-\mu)^2}d\mu. \tag{6.20}$$

Here we have made use of the definition given by Eq. (2.8). Also, the change of variables $\eta=\nu-\mu$ gives

$$\int_{-1}^{1}\mu\delta'(\nu-\mu)\,d\mu=\int_{\nu-1}^{\nu+1}(\nu-\eta)\delta'(\eta)d\eta=1 \tag{6.21}$$

and

$$\frac{c\nu}{2}\int_{-1}^{1}\mu\,P\frac{1}{(\nu-\mu)^2}d\mu=\frac{c\nu}{2}P\int_{-1}^{1}\frac{\mu-\nu}{(\nu-\mu)^2}d\mu=-\frac{c\nu}{2}P\int_{-1}^{1}\frac{1}{(\nu-\mu)}d\mu=\lambda(c,\nu)-1. \tag{6.22}$$

Combining Eqs. (6.20)–(6.22), we obtain

$$\rho(\nu)=b-a[1-\lambda(c,\nu)]-\nu. \tag{6.23}$$

We now summarize our results. For any values of a,b, and c, continuum ($-1\leqslant\nu\leqslant1$) solutions of Eq. (6.1) exist of the form (6.2), where $g_\nu(\mu)$ is the usual homogeneous medium eigenfunction defined by Eq. (6.9), and $f_\nu(\mu)$ is defined by Eqs. (6.19) and (6.23). If a and b are related by Eqs. (6.11) and (6.12), then a discrete solution exists of the form given by Eq. (6.2) with $\nu=\nu_0$, $g_{\nu_0}(\mu)$ defined by Eq. (6.9), and $f_{\nu_0}(\mu)$ defined by Eq. (6.10). If instead, a and b are related by Eqs.

(6.12) and (6.13), then a discrete solution exists of the form Eq. (6.2) with $\nu = -\nu_0$, $g_{-\nu_0}(\mu)$ defined by Eq. (6.9), and $f_{-\nu_0}(\mu)$ defined by Eq. (6.10) where ν_0 is replaced by $-\nu_0$.

The discrete and continuum solutions given by Eq. (6.2) satisfy the full-range orthogonality conditions Eq. (3.19) for $\nu, \nu' > 0$. Therefore, if one can establish half-range completeness (we discuss this below), Siewert's integral equation method[15] (the F_N method) can presumably be used to solve the half-space albedo problem.

The continuum modes constructed above contain higher-order distributions than have occurred previously in transport theory problems. Nevertheless, the theoretical foundations of these distributions have been developed in the literature.[17] Before applying these modes to actual transport problems, a number of technical details will have to be examined. For example, to multiply the continuum modes by an expansion coefficient $a(\nu)$ and integrate over ν, it is probably necessary that $a(\nu)$ have a *derivative* which is Hölder-continuous (or L_p integrable). We shall not examine this detail here.

Also, a question can be asked concerning the "consistency" of the continuum modes. These modes were constructed, in effect, by separating the distribution ψ_ν into a "constant" (f_ν) and a "linear" (g_ν) part with respect to z. Yet, the "constant" part contains the distribution $\delta'(\nu - \mu)$ which acts on the function $e^{-z/\nu}$ and creates a new "linear" term with respect to z. To determine whether this is an actual, or just an apparent, inconsistency, let us consider the general solution of the transport equation which is constructible from the continuum modes

$$\psi(z,\mu) = \int_{-1}^{1} a(\nu)[f_\nu(\mu) + z g_\nu(\mu)] e^{-z/\nu}\, d\nu, \quad (6.24)$$

or

$$\begin{aligned}\psi(z,\mu) = {} & a(\mu)\rho(\mu)e^{-z/\mu} + P\int_{-1}^{1} a(\nu)\frac{c\nu}{2}\frac{1}{\nu-\mu}e^{-z/\nu}\,d\nu \\ & - \mu\frac{\partial}{\partial\mu}[a(\mu)\mu\lambda(c,\mu)e^{-z/\mu}] \\ & - \frac{c\mu}{2}P\int_{-1}^{1}\frac{\nu^2 a(\nu)e^{-z/\nu} - \mu^2 a(\mu)e^{-z/\mu}}{(\nu-\mu)^2}\,d\nu \\ & + z a(\mu)\lambda(c,\mu)e^{-z/\mu} \\ & + z P\int_{-1}^{1} a(\nu)\frac{c\nu}{2}\frac{1}{\nu-\mu}e^{-z/\nu}\,d\nu. \quad (6.25)\end{aligned}$$

Here $a(\nu)$ is taken to be sufficiently smooth that all of the integrals converge pointwise, and in this form there are no expressions more singular than a principal value integral. We have introduced this form into the transport equation (6.1) and have verified, by direct but tedious calculations, that it, indeed, is a solution. Therefore the continuum modes, derived formally above, actually produce solutions of the transport equation when multiplied by $a(\nu)$ and integrated over ν.

Another item which we mention is that although a and b are constrained by Eq. (6.11) or (6.13), one can take the limit $a \to \infty$, and then $c(z) \to c = \text{const}$. We have verified that in this limit the discrete and continuum solutions developed above reduce to the standard homogeneous medium eigenfunctions.

Finally, the equation of half-range completeness is one which we have only briefly considered. If one wishes to consider the half-space $z > z_0$, then one must impose Eq. (6.11), adjoin to Eq. (6.25) the discrete mode for $\nu = \nu_0$, and set $a(\nu) = 0$ for $\nu < 0$. Then one obtains the "general" solution of the transport equation (6.1) which decays as $z \to +\infty$. Then, from Eq. (6.25), it is clear that applying the boundary condition (i.e., prescribing the incident flux) will lead to a singular integro-differential equation for the half-range coefficients $a(\nu)$. This equation is more complex than the usual singular integral equations which occur for homogeneous media, or the equation encountered in Secs. II and III, and we have not yet examined it in detail.

VII. CONCLUDING REMARKS

We have shown in this paper that eigensolutions of the linear transport equation, both discrete and continuum, can be constructed for spatially varying media, characterized by a nonconstant $c(z)$. The number of functions $c(z)$ for which such eigenfunctions can be found appears to be very large. For the most part, our results indicate that it is relatively easy to construct one discrete mode and continuum modes, and that these modes appear to be complete on the half-range. For the "exponential" medium, a second discrete mode exists for certain restricted values of the parameters in $c(z)$. Whether this second discrete eigenfunction extends the range of completeness of the eigenfunctions has not been investigated.

To our knowledge, no other work has been reported in the literature concerning the explicit construction of elementary solutions of the transport equation for a continuously varying $c(z)$, with the exception of the work of Mullikin and Siewert.[12] These authors found continuum solutions for $c(z) = c_0 \exp(-z/s)$, but did not find discrete solutions. By using the F_N integral equation method of Siewert[15] and the continuum solutions, they numerically computed the albedo for the half-space problem, for $c_0 = 1$ and various values of s. Their results were in excellent agreement with independent numerical calculations utilizing the standard S_N method for (directly) solving the transport equation. However, Mullikin and Siewert also attempted to compute the albedo transmission for finite slab problems, but reported difficulties in obtaining agreement between the F_N and S_N methods. There seems to be two likely causes for this: Either the continuum modes are incomplete, or the F_N method is numerically unstable (for this problem). However, with regard to the half-space albedo problem for which there was numerical agreement, the integral equation for the outgoing flux which was solved by Mullikin and Siewert was shown by Martin[13] to possess a unique solution for c_0 and s satisfying a certain inequality. Thus, it seems likely that, at least for such c_0 and s, the continuum modes are half-range complete. We intend to investigate this completeness question in a future article.

Much work is required to put the analysis for $c(z) \neq \text{const}$ on the same firm foundation as the $c(z) = \text{const}$

case. For any given $c(z)$ and the associated eigenfunctions, half-range and/or full-range completeness must be investigated. It would also be convenient, for the actual solution of problems, to have available half-range and full-range orthogonality relationships to determine the expansion coefficients. A completely open question is: For what class of functions $c(z)$ can one construct a set of eigenfunctions which are complete on either the half-range or the full-range? Generalizations to anisotropic scattering, multigroup problems, multiple-region problems, etc., also seem possible. We feel the ideas described in this paper should open up many fruitful avenues for theoretical research in both linear transport theory and in the theory and numerical solution of singular integral equations.

ACKNOWLEDGMENTS

The authors are indebted to Mr. V. C. Badham for performing the numerical work summarized in Table I, and to Professors T. W. Mullikin and C. E. Siewert for a number of illuminating conversations and helpful suggestions regarding the manuscript. We would also like to thank Professors Mullikin and Siewert for communicating their unpublished work[12] to us. One of the authors (GCP) would like to acknowledge the hospitality of the transport theory group at the Los Alamos Scientific Laboratory. It was during an enjoyable visit to LASL that the work reported here was completed. Earlier efforts by this author were partially supported by the National Science Foundation under grant ENG-77-24992. The work at LASL was performed under the auspices of the United States Department of Energy.

[1]B. Davidson, *Angular Distribution Due to an Isotropic Point Source and Spherically Symmetric Eigensolutions of the Transport Equation*, Report MT-112, National Research Council of Canada (1945).
[2]E. P. Wigner, in *Nuclear Reactor Theory* (edited by G. Birkhoff and E. P. Wigner) Proc. Symp. Appl. Math. **11**, 89 (1961).
[3]N. G. Van Kampen, Physica (Utrecht) **21**, 949 (1955).
[4]N. G. Van Kampen, Proc. K. Ned. Akad. Wet. Se B **63**, 92 (1960).
[5]K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[6]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison–Wesley, Reading, Massachusetts, 1967).
[7]N. J. McCormick and I. Kuscer, Adv. Nucl. Sci. Technol. **7**, 181 (1973).
[8]W. Greenberg and P. F. Zweifel, Trans. Theor. Statis. Phys. **5**, 219 (1976).
[9]M. D. Arthur, W. Greenberg, and P. F. Zweifel, Phys. Fluids **22**, 1465 (1979).
[10]C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966); J. Math. Phys. **7**, 2092 (1966).
[11]C. Cercignani, *Mathematical Methods in Kinetic Theory* (Plenum, New York, 1969).
[12]C. E. Siewert, *Sixth Conference on Transport Theory*, Tucson, Arizona April 2–6, 1979. Also, see T. W. Mullikin and C. E. Siewert, "Radiative Transfer in Inhomogeneous Atmospheres," Ann. Nucl. Energy (in press).
[13]B. J. Martin, SIAM J. Appl. Math. **20**, 703 (1971).
[14]E. W. Larsen, S. Sancaktar, and P. F. Zweifel, J. Math. Phys. **16**, 1117 (1975).
[15]C. E. Siewert and P. Benoist, Nucl. Sci. Eng. **69**, 156 (1979).
[16]G. C. Pomraning, J. Quant. Spectros. Radiat. Transfer **19**, 607 (1978).
[17]I. M. Gel'fand and G. E. Shilov, *Generalized Functions*, (Academic, New York, 1964), Vol. 1.

Solutions of the sine-Gordon equation in higher dimensions

George Leibbrandt
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada

Rudolf Morf[a)]
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

Shein-shion Wang
Institute of Computer Science, University of Guelph, Guelph, Ontario, Canada

(Received 25 September 1978; accepted for publication 8 June 1979)

Exact solutions are derived for the three-dimensional and four-dimensional sine-Gordon equation $[\nabla^2 - \partial^2/(c^2\partial t^2)]\chi = \sin\chi$. The principal tools in the derivation are a new Bäcklund transformation and the appropriate generating formulas which allow us to generate an infinite number of real classical solutions. Three-dimensional computer plots depicting the time evolution are presented (a) for the two-soliton and four-soliton solutions of the sine-Gordon equation, and (b) for the three-wave interaction of the associated sinh-Gordon equation.

I. INTRODUCTION

During the past few years there has been remarkable interest among physicists in the analysis of nonlinear dispersive wave equations. Among the partial differential equations[1] that have gained prominence is the famous sine-Gordon equation in $1+1$ dimensions (one space and one time dimension). This amazing equation plays a central role in such diverse fields as differential geometry,[2] nonlinear optics,[3] plasma physics,[4] superconductivity,[5] and particle physics.[6] The present popularity of the sine-Gordon equation (SGE) among particle physicists is primarily due to its intimate connection with soliton theory[1] and to the equivalence of the *quantized* SGE with the charge-zero sector of the massive Thirring model.[7]

Despite this emphasis on the traditional SGE,[1,8] there has been considerable research not only on related equations such as the double SGE,[9] but also on other procedures of solution such as the base equation technique[10] and the method of prolongation structures.[11]

The purpose of this note is to discuss for the classical SGE

$$(\nabla^2 - \partial_t^2)\chi = \sin\chi, \quad c = 1, \tag{1}$$

the new Bäcklund transformations in three and four dimensions,[12] and to exploit these transformations to generate an infinite class of exact solutions in both $(2+1)$, $(3+0)$ and $(3+1)$ dimensions. We recall that a Bäcklund transformation[2] may be viewed geometrically as a transformation of a solution surface S into a new solution surface S', where S is a solution of a given partial differential equation, but where S' may either be a solution of the original equation or of some other differential equation.

The two principal procedures for solving nonlinear partial differential equations in $1+1$ dimensions are the method of Bäcklund transformations and the inverse scattering technique.[13] Since the latter has, as yet, only been developed in *one* space dimension, this technique is inapplicable to equations such as (1). One is accordingly restricted, in higher dimensions, to the method of Bäcklund transformations which yields not only the crucial generating formulas for multiple solutions, but also provides the key to a systematic analysis of the conserved currents.

2. REVIEW OF SGE IN 2 +0 DIMENSIONS

The Bäcklund transformation for the SGE in two space dimensions and no time dimension $(2+0$ dimensions),

$$(\partial_x^2 + \partial_y^2)\chi = \sin\chi, \quad \partial_x \equiv \partial/\partial x, \quad \partial_y \equiv \partial/\partial y, \tag{2}$$

was shown to be a scalar, albeit complex, equation of the form[14]

$$(\partial_x + i\partial_y)(\alpha - i\beta)/2 = \sin((\alpha + i\beta)/2)\exp i\theta, \tag{3}$$

where α and $i\beta$ are solutions of (2), θ is a real Bäcklund transformation parameter and χ a massless scalar field. Equation (3) represents a transformation from the "old" solution α to the "new" solution $i\beta$, $i\beta = B(\theta)\alpha$, with $B(\theta)$ known as the Bäcklund transformation operator (see the Bianchi diagram in Fig. 1). The simplest nontrivial solutions of (2) read[14]

$$\alpha_1(x,y;\theta) = 4\tan^{-1}(c_1\exp J), \quad J = x\cos\theta + y\sin\theta, \tag{4a}$$

$$\beta_1(x,y;\theta) = \begin{cases} 4\tanh^{-1}(c_2\exp J), & J \leqslant 0, \quad (4b) \\ 4\coth^{-1}(c_2\exp J), & J > 0, \quad (4c) \end{cases}$$

where c_1, c_2 are constants of integration.

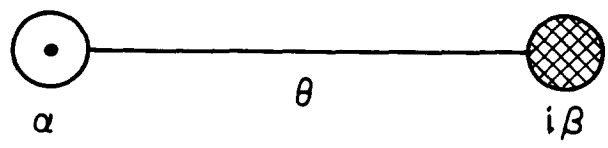


FIG. 1. Bianchi diagram for the Bäcklund transformation in $2+0$ dimensions as given by Eq. (3). The transformation is characterized by the single real parameter θ.

[a)]Present address: RCA-Laboratories, Badenerstrasse, Zurich.

 0022-2488/80/071613-12$1.00

3. SGE IN 3 +1 DIMENSIONS

Rather than discuss next, as would seem logical, the three-dimensional case, we shall immediately tackle the four-dimensional SGE, because its associated Bäcklund transformation and generating formulas reduce quite naturally to the corresponding expressions in three (i.e., 3 +0 or 2 +1) dimensions.

The Bäcklund transformation connected with the SGE

$$(\nabla^2 - \partial_t^2)\chi = \sin\chi, \quad \nabla^2 = \partial_x^2 + \partial_y^2 + \partial_z^2, \quad c = 1, \tag{5}$$

reads

$$(I\partial_x + i\sigma_1\partial_y + i\sigma_3\partial_z + \sigma_2\partial_t)(\alpha - i\beta)/2 = \sin((\alpha + i\beta)/2)\exp\{i\theta\sigma_1\exp[(-i\phi\sigma_2)\exp(-\tau\sigma_1)]\}, \tag{6}$$

where $\sigma_1, \sigma_2, \sigma_3$ are the usual Pauli matrices and I is the 2×2 identity matrix. Observe that the right-hand side of (6) is characterized by three *nested* exponential functions. The Bäcklund parameters (θ,ϕ,τ) are restricted to the domains $0\leqslant\theta\leqslant 2\pi$, $0\leqslant\phi\leqslant 2\pi$, $-\infty < \tau < +\infty$, while the real functions α,β satisfy

$$(\nabla^2 - \partial_t^2)\begin{cases}\alpha(x,y,z,t) = \sin\alpha(x,y,z,t), & (7)\\ \beta(x,y,z,t) = \sinh\beta(x,y,z,t), & (8)\end{cases}$$

respectively. The matrix equation (6) describes, in analogy with (3), a transformation from α to $i\beta$,

$$i\beta = B(\theta,\phi,\tau)\alpha, \tag{9}$$

which may be depicted by the Bianchi diagram drawn in Fig. 2.

To solve (5) we rewrite (6) succinctly as

$$(I\partial_x + iP)(\alpha - i\beta)/2 = \sin((\alpha + i\beta)/2)\exp(i\theta S), \tag{10a}$$

where $P = \sigma_1\partial_y + \sigma_3\partial_z - i\sigma_2\partial_t$ and $\exp(i\theta S) = A_1 + iA_2$. Abbreviating

$$\exp[-i\phi\sigma_2(\exp(-\tau\sigma_1))] = \exp(-i\phi W),$$

where $W = \sigma_2(I\cosh\tau - \sigma_1\sinh\tau)$ with $W^2 = I$, we finally obtain

$$A_2 = \begin{pmatrix} \sin\theta\sin\phi\cosh\tau & (\cos\phi - \sin\phi\sinh\tau)\sin\theta \\ (\cos\phi + \sin\phi\sinh\tau)\sin\theta & -\sin\theta\sin\phi\cosh\tau \end{pmatrix}, \tag{10b}$$

and $A_1 = I\cos\theta$. Separation of (10a) into real and imaginary components yields

$$I\partial_x(\alpha/2) + P(\beta/2) = A_1\sin(\alpha/2)\cosh(\beta/2) - A_2\cos(\alpha/2)\sinh(\beta/2), \tag{11}$$

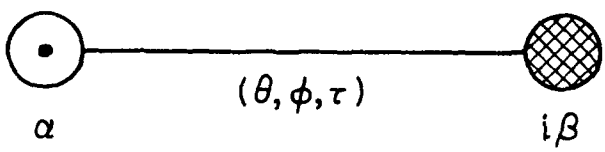


FIG. 2. Bianchi diagram for the Bäcklund transformation in 3 +1 dimensions as given by Eq. (6). The transformation is characterized by the real parameters (θ,ϕ,τ).

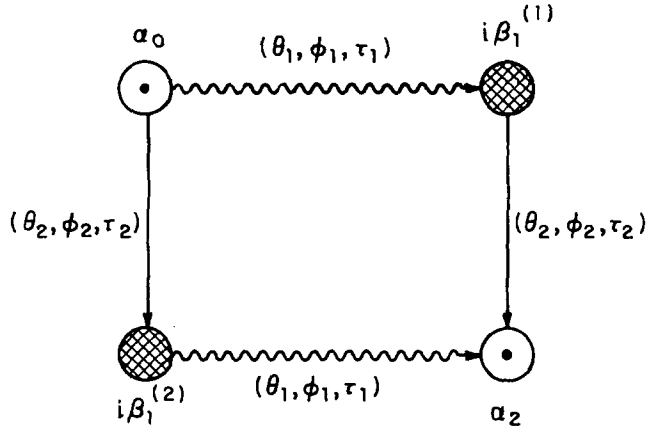


FIG. 3. Bianchi diagram used in the derivation of the generating formula Eq. (14).

$$P(\alpha/2) - I\partial_x(\beta/2) = A_1\cos(\alpha/2)\sinh(\beta/2) + A_2\sin(\alpha/2)\cosh(\beta/2). \tag{12}$$

The lowest, nontrivial solutions for α,β emerge easily from Eqs. (11) and (12) by setting first $\beta \equiv \beta_0 = 0$, then $\alpha \equiv \alpha_0 = 0$. Thus

$$\alpha_1(x,y,z,t;\theta,\phi,\tau) = 4\tan^{-1}(a_0\exp R), \tag{13a}$$

$$\beta_1(x,y,z,t;\theta,\phi,\tau) = \begin{cases} 4\tanh^{-1}(a_1\exp R), & \text{if } R\leqslant 0, \quad (13b)\\ 4\coth^{-1}(a_1\exp R), & \text{if } R > 0, \quad (13c)\end{cases}$$

$$R = x\cos\theta + y\sin\theta\cos\phi + \sin\theta\sin\phi\,(z\cosh\tau + t\sinh\tau), \tag{13d}$$

where a_0, a_1 are integration constants. Since (13a) reduces, for suitable choices of (θ,ϕ,τ), to the *one-soliton* solution in 1 +1 dimensions, we shall also call α_1 a *soliton* solution of Eq. (5); the same terminology will be applied to multiple soliton solutions [cf. Eq. (21)]. The β solutions, by contrast, lack soliton character.

The existence of the Bäcklund transformation (6) enables us to derive the *generating formula*[12] (see Fig. 3 and Ref. 15)

$$\tan((\alpha_2 - \alpha_0)/4) = D_{12}\tanh((\beta_1^{(1)} - \beta_1^{(2)})/4), \tag{14}$$

$$D_{12} = \pm\sqrt{(1 + L_{12})/(1 - L_{12})}, \quad |L_{12}| < 1, \tag{15a}$$

$$L_{12} = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2 \times [\cos\phi_1\cos\phi_2 + \sin\phi_1\sin\phi_2\cosh(\tau_1 - \tau_2)], \tag{15b}$$

where α_0 and α_2 both satisfy (7). Expression (14) was specifically derived from $\alpha_0 = 0$ and its Bäcklund-generated solutions $\beta_1^{(j)} \equiv \beta_1(\theta_j,\phi_j,\tau_j)$, $j = 1,2$. Formula (14) permits us to generate explicitly, and without additional quadratures, an infinite class of exact α_{2n} solutions, provided the appropriate constraint equations involving $L_{2n-1,2n}$ are fulfilled whenever $n\geqslant 2$. We shall discuss these constraint equations in Sec. 4(B) in connection with the four-soliton solution α_4, i.e., when $n = 2$.

4. SGE IN THREE DIMENSIONS

A. 3 +0 dimensions

For the SGE

$$(\partial_x^2 + \partial_y^2 + \partial_z^2)\chi = \sin\chi, \tag{16a}$$

the choice $\tau = 0$ in Eq. (6) yields the Bäcklund transformation

$$(I\partial_x + i\sigma_1\partial_y + i\sigma_3\partial_z)(\alpha - i\beta)/2 = \sin((\alpha + i\beta)/2)\exp[i\theta\sigma_1\exp(-i\phi\sigma_2)], \tag{16b}$$

with $0\leqslant\theta\leqslant 2\pi$, $0\leqslant\phi\leqslant 2\pi$. The α_1 solution reads [cf. Eq. (13a)]

$$\alpha_1(x,y,z;\theta,\phi) = 4\tan^{-1}\{a_0'\exp[x\cos\theta + \sin\theta(y\cos\phi + z\sin\phi)]\}, \tag{16c}$$

while all higher α solutions can be derived from generating formulas similar to Eqs. (14) and (15). For α_2 we have

$$\tan((\alpha_2 - \alpha_0)/4) = \pm\sqrt{(1 + L'_{12})/(1 - L'_{12})}\tanh((\beta_1^{(1)} - \beta_1^{(2)})/4), \tag{17}$$

provided $|L'_{12}| < 1$,

$$L'_{12} = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\phi_1 - \phi_2).$$

B. 2 +1 dimensions

Setting $\tau = 0$ again in Eq. (6) and replacing $\phi\rightarrow i\lambda$, $\sigma_2\rightarrow\sigma_3$, we get the Bäcklund transformation

$$(I\partial_x + i\sigma_1\partial_y + \sigma_2\partial_t)(\alpha - i\beta)/2 = \sin((\alpha + i\beta)/2)\exp[i\theta\sigma_1\exp(\lambda\sigma_3)] \tag{18a}$$

for the (2 +1)-dimensional SGE

$$(\partial_x^2 + \partial_y^2 - \partial_t^2)\chi = \sin\chi, \quad c = 1. \tag{18b}$$

1. The one-soliton solution $\alpha_1(x,y,t)$

Equation (18b) admits solutions of the form

$$\alpha_1(x,y,t;\theta,\lambda) = 4\tan^{-1}\{a_0''\exp[x\cos\theta + \sin\theta(y\cosh\lambda + t\sinh\lambda)]\}, \tag{19}$$

where a_0'' is an integration constant, $0\leqslant\theta\leqslant 2\pi$, and $-\infty < \lambda < +\infty$. The generating formulas for higher solutions α_{2n}, $n = 1,2,3,\cdots$, are similar to Eq. (17) and its generalizations, but with L'_{12} replaced by $l_{12} = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cosh(\lambda_1 - \lambda_2)$.

2. The two-soliton solution $\alpha_2(x,y,t)$

The functional form of α_2 follows from Eqs. (14) and (15), with

$$L_{12}\rightarrow l_{12} = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cosh(\lambda_1 - \lambda_2),$$

$$D_{12} = [(1 + l_{12})/(1 - l_{12})]^{1/2},$$

and

$$\tanh(\beta_1^{(j)}/4) = a_j\exp[x\cos\theta_j + \sin\theta_j(y\cosh\lambda_j + t\sinh\lambda_j)], \quad j = 1,2. \tag{20}$$

There are no constraints on D_{12}, other than $|l_{12}| < 1$, which implies that nonlinear superposition imposes no further restrictions on the manner in which the two soliton waves may be combined.

We have verified numerically on the computer and by explicit calculation that α_2 is indeed a solution of the three-dimensional sine-Gordon equation (18b) and that this solution ranges in value from -2π to $+2\pi$, as expected.

3. The four-soliton solution $\alpha_4(x,y,t)$

From theory[12] and Eqs. (14) and (15) the solution α_4 reads

$$\tan((\alpha_4 - \alpha_2^{(2)})/4) = D_{14}\tanh((\beta_3^{(1)} - \beta_3^{(2)})/4), \tag{21}$$

where

$$\tanh((\beta_3^{(s)} - \beta_1^{(s+1)})/4) = D_{s,s+2}\tan((\alpha_2^{(s+1)} - \alpha_2^{(s)})/4), \quad s = 1,2, \tag{22}$$

and

$$\tan((\alpha_2^{(p)} - \alpha_0)/4) = D_{p,p+1}\tanh((\beta_1^{(p)} - \beta_1^{(p+1)})/4), \quad p = 1,2,3, \tag{23}$$

$$\tanh(\beta_1^{(j)}/4) = a_j\exp[x\cos\theta_j + \sin\theta_j(y\cosh\lambda_j + t\sinh\lambda_j)], \quad j = 1,2,3,4; \tag{24}$$

the coefficients D read

$$D_{rq} = \sqrt{(1 + l_{rq})/(1 - l_{rq})}, \quad |l_{rq}| < 1, \ r\neq q, \ r,q = 1,2,3,4,$$

$$l_{rq} = \cos\theta_r\cos\theta_q + \sin\theta_r\sin\theta_q\cosh(\lambda_r - \lambda_q). \tag{25}$$

In what follows we shall set $\alpha_0 = 0$ and $a_j = +1$, $j = 1,2,3,4$. Let us next discuss the *constraint equations* associated with the solutions (21) and (22) (cf. Gibbon and Zambotti, Ref. 8).

For $\beta_3^{(1)}$ and $\beta_3^{(2)}$, the constraint equations read, respectively,

$$1 + 2l_{12}l_{23}l_{31} = (l_{12})^2 + (l_{23})^2 + (l_{31})^2, \tag{26a}$$

$$1 + 2l_{23}l_{34}l_{42} = (l_{23})^2 + (l_{34})^2 + (l_{42})^2. \tag{26b}$$

The constraints (26) are obtained by substituting $\beta_3^{(s)}$, $s = 1,2$, in Eq. (22) back into Eq. (8). [A similar procedure yields the four constraint equations listed below in (27).] In the case of α_4, Eq. (21), there are *four* coupled constraint equations:

$$1 + 2l_{12}l_{23}l_{31} = (l_{12})^2 + (l_{23})^2 + (l_{31})^2, \tag{27a}$$

$$1 + 2l_{12}l_{24}l_{41} = (l_{12})^2 + (l_{24})^2 + (l_{41})^2, \tag{27b}$$

$$1 + 2l_{23}l_{34}l_{42} = (l_{23})^2 + (l_{34})^2 + (l_{42})^2, \tag{27c}$$

$$1 + 2l_{13}l_{34}l_{41} = (l_{13})^2 + (l_{34})^2 + (l_{41})^2. \tag{27d}$$

System (27) contains *six* distinct l's as functions of the eight Bäcklund parameters θ_i, λ_i, $i = 1,2,3$, and 4. It is possible to solve (27) consistently for the l's by choosing, for example, $l_{12} = 0.8$, and then solving for l_{41} from (27b). Then l_{13}, l_{23}, l_{34} are obtained from the other three equations by iteration. The six equations relating the l's to the θ's and λ's may then be solved by choosing, for example, $\theta_1 = 35°$ and $\lambda_1 = 0.25$;

the remaining θ's and λ's are obtained from the six equations by iteration. Thus we find for the initial choices $l_{12} = 0.6$ and $l_{24} = 0.8$, the following set of l's:

$$l_{12} = 0.6, \quad l_{24} = 0.8, \quad l_{14} = 0, \quad l_{43} = 0.9, \tag{28}$$
$$l_{13} = 0.4358898943540674, \quad l_{23} = 0.9815339366124404,$$

which yield, upon choosing $\theta_1 = 35.0°$, $\lambda_1 = 0.25$ initially, the following self-consistent set of Bäcklund parameters:

$$\theta_1 = 35.0°, \quad \lambda_1 = 0.25,$$
$$\theta_2 = 88.94992304743539°, \quad \lambda_2 = 0.45, \tag{29}$$
$$\theta_3 = 100.093234432231774°, \quad \lambda_3 = 0.478043195386125,$$
$$\theta_4 = 126.260412122729°, \quad \lambda_4 = 0.5572076405471.$$

When verifying α_4 as a solution on the computer, it is important that the θ's and λ's, emerging from system (28), be arranged in *ascending* order as in (29), otherwise α_4 does not satisfy Eq. (7). The coefficients D possess the values:

$$\begin{aligned} D_{12} &= 2.000\,000\,000\,000\,0000, \\ D_{13} &= 1.595\,433\,215\,948\,9636, \\ D_{23} &= 10.358\,898\,943\,540\,6677, \\ D_{24} &= 3.000\,000\,000\,000\,053, \\ D_{34} &= 4.358\,898\,943\,540\,780, \\ D_{14} &= 1.000\,000\,000\,000\,011. \end{aligned} \tag{30}$$

The constraint equations (27) play a dual role. On the one hand, they ensure that β_3 and α_4 turn out to be *exact* solutions of their respective differential equations. On the other hand, they are sufficiently powerful to force the functions β_3 and α_4 to become *coplanar*, as seen from the accompanying diagrams. (See, in this connection, also the discussion in Gibbon–Zambotti and Kobayashi–Izutzu in Ref. 8, as well as in Christiansen's recent preprint.[16]) It is amusing to realize that in the absence of constraints, β_3 and α_4 would turn out to be *nonplanar*, i.e., they would be genuinely three-dimensional functions of x,y,t.

Our equations were solved in quadruple precision on an AMDAHL 470/V5 and the following accuracies were obtained. The contraint equations were solved for the l_{ij}'s to about 25 places accuracy. The θ's and λ's were accurate to 15 places after the decimal. We have also verified numerically that for $x = -5$, $t = 1$ and $y \in [-30, +30]$, the two β's, which are calculated from the θ's and λ's, satisfy Eq. (8) to better than 1 part in 10^{24}. The corresponding α_4 satisfies Eq. (7) to better than 1 part in 10^{15}.

5. THREE-DIMENSIONAL PLOTS OF α_2, α_4, AND β_3 IN 2 +1 DIMENSIONS

Figures 4, 6 and 7 display the structure of the two-soliton and four-soliton solutions of Eq. (18b), while Fig. 5 depicts the behavior of β_3 which solves $(\nabla^2 - \partial_t^2)\beta_3 = \sinh\beta_3$. Since these solutions depend on the three variables x, y and t, we thought it best to plot them against the space variables x, y for different values of the time t. The three-dimensional plots were traced on a CALCOMP 770 plotter using 51 evenly spaced points. We found it convenient, for plotting purposes, to plot the functions $|\sin(\alpha/2)|$ and $|\tanh\beta_3|$ rather than $\sin(\alpha/2)$ and $\tanh\beta_3$, respectively. The relation between (x,y) in the text [e.g., Eq. (20)], and $(X2, X1)$ in the various graphs is $x = X2 - a, y = X1 - a, a > 0$. For example, in Fig. 4(b), $a = 12.5$, so that $X2 \in [0,25]$, $X1 \in [0,25]$. Here is a brief description of the various plots. (Observe that $X\,\mathrm{MAX} \equiv 2a$.)

Fig. 4: Figure 4(b) depicts, for 3 different values of the time $t \equiv T$, the evolution of the function $|\sin(0.5\alpha_2^{(3)}(x,y,T))|$ for $T = -15, 1, +15$ over $X2 \in [0,25]$, $X1 \in [0,25]$. In order to facilitate comparison between $\alpha_2^{(3)}, \beta_3^{(1)}$ and α_4, we have chosen the same θ, λ values as in (29), namely
$\theta_3 = 100.0932344\cdots°$, $\theta_4 = 126.26041212\cdots°$,
$\lambda_3 = 0.47804319\cdots$, $\lambda_4 = 0.55720764\cdots$.

Fig. 5: Fig. 5(a) shows the structure of $|\tanh\beta_3^{(1)}|$ at $T = -20$ for $X2 \in [0,20]$ and $X1 \in [0,20]$. For $\beta_3^{(1)}$ the relevant Bäcklund parameters are $\theta_1, \theta_2, \theta_3$ and $\lambda_1, \lambda_2, \lambda_3$, as given in system (29). Notice that the surface $|\tanh\beta_3^{(1)}|$ is almost flat, i.e., close to unity, near the center $[X2, X1] = [10,10]$. Fig. 5(b) shows the structure of $|\tanh\beta_3^{(1)}|$ for $T = -20, +1, +20$ over the domain $[0,30] \cup [0,30]$.

Fig. 6: This diagram gives a "close-up" of $|\sin(0.5\alpha_4)|$ for $T = +1$ and $X2 \in [0,5]$, $X1 \in [0,5]$. The eight Bäcklund parameters are listed in (29). We also obtained the following asymptotic values for α_4:

$$\alpha_4^{(\min)} = 0.1975,$$
$$\alpha_4^{(\max)} = 25.155,$$

confirming so to speak that α_4 does indeed carry topological charge $Q = +4$.

Fig. 7: This figure shows the time evolution of $|\sin(0.5\alpha_4)|$ as a function of $X1$, $X2$ for $T = -15, -5, 0, +5$, and $+15$.

6. APPLICATION TO SUPERCONDUCTIVITY

The above information suffices to construct time-dependent solutions of Josephson's equation

$$[\partial_x^2 + \partial_y^2 - \partial^2/(c_0^2 \partial t^2)]\psi(x,y,t) = (\lambda_J)^{-2}\sin\psi(x,y,t), \tag{31}$$

which plays an important role in superconductivity,[5,17,18] where it describes the propagation of magnetic flux through a Josephson tunneling junction. A Josephson junction consists basically of two superconducting metals A, B which are separated by a thin nonsuperconducting barrier. In Eq. (31), ψ denotes the phase *difference* between metals A, B which causes a supercurrent to flow across the barrier, λ_J is the Josephson penetration depth and c_0 the speed of electromagnetic waves along the surface of the barrier (in the absence of a Josephson current). Since Eq. (31) has the same structure as (18b), its solutions will possess the same functional form as in (19) and (14). The first solution, for example, is

$$\psi_1(x,y,t;\theta;\lambda) = 4\tan^{-1}Q, \tag{32}$$

with $Q \equiv a_0 \exp\{(x/\lambda_J)\cos\theta + \sin\theta[(y/\lambda_J)\cosh\lambda + (c_0 t/\lambda_J)\sinh\lambda]\}$, which is analogous to the single-fluxon solution in 1 +1 dimensions,[18] it is characterized by the parameters (θ,λ) and the constant of integration a_0. The second solution, analogous to the two-fluxon solution in 1 +1 di-

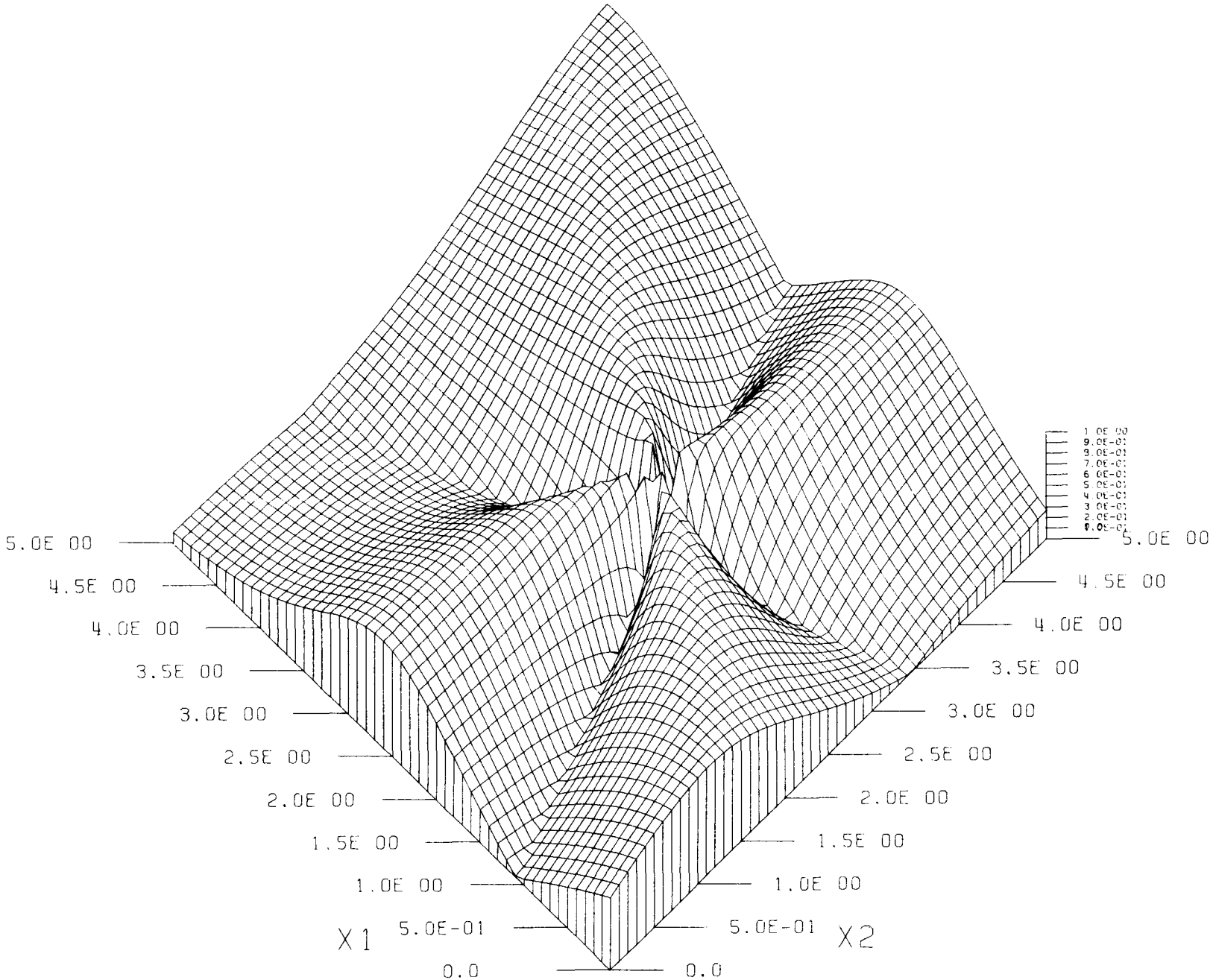


FIG. 4. (a) Behavior of $|\sin(0.5\alpha_2^{(3)})|$ for $T = +1$ and $X\,\mathrm{MAX} = 5$.

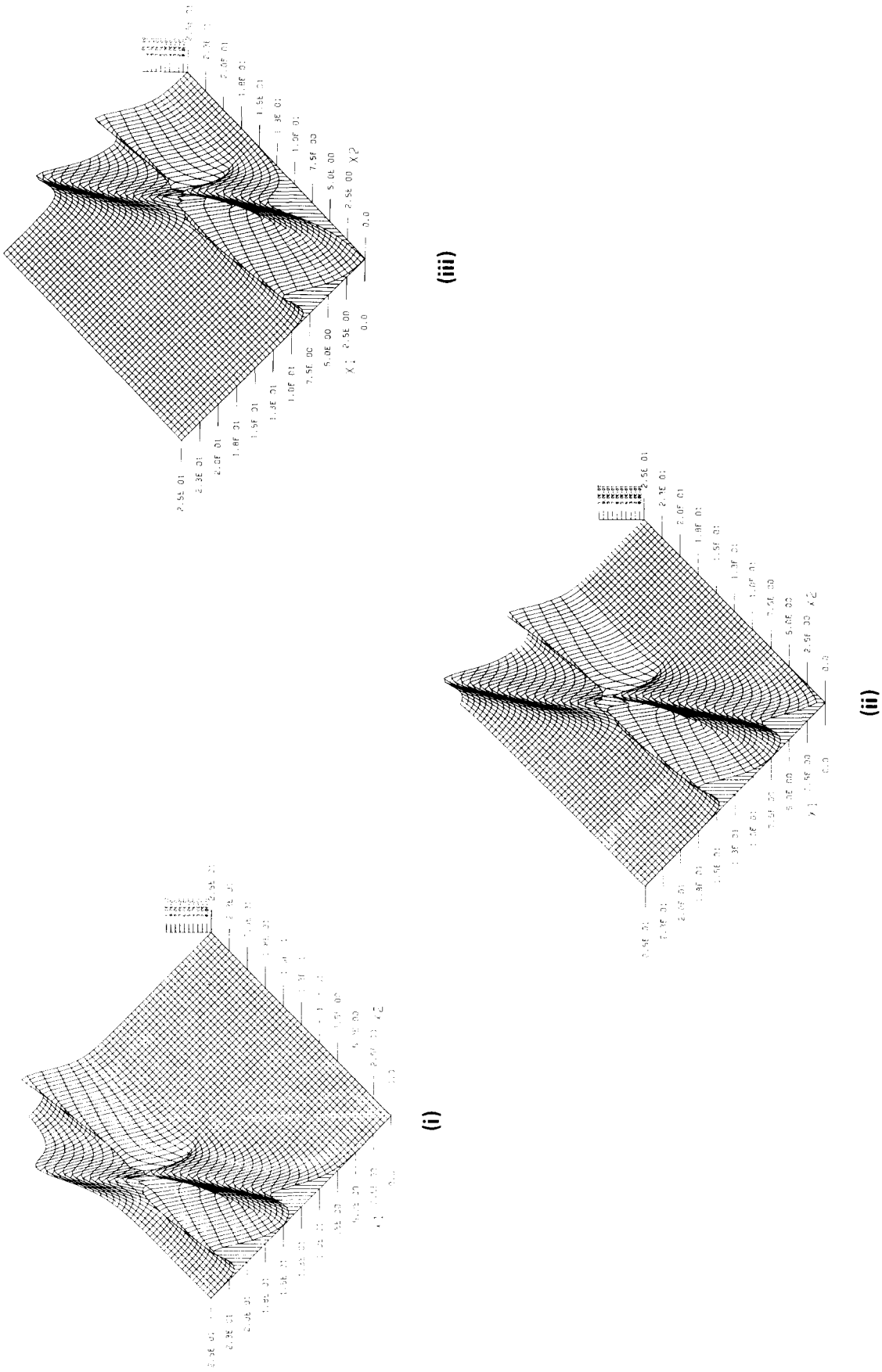


FIG. 4. (b) Time evolution of the function $|\sin(0.5\alpha_2^{(3)})|$ at $XMAX = 25$ for (i) $T = -15$, (ii) $T = 1$, and (iii) $T = +15$.

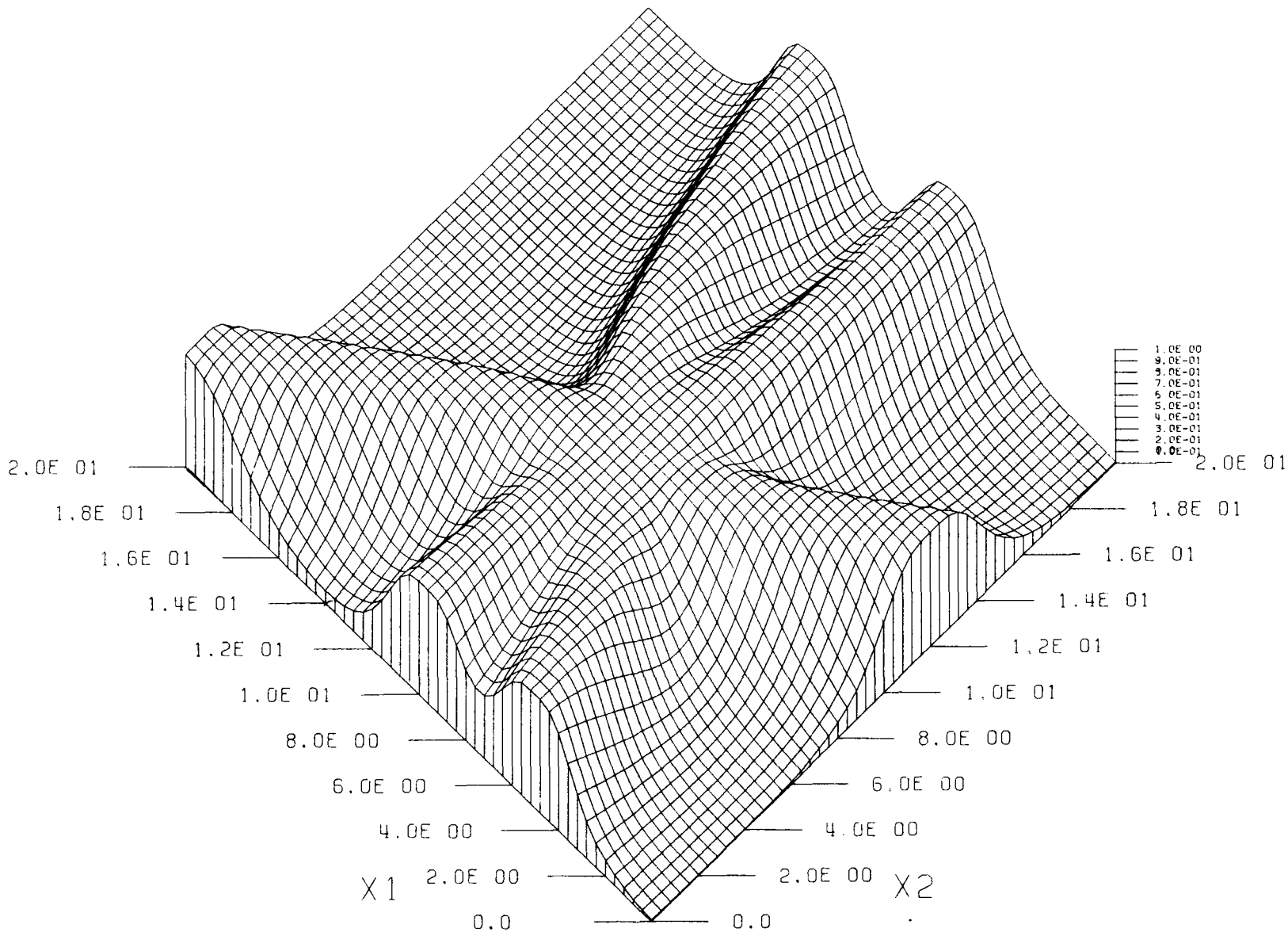


FIG. 5. (a) Three-wave interaction function $|\tanh\beta_3^{(1)}|$ at $X\mathrm{MAX} = 20$, $T = -20$.

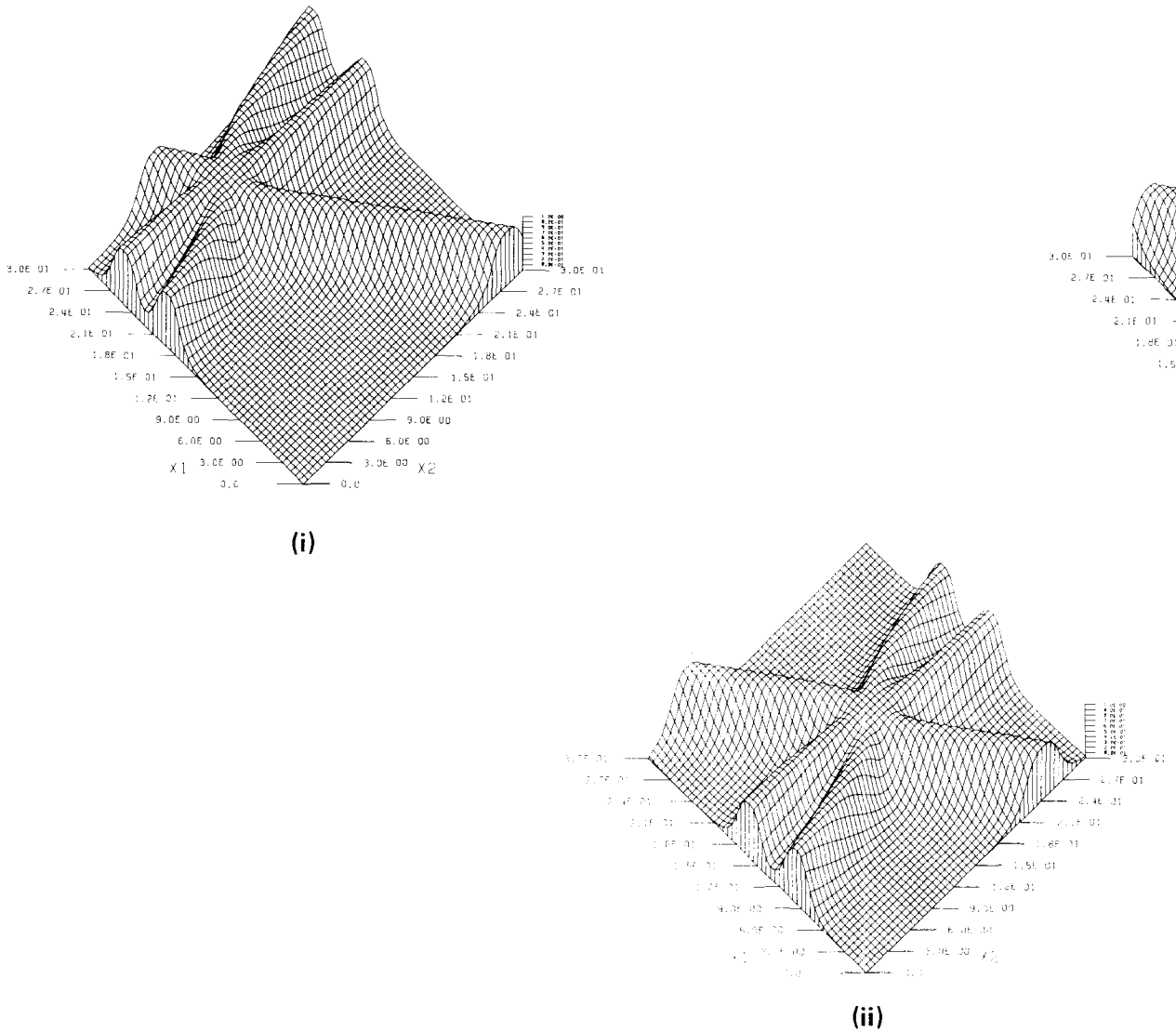


FIG. 5. (b) Three-wave interaction function $|\tanh\beta_{\lambda}^{(1)}|$ at $X\,\mathrm{MAX} = 30$ for (i) $T = -20$, (ii) $T = +1$, and (iii) $T = +20$.

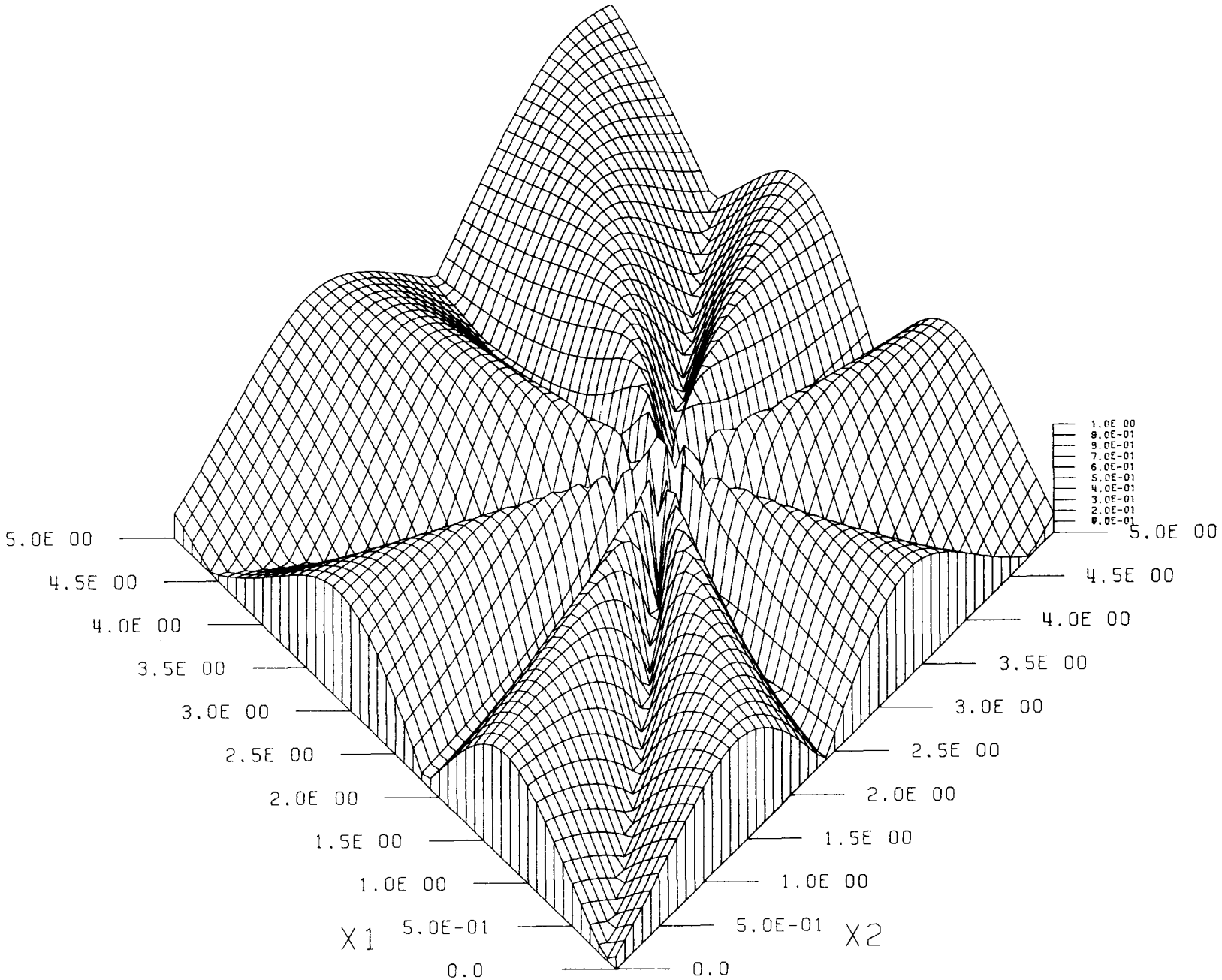


FIG. 6. The function $|\sin(0.5\alpha_4)|$ at $X\,\mathrm{MAX} = 5$, $T = +1$.

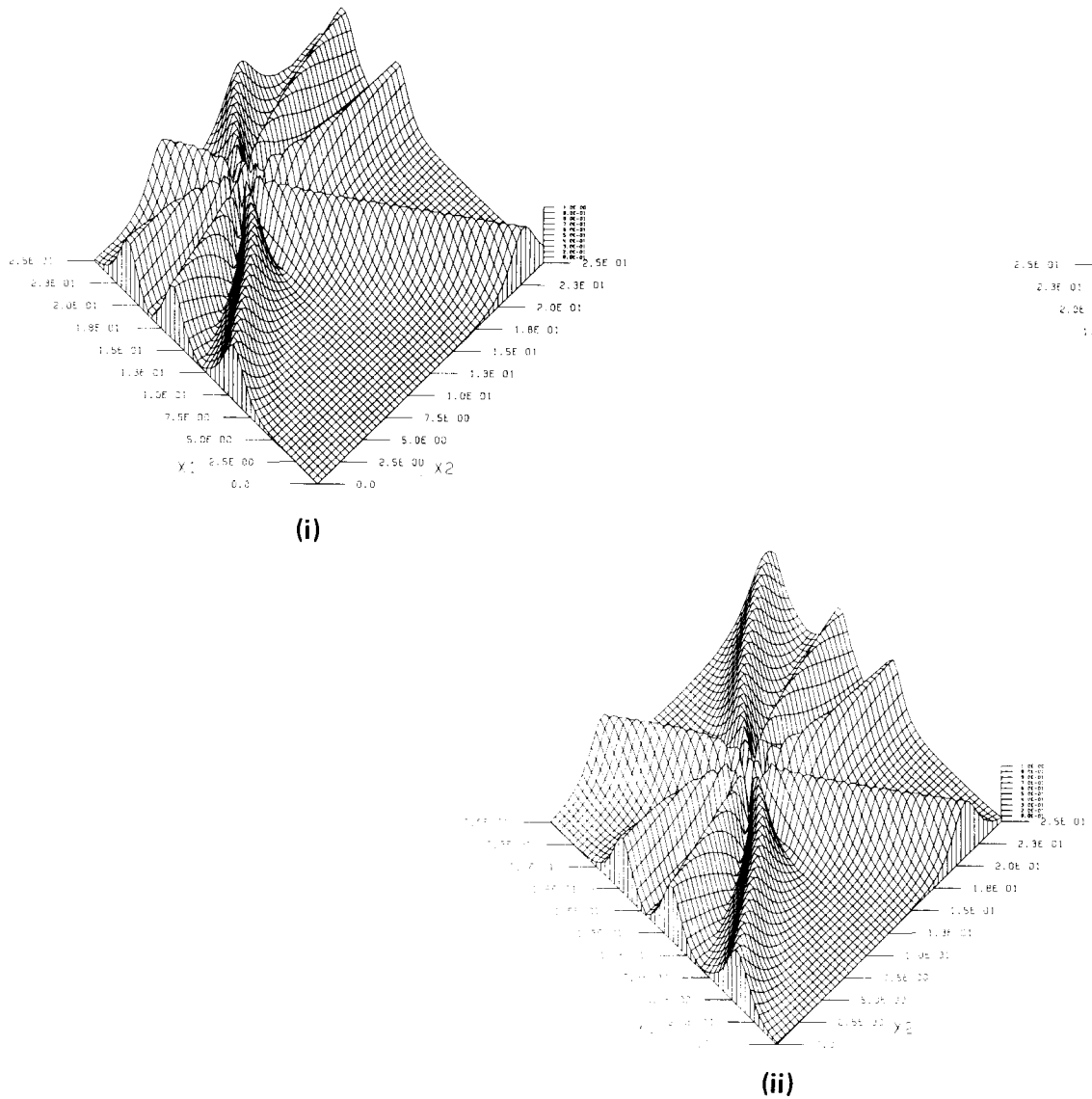


FIG. 7. Time evolution of $|\sin(0.5\alpha_4)|$ at $X\,\mathrm{MAX} = 25$ for (i) -15, (ii) $T = -15$, (iii) $T = 0$, (iv) $T = +5$, and (v) $T = +15$.

(v)
(iv)
(iii)

mensions, reads [cf. Eq. (14)]

$$\psi_2(x,y,t;\theta_1,\theta_2,\lambda_1,\lambda_2) = 4\tan^{-1}[D''_{12}(Q_1 - Q_2)/(1 - Q_1Q_2)], \tag{33a}$$

$$Q_j = a_j \exp\{(x/\lambda_J)\cos\theta_j + \sin\theta_j \times [(y/\lambda_J)\cosh\lambda_j + (c_0 t/\lambda_J)\sinh\lambda_j]\}, \quad j = 1,2, \tag{33b}$$

with D''_{12} given by

$$D''_{12} = \pm\sqrt{(1 + L''_{12})/(1 - L''_{12})}, \quad |L''_{12}| < 1.$$

A more interesting solution of Josephson's equation in 2 +1 dimensions is the four-fluxon solution $\psi_4(x,y,t)$ which can be determined from a system of equations analogous to Eqs. (21)–(27), with $\psi_0 \equiv 0$. We shall not discuss these equations any further.

7. DISCUSSION

We have examined the classical sine-Gordon equation in three and four dimensions. Employing Bäcklund transformations and the appropriate generating formulas, we have indicated how to derive an infinite class of exact solutions α_{2n}, $n = 1,2,3,\cdots$, in both 2 +1, 3 +1 and 3 +0 dimensions. For $n = 1$, no constraint equations are needed to make α_2 exact. For $n = 2$, on the other hand, the four-soliton function α_4 will become an exact solution of $(\nabla^2 - \partial_t^2)\alpha_4 = \sin\alpha_4$ only if we impose four coupled constraint equations [Eqs. (27)]. The effect of these equations is to make α_4 *coplanar*. Several three-dimensional computer plots are presented which depict the time evolution of $\alpha_2(x,y,t)$, $\beta_3(x,y,t)$ and $\alpha_4(x,y,t)$.[19]

Working specifically in 2 +1 dimensions, we find that the Bäcklund transformation (18a), together with the appropriate constraint equations, yields exact multiple soliton solutions such as α_4. The latter function is coplanar and may be interpreted as a (2 +0)-dimensional object in a moving frame. Our analysis seems to suggest, therefore, that the Bäcklund transformation (18a) is not powerful enough to generate multiple soliton solutions α_{2n} ($n = 2,3,4,\cdots$) which are truly three-dimensional. We wish to emphasize that this result differs from the conclusion drawn on p. 438 of Ref. 12.

ACKNOWLEDGMENTS

One of us (G.L.) would like to thank Howard Georgi and Sidney Coleman, as well as the secretarial staff, for their hospitality and assistance during his stay in the Theoretical Physics Group; he would also like to thank Sidney Coleman for his constructive criticism at various stages of this work. This research was supported in part by the National Science Foundation under Grant No. PHY77-22864, and by the National Research Council of Canada, under Grant No. A8063.

[1] A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973), and references therein.
[2] L.P. Eisenhart, *Differential Geometry of Curves and Surfaces* (Dover, New York, 1960).
[3] A.J. DeMaria, D.A. Stetser, and W.H. Glenn, Jr., Science **156**, 1557 (1967); G.L. Lamb, Jr., Phys. Lett. A **25**, 181 (1967); Rev. Mod. Phys. **43**, 99 (1971).
[4] H. Washimi and T. Taniuti, Phys. Rev. Lett. **17**, 996 (1966); F.D. Tappert, Phys. Fluids **15**, 2446 (1972).
[5] B.D. Josephson, Phys. Lett. **1**, 251 (1962); P.W. Anderson and J.M. Rowell, Phys. Rev. Lett. **10**, 230 (1963).
[6] R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **10**, 4114, 4130, 4138 (1974), and D **11**, 3424 (1975); R. Rajaraman, Phys. Rep. C **21**, 227 (1975). See also "Extended Systems in Field Theory," edited by J.L. Gervais and A. Neveu, Phys. Rep. C **23**, 236 (1976).
[7] S. Coleman, Phys. Rev. D **11**, 2088 (1975), and lectures delivered at the 1975 International School of Subnuclear Physics "Ettore Majorana," Harvard Report, Harvard University, Cambridge, Mass., 1975 (to be published).
[8] A Barone, F. Esposito, and C.J. Magee, Riv. Nuovo Cimento **1**, No. 2, 227 (1971); R. Hirota, J. Phys. Soc. Jpn. **35**, 1566 (1973); J.D. Gibbon and G. Zambotti, Nuovo Cimento B **28**, 1 (1975); K.K. Kobayashi and M. Izutzu, J. Phys. Soc. Jpn. **41**, 1091 (1976); P.J. McCarthy, Lett. Math. Phys. **2**, 167 (1977); G. Leibbrandt and G. McKeon, Phys. Lett. A **61**, 83 (1977); and G. Leibbrandt, J. Math. Phys. **19**, 960 (1978).
[9] P.B. Burt, Proc. R. Soc. London Ser. A **359**, 479 (1978); R.K. Dodd and R.K. Bullough, Proc. R. Soc. London. Ser. A **351**, 499 (1976).
[10] E. Pinney, Proc. Am. Math. Soc. **1**, 681 (1950); J.L. Reid and P.B. Burt, J. Math. Anal. Appl. **47**, 520 (1974); M.S. Klamkin and J.L. Reid, SIAM J. Math. Anal. **1**, 305 (1976).
[11] H.D. Wahlquist and F.B. Estabrook, J. Math. Phys. **16**, 1 (1975); **17**, 1293 (1976); R.K. Dodd and J.D. Gibbon, Proc. R. Soc. London Ser. A **359**, 411 (1978).
[12] G. Leibbrandt, Phys. Rev. Lett. **41**, 435 (1978).
[13] M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1462 (1973); **31**, 125 (1973).
[14] G. Leibbrandt, Phys. Rev. B **15**, 3353 (1977); see also Phys. Rev. D **16**, 970 (1977).
[15] See, for example, Fig. 4 in the first paper of Ref. 14.
[16] P.L. Christiansen, "On Bäcklund transformations and solutions to the 2 +1- and 3 +1-dimensional sine-Gordon equation," in Proceedings from the Fourth Scheveningen Conference on Differential Equations, 1979 (to be published).
[17] B.D. Josephson, Adv. Phys. **14**, 419 (1965).
[18] A.C. Scott, Nuovo Cimento B **69**, 241 (1970); D.W. McLaughlin and A.C. Scott, Appl. Phys. Lett. **30**, 545 (1977).
[19] A three-dimensional plotting routine, developed by Peter P. Zachar (McMaster University), was used.

Solution of the Chandrasekhar H-equation by Newton's Method [a)]

C. T. Kelley
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650

(Received 1 February 1980; accepted for publication 20 March 1980)

For any value of the parameter c, $c \neq 1$, for which a solution to the Chandrasekhar H-equation exists, Newton's method may be used to compute the solution by iteration.

I. INTRODUCTION

The Chandrasekhar H-equation arises naturally in many branches of the theory of radiative transfer.[1-3] The equation is

$$H(\mu,c) = 1 + c\,H(\mu c)\int_0^1 \frac{\mu}{\mu+\nu} H(\nu,c)\psi(\nu)d\nu. \quad (1)$$

In Eq. (1), c is a complex parameter, $\psi \in L^1(0,1)$ is given and real valued, and H is the function to be found. In this paper we normalize ψ by requiring that

$$\int_0^1 \psi(\mu)d\mu = \tfrac{1}{2}. \quad (2)$$

Equation (2) will be assumed to hold throughout this paper.

The solution of Eq. (1) by iteration has been the subject of much recent work.[4-8] These papers consider the case $|c| \leqslant 1$. The difficulty encountered is that Eq. (1) has more than one solution for $0 < |c| < 1$, only one of which is of physical interest. Various methods are employed to show that the solution to Eq. (1) obtained by iteration is the one of physical importance.

In Refs. 4 and 5, convergence of the iteration scheme requires one of the following assumptions:

$$\psi \geqslant 0, \quad \int_0^1 \psi(\mu)d\mu = \tfrac{1}{2}, \quad (3)$$

$$\int_0^1 |\psi(\mu)|d\mu \leqslant \tfrac{1}{2}. \quad (4)$$

Reference 8 deals with a somewhat more general situation; in the case considered here, the result of that paper is that Eq. (1) may be solved by iteration for $|c| \leqslant 1$ if Eq. (2) holds and

$$\int_0^1 e^{-x/\nu}\psi(\nu)\frac{d\nu}{\nu} \geqslant 0 \text{ for all } x > 0. \quad (5)$$

In this paper we assume only that Eq. (2) holds.

It can often be shown[3,9] that Eq. (1) has solutions for values of c with $|c| > 1$. The purpose of this paper is to give iterative methods of solution to Eq. (1) that converge for any $c \neq 1$ for which a solution to Eq. (1) exists, and do not require positivity assumptions like (3) or (5).

We work in the Banach space $C[0,1]$ of complex-valued continuous functions on $[0,1]$ endowed with the supremum norm. We define a bounded linear map L and a bounded bilinear form B on $C[0,1]$ by

$$(Lf)(\mu) = \int_0^1 \frac{\mu}{\mu+\nu}\psi(\nu)f(\nu)d\nu, \quad (6)$$

$$B(f,g)\ (\mu) = f(\mu)(Lg)(\mu) + g(\mu)(Lf(\mu)). \quad (7)$$

We let $\mathbb{C}$ denote the complex numbers and define, for each $c \in \mathbb{C}$, a map F on $C[0,1]$ by

$$F(f,c) = f - 1 - \tfrac{1}{2}c\,B(f,f). \quad (8)$$

We may then write Eq. (1) as

$$F(H(c),c) = 0. \quad (9)$$

For fixed c, the Frechét derivative[10] of $F(f,c)$ with respect to f is a bounded linear map on $C[0,1]$. We denote this map by $F'(f,c)$. For f and u in $C[0,1]$ we have

$$F'(f,c)u = u - cB(f,u). \quad (10)$$

Crucial to all our results is the following lemma.

Lemma A. Assume that $\psi \in L^1[0,1]$ is real valued and satisfies Eq. (2). Let $H(c_0)$ be any solution to Eq. (1) with $c = c_0$. Then $F'(H(c_0),c_0)$ is an invertible map on $C[0,1]$ with bounded inverse if $c_0 \neq 1$. If $c_0 = 1$, $F'(H(1),1)$ has a one-dimensional null space spanned by the function $f(\mu) = \mu H(\mu,1)$. The range of $F'(H(1(,1))$ is

$$\left\{F \in C\ [0,1]\ \middle|\ \int_0^1 F(\nu)\psi(\nu)d\nu = 0\right\}. \quad (11)$$

Before describing our iteration schemes we remark on two interesting immediate consequences of the lemma. In Ref. 11, Mullikin showed that if $\psi \geqslant 0$ and satisfies Eq. (2) then bifurcation of Eq. (1) takes place at $c = 1$. More precisely, if $H(\mu,c)$ is the physical solution to Eq. (1) for $|c| \leqslant 1$, then there is $\epsilon > 0$ so that $H(\mu,c)$ is an analytic $C[0,1]$-valued function of $(1-c)^{1/2}$ in the set

$$\{c\ |\ |1-c| \leqslant \epsilon, \quad \arg(1-c) \neq \pi\}. \quad (12)$$

Hence the physical solution becomes double-valued near $c = 1$. Mullikin's proof required only that $F'(H(1),1)$ have one-dimensional null space, that the range be given by Eq. (11), and that $f(\mu) = \mu H(\mu,1)$ not be in range of $F'(H(1),1)$. This last requirement may be written as

$$\int_0^1 \mu H(\mu,1)\psi(\mu)d\mu \neq 0. \quad (13)$$

Equation (13) is not a consequence of Eq. (2), but will follow[8] from Eq. (5). We have the following.

Theorem 1.1 Let $H(\mu,c)$ be the physical solution to Eq. (1) for $|c| < 1$. Assume that $\psi \in L^1[0,1]$ is real valued and that Eq. (2) holds. Then, if Eq. (13) holds, $H(\mu,c)$ is analytical in $(1-c)^{1/2}$ for c in the set given by Eq. (12) for some $\epsilon > 0$.

The next result is an immediate consequence of the

[a)]This work was sponsored by the National Science Foundation under Grant No. MCS-7902659.

 0022-2488/80/071625-04$1.00

lemma and the implicit function theorem.[12]

Theorem 1.2 Let $H_0(\mu)$ be any solution to Eq. (1) for $c = c_0 \neq 1$. Then there is $\epsilon > 0$ and a function $H(\mu,c)$ that is analytic and $C[0,1]$-valued for $c \in \{c_0 - c| < \epsilon\}$ so that

$$F'(H(c),c) = 0$$

and

$$H(\mu,c_0) = H_0(\mu).$$

Hence the set $\{c|$ Eq. (1) has a solution, $c \neq 1\}$ is open.

We now describe our iteration schemes. Convergence of both is a consequence of the lemma and the general theory developed in Ref. 10. However, we give proofs of convergence in Sec. II because in our case the proofs in Ref. 10 may be simplified as we assume existence of a solution and seek only to show that the solution may be computed by iteration.

The first scheme is Newton's method.[10] Let $c \neq 1$ be given. Let H_0 by an initial guess to the solution of Eq. (1). For $n \geqslant 1$, the iterates are given by

$$H_n = H_{n-1} - (F'(H_{n-1},c))^{-1} F(H_{n-1},c). \tag{14}$$

Note that in Eq. (14) one must prove that all of the inverses $(F'(H_{n-1},c))^{-1}$ exist. As we will see this requires that H_0 be sufficiently close to the actual solution. The advantage of Newton's method is rapid convergence. The disadvantage is that $F'(H_{n-1},c)$ must be inverted in order to compute H_n.

Our second scheme is the Chord method. This is called "Modified Newton's method" in Ref. 10. Here, for $n \geqslant 1$,

$$H_n = H_{n-1} - (F'(H_0,c))^{-1} F(H_{n-1},c). \tag{15}$$

Note that for this method only one inverse, $(F'(H_0,c))^{-1}$, must be computed. However, convergence of this scheme is slower that for the Newton's method.

For Newton's method, the result is the following.

Theorem 1.3 Let $c \neq 1$ be such that a solution $H(\mu,c)$ to Eq. (1) exists. Then there is $\epsilon_0 > 0$ so that if $\|H_0 - H(c)\| < \epsilon_0$ the iteration scheme given by Eq. (14) converges and

$$\|H_{n+1} - H(c)\| \leqslant 3|c| \|F'(H(c),c)^{-1}\| \ \|H_n - H(c)\|^2. \tag{16}$$

For the Chord method we have the following.

Theorem 1.4 Let $c \neq 1$ be such that a solution $H(\mu,c)$ to Eq. (1) exists. Then there is $\epsilon_0 > 0$ so that if $\|H_0 - H(c)\| < \epsilon_0$ the iteration scheme given by Eq. (15) converges and

$$\|H_{n+1} - H(c)\| \leqslant 2|c|\epsilon_0 \|F'(H(c),c)^{-1}\| \ \|H_n - H(c)\|. \tag{17}$$

We may rewrite Eq. (1) as

$$G(H,c) = H - (1 - cLH)^{-1} = 0. \tag{18}$$

If H is a solution to Eq. (18) and $u \in C[0,1]$, we have

$$G'(H,c)u = u - cH^2 Lu. \tag{19}$$

A corollary of the proof of Lemma A will be that if $c_0 \neq 1$ and H is a solution to Eq. (18) for $c = c_0$, then $(G'(H,c))^{-1}$ exists as a bounded operator on $C[0,1]$. The proofs for Theorems 1.3 and 1.4 extend directly to this setting and we have the following.

Theorem 1.5 The results of Theorems 1.3 and 1.4 remain valid if F is replaced by G.

The reader should note that these results do not draw a distinction between the solution to Eq. (1) that is of physical significance and the other solutions. Any solution may be computed if the initial guess is sufficiently close to the desired solution. For $c > 1$, if Eq. (13) holds, there are two equally good candidates for the solution of physical interest, and they are complex conjugates of each other.

However, if one is interested in a particular solution, $H(\mu,c)$, to Eq. (1), and one knows $H(\mu,c_0)$ and wishes to compute $H(\mu,c_1)$, then, if $c_0 \neq 1$ and c_1 is sufficiently close to c_0, one may let $H(\mu,c_1)$ be the initial guess and Theorem 1.2 will imply that the solution $H(\mu,c_1)$ obtained by either of the schemes described above will be the solution of interest.

II.

In this section we assume Lemma A and prove Theorems (1.3) and (1.4). Let $c \neq 1$ be such that a solution $H(\mu,c)$ to Eq. (1) exists. Then $(F'(H(c),c))^{-1}$ exists by Lemma A. As $F'(f,c)$ is continuous in both f and c, there is $\epsilon > 0$ so that if $\|H(c) - f\| < \epsilon$ then $(F'(f,c))^{-1}$ exists.

Let $H_0 \in C[0,1]$ be such that

$$\|H(c) - H_0\| < \epsilon_0 < \epsilon. \tag{20}$$

We show that both of the iteration schemes (14) and (15) converge, for ϵ_0 sufficiently small, if H_0 is the initial guess. We consider Newton's method first.

By Eq. (20), $(F'(H_0,c))^{-1}$ exists. We show that if $\|H(c) - H_n\| < \epsilon_0$, and ϵ_0 is sufficiently small, then $\|H_{n+1} - H(c)\| < \epsilon_0$ and Eq. (16) holds. This will prove Theorem 1.3.

Let $H_n = H(c) + q_n$ for $n \geqslant 0$. We have, by Eq. (14), for ϵ_0 sufficiently small,

$$\begin{aligned}
q_{n+1} &= q_n - [F'(H(c) + q_n(c),c)]^{-1} \ [H(c) \\
&\quad + q_n - 1 - c \ H(c) \ LH(c) \\
&\quad - cB(H(c),q_n) - cq_n Lq_n] \\
&= q_n - [F'(H(c),c) - cB(q_n,\cdot)]^{-1} \\
&\quad \times [q_n - cB(H,q_n) - cq_n Lq_n] \\
&= q_n - \{I - c[F'(H(c),c)]^{-1} B \ (q_n,\cdot)\}^{-1} \\
&\quad \times \{q_n - c[F'(H(c),c)]^{-1} q_n Lq_n\} \\
&= c[F'(H(c),c)]^{-1} \ q_n Lq_n - \\
&\quad \sum_{m=1}^{\infty} [cF'(H(c),c)^{-1} B(q_n,\cdot)]^m \\
&\quad \times \{q_n - c[F'(H(c),c)]^{-1} q_n Lq_n\}.
\end{aligned}$$

Hence, as $\|fLf\| \leqslant \|f\|^2$, we have

$$\begin{aligned}
\|q_{n+1}\| &\leqslant |c| \ \|[F'(H(c),c)]^{-1}\| \ \|q_n\|^2 \\
&\quad + |c| \ \|F'(H(c),c)^{-1}\| \ \|B(q_n,\cdot)\| \ \|q_n\| \\
&\quad + O(\|q_n\|^3).
\end{aligned}$$

Hence, if $\|q_n\| < \epsilon_0$ is sufficiently small, we have

$$\|q_{n+1}\| \leqslant 3|c| \ \|[F'(H(c),c)]^{-1}\| \ \|q_n\|^2 < \epsilon_0,$$

and this is Eq. (16).

The proof of Theorem 1.4 is similar. If we let H_0 satisfy Eq. (20) and let $H_n = H(c) + p_n$ be given by Eq. (15), we have

$$
\begin{aligned}
p_{n+1} &= p_n - [F'(H_0,c)]^{-1}\ [p_n - cB(H(c),p_n) - cp_nLp_n] \\
&= p_n - [F'(H_0,c)]^{-1}\ [F'(H(c),c)p_n - cp_nLp_n] \\
&= p_n - [F'(H(c),c) - cB(p_0,\cdot)]^{-1} \\
&\quad \times [F'(H(c),c)p_n - cp_nLp_n] \\
&= p_n - \{I - c[F'(H(c),c)]^{-1}B(p_0,\cdot)\}^{-1} \\
&\quad \times \{p_n - c[F'(H(c),c)]^{-1}p_nLp_n\}.
\end{aligned}
$$

Hence, as $\|B(p_0,\cdot)\| \leqslant \|p_0\|$,

$$
\|p_{n+1}\| \leqslant |c|\ \|p_0\|\ \|p_n\|\ \|[F'(H(c),c)]^{-1}\| + O(\|p_n\|^2).
$$

Therefore, if ϵ_0 is sufficiently small, we have

$$
\|p_{n+1}\| \leqslant 2|c|\epsilon_0\|F'(H(c),c)^{-1}\|\ \|p_n\|.
$$

This is Eq. (17), and the proof is complete.

III. PROOF OF THE LEMMA

We consider first the case $c \neq 1$. We let H be any solution to Eq. (1) or, equivalently, Eq. (18). By Eq. (18) we have that H is analytic in μ for $\mathrm{Re}\mu > 0$ and

$$
\begin{aligned}
&\lim_{\mu\to\infty} H(\mu,c) \\
&\quad = \left[1 - c\int_0^1 H(\mu,c)\psi(\mu)d\mu\right]^{-1} = (1-c\alpha)^{-1}, \quad (21)
\end{aligned}
$$

where

$$
\alpha = (1/c)[1 \pm (1-c)^{1/2}]. \quad (22)
$$

For our purposes it will not matter which of the $+$ or $-$ signs occurs in Eq. (22).

Let $u \in C[0,1]$ be such that $F'(H(c),c)u = 0$. We have $0 = u - cB(H(c),u) = u - cuLH - cHLu = H^{-1}u - cLu$. Hence,

$$
u = cH^2Lu = G'(H(c),c)u, \quad (23)
$$

where G is given by Eq. (18). Let u be any solution to $F'(H(c),c)u = 0$. We have

$$
\begin{aligned}
&\int_0^1 u(\mu)\psi(\mu)d\mu \\
&\quad = c\int_0^1 \psi(\mu)B(H(c),u)(\mu)d\mu \\
&\quad = c\left[\int_0^1 \psi(\mu)H(\mu)d\mu\right]\int_0^1 u(\mu)\psi(\mu)d\mu. \quad (24)
\end{aligned}
$$

Hence either $\int_0^1 u(\mu)\psi(\mu)d\mu = 0$ or $\int_0^1 \psi(\mu)H(\mu)d\mu = 1/c$. Equation (22) implies that $\int_0^1 \psi(\mu)H(\mu)d\mu \neq 1/c$, as $c \neq 1$, hence

$$
\int_0^1 u(\mu)\psi(\mu)d\mu = 0. \quad (25)
$$

If we let $g(\mu) = H^{-1}(\mu,c)u(\mu)$ we have

$$
g = cHL(Hg) \quad (26)
$$

and

$$
\int_0^1 g(\mu)H(\mu)\psi(\mu)d\mu = 0. \quad (27)
$$

Equations (26) and (27) imply that g also satisfies

$$
g = -cHM(Hg), \quad (28)
$$

where the operator M is defined for $f \in C[0,1]$ by

$$
(Mf)(\mu) = \int_0^1 \frac{\nu}{\mu+\nu}\psi(\mu)f(\nu)d\nu. \quad (29)
$$

We define a subspace X of $C[0,1]$ by

$$
X = \left\{g \in C[0,1] \mid \int_0^1 g(\mu)H(\mu,c)\psi(\mu)d\mu = 0\right\}. \quad (30)
$$

We will be done if we show that Eq. (26) or, equivalently, Eq. (28) has only the trivial solution in X. For $f \in C\{0,1\}$ define f^* by

$$
f^*(\mu) = \mu f(\mu). \quad (31)
$$

Define $\beta(f)$ by

$$
\beta(f) = -c(1-c\alpha)^{-1}\int_0^1 H(\mu)f^*(\mu)\psi(\mu)d\mu. \quad (32)
$$

Note that $\beta(f)$ is well defined as $1 - c\alpha \neq 0$ since $c \neq 1$.

For $f \in C[0,1]$ define Tf by

$$
Tf = (f^* - \beta(f))^*. \quad (33)
$$

We have, as $(Mf)^* = Lf^*$ for all $f \in C[0,1]$,

$$
\begin{aligned}
cHL(HTg) &= cHL(H(g^* - \beta(g))^*) \\
&= cHL(Hg^* - H\beta(g))^* \\
&= (cHM(H(g^* - \beta(g))))^*. \quad (34)
\end{aligned}
$$

Therefore, by Eq. 28,

$$
\begin{aligned}
g^* &= -c(HM(Hg))^* = -cH(M(Hg))^* \\
&= -cHL(Hg^*) = (1-c\alpha)\beta(g)H + cHM(Hg^*). \quad (35)
\end{aligned}
$$

Hence, as $cHMH = c\alpha H - cHLH = c\alpha H + 1 - H = 1 - (1-c\alpha)H$, we have, by Eq. (35),

$$
\begin{aligned}
cHM(H(g^* - \beta(g)) &= g^* - (1-c\alpha)\beta(g)H - \beta(g)cHMH \\
&= g^* - \beta(g). \quad (36)
\end{aligned}
$$

Therefore,

$$
cHL(HTg) = [cHM(Hg^* - \beta(g))^*] = (g^* - \beta(g))^* = Tg. \quad (37)
$$

Hence if $c \neq 1$ and g is a solution to Eq. (26), then $g \in X$, Tg is also a solution to Eq. (26), and $Tg \in X$. Now as L is compact, Eq. (26) has a finite-dimensional solution space. Hence, there are an integer $k \geqslant 1$ and numbers $\{a_i\}_{i=1}^k$ with $a_k \neq 0$ so that

$$
\sum_{i=1}^k a_i T^i g = 0. \quad (38)
$$

Equation (38) and the definition of T imply that there are polynomials p and q so that

$$
g(\mu) = \mu p(\mu^2)/q(\mu^2). \quad (39)
$$

As T is linear, Eq. (38) also implies that

$$
\sum_{i=1}^k a_i T^{i+1}(g) = 0. \quad (40)
$$

Hence $Tg = g$. This implies that

$$
g(\mu) = \frac{\mu\beta(g)}{\mu^2 - 1}, \quad (41)
$$

and hence $g = 0$ as $g \in C[0,1]$.

It remains to consider the case $c = 1$. In this case we have

$$HMH = 1 \tag{42}$$

and

$$\int_0^1 H(\mu,1)\psi(\mu)d\mu = 1.$$

As before, we let g be any solution to Eq. (26). We will be done if we show that g is of form

$$g(\mu) = a\mu, \tag{43}$$

for some constant a. As in Ref. 11, this will imply that the range of $F'(H(1),1)$ is given by Eq. (11).

For $f \in C[0,1]$ we define f_* by

$$f_*(\mu) = (1/\mu)f(\mu). \tag{44}$$

If g is a solution to Eq. (26), then $g_* \in L^2[0,1]$, and we have

$$g_* = (HL(Hg))^* = HM(Hg_*). \tag{45}$$

Hence $g_* \in C[0,1]$. We define $\alpha(g)$ by

$$\alpha(g) = \int_0^1 \psi(\nu)H(\nu,1)g_*(\nu)d\nu. \tag{46}$$

We define $S(g)$ by

$$S(g) = (g_* - \alpha(g))_*. \tag{47}$$

To see that $S(g) \in L^2[0,1]$ note that $g_* - \alpha(g) \in X$ and hence

$$\begin{aligned} -HL(g_* - \alpha(g)) &= HM(H(g_* - \alpha(g)) \\ &= HM(Hg)_* - \alpha(g) = (HLHg)_* - \alpha(g) = g_* - \alpha(g). \end{aligned} \tag{48}$$

Hence $HM(HS(g))$ is defined, we have

$$HM(HS(g)) = (HL(H(g_* - \alpha(g))))_* = -S(g). \tag{49}$$

We have, therefore,

$$\begin{aligned} S(g) = &-HM(HS(g)) = HL(H(S(g))) \\ &- H\alpha(g_* - \alpha(g)). \end{aligned} \tag{50}$$

We multiply both sides of Eq. (50) by ψ and integrate to obtain

$$\alpha(g_* - \alpha(g)) = \int_0^1 S(g)(\mu)(1 - HMH)(\mu)d\mu = 0. \tag{51}$$

Hence $S(g) \in X$ and $S(g)$ is a solution to Eq. (26). The remainder of the proof follows that for the case $c \neq 1$. We find $S(g) = g$ or $S(g) = 0$ and

$$g(\mu) = \frac{1}{\mu}\left(\frac{g(\mu)}{\mu} - \alpha(g)\right) \quad \text{or} \quad S(g) = 0. \tag{52}$$

As $g \in C[0,1]$ we must have $S(g) = 0$ and hence

$$g(\mu) = \mu\alpha(g). \tag{53}$$

This completes the proof.

[1] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).
[2] I. W. Busbridge, *The Mathematics of Readiative Transfer*, Cambridge tracts No. 50 (Cambridge U.P., Cambridge, 1960).
[3] K. M. Case and P. F. Zwiefel, *Linear Transport Theory* (Addison–Wesley, Reading, Masachusetts, 1967).
[4] R. L. Bowden and P. F. Zwiefel, Astrophys. J. **210**, 178–83 (1976).
[5] G. A. Hively, SIAM J. Math. Anal. **9**, 787–92 (1978).
[6] L. B. Rall, Rend. Circ. Math. Palermo **10**, 314 (1961); also, *Computational Solution of Nonlinear Operator Equations* (Wiley, New York, 1969).
[7] R. L. Bowden, J. Math. Phys. **4**, 608 (1979).
[8] C. T. Kelley, SIAM J. Math. Anal. **10**, 844–9 (1979).
[9] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Gröningen, The Netherlands, 1953).
[10] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, New York, 1964).
[11] T. W. Mullikin, J. Appl. Prob. **5**, 357–74 (1968).
[12] M. G. Crandall and P. H. Rabinowitz, J. Funct. Anal. **8**, 321–40 (1971).

Generalized method of a resolvent operator expansion. III

M. Znojil
Nuclear Physics Institute, Czechoslovak Academy of Sciences, 250 68 Řez near Prague, Czechoslovakia

(Received 17 May 1979; accepted for publication 10 August 1979)

We show that our Born-like parametrized expansion of $R(E)=(E-H)^{-1}$ is, after minor modifications, well defined even at the pole of $R(E)$ and provides a new method of solving linear homogeneous equations. Three different possibilities of application are discussed here:
(1) An analytic method of solving the differential eauations. It is based on the partitioning of generalized power series and illustrated by the new solution of the s-wave Schrödinger equation.
(2) A consequent model space reduction of Schrödinger equation. In terms of the matrix moments of H, the effective interaction is defined as an operator continued fraction.
(3) A new form of the perturbation theory which dispenses with the solution of the unperturbed problem.

1. INTRODUCTION

There exists a particularly compact unification of various perturbation theories based on the search for the poles $E=\lambda_0$ of the resolvent operator (Green function) $R(E)=(E-H)^{-1}$. It is described, e.g., by Kumar[1] and consists of the formal factorization of $R(E)$ into the singular and regular parts $W(E)$ and $\mathscr{M}(E)$, respectively, in the vicinity of the pole. Putting

$$W(E)=\text{const}/\det(E-H), \tag{1.1}$$

we may also include some numerical diagonalization methods[2] in this scheme since the projector which defines an eigenstate of full ("perturbed") Hamiltonian H coincides with the regular part $\mathscr{M}(E)$ of $R(E)$ in the limit $E\to\lambda_0$. In the present paper, we also include in the same scheme our method of expansion of $R(E)$ (Refs. 3 and 4 quoted here as I and II, respectively). In this way we obtain a new, very flexible method of solving linear homogeneous equations

$$(E-H)|\psi\rangle=0. \tag{1.2}$$

The corresponding modified form of expansion is described in Sec. 2. In the spirit of the method I, we interpret Eq. (1.1) simply as an additional restriction for the choice of the free parameters. The necessity of such a restriction follows from the fact that after an arbitrary (parametrized) rearrangement of the Born series (I), we cannot arrive at the convergent representation of $R(E)$ near the pole. As a consequence of the restriction, the natural factorization [definition of $W(E)$] immediately follows and uniquely defines $\mathscr{M}$.

An example showing how the method works in its generality is given in Sec. 3. The new analytic solution obtained for the s-wave Schrödinger equation with p exponentials, $p\geqslant 1$, has a form of a generalized Bessel function.

When the results of II are taken into account, the projector operator formalism leads to the more restrictive specification of parameters. In this spirit, we complete the standard model space reduction of the Schrödinger equation by the operator continued fraction definition of the effective interaction. The formalism assumes the knowledge of the model space moments $\langle m\rangle=PH^mP$ of H and is described in Sec. 4.

There is an interesting application of the moment method of Sec. 4 based on the Taylor series expansion of the moments $\langle m\rangle$. This generalizes the standard perturbation techniques and is developed in Sec. 5.

Finally, in Sec. 6 we discuss the numerical aspects and possible applications of the methods as algorithms for the diagonalization of some special matrices.

2. BORN-LIKE EXPANSION FOR BOUND STATES

Let us write the Hamiltonian H in the representation

$$H=\sum_{\lambda}|\lambda><\lambda|+H_c, \tag{2.1}$$

where the $|\lambda\rangle$'s correspond to bound states ($\lambda<0$) and H is the continuous spectrum part of H. In the vicinity of $\lambda=\lambda_0$, the pole term in the resolvent operator $R(E)=(E-H)^{-1}$ dominates so that the regularized operator

$$(E-\lambda_0)R(E)=\sum_{\substack{i=1,\\ \lambda_0=\lambda_{0i}}}^{m(\lambda_0)}|\lambda_{0i}|\rangle\langle\lambda_{0i}||+O(E-\lambda_0) \tag{2.2}$$

approximates the projector on bound states $|\lambda_{0i}\rangle$ corresponding to the energy λ_0. We intend to apply in Eq. (2.2) the rearranged Born expansion I.

We must extend its validity to the singular point λ_0. Introducing the functions (vectors) $|Y_0^i(E,\lambda_0)\rangle$ by the equation

$$(E-H)|Y_0^i(E,\lambda_0)\rangle=(E-\lambda_0)|X_1^i\rangle,\quad i\in\overline{1,M_1}, \tag{2.3}$$

we may write formally (for $E\neq\lambda_0$, cf. I)

$$\begin{aligned}|Y_0^i(E,\lambda_0)\rangle&=(E-\lambda_0)R(E)|X_1^i\rangle\\&=(E-\lambda_0)\sum_{l=1}^{\infty}\sum_{s=1}^{M_l}D_{il}^{is}(E)|X_s^l\rangle,\end{aligned} \tag{2.4}$$

$$D_{il}^{is}(E)=\sum_{\substack{s_j=1\\ j=1,2,\ldots,l-1}}^{M_j}\sum_{i_{j+1}=1}^{M_{j+1}}F_1^{is_1}(E)B_1^{s_1i_2}F_2^{i_2s_2}(E)\cdots B_{l-1}^{s_{l-1}i_l}F_l^{i_ls}(E).$$

Here, the $(M_k\times M_k)$-dimensional matrix continued fractions $(MCF)F_k(E)$ are defined by the recurrence

$$F_k(E)=[EI-A_k-B_kF_{k+1}(E)C_{k+1}]^{-1},\quad k=1,2,\cdots, \tag{2.5}$$

with $(M_k\times M_k)$-,$(M_k\times M_{k+1})$-, and $(M_{k+1}\times M_k)$- dimen-

 0022-2488/80/071629-07$1.00

sional matrices of free parameters A_k, B_k, and C_{k+1}, respectively, $M_k \leqslant \infty$. The relation

$$H|X_k^s\rangle = \sum_{r=1}^{M_{k-1}} C_k^{sr}|X_{k-1}^r\rangle + \sum_{r=1}^{M_k} A_k^{sr}|X_k^r\rangle + \sum_{r=1}^{M_{k+1}} B_k^{sr}|X_{k+1}^r\rangle, \quad k = 1,2,\cdots, \tag{2.6}$$

specifies (possibly nonuniquely) the expansion set $\{|X_k^r\rangle\}$ in terms of H and free parameters.

Now, we consider the limiting transition $E \to \lambda_0$. Because the vector

$$|Y_0^i(E,\lambda_0)\rangle \to \sum_{\substack{j=1 \\ \lambda_0 = \lambda_{0j}}}^{m(\lambda_0)} |\lambda_{0j}\rangle c_j^i, \quad c_j^i = \langle \lambda_{0j}|X_1^i\rangle \tag{2.7}$$

becomes equal to the solution of the homogeneous equation

$$(E-H)|Y_0^i(E,\lambda_0)\rangle = 0 \tag{2.8}$$

for $E = \lambda_0$, our aim is to give the precise meaning to the $E \to \lambda_0$ limit on the right-hand side of Eq. (2.4). It is clear that the difficulties are similar to those in the Born series[5]: The existence of the auxiliary sequence F_k, $k = 1,2,...,N$, for any cutoff $N \leqslant \infty$ implies that the expansion (2.4) ceases to converge because $(E-\lambda_0)D_{1l}(E) \to 0$ and $|Y_0^i(E,\lambda_0)\rangle - (E-\lambda_0)\Sigma_{l,s} D_{1l}^{is}|X_l^s\rangle \to |Y_0^i(\lambda_0,\lambda_0)\rangle$ for $E \to \lambda_0$.

We shall show first that it is sufficient to demand the existence of F_k's for $k \geqslant 2$ only. Really, putting

$$\begin{aligned} |Y_0^i(E,\lambda_0)\rangle &= \sum_{s_1=1}^{M_1} \mathcal{N}_0^{is_1} |\psi_0^{s_1}(E)\rangle, \\ |\psi_0^i(E)\rangle &= \sum_{l=1}^{N} \sum_{s=1}^{M_l} d_l^{is}(E)|X_l^s\rangle, \\ \mathcal{N}_0^{is_1} &= (E-\lambda_0)F_1^{is}(E), \\ d_l(E) &= B_1 F_2(E)\cdots B_{l-1}F_l(E), \\ l &= 2,...,N = \infty, \quad E \neq \lambda_0, \quad d_1(E) = I, \quad i \in \overline{1,M_1}, \end{aligned} \tag{2.9}$$

and assuming

$$0 = \det[\lambda_0 I - A_1 - B_1 F_2(\lambda_0)C_2] = \det\{[F_1(\lambda_0)]^{-1}\}, \tag{2.10}$$

we see that the limit $(E-\lambda_0)F_1(E) = 0 \times \infty$ is indeterminate and hence $|Y_0^i(\lambda_0,\lambda_0)\rangle$ may become nozero.

The simpler regularization of the series $|Y_0^i(\lambda_0,\lambda_0)\rangle$ may be achieved in analogy with the Lanczos numerical method[6]: We replace $\mathcal{N}_0$ in Eq. (2.9) [and Eq. (2.4)] by the energy-independent normalization matrix $\mathcal{N}$ that satisfies the condition

$$\mathcal{N}\cdot[\lambda_0 I - A_1 - B_1 F_2(\lambda_0)C_2] = 0. \tag{2.11}$$

It follows from Eq. (2.10) that $\mathcal{N} \neq 0$ exists. In fact, the substitution $\mathcal{N}_0 \to \mathcal{N}$ corresponds just to another (energy-dependent) normalization chosen on the right-hand side of Eq. (2.3) from the very beginning.

Of course, the new form

$$(E-H)|Y^i(E,\lambda_0)\rangle = \sum_{r,s=1}^{M_1} \mathcal{N}^{ir}[F_1^{-1}(E)]^{rs}|X_1^s\rangle \tag{2.12}$$

of Eq. (2.3) represents the same homogeneous Eq. (2.8) or (1.2) in the limit $E \to \lambda_0$.

The multiplicity $m'(\lambda_0)$ of the root λ_0 in Eq. (2.10) determines the possible number of independent rows in the matrix $\mathcal{N}$. We may therefore conclude the following:

Proposition: Provided that

(D1) the auxiliary sequence of inverted matrices $[F_k(E)]^{-1}$ is uniquely defined for all $k \geqslant 1$ by the free parameters according to Eq. (2.5),

(D2) the relation (2.6) between H and $|X\rangle$'s holds,

(D3) the series $|\psi_0^i(\lambda_0)\rangle$ defined in Eq. (2.9) converges, and

(D4) the multiplicity $m'(\lambda_0) \leqslant M_1$ of the root in Eq. (2.10) is less than or equal to the multiplicity $m(\lambda_0)$ of the same eigenvalue λ_0 of H, then, with $m'(\lambda_0)$ independent rows of the "normalization" matrix $\mathcal{N}^{is}$, $i \in \overline{1,m'(\lambda_0)}$, $s \in \overline{1,M_1}$, subjected to the finite number of linear algebraic Eqs. (2.11), the vectors

$$|Y^i(\lambda_0,\lambda_0)\rangle = \sum_{j=1}^{M_1} \mathcal{N}^{ij}|\psi_0^j(\lambda_0)\rangle, \quad i \in \overline{1,m'(\lambda_0)} \tag{2.13}$$

represent the $m'(\lambda_0)$ independent solutions of the homogeneous Eq. (1.2) corresponding to the eigenvalue λ_0.

Remarks:

Ad (D1): When the convergence of MCF (2.5) cannot be proved for our choice of free parameters, we may still employ an alternative (polynomial) auxiliary sequence Γ (with possibly different convergence domain) introduced in Appendix II of Ref. 7.

Ad (D2): The main assumption defining our method is just the "clever" fulfillment of the relation (2.6) which represents the generalization of the Lanczos recurrent construction of the basis (cf. Refs. 2, I, II and Secs. 3 and 6).

Ad (D3): Again, the typical proof of convergence is given in Sec. 3. Another (iterative) approach was described in detail in Ref. 7.

Ad (D4): This is the only assumption pertinent to the $E = \lambda_0$ case of our expansion method. It restricts further the freedom in the choice of free parameters (e.g., fixing ξ in Sec. 3). In this respect, it is not independent of (D3) (cf. how the knowledge of some eigenstates of H is used for the acceleration of convergence of the Born series[5]) and influences also (D2) and (D1) (both F_k and $|X\rangle$ depend on the parameters).

3. APPLICATION TO DIFFERENTIAL EQUATIONS: NEW ANALYTIC SOLUTION OF THE SCHRÖDINGER EQUATION

Certain classes of differential equations may be solved by the Taylor series ansatz leading to the solvable two-term recurrence relations for coefficients.[8] The matrix generalization of these relations enables us to cover the more complicated differential equations by the analogous ansatz which may contain more free parameters corresponding to the constants in the differential equation. We present an illustrating example in what follows.

In the Schrödinger equation (1.2) we consider the Hamiltonian acting in the s wave by the superposition of the exponential potentials,

$$H = H(p) = -\frac{d^2}{dr^2} - \sum_{i=1}^{p} g_i e^{-\gamma_i r}, \tag{3.1}$$

where $0<\gamma_1<\gamma_2<\cdots<\gamma_p$, $0\leqslant r, g_i\in\mathbb{C}$. This form of the potential [discrete Laplace transform of an arbitrary function $g(\gamma)$] may be used in many realistic situations. The solution with $p=1$ is known. It is represented by the Bessel function.[5] For $p>1$, the solution was known only in an implicit form.[5,9]

We apply the method of Sec. 2. The relation (2.6) may be fulfilled when we use $|X_k^m\rangle$ in the form of exponentials

$$\langle r|X_k^m\rangle = e^{-\varphi(k,m,\xi)\cdot r}. \tag{3.2}$$

For simplicity we shall number them by the p nonnegative integers $i_1, i_2, \ldots, i_p$ and define

$$\begin{aligned}
&\varphi(k,m,\xi) = \xi + \sum_{n=1}^{p} i_n\gamma_n,\\
&k = 1 + \sum_{n=1}^{p} i_n \leqslant N = \infty,\\
&m = 1 + \binom{i_1}{1} + \binom{i_1+i_2+1}{2} + \cdots\\
&\quad + \binom{k-1-i_p+p-2}{p-1}\\
&\leqslant M_k = M_k(p) = \binom{k+p-2}{p-1},
\end{aligned} \tag{3.3}$$

where ξ is a free (complex) parameter. The one-to-one mapping $\{i_n\}_{n=1}^{p} \leftrightarrow (k,m)$ may be inverted by the prescription

$$i_p = \max j, \quad \binom{k+p-3-j}{p-1} \leqslant m-1,$$

with the $(p-2)$ times repeated formal substitution

$$\begin{aligned}
&m_{\text{new}} = m - \binom{k+p-3-i_p}{p-1} \geqslant 1,\\
&k_{\text{new}} = k - i_p \geqslant 1, \quad p_{\text{new}} = p-1 \geqslant 1.
\end{aligned}$$

Equation (2.6) determines the form of the parametric matrices

$$\begin{aligned}
&C_{k+1} = 0, \quad A_k = A_k^{ml}(p) = -\delta_{ml}\varphi^2(k,m,\xi),\\
&F_k(E) = F_k^{ml}(E,p) = \delta_{ml}[E+\varphi^2(k,m,\xi)],\\
&m,l\in\overline{1,M_k(p)},
\end{aligned} \tag{3.4}$$

$$B_k = B_k(p) = \begin{pmatrix}
-g_pI_1, & B_1(p-1) & 0, \cdots & 0\\
0, & -g_pI_2, & B_2(p-1), & 0,\cdots\ 0\\
\cdots & & & \\
0, & \cdots\ 0, & -g_pI_k, & B_k(p-1)
\end{pmatrix}$$

$p>1$,

$$B_k(1) = -g_1, \quad I_k = I_k^{ml}(p) = \delta_{ml}, \quad m,l\in\overline{1,M_k(p-1)},$$

$k=1,2,\cdots$,

so that the conditions (D1) and (D2) of the Proposition are satisfied. Since $M_1(p)=1$, it is easy to satisfy (D4) when we specify $E=\xi^2$. To verify the Proposition and obtain a formal solution of the Schrödinger equation, it remains to prove the convergence (D3) of the series $|\psi_0\rangle$ for $N=\infty$.

Lemma: The series

$$\begin{aligned}
&\langle r|Y\rangle = \mathcal{N}\sum_{l=1}^{N}\sum_{j=1}^{M_l} d_l^j(\xi,p)e^{-\varphi(l,j,\xi)r},\\
&N=\infty, \quad \mathcal{N} = d_1^1(\xi,p) = 1,\\
&d_l^j(\xi,p) = \sum_{i=1}^{M_{l-1}} d_{l-1}^i(\xi,p)B_{l-1}^{ij}(p)F_l^{ij}(-\xi^2,p),
\end{aligned} \tag{3.5}$$

$j\in\overline{1,M_l}$, $l\in\overline{2,N}$,

i.e., Eq. (2.9) written in terms of functions (3.2) and parameters (3.4) is defined and convergent for arbitrary complex $\xi=(-E)^{1/2}\neq\xi(k,m) = -\Sigma_{l=1}^{p} i_l\gamma_l/2$, $k>1$, and bounded for $\mathrm{Re}\,\xi\geqslant 0$.

Proof: We prove the absolute convergence of $\langle r|Y\rangle$ by majorizing the series (3.5) and avoiding the points $\xi(k,m)$ where $\det F_k^{-1}=0$, $k>1$, and the matrix F_k ceases to exist. For $r\in\langle r,\infty)$, we have $|\langle r|X_k^m\rangle|\leqslant\exp(-r\mathrm{Re}\,\xi)$ so that we may put $r=0$ without loss of generality. For large k, we replace the elements of matrix F_k by the expression $(k\gamma_1)^{-2} + O(k^{-3})$. Thus the action of $B_{k-1}F_k$ on $\langle 0|X_k^m\rangle$ may be represented by identity matrix multiplied by the overall numerical factor $hk^{-2}+O(k^{-3})$, $h=\Sigma_{n=1}^{p}(-g_n)/\gamma_1^2$ since the number of nonzero elements in each row of B_k is equal to p. The majorizing series may therefore be constructed when we substitute $|g_i|$ for $-g_i$ in B_k and h. This is the power series in h with an infinite radius of convergence. QED

Consequence: Provided that the choice of the free parameter $\xi>0$ is given by the transcendental equation

$$0 = 1 + \sum_{l=2}^{N}\sum_{j=1}^{M_l} d_j^l(\xi,p), \quad N=\infty, \tag{3.6}$$

and the normalization coefficient (number) $\mathcal{N}$ is defined by the expression

$$\begin{aligned}
\mathcal{N}^{-2} = &\sum_{k,k'=1}^{N}\sum_{n=1}^{M_k}\sum_{n'=1}^{M_{k'}} d_k^n(\xi,p)d_{k'}^{n'}(\xi,p)\\
&\times[\varphi(k,n,\xi)+\varphi(k',n',\xi)]^{-3},
\end{aligned} \tag{3.7}$$

the series $\langle r|Y\rangle$ represents a bound state corresponding to the s-wave Hamiltonian (3.1) and energy $E=-\xi^2<0$.

Proof: Equation (3.6) represents the standard boundary condition $\langle r|Y\rangle=0$ for $r=0$ and $\xi>0$ guarantees the integrability for $r=\infty$. From the definition $\langle Y|Y\rangle=1$ with the insertion of formula[10] $\langle X_k^m|X_{k'}^{m'}\rangle = [\varphi+\varphi']^{-3}$ we get Eq. (3.7). QED

In the known special case $p=1$, the solution $\langle r|Y\rangle$ with the normalization Eq. (3.7) coincides with the Taylor series for the Bessel function[5,10] $J_{2\xi/\gamma}(2g^{1/2}\exp(-\gamma r/2)/\gamma)$. For $p>1$, the analogous compact form of solution $\langle r|Y\rangle$ seems to be new (cf. Refs. 5, 9).

Equation (3.6) defines the binding energies as zeros of transcendental functions. We have verified numerically the consistency and stability of such a procedure. Putting $\gamma_i=\gamma$ and $g_i=g/p$, $i\in\overline{1,p}$ for $p=1$, $p=2$, and $p=4$, we have reproduced, within the range of the rounding errors, the known eigenvalues of the classical exponential potential. The convergence of partial sums ($N<\infty$ approximations) was also good for different samples of potentials.

With $\mathcal{N}=1$, $\xi=-iE^{1/2}$, $E>0$, the function $\langle r|Y\rangle$

represents the irregular solution of Eq. (1.2) and $\langle 0|Y\rangle$ is the Jost function.[5] Thus, the use of the pure imaginary values of ξ provides the scattering solutions and may be useful in the realistic fit of any s-wave phase shifts by the exponential potentials (couplings g_i and ranges γ_i, $i\in\overline{1,p}$).

4. MCF DEFINITION OF THE EFFECTIVE INTERACTION IN THE SCHRÖDINGER EQUATION

We assume that the exact Schrödinger equation (1.2) may in principle be investigated by means of the limiting transition $N\to\infty$ in its separable approximation

$$I_N H I_N|\psi_N\rangle = E_N|\psi_N\rangle, \tag{4.1}$$

where N denotes the cutoff in the approximate identity projector

$$I_N = \sum_{k=1}^{N} P_k, \quad P_k = \sum_{i,j=1}^{M} |X_k^i\rangle (D_k^{-1})^{ij}\langle X_k^j|,$$
$$\langle X_k^j|X_{k'}^{j'}\rangle = \delta_{kk'} D_k^{ij}, \tag{4.2}$$

and $I_\infty = 1$. Further, we assume that this transition $N\to\infty$ is technically impossible (e.g., due to the computer limitations) and that only the M-dimensional ($M\leqslant 0$) model space $\mathscr{H}_P$ specified by the projector $P = P_1$ is tractable in practice. It is known[11] that the exact solution of linear Eq. (4.1) is still possible within the model space $\mathscr{H}_P$ because the nonlinear equation

$$[E_N - PHP - V_N(E_N)]P|\psi_N\rangle = 0,$$
$$V_N(E) = PHQ_N \frac{1}{E - Q_N H Q_N} Q_N HP, \tag{4.3}$$
$$Q_N = I_N - P = P_2 + P_3 + \cdots + P_N,$$

accompanied by the definition

$$Q_N|\psi_N\rangle = (E_N - Q_N H Q_N)^{-1} Q_N HP\cdot P|\psi_N\rangle,$$

is fully equivalent to Eq. (4.1). Only the definition of the so-called effective Hamiltonian $H_{\text{eff}} = PHP + V_N(E)$ necessitates leaving the model space.

In the spirit of Sec. 2, Eq. (4.1) may be solved by the MCF method. Using the (generalized) Lanczos idea of generating $|X_k\rangle$'s by H, and their elimination for $k = N, N-1$, ..., 2 according to II, we are able to specify the resulting equation [Eq. (2.11)] in terms of the model space quantities only. It is exciting that the change of notation "translates" Eq. (2.11) just into Eq. (4.3) where the effective interaction correction $V_N(E)$ is redefined and given as the model space [$(M\times M)$-dimensional, $M\leqslant\infty$], recursively defined operator. In this way the manageable explicit MCF formulas for the effective interaction are obtained not necessitating to leave the model space.

This idea is not entirely new because it coincides with the classical moment method[12] for $M = 1$. The great advantage of using $M\geqslant 1$ lies in the sufficiency of low N's to determine the corrections to Eq. (4.1). Even the lowest $N = 2$ case of Eq. (4.1) with the effective interaction correction

$$V_2(E) = \beta_2 \frac{1}{E\beta_2 - \alpha_2}\beta_2,$$

$$\beta_2 = PH^2P - PHP\cdot PHP = PHQHP, \quad Q = 1 - P = Q_\infty, \tag{4.4}$$
$$\alpha_2 = PH^3P - PH^2PPHP - PHPPH^2P + PHPPHPPHP = PHQHQHP$$

provides a useful alternative to the standard doubling of dimension of arbitrary basis in Eq. (4.1). Provided that the correction $\epsilon_{21} = E_2 - E_1$ is small, it is also easy to show that the first-order Taylor expansion of the secular determinant leads to the compact approximate prescription

$$E_2 = E_1 - 1/\mathrm{tr}[G(E_1)] + O(\epsilon_{21}^2),$$
$$G(E) = [1 + V_2(E)(1/PHQHP)V_2(E)] \times [E - PHP - V_2(E)]^{-1} \tag{4.5}$$

for an improved energy.

An entirely analogous procedure is to be applied for any higher $N > 2$. The algorithm reads:

(a) Calculate the model space moments $\langle m\rangle = PH^mP$, $m\in\overline{1,2N-1}$ of the Hamiltonian operator H.

(b) Evaluate the auxiliary matrices α_k, β_k from the formulas in II using the auxiliary sequences $\eta_k^{[l,r]}$, $\theta_k^{[l,r]}$.

(c) Find the solution of Eq. (4.3) where the explicit MCF formula for the effective interaction correction (cf. II) reads

$$V_N(E) = \beta_2 w_2(E)\beta_2,$$
$$w_k(E) = [E\beta_k - \alpha_k - \beta_{k+1} w_{k+1}(E)\beta_{k+1}]^{-1}, \tag{4.6}$$

$k\in\overline{2,N}$.

Besides the trivial initialization we may also use an improved one,[13] $w_{N+1}\neq 0$, based on our physical interpretation of P_{N+1}.[14]

It is useful to use some symbolic manipulation language and the computer-produced codes since the explicit formulas become quite involved even at the next $N = 3$ cutoff for $D = I$ where, $g_2 = 1/\beta_2$, $\langle m\rangle = \langle X_1|H^m|X_1\rangle$, and

$$\beta_3 = \langle 4\rangle - \langle 2\rangle\langle 2\rangle - (\langle 3\rangle - \langle 2\rangle\langle 1\rangle)g_2(\langle 3\rangle - \langle 1\rangle\langle 2\rangle),$$
$$\alpha_3 = \langle 5\rangle - \langle 2\rangle\langle 1\rangle\langle 2\rangle - (\langle 3\rangle - \langle 2\rangle\langle 1\rangle)g_2 \times(\langle 4\rangle - \langle 1\rangle\langle 1\rangle\langle 2\rangle) - (\langle 4\rangle - \langle 2\rangle\langle 1\rangle\langle 1\rangle)g_2 \times(\langle 3\rangle - \langle 1\rangle\langle 2\rangle) + (\langle 3\rangle - \langle 2\rangle\langle 1\rangle g_2 \times(\langle 3\rangle - \langle 1\rangle\langle 1\rangle\langle 1\rangle)g_2(\langle 3\rangle - \langle 1\rangle\langle 2\rangle).$$

By this complexity (and by the computer time consumption) we pay for lowered dimension of P with repect to I_N (saving computer storage). The comparable complexity of formulas survives even at $M = 1$, where the diagrammatic methods are used for α_k, β_k and their evaluation,[15] and it increases with N. We may say that without the significant lowering of the cutoff N (and hence the number of moments) by increasing M (dimension $M\times N$ of I_N is fixed), the method would suffer also from cancellations (loss of precision in the computer), and in the classical $M = 1$ case it seems to be quite impracticable [cf. paragraph (g) in Sec. 6].

While the energy E_N is a good approximation to the exact $E = E_\infty$ for sufficiently high M, N, only the M-dimensional projection $P|\psi_N\rangle$ of the full solution $|\psi_N\rangle$ is obtained from Eq. (4.3) for any N. If $Q|\psi_N\rangle$ cannot be neglected, we should not eliminate the rest of the basis. More in the spirit of the moment approach is the repartitioning of the basis such

that $|X_1^i\rangle_{\rm new}$, $i\in\overline{1,M_{\rm new}}$, will include further interesting components from $|X_k^l\rangle_{\rm old}$, $k>1$. Because an increase of dimension M would make the evaluation of the matrix continued fraction $w_2(E)$ too time-consuming, the dimension reduction is to be used. It is based on the investigation of β_k's and resembles (generalizes) one of the practical advantages of the standard Lanczos technique, namely the natural truncation of the basis.[2,16] Really, once $\beta_k=\langle\xi_k|\xi_k\rangle$ (cf. II) is singular ($\det\beta_k=0$), it is obvious that the set of $M=M_{\rm old}$ Lanczos vectors $|\xi_k\rangle$ ($=|X_k\rangle$ without normalization) is linearly dependent. The unitary $M\times M$ transformation U of $|\xi_k\rangle$ is to be found (e.g., by diagonalization of β_k) which annihilates the last vector, $|\xi_k^M\rangle_{\rm new}\equiv 0$. The same U transformation applied to all vectors simultaneously, enables us either to cross out $|X_1^M\rangle_{\rm new}$ and start, from the very beginning, with $M_{\rm new}=M-1$ initial vectors $|X_1^i\rangle_{\rm new}=\Sigma_{j=1}^{M}|X_i^j\rangle_{\rm old}U^{ji}$, $i\in\overline{1,M_{\rm new}}$, or rather lower just the dimension of higher partitions, $M_{k+i}=M-1$, $i\geqslant 0$, simplifying the structure of MCF.

In principle, the elimination of the (hidden) dependence of the initial vectors $|X_1^i\rangle$ may occur repeatedly. In full analogy with the Lanczos $M=1$ case, we may arrive at the finitization of the basis (i.e., $\beta_N\equiv 0$ at some N) provided that all the initial vectors $|X_1^i\rangle$, $i\in\overline{1,M_1}$, $M_1\leqslant\infty$, span only a finite invariant space of H. In practice, even the smallness of $\det\beta_k$ should fully be employed in lowering the dimension M thus providing the most natural truncation procedure at all.

5. GENERALIZED PERTURBATION THEORY

Knowledge of the moments $\langle m\rangle^{ij}=\langle X_1^i|H^m|X_1^j\rangle$, $i,j\in\overline{1,M}$ is sufficient for the method of the preceding paragraph to work. Since the calculation of moments $\langle m\rangle$ may represent the serious difficulty in practice, we shall now require that just the perturbation expansion ($\gamma\ll 1$)

$$\langle m\rangle=\langle m\rangle_{[0]}+\gamma\langle m\rangle_{[1]}+\gamma^2\langle m\rangle_{[2]}+\cdots \tag{5.1}$$

of moments is given. Let us investigate the consequences.

We apply the method of Sec. 4 to exact H denoting the perturbation order in all the formal expansions like Eq. (5.1) by adding the subscript in brackets. According to the general methodology described in II, we have first to construct the auxiliary sequence

$$\eta_k^{[l,r]}=\sum_{n=0}^{\infty}\gamma^n\eta_{k[n]}^{[l,r]},\quad l,r\in\overline{1,k-1},\quad k=1,2,\cdots.$$

We put $\eta_k^{[l,r]}=\langle l+r\rangle_{[n]}$, $n=0,1,\cdots$, and define

$$\begin{aligned}
&\beta_{k[n]}=\eta_{k[n]}^{[k-1,k-1]},\quad n=0,1,\cdots,\\
&g_{k[0]}=1/\beta_{k[0]},\\
&g_{k[n]}=-g_{k[0]}[\beta_{k[1]}g_{k[n-1]}+\cdots+\beta_{k[n]}g_{k[0]}],\quad n=1,2,\cdots,\\
&\eta_{k+1[n]}^{[l,r]}=\eta_{k[n]}^{[l,r]}-\sum_{n_1+n_2+n_3=n}\eta_{k[n_1]}^{[l,k-1]}g_{k[n_2]}\eta_{k[n_3]}^{[k-1,r]},\quad n=0,1,\cdots,\quad k=1,2,\cdots,
\end{aligned} \tag{5.2}$$

in a recurrent way, so that the parameters $\beta_k=\Sigma_{n=0}^{\infty}\gamma^n\beta_{k[n]}$ are generated from Eq. (5.1). In an entirely similar way we define also $\theta_{k[n]}^{[l,r]}$ and $\alpha_k=\Sigma_{n=0}^{\infty}\gamma^n\alpha_{k[n]}$ (cf. II). Then, the perturbation expansion of MCF in Eq. (4.6) leads to the general formula

$$\begin{aligned}
w_k(E)_{[0]}&=[E\beta_{k[0]}-\alpha_{k[0]}-\beta_{k+1[0]}\\
&\quad\times w_{k+1}(E)_{[0]}\beta_{k+1[0]}]^{-1},\\
w_k(E)_{[n]}&=-w_k(E)_{[0]}[\rho_{k[1]}w_k(E)_{[n-1]}\\
&\quad+\cdots+\rho_{k[n]}w_k(E)_{[0]}],\\
\rho_{k[n]}&=E\beta_{k[n]}-\alpha_{k[n]}\\
&\quad-\sum_{n_1+n_2+n_3=n}\beta_{k+1[n_1]}w_{k+1}(E)_{[n_2]}\beta_{k+1[n_3]},\\
&\qquad n=1,2,\cdots,\quad k=N,N-1,\ldots,2,
\end{aligned} \tag{5.3}$$

which defines simultaneously all the perturbation contributions in $w_k(E)=\Sigma_{n=0}^{\infty}\gamma^n w_k(E)_{[n]}$ by the polynomial recurrence. At the finite cutoff N, the trivial initialization $w_{N+1}(E)_{[n]}=0$, $n=0,1,\cdots$, defines, in a recurrent way again, the sequence $w_N(E)$, $w_{N-1}(E),\ldots,w_2(E)$. Finally, the perturbation expansion of the effective interaction correction

$$\begin{aligned}
V_N(E)&=\sum_{n=0}^{\infty}\gamma^n V_N(E)_{[n]},\\
V_N(E)_{[n]}&=\sum_{n_1+n_2+n_3=n}\beta_{2[n_1]}w_2(E)_{[n_2]}\beta_{2[n_3]}
\end{aligned} \tag{5.4}$$

is to be inserted into the model space Schrödinger equation (4.3). This provides the desired solution so that the perturbation theory based on the moment expansion (5.1) is completed.

The connection of our scheme based on Eq. (5.1) with the standard perturbation decomposition is the following: First, the straightforward insertion of $H=H_0+\gamma V$, $\gamma\ll 1$, into $\langle m\rangle$ gives

$$\begin{aligned}
\langle m\rangle_{[0]}&=\langle X_1|H_0^m|X_1\rangle,\\
\langle m\rangle_{[1]}&=\langle X_1|H_0^{m-1}V|X_1\rangle+\langle X_1|H_0^{m-2}VH_0|X_1\rangle\\
&\quad+\cdots+\langle X_1|VH_0^{m-1}|X_1\rangle,\\
&\cdots\\
\langle m\rangle_{[m]}&=\langle X_1|V^m|X_1\rangle,\\
\langle m\rangle_{[m+r]}&=0,\quad r=1,2,\cdots.
\end{aligned} \tag{5.5}$$

Second, the diagonality of H_0 (and $D_1=1$) implies,

$$\langle X_1|H_0^k=\langle X_1|H_0^k|X_1\rangle\langle X_1| \tag{5.6}$$

so that the knowledge of moments of the interaction (perturbation) V is a sufficient input for obtaining the (nonlinear) equation

$$\det[E_n-PHP-W_N(E_N)]=0 \tag{5.7}$$

[i.e., Eq. (2.10)] for energies and Eq. (4.3) for the eigenvector projections. It is also possible (third) to work with the formal expansion $E=\Sigma_{n=0}^{\infty}\gamma^n E_{[n]}$, but this will not be done here. We may conclude that our formulas represent just the modification of the standard Brillouin–Wigner (BW) perturbation theory.[1]

It is seen that the second assumption (diagonality of H_0, i.e., specification of the basis $|X_k^i\rangle$ as eigenvectors of H_0) may be omitted in our approach for such H's and H_0's

that directly permit us to evaluate the matrix elements in Eq. (5.5), e.g., by analytical means when $H_0 =$ kinetic energy $=$ differential operator and $V =$ local potential. In this way, we skip one step in the standard algorithm, namely the diagonalization of H_0.

The MCF character of $V_N(E)$ resembles strongly the Padé approximation techniques[17] (resummation[18]) and occurs here in the most natural way from the very beginning: We may place all the large matrix elements of H (e.g., kinetic operator T and any polynomial approximation to V in the oscillator basis) into H_0 and guarantee the smallness of the perturbation $\gamma V = H - H_0$, permitting H to be block-tridiagonal. Then, in the lowest order, the weakened form of Eq. (5.6), namely

$$PH_0^k = PH_0^k I_{k+1}, \tag{5.8}$$

completes the analogy with the modified BW theory because the formal extension of the model space projector $P \to I_{2N-1} = P_{\text{new}}$ and the knowledge of the first moments $I_m V I_n$, $m + n = 2N$, of perturbation reduces the evaluation of $\langle m \rangle_{[0]}$ and $\langle m \rangle_{[1]}$ in Eq. (5.5) to the finite matrix multiplication again. In higher orders the exact matrix form of $\langle m \rangle_{[k]}$, $k \geqslant 2$, contains the infinite summations, but this is present in the BW case as well.

In the important first order case, the simplified form of Eq. (5.4) (the effective interaction correction) reads

$$\begin{aligned} V_N(E) = {} & \beta_{2[0]} w_2(E)_{[0]} \beta_{2[0]} + \gamma \beta_{2[1]} w_2(E)_{[0]} \beta_{2[0]} \\ & + \gamma \beta_{2[0]} w_2(E)_{[0]} \beta_{2[1]} + \gamma \beta_{2[0]} w_2(E)_{[1]} \beta_{2[0]}, \end{aligned} \tag{5.9}$$

$$\begin{aligned} w_k(E)_{[1]} = {} & -w_k(E)_{[0]} \Big[E\beta_{k[1]} - \alpha_{k[1]} \\ & - \sum_{n_1 + n_2 + n_3 = 1} \beta_{k+1[n_1]} w_{k+1}(E)_{[n_2]} \\ & \times \beta_{k+1[n_3]} \Big] w_k(E)_{[0]}, \quad k = 2,3,\ldots,N. \end{aligned}$$

It should be complemented by the MCF [the first row of Eq. (5.3)] and by the definition of parameters [Eq. (5.2) for $n = 0,1$ plus similar equation for α's].

In practice, the cutoff N is to be fixed by the properties of H_0 but the classical dependence among H_0, E_0 (be eigenvalue of H_0) and $|X_k^m\rangle$ (be eigenstates of H_0) is removed in the spirit of the realistic situation:

(a) The reliable (variational) initial approximation E_0 to E usually exists, shortening the root search in Eq. (5.7).

(b) The good physical models of $|\psi_N\rangle$ simplify the construction of model space and may lower significantly the cutoff N (cf. the truncation procedure mentioned at the end of Sec. 4).

(c) The rest of the unperturbed basis $|X_2\rangle$, $|X_3\rangle$, $\cdots$ is almost arbitrary and the flexibility of partitioning enables us to keep the magnitude of perturbation $H - H_0$ within any predetermined limits. Thus, we may avoid the convergence problems and the necessity of using the higher perturbation orders for the standard two-particle interaction V even if it is not small as a whole. This is very useful since the higher moments V^k, $k \geqslant 2$, represent the more-particle forces.

6. A FEW REMARKS FROM THE NUMERICAL POINT OF VIEW

When we interpret H as the block-tridiagonal or finite matrix, then Sec. 2 or 4, respectively, defines the methods of its numerical diagonalization. Nevertheless, they are not designed to replace the standard computer codes for diagonalization of the matrix form of H. The reason is simple: From the computer time point of view they are less effective than their special case with $M = 1$ suggested by Lanczos.[6] The exceptions are as follows.

(a) We want to reduce the loss of precision ϵ. In the generalized Lanczos algorithm generating the basis $|X_k^i\rangle$ from $|X_1^i\rangle$ by H according to Eq. (2.6) (cf. also II), the value of ϵ is proportional to the number of subtractions[2]; it decreases as $1/M$ (at fixed dimension $n = M \times N$). In some cases, the necessity of reorthogonalizing[2] the basis may thus be avoided.

(b) H is a band matrix [e.g., from (a)]. Then, the parameters A_k, B_k, and C_{k+1}, $k = 1,2,\cdots$, are not free and are fully determined by the matrix elements of H^+ (cf. I, especially Sec. 2). Here, Eq. (2.10) [the condition (D4)] has a standard meaning of the secular equation $\det(E - H) = 0$. It determines not only the eigenvalues λ_0 (cf. Ref. 19) but also MCF $F_k(\lambda_0)$ and hence the eigenvectors $|Y\rangle$ at the same time. Physically, the advantage lies in nonmixing of the different configurations in the basis which remains fixed.

(c) H is a sparse matrix and has the block-tridiagonal form after the basis is appropriately renumbered. This generalizes the band method (b). Some important technical details (regularization of random singularities, etc.) were discussed elsewhere.[7]

(d) H may be approximated by the block-tridiagonal matrix H_0. E.g., the many-body Hamiltonian $H = T + V$ in the oscillator basis may be truncated in the nonstandard way preserving only the T matrix elements outside the finite (M-dimensional, $M < \infty$) model space. Then $H_0 = T + PVP$ is the block-tridiagonal matrix (the fanlike one, with $M_{k+1} > M_k$). MCF formulas represent a direct generalization of the ($M = 1$) continued fraction method of Bassichis and Strayer.[20,13] The perturbation theory of Sec. 5 contains an explicit prescription for including the corrections $V - PVP$ in a systematic way.

(e) The matrix H may be truncated to the block-tridiagonal form H_0. The smallness of the error should be tested iteratively, increasing partition dimensions M_k (i.e., "opening the fan").

(f) The $n \times n$ matrix H is entirely general and it is too large to be stored in the computer. In this case, the choice of the model space may iteratively be improved in the spirit of the Lanczos method and in a way suggested for $M = 1$ by Berger *et al.*[21] They use the lowest cutoff $N = 2$. The $M > 1$ generalization of this is quite straightforward: From the old basis $|X_1^k\rangle$, $k \in \overline{1,M_1}$, $|X_2^k\rangle$, $k \in \overline{1,M_2}$, we define the new vectors $|X_1^k\rangle_{\text{new}} = |Y\rangle = \mathcal{N}|\psi_0\rangle$ according to Eq. (2.9), and arbitrary $|X_2^{k'}\rangle_{\text{new}}$ orthogonal to $|X_1^k\rangle_{\text{new}}$. All the necessary nonzero parameters A_1^{ij}, B_1^{ij}, C_2^{ij}, and A_2^{ij} are defined by the lowest moment formulas (II) adapted for convenience as

$$PHP = A_1 (= \alpha_1),$$
$$PHQHP = B_1 C_2 (= \beta_2), \quad Q = 1 - P, \tag{6.1}$$
$$PHQHQHP = B_1 A_2 C_2 (= \alpha_2).$$

As well as for $M = 1$, this definition of A_1, B_1, C_2, and A_2 is not unique and admits some convenient conventions (cf. $C_2 = B_1^+$ and the choice of positive sign in Ref. 21).

(g) In paragraph (f), entering of QHQ {[$(n - M) \times (n - M)$]-dimensional matrix} into the storage of the computer and the matrix multiplication on the left-hand side of Eq. (6.1) are to be done "row-wise" since we assume $n^2 >$ storage capacity $> n \geqslant N \times M$. Hence, the restriction to $N = 2$ is not substantial. Using the general moment formulas (II) in the form of the algorithm initialized by the three $M \times M$ matrices $g_1 = 0, \beta_1 = 1, \alpha_1 = PHP$, two $(n - M) \times M$ matrices $T_0 = 0$, $W_1 = QHP$, and two $(n - M) \times (n - M)$ matrices $Y_0 = 1$, $U = QHQ$ (to be manipulated row-wise), we define the auxiliary sequences α_k and β_k by the recurrent prescription

$$g_{k-1} = (\beta_{k-1})^{-1},$$
$$Y_{k-1} = Y_{k-2} - T_{k-2} g_{k-1} T_{k-2}^+,$$
$$T_{k-1} = Y_{k-1} W_{k-1}, \quad \beta_k = W_{k-1}^+ T_{k-1}, \tag{6.2}$$
$$W_k = U T_{k-1}, \quad \alpha_k = T_{k-1}^+ W_k, \quad k = 2,3,\ldots,N.$$

The first three items may be skipped for $k = 2$ when we change the initialization to $g_1 = 1, T_1 = W_1, Y_1 = 1$. It may be shown that the algorithm (6.2) is an equivalent reformulation of the method of Sec. 4 when the moments of H are not known and are to be approximately evaluated by the (truncated) matrix multiplication.

7. CONCLUSIONS

The main feature of our method is its structure of the Born series reordered by means of the introduction of free parameters (idea of Haydock[22]). In the present third part of the series (I, II), we may conclude that it treats both the homogeneous and nonhomogeneous equations in the same way.

The flexibility of the method is connected with the *a priori* freedom in the choice of parameters. In this sense, we may interpret and relate the special cases (Taylor series + orthogonalization = Lanczos method[6]) as well as the generalizations ($M = 1$ Lanczos method $\rightarrow M > 1$ Graffi and Grecchi[19] band matrix MCF method). It seems to us that the practical value of such a general formalism is in filling the gaps between the known methods (e.g., between the classical theory of moments[12] and the matrix of Padé approximants[23] to the Born series or the generalized Brillouin–Wigner perturbation theory[24]), but the main merit will be its heuristic power. We believe that it is well illustrated by our generalizations of the Bessel functions ($M = 1$ power series $\rightarrow M > 1$ Dirichlet series[25]), by the model space formulas for the effective interaction corrections, and by a new formulation of the perturbation theory. Because of the methodical character of the paper, we have deferred further applications as well as any detailed calculations to later publications.

[1] K. Kumar, *Perturbation Theory and the Nuclear Many Body Problem* (North-Holland, Amsterdam, 1962).
[2] J.H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).
[3] M. Znojil, J. Math. Phys. **18**, 717 (1977).
[4] M. Znojil, J. Math. Phys. **20**, 2330 (1980).
[5] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[6] C. Lanczos, J. Res. NBS **45**, 255 (1950).
[7] M. Znojil, J. Phys. A: Math. Gen. **11**, 1501 (1978).
[8] E. Kamke, *Differentialgleichungen I* (Leipzig, 1959).
[9] A. Martin, Nuovo Cimento **14**, 403 (1959); L. Trlifaj, J. Math. Phys. **21**, 1724 (1980).
[10] I.S. Gradshteyn and I.M. Ryzhik, *Tablicy Integralov, Summ i Proizv* (Nauka, Moscow, 1971).
[11] P.O. Löwdin, J. Math. Phys. **3**, 969 (1962).
[12] N.I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, Edinburgh, 1976).
[13] M. Znojil, J. Phys. A: Math. Gen. **9**, 1 (1976).
[14] H. Feshbach, A. Kerman, and S. Koonin, "Theory of statistical multistep reactions," preprint 1977, to be published.
[15] R.R. Whitehead and A. Watt, J. Phys. G: Nucl. Phys. **4**, 835 (1978).
[16] E.g., R.R. Whitehead, Nucl. Phys. A **182**, 290 (1972).
[17] G. Baker, *Padé Approximants in Theoretical Physics* (Academic, New York, 1970).
[18] J. Killingbeck, Rep. Prog. Phys. **40**, 963 (1977).
[19] S. Graffi and V. Grecchi, Lett. Nuovo Cimento **12**, 425 (1975).
[20] W.H. Bassichis, M.R. Strayer, Phys. Rev. C **18**, 1505 (1978).
[21] W.A. Berger, H.G. Miller, K.G. Kreuzer, and R.M. Dreizler, J. Phys. A: Math. Gen. **10**, 1089 (1977).
[22] R. Haydock, J. Phys. A: Math. Nucl. Gen. **7**, 2120 (1974).
[23] G. Turchetti, Fortschr. Phys. **26**, 1 (1978).
[24] G. Turchetti, F. Ortolani, and C. Sagretti, Nuovo Cimento A **44**, 211 (1978).
[25] A.F. Leontiev, *Ryady Exponent* (Nauka, Moscow, 1976).

Regularization of Hamiltonians and processes

S. Albeverio, R. Høegh-Krohn, and L. Streit
Fakultäten für Mathematik und Physik, Universität Bielefeld, 4800 Bielefeld 1, West Germany,
CNRS-CPT, Marseille, France,
U.E.R, Université d'Aix-Marseille, Luminy, France,
and Matematisk Institutt, Universitetet i Oslo, Norway

(Received 6 February 1979; accepted for publication 6 August 1979)

For Markov processes and Hamiltonians given by energy forms, singular drift coefficients and potentials more general than distributions are allowed. We show that the forms can be regularized in such a way tha the approximating processes and semigroups converge as the regularization is removed.

I. INTRODUCTION

The theory of Dirichlet forms on L^2 spaces, which has been developed in connection with potential theory and the theory of symmetric Markov processes (see, e.g., Refs. 1–6) has been applied to quantum mechanics in our previous paper.[7] We have shown in particular that by means of this theory it is possible to define the quantum dynamics for cases where the Hamiltonian cannot be written as the perturbation (in the sense of operator — or form — sums) of the kinetic energy by a potential term, supposed to be a function or a distribution. In fact Dirichlet forms permit to handle cases, like zero range potentials in R^d, which have no meaning as distributions (for $d > 1$) ([7]; see also Ref. 8 for another treatment). By the method of Dirichlet or energy forms the Hamiltonian is defined as the self-adjoint positive operator H_μ associated with the positive form $f \to \int (\nabla f)^2 d\mu$ in $L^2(d\mu)$, where $d\mu$ is a measure on R^d, satisfying some mild regularity condition ([2]; "admissibility" in the sense of Ref. 7). The measure μ is associated with the infimum of the spectrum of H_μ, which need not be an eigenvalue. The associated eigenfunction ("ground state function") determines H_μ. The idea of determining Hamiltonians by "ground state functions" had occurred in connection with the canonical formalism (see e.g., Refs. 9, 10) and has received renewed attention in recent years in connection also with questions of quantum field theory[11–20] and stochastic mechanics.[21–23] Some of the methods developed found applications in the theory of stochastic processes and equations. The method of Dirichlet forms, being a way of introducing Hamiltonians directly by energy forms $\int (\nabla f)^2 \, d\mu$ rather than sums of operators or forms, should be well suited for the handling of applications like, e.g., approximation and stability questions. Some of these questions have been discussed by Ezawa, Klauder, Shepp and Narnhofer.[24–26] In particular these authors discussed in several examples the possibility of approximating the processes given by ground state measures corresponding to smooth potentials.

In the present paper we obtain some general results about approximation of the processes given by Dirichlet forms. We show that energy forms and Hamiltonians given by a measure of the type $d\mu(x) = \rho(x)\,dx$ on R^d can be approximated, in the strong resolvent, i.e., operator semigroup sense, by energy forms given by $d\mu_n(x) = \rho_n(x)\,dx$, where ρ_n converges suitably to ρ, in such a way that ρ can, e.g., correspond to "singular potentials" whereas the ρ_n correspond to smooth potentials. We then obtain weak convergence of the corresponding Markov processes. It is to be noticed that our stress is on results, which do not require (resp. require a minimum of) smoothness or restrictions about growth at infinity for the densities ρ and ρ_n, and hence for the associated drift coefficients or potentials. This is in contrast to the types of approximation results, which are obtained by other methods, like, e.g., in the theory of stochastic equations,[27,28] or in methods based on the existence of a potential as a measurable function,[29–33] or like a distribution.[34] Note that most of our results are valid for processes on R^d, for arbitrary d. Some more detailed results for $d = 1$ are also discussed and they should point, once the L^2-potential theory for arbitrary regions R^d is further developed, to stronger results in the d-dimensional case, too.

II. REGULARIZATION OF INTERACTIONS

Let φ be a real-valued measurable function on R^d such that $\varphi \in L^2_{\text{loc}}(R^d)$ and the measure $d\mu(x) = \varphi(x)^2\,dx$ is an admissible measure in the sense of Ref. 7, i.e., such that the positive quadratic form on $C_0^1(R^d)$, $\int_{R^d} (\nabla f)^2 \, d\mu$ (∇ the gradient), is closable in $L^2(R^d, d\mu) \equiv L^2(d\mu)$. We shall also say for simplicity that such a φ is *admissible*. In Refs. 2 and 7 several sufficient conditions for a function φ to be admissible were given. In particular we recall that $\varphi, \nabla\varphi \in L^2_{\text{loc}}(R^d)$ is a sufficient condition; for other such conditions see Theorems 2.1–2.4 in Ref. 7. Now let φ be admissible; then the closure of the form $\int_{R^d} (\nabla f)^2 \, d\mu$ is a closed positive quadratic form on $L^2(R^d, d\mu)$ called the *energy form* associated with μ. Such a form is a diffusion Dirichlet form in the sense of Ref. 15. Let $H = \nabla^* \nabla$ be the positive self-adjoint operator in $L^2(d\mu)$ uniquely associated with the energy form given by μ, where ∇ is now the closed gradient operator and ∇^* its adjoint in $L^2(d\mu)$. In the case where $\varphi > 0$, almost everywhere with respect to Lebesgue measure, the map $f \to \varphi^{-1} f$ gives a unitary equivalence between $L^2(R^d, dx) \equiv L^2(dx)$ and $L^2(R^d, d\mu) \equiv L^2(d\mu)$ and by this equivalence to H_μ there corresponds a self-adjoint positive operator H in $L^2(R^d, dx)$ such that

$$\varphi^{-1} H f = H_\mu \varphi^{-1} f \tag{2.1}$$

 0022-2488/80/071636-07$1.00

for any $f \in \varphi D(H_\mu)$. We recall that, as shown in Ref. 7, if φ is in addition sufficiently smooth (e.g., $\nabla\varphi$ and $\varphi^{-1}\nabla\varphi \in L^2_{\rm loc}(R^d)$) then the quadratic form associated with H is given for $f \in C_0^1(R^d) \subset L^2(dx)$ by $\int_{R^d} |\nabla f|^2\, dx + \int_{R^d} V f^2 dx$, with $V \equiv \frac{1}{2}\nabla\beta + \frac{1}{4}\beta^2, \beta = 2\varphi^{-1}\nabla\varphi$ in the sense of distributions. In this case, H is a Schrödinger operator, corresponding to a potential V, which is a generalized function. In case φ is smoother, e.g., such that in addition $\varphi^{-1}\Delta\varphi \in L^2_{\rm loc}(R^d)$, where Δ is the Laplacian, then V is a Schrödinger potential in the usual sense of a measurable function on R^d and H is equal to $-\Delta + V$ on $C_0^2(R^d)$. We shall call the operator H given by (2.1) with $H_\mu = \nabla^*\nabla$ a (singular) Schrödinger operator. In general, for φ only supposed to be admissible, it is not possible to write H as the perturbation of the Laplacian Δ by a potential V, which should be a measurable or generalized function. It is, however, natural to ask whether even in this singular case it is possible to find approximations $d\mu_n(x) = \varphi_n^2(x)\, dx$ the measure $d\mu(x)$ so that the associated operators H_n — defined as in (2.1),

$$\varphi_n^{-1} H_n f = H_{\mu_n} \varphi_n^{-1} f, \tag{2.2}$$

with $H_{\mu_n} = \nabla^*\nabla$ in $L^2(R^d, d\mu_n) \equiv L^2(d\mu_n)$, where ∇ is the gradient operator in $L^2(d\mu_n)$ and ∇^* is its adjoint — are "smoother" than H and converge in the resolvent sense to H as $n \to \infty$. In particular we are interested in the situation where H_n can be written in the form

$$H_n = -\Delta + V_n, \tag{2.3}$$

where the V_n are measurable functions, so that the H_n are Schrödinger operators in the usual sense. We shall now examine in general the question of approximants H_n given by (2.2) for H.

Let $\varphi > 0$ Lebesgue-almost everywhere on R^d. Let φ_n be an increasing sequence of positive admissible functions on R^d, bounded pointwise Lebesgue-almost everywhere by φ, i.e.,

$$0 < \varphi_1 \leqslant \varphi_2 < \cdots \leqslant \varphi \quad \text{(a.e.)}. \tag{2.4}$$

Suppose $\varphi \in L^2_{\rm loc}(R^d)$. The operator ∇ is closed in $L^2(d\mu_n)$, by the assumption that φ_n is admissible; thus we have that the operator $\varphi_n \nabla \varphi_n^{-1}$ is the operator in $L^2(R^d, dx)$ unitarily equivalent to ∇ by the unitary transformation $f \to \varphi_n^{-1} f$ from $L^2(dx)$ onto $L^2(d\mu_n)$. Hence $\varphi_n \nabla \varphi_n^{-1}$ is closable when restricted to functions of the form $\varphi_n C_0^1(R^d)$ and $\varphi_n \nabla \varphi_n^{-1}$ is closed. Since, by (2.1), φ_n/φ is a bounded operator defined everywhere in $L^2(dx)$ we then have that $(\varphi_n \nabla \varphi_n^{-1})\varphi_n/\varphi$ is closed in $L^2(dx)$, with domain containing $\varphi C_0^1(R^d)$. Let now $h^{(n)}$ be the positive quadratic form on $L^2(dx)$ with domain $D(h^{(n)}) = \{f \in L^2(dx) | \int (\nabla[f/\varphi])^2 \varphi_n^2\, dx < \infty\}$, defined for $f \in D(h^{(n)})$ by

$$h^{(n)}(f) \equiv \int (\nabla[f/\varphi])^2 \varphi_n^2\, dx. \tag{2.5}$$

Since

$$h^{(n)}(f) = \int (\varphi_n \nabla \varphi_n^{-1}(\varphi_n/\varphi) f)^2\, dx \tag{2.6}$$

by the above considerations we have that $h^{(n)}$ is a closed form. Moreover we have by (2.4),

$$0 \leqslant h^{(n)} \leqslant h^{(n+1)} \leqslant \cdots \leqslant h, \tag{2.7}$$

where h is the energy form in $L^2(dx)$ given by $d\mu = \varphi^2 dx$, i.e.,

$$h(f) = \int (\nabla[f/\varphi])^2 \varphi^2\, dx. \tag{2.8}$$

In particular we then have

$$D(h^{(n)}) \supset D(h) \supset \varphi C_0^1(R^d), \tag{2.9}$$

which shows in particular that the forms $h^{(n)}$ are all densely defined quadratic forms on $L^2(R^d, dx)$. Suppose now $\varphi_n \to \varphi$ Lebesgue-almost everywhere and let $h^{(\infty)}$ be the positive quadratic form defined by

$$h^{(\infty)}(f) = \lim_{n\to\infty} h^{(n)}(f) \tag{2.10}$$

for all f in the domain

$$D(h^{(\infty)}) = \left\{ f \in \cap_n D(h^{(n)}) \,\middle|\, \lim_{n\to\infty} h^{(n)}(f) < \infty \right\}.$$

By a Theorem of Kato[35] and Simon[36] (Theorem 3.1) on monotone sequences of positive closed densely defined quadratic forms we have that $h^{(\infty)}$ is a densely defined, closed, positive quadratic form and that $H^{(n)} \to H^{(\infty)}$ in the strong resolvent sense, where $H^{(n)}$ and $H^{(\infty)}$ are the positive self-adjoint operators in $L^2(dx)$ uniquely associated with the closed positive forms $h^{(n)}$ resp. $h^{(\infty)}$.

By the equality of the domains of the form $h^{(\infty)}$ and the form $(H_{(\varphi)}^{1/2} f, H_{(\varphi)}^{1/2} f)$, with $H_{(\varphi)} \equiv (\varphi\nabla\varphi^{-1})^* \varphi \nabla \varphi^{-1}$, the latter having domain $D(H_{(\varphi)}^{1/2}) = D(\varphi\nabla\varphi^{-1})$ in $L^2(dx)$, we have that $H^{(\infty)} = H_{(\varphi)}$. But $H_{(\varphi)}$ is the positive self-adjoint operator in $L^2(dx)$ associated with the energy form $\int (\nabla[f/\varphi])^2 \varphi^2\, dx$ given by $d\mu = \varphi^2 dx$; hence $H_{(\varphi)} = H$. Thus $H^{(\infty)} = H_{(\varphi)} = H$, $h = h^{(\infty)}$ and

$$H^{(n)} \to H \tag{2.11}$$

in the strong resolvent sense.

We have thus proven the following:

Theorem 1: Let φ_n, φ be admissible functions on R^d such that $0 < \varphi_1 \leqslant \varphi_2 \leqslant \cdots \leqslant \varphi$, $\varphi_n \uparrow \varphi$, Lebesgue-almost everywhere. Let $h^{(n)}(f) \equiv \int (\nabla[f/\varphi])^2 \varphi_n^2\, dx$ and let $H^{(n)}$ be the associated self-adjoint operator on $L^2(R^d, dx)$. Then $h^{(n)}(f) \uparrow h(f)$ as $n \to \infty$ and $H^{(n)} \to H$ in the strong resolvent sense.

Remark: The assumption that φ be admissible can be replaced by the assumption that $\varphi \leqslant \psi$ almost everywhere, with ψ admissible. The a.e. convergence of φ_n to φ yields automatically that φ is admissible, by Theorem 3.1 of Ref. 36. We shall now consider the positive self-adjoint operator

$$H_n \equiv (\varphi_n \nabla \varphi_n^{-1})^* (\varphi_n \nabla \varphi_n^{-1}) \tag{2.12}$$

in $L^2(dx)$, which is the operator in $L^2(dx)$ associated with the closed positive form

$$(H_n^{1/2} f, H_n^{1/2} f) = \int (\nabla[f/\varphi_n])^2 \varphi_n^2\, dx, \tag{2.13}$$

i.e., in case $\varphi_n > 0$ a.e., with the energy form in $L^2(d\mu_n)$ given by $d\mu_n = \varphi_n^2\, dx$. One has, for any $f \in D(H^{(n)1/2}) = D(h^{(n)})$

$$\begin{aligned} h^{(n)}(f) &= (H^{(n)1/2} f, H^{(n)1/2} f) \\ &= \int [\varphi_n \nabla \varphi_n^{-1})(\varphi_n/\varphi) f]^2\, dx \\ &= (H_n^{1/2}(\varphi_n/\varphi) f, H_n^{1/2}(\varphi_n/\varphi) f), \end{aligned} \tag{2.14}$$

hence, by the uniqueness of the self-adjoint operator associated with a given closed positive quadratic form,

$$H^{(n)} = [H_n^{1/2}(\varphi_n/\varphi)]^* H_n^{1/2}(\varphi_n/\varphi). \tag{2.15}$$

We shall now see that from (2.15) follows

$$H^{(n)} = (\varphi_n/\varphi) H_n (\varphi_n/\varphi). \tag{2.16}$$

In fact this is a consequence of the following.

Lemma 2: Let A,B be self-adjoint operators in a Hilbert space $\mathscr{H}$ and suppose that the domain $D(AB)$ is dense, that AB is closable, that $D(B)$ contains the range of A, that B^{-1} exists, and that one has $D(B^{-1}(AB)^*) \supset D(AB)^*$. Then $(AB)^* = B^* A^*$.

Corollary: If A,B are self-adjoint operators in a Hilbert space $\mathscr{H}$, with B, B^{-1} bounded and $D(AB)$ dense, AB closable, then $(AB)^* = B^* A^*$.

Proof: That $(AB)^* \supset B^* A^* = BA$ is obvious. By the assumption that B^{-1} exists on a domain containing $D(A)$ we have

$$(AB)^{**} B^{-1} \supset A. \tag{2.17}$$

From this it follows easily, using the self-adjointness of B^{-1}, that

$$((AB)^{**} B^{-1})^* \subset A^* = A. \tag{2.18}$$

On the other hand,

$$((AB)^{**} B^{-1})^* \supset B^{-1}(AB)^*; \tag{2.19}$$

hence, from (2.18),

$$A\chi = B^{-1}(AB)^* \chi \tag{2.20}$$

for all $\chi \in D(B^{-1}(AB)^*)$. Since by assumption the range of A is contained in $D(B)$, (2.20) yields

$$BA\chi = (AB)^* \chi \tag{2.21}$$

for all $\chi \in D(B^{-1}(AB)^*)$; thus *a fortiori*, due to the assumption $D(B^{-1}(AB)^*) \supset D((AB)^*)$,

$$BA \supset (AB)^*. \tag{2.22}$$

But (2.22) yields then the Lemma.■ As a consequence of the Corollary to Lemma 2 we have the following

Lemma 3: Let φ_n be admissible, $\varphi_n > 0$ a.e., and assume $\varphi/\varphi_n \in L^\infty(R^d)$. Then

$$H^{(n)} = (\varphi_n/\varphi) H_n (\varphi_n/\varphi)$$

as self-adjoint operators in $L^2(d\mu_n)$.

Proof: Take $= L^2(d\mu_n)$, $A = H_n^{1/2}$, and $B = \varphi_n/\varphi$ in Corollary of Lemma 2. Then, $D(H_n^{1/2}(\varphi_n/\varphi))$ $= D(H^{(n)1/2})$ being dense, we see that the assumption $D(AB)$ dense is fulfilled. Also $(AB)^*$ exists and is densely defined, since $(\varphi_n/\varphi)H_n^{1/2}$ is densely defined, φ_n/φ being bounded. Hence the assumption that AB be closable is satisfied. The lemma follows from the Corollary to Lemma 2. ■

In order now to get a corresponding result to Theorem 1 for the Schrödinger operator H_n instead of $H^{(n)}$ we shall use Lemma 2 together with the following result, which follows immediately from a general result of Kurtz[37] whose Hilbert space version says that strong resolvent convergence is equivalent to the existence of the strong graph limit (see e.g., Ref. 38, p. 293).

Lemma 4: For all $g \in D(H)$ there exists a sequence $g_n \in D(H^{(n)})$, converging strongly in $L^2(dx)$ to g and such that $H^{(n)} g_n$ converges strongly in $L^2(dx)$ to Hg as $n \to \infty$.■ Let now g be arbitrary in $D(H) \subset L^2(dx)$ and let g_n be as in Lemma 4. Consider the sequence $f_n = (\varphi_n/\varphi) g_n$ in $L^2(dx)$. Suppose $\varphi_n/\varphi \uparrow 1$ in $L^\infty(dx)$, then we have that g is the strong limit in $L^2(dx)$ of f_n. Consider now $H_n f_n$. We have $H_n f_n = H_n(\varphi_n/\varphi) g_n$ and from Lemma 3,

$$H_n(\varphi_n/\varphi) g_n = (\varphi/\varphi_n) H^{(n)} g_n, \tag{2.23}$$

since $g_n \in D(H^{(n)})$.

But $H^{(n)} g_n \to Hg$ strongly as $n \to \infty$, by Lemma 4. On the other hand $\varphi_n/\varphi \uparrow 1$ in $L^\infty(dx)$ implies $(\varphi/\varphi_n) \downarrow 1$ in $L^\infty(dx)$. Then $H^{(n)} g_n \to Hg$ implies $(\varphi/\varphi_n) H^{(n)} g_n \to Hg$ strongly as $n \to \infty$. Thus, from (2.23), $H_n f_n \to Hg$, strongly as $n \to \infty$. We now want to use the result of Kurtz mentioned above that the strong resolvent convergence is equivalent to the existence of the strong graph limit. First we remark that for a pair $\{g, Hg\}$, $g \in D(H)$ we have proven the existence of a sequence $f_n \in D(H_n)$ with $f_n \to g$ and $H_n f_n \to Hg$, strongly as $n \to \infty$. This is, however, the same as to say that $\{g, Hg\}$ is in the strong graph limit of H_n and hence one has the strong resolvent convergence of H_n to H. Hence we have proven the following

Theorem 5: Let φ_n, φ be admissible functions on R^d such that $0 < \varphi_1 \leqslant \varphi_2 \leqslant \cdots \leqslant \varphi$ a.e., $\varphi_n/\varphi \uparrow 1$ in $L^\infty(dx)$. Let H_n be the self-adjoint operator in $L^2(dx)$ associated with the energy form in $L^2(d\mu_n)$ given by $d\mu_n = \varphi_n^2\, dx$. Let H be the self-adjoint operator in $L^2(dx)$ associated with the energy form in $L^2(d\mu)$ given by $d\mu = \varphi^2\, dx$. Then one has that H_n converges to H as $n \to \infty$ in the strong resolvent sense.■

Remark: As in the Remark following Theorem 1 we can replace the assumption that φ be admissible by the assumption that $\varphi \leqslant \psi$ a.e., with ψ admissible. We shall now consider the situation where φ_n is a decreasing sequence of positive admissible functions on R^d, such that $\varphi_1 \geqslant \varphi_2 \geqslant \cdots \geqslant \varphi_n \geqslant \cdots \geqslant \varphi > 0$ a.e. Then we have, defining $h^{(n)}$, h as in (2.6), (2.8)

$$h^{(1)} \geqslant h^{(2)} \geqslant \cdots \geqslant h^{(n)} \geqslant \cdots \geqslant h \geqslant 0.$$

Suppose $\varphi_n \downarrow \varphi$ a.e. Let $D(h_\infty) \equiv \cup_n D(h^{(n)}) \subset D(h)$ and for $f \in D(h_\infty)$, $h_\infty(f) \equiv \lim_{n\to\infty} h^{(n)}(f)$. From a result of Simon (Theorem 3.2 in Ref. 36) one has that $H^{(n)}$ converges in the strong resolvent sense as $n \to \infty$ to the closure of the largest closable quadratic form less than h_∞. Assume φ is admissible. Then h is closed being the energy form associated with φ, hence h restricted to $D(h_\infty)$ is closable and coincides with h_∞. By a result of Kato (Ref. 35, Theorem VIII, 3.11) one has then that the closure of h_∞ is actually equal to h and $H^{(n)}$ converges in the strong resolvent sense as $n \to \infty$ to H. If we suppose moreover that φ_n/φ is in $L^\infty(dx)$, then we have, analogously as in the proof of Lemma 3, that $(\varphi_n/\varphi) H_n (\varphi_n/\varphi) = H^{(n)}$ and if $\varphi_n/\varphi \downarrow 1$ in $L^\infty(dx)$ we have then, as in Theorem 5, that $H_n \to H$ in the strong resolvent sense. Hence we have the following theorem, which corresponds to Theorems 1 and 5 in the case of a monotonic decreasing sequence φ_n rather than a monotonic increasing one.

Theorem 6: Let φ_n be admissible and such that $\varphi_1 \geqslant \varphi_2 \geqslant \varphi_n \geqslant \cdots > \varphi > 0$ a.e. Suppose moreover $\varphi_n \downarrow \varphi$ a.e. Let $h^{(n)}(f) = \int (\nabla[f/\varphi])^2 \varphi_n^2 (dx)$ and let $H^{(n)}$ be the associated self-adjoint operator on $L^2(dx)$. Then $h^{(n)}(f) \downarrow h(f)$ as $n \to \infty$, where $h(f) = \int (\nabla[f/\varphi])^2 \varphi^2\, dx$ is the energy form given by $\varphi^2\, dx$ and H is the associated self-adjoint operator on $L^2(dx)$. Moreover if $\varphi_n/\varphi \downarrow 1$ in $L^\infty(dx)$ then $H_n \to H$ in the strong resolvent sense, where H_n is the energy operator in $L^2(dx)$ associated with the energy form given by $\varphi_n^2\, dx$. ■

Examples: (1) Let $d = N$, $\varphi(x) = e^{-\lambda \sum_{i<j} |x_i - x_j|}$, $x = (x_1, \ldots, x_N), x_i \in R, \lambda \in R$. φ is an admissible function and the corresponding energy form is

$$\int (\nabla[f/\varphi])^2 \varphi^2\, dx = \int (\nabla f)^2\, dx - 4\lambda \sum_{i<j} \int f(x_1 \ldots x_i = x_j \ldots x_N)^2 \times \prod_{k \neq i} dx_k + N(N-1)\lambda^2 \int f^2\, dx \,. \tag{2.24}$$

The associated Hamiltonian, formally described by $-\Delta - 4\lambda \Sigma_{i<j} \delta(x_i - x_j) + N(N-1)\lambda^2$, has been studied before in several connections.[39,40] If we take $\varphi_n(x) = \exp[-\lambda \Sigma_{i<j} (|x_i - x_j|^2 + 1/n)^{1/2}]$, then we have that $V_n(x) = \Delta\varphi_n(x)/\varphi_n(x)$ are smooth potentials approaching zero at infinity; hence $H_n = -\Delta + \Sigma_{i<j} V_n(x_i - x_j) + N(N-1)\lambda^2$ is a Hamiltonian, self-adjoint on the domain of Δ, with two-particle smooth interaction. We can apply Theorem 5 to this case to obtain the strong resolvent convergence of the regularized Hamiltonian H_n to H, H being the self-adjoint operator associated with the energy form given by (2.24).

(2) Consider now the equation $-\varphi'' + \lambda |x|^{-\alpha} \varphi = 0$ on R, with $1 < \alpha < 2$. It is noted in Ref. 7 that one has a one-parameter family of solutions lying in $L^2_{\text{loc}}(R)$, $\varphi^\beta = (1-\beta)\varphi^0 + \beta\varphi^1$, $\beta \in R$, with $\varphi^0(x) = x + O(x^{3-\alpha})$ and $\varphi^1(x) = 1 + O(x^{2-\alpha})$ in a neighborhood of $x = 0$, so that φ^β are admissible. Consider, e.g., $\varphi(x) \equiv \varphi^1(x)$, then

$$\varphi(x) = 1 - \frac{\lambda}{(2-\alpha)(\alpha-1)} \times |x|^{2-\alpha} + O(|x|^{2-\alpha})$$

in a neighborhood of $x = 0$. For $\lambda > 0$ we see that we can approximate $\varphi(x)$ by smooth strictly positive functions $\varphi_n(x)$ such that $0 < \varphi_n \leqslant \varphi$ and such that for $|x_n| \leqslant a_n$ one has $\varphi_n(x) = 1 - 1/n - d_n x^2$, where a_n, d_n are suitable constants and $\varphi_n(x) = \varphi(x)$ for $|x| \geqslant a_n$. For the continuity of φ_n and φ'_n at $x = a_n$ one has to require

$$d = \frac{\lambda}{2(\alpha-1)} \left(\frac{\lambda\alpha}{(\alpha-1)(2-\alpha)n} \right)^{-\alpha/(2-\alpha)}$$

and

$$a_n = \left(\frac{\lambda\alpha}{(\alpha-1)(2-\alpha)n} \right)^{1/(2-\alpha)}.$$

For $|x| \leqslant a_n$ one has then

$$V_n = \frac{\Delta\varphi_n}{\varphi_n} = \frac{-2d_n}{1 - (1/n) - d_n x^2},$$

a smooth "approximation" of a square well potential of width a_n and depth $-2d_n/(1-n^{-1})$.

Remark: The assumptions for the results in this section involve convergence of φ_n to φ in a strong sense; however, they do not involve smoothness and/or growth conditions on φ_n and φ, hence also not on the corresponding drift coefficients $\beta_n = 2\nabla\varphi_n/\varphi_n$, $\beta = 2\nabla\varphi/\varphi$ (see Ref. 7) nor the corresponding potentials $V_n = \Delta\varphi_n/\varphi_n$, $V = \Delta\varphi/\varphi$, which might very well not be measurable functions or distributions. Let us mention that there are many results in the literature, involving other techniques, for the cases where V_n, V exist as measurable functions or distributions. Sufficient conditions for strong resolvent convergence are e.g.,

(a)Refs. 32,33: $|V_n| \leqslant |V_{n+1}| \leqslant |V|$, $V_n \to V$ pointwise, $|V|^{1/2} D(\nabla) \cap L^2(dx)$ dense in $L^2(dx)$, $V_- \leqslant A(-\Delta) + B$, as forms on $D(\nabla)$, with $V_- = -\min(V,0)$;

(b)Refs. 29,30: V_n, $V \in L^1_{\text{loc}}$, V_n, $V \geqslant 0$, $V \to V_n$ in L^1_{loc}

(c)Ref. 34: $V = V_+ + V_-$, V_- distributional and "small" with respect to $(-\Delta)$ in the sense of forms, V_n involving smooth approximations of V_- (see Ref. 34 for details).

Other results obtained by methods of stochastic equations and/or for the case $d = 1$ will be mentioned in the next section.

III. WEAK CONVERGENCE OF REGULARIZED PROCESSES

We first consider the same situation as in Theorems 5 and 6 in Sec. II. Let φ_n be a sequence of admissible functions. Then the energy form $f \to \int (\nabla f)^2 d\mu_n = (H_{\mu_n}^{1/2} f, H_{\mu_n}^{1/2} f)$ is a closed positive quadratic form on the domain $D(H_{\mu_n}^{1/2}) \subset L^2(d\mu_n)$ where H_{μ_n} is the self-adjoint positive operator given by $H_{\mu_n} = \nabla^*\nabla$, where ∇ is the closure in $L^2(d\mu_n)$ of the gradient operator considered as acting on $C_0^1(R^d)$ functions in $L^2(d\mu_n)$. Let $\varphi_n^2 > 0$ a.e., then $L^2(d\mu_n)$ is unitarily equivalent by the map $f \to \varphi_n f$ to $L^2(dx)$. By this equivalence, to H_{μ_n} there corresponds the self-adjoint positive operator H_n such that $\varphi_n^{-1} H_n \varphi_n f = H_{\mu_n} f$ for any $f \in D(H_{\mu_n})$. Under the assumptions of Theorem 5 or 6 we have that the operator H_n in $L^2(dx)$ converges in the strong resolvent sense to the operator H in $L^2(dx)$, where φ is admissible and H is defined by $\varphi^{-1} H \varphi f = H_\mu f$, for any $f \in D(H_\mu)$, $H_\mu = \nabla^*\nabla$ being the energy operator uniquely associated with the energy form given by the admissible function φ. By Trotter's theorem the strong resolvent converges of H_n to H is equivalent to the strong convergence of the semigroups e^{-tH_n} in $L^2(dx)$ to the semigroup e^{-tH} in $L^2(dx)$, $t > 0$. By Fukushima's results we have that $e^{-tH_{\mu_n}}$ and e^{-tH_μ}, and hence e^{-tH_n} and e^{-tH}, are Markov semigroups in $L^2(d\mu_n)$, $L^2(d\mu)$ resp. $L^2(dx)$. Suppose now first that $\int d\mu_n < \infty$ so that we can assume $\int d\mu_n = 1$. Let $t \to \xi_n(t)$ be a realization of the homogeneous Markov process given by the invariant measure μ_n and the Markov semigroup $e^{-tH_{\mu_n}}$. Let f_i, $i = 1, \ldots, k$ be bounded Borel func-

tions on R^d. Consider the expectations

$$\int f_1(\xi_n(t_1))\cdots f_k(\xi_n(t_k))d\mu^*{}_n$$

for $t_1 \geqslant \cdots \geqslant t_k$, where $\mu^*{}_n$ is the measure associated to the process $\xi_n(t)$ i.e., on the space of trajectories of $\xi_n(t)$. Since $\xi_n(t)$ is a Markov process with semigroup $e^{-tH_{\mu_n}}$ and invariant measure μ_n we have

$$\begin{aligned}&\int f_1(\xi_n(t_1))\cdots f_k(\xi_n(t_k))d\mu^*{}_n\\ &= (f_1,\exp[-(t_1-t_2)H_{\mu_n}]f_2\exp[-(t_2-t_3)H_{\mu_n}]\\ &\quad\times\exp[-(t_{k-1}-t_k)H_{\mu_n}]f_k)_n, \qquad (3.1)\end{aligned}$$

where $(\ ,\ )_n$ is the scalar product in $L^2(d\mu_n)$. But

$$\begin{aligned}&(f_1,\exp[-(t_1-t_2)H_{\mu_n}]f_2\cdots\\ &\quad\times\exp[-(t_{k-1}-t_k)H_{\mu_n}]f_k)_n\\ &\quad=(\varphi_n f_1,\varphi_n\exp[-(t_1-t_2)H_{\mu_n}]\varphi_n^{-1}f_2\varphi_n\\ &\qquad\times\exp[-(t_2-t_3)H_{\mu_n}]\cdots\\ &\qquad\times\exp[-(t_{k-1}-t_k H_{\mu_n}]\varphi_n^{-1}\varphi_n f_k), \qquad (3.2)\end{aligned}$$

where $(\ ,\ )$ is the scalar product in $L^2(dx)$. Using that, from (2.1),

$$\varphi_n e^{-tH\mu_n}\varphi_n^{-1} = e^{-tH_n}, \qquad (3.3)$$

we get that (3.2) is equal to

$$\begin{aligned}&(\varphi_n f_1,\exp[-(t_1-t_2)H_n]f_2\exp[-(t_2-t_3H_n]\cdots\\ &\quad\times\exp[-(t_{k-1}-t_k)H_n]\varphi_n f_k). \qquad (3.4)\end{aligned}$$

Suppose now that the assumptions of Theorem 5 or 6 are satisfied, so that one has strong convergence of the semigroup e^{-tH_n}, $t \geqslant 0$ to the semigroup e^{-tH}, $t \geqslant 0$, as $n \to \infty$. Moreover, under the assumptions of Theorem 5 or 6 we have $\varphi_n h \to \varphi h$ strongly as $n \to \infty$, for all $h \in D(\varphi)$; hence we have that (3.4) converges as $n \to \infty$ to $(\varphi f_1,$ $\exp[-(t_1-t_2)H]f_2\ldots\exp[-(t_{k-1}-t_k)]\varphi f_k)$; hence from (3.1), (3.2) using also that e^{-tH} is a contraction, we get

$$\begin{aligned}&\lim_{n\to\infty}\int f_1(\xi_n(t_1))\cdots f_k(\xi_n(t_k))\,d\mu^*{}_n\\ &\quad=(\varphi f_1,\exp[-(t_1-t_2)H]f_2\cdots\\ &\qquad\times\exp[-(t_{k-1}-t_k)H]\varphi f_k); \qquad (3.5)\end{aligned}$$

thus using

$$\varphi e^{-tH_n}\varphi^{-1} = e^{-tH},$$

which follows from (2.1), we get

$$\begin{aligned}&\lim_{n\to\infty}\int f_1(\xi_n(t_1))\cdots f_k(\xi_n(t_k))d\mu^*{}_n\\ &\quad=\int f_1(\xi(t_1))\cdots f_k(\xi(t_k))\,d\mu^*, \qquad (3.6)\end{aligned}$$

where $\xi(t)$ is the Markov process given by the invariant measure μ and the Markov semigroup $\exp(-tH_\mu)$ in $L^2(d\mu)$ and μ^* is the corresponding measure on path space. From this it follows that the convergence (3.6) holds also when $\Pi_{j=1} f_j(\xi_n(t_j))$ is replaced by $g(\xi_n(t_1)),\ldots,\xi_n(t_k)$ for arbitrary $g \in C_b(R^{kd})$. In particular the finite dimensional distributions of ξ_n converges to those of ξ. Thus we have proven the following.

Theorem 7: Let the assumptions of Theorem 5 or Theorem 6 be satisfied. Let $d\mu_n = \varphi_n^2\,dx$ be probability measures and let ξ_n, ξ be the homogeneous conservative Markov processes given by the Markov semigroups $\exp(-tH_{\mu_n})$ resp. $\exp(-tH_\mu)$, $t \geqslant 0$, and the invariant measures μ_n resp. μ. Then $\xi_n \to \xi$ in the sense that the finite dimensional distributions converge. ■

In the case where we do not necessarily have $\int d\mu_n < \infty$ we remark that according to Ref. 7 we have, for $\varphi_n > 0$, a.e., that the energy forms associated with φ_n resp. φ are regular in the sense of Fukushima.[2] As shown by Fukushima[2-4] (see also Refs. 5 and 6), we can associate properly to these forms Hunt processes $\xi_n^x(t)$ resp. $\xi^x(t)$, started at x in R^d modulo polar sets. Moreover, for quasi-every x in R^d and all $\lambda > 0$, the kernels $G_\lambda^n(x,dy)$ resp. $G_\lambda(x,dx)$ of $G_\lambda^n \equiv (H_{\mu_n}+\lambda)^{-1}$ and $G_\lambda = (H_\mu+\lambda)^{-1}$ are defined, so that for $t > 0$ and $f \in C_0(R^d)$ (continuous functions of bounded support), one has

$$(G_\lambda^n f)(x) = \int G_\lambda^n(x,dy)f(y)$$

resp.

$$(G_\lambda f)(x) = \int G_\lambda(x,dy)f(y).$$

Because of $\varphi_n(H_{\mu_n}+\lambda)^{-1}\varphi_n^{-1} = (H_n+\lambda)^{-1}$, $\varphi(H_\mu+\lambda)\varphi^{-1} = (H+\lambda)^{-1}$, for all $\alpha > 0$ [which is a consequence of (2.1)] we see that the kernels $G_\lambda^{n,\varphi_n}(x,dy)$ and $G_\lambda^{\varphi}(x,dy)$ of $(H_n+\lambda)^{-1}$, $(H+\lambda)^{-1}$ are defined by

$$G_\lambda^{n,\varphi_n}(x,dy) = \varphi_n(x)G_\lambda^n(x,dy)\varphi_n^{-1}(y) \qquad (3.7)$$

and

$$G_\lambda^{\varphi}(x,dy) = \varphi(x)G_\lambda(x,dy)\varphi^{-1}(y). \qquad (3.8)$$

Now let φ_n, φ be as in Theorems 5 or 6, so that one has the strong resolvent convergence of H_n to H. Then, strongly in $L^2(dx)$

$$\begin{aligned}&(H_n+\lambda)^{-1}f_1(H_n+\lambda)^{-1}f_2\cdots(H_n+\lambda)^{-1}f_k\\ &\quad\to(H+\lambda)^{-1}f_1(H+\lambda)^{-1}f_2\cdots(H+\lambda)^{-1}f_k\end{aligned}$$

as $n \to \infty$, for all $f_i \in L^2(dx)$, $(H_n+\lambda)^{-1}$ being uniformly bounded. Thus

$$\begin{aligned}&\int_{R^{kd}} G^{n,\varphi_n}(x,dy_1)f_1(y_1)G_\lambda^{n,\varphi_n}(y_1,dy_2)f_2(y_2)\\ &\qquad\times\cdots G^{n,\varphi_n}(y_{k-1},dy_k)f_k(y_k)\\ &\quad\to\int G_\lambda^{\varphi}(\cdot,dy_1)f_1(y_1)G_\lambda^{\varphi}(y_1,dy_2)f_2(y_2)\\ &\qquad\times G_\lambda^{\varphi}(y_{k-1},dy_k)f_k(y_k)\end{aligned}$$

as $n \to \infty$, strongly in $L^2(dx)$. In particular the finite dimensional distributions converge as $n \to \infty$, strongly in $L^2(dx)$. By subsequences we have then also convergence of the finite dimensional distributions, for a.e. initial condition x. We have thus proven the following.

Theorem 8: Let the assumptions of Theorem 5 or Theorem 6 be satisfied. Let $\xi_n^x(t), \xi^x(t)$ be the regular Hunt processes properly associated in the sense of Fukushima with the regular energy forms given by $d\mu_n = \varphi_n^2\,dx$. Then ξ_n^x

$\to \xi^x$ in the sense of convergence of finite dimensional distributions, strongly in $L^2(dx)$ and, by subsequences, pointwise almost surely for every x.■ We shall conclude with some remarks. In the case, where the assumptions on φ_n ,φ are not strong enough to allow to conclude that H_n necessarily converges strongly to H, we still have some control on the limit by the following theorem:

Theorem 9: Let φ_n be admissible and suppose $\varphi_n \uparrow \varphi$ a.e., with φ admissible and in $L^2(dx)$. Suppose $C_0^2(R^d) \subset D(H_n)$, then $(f, H_n g) \to (f, Hg)$, for all $f,g \in C_0^2(R^d)$, where $(\ ,\ )$ is the scalar product in $L^2(dx)$. If, moreover, $\beta \in L^2(dx)$, then $E_{\mu_n}(f(\xi_n(0))\, g(\xi_n(t)))$ has uniformly bounded second derivative with respect to t, where E_{μ_n} is the expectation with respect to the stationary Markov process ξ_n with infinitesimal generator H_{μ_n} and invariant finite measure $d\mu_n$, and $f,g \in C_0^2(R^d)$. If, moreover, $\xi_n(t)$ converges weakly to some process, $\eta(t)$ (in the sense of convergence of the finite dimensional distributions), then $E_\mu((f(\eta(0))g(\eta(0)))$ is a twice differentiable function of t and

$$-\frac{d}{dt}E(f(\eta(0))g(\eta(t)))\Big|_{t=0} = (f, Hg)$$

for any $f,g \in C^2(R^d)$. If $\eta(t)$ is a Markov process with infinitesimal generator H_η we have $H_\eta = H$ on $C_0^2(R^d)$.

Proof: The proof follows from the proof of Theorem 3.7 in Ref. 11. ■

Remark: The assumption that φ be admissible is e.g., satisfied when the φ_n are such that $\int (\nabla f)^2 \varphi_n^2\, dx \leqslant C(\int f^2 \times \varphi_n^2\, dx)^{1/2}$ for all $f \in C_0^\infty(R^n)$. In this case one has namely $C_0^1(R^d) \subset D(\nabla^*)$ as follows from Ref. 12.

Remark: The case of $d = 1$ gives an indication that the assumptions of Theorems 5 and 6 are not optimal. From a result of Rosenkrantz[41] we can prove, e.g., the following result.

Theorem 10: Let $d = 1$ and let φ_n be admissible such that the corresponding Markov process ξ_n has continuous paths. Assume $1/\varphi_n \in L^2_{\rm loc}(R^d)$ and

$$\int_0^{\pm\infty} dx\, \varphi_n(x)^2 \int_0^x \varphi_n(y)^{-2}\, dy = \infty$$

and

$$\int_0^{\pm\infty} dx\, \varphi_n(x)^{-2} \int_0^x \varphi_n(y)^2\, dy = \infty$$

(so that $\pm\infty$ are natural boundaries for the process ξ_n). Assume $\varphi_n(x) \to \varphi(x)$ pointwise for all point of continuity of $\varphi(x)$. Then one has for every $\lambda > 0$

$$\lim_{n\to\infty} \sup_{x\in R} \left| [(\lambda + H_{\mu_n})^{-1} f](x) - [(\lambda + H_\mu)^{-1} f](x) \right| = 0,$$

for all $f \in C_\infty(R)$, $C_\infty(R)$ being the Banach space of bounded continuous functions vanishing at infinity. This implies

$$\lim_{n\to\infty} \sup_{x\in R} \left| \exp(-tH_{\mu_n}) f - \exp(-tH_\mu) f \right| = 0.$$

Moreover $\xi_n \to \xi$ weakly in the sense of weak convergence of the finite dimensional distributions.

Proof: In order to apply the result of Rosenkrantz it suffices to verify that the assumptions

$$\int_0^{\pm\infty} dx\, \varphi_n(x)^{-2} \int_0^x dy \varphi_n(y)^2\, dy = \int_0^{\pm\infty} dx\, \varphi_n(x)^2 \int_0^x dy\, \varphi_n(y)^{-2} = \infty$$

imply that $\pm\infty$ are natural boundaries for the process ξ_n. This is, however, easily done noticing that, in standard notations (see e.g., Refs. 41 and 42) the scale function to ξ_n (assuming, without loss of generality, that 0 is a point of continuity) is $p_n(x) = \int_0^x (\varphi_n(s)/\varphi_n(0))^2\, ds$ and the speed measure to ξ_n is $dm_n(x) = (\varphi_n(0)/\varphi_n(x))^2\, dx$. Then it suffices to apply the known criteria for natural boundaries. The rest follows from Ref. 41 and e.g., Ref. 43. ■

Remark: Results concerning resolvent convergence involving, however, growth or smoothness conditions on β or V follow also from other methods of stochastic equations.[44] The case $d = 1$ with smooth φ [namely $\varphi \in \mathcal{S}(R)$] has been treated in details in Ref. 20. We shall give elsewhere a detailed treatment of the general $d = 1$ case.

ADKNOWLEDGMENTS

We are grateful to Prof. Derek Robinson for an instructive discussion. This work was initiated during stays of the authors at the Zentrum für Interdisziplinäre Forschung der Universität Bielefeld. Further work has been done during a common stay at the Centre de Physique Théorique, CNRS, Marseille and U.E.R. Université d'Aix-Marseille, Luminy. The authors are grateful to these institutions for their invitations and hospitality. We would also like to thank Hedwig Sorhage for her patience and her nice typing of the manuscript.

[1] J. Deny, Méthodes Hilbertiennes en théorie du potentiel, in *Potential Theory* (C.I.M.E, Ed. Cremonese, Roma, 1970), pp. 121–201.
[2] M. Fukushima, in Proceedings of the Second Japan-USSR Symposium on Probability Theory, *Lecture Notes in Mathematics* 330, edited by G. Maruyama and Yu. V. Prokhorov (Springer, Berlin, 1976), pp. 46–79.
[3] M. Fukushima, in Proceedings of the Third Japan-USSR Symposium on Probability Theory, Lecture Notes in Mathematics **550** (Springer, Berlin, 1976).
[4] M. Fukushima, *Dirichlet Forms and Markov Processes* (Kodansha, Tokyo; North-Holland, Amsterdam) (to appear).
[5] M.L. Silverstein, Symmetric Markov Processes, *Lecture Notes in Mathematics* **426** (Springer, Berlin, 1974).
[6] M.L. Silverstein, Boundary Theory for Symmetric Markov Processes, *Lecture Notes in Mathematics* **516** (Springer, Berlin, 1976).
[7] S. Albeverio, R. Høegh-Krohn, and L. Streit, Energy forms, Hamiltonians, and distorted Brownian paths, J. Math. Phys. **18**, 907–917 (1977).
[8] S. Albeverio, J.E. Fenstadt, and R. Høegh-Krohn, "Singular perturbations and non-standard analysis," Trans. Am. Math. Soc. **252**, 275–95 (1979).
[9] F. Coester, and R. Haag, "Representation of states in a field theory with canonical variables," Phys. Rev. **117**, 1137–1145 (1970).
[10] H. Araki, "Hamiltonian formalism and the canonical commutation relations in quantum field theory," J. Math. Phys. **1**, 492–504 (1960).
[11] S. Albeverio, and R. Høegh-Krohn, Quasi-invariant measures, symmetric diffusion processes and quantum fields, in: Les méthodes mathématiques de la théorie quantique des champs, Colloq. Int. CNRS **248**, 11–59 (1976).
[12] S. Albeverio, and R. Høegh-Krohn, Dirichlet forms and diffusion processes on rigged Hilbert spaces, Zeitschr. f. Wahrscheinlichkeitstheorie verw. geb. **40**, 1–57 (1977).
[13] S. Albeverio, and R. Høegh-Krohn, Hunt processes and analytic potential theory on rigged Hilbert spaces, Ann. Inst. H. Poincaré B **13**, 269–291 (1977).

[14]T. Hida, and L. Streit, "On quantum theory in terms of white noise," Nagoya Math. J. **68**, 21–34 (1977).
[15]S. Albeverio, and R. Høegh-Krohn, "The structure of diffusion processes," Marseille Preprint; S. Albeverio, and R. Høegh-Krohn, Topics in infinite dimensional analysis, pp. 279–302 in "Mathematical Problems in Theoretical Physics," edited by G.F. Dell'Antonio, S. Doplicher, and G. Jona-Lasinio, *Lecture Notes in Physics* (Springer, Berlin, 1978), Vol. 80.
[16]L. Gross, "Logarithmic Sobolev inequalities," Amer. J. Math. **97**, 1061–1083 (1976).
[17]J.P. Eckmann,"Hypercontractivity for anharmonic oscillators," J. Funct. Anal. **16**, 388–404 (1974).
[18]I. Herbst, "On canonical quantum field theories," J. Math. Phys. **17**, 1210–1221 (1976).
[19]P. Courrège, and R. Renouard, "Oscillateur anharmonique, mesures quasi-invariantes sur $C(R,R)$ et théorie quantique des champs en dimension $d = 1$," Astérisque **22–23**, 3–245 (1975).
[20]P. Priouret, and M. Yor, "Processus de diffusion à valeurs dans R et mesures quasi-invariantes sur $C(R,R)$," Astérisque **22–23**, 247–290 (1975).
[21]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U. P., Princeton, N.J., 1967).
[22]F. Guerra, On stochastic field theory, Proc. II. Aix-en-Provence Int. Conf. Elem. Part. 1973, J. Phys. Suppl. *T* **34**, Fasc. 11-12, Coll. C1-95-98.
[23]S. Albeverio, and R. Høegh-Krohn, "A remark on the connection between stochastic mechanics and the heat equation," J. Math. Phys. **15**, 1745–1747 (1974).
[24]H. Ezawa, J.R. Klauder, and L.A. Shepp, A path space picture for Feynman–Kac averages, Ann. Phys. **88**, 588–620 (1974).
[25]H. Ezawa, J.R. Klauder, and L.A. Shepp, Vestigial effects of singular potentials in diffusion theory and quantum mechanics, J. Math. Phys. **16**, 783–799 (1975).
[26]H. Narnhofer, and J.R. Klauder, Boundary conditions and singular potentials in diffusion theory, J. Math. Phys. **17**, 1201–1209 (1976).
[27]I.I. Gihman, and A.V. Skorohod, *Stochastic Differential Equations* (Springer, Berlin, 1972).
[28]A. Friedman, *Stochastic Differential Equations and Applications* (Academic, New York, 1975).
[29]T. Kato, Remarks on Schrödinger operators with vector potentials, to be published in "Integral Equations and Operator Theory."
[30]B. Simon, "Maximal and minimal Schrödinger forms," J. Operator Theory **1** 37–49 (1979).
[31]W. Faris, Self-adjoint operators, *Lecture Notes in Mathematics* **433**, (Springer, Berlin, 1975).
[32]D. Robinson, "Scattering theory with singular potentials, I. The two-body problem," Ann. Inst. Henri Poincaré A **21**, 185–215 (1974).
[33]M. Schechter, "Cut-off potentials and form extensions," Lett. Math. Phys. **1**, 265–273 (1976).
[34]I.W. Herbst, and A.D. Sloan, Perturbation of translation invariant positivity preserving semigroups on $L(R^N)$, Trans. Am. Math. Soc. **236**, 325–60 (1978).
[35]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966), Chap. VIII Sec. 3.
[36]B. Simon, "A canonical decomposition for quadratic forms with applications to monotone convergence theorems," J. Funct. Anal. **28**, 377–85 (1978).
[37]T. Kurtz, "Extensions of Trotter's operator semigroup approximation theorems," J. Funct. Anal. **3**, 354–375 (1969).
[38]M. Reed, and B. Simon, *Methods of Modern Mathematical Physics*, I Functional Analysis (Academic, New York, 1972).
[39]J.B. McGuire, "Study of exactly soluble one-dimensional N-body problems," J. Math. Phys. **5**, 622–636 (1964).
[40]C.N. Yang, S-matrix for the one-dimensional N-body problem with repulsive or attractive δ-function interaction, Phys. Rev. **168**, 1920–1923 (1967).
[41]W.A. Rosenkrantz, "Limit theorems for solutions of a class of stochastic differential equations," Ind. University, Math. J. **24**, 613–625 (1975).
[42]P. Mandl, *One-dimensional Markov Processes* (Springer, Berlin, 1968).
[43]A.V. Skorokhod, Limit theorems for Markov processes, Theory Prob. Appl. **3**, 202–246 (1958).
[44]D.W. Stroock, S.R.S. Varadhan, "Diffusion processes with continuous coefficients I, II," Commun. Pure Appl. Math. **22**, 345–400 (1969).

Absorption time by a random trap distribution

Herbert B. Rosenstock
U.S. Naval Research Laboratory, Washington, D.C. 20375

(Received 25 January 1980; accepted for publication 4 March 1980)

Consider a particle performing a random walk to nearest neighbors on a simple cubic lattice in any number of dimensions D. The lattice contains a fraction q of randomly located "trapping" sites which absorb the walker when stepped on. We calculate the mean time to trapping. This involves the expected number $V(t)$ of distinct sites that a walker would visit in t steps in absence of traps, a quantity known only asymptotically for large t; however, an exact calculation possible in one dimension suggests that the results are unexpectedly precise. The trapping time is proportional to q^{-1} for D = 3 or greater, and more complicated for D = 1,2.

1. INTRODUCTION

Consider a simple cubic (SC) lattice in any number D of dimensions and infinite in extent on which a particle is performing a random walk by stepping, with equal probability $\frac{1}{2}$D, on one of the nearest neighbors after a specified time interval. The lattice contains two kinds of sites: ordinary ones, and a small fraction q of "traps," randomly located and with the property of "trapping" the walker and presumably ending the walk. Physical applications that come to mind include energy transfer in luminescing organic solids,[1] cluster formation in silver halide grains,[2,3] as well as the classical gambler's ruin problem.[4]

The following formulas gives the mean time to absorption in a random walk among randomly located traps of density q, valid in any dimension:

$$\langle n\rangle = \sum_{t=1}^{\infty} t\,q\{V(t) - V(t-1)\}(1-q)^{V(t-1)}. \tag{1}$$

Here $V(t)$, usually called the "range" of walk, means the number of distinct sites that would be visited in t steps if there were no traps. In the above formula, the last term $(1-q)^{V(t-1)}$ is the probability of not stepping on a trap in the first $t-1$ steps; the central one is the probability $[V(t) - V(t-1)]$ that the t th step lead to a new site (since if it led to an old site, that site certainly would not be a trap), multiplied by the probability q that the new site indeed is a trap. Finally, the front term Σt sums the duration t over this probability distribution.

According to Ref. 5, Eq. (III.15), V is asymptotically given by

$$V(t) = \begin{cases} (8t/\pi)^{1/2}, & \text{1D}, \\ (\pi t/\log t), & \text{2D}, \\ (1-F)t, & \text{3D or higher}, \end{cases} \tag{2}$$

where [6–8] F, a numerical constant, is the probability of (eventually) returning to the origin in the absence of traps.

To evaluate Eq. (1), we first replace $V(t-1)$ in the exponent (only) by $V(t)$:

$$\langle n\rangle = \sum_{t=1} tq[V(t) - V(t-1)](1-q)^{V(t)} \tag{3}$$

and change the sum to an integral,

$$\langle n\rangle = q\int_0^{\infty} t\,dt\,\frac{dV}{dt}(1-q)^{V} \tag{4}$$

or

$$\langle n\rangle = q\int_0^{\infty} t\,dV e^{-qV}, \tag{5}$$

where $t = t(V)$ is now the solution of whichever of Eqs. (2) is appropriate to our dimensionality.

In evaluating this, we should remember that the formulas (2) are valid asymptotically, i.e., for long walks (large t). However, most walks are quite long. (For example, if traps are located periodically a distance $q^{-1/\mathrm{D}}$ apart, a walk starting at a site next to a trap will last $q^{-1} - 1$ steps; see Ref. 3, Appendix H.) So we expect the results to be valid for small q, perhaps below 0.1. More about this will be discussed later.

2. THREE OR MORE DIMENSIONS

This is the easiest case. The integral approximation (5) is not needed, for we can substitute the last of Eqs. (2) into Eq. (1) and do the sum exactly. The result is

$$\langle n\rangle = \frac{q(1-F)}{[1-(1-q)^{1-F}]^2} \tag{6}$$

or, for small q,

$$\langle n\rangle = 1/q(1-F). \tag{7}$$

On the attached Fig. 1 the two formulas are indistinguishable for $q < 0.1$. We have plotted it for the 3D SC lattice, for which $F = 0.340537$. In 4D, F would be[9] 0.20; in 5D, it would be 0.13. F is also known[9] for some 3D lattices that are not SC, and is numerically not very different. In Fig. 1, any such change would produce no more than a slight shift in the curve; its slope would not change.

3. TWO DIMENSIONS

This is the most difficult case, because of the complicated form of Eq. (2) in 2D. Before we can carry out the integration (5), we must express t in terms of V by solving the second of Eqs. (2), which we rewrite as

$$f(t) = 0, \tag{8}$$

where

 0022-2488/80/071643-03$1.00

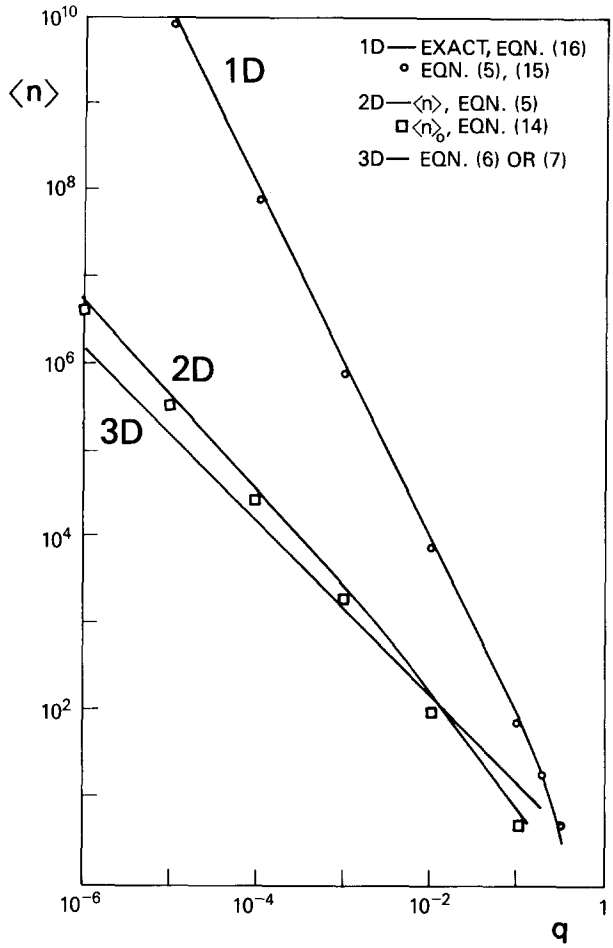


FIG. 1. Mean trapping times in one, two, and three dimensions. In each case, the "best" result is shown as a solid curve.

$$f(t) = t - u\log(t) \text{ and } u = V/\pi. \tag{9}$$

We use an iterative scheme called Newton's method, which involves guessing an initial solution t_0 and produces sucessive approximations $t_1, t_2, \cdots$ according to the formula

$$t_{i+1} = t_i - \frac{f(x_i)}{f'(x_i)}, \tag{10}$$

which with Eqs. (9) becomes

$$t_{i+1} = t_i u(\log t_i - 1)/(t_i - u). \tag{11}$$

We start with

$$t_0 = u\log u \tag{12}$$

and find that for all $V > 8$, three iterations yield accuracy to six significant figures. Equation (5) can now be evaluated numerically. If instead of the correct t we put our initial guess t_0 into Eq. (5) and write

$$\langle n \rangle_0 = q\int_0^\infty t_0 dV e^{-qV}, \tag{13}$$

we find that this can be evaluated analytically,

$$\langle n \rangle_0 = \frac{1}{\pi q}(-\log\pi q + 1 - C), \quad \text{with } C = 0.577216. \tag{14}$$

The "correct" $\langle n \rangle$ evaluated numerically and $\langle n \rangle_0$ given by Eq. (14) are both ploted in Fig. 1, the former as a curve, the latter as squares. The approximation (13) is very close.

4. ONE DIMENSION

Straightforward substitution of the first of Eqs. (2) into Eq. (5) gives

$$\langle n \rangle = \pi/4q^2. \tag{15}$$

However, this problem can also be solved exactly. [Remember that Eq. (15) is not exact because Eqs. (2) are not valid for short walks.] In 1D, only two traps can absorb any particular walker: the first one to his left and the first one to his right. We therefore have $\langle n \rangle = \Sigma_a D_a s_a(q)$, where D_a is the expected duration of a walk on a 1D lattice of a sites with absorbing boundaries, and $s_a(q)$ is the probability that, given an overall trap density q, a walker finds himself between two absorbing sites a distance a apart. This first of these quantities[10] is $(a-1)(a+1)/6$ and the second[11] $aq^2(1-q)^{a-1}$. Substitution and summation gives

$$\langle n \rangle = (1-q)/q^2, \tag{16}$$

which for small q is very close to $1/q^2$. So we see that our expression (15) differs from the exact one (16) by a factor $\pi/4 = 0.78$. On our plot, the two are barely distinguishable throughout the range shown. When q increases beyond 0.1, the asymptotic expression actually becomes better (surprisingly!), becomes precisely correct for q near 0.2146, and only then begins to deviate substantially.

The good agreement in 1D of the approximate expression (15) based on our dimension-independent approximation (1)–(5) with the exact value (16) suggests that the same should be true in higher dimensions. In fact, recent work suggests[3] that one should expect precision to be even better in higher dimensions. We have determined the mean length as well as the dispersion of walks for a model differing from the present one only in the details of the trap distribution[12] in both 1D and 2D. The result shows very little dispersion in 2D, and appreciably more dispersion in 1D. It is precisely this dispersion—the comparatively large number of walks in 1D that are much shorter than average—that causes the deviation between Eqs. (15) and (16). Since the present figure shows this discrepany to be quite small in 1D, we can expect the discrepancy to be even smaller in 2D.

Finally, we note the unexpected crossover between the 2D and 3D curve at $q \sim 0.015$. We can think of no obvious physical explanation for this.

5. SUMMARY

The mean time to absorption in a random walk among randomly located traps of density q is given by

$$\langle n \rangle = \begin{cases} (1-q)/q^2, & \text{in 1D} \\ \dfrac{1}{\pi q}(-\log\pi q + 1 - C) + f_2(q), & \text{in 2D} \\ 1/(1-F)q, & \text{in 3D,4D,5D,}\cdots. \end{cases} \tag{17}$$

Here the quantity $f_2(q)$ is given by

$$f_2(q) = q \int_0^\infty (t - t_0) dV e^{-qV},$$

with $t_0 = u \log u$, $u = V/\pi$, and t the solution of $u = t/\log t$; f_2 is small enough to be omitted in subsequent calculations. F is a constant, equal to 0.340537 for the 3D SC lattice, and C is Euler's constant, 0.577216···.

ACKNOWLEDGMENT

I would like to thank Dr. George H. Weiss for a very fruitful discussion.

[1]H. B. Rosenstock, Phys. Rev. **187**, 1166 (1969).
[2]C. L. Marquardt and G. Gliemeroth, J. Appl. Phys. **50**, 4584 (1979).
[3]H. B. Rosenstock (in print).
[4]The best general reference for random walk problems treated here is W. Feller, *Introduction to Probability Theory and Applications*, (Wiley, New York, 1957) Vol I, 2nd edition, Chap. 3, 9, and 14.
[5]E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).
[6]H. B. Rosenstock, J. Math. Phys. **11**, 487 (1970).
[7]The behavior of $V(t)$ is discussed in greater detail by N. C. Jain and W. E. Pruitt, *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability* (University of California, Berkeley, 1972), Vol. III, p. 31; and Ref. 8.
[8]G. H. Weiss (in print).
[9]E. W. Montroll, J. Soc. Ind. Appl. Math. **4**, 241 (1956).
[10]This can be obtained from Eq. (3.5) on p. 318 of Ref. 4 by averaging over a: $a^{-1} \sum_{z=0}^{a-1}$.
[11]H. B. Rosenstock, J. Soc. Ind. Appl. Math. **9**, 169 (1961), Appendix 2. Note that the distance between traps is called $b + 1$ there.
[12]In that model, the traps appear on the ordered sublattice rather than at random. See Ref. 3, Sec. III and Fig. 2.

Unidirectional wave propagation in one-dimensional first order Hamiltonian systems

E. W. C. van Groesen
Mathematical Institute, Catholic University, Toernooiveld, Nijmegen, The Netherlands

(Received 18 July 1979; accepted for publication 22 August 1979)

Defining the velocity of a conserved density to be the velocity of the center of gravity of this density, it is shown that for linear equations this velocity equals the weighted group velocity (with the density as weight function). For nonlinear equations, expressions for the centrovelocity of several conserved densities are derived. In particular, for a class of nonlocal equations, the centrovelocity of the energy density turns out to be some weighted average of the group velocity of the corresponding linearized equation. For a specific equation of this type, viz, the BBM equation, it is shown that upon restricting it to solutions whose initial form represents a long, low wave, the centrovelocity of the energy density is positive for all positive time.

1. INTRODUCTION

In this paper we shall consider equations of the form

$$\partial_t u = -\partial_x h'(u), \tag{1.1}$$

wherein u is a scalar function of one space variable $x \in Rl$ and the time $t \in Rl$. Equation (1.1) may be envisaged as an evolution equation in a function space U consisting of sufficiently smooth functions which tend, together with their derivatives, to zero for $|x| \to \infty$. Given a functional h on this space, $h'(u)$ denotes the functional derivative of h at the point u (i.e., the formal Gateaux derivative with respect to the L_2 inner product $(d/d\epsilon)h(u+\epsilon v)|_{\epsilon=0} = \langle h'(u), v\rangle$ for $u, v \in U$, where $\langle,\rangle$ is the L_2 inner product.

Equation (1.1) is called a first order Hamiltonian system, with the functional h as the Hamiltonian. Equations of this type often occur in mathematical physics. Probably the oldest example is the KdV equation

$$\partial_t u = -\partial_x (u + u^2 + u_{xx}), \tag{1.2}$$

which is of the form (1.1) with Hamiltonian

$$h(u) = \int \left\{ \frac{1}{2}u^2 + \frac{1}{3}u^3 - \frac{1}{2}u_x^2 \right\} dx \tag{1.3}$$

(integration is over the whole real line). Korteweg and de Vries[1] derived this equation in 1895 as an approximate description of the evolution of surface waves on a layer of fluid, under the assumption that the waves were mainly running in one direction. Due to this last assumption, they obtained an equation which is of first order in time, whereas Boussinesq had already obtained an equation of second order in time for the evolution of surface waves which admits solutions representing waves running in both directions. It turned out that this is just a particular instance of a much more general situation: Starting with a Hamiltonian system with canonically conjugate variables q and p, for which q enters as a potential function (in which form Boussinesq's equation can be rephrased), the restriction to solutions which mainly run in one direction leads, within some approximation, to an equation of the form (1.1) (cf. Whitham[2], Benjamin[3], and in particular Broer[4], Broer *et al.*[5], and van Groesen[6]).

Initiated by the pioneering work of Gardner *et al.* (cf. Ref. 7 for references) and Lax[8], the interest in the KdV and related equation(s) has considerably increased during the past ten years because it turned out that this equation is a completely integrable system (Zacharov and Faddeev[9]) which is soluble with the aid of inverse scattering theory.

In this paper we shall not deal with these aspects of (some) equations of the form (1.1). Our aim is to investigate what sense can be given to such statements as "Eq. (1.1) (approximately) describes unidirectionally propagating waves" and to derive conditions under which an equation of this type deserves such a qualification. Although several notions of propagativity are mentioned in the literature (such as phase, group, front, and signal velocity amongst many others), especially for nonlinear equations this point seems to have been overlooked somewhat.

Therefore, in Sec. 2 we shall define and comment on a notion of propagativity which can be used for any (nonlinear) system which has a conserved density and which is essentially nothing but the velocity of the center of gravity (centrovelocity) of this conserved density. Adopting this definition, it is shown in Sec. 3 that for *linear* equations of the form (1.1), the centrovelocity of any definite quadratic density is equal to the weighted group velocity, with the density as the weight function. This result shows that for linear equations the group velocity plays a fundamental role in the propagation of arbitrary wave forms (not necessarily "small wave packets" or domains of only "weak dispersion"), in contradistinction to a statement by Krotscheck and Kundt[10] who pose that "... the group velocity cannot be used for a rigorous discussion of propagation speeds: it has only approximate character, and loses its meaning in domains of strong dispersion".

In Sec. 4 we shall derive expressions for the centrovelocity of several conserved densities of Eq. (1.1). In Sec. 5 we shall specialize these results to a class of equations for which Eq. (1.1) is a local equation (i.e., h' is a local operator), and give sufficient conditions which assure that such an equation is unidirectionally propagative for all solutions. In Sec. 6 we consider nonlocal equations. In particular, we shall show that for a restricted class of such equations the centrovelocity of the conserved energy density [i.e., the density corre-

 0022-2488/80/071646-10$1.00

sponding to the invariant integral $h(u)$] is given by a weighted average of the group velocity of the corresponding linearized equation. An important example of a nonlocal equation of the form (1.1) is the BBM equation

$$(1-\partial_x^2)\partial_t u = -\partial_x(u+u^2), \tag{1.4}$$

which was proposed by Benjamin, Bona, and Mahony[11] as an alternative to the KdV equation to describe long, low waves.

In the final section we shall rigorously prove that the centrovelocity of the energy density of Eq. (1.4) is positive for *all* $t>0$ for any solution of Eq. (1.4) whose initial value belongs to a specified class of functions, where this class can be interpreted as the class of long, low wave forms.

2. PRELIMINARIES AND DEFINITIONS

In order to assure that with every evolution equation of the form

$$\partial_t u = -\partial_x h'(u) \tag{2.1}$$

there corresponds a unique Hamiltonian $h\in C^1(U,Rl)$, we require

$$h(0) = h'(0) = 0.$$

Moreover, we only consider translational invariant functionals h (i.e., h' does not depend explicitly on $x\in Rl$).

In the following we shall call Eq. (2.1) a local equation if h' is a local operator, i.e., if $h'(u)(x)$ depends on the function u and its derivatives at the place x only. Equation (2.1) is said to be nonlocal if h' is not a local operator. Clearly, the KdV equation (1.2) is a local equation. The BBM equation (1.4) is not of the form (2.1) but it can be brought to this form by a simple linear transformation.

More generally, we shall consider equations of the form

$$\partial_t Dv = -\partial_x k'(v) \quad [k(0)=k'(0)=0], \tag{2.2}$$

wherein D is some symmetric, regular operator on U, commuting with ∂_x. Transformation properties between Eqs. (2.1) and (2.2) can easily be described.

Proposition 2.1: (i) *Let L be a regular operator on U, commuting with ∂_x. Under the transformation $v:=L^{-1}u$, Eq.* (2.1) *transforms into Eq.* (2.2) *with*

$$D = L^*L \text{ and } k(v) := h(Lv).$$

(ii) *If the operator D admits the representation $D=L^*L$ for some regular operator L, then under the transformation $u:=Lv$, Eq.* (2.2) *transforms into Eq.* (2.1) *with*

$$h(u) = k(L^{-1}u).$$

Proof: If $k(v)=h(Lv)$, then on differentiating with respect to v there results

$$\langle k'(v),w\rangle = \langle h'(Lv),Lw\rangle, \quad \forall w\in U,$$

or

$$k'(v) = L^*h'(Lv).$$

Hence, if $u=Lv$,

$$L^*[\partial_t u + \partial_x h'(u)] = \partial_t L^*Lv + \partial_x k'(v),$$

which proves (i). Part (ii) is proved analogously. □

The BBM equation (1.4) is of the form (2.2) with

$$D = I-\partial_x^2, \quad k(v) = \int\left(\frac{1}{2}v^2+\frac{1}{3}v^3\right)dx. \tag{2.3}$$

As follows from proposition 2.1, it can be brought to the form (2.1) by the linear transformation $u:=D^{1/2}v$, where $D^{1/2}$ is the symmetric, positive square root of the operator D. Then it is seen that the BBM equation is a nonlocal equation.

However, for many purposes it is simpler to deal directly with equations in the form (2.2) than with the corresponding transformed form of the equations. When dealing with nonlocal operators, the following definition is not completely standard.

Definition 2.2: A functional $e\in C^1(U,Rl)$ is said to be an *invariant integral* for Eq. (2.1) if

$$\partial_t e(u) = 0$$

for every solution of Eq. (2.1). Any operator E on U for which

$$e(u) = \int E(u)dx$$

is an invariant integral will be called a *conserved density* for Eq. (2.1).

Remark 2.3: From this definition it follows that if E is a conserved density, then there exists a *flux density* T such that

$$\partial_t E(u) + \partial_x T(u) = 0,$$
$$T(u)\to 0 \quad \text{for} \quad |x|\to\infty \tag{2.4}$$

for every solution of Eq. (2.1). The expressions (2.4) are of the form of a local conservation law. However, only if T is a local operator do we obtain the usual result that for arbitrary interval $(a,b)\subset Rl$, $\partial_t\int_a^b E(u)$ depends only on the value of u and its derivatives with respect to x at the points $x=a$ and $x=b$.

Lemma 2.4: *The functionals $\int u\,dx$, h, and m, with*

$$m(u) := \int u^2 dx,$$

are invariant integrals for Eq. (2.1), *whereas the functionals $\int v\,dx$, k, and*

$$m(v) := \int v\cdot Dv dx$$

are invariant integrals for Eq. (2.2).

Proof: Integrating the equations over the whole real axis, the statement for the linear densities follows immediately. Furthermore, for solutions of Eq. (2.1) we have

$$\partial_t h(u) = \langle h'(u),\partial_t u\rangle = \langle h'(u), -\partial_x h'(u)\rangle = 0$$

and

$$\partial_t m(u) = 2\langle u,\partial_t u\rangle$$
$$= -2\langle u,\partial_x h'(u)\rangle = 2\langle h'(u),u_x\rangle = 0$$

(the last equality because h is a translational invariant functional). The corresponding results for Eq. (2.2) now follow because of proposition 2.1. □

Remark 2.5: If Eq. (2.1) is meant to describe a specific physical system (such as the water wave problem), invariant integrals often admit a physical interpretation (such as conservation of mass, energy, and momentum).

We now come to the main ideas concerning the propagation of conserved densities.

Definition 2.6: Let E be any conserved density of (2.1) with $e(u) = \int E(u)dx$ the corresponding invariant integral. For $u \in U$, the *center of gravity* of $E(u)$ will be denoted by $X_e(u)$ and is defined if $e(u) \neq 0$ by

$$\int [x - X_E(u)] \cdot E(u)dx = 0. \tag{2.5}$$

For a solution u of Eq. (2.1), the *centrovelocity* of E is defined to be the velocity of the center of gravity of $E(u)$:

$$V_E(u)(t) := \partial_t X_E[u(t)]. \tag{2.6}$$

The density E is said to be propagated to the right by a solution u of Eq. (2.1) at time t if $V_E(u)(t) > 0$, and Eq. (2.1) is said to be unidirectionally propagative (to the right) with respect to E if $V_E(u)(t) > 0$ for every nontrivial solution of Eq. (2.1) and all $t \in Rl$.

With respect to this definition some remarks have to be made.

Remark 2.7: If T is the flux density corresponding to the conserved density E [i.e., E and T satisfy Eq. (2.4) for every solution of Eq. (2.1)], the V_E is easily seen to be given by

$$V_E(u) = \frac{I}{e(u)} \cdot \int T(u)dx. \tag{2.7}$$

Remark 2.8: Although the centrovelocity of a conserved density as defined above has some physical significance, it is of course by no means the only possible way to describe propagation phenomena. However, some advantages of the proposed definition may be noticed: (i) it is not necessary to restrict to a special class of solutions (such as monochromatic solutions in linear equations); (ii) the definition is the same for nonlinear and for linear equations; (iii) it is possible (as will become clear in the next sections) to formulate, with relative ease, general conditions on the functional h (or on the operator D and the functional k) which assure that Eq. (2.1) [Eq. (2.2)] is unidirectionally propagative with respect to some density; (iv) if the equation admits a solution which travels undisturbed in shape with constant velocity c, say $u(x,t) = \varphi(x - ct)$, then $V_E(\varphi)(t) = c$ for all time and every conserved density E.

Remark 2.9: As a consequence of the proposed definition, with every conserved density there is associated a velocity for every solution. Suppose E_1 and E_2 are two different conserved densities [possibly with the same invariant functional (!); cf. Remark 2.10]. Then, if V_1 and V_2 denote the corresponding velocities, the velocity V_{12} of the conserved density $E_{12} = E_1 + E_2$ is given by

$$V_{12} = \frac{e_1 \cdot V_1 + e_2 \cdot V_2}{e_1 + e_2}$$

for every solution.

Furthermore, in general there is no evidence at all that if the density E_1 is being propagated to the right by some solution u, the same is true for the density E_2. However, for *linear* equations with constant coefficients it will be shown in the next section that if Eq. (2.1) is unidirectionally propagative with respect to some definite, quadratic conserved density, then the same is true for every quadratic density. For nonlinear equations no such strong relationship between the propagativities of different conserved densities has been found (nor can be expected to hold).

Remark 2.10: Closely related with the foregoing remark is the following observation: If E is a conserved density, with T the corresponding flux density, then E^*, defined by

$$E^*(u) := E(u) + \partial_x F(u),$$

where $F(u)$ is any expression in u satisfying $F(u) \to 0$ for $|x| \to \infty$ if $u \in U$, is also a conserved density with the same invariant integral

$$e(u) = \int E(u)dx = \int E^*(u)dx.$$

The flux density T^* corresponding to E^* is given by

$$T^*(u) = T(u) - \partial_t F(u),$$

and if X and X^* denote the centers of gravity of E and E^*, respectively, with corresponding velocities V and V^*, we have

$$X(u) - X^*(u) = \frac{1}{e(u)} \cdot \int F(u)dx$$

and

$$V(u) - V^*(u) = \frac{1}{e(u)} \cdot \int [T(u) - T^*(u)]dx.$$

From this it follows that, upon adding a term $\partial_x F$ to the density, the corresponding velocity will change in general: Only if the total flux is not altered will the velocity remain the same.

Remark 2.11: In general, the velocity functional V_E is not an invariant integral. However, if the conserved density E has a conserved flux, i.e., if the total flux itself if an invariant integral

$$\partial_t \int T(u)dx = 0,$$

then V_E as given by Eq. (2.7) is an invariant integral. In that case, the center of gravity is a linear function of t:

$$X(u)(t) = t \cdot V(u) + X_0(u), \tag{2.8}$$

where X_0 is an invariant integral (the position of the center of gravity at $t = 0$). Inserting Eq. (2.8) into (2.5) gives

$$\int [xE(u) - tT(u)]dx = X_0(u) \cdot \int E(u)dx,$$

which leads to the following invariant functional which contains the x and t variable explicitly:

$$\partial_t \int [xE(u) - tT(u)]dx = 0. \tag{2.9}$$

3. PROPAGATION IN LINEAR SYSTEMS

Linear first order Hamiltonian systems are described by an equation of the form

$$\partial_t u = -\partial_x L u, \tag{3.1}$$

where L is some symmetric operator. The Hamiltonian for such equations is the quadratic functional

$$h(u) = \tfrac{1}{2}\langle u, Lu\rangle. \tag{3.2}$$

In the following we shall restrict ourselves to the simplest class of operators, viz., the class of pseudodifferential operators (with constant coefficients). If M is a pseudodifferential operator, we shall denote the symbol of M by $\hat{M}$ as follows:

$$\widehat{Mu}(k) = \hat{M}(k)\cdot\hat{u}(k),$$

wherein $\hat{u}$ denotes the Fourier transform of the function u.

Theorem 3.1: *Any linear density $E(u) = Pu$, where P is a pseudodifferential operator, is conserved. The center of gravity and its velocity can be defined for solutions for which $\int Pu\,dx \neq 0$ and we have*

$$V_E(u) = \hat{L}(0). \tag{3.3}$$

Hence, all linear densities are being propagated with the same constant speed $\hat{L}(0)$, independent of the particular solution.

Proof: Since P commutes with ∂_x, we have

$$\partial_t Pu + \partial_x PLu = 0.$$

From this it follows with Eq. (2.7) and Fourier transformation that

$$V_E(u)(t) = \frac{\int PLu\,dx}{\int Pu\,dx} = \frac{\hat{P}(0)\cdot\hat{L}(0)\cdot\hat{u}(0)}{\hat{P}(0)\cdot\hat{u}(0)} = \hat{L}(0). \quad \square$$

Quadratic conserved densities are more interesting and have been studied in great detail. It is at this point that the concept of group velocity enters the discussion of propagation. The dispersion relation for Eq. (3.1) is

$$\omega = k\cdot\hat{L}(k), \tag{3.4}$$

and corresponding to the group velocity $d\omega/dk$, we define an operator G.

Definition 3.2: The *group-velocity operator* G is defined to be the pseudodifferential operator with symbol

$$\hat{G}(k) := \frac{d\omega}{dk}(k) = \frac{d}{dk}[k\cdot\hat{L}(k)]. \tag{3.5}$$

Lemma 3.3: *As L is symmetric, the operator G is symmetric.*

Proof: The proof is immediate from the fact that a pseudodifferential operator M is symmetric if and only if its symbol satisfies $\hat{M}(k) \in Rl$ for $k \in Rl$. $\square$

Lemma 3.4: *Any quadratic density $E(u) = Pu\cdot Qu$, where P and Q are pseudodifferential operators, is a conserved density for Eq.* (3.1).

Proof: Generally, $E(u) = Pu\cdot Qu$ is a conserved density of Eq. (3.1) if P and Q satisfy

$$P^*Q\partial_x L = L\partial_x P^*Q.$$

Clearly, for pseudodifferential operators P and Q, this condition is satisfied. $\square$

As a typical result concerning the relation between group velocity and the propagation of conserved densities by *monochromatic* solution of Eq. (3.1), we quote the following result:

Theorem 3.5: *Let E be a quadratic conserved density, with corresponding flux density T. Then E is being propagated with the group velocity in the following sense: For monochromatic solutions $\phi(x,t) = \phi_0 \exp i(k_0 x - \omega_0 t)$, where ϕ_0 is a constant and $\omega_0 = \omega(k_0)$, the following relation holds:*

$$\frac{T(\phi)}{E(\phi)} = \frac{d\omega}{dk}(k_0). \tag{3.6}$$

This theorem is well known and can be found, for example, in de Graaf and Broer[12]. With the proposed definition 2.6, it turns out that the group velocity plays an equally important role in the propagation of quadratic densities by nonperiodic solutions.

Theorem 3.6: *Consider a definite (conserved) quadratic density $E(u) = Au\cdot Au$, where A is some pseudodifferential operator. Then, the centrovelocity of this density is (an invariant integral and is) given by*

$$V_E(u) = \langle Au, Au\rangle^{-1}\cdot\langle Au, GAu\rangle, \tag{3.7}$$

wherein G is the group-velocity operator. Hence, $V_E(u)$ equals the weighted group velocity, with weight function $|\widehat{Au}|^2$.

Proof: Inserting Eq. (3.1) directly into definition (2.6), it follows that

$$V_E(u) = 2\cdot\langle\widehat{Au}, \widehat{Au}\rangle^{-1}\cdot\langle\widehat{Au}, -x\partial_x L\widehat{Au}\rangle.$$

With Parseval's theorem and Eq. (3.4) we find

$$V_E(u) = 2\cdot\langle\widehat{Au}, \widehat{Au}\rangle^{-1}\cdot\langle\widehat{Au}, \partial_k[\omega\cdot\widehat{Au}]\rangle,$$

and after some straightforward manipulations

$$\begin{aligned} V_E(u) &= \langle\widehat{Au}, \widehat{Au}\rangle^{-1}\cdot\left\langle\widehat{Au}, \frac{d\omega}{dk}\cdot\widehat{Au}\right\rangle \\ &= \langle Au, Au\rangle^{-1}\cdot\langle Au, GAu\rangle. \end{aligned} \tag{3.8}$$

This proves Eq. (3.7), and as V_E is a quadratic functional, it is an invariant integral according to Lemma 3.4. $\square$

An immediate consequence of this theorem is the following:

Corrolary 3.7: *Equation* (3.1) *is unidirectionally propagative (to the right) with respect to any definite, quadratic density if and only if the group velocity is nonnegative for all wave numbers*:

$$\hat{G}(k) \geqslant 0, \quad \textit{for all } k \in Rl. \tag{3.9}$$

Remark 3.8: For more general quadratic densities of the form $E(u) = Au\cdot Bu$, with A and B pseudodifferential operators, the corresponding centrovelocity is easily found to be the invariant integral

$$V_E(u) = \langle Au, Bu\rangle^{-1}\cdot\langle Au, GBu\rangle \tag{3.10}$$

for solutions for which $e(u) = \langle Au, Bu\rangle \neq 0$.

Remark 3.9: For two specific densities, viz., the momentum density [$A = I$ in Eq. (3.7)] and the energy density [$A^2 = L$ in Eq. (3.7)], the result (3.7) can already be found in Wehausen and Laitone (Ref. 13, Sec. 15).

Our more general formulation makes their results somewhat more transparent.

4. PROPAGATION IN NONLINEAR SYSTEMS

For the first order Hamiltonian equations under consideration

$$\partial_t u = -\partial_x h'(u), \tag{4.1}$$

the Hamiltonian h can be written as

$$h(u) = \int H(u)\,dx, \tag{4.2}$$

where H is some (nonlinear) operator on U. Only differentia-

ble operators H will be considered, i.e., operators H for which, for every $u \in U$, there exists a linear operator, depending on u, and denoted by $H'(u)$, the derivative of H at the point u, such that

$$\frac{d}{d\epsilon} H(u+\epsilon v)|_{\epsilon=0} = H'(u)v,$$

for every function $v \in U$. As we also have

$$\frac{d}{d\epsilon} h(u+\epsilon v)|_{\epsilon=0} = \int h'(u)\cdot v dx,$$

it follows that

$$\int h'(u)\cdot v dx = \int H'(u) v dx$$

for every $v \in U$. Hence, there exists an operator B such that

$$h'(u)\cdot v = H'(u)v - \partial_x B(u,v). \tag{4.3}$$

Note that $B(u,v)$ is linear in v, nonlinear in u (except when H is quadratic), and that $B(u,v) \to 0$ for $|x| \to \infty$ on the considered class of functions. As h is assumed to be a translational invariant functional, it follows that

$$h'(u)\cdot u_x = \partial_x [H(u) - B(u,u_x)] \tag{4.4}$$

and, as h does not depend explicitly on t, we also have

$$\partial_t H(u) = h'(u)\cdot u_t + \partial_x B(u,u_t). \tag{4.5}$$

With these results it is not difficult to find the local conservation laws for the conserved densities of Eq. (4.1). In fact, as it stands, Eq. (4.1) is of the form of a local conservation law for the density u. For the conserved density u^2, the corresponding flux density follows with the aid of Eq. (4.4):

$$\begin{aligned}\partial_t u^2 &= 2u\cdot u_t = -2u\cdot\partial_x h'(u) \\ &= -2\partial_x [u\cdot h'(u) - H(u) + B(u,u_x)].\end{aligned}$$

In the same way, using Eq. (4.5), we find the following for the conserved density $H(u)$:

$$\begin{aligned}\partial_t H(u) &= -h'(u)\cdot\partial_x h'(u) - \partial_x B(u,\partial_x h'(u)) \\ &= -\partial_x \{\tfrac{1}{2}[h'(u)]^2 + B(u,\partial_x h'(u))\}.\end{aligned}$$

Having found the flux densities corresponding to the conserved densities of Eq. (4.1), the centrovelocities of these densities follow immediately with Eq. (2.7) and we have obtained theorem 4.1.

Theorem 4.1: *For Eq.* (4.1) *we have the following*: (i) *The centrovelocity of the conserved density u is given by*

$$V(u) = \left(\int u dx\right)^{-1} \cdot \int h'(u) dx. \tag{4.6}$$

(ii) *The centrovelocity of the conserved quadratic density* u^2 *is given by*

$$V(u) = 2\cdot\|u\|^{-2}\cdot\int \{u\cdot h'(u) - H(u) + B(u,u_x)\} dx. \tag{4.7}$$

(iii) *The centrovelocity of the (energy) density* $H(u)$ *is given by*

$$V(u) = h(u)^{-1}\cdot\int \{\tfrac{1}{2}[h'(u)]^2 + B(u,\partial_x h'(u))\} dx. \tag{4.8}$$

With these general results it is possible, at least in principle, to investigate whether one of the conserved densities is being propagated to the right by a particular solution. However, the possibility to find conditions on the functional h which assure that some of these velocity functionals are positive depends very much on the "nature" of the operator h'. In fact, for local equations it is possible to investigate this matter completely. By way of example we shall consider a class of local equations in the next section. This class is relatively simple but has nonetheless all the peculiarities of more general local equations. In contradistinction, even the most simple nonlocal equations are difficult to deal with. This is due to the mathematical difficulties encountered in the investigation of the positivity of functionals with nonquadratic nonlocal integrands. Nevertheless, for a special class of nonlocal equations we shall derive some remarkable results concerning the propagation of the energy density in Secs. 6 and 7.

5. A CLASS OF LOCAL EQUATIONS

We consider local equations of the form (4.1) for which the Hamiltonian $h(u)$ is given by

$$h(u) = \int \{N(u) + S(u_x)\} dx, \tag{5.1}$$

where N and S are smooth (C^2) functions of their arguments with derivatives n and s, respectively:

$$n(y) := N'(y) = \frac{dN}{dy}(y), \quad y \in Rl,$$

$$s(z) := S'(z) = \frac{dS}{dz}(z), \quad z \in Rl$$

(primes denote differentiations with respect to the arguments). We assume that

$$N(0) = S(0) = n(0) = s(0) = 0,$$

$0 = N(0) = S(0) = n(0)$ to assure that $h(0) = h'(0) = 0$ and $s(0) = 0$ is no restriction. Equation (4.1) with h as in Eq. (5.1) then reads

$$\partial_t u = -\partial_x \{n(u) - \partial_x s(u_x)\}. \tag{5.2}$$

In a straightforward way, the following results are obtained from theorem 4.1.

Theorem 5.1: *The centrovelocities of the conserved densities*

$$u,\ u\cdot u,\ N(u) + S(u_x) \tag{5.3}$$

for Eq. (5.2) *are given by*

$$V(u) = \left(\int u dx\right)^{-1} \cdot \int n(u) dx, \tag{5.4}$$

$$\begin{aligned}V(u) = 2\cdot\|u\|^{-2}\cdot\int \{&u\cdot n(u) - N(u) + 2u_x\cdot s(u_x) \\ &- S(u_x)\} dx,\end{aligned} \tag{5.5}$$

$$\begin{aligned}V(u) = h(u)^{-1}\cdot\int \Big\{&\frac{1}{2} n^2(u) + 2u_x\cdot s(u_x)\cdot n'(u) \\ &+ \frac{3}{2}[\partial_x s(u_x)]^2\Big\} dx.\end{aligned} \tag{5.6}$$

To assure that Eq. (5.2) is unidirectionally propagative with respect to the densities u^2 and/or $N(u) + S(u_x)$ it suffices to

state conditions for the functions N and S which assure that the velocity functionals (5.5) and/or (5.6) are nonnegative on the considered class of functions U. In that way we get Theorem 5.2.

Theorem 5.2: *Equation* (5.1) *is unidirectionally propagative to the right with respect to* (i) *the density* u^2 *if*

$$y\cdot n(y) - N(y) \geqslant 0, \quad \forall y \in Rl,$$

$$2z\cdot s(z) - S(z) \geqslant 0, \quad \forall z \in Rl;$$

(ii) *the (energy) density* $N(u) + S(u_x)$ *(required to be positive) if*

$$N(y) \geqslant 0, \; n'(y) \geqslant 0, \; \forall y \in Rl,$$

$$S(z) \geqslant 0, \; 2z\cdot s(z) \geqslant 0, \; \forall z \in Rl;$$

(iii) *both the density* u^2 *and the energy density if*

$$N(y) \geqslant 0, \; y\cdot n(y) - N(y) \geqslant 0, \; n'(y) \geqslant 0 \quad \forall y \in Rl,$$

$$S(z) \geqslant 0, \; 2z\cdot s(z) - S(z) \geqslant 0, \quad \forall z \in Rl.$$

Remark 5.3: Linearizing Eq. (5.2) gives

$$\partial_t u = -\partial_x \{n'(0)\cdot u - s'(0)\cdot \partial_x^2 u\}, \tag{5.7}$$

which has the dispersion relation

$$\omega(k) = n'(0)\cdot k + s'(0)\cdot k^3$$

and group velocity

$$\frac{d\omega}{dk}(k) = n'(0) + 3s'(0)\cdot k^2.$$

The velocity functionals of the density u^2 and the linearized energy density $\frac{1}{2}n'(0)\cdot u^2 + \frac{1}{2}s'(0)\cdot u_x^2$ of this linear equation (5.7) [as given by Eq. (3.7)] are easily seen to be the quadratic terms in a Taylor expansion of the integrands of Eqs. (5.5) and (5.6), respectively.

Remark 5.4: As u and u^2 are conserved densities for Eq. (5.2), the velocity (5.4) is an invariant functional if

$$n(u) = \alpha u + \beta u^2, \quad \alpha,\beta \in Rl.$$

In that case, u is a conserved density with conserved flux and according to remark 2.11, there exists an invariant functional which depends on x and t explicitly; in this case

$$\partial_t \int \{xu - tn(u)\}dx = 0. \tag{5.8}$$

Remark 5.5: The KdV equation (1.2) belongs to the considered class of equations with

$$N(y) = \tfrac{1}{2}y^2 + \tfrac{1}{6}y^3, \quad S(z) = \tfrac{1}{2}z^2.$$

This equation is neither unidirectionally propagative with respect to the density u^2 nor with respect to its energy density, as may be confirmed from the expressions (5.5) and (5.6). The velocity functional (5.4) is an invariant functional, and Eq. (5.8) reads

$$\partial_t \int \{xu - t(u + \tfrac{1}{2}u^2)\}dx = 0. \tag{5.9}$$

Apart from a trivial transformation, this functional was already given by Miura *et al*[14].

6. A CLASS OF NONLOCAL EQUATIONS

Here we shall examine equations of the form

$$\partial_t Du = -\partial_x e'(u), \tag{6.1}$$

where the functional e is given by

$$e(u) = \int N(u)dx, \tag{6.2}$$

with N a smooth (C^2) function of its argument, $n(u)$: $= (dN/du)(u)$, $N(0) = n(0) = 0$, and where $D = A^2$ with A some positive symmetric pseudodifferential operator. From Sec. 2 it follows that, via a simple linear transformation, Eq. (6.1) can be brought into an equation of the form (4.1), but except when A^{-1} is an ordinary differential operator, this equation will be of nonlocal type. In these cases it is somewhat simpler to deal directly with the form (6.1).

For Eq. (6.1) we have three invariant integrals:

$$\int u dx, \quad \int \langle u, Du\rangle, \quad \int e(u)$$

(cf. lemma 2.4). The velocity of the linear density u is again given by Eq. (5.2) and remark 5.4 applies as well: If $\int n(u)dx$ is an invariant functional, then

$$\partial_t \int \{xDu - tn(u)\}dx = 0.$$

For the following we define the symmetric operator G as the pseudodifferential operator with symbol $\hat{G}$, where

$$\hat{G}(k) = \partial_k [k\cdot \hat{D}^{-1}(k)]$$
$$= \hat{D}^{-1}(k)\cdot[1 - k\cdot \hat{D}^{-1}(k)\cdot \partial_k \hat{D}(k)]. \tag{6.3}$$

Remark 6.1: Note that if u has a linear term, say $n'(0) = 1$, then $\omega(k) := k\cdot \hat{D}^{-1}(k)$ is the dispersion relation of the linearized equation (6.1), and then $\hat{G}(k)$ is the corresponding group velocity. However, if Eq.(6.1) does not admit a formal linearization, i.e., if $n' = (0) = 0$, this interpretation of ω and $\hat{G}$ no longer makes sense, but the results to be derived remain valid!

Theorem 6.2: *The centrovelocity of the energy density* $N(u)$ *is given by*

$$V(u) = \tfrac{1}{2}e(u)^{-1}\cdot\langle n(u), Gn(u)\rangle. \tag{6.4}$$

Consequently, Eqs. (6.1) *and* (6.2) *are unidirectionally propagative to the right with respect to the energy density (assumed to be positive) if and only if*

$$\hat{G}(k) \geqslant 0, \quad \forall k \in Rl. \tag{6.5}$$

Proof: The simplest way to derive this result is analogous to the proof of Theorem 3.6: Using Fourier-transform techniques and writing $\hat{n}$ for the Fourier transform of the expression $n(u)$, we find

$$\partial_t \int xN(u)dx = \langle x\cdot n(u), \partial_t u\rangle$$
$$= \langle n(u), -x\partial_x D^{-1} n(u)\rangle$$
$$= \tfrac{1}{2}\langle \hat{n}(k), \hat{G}(k)\cdot \hat{n}(k)\rangle.$$

Hence,

$$V(u) = \tfrac{1}{2}e(u)^{-1}\cdot\langle \hat{n}, \hat{G}\hat{n}\rangle$$
$$= \tfrac{1}{2}e(u)^{-1}\cdot\langle n(u), Gn(u)\rangle$$

and the theorem follows. □

Remark 6.3: The BBM equation (1.4) belongs to the considered class of equations with

$$D = 1 - \partial_x^2, \; N(u) = \tfrac{1}{2}u^2 + \tfrac{1}{6}u^3. \tag{6.6}$$

For this operator D, the function $\hat{G}$ is not positive for all $k \in Rl$:

$$\hat{G}(k) = (1 + k^2)^{-2} \cdot (1 - k^2). \tag{6.7}$$

Hence, the BBM equation is not unidirectionally propagative with respect to its (nondefinite) energy density. (See however the results of the next section for a restricted set of solutions.)

Remark 6.4: More generally, if D can be written as

$$\hat{D} = (1 + k^2)^{\sigma}, \quad \sigma \in Rl, \tag{6.8}$$

then $\hat{G}$ is given by

$$\hat{G}(k) = (1 + k^2)^{-\sigma - 1} \cdot [1 + (1 - 2\sigma) \cdot k^2]. \tag{6.9}$$

Hence, for such operators, condition (6.5) is satisfied if and only if

$$\sigma \leqslant \tfrac{1}{2}. \tag{6.10}$$

The energy velocity as given by Eq. (6.4) is remarkably simple. However, matters are much more complicated for the velocity of the (positive) quadratic density $Au \cdot Au$. Using Fourier-transform techniques in intermediate steps, it is not difficult to derive the following result in a direct way.

Theorem 6.5: *The centrovelocity of the positive density* $Au \cdot Au$ *can be expressed with the operator* G *as*

$$V(u) = \langle Au, Au \rangle^{-1} \cdot \int \{u \cdot n(u) - 2N(u) + n(u) \cdot DGu\} dx. \tag{6.11}$$

Note that for linear equations $N(u) = \frac{1}{2}u^2$, Eq. (6.11) agrees with Eq. (3.7). However, in the more interesting case on nonlinear equations it seems to be impossible to derive conditions on N and D such that Eq. (6.11) is a positive functional.

7. UNIDIRECTIONAL PROPAGATION IN LONG, LOW WAVE MODELS

In this section we shall once again examine equations of the form

$$\partial_t Du = -\partial_x n(u), \tag{7.1}$$

where D is a pseudodifferential operator and $n(u)$ a smooth function of its argument. The energy density $N(u)$, with $n(u) = (dN/du)(u)$, $N(0) = n(0) = 0$, is no longer required to be positive. In view of the results of the foregoing section we shall only consider the centrovelocity of the energy density. This velocity is given by Eq. (6.4):

$$V(u) = \tfrac{1}{2} \cdot e(u)^{-1} \cdot \langle n(u), Gn(u) \rangle, \tag{7.2}$$

where G is the pseudodifferential operator with symbol given by Eq. (6.3). We shall suppose that

$$n'(0) = 1, \tag{7.3}$$

such that G can be interpreted as the group-velocity operator of the linearized problem. In the foregoing section it was shown that the BBM equation

$$(1 - \partial_x^2)\partial_t u = -\partial_x (u + \tfrac{1}{2}u^2) \tag{7.4}$$

is not unidirectionally propagative with respect to the energy density. However, the BBM equation (as many other equations of this type) is derived as an approximate equation for the description of "fairly long, fairly low" waves (cf. Broer,[4] Broer *et al.*,[5] and Benjamin *et al.*[11] for details about this approximate character of the equation).

Therefore, it is reasonable to investigate the positivity of the functional (7.2) on the restricted class of functions which can be described as long, low waves. To make this idea more concrete, let us suppose that we can define two functionals ϵ and λ whose values $\epsilon(u)$ and $\lambda(u)$ are a measure of the height and of the "length" of the function u, respectively. Then the class of long, low waves can be described as the set of functions satisfying

$$\epsilon(u) < \epsilon_0, \quad \lambda(u) > \lambda_0, \tag{7.5}$$

where ϵ_0 and λ_0^{-1} are small positive numbers. Now suppose that numbers ϵ_0 and λ_0 can be found such that $V_E(u)$ is of the same sign (positive say) for every function u which satisfies Eqs. (7.5). Then, if u is a solution of Eq. (7.1) which satisfies Eqs. (7.5) at some instant t_0, $V_E(u)(t)$ will be positive at $t = t_0$ and for times $t > t_0$ as long as $u(t)$ satisfies conditions (7.5). *A priori*, it is by no means clear that solutions corresponding to initial data which satisfy Eqs. (7.5) satisfy this condition for all $t > 0$.

Especially for nonlinear equations this is a critical point. To demonstrate this for the long wavelength condition for instance, consider the solution of Eq. (7.1) corresponding to an initial value $g(x)$ whose Fourier transform $\hat{g}$ satisfies $\hat{g}(k) = 0$ for $|k| \geqslant k_0$, $k \in Rl$ (i.e., g consists of long wave components only). A Fourier transformation of Eq. (7.1) shows that if the equation is *linear*, then $\hat{u}(k,t) = 0$ for $|k| \geqslant k_0$ for all $t \geqslant 0$, but if the equation is *nonlinear*, then $\hat{u}(k,t) \neq 0$ for almost all $k \in Rl$, no matter how small $t > 0$: *Initial long wave components generate short wave components instantly*. From these remarks and observations the following definition will be acceptable.

Definition 7.1: Let there be given two positive functionals ϵ and λ [for which $\epsilon(u)$ and $\lambda(u)$ are a measure for the height and the length of a function $u \in U$, respectively]. Then, Eq. (7.1) is said to be unidirectionally propagative with respect to the energy density $N(u)$ for (λ) long and (ϵ) low waves if positive numbers ϵ_0 and λ_0 can be found such that $V(u)(t) \geqslant 0$ for all $t > 0$ for every solution u whose initial value u_0 satisfies

$$\epsilon(u_0) < \epsilon_0, \quad \lambda(u_0) > \lambda_0. \tag{7.6}$$

For the following we suppose that the symbol of the group-velocity operator G can be estimated as

$$\hat{G}(k) \geqslant \hat{G}(0) \cdot [1 - \ell^2 k^2], \quad k \in Rl, \tag{7.7}$$

where $\hat{G}(0)$ and ℓ are positive numbers. For long wave models such an estimate is generally possible: The long wave components propagate with the largest, positive speed (the group velocity has a positive maximum at $k = 0$). With Eq. (7.7) the velocity functional (7.2) can be estimated as

$$V(u) \geqslant \tfrac{1}{2}\hat{G}(0) \cdot \frac{\|n(u)\|^2}{e(u)} \cdot \left\{1 - \ell^2 \cdot \frac{\|\partial_x n(u)\|^2}{\|n(u)\|^2}\right\}. \tag{7.8}$$

From this it immediately follows that $V(u) \geqslant 0$ if

$$\lambda(u) \geqslant \ell, \tag{7.9}$$

if the functional λ is defined by

$$\lambda(u)^{-1} := \frac{\|\partial_x n(u)\|}{\|n(u)\|} \tag{7.10}$$

Remark 7.2: The functional λ defined by Eq. (7.10) can indeed be interpreted as an averaged wave length: $\lambda(u)^{-2}$ is the weighted average of k^2 with weight function $|\widehat{n(u)}|^2$. Another way to interpret $\lambda(u)$ as a measure of the "length" of the function u follows from the observation

$$\lambda(u_\mu)^{-2} = \mu^2 \cdot \lambda(v)^2, \quad \text{for } u_\mu(x) := v(\mu x); \tag{7.11}$$

hence, $\lambda(u_\mu) \to \infty$ for $\mu \to 0$.

In the following we shall show that it is sometimes possible to find conditions of the form (7.6), i.e., conditions imposed on the initial data only, which assure that the resulting solutions satisfy condition (7.9) for *all* $t \geqslant 0$. For simplicity we shall restrict ourselves in the first instance to a specific equation, viz., the BBM equation (7.4). Note that this equation satisfies Eq. (7.7) with

$$\hat{G}(0) = 1, \quad \ell^2 = 3. \tag{7.12}$$

For what follows we have to recall that $H^1(Rl)$ denotes the first Sobolev space of functions $u \in L_2(Rl)$ which have (generalized) derivatives $u_x \in L_2(Rl)$. Supplied with the norm $\| \ \|_1$ it is a Banach space which is continuously embedded in $C^0(Rl)$:

$$|u|_\infty^2 \leqslant \tfrac{1}{2}\|u\|_1^2, \quad u \in H^1(Rl), \tag{7.13}$$

where

$$|u|_\infty := \sup_{x \in Rl} |u(x)|,$$
$$\|u\|_1^2 := \|u\|^2 + \|u_x\|^2.$$

Concerning the existence of a classical solution of the initial value problem for the BBM equation, we quote the following result:

Lemma 7.3: *Let* $u_0 \in C^2(Rl) \cap H^1(Rl)$. *Then there exists a unique (classical) solution* u *of Eq.* (7.4) *with*

$$u(x,0) = u_0(x)$$

and

$$u(\cdot,t), \partial_t u(\cdot,t) \in C^2(Rl) \cap H^1(Rl), \quad \text{for all } t \geqslant 0.$$

Consequently, the (momentum) functional

$$m(u) := \tfrac{1}{2}\langle u, Du \rangle$$

and the energy functional $e(u)$ *are neatly defined and are invariant integrals*:

$$m(u(t)) = m(u_0), \quad e(u(t)) = e(u_0), \quad \forall\, t \geqslant 0.$$

Proof: The proof of this result can be found in Benjamin *et al.*[11] □

We are now in a position to formulate the main result.

Theorem 7.4: *The BBM equation* (7.4) *is unidirectionally propagative to the right with respect to the energy density for the class of long, low waves, which is characterized as the solutions whose initial value* u_0 *satisfy*

$$u_0 \in C^2(Rl) \cap H^1(Rl),$$
$$\lambda(u_0) > \lambda_0,$$
$$\epsilon(u_0) < \epsilon_0, \tag{7.14}$$

for sufficiently small positive numbers ϵ_0 *and* λ_0^{-1}. *Here,* λ *is the functional defined by Eq.* (7.10) *and*

$$\epsilon(u) := \|u\|_1.$$

Proof: In view of the estimate (7.8) and result (7.12) we have to show that λ_0 and ϵ_0 can be found such that

$$\lambda(u)^{-2} \leqslant \tfrac{1}{3} \tag{7.15}$$

for every $t \geqslant 0$ and every solution with initial data satisfying Eq. (7.14). Let u_0 denote the initial value and define $\delta > 0$ by

$$\|u_0\|_1^2 = \tfrac{1}{2}\delta^2.$$

As

$$m(u) = \tfrac{1}{2}\langle u, (1 - \partial_x^2)u \rangle = \tfrac{1}{2}\|u\|_1^2 \tag{7.16}$$

is an invariant functional, it follows that

$$\|u\|_1^2 = \tfrac{1}{2}\delta^2, \quad \forall t \geqslant 0$$

and with Eq. (7.13), that

$$|u|_\infty < \delta, \quad \forall t \geqslant 0. \tag{7.17}$$

Then we can derive the following useful estimates for the functionals e and λ:

$$\tfrac{1}{2}(1 - \tfrac{1}{3}\delta) \cdot \|u\|^2 \leqslant e(u) \leqslant \tfrac{1}{2}(1 + \tfrac{1}{3}\delta) \cdot \|u\|^2, \quad \forall t \geqslant 0 \tag{7.18}$$

and, provided $\delta < 2$,

$$\left(\frac{1-\delta}{1+\tfrac{1}{2}\delta}\right)^2 \cdot \left(\frac{\|u\|_1^2}{\|u\|^2} - 1\right) \leqslant \lambda(u)^{-2} \leqslant \left(\frac{1+\delta}{1-\tfrac{1}{2}\delta}\right)^2 \cdot \left(\frac{\|u\|_1^2}{\|u\|^2} - 1\right), \quad \forall t \geqslant 0. \tag{7.19}$$

As e is an invariant functional, it follows from Eq. (7.18) that

$$(1 - \tfrac{1}{3}\delta) \cdot \|u_0\|^2 \leqslant (1 + \tfrac{1}{3}\delta) \cdot \|u\|^2, \quad \forall t \geqslant 0. \tag{7.20}$$

Writing $\lambda_0 = \lambda(u(x,0))$, it follows from Eq. (7.19) that

$$\left(\frac{1-\delta}{1+\tfrac{1}{2}\delta}\right)^2 \cdot \left(\frac{\|u_0\|_1^2}{\|u_0\|^2} - 1\right) \leqslant \lambda_0^{-2},$$

and, as $\|u\|_1$ is invariant, we obtain, provided $\delta < 1$,

$$\frac{\|u\|_1^2}{\|u_0\|_2} \leqslant 1 + \lambda_0^{-2} \cdot \left(\frac{1 + \tfrac{1}{2}\delta^2}{1 - \delta}\right), \quad \forall t \geqslant 0. \tag{7.21}$$

With Eqs. (7.20) and (7.21) we can majorize the right hand side of Eq. (7.19) and obtain

$$\lambda(u)^{-2} \leqslant \left(\frac{1+\delta}{1-\tfrac{1}{2}\delta}\right)^2 \cdot \left\{ \left(\frac{1+\tfrac{1}{3}\delta}{1-\tfrac{1}{3}\delta}\right) \cdot \left[1 + \lambda_0^{-2} \cdot \left(\frac{1+\tfrac{1}{2}\delta}{1-\tfrac{1}{2}\delta}\right)^2\right] - 1 \right\}, \quad \forall t \geqslant 0. \tag{7.22}$$

This result shows that $\lambda(u)^{-2}$ can be majorized uniformly with respect to t in terms of initial value δ and λ_0. Moreover, it is easily seen that the right hand side of Eq. (7.22) can be bounded by $\frac{1}{3}$ if δ and λ_0^{-1} are taken sufficiently small. This shows that condition (7.15) is satisfied for δ (and hence ϵ_0) and λ_0 sufficiently small. With the extra observation that $e(u)$ is positive if $\delta < 3$, as follows from Eq. (7.18), this proves the theorem. □

Remark 7.5: From a physical point of view the foregoing theorem is satisfactory because the requirements define the functions to be low waves, as follows from the estimate

(7.13), and to be long waves in the sense of remark 7.2. However, it is possible to show that the velocity functional (7.2) is positive on a larger class of functions. Therefore, define the functional Λ by

$$\Lambda(u) := \frac{m(u)}{e(u)}. \tag{7.23}$$

Then it can be shown that

$$V(u) > 0, \quad \text{for every } u \in S_\gamma, \tag{7.24}$$

where S_γ is the set of functions for which

$$\epsilon(u) := \|u\|_1 < \gamma,$$

for some γ, $0 < \gamma < 2$, where the function $\Gamma(\gamma)$ is given by

$$\Gamma(\gamma) = \left(1 + \frac{1}{3}\gamma\right)^{-1} \cdot \left(1 + \frac{1}{3}\frac{1 - \frac{1}{2}\gamma}{1 + \gamma}\right).$$

[Note that $\Gamma(0) = \frac{4}{3}$, $\Gamma(2) = \frac{3}{5}$, and

$$\Gamma(\gamma) > 1, \quad \text{for } 0 < \gamma < \gamma_0 = \frac{1}{2}\left(\sqrt{\frac{11}{2}} - \frac{3}{2}\right). \tag{7.26}$$

As the functionals ϵ and Λ are invariant functionals for the BBM equation, it follows that $V(u)(t) > 0$ for all $t \geqslant 0$ for every solution whose initial value satisfies conditions (7.25). Although the functional Λ has the advantage of being an invariant functional, its relevance as a measure of the "length" of a function is less clear. Nevertheless, for functions

$$u_{\delta,\mu}(x) := \delta v(\mu x),$$

we have

$$\Lambda(u_{\delta,\mu}) = \left[\|v\|^2 + \mu^2\|\partial_x v\|^2\right] \cdot \left[\|v\|^2 + \frac{1}{6}\delta\int_{-\infty}^{\infty} dy v^3(y)\right]^{-1},$$

such that

$$\Lambda(u_{\delta,\mu}) \to 1 \quad \text{for } \delta, \mu > 0.$$

From this it follows that for the class of long, low waves $\Lambda \approx 1$, and hence, because of Eq. (7.26), this class is included in the set S_γ for $\gamma < \gamma_0$. This shows that the result stated above includes the contents of Theorem 7.5.

Remark 7.6: It is illustrative to apply the above described method to more general equations of the form (7.1), where D is given by Eq. (6.8). The first problem to be considered concerns the existence of global solutions of the initial value problem for such equations. For $\sigma > 0$, let H^σ denote the fractional order Sobolev space, defined as the set

$$H^\sigma := \{u \in L_2 \mid \hat{u} \cdot (1 + k^2)^{\sigma/2} \in L_2\},$$

supplied with the norm

$$\|u\|_{H^\sigma}^2 = \int_{-\infty}^{\infty} (1 + k^2)^\sigma \cdot |\hat{u}|^2 dk.$$

For arbitrary $\sigma > 0$, H^σ is a Banach space, but only if $\sigma > \frac{1}{2}$ is H^σ continuously embedded in C^0 and have the property that multiplication of elements from H^σ is a continuous operation in H^σ.

Along the same lines as in the proof of Lemma 7.3, it is possible (using a contraction mapping principle, now in a space based on H^σ) to prove the existence of a unique solution, over some time interval $[0,T]$ of the integral equation corresponding to Eq. (7.1):

$$u(x,t) = u_0(x) - \int_0^t \partial_x D^{-1} n(u(\tau)) d\tau, \tag{7.27}$$

where T depends on the H^σ norm of the initial value u_0 only:

$$T = T(\|u_0\|_{H^\sigma}). \tag{7.28}$$

Assuming $u_0 \in H^s$, for $s > \sigma$, it can be shown, using well-known bootstrap arguments, that the solution of Eq. (7.27) is also in H^s. This implies that if s is sufficiently large, u is also a solution of the original equation over the time interval $[0,T]$. Moreover, in that case the functional $m(u) = \frac{1}{2}\langle u, Du\rangle = \frac{1}{2}\|u\|_{H^\sigma}^2$ is in fact an invariant integral, from which it follows that because of Eq. (7.28) the local solution can be continued over an arbitrary time interval, which gives us the desired existence result.

To arrive at an useful estimate for the functional V, note that the group velocity (6.9) can be estimated as

$$\hat{G}(k) \geqslant 1 - 3[(1 + k^2)^\sigma - 1]. \tag{7.29}$$

Then V as given by Eq. (7.2) can be estimated as

$$V_\sigma(u) \geqslant \tfrac{1}{2} e(u)^{-1} \cdot \{\|n(u)\|^2 - 3[\|n(u)\|_{H^\sigma}^2 - \|n(u)\|^2]\} = \frac{1}{2}\frac{\|n(u)\|^2}{e(u)} \cdot [1 - 3\lambda_\sigma^{-2}(u)],$$

if the "averaged wavelength" functional λ_σ is defined by

$$\lambda_\sigma(u) = \left(\frac{\|n(u)\|_{H^\sigma}^2}{\|n(u)\|^2} - 1\right)^{-1/2}$$

In much the same way as was done in the proof of Theorem 7.4, it is then possible to show that $V_\sigma(u) \geqslant 0$ for all time $t > 0$ if u is a solution with initial data satisfying $\lambda_\sigma(u_0) > \lambda^0$ and $m(u) = \frac{1}{2}\|u\|_{H^\sigma}^2 < \epsilon_0$, for λ^0 and ϵ_0 sufficiently small. [It is intriguing to observe that no value of σ for which these results can be obtained along the lines indicated above ($\sigma > \frac{1}{2}$) corresponds to an operator D which leads to a group velocity $\hat{G}(k)$ which is positive for all real k; cf. Eq. (6.10)].

ACKNOWLEDGMENTS

This work was part of my doctoral thesis, prepared at Eindhoven University of Technology. I would like to thank my supervisor Professor L.J.F. Broer of the Physics Department for his encouragement and some helpful remarks.

[1] D.J. Korteweg and G. De Vries, Philos. Mag. **39**, 422 (1895).
[2] G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[3] T.B. Benjamin, "Lectures on Nonlinear Wave Motion," in *Nonlinear Wave Motion*, Lectures in Applied Mathematics (American Mathematical Society, Providence, 1974), Vol. 15.
[4] L.J.F. Broer, Appl. Sci. Res. **30**, 430 (1974).
[5] L.J.F. Broer, E.W.C. van Groesen, and J.M.W. Timmers, Appl. Sci. Res. **32**, 619 (1976).
[6] E.W.C. van Groesen, "Variational Methods in Mathematical Physics," Thesis, Eindhoven University of Technology.
[7] C.S. Gardner, J.M. Greene, M.D. Kruskal, and R.M. Miura, Commun. Pure Appl. Math. **29**, 97 (1974).

[8]P.D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).
[9]V.E. Zakharov and L.D. Faddeev, Funct. Anal. Appl. **5**, 280 (1972).
[10]E. Krotscheck and W. Kundt, Commun. Math. Phys. **60**, 171 (1978).
[11]T.B. Benjamin, J.L. Bona, and J.J. Mahony, Philos. Trans. R. Soc. London Ser. A **272**, 47 (1972).
[12]J. de Graaf and L.J.F. Broer, Rep. Math. Phys. **3**, 109 (1972).
[13]J.V. Wehausen and E.V. Laitone, "Surface Waves," in *Encyclopedia of Physics*, edited by S. Flügge (Springer, Berlin, 1960), Vol. 9.
[14]R.M. Miura, C.S. Gardner, and M.D. Kruskal, J. Math. Phys. **9**, 1204 (1968).

Symmetries and vacuum Maxwell's equations

W.-H. Steeb
Universität Paderborn, Theoretische Physik, D-4790 Paderborn, West Germany

(Received 22 January 1980; accepted for publication 22 February 1980)

A new class of infinitesimal symmetries are given for Maxwell's equations in the absence of sources.

Let $M = \mathbb{R}^4$ [space–time (x_1, x_2, x_3, ct)] and suppose that M is a pseudo-Riemannian manifold, where the metric tensor field is given by (Minkowski space)

$$g = dx_1 \otimes dx_1 + dx_2 \otimes dx_2 + dx_3 \otimes dx_3 - dx_4 \otimes dx_4 , \tag{1}$$

i.e., $g_{11} = g_{22} = g_{33} = 1$, $g_{44} = -1$, $(x_4 = ct)$. Consider the 2-form

$$\begin{aligned}\beta = {} & B_3 dx_1 \wedge dx_2 + B_1 dx_2 \wedge dx_3 + B_2 dx_3 \wedge dx_1 \\ & + E_1 dx_1 \wedge dt + E_2 dx_2 \wedge dt + E_3 dx_3 \wedge dt\end{aligned} \tag{2}$$

defined on M, where $\wedge$ is the exterior, or wedge, product.

For Riemannian or pseudo-Riemannian manifolds the Hodge *-operator (duality operation) is defined on differential forms: * is an f-linear mapping and transforms a p-form into its dual $(m-p)$-form (dim $M = m$). The *-operator applied to a p-form defined on an arbitrary Riemannian (or pseudo-Riemannian) manifold with metric tensor field g is defined by

$$\begin{aligned}&*(dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}): \\ &= \sum_{j_1 \cdots j_m = 1}^{m} g^{i_1 j_1} \cdots g^{i_p j_p} \frac{1}{(m-p)!} \frac{g}{\sqrt{|g|}} \\ &\quad \times \epsilon_{j_1 \cdots j_m} dx_{j_{p+1}} \wedge \cdots \wedge dx_{j_m} ,\end{aligned} \tag{3}$$

where $\epsilon_{j_1 \cdots j_m}$ is the total antisymmetric tensor $(\epsilon_{1,2,\ldots,m} = +1)$, $g \equiv \det(g_{ij})$ and $\Sigma_j g^{ij} g_{jk} = \delta^i_k$ (Kronecker symbol).

For the present case (Minkowski space) we obtain

$$\begin{aligned}*\beta = {} & \frac{E_1}{c} dx_2 \wedge dx_3 + \frac{E_2}{c} dx_3 \wedge dx_1 + \frac{E_3}{c} dx_1 \wedge dx_2 \\ & - B_3 c\, dx_3 \wedge dt - B_1 c\, dx_1 \wedge dt - B_2 c\, dx_2 \wedge dt.\end{aligned} \tag{4}$$

Observe that $**\beta = -\beta$. Now Maxwell's equations in the absence of sources are given by

$$d\beta = 0 \tag{5}$$

and

$$d(*\beta) = 0. \tag{6}$$

Moreover, note that

$$\beta \wedge \beta = 2(B_1 E_1 + B_2 E_2 + B_3 E_3)\, \Omega \tag{7}$$

and

$$\begin{aligned}\beta \wedge *\beta = {} & \left(\frac{E_1^2}{c} + \frac{E_2^2}{c} + \frac{E_3^2}{c} \right. \\ & \left. - B_1^2 c B_2^2 c - B_3^2 c \right) \Omega,\end{aligned} \tag{8}$$

where Ω is the volume element in space–time, i.e., $\Omega = dx_1 \wedge dx_2 \wedge dx_3 \wedge dt$.

The aim of the present paper is to give a new class of transformations which leave invariant the given set of partial differential equations. The invariance of the vacuum Maxwell equations has been investigated by Harrison and Estabrook.[1] We study additional transformations which have not been given by Harrison and Estabrook.[1]

Let us introduce the abbreviation

$$\begin{aligned}(\cdots) \equiv \Bigg(& B_1, B_2, B_3, E_1, E_2, E_3, \frac{\partial B_1}{\partial x_1}, \ldots, \\ & \times \frac{\partial B_1}{\partial t}, \frac{\partial B_2}{\partial x_1}, \ldots, \frac{\partial B_3}{\partial t}, \frac{\partial E_1}{\partial x_1}, \ldots, \frac{\partial E_3}{\partial t} \Bigg)\end{aligned} \tag{9}$$

The conditions $d\beta = 0$ and $d(*\beta) = 0$ lead to the following set of linear partial differential equations of first order:

$$\begin{aligned}
F_1(\cdots) &\equiv \frac{\partial B_3}{\partial x_2} - \frac{\partial B_2}{\partial x_3} - \frac{1}{c^2} \cdot \frac{\partial E_1}{\partial t} = 0 , \\
F_2(\cdots) &\equiv \frac{\partial B_1}{\partial x_3} - \frac{\partial B_3}{\partial x_1} - \frac{1}{c^2} \cdot \frac{\partial E_2}{\partial t} = 0 , \\
F_3(\cdots) &\equiv \frac{\partial B_2}{\partial x_1} - \frac{\partial B_1}{\partial x_2} - \frac{1}{c^2} \cdot \frac{\partial E_3}{\partial t} = 0 , \\
F_4(\cdots) &\equiv \frac{\partial E_3}{\partial x_2} - \frac{\partial E_2}{\partial x_3} + \frac{\partial B_1}{\partial t} = 0 , \\
F_5(\cdots) &\equiv \frac{\partial E_1}{\partial x_3} - \frac{\partial E_3}{\partial x_1} + \frac{\partial B_2}{\partial t} = 0 , \\
F_6(\cdots) &\equiv \frac{\partial E_2}{\partial x_1} - \frac{\partial E_1}{\partial x_2} + \frac{\partial B_3}{\partial t} = 0 , \\
F_7(\cdots) &\equiv \frac{\partial B_1}{\partial x_1} + \frac{\partial B_2}{\partial x_2} + \frac{\partial B_3}{\partial x_3} = 0 , \\
F_8(\cdots) &\equiv \frac{\partial E_1}{\partial x_1} + \frac{\partial E_2}{\partial x_2} + \frac{\partial E_3}{\partial x_3} = 0 .
\end{aligned} \tag{10}$$

Let us briefly describe the method for investigating the symmetries. We put $\partial B_1/\partial x_1 \to p_{11}$, $\partial B_1/\partial x_2 \to p_{12}$, $\ldots, \partial E_3/\partial t \to p_{64}$. Following Dieudonné[2] we consider for investigation the infinitesimal symmetries of the following differential forms:

$$F_1(B_1, B_2, B_3, E_1, E_2, E_3, p_{11}, \ldots, p_{14}, p_{21}, \ldots, p_{34}, p_{41}, \ldots, p_{64}),$$

$$\vdots$$

$$F_8(\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots),$$

$$dF_1, dF_2, \ldots, dF_8,$$

$$\alpha_1 = dB_1 - (p_{11}\, dx_1 + p_{12}\, dx_2 + p_{13}\, dx_3 + p_{14}\, dt),$$

 0022-2488/80/071656-03$1.00

$$\alpha_2 = dB_2 - (p_{21}\,dx_1 + p_{22}\,dx_2 + p_{23}\,dx_3 + p_{24}\,dt),$$
$$\alpha_3 = dB_3 - (p_{31}\,dx_1 + p_{32}\,dx_2 + p_{33}\,dx_3 + p_{34}\,dt),$$
$$\beta_1 = dE_1 - (p_{41}\,dx_1 + p_{42}\,dx_2 + p_{43}\,dx_3 + p_{44}\,dt),$$
$$\beta_2 = dE_2 - (p_{51}\,dx_1 + p_{52}\,dx_2 + p_{53}\,dx_3 + p_{54}\,dt),$$
$$\beta_3 = dE_3 - (p_{61}\,dx_1 + p_{62}\,dx_2 + p_{63}\,dx_3 + p_{64}\,dt),$$
$$d\alpha_1,\ d\alpha_2,\ d\alpha_3,\ d\beta_1,\ d\beta_2,\ d\beta_3. \tag{11}$$

Hence $F_i (i = 1,...,8)$ is a 0-form (function), $dF_i (i = 1,...,8)$ a 1-form; α_i and $\beta_i (i = 1,2,3)$ are 1-forms, and $d\alpha_i$ and $d\beta_i (i = 1,2,3)$ 2-forms; d denotes the exterior derivative.

Thus

$$\begin{aligned}
F_1(\cdots) &= p_{32} - p_{23} - \frac{1}{c^2} p_{44},\\
F_2(\cdots) &= p_{13} - p_{31} - \frac{1}{c^2} p_{54},\\
F_3(\cdots) &= p_{21} - p_{12} - \frac{1}{c^2} p_{64},\\
F_4(\cdots) &= p_{62} - p_{53} + p_{14},\\
F_5(\cdots) &= p_{43} - p_{61} + p_{24},\\
F_6(\cdots) &= p_{51} - p_{42} + p_{34},\\
F_7(\cdots) &= p_{11} + p_{22} + p_{33},\\
F_8(\cdots) &= p_{41} + p_{52} + p_{63}.
\end{aligned} \tag{12}$$

Now let

$$\langle F_1,...,F_8,\ dF_1,...,\ dF_8,\ \alpha_1,...,\beta_3,\ d\alpha_1,...,d\beta_3\rangle \tag{13}$$

denote the ideal generated by $F_1,...,\ d\beta_3$. Moreover, let Z denote a vector field (infinitesimal generator) defined on $\mathbb{R}^{6+24} \equiv \mathbb{R}^{30}$, i.e.,

$$Z = \sum_{i=1}^{3} \left[V_i(\cdots) \frac{\partial}{\partial E_i} + W_i(\cdots) \frac{\partial}{\partial B_i} \right] + \sum_{i=1}^{6} \sum_{j=1}^{4} T_{ij}(\cdots) \frac{\partial}{\partial p_{ij}}. \tag{14}$$

The partial differential equations given above are called invariant with respect to Z if

$$\begin{aligned}
&L_Z F_i \in \langle F_i\rangle \quad (i = 1,...,8),\\
&L_Z dF_i \in \langle F_i, dF_i, \alpha_j, \beta_j\rangle \quad (i = 1,...,8; j = 1,2,3),\\
&L_Z \alpha_j \in \langle F_i, dF_i, \alpha_j, \beta_j\rangle \quad (i = 1,...,8; j = 1,2,3),\\
&L_Z \beta_j \in \langle F_i, dF_i, \alpha_j, \beta_j\rangle \quad (i = 1,...,8; j = 1,2,3),\\
&L_Z\, d\alpha_j \in \langle F_i, dF_i, \alpha_j, \beta_j, d\alpha_j\ d\beta_j\rangle \quad (i = 1,...,8; j = 1,2,3),\\
&L_Z d\beta_j \in \langle F_i, dF_i, \alpha_j, \beta_j, d\alpha_j, d\beta_j\rangle \quad (i = 1,...,8; j = 1,2,3),
\end{aligned} \tag{15}$$

where $L_Z(\cdot)$ denotes the Lie derivative with respect to Z.

Let

$$Z_1 = \sum_{i=1}^{3} \left(B_i \frac{\partial}{\partial B_i} + E_i \frac{\partial}{\partial E_i} \right) \tag{16}$$

Then the once-extended vector field is given by[3]

$$\bar{Z}_1 = Z_1 + \sum_{k=1}^{6} \sum_{j=1}^{4} p_{kj} \frac{\partial}{\partial p_{kj}} \tag{17}$$

We find that the system of partial differential equations (10) is invariant under $\bar{Z}_1$. A straightforward calculation shows that

$$\begin{aligned}
&L_{\bar{Z}_1} F_i = F_i\,,\\
&L_{\bar{Z}_1} \alpha_j = \alpha_j; \quad L_{\bar{Z}_1} \beta_j = \beta_j\,.
\end{aligned} \tag{18}$$

In order to obtain $L_{\bar{Z}_1} dF_i$, $L_{\bar{Z}_1} d\alpha_j$, and $L_{\bar{Z}_1} d\beta_j$, we refer to the rule $d(L_X\gamma) \equiv L_X(d\gamma)$ for an arbitrary vector field x and an arbitrary form γ. Z_1 describes the field scale change. Now let

$$Z_2 = \sum_{i=1}^{3} \left(\frac{E_i}{c} \cdot \frac{\partial}{\partial B_i} - cB_i \frac{\partial}{\partial E_i} \right)$$

Then the once-extended vector field is given by[3]

$$\bar{Z}_2 = Z_2 + \frac{1}{c} \left\{ \sum_{j=1}^{4} \left(p_{4j} \frac{\partial}{\partial p_{1j}} - p_{1j} \frac{\partial}{\partial p_{4j}} \right) + \left(p_{5j} \frac{\partial}{\partial p_{2j}} - p_{2j} \frac{\partial}{\partial p_{5j}} \right) + \left(p_{6j} \frac{\partial}{\partial p_{3j}} - p_{3j} \frac{\partial}{\partial p_{6j}} \right) \right\}.$$

A straightforward calculation shows that the system of partial differential equations (10) is invariant under $\bar{Z}_2$. For example, we find

$$L_{\bar{Z}_2} F_1 = F_4,...,L_{\bar{Z}_2} F_6 = F_7, \quad L_{\bar{Z}_2} F_7 = -F_6$$

and

$$L_{\bar{Z}_2} \alpha_i = \frac{1}{c} \beta_i; \quad L_{\bar{Z}_2} \beta_i = -c\alpha_i\,.$$

The vector fields Z_1 and Z_2 commute, i.e.,

$$[Z_1, Z_2] = 0.$$

Consequently

$$[\bar{Z}_1, \bar{Z}_2] = 0.$$

The infinitesimal generators described above have been given by Harrison and Estabrook.[1] Z_2 generates, via a Lie series, the rotations

$$E_i \to E_i \cos\alpha + cB_i \sin\alpha; \quad B_i \to B_i \cos\alpha - (E_i/c) \sin\alpha.$$

Finally, note that $Z_1[*(\beta\wedge\beta)] = 2[*(\beta\wedge\beta)]$ and $Z_1[*(\beta\wedge *\beta)] = 2[*(\beta\wedge *\beta)]$.

Let us now give further infinitesimal generators which leave invariant the system of partial differential equations (10). Consider

$$Z_3 = \left(p_{23} - p_{32} + \frac{1}{c^2} p_{44}\right) \frac{\partial}{\partial B_1} + \left(p_{31} - p_{13} + \frac{1}{c^2} p_{54}\right) \times \frac{\partial}{\partial B_2} + \left(p_{12} - p_{21} + \frac{1}{c^2} p_{64}\right) \frac{\partial}{\partial B_3},$$

$$Z_4 = (p_{53} - p_{62} - p_{14}) \frac{\partial}{\partial E_1} + (p_{61} - p_{43} - p_{24}) \frac{\partial}{\partial E_2} + (p_{42} - p_{51} - p_{34}) \frac{\partial}{\partial E_3},$$

$$Z_5 = \left(p_{23} - p_{32} + \frac{1}{c^2} p_{44}\right) \frac{\partial}{\partial E_1} + \left(p_{31} - p_{13} + \frac{1}{c^2} p_{54}\right) \times \frac{\partial}{\partial E_2} + \left(p_{12} - p_{21} + \frac{1}{c^2} p_{64}\right) \frac{\partial}{\partial E_3},$$

$$Z_6 = (p_{53} - p_{62} - p_{14}) \frac{\partial}{\partial B_1} + (p_{61} - p_{43} - p_{24}) \frac{\partial}{\partial B_2}$$

$$+(p_{42}-p_{51}-p_{34})\frac{\partial}{\partial B_3},$$

$$Z_7=(p_{11}+p_{22}+p_{33})\left(\frac{\partial}{\partial E_1}+\frac{\partial}{\partial E_2}+\frac{\partial}{\partial E_3}\right),$$

$$Z_8=(p_{11}+p_{22}+p_{33})\left(\frac{\partial}{\partial B_1}+\frac{\partial}{\partial B_2}+\frac{\partial}{\partial B_3}\right),$$

$$Z_9=(p_{41}+p_{52}+p_{63})\left(\frac{\partial}{\partial E_1}+\frac{\partial}{\partial E_2}+\frac{\partial}{\partial E_3}\right),$$

$$Z_{10}=(p_{41}+p_{52}+p_{63})\left(\frac{\partial}{\partial B_1}+\frac{\partial}{\partial B_2}+\frac{\partial}{\partial B_3}\right).$$

A straightforward calculation shows that the system of partial differential equations is invariant under Z_3, Z_4,..., Z_{10}. In particular, we find

$$L_{Z_k} F_j = 0 \quad (k=3,\ldots,10; j=1,\ldots,7),$$

$$L_{Z_1}\alpha_i = -dF_i, \quad L_{Z_1}\beta_i = 0,$$

$$L_{Z_2}\alpha_i = 0, \quad L_{Z_2}\beta_i = dF_{i+3},$$

$$L_{Z_3}\alpha_i = 0, \quad L_{Z_3}\beta_i = -dF_i,$$

$$L_{Z_4}\alpha_i = -dF_{i+3}, \quad L_{Z_4}\beta_i = 0, \qquad (i=1,2,3)$$

$$L_{Z_5}\alpha_i = 0, \quad L_{Z_5}\beta_i = dF_7,$$

$$L_{Z_6}\alpha_i = dF_7, \quad L_{Z_6}\beta_i = 0,$$

$$L_{Z_7}\alpha_i = 0, \quad L_{Z_7}\beta_i = dF_8,$$

$$L_{Z_{10}}\alpha_i = dF_8, \quad L_{Z_{10}}\beta_i = 0.$$

In contrast to the vector fields Z_1 and Z_2, the vector fields Z_3,...,Z_{10} lead to a new class of invariances. Whereas the vector fields Z_1 and Z_2 generate tangent transformation groups of Lie, the vector fields Z_3,...,Z_{10} generate Lie–Bäcklund tangent transformation groups. In order to obtain the Lie–Bäcklund tangent transformation group, we must calculate the extension of the infinitesimal generator up to infinite order.[4]

[1] B.K. Harrison and F.B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[2] J. Dieudonné, *Treatise on Analysis* (Academic, New York, 1974), Chap. 18.
[3] G.W. Bluman and J.D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974), p. 158.
[4] R.L. Anderson and N.H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979), pp. 41,49.

Para-Fermi quantization in the representation of SO(n)

T. Maekawa and N. Noguchi
Department of Physics, Kumamoto University, Kumamoto 860, Japan

(Received 19 December 1979; accepted for publication 14 March 1980)

The result of the representation theory of SO(n) is applied to the quantization of f para-Fermi oscillators, and it is shown that a special irreducible representation of SO($2f+1$) is realized in the quantization of the field with order p when the condition of vacuum is used. The group theoretical meaning of the quantities such as vacuum, order p, and the space of the state in the para-Fermi quantization is made clear, and it is seen that the space of the state vectors corresponds to that of the single- or double-valued representations according to p even or odd.

I. INTRODUCTION

Paraquantization of field is studied by many authors[1-3] and two methods corresponding to the generalization of Fermi and Bose statistics are known useful. The f parafield operators are arranged to satisfy the commutation relations of SO($2f+1$) [4] or Sp($2f,R$) [5] according to the generalized Fermi or Bose operators. Therefore, the close connection between the parafield quantization and the representation theory of the groups is expected, but it seems that we know a few of the connections such as the Green decomposition and the Kronecker product representation of the fundamental spinor representation.[4] As the representation of SO(n) is well known, it will be important to clarify the connection between the para-Fermi quantization and the representation of SO(n) in order to make further developments of the parafield.

The purpose of this article is to apply the representation of SO(n) to the para-Fermi-quantization and to make clear the meaning of vacuum, order p of the para-Fermi quantization and the space of the state vectors with the fixed order p.[2] In Sec. II, a brief summary of the representation of SO(n) is given for fixing the notations and for the following discussion. In Sec. III, the result of the representation of SO(n) is used to define the vacuum state, and then the meaning of the vacuum and order of the quantization and space with the fixed order, which are introduced in earlier papers[2] independently of the unitary irreducible representation (UIR) of SO(n), are made clearer.

II. BRIEF SUMMARY OF THE REPRESENTATION OF SO(n)

In this section, the results needed for the following section on the group SO(n) are summarized briefly.[6,7]

The infinitesimal generators D_{jk} ($=-D_{kj}$, Hermitian) of the representation of SO(n) satisfy the commutation relations

$$[D_{jk},D_{lm}]=i(\delta_{jl}D_{km}+\delta_{km}D_{jl}-\delta_{kl}D_{jm}-\delta_{jm}D_{kl})\,. \tag{2.1}$$

The basis vectors for the UIR of SO(n), classified by the group chain SO(n)$\supset$SO($n-1$)$\supset\cdots\supset$SO(2), are given by the Gel'fand and Tsetlin[6] labels as follows:

$$|m_{ij}\rangle=\left|\begin{array}{l} m_{n1},m_{n2},\ldots,m_{n[n/2]} \\ m_{n-1\,1},m_{n-1\,2},\ldots,m_{n-1[(n-1)/2]} \\ \cdot \\ \cdot \\ \cdot \\ m_{51},m_{52} \\ m_{41},m_{42} \\ m_{31} \\ m_{21} \end{array}\right\rangle, \tag{2.2}$$

where $[j/2]$ means the largest integer smaller or equal to $j/2$. The numbers m_{jk} are simultaneously integers or half-integers and are subject to the restrictions

$$\begin{aligned} &m_{2j+1\,i+1}\leqslant m_{2ji}\leqslant m_{2j+1\,i} \quad (i=1,2,\ldots,j-1)\,, \\ &m_{2j\,i+1}\leqslant m_{2j-1\,i}\leqslant m_{2ji} \quad (i=2,3,\ldots,j-1), \\ &|m_{2jj}|\leqslant m_{2j-1\,j-1}\leqslant m_{2j\,j-1}\,, \\ &|m_{2jj}|\leqslant m_{2j+1\,j}\,. \end{aligned} \tag{2.3}$$

The UIR of SO(n) is characterized by the $[n/2]$ numbers m_{nj} in the top row in (2.2) and the rows and columns of the matrix elements of D_{jk} are labeled by those below the second row in (2.2).

The second-order Casimir operator is given by

$$F^{(n)}=\sum_{j>k}^{[n/2]}D_{jk}^2\,, \tag{2.4}$$

with the eigenvalue with respect to (2.2)[7]

$$F^{(n)}=\sum_{j=1}^{[n/2]}m_{nj}(m_{nj}+n-2j)\,. \tag{2.5}$$

Making use of (2.1), we obtain the relation for $j\leqslant n-2$,

$$\begin{aligned}[D_{jn},F^{(n-1)}-F^{(n-2)}]&=D_{jn}-2iD_{j\,n-1}D_{n-1\,n} \\ &=-D_{jn}-2iD_{n-1\,n}D_{j\,n-1}\,. \end{aligned} \tag{2.6}$$

It follows from (2.6) together with (2.5) that all matrix elements of the Lie algebra, i.e., D_{jk}, are determined from those of $D_{j\,j-1}$ ($j=2,3,\ldots,n$).

The dimension of the UIR of SO(n) is given with m_{nj} as follows[7]:

 0022-2488/80/071659-03$1.00

$$N(m_{n1},m_{n2},\ldots,m_{n(n-1)/2}) = \left[(n-2)!(n-4)!!\prod_{j=1}^{(n-5)/2}(n-2j-3)!\right]^{-1} \times\prod_{i=1}^{(n-1)/2}(2l_{ni}-1) \times\prod_{j=1}^{(n-3)/2}\prod_{k=j+1}^{(n-1)/2}[(l_{nj}-\tfrac{1}{2})^2-(l_{nk}-\tfrac{1}{2})^2], \quad (2.7)$$

for n odd, and

$$N(m_{n1},m_{n2},\ldots,m_{nn/2}) = 2^{(n-2)/2}\left[(n-2)!(n-4)!!\prod_{j=1}^{(n-4)/2}(n-2j-3)!\right]^{-1} \times\prod_{j=1}^{(n-2)/2}\prod_{k=j+1}^{n/2}(l_{nj}^2-l_{nk}^2), \quad (2.8)$$

for n even. The quantity l_{jk} is defined by

$$l_{jk} = m_{jk} + [(j+1)/2] - k \quad (k = 1,2,\ldots,[j/2]). \quad (2.9)$$

The action of the generators D_{jj+1} $(1\leqslant j\leqslant n-1)$ on the bases (2.2) is given as follows[6,7]:

$$D_{2k\,2k+1}|m_{ij}\rangle = \sum_{j=1}^{k}A(m_{2kj})|m_{2kj}+1\rangle - \sum_{j=1}^{k}A(m_{2kj}-1)|m_{2kj}-1\rangle, \quad (2.10)$$

$$D_{2k-1\,2k}|m_{ij}\rangle = \sum_{j=1}^{k-1}B(m_{2k-1\,j})|m_{2k-1\,j}+1\rangle - \sum_{j=1}^{k-1}B(m_{2k-1\,j}-1)|m_{2k-1\,j}-1\rangle + C_{2k}|m_{ij}\rangle, \quad (2.11a)$$

$$D_{21}|m_{ij}\rangle = m_{21}|m_{ij}\rangle. \quad (2.11b)$$

It is noted that in the bases on the right-hand side only the numbers, which change under the action of the generator, are written except the base $|m_{ij}\rangle$. The matrix elements A, B, and C are given in the form

$$A(m_{2kj}) = \tfrac{1}{2}i\left\{\prod_{i=1}^{k-1}[(l_{2k-1\,i}-\tfrac{1}{2})^2-(l_{2kj}+\tfrac{1}{2})^2] \times\prod_{i=1}^{k}[(l_{2k+1\,i}-\tfrac{1}{2})^2-(l_{2kj}+\tfrac{1}{2})^2]\right\}^{1/2} \times\left\{\prod_{i=1}^{k}{}'(l_{2ki}^2-l_{2kj}^2)[l_{2ki}^2-(l_{2kj}+1)^2]\right\}^{-1/2}, \quad (2.12)$$

$$B(m_{2k-1\,j}) = i\left[\prod_{i=1}^{k-1}(l_{2k-2\,i}^2-l_{2k-1\,j}^2)\prod_{i=1}^{k}(l_{2ki}^2-l_{2k-1\,j}^2)\right]^{1/2} \times\left\{l_{2k-1\,j}^2(4l_{2k-1\,j}^2-1)\prod_{i=1}^{k-1}{}'(l_{2k-1\,i}^2-l_{2k-1\,j}^2) \times[(l_{2k-1\,i}-1)^2-l_{2k-1\,j}^2]\right\}^{-1/2}, \quad (2.13)$$

$$C_{2k} = \prod_{i=1}^{k-1}l_{2k-2\,i}\prod_{i=1}^{k}l_{2ki}\left[\prod_{i=1}^{k-1}l_{2k-1\,i}(l_{2k-1\,i}-1)\right]^{-1}, \quad (2.14)$$

where the prime on Π means the product of factors except for $i=j$.

In the next section, we need the representation in the case of n odd. Therefore, we define the base $|\min\rangle$ in the case of $n\,(=2f+1)$, odd by the base (2.2), with the possible minimum numbers m_{ij} given by (2.3), i.e.,

$$|\min\rangle = \left|\begin{array}{c} m_{n1},m_{n2},\ldots,m_{nf-1},m_{nf} \\ m_{n2},m_{n3},\ldots,m_{nf},-m_{nf} \\ \cdot \\ \cdot \\ \cdot \\ m_{nf},m_{nf} \\ m_{nf},-m_{nf} \\ m_{nf} \\ -m_{nf} \end{array}\right\rangle \quad (2.15)$$

It, then, follows from (2.11) and (2.14) that the base (2.15) is the eigenvector of $D_{2j-1\,2j}$ with the eigenvalue, i.e.,

$$D_{2j-1\,2j}|\min\rangle = \bar{C}_{2j}|\min\rangle, \quad (2.16)$$

where the $\bar{C}_{2j}$ is the C_{2j} evaluated at the base (2.15). It is easy to give the explicit expression for the $\bar{C}_{2j}$, but we do not need it here. These results will be used in the next section.

III. PARA-FERMI QUANTIZATION AND THE UIR OF SO(n)

Let a_j $(j=1,2,\ldots,f)$ be a set of operators and $a_j^\dagger$ their adjoints. The para-Fermi quantization of these operators is defined by the commutation relations[1]

$$[a_j,[a_k^\dagger,a_l]] = 2\delta_{jk}a_l, \quad [a_j,[a_k,a_l]] = 0, \quad (3.1)$$

where $[A,B] = AB - BA$. If we define the quantities J_{jk} in terms of a_j and $a_j^\dagger$ as follows,

$$\begin{aligned} J_{2j\,2k-1} &= -J_{2k-1\,2j} = \tfrac{1}{2}(N_{jk}+N_{kj}+L_{jk}-M_{jk}), \\ J_{2j-1\,2k-1} &= -J_{2k-1\,2j-1} = -\tfrac{1}{2}i(N_{jk}-N_{kj}+L_{jk}+M_{jk}), \\ J_{2j2k} &= -J_{2k\,2j} = -\tfrac{1}{2}i(N_{jk}-N_{kj}-L_{jk}-M_{jk}), \\ J_{2j-1\,2f+1} &= -J_{2f+1\,2j-1} = \tfrac{1}{2}(a_j+a_j^\dagger), \\ J_{2j\,2f+1} &= -J_{2f+1\,2j} = -\tfrac{1}{2}i(a_j-a_j^\dagger), \end{aligned} \quad (3.2)$$

where

$$N_{jk} = \tfrac{1}{2}[a_j^\dagger,a_k], \quad L_{jk} = \tfrac{1}{2}[a_j^\dagger,a_k^\dagger], \quad M_{jk} = \tfrac{1}{2}[a_j,a_k],$$

we have $f(2f+1)\,[=n(n-1)/2]$ Hermitian operators J_{jk} and it follows from (3.1) and their adjoints that the quantities J_{jk} satisfy the same commutation relations as (2.1) for the generators of SO(n) ($n=2f+1$).[4]

Usual para-Fermi quantization may be summarized as follows.[2] The unique no-particle state $|0\rangle$ (vacuum) is defined by the eigenstate of the operators N_{jj} $(=-J_{2j-1\,2j})$ with the minimum eigenvalue, and due to (3.1)

$$a_j|0\rangle = (J_{2j-1\,2f+1}+iJ_{2j\,2f+1})|0\rangle = 0, \quad (3.3)$$

for all $j=1,2,\ldots,f$. Then, the order p of the para-Fermi quantization is defined by

$$a_j a_l^\dagger|0\rangle = \delta_{jl}\,p|0\rangle \quad (3.4)$$

It is, then, shown that the space of the para-Fermi quantization with the fixed p is spanned by the vectors which are obtained by operating all possible $a_j^\dagger$ on $|0\rangle$ and p becomes a nonnegative integer due to the positive definiteness of the norm of the state vectors.[2] However, it seems necessary to

clarify the group-theoretical meaning of the vacuum, order p, irreducibility, and single- or double-valued representation.

If we use the representation theory of SO(n) and definition of the vacuum, the meaning of the above quantities such as the vacuum, order p, etc., becomes evident. In what follows, we show the close connection.

Making use of the same notations as in Sec. 2, we obtain from (2.16)

$$N_{jj}|\min\rangle = -J_{2j-1\,2j}|\min\rangle = -\bar{C}_{2j}|\min\rangle . \tag{3.5}$$

It is, thus, expected that the state $|\min\rangle$ corresponds to the vacuum state because the vacuum state is defined by the minimum eigenstate of N_{jj} ($j = 1,2,...,f$). Therefore, in order that the state $|\min\rangle$ denotes the vacuum, it is necessary to be subjected to the following conditions due to (3.1):

$$a_j|\min\rangle = (J_{2j-1\,2f+1} + iJ_{2j\,2f+1})|\min\rangle = 0 , \tag{3.6}$$

for all j ($j = 1,2,...,f$). Operating $a_j^\dagger$ from the left of (3.6) and summing over j after making use of the relations (2.1) and (2.4), we obtain

$$\left[F^{(2f+1)} - F^{(2f)} - \sum_{j=1}^{f} D_{2j-1\,2j}\right]|\min\rangle = 0 . \tag{3.7}$$

(3.7) leads to the conditions

$$\sum_{j=1}^{f} m_{nj}(m_{nj} + 2f + 1 - 2j)$$
$$- \sum_{j=1}^{f-1} m_{nj+1}(m_{nj+1} + 2f - 2j) - m_{nf}(m_{nf} + f) = 0 ,$$
$$B(m_{2j-1k}) = 0 \quad (j = 2,3,...,f;\ k = 1,2,...,j-1) , \tag{3.8}$$

where (2.5), (2.16), and (2.11) are used and $B(m_{2j-1k})$'s are given at the possible minimum values (2.15) of m_{jk}. The following result is obtained from (3.8) with (2.13):

$$m_{n1} = m_{n2} = m_{n3} = \cdots = m_{nf} . \tag{3.9}$$

Thus, the vacuum state, which will be written with the same symbol $|0\rangle$ as in (3.3), is completely characterized by a number m_{n1} which is written as $p/2$, and it follows that $|0\rangle$ is explicitly given by (2.15) with all $m_{nj} = p/2$ ($j = 1,2,...,f$) and becomes the eigenvector of N_{jj} with the eigenvalue $-m_{n1}$ ($= -p/2$)

$$N_{jj}|0\rangle = -\tfrac{1}{2}p|0\rangle , \tag{3.10}$$

for we have $\bar{C}_{2j} = m_{n1}$ ($j \geqslant 2$) in the case of (3.9). It is obvious from (3.5), (3.6), and the expression of N_{jj} given below (3.2) that the vacuum state $|0\rangle$ satisfies the relation

$$a_j a_k^\dagger|0\rangle = \delta_{jk}\, p|0\rangle , \tag{3.11}$$

which agrees with (3.4). Thus, we see that p is the order of the para-Fermi quantization defined by (3.4).

Therefore, it is seen that the f para-Fermi oscillators with the order p ($= 2m_{n1}$) is given by the representation of SO(n) characterized by a number p, and the single- or double-valuedness of the representation is trivial according to p even or odd. It is easy to give the Gel'fand and Tsetlin bases by operating the raising (or lowering) operators on $|0\rangle$ (or $|\max\rangle$),[8] but we remark briefly about the representation diagonalizing N_{jj}.

From (3.1), we obtain for $s = 1,2,\cdots$

$$[N_{jj},(a_k^\dagger)^s] = s\delta_{jk}(a_k^\dagger)^s \tag{3.12}$$

Thus, the following relation holds from (3.5) and (3.12):

$$N_{jj}\prod_{k=1}^{f}(a_k^\dagger)^{s_k}|0\rangle = (-\tfrac{1}{2}p + s_j)\prod_{k=1}^{f}(a_k^\dagger)^{s_k}|0\rangle , \tag{3.13}$$

with nonnegative integers s_k and the action of $(a_k^\dagger)^s$ on $|0\rangle$ becomes identically zero for $s > p$ due to (2.10) and (2.11). It follows from (3.12) that in general the eigenvectors of N_{jj} are degenerate. In order to specify the bases diagonalizing N_{jj} uniquely, we need $f(f-1)/2$ commuting operators more because we have only f commuting operators N_{jj} but $f(f+1)/2$ operators are needed to specify the matrix elements uniquely.

It is obvious that the result of the action of a_j or $a_j^\dagger$ on the states such as (3.12) can be expressed in terms of the linear combination of the Gel'fand and Tsetlin bases given by (2.2) with $m_{n1} = m_{n2} = \cdots = m_{nf} = p/2$. The dimension of the space with the order p is, of course, given from (2.7) ($n = 2f+1$) by

$$N(p) = \left[(2f-1)!\,(2f-3)!!\prod_{j=1}^{f-2}(2f-2j-2)!\right]^{-1}$$
$$\times\frac{(p+2f-1)!!}{(p-1)!!}\prod_{j=1}^{f-1}\frac{(f-j)!\,(p+2f-2j)!}{(p+f-j)!} \tag{3.14}$$

Thus, we may conclude that there exists the close connection between the quantities such as the vacuum, order p, irreducible representation space with p in the para-Fermi quantization and those such as the minimum state defined in Sec. 2, characteristic number m_{n1} ($= p/2$), and irreducible representation space of the group SO(n). The result will play an important role in the further development of the para-Fermi quantization.

[1] H. S. Green, Phys. Rev. **90**, 270 (1953).
[2] O. W. Greenberg and A. M. L. Messiah, Phys. Rev. **138**, B1155 (1965).
[3] Soryushiron Kenkyu (Mimeographed Circular in Japanese) (Kyoto). Many related papers are cited there.
[4] C. Ryan and E. C. G. Sudarshan, Nucl. Phys. **47**, 207 (1963).
[5] S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1962); M. Omote, Y. Ohnuki, and S. Kamefuchi, Prog. Theor. Phys. **56**, 1948 (1976).
[6] I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk SSSR **71**, 840 (1950).
[7] T. Maekawa, J. Math. Phys. **20**, 691 (1979).
[8] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967); M. K. F. Wong, *ibid.* **8**, 1899 (1967).

High-energy behavior of renormalized Feynman amplitudes. II

Edward B. Manoukian
Department of National Defence, Royal Military College of Canada, Kingston, Ontario K7L 2W3

(Received 8 January 1980; accepted for publication 22 February 1980)

The high-energy polynomial and logarithmic behavior of renormalized Feynman amplitudes, involving subtractions, is studied with total generality when some or all of the external momenta of the graphs in question become large in Euclidean space nonexceptionally for theories which on experimental grounds may contain *zero* mass particles. Initially, all the propagators are defined to have nonzero masses in their denominators. The vanishing masses are led to vanish, and in general, at different rates. Rules are given for applications and examples are then worked out to demonstrate the application of these rules.

I. INTRODUCTION

On dimensional grounds the high-energy behavior of a Feynman amplitude and its zero-mass behavior are expected to be intimitaly connected. The actual connection is, however, far from obvious especially when several momenta and several masses are involved in an amplitude and when some, not necessarily all, of the masses approach zero, and in general, at different rates. A thorough and quite general study of the zero-mass behavior of Feynman amplitudes has been recently carried out in Euclidean space[1] and we shall spell out in this paper the basic conclusions needed here. On the other hand, our previous high-energy study[2] of subtracted amplitudes relied on a nonzero mass constraint in the theory. The purpose of this work is to carry out a general study of the high-energy behavior of subtracted-out Feynman amplitudes for theories that may contain zero-mass particles on experimental grounds. Initially we write for the denominators of the propagators, corresponding to a pole term Q^2, $Q^2+\mu^2$ in the expression for the (subtracted) Feynman integrand. The study is general and considers:

(i) The cases when the vanishing masses vanish, in general, at different rates in conformity with experiments. Experimentally, different small upper bounds are set on masses that, theoretically, are taken to be zero.

(ii) The high-energy behavior of the amplitudes corresponds also to cases when some or all the external (nonexceptional) momenta become large, and in general, at different rates. To obtain the final expression for the renormalized Feynman integrand one may use the subtractions as given in Ref. 3.

Section II deals with the main analysis. In Sec. III we give some examples as illustrations of the rules given in Sec. II.

II. HIGH-ENERGY ANALYSIS

A Feynman amplitude associated with a proper and connected graph G is of the form

$$A(p_1,p_2,\cdots;\mu_1,\mu_2,\cdots) = \int dk_1\cdots dk_{4n}\, R(p_1,p_2,\cdots;k_1,k_2,\cdots;\mu_1,\mu_2,\cdots), \tag{1}$$

where

$$R(p_1,p_2,\cdots;k_1,k_2,\cdots;\mu_1,\mu_2,\cdots) = \sum \frac{A^{i}_{m_{i1},\cdots,n_{i1},\cdots,\tau_{i1}}}{\Pi_l (Q_l^2+\mu_l^2)^{\sigma_l}} \times \prod_j (p_j)^{m_{ij}} \prod_t (k_t)^{n_{it}} \prod_n (\mu_n)^{\tau_{in}}. \tag{2}$$

$\sigma_l > 0$, the sum is over nonnegative integers $m_{i1},\ldots,n_{i1},\ldots,\tau_{i1},\ldots$. The A^i's are some suitable coefficients and $Q_l = q_l^G + k_l^G$, with $q_l^G = \Sigma_j a_j^l p_j$ and $k_l^G = \Sigma_j b_j^l k_j$, respectively, denote the external and internal momenta of the graph G. In reference to a subdiagram $g \subset G$, we write $Q_l = q_l^g + k_l^g$, where, of course, q_l^G, k_l^G and q_l^g, k_l^g may not necessarily coincide and, in general, an external momentum of g may be an internal momentum of G. Let

$$\mathscr{L}_p = \{p_1,p_2,\ldots,p_r\}, \quad \mathscr{L}_\mu = \{\mu_1,\mu_2,\ldots,\mu_s\}$$

be subsets of the external momenta and masses of the graph G. We consider the following scalings by positive numbers $\eta_1,\eta_2,\ldots,\eta_r,\lambda_1,\lambda_2,\ldots,\lambda_s$:

$$p_1\to\eta_1\cdots\eta_r p_1, \quad p_2\to\eta_2\cdots\eta_r p_2,\cdots,p_r\to\eta_r p_r,$$
$$\mu_1\to\lambda_1\mu_1, \quad \mu_2\to\lambda_1\lambda_2\mu_2,\ldots, \quad \mu_s\to\lambda_1\lambda_2\cdots\lambda_s\mu_s, \tag{3}$$

and consider the behavior of A when $\eta_1,\eta_2,\ldots,\eta_r\to\infty$ and $\lambda_1,\lambda_2,\ldots,\lambda_s\to 0$, all approaching their limits independently. Without loss of generality, we may assume that the momenta (nonexceptional) that become large at the same rate have been grouped into classes, with p_1 or $p_2,\ldots$, or p_r as their representative. Similarly, the vanishing masses that vanish at the same rate may have been assumed to be identified with μ_1 or $\mu_2,\ldots$, or μ_s. Under the scalings in (4), the integrand R, as defined in (2), is transformed to

$$R(p_1',p_2',\cdots;k_1',k_2',\cdots;\mu_1',\mu_2',\cdots)(\lambda_1\lambda_2\cdots\lambda_s)^{d(N)-2\Sigma_l\sigma_l}, \tag{4}$$

where

$$p_j' = \frac{(\eta_1\cdots\eta_j)^{c_{jr}}}{\lambda_1\lambda_2\cdots\lambda_s} p_j, \quad k_l' = (\lambda_1\cdots\lambda_s)^{-1}k_l,$$
$$\mu_n' = \frac{(\lambda_1\cdots\lambda_n)^{c_{ns}}}{\lambda_1\cdots\lambda_s}\mu_n, \tag{5}$$

$$c_{ab} = \begin{cases} 1, & a\leqslant b, \\ 0, & a>b, \end{cases} \tag{6}$$

 0022-2488/80/071662-03$1.00

$$(Q_l^2+\mu_l^2)\to(\lambda_1\cdots\lambda_s)^2\left[\left(\sum_j a_j^l p_j' + \sum_j b_j^l k_j'\right)^2+\mu_l'^2\right], \quad (7)$$

and $d(N)$ is the dimensionality of the numerator of R. Accordingly, the amplitude A is transformed to

$$A(\eta_1\cdots\eta_r\, p_1,\ldots,\eta_r\, p_r, p_{r+1},\cdots;\lambda_1\mu_1,\ldots,\lambda_1\cdots\lambda_s\mu_s,\mu_{s+1},\cdots)$$
$$=(\lambda_1\lambda_2\cdots\lambda_s)^{d(G)}A', \quad (8)$$

with

$$A'=A(p_1',p_2',\cdots;\mu_1',\mu_2',\cdots)$$
$$=\int dk_1\cdots dk_4\, R(p_1',\cdots;k_1,k_2,\cdots;\mu_1',\mu_2',\cdots), \quad (9)$$

and we may write

$$p_1'=(\eta_1\cdots\eta_r\lambda_1^{-1}\cdots\lambda_s^{-1})p_1,\ldots,p_r'=(\eta_r\lambda_1^{-1}\cdots\lambda_s^{-1})p_r,$$
$$p_j'=(\lambda_1^{-1}\cdots\lambda_s^{-1})p_j, \quad r<j,$$
$$\mu_k'=(\lambda_1^{-1}\cdots\lambda_s^{-1})\mu_k, \quad s<k, \quad (10)$$
$$\mu_1'=(\lambda_2^{-1}\cdots\lambda_s^{-1})\mu_1,\ldots,\mu_s'=\mu_s.$$

In Eq. (8) we have identified $d(N)-2\Sigma_l\sigma_l+4n$ with the dimensionality of the graph G. Let N be the number of p,k,μ components and consider $p_1',p_2',\cdots;k_1,k_2,\cdots;\mu_1',\mu_2',\cdots$ as the components of an N-dimensional vector $\mathbf{P}'$ in an N-dimensional Euclidean space R^N. Let I be a $4n$-dimensional subspace of R^N, associated with the $4n$ integration variables. Let E be the orthogonal complement of I in R^N. Let $\Lambda(I)$ denote the projection operator onto E along I.[4] In particular, p_1', $p_2',\cdots;\mu_1',\mu_2',\cdots$ may be considered as the components of a vector $\mathbf{P}$, with $\mathbf{P}=\Lambda(I)\mathbf{P}'$. The vector $\mathbf{P}$ may be written in the form:

$$\mathbf{P}=(\lambda_1\cdots\lambda_s)^{-1}\left\{\sum_{i=1}^{r}(\eta_i\cdots\eta_r)\mathbf{L}_i+\mathbf{L}_{r+1}\right\}$$
$$+\sum_{j=2}^{s}(\lambda_j\cdots\lambda_s)^{-1}\mathbf{L}_{j+s}+\mathbf{C}, \quad (11)$$

with

$\mathbf{L}_1$ with nonvanishing component p_1,

$\vdots$

$\mathbf{L}_r$ with nonvanishing component p_r,

$\mathbf{L}_{r+1}$ with nonvanishing components $p_{r+1},p_{r+2},\cdots;$

$$\mu_{s+1},\mu_{s+2},\cdots, \quad (12)$$

$\mathbf{L}_{r+2}$ with nonvanishing component μ_1,

$\vdots$

$\mathbf{L}_{r+s}$ with nonvanishing component μ_{s-1},

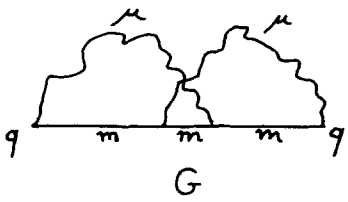


FIG. 1. A fourth-order electron self-energy graph.

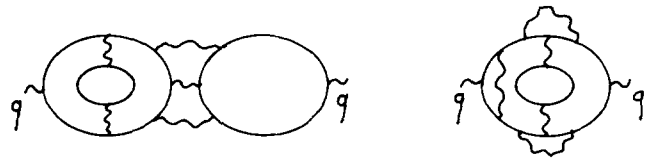


FIG. 2. Some twelfth-order photon self-energy graphs.

$\mathbf{C}$ with nonvanishing component μ_s.

The transformed amplitude A may be rewritten as $(\lambda_1\cdots\lambda_s)^{d(G)}A(\mathbf{P})$ we denote by $S_j=\{\mathbf{L}_1,\mathbf{L}_2,\ldots,\mathbf{L}_j\}$ a subspace of E spanned by the vectors $\mathbf{L}_1,\mathbf{L}_2,\ldots,\mathbf{L}_j$. We say that the parameter η_i appearing in (11) is associated with the subspace S_j, and λ_i^{-1}, also appearing in (11), is associated with the subspace S_{r+i}, $1\leqslant i\leqslant s$. For each line l in G carrying a momentum-mass $(Q_l^\alpha)=(Q_l^0,Q_l^1,Q_l^2,Q_l^3,\mu_l)$ we introduce a vector $\mathbf{V}_l^\alpha$ such that $\mathbf{V}_l^\alpha\cdot\mathbf{P}'=Q_l^\alpha$. In particular, we note that $(\mathbf{V}_l^\alpha\cdot\mathbf{P}')(\mathbf{V}_l^\alpha\cdot\mathbf{P}')=Q_l^2+\mu_l^2$. As $Q_l=q_l+k_l$, we may also introduce, separately, $\mathbf{V}$ vectors for the external momenta, internal momenta, and the masses carried by each line.

We give the analysis first (a) for the behavior of A for $\eta_1,\ldots,\eta_r\to\infty,\lambda_1,\ldots,\lambda_s\to0$ for the general cases. We then consider the behavior of A in the cases when the limits $\lambda_1,\lambda_2,\ldots,\lambda_s$ of A [(b)] may be rigorously taken.

(a) Let $(\lambda_i^{-1})\equiv\eta_{r+i}$. Let j be a fixed integer in $1\leqslant j\leqslant r+s$. We consider the behavior of $A(\mathbf{P})$ in reference to the parameter η_j. Let $T_j=\{G_j',G_j'',\cdots\}$ be the set of all subdiagrams $\subseteq G$, with $d(G_j')=d(G_j'')=\cdots$, and let $\{S_j',S_j'',\cdots\}$ be subspaces in R^N, with which the subdiagrams $\{G_j',G_j'',\cdots\}$ are associated; i.e., all the $\mathbf{V}$ of a $G_j'\in T_j$ are not orthogonal to S_j' and $\Lambda(I)S_j'=\Lambda(I)S_j''=\cdots=S_j$, such that any other subdiagram G', similarly defined, is such that $d(G')\leqslant d(G_j')$. If the equality holds, then $G'\in T_j$, by definition. For each $G_j'\in T_j$ let $\{g_j^i, i=1,\cdots\}$ be the set of all proper but not necessarily connected subdiagrams $\subseteq G_j'$ such that for each g_j^i each connected part of g_j^i is divergent. For each g_j^i consider the subdiagram $(G_j'\mid g_j^i)$ obtained from G_j' by shrinking g_j^i in it to a point and replacing it by a polynomial of degree $\leqslant d(g_j^i)$ in its external momenta. Let $S_j'^i$ be a subspace of R^N with which the subdiagram $(G_j'\mid g_j^i)$ is associated, i.e., all the $\mathbf{V}$ of the lines in G_j'/g_j^i are not orthogonal to $S_j'^i$, all the $\mathbf{V}$ of the internal momenta and the masses in g_j^i are orthogonal to $S_j'^i$, and at least some of the external momenta of g_j^i are not orthogonal to $S_j'^i$. (G/g simply denotes a diagram G with g, a subdiagram ($g\subseteq G$) in G shrunk to a point.) In particular, we note that for $j>r$ all the masses in the lines in each g_j^i must be from the set $\{\mu_{j-r+1},\ldots,\mu_s\}$. Accordingly, the behavior of A may be quite generally stated as follows:

$$\lim_{\substack{\lambda_1,\ldots,\lambda_s\to0\\ \eta_1,\ldots,\eta_r\to\infty}} A$$
$$=O\left(\eta_1^{\delta_1}\cdots\eta_{s+r}^{\delta_{s+r}}\sum_{\gamma_1,\ldots,\gamma_{r+s}}(\ln\eta_{\pi_1})^{\gamma_1}\cdots(\ln\eta_{\pi_{s+r}})^{\gamma_{s+r}}\right), \quad (13)$$

where $\lambda_1^{-1}=\eta_{r+1},\ldots,\lambda_s^{-1}=\eta_{r+s}$,

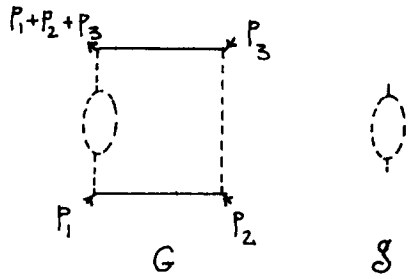


FIG. 3. (a) Fermion–Fermion (spin $\frac{1}{2}$) scattering graph. The dotted lines denote scalar bosons. (b) Proper divergent subdiagram g of G with masses from the set $\{\mu\}$.

$$\delta_i = d(G_i'), \quad 1 \leqslant i \leqslant r,$$
$$\delta_j = -d(G/G_j'), \quad r+1 \leqslant j \leqslant r+s, \tag{14}$$

and the sums are over all nonnegative integers $\gamma_1,...,\gamma_{r+s}$ satisfying

$$\sum_{i=1}^{k} \gamma_i \leqslant \beta_I(\{\mathbf{L}_1,\mathbf{L}_2,...,\mathbf{L}_{\pi_k}\}), \tag{15}$$

for all $1 \leqslant k \leqslant s+r$, and where we ordered the so-called logarithmic asymptotic coefficients[2,5] β_I in an increasing

$$\beta(\{\mathbf{L}_1,...,\mathbf{L}_{\pi_1}\}) \leqslant \cdots \leqslant \beta(\{\mathbf{L}_1,...,\mathbf{L}_{\pi_{s+r}}\}), \tag{16}$$

where $\pi_1,...,\pi_{s+r}$ is a permutation of $1,2,...,s+r$.

(b) Here it is assumed that the limits $\lambda_1,...,\lambda_s \to 0$ in A may be rigorously taken. According to Ref. 1, the limits $\lambda_1,\lambda_2,...,\lambda_s \to 0$ of A exist if in reference to each parameter λ_i, $1 \leqslant i \leqslant s$, the following two conditions are true:

(i) Any subdiagram $G_0 \subseteq G$ containing all the external vertices of G, with all the masses in G/G_0 (if not empty) from the set $\{\mu_i,\mu_{i+1},...,\mu_s\}$, and any line in G_0 not participating in a closed loop and not carrying an external momenta of G carrying a mass from the set $\{\mu_1,...,\mu_{i-1},\mu_{s+1},\mu_{s+2},\cdots\}$ must be *such that* $d(G_0) \leqslant d(G)$. This must be true for all G_0's as just defined. By an external vertex of G, it is meant a vertex of G to which an external line to G would be attached. In particular, we note that the definition of G_0 as just given implies that if there is an internal vertex in G to which is attached just one line of G_0, then the mass carried by this line must be from the set $\{\mu_1,...,\mu_{i-1},\mu_{s+1},\mu_{s+2},\cdots\}$.

(ii) For those subdiagrams G_0, as defined above, with $d(G_0) = d(G)$, G_0 must not contain a divergent proper subdiagram ($\subseteq G_0$) with all the masses in its lines from the set $\{\mu_i,\mu_{i+1},...,\mu_s\}$.

In particular, we note that $G(=G_0)$ itself must trivially satisfy these two conditions, for all $1 \leqslant i \leqslant s$, to be in the category (b). According to such a hypothesis limit $A = O(1)$ for $\lambda_1,\lambda_2,...,\lambda_s \to 0$, as any subdiagram G_0, as defined above, is such that $d(G_0) \leqslant d(G)$ (by hypothesis). This means that $\beta_I(S_{r+1}) = \cdots = \beta_I(S_{r+s}) = 0$. According to such a hypothesis, the behavior of A simplifies to

$$\lim_{\substack{\lambda_1,...,\lambda_s \to 0,\\ \eta_1,...,\eta_r \to \infty}} A$$
$$= O\left(\eta_1^{d(G_1')}\cdots\eta_r^{d(G_r')} \sum_{\gamma_1,...,\gamma_r} (\ln\eta_{\pi_1})^{\gamma_1}\cdots(\ln\eta_{\pi_r})^{\gamma_r}\right), \tag{17}$$

where

$$\sum_{i=1}^{k} \gamma_i \leqslant \beta(\{\mathbf{L}_1,...,\mathbf{L}_{\pi_k}\}) \tag{18}$$

for all $1 \leqslant k \leqslant r$, and

$$\beta(\{\mathbf{L}_1,...,\mathbf{L}_{\pi_1}\}) \leqslant \cdots \leqslant \beta(\{\mathbf{L}_1,...,\mathbf{L}_{\pi_r}\}), \tag{19}$$

where $\pi_1,...,\pi_r$ is a permutation of $1,...,r$.

In the next section examples will be worked out to illustrate these rules for graphs in categories (a) and (b).

III. EXAMPLES

Consider the classic example of the self-energy graph in Fig. 1 with $d(G) = 1$. Here we are interested in the amplitude $A(\eta q,m,\lambda\mu)$ for $\lambda \to 0$, $\eta \to \infty$. The limit $\lambda \to 0$ exists,[1] as any subdiagram $G' \subseteq G$ is such that $d(G') \leqslant d(G)$ and G does not contain any divergent proper subdiagram $\subseteq G$ having all its masses from the set $\{\mu\}$. On the other hand, the asymptotic coefficients of η are (c.f. Ref. 1) $\alpha_I(S_1) = 1$ and $\beta_I(S_1) = 1$, hence

$$\lim_{\lambda \to 0,\ \eta \to \infty} A(\eta q,m,\lambda\mu) = O(\eta[a_0 + a_1 \ln\eta]), \tag{20}$$

where a_0 and a_1 are constants.

Similarly, for the photon self-energy graphs in Fig. 2 the corresponding amplitudes $A(\eta q,m,\lambda\mu)$ exist for $\lambda \to 0$,[1] and the asymptotic coefficients of η are $\alpha_I(S_1) = 2$, $\beta_I(S_1) = 6$, i.e.,

$$\lim_{\lambda \to 0,\ \eta \to \infty} A(\eta q,m,\lambda\mu) = O\left(\eta^2 \sum_{n=0}^{6} a_n (\ln\eta)^n\right), \tag{21}$$

with $a_0,...,a_6$ some constants, some of which may be zero.

Finally consider the graph in Fig. 3 with $d(G) = -4$. Here we are interested in the behavior of the corresponding amplitude $A(p_1,p_2,\eta p_3,m,\lambda\mu)$ for $\lambda \to 0$ and $\eta \to \infty$. First we note that the limit $\lambda \to 0$ cannot be rigorously taken, as G contains a divergent proper diagram g with its masses from the set $\{\mu\}$. Consider the amplitude $A(p_1/\lambda, p_2/\lambda, \eta p_3/\lambda, m/\lambda, \mu)$. It is easy to see that the asymptotic coefficients of λ^{-1} are $\alpha_I(S_2) = -4$ and $\beta_I(S_2) = 1$. On the other hand, the asymptotic coefficients of η are $\alpha_I(S_1) = 1$ and $\beta_I(S_1) = 0$. Accordingly,

$$\lim_{\lambda \to 0,\ \eta \to \infty} A(p_1,p_2,\eta p_3,m,\lambda\mu)$$
$$= O\left\{\lambda^{-4}\left(\frac{1}{\lambda}\right)^{-4} \frac{1}{\eta}\left[a_0 + a_1 \ln\left(\frac{1}{\lambda}\right)\right]\right\}$$
$$= O\left\{\frac{1}{\eta}\left[a_0 + a_1 \ln\left(\frac{1}{\lambda}\right)\right]\right\}. \tag{22}$$

The above examples, although elementary, are rich enough to demonstrate the rules in Sec. II at work.

[1] E. B. Manoukian, J. Math. Phys. **21**, 1218 (1980).
[2] E. B. Manoukian, J. Math. Phys. **19**, 917 (1978).
[3] E. B. Manoukian, Phys. Rev. D **14**, 966, 2202(E) (1976).
[4] S. Weinberg, Phys. Rev. **118**, 838 (1960).
[5] J. P. Fink, J. Math. Phys. **9**, 1389 (1968).

Simplified Bose description of para-Bose operators

M. Schmutz
Institut für Theorie der Kondensierten Materie, Universität Karlsruhe, 7500 Karlsruhe, West Germany

(Received 27 November 1979; accepted for publication 15 February 1980)

A set of n para-Bose operators of order p is explicitly constructed in terms of $n \times p$ Bose operators.

1. INTRODUCTION

The treatment of para-operators is intricate because the defining commutator relations are trilinear. Therefore, it is very convenient to use the so-called Green decomposition,[1,2] where the para-Fermi and para-Bose operators are formed as linear combinations of several different commuting Fermi operators and anticommuting Bose operators, respectively. This construction has the peculiarity that in either case the basic bilinear relations include commutation as well as anticommutation relations.

In recent contributions R. Jagannathan and R. Vasudevan[3,4] have given representations of para-Bose (para-Fermi) operators of any order entirely in terms of Bose (Fermi) operators. In the case of n para-Bose operators of order p these authors form the np operators which enter in the Green decomposition and which obey mixed commutation relations by the use of $np + [p/2]$ Bose operators, where $[p/2]$ denotes the integer part of $p/2$. Using a different principle of construction we show in this note that an alternative Bose description can be given where the formation of np operators with the desired mixed algebraic properties requires just np Bose operators.

In Sec. 2 we consider the case of a single para-Bose operator of aribitrary order p. In Sec. 3 we generalize our approach to the situation of an arbitrary number of para-Bose operators of order p.

2. SINGLE PARA-BOSE OPERATOR OF ORDER P

Let $\{a^{(\alpha)} | \alpha = 1,2,...,p\}$ be of a set of p Bose-operators $[a^{(\alpha)},a^{(\beta)+}] = \delta_{\alpha\beta}, [a^{(\alpha)},a^{(\beta)}] = 0,$

$$\alpha,\beta = 1,2,...,p. \tag{2.1}$$

Then the unitary operators $G^{(\alpha)}$ $(\alpha = 1,2,...,p)$ defined by

$$G^{(\alpha)} = \exp(i\pi a^{(\alpha)+} a^{(\alpha)}) \tag{2.2}$$

have the properties

$$G^{(\alpha)} = G^{(\alpha)+} = (G^{(\alpha)})^{-1}, \quad (G^{(\alpha)})^2 = 1, \tag{2.3}$$

and they satisfy the relations

$$[G^{(\alpha)},G^{(\beta)}] = 0, \tag{2.4}$$

$$\{a^{(\alpha)},G^{(\alpha)}\} = \{a^{(\alpha)+},G^{(\alpha)}\} = 0, \tag{2.5}$$

$$[a^{(\alpha)},G^{(\beta)}] = [a^{(\alpha)+},G^{(\beta)}] = 0, \quad \alpha \neq \beta. \tag{2.6}$$

We mention that according to Eq. (2.5) the unitary operator $G^{(\alpha)}$ generates the canonical transformation $a^{(\alpha)},a^{(\alpha)+} \rightarrow -a^{(\alpha)}, -a^{(\alpha)+}$.

From Eqs. (2.3)–(2.6) it follows immediately that the operators $\overline{G}^{(\alpha)}$ $(\alpha = 1,2,...,p)$ defined by

$$\overline{G}^{(1)} = 1, \quad \overline{G}^{(\alpha)} = \prod_{\nu=1}^{\alpha-1} G^{(\nu)}, \quad \alpha > 1, \tag{2.7}$$

obey the relations

$$\overline{G}^{(\alpha)} = \overline{G}^{(\alpha)+} = (\overline{G}^{(\alpha)})^{-1}, (\overline{G}^{(\alpha)})^2 = 1, [\overline{G}^{(\alpha)},\overline{G}^{(\beta)}] = 0, \tag{2.8}$$

$$\{a^{(\alpha)},\overline{G}^{(\beta)}\} = \{a^{(\alpha)+},\overline{G}^{(\beta)}\} = 0, \quad \alpha < \beta, \tag{2.9}$$

$$[a^{(\alpha)},\overline{G}^{(\beta)}] = [a^{(\alpha)+},\overline{G}^{(\beta)}] = 0, \quad \alpha \geqslant \beta. \tag{2.10}$$

If we now define the operators $\{b^{(\alpha)} | \alpha = 1,2,...,p\}$ by

$$b^{(\alpha)} = \overline{G}^{(\alpha)} a^{(\alpha)}, \tag{2.11}$$

we find from Eqs. (2.8) and (2.10) that

$$b^{(\alpha)+} = \overline{G}^{(\alpha)} a^{(\alpha)+}, \tag{2.12}$$

and from Eqs. (2.8)–(2.10) it follows that the operators $b^{(\alpha)}$ form a set of p anticommuting Bose operators

$$[b^{(\alpha)},b^{(\alpha)+}] = 1, [b^{(\alpha)},b^{(\alpha)}] = 0, \tag{2.13}$$

$$\{b^{(\alpha)},b^{(\beta)}\} = \{b^{(\alpha)},b^{(\beta)+}\} = 0, \quad \alpha \neq \beta. \tag{2.14}$$

Equation (2.13) is an immediate consequence of Eqs. (2.8) and (2.10). For a proof of Eq. (2.14) let us consider, for example, the anticommutator of the operators $b^{(\alpha)},b^{(\beta)+}$ with $\alpha \neq \beta$. Using Eqs. (2.8)–(2.12) we have

$$\begin{aligned} \{b^{(\alpha)},b^{(\beta)+}\} &= \overline{G}^{(\alpha)} a^{(\alpha)} \overline{G}^{(\beta)} a^{(\beta)+} + \overline{G}^{(\beta)} a^{(\beta)+} \overline{G}^{(\alpha)} a^{(\alpha)} \\ &= \begin{cases} -\overline{G}^{(\alpha)}\overline{G}^{(\beta)}[a^{(\alpha)},a^{(\beta)+}], & \alpha < \beta \\ +\overline{G}^{(\alpha)}\overline{G}^{(\beta)}[a^{(\alpha)},a^{(\beta)+}], & \alpha > \beta \end{cases} \\ &= 0, \quad \alpha \neq \beta. \end{aligned} \tag{2.15}$$

Following Green's ansatz we arrive at the desired para-Bose operator $A^{(p)}$ of order p

$$A^{(p)} = \sum_{\alpha=1}^{p} b^{(\alpha)}, \tag{2.16}$$

which satisfies the general para-Bose commutation relation for a single degree of freedom:

$$[A,\{A^+, A\}] = 2A. \tag{2.17}$$

3. ARBITRARY NUMBER OF PARA-BOSE OPERATORS OF ORDER P

If we now consider a set $\{a_j^{(\alpha)} | \alpha = 1,2,...,p; j = 1,2,...,n\}$ of np Bose operators

$$[a_j^{(\alpha)},a_k^{(\beta)+}] = \delta_{\alpha\beta}\delta_{jk}, \quad [a_j^{(\alpha)},a_k^{(\beta)}] = 0,$$

$$\alpha,\beta = 1,2,...,p; \quad j,k = 1,2,...,n, \tag{3.1}$$

we can define the operators analogous to those of Eq. (2.2) by

$$G_j^{(\alpha)} = \exp(i\pi a_j^{(\alpha)+} a_j^{(\alpha)}), \tag{3.2}$$

 0022-2488/80/071665-02$1.00

such that

$$G_j^{(\alpha)} = G_j^{(\alpha)+} = (G_j^{(\alpha)})^{-1}, (G_j^{(\alpha)})^2 = 1,$$
$$[G_j^{(\alpha)}, G_k^{(\beta)}] = 0, \tag{3.3}$$
$$\{a_j^{(\alpha)}, G_j^{(\alpha)}\} = \{a_j^{(\alpha)+}, G_j^{(\alpha)}\} = 0,$$
$$[a_j^{(\alpha)}, G_k^{(\beta)}] = [a_j^{(\alpha)+}, G_k^{(\beta)}] = 0, \ \alpha \neq \beta \text{ and/or } j \neq k.$$

If we arrange the Bose operators $a_j^{(\alpha)}$ into the n sets $\{a_j^{(\alpha)} | \alpha = 1,2,...,p\}$, we may apply to each set the procedure of Sec. 2. As a result we get the n sets of operators $\{\overline{b}_j^{(\alpha)} | \alpha = 1,2,...,p\}$ with

$$\overline{b}_j^{(\alpha)} = \overline{G}_j^{(\alpha)} a_j^{(\alpha)}, \tag{3.4}$$

$$\overline{G}_j^{(1)} = 1, \overline{G}_j^{(\alpha)} = \prod_{\nu=1}^{\alpha-1} G_j^{(\nu)}, \quad \alpha > 1. \tag{3.5}$$

According to Sec. 2 the operators $\overline{b}_j^{(\alpha)}, \overline{b}_j^{(\alpha)+}$ within each set, i.e., for fixed j, obey the mixed commutation relations (2.13) and (2.14) and operators from different sets commute by the construction

$$[\overline{b}_j^{(\alpha)}, \overline{b}_k^{(\beta)}] = [\overline{b}_j^{(\alpha)}, \overline{b}_k^{(\beta)+}] = 0, \quad j \neq k. \tag{3.6}$$

Although the operators $\overline{b}_j^{(\alpha)}, \overline{b}_j^{(\alpha)+}$ satisfy some sort of mixed commutation relations, these relations are not of the type which is necessary in order to form from these operators n para-Bose operators of order p by the use of Green's ansatz.

This deficiency can easily be cured by defining the np operators $b_j^{(\alpha)}$ by

$$b_j^{(\alpha)} = \overline{G}_{(n)}^{(\alpha)} a_j^{(\alpha)}, b_j^{(\alpha)+} = \overline{G}_{(n)}^{(\alpha)} a_j^{(\alpha)+}, \tag{3.7}$$

where the operators

$$\overline{G}_{(n)}^{(1)} = 1, \overline{G}_{(n)}^{(\alpha)} = \prod_{i=1}^{n} \overline{G}_i^{(\alpha)}$$
$$= \prod_{i=1}^{n} \prod_{\nu=1}^{\alpha-1} G_i^{(\nu)}, \quad \alpha > 1, \tag{3.8}$$

obey the relations [compare with (2.8)–(2.10)]

$$\overline{G}_{(n)}^{(\alpha)} = \overline{G}_{(n)}^{(\alpha)+} = (\overline{G}_{(n)}^{(\alpha)})^{-1},$$
$$(\overline{G}_{(n)}^{(\alpha)})^2 = 1, [\overline{G}_{(n)}^{(\alpha)}, \overline{G}_{(n)}^{(\beta)}] = 0, \tag{3.9}$$
$$\{a_j^{(\alpha)}, \overline{G}_{(n)}^{(\beta)}\} = \{a_j^{(\alpha)+}, \overline{G}_{(n)}^{(\beta)}\} = 0, \quad \alpha < \beta, \tag{3.10}$$
$$[a_j^{(\alpha)}, \overline{G}_{(n)}^{(\beta)}] = [a_j^{(\alpha)+}, \overline{G}_{(n)}^{(\beta)}] = 0, \quad \alpha \geqslant \beta. \tag{3.11}$$

It then follows from these relations that the operators $b_j^{(\alpha)}$ may be arranged into p sets $\{b_j^{(\alpha)} | j = 1,2,...,n\}$ of anticommuting Bose operators

$$[b_j^{(\alpha)}, b_k^{(\alpha)+}] = \delta_{jk}, [b_j^{(\alpha)}, b_k^{(\alpha)}] = 0,$$
$$\{b_j^{(\alpha)}, b_k^{(\beta)}\} = \{b_j^{(\alpha)}, b_k^{(\beta)+}\} = 0, \quad \alpha \neq \beta,$$
$$\alpha, \beta = 1,2,...,p; \quad j,k = 1,2,...,n. \tag{3.12}$$

Hence, by Green's ansatz we find the n para-Bose operators of order p

$$A_j^{(p)} = \sum_{\alpha=1}^{p} b_j^{(\alpha)}, \quad j = 1,2,...,n, \tag{3.13}$$

which satisfy the general para-Bose commutation relations

$$[A_i, \{A_j^+, A_k\}] = 2\delta_{ij} A_k,$$
$$[A_i, \{A_j, A_k\}] = 0, \tag{3.14}$$
$$i, j, k = 1,2,...,n.$$

The method we have used in order to construct from np Bose operators $a_j^{(\alpha)}, a_j^{(\alpha)+}$ a set of np operators which satisfy mixed commutation relations relies on the special properties (3.3) of the operators $G_j^{(\alpha)}$. If we compare the definitions (3.4), (3.5) and (3.7), (3.8) we see that the different types of mixed commutation relations of the operators $\overline{b}_j^{(\alpha)}, \overline{b}_j^{(\alpha)+}$ and $b_j^{(\alpha)}, b_j^{(\alpha)+}$ result from the use of different combinations of the operators $G_i^{(\nu)}$. When using other combination of these operators we obviously can form operators with other types of mixed algebraic properties. As a final remark we mention that the special properties of the operator G, associated with a Bose degree of freedom, play an essential role in the Klein transformation,[5] where a set of pure Fermi operators is constructed from a set of pure Bose operators. For a discussion of this point we refer to Ref. 6.

[1] H.S. Green, Phys. Rev. **90**, 270 (1953).
[2] O.W. Greenberg and A.M.L. Messiah, Phys. Rev. B **138**, 1155 (1965).
[3] R. Jagannathan and R. Vasudevan, J. Math. Phys. **19**, 1493 (1978).
[4] R. Jagannathan and R. Vasudevan, J. Math. Phys. **20**, 390 (1979).
[5] O. Klein, J. Phys. (USSR) **9**, 1 (1938).
[6] S. Naka, Prog. Theor. Phys. **59**, 2107 (1978).

Quadratic Hamiltonians in phase space and their eigenstates[a)]

M. Moshinsky[b)] and P. Winternitz
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada

(Received 12 January 1979; accepted for publication 10 August 1979)

The solution of the Schrödinger equation for a Hamiltonian H that is a general second order polynomial in the canonical coordinates q_i and momenta p_i is discussed. Examples of such Hamiltonians abound in various fields of physics and, e.g., the problem of small vibrations has been solved a long time ago. In the general case we first use linear canonical transformations [the group Sp(2n, R)] to reduce H to a simplified representative form H_R. Next we imbed each H_R into a complete set of commuting second order integrals of motion, making use of a recently obtained classification of maximal Abelian subalgebras of the algebra sp(2n,R). This imbedding is then used to separate variables in the representative Schrödinger equations and to obtain complete sets of their eigenfunctions. Finally the solutions of the representative equations can be transformed back into those of the original one making use of repreentations of the canonical transformations. The program is implemented fully for $n = 1,2$ and its structure is analyzed for arbitrary n. Special cases of interest are treated in detail, e.g., Hamiltonians for a system of n particles invariant under rotations, translations and permutations.

1. INTRODUCTION

The purpose of this article is to discuss the problem of finding the solutions of the Schrödinger equation

$$\mathscr{H}\psi(q_1,...,q_n) = E\psi(q_1,...,q_n), \tag{1}$$

where $\mathscr{H}$ is a quite general quadratic Hamiltonian in phase space, i.e.,

$$\mathscr{H} = z^T H z, \tag{2}$$

where $z^T = (q_1,...,q_n, p_1,...,p_n)$ and $H = H^T$ a real symmetric $2n \times 2n$ matrix (the superscript T denotes transposition) which we write in the form

$$H = \begin{pmatrix} C & A^T \\ A & B \end{pmatrix}, \quad B = B^T, \quad C = C^T. \tag{3}$$

This type of problem makes its appearance in many fields of physics. The simplest example is that of the harmonic oscillator,[1] which is of fundamental importance in many body physics, in particular nuclear theory. Classical[2] and quantum mechanical[3] problems involving small vibrations provide further examples that have been solved using well known methods. More general Hamiltonians of the type (1)–(3) occur in problems of molecular physics and quantum chemistry[4] either directly or as the result of suitable approximations. A study of nuclear forces and nuclear structure[5] also leads to quadratic Hamiltonians, as do investigations of the motion of electrons in magnetic fields.[6] Quadratic Hamiltonians in canonical coordinates and momenta arise in field theories[7]; the problem of constructing, for example, vacuum states is then isomorphic to the problem stated above.

It should also be mentioned that solvable models are hard to come by in nonrelativistic and relativistic quantum theory. A general solution of problems (1)–(3) would provide us with entire families of solvable problems, complementing the harmonic oscillator, Coulomb potential, and a few others that are known.

Our approach involves the following mathematical steps.

(1) We apply the real linear canonical transformations[8-10] Sp(2n,$\mathbb{R}$) to classify the matrices $X = KH$, with H as in Eq. (3) and

$$K = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \tag{4}$$

into orbits and then choose a representative element X_R on each orbit. Each of these elements provides us with a representative Hamiltonian $\mathscr{H}_R$. Any Hamiltonian of type (1) can then be transformed into one of the representative Hamiltonians by a canonical transformation and these $\mathscr{H}_R$ can be chosen in relatively simple forms. As a mathematical problem this amounts to a classification of elements of the symplectic Lie algebra sp(2n,$\mathbb{R}$) into conjugacy classes, a problem that has received much attention in the literature and been essentially completely solved[8,11-14].

(2) We classify all maximal Abelian subalgebras (MASA) of sp(2n,$\mathbb{R}$) into conjugacy classes under the group Sp(2n,$\mathbb{R}$) and choose a representative subalgebra D_R for each class. This makes it possible to imbed each matrix X_R into one (or more) MASA D_R, i.e., to imbed each Hamiltonian $\mathscr{H}_R$ into a complete set of commuting second order (in phase space) integrals of motion. Much work has been done on the problem of classifying commuting sets of matrices,[15] in particular on MASA of semisimple Lie algebras[16,17]. Considerable progress has been achieved in this direction quite recently[18,19], partly in connection with the present paper.

(3) We solve the Schrödinger equation (1) for each representative Hamiltonian $\mathscr{H}_R$. Each embedding of the Hamiltonian $\mathscr{H}_R$ into a complete set of commuting second or-

[a)]Work supported in part by the National Research Council of Canada, Ministère de l'éducation du gouvernement du Québec, and the Canada Council.

[b)]Permanent address: Instituto de Fisica, UNAM, Apdo Postal 20364, Mexico 20, D.F. Member of the Instituto Nacional de Energia Nuclear and El Colegio Nacional.

 0022-2488/80/071667-16$1.00

der operators D_R provides us with a means of separating variables in the corresponding Schrödinger equation. If several different embeddings of $\mathscr{H}_R$ are possible, then several different separable coordinate systems will exist. The connection between the separation of variables in partial differential equations and sets of commuting second order operators has been the subject of numerous publications.[20-27]

(4) We construct the explicit Sp(2n,$\mathbb{R}$) transformation g that takes a given Hamiltonian $\mathscr{H}$ into the corresponding representative Hamiltonian $\mathscr{H}_R$ and the matrix elements of the operator T_g, representing g in the space of wave functions. These matrix elements make it possible to express the eigenfunctions of the Hamiltonian (1) in terms of those of a representative Hamiltonian as linear combinations with known coefficients (in some cases these will be continuous linear combinations, i.e., integral representations with known kernels). The construction of g and T_g is a standard problem of group representation theory which in general has, however, not been solved.

To put the results of this paper into a somewhat broader context, they should be viewed as a meeting point of three different but related research programs. These are (1) a systematic study and application of groups of canonical transformations in quantum mechanics,[10,20,28-33] (2) a systematic study of the subgroup structure of Lie groups,[18-20,34-36] (3) a systematic study of the separation of variables in partial differential equations.[20-27]

In Sec. 2 we present some of the general theory involved in the problem for $2n$ dimensional phase space. Section 3 is devoted to the simple case when $n = 1$ and the program can be implemented completely. The case of two degrees of freedom ($n = 2$) is analyzed in Sec. 4. Comments on the case of $n = 3$ are also presented in this section. In Secs. 5 and 6 we return to interesting special cases in $2n$ dimensional phase space, namely, to the problem of small vibrations on the one hand and to that of quadratic Hamiltonians invariant under rotations, translations, and permutations on the other.

2. GENERAL COMMENTS ON QUADRATIC HAMILTONIANS FOR SYSTEMS WITH n DEGREES OF FREEDOM

Let us now discuss in somewhat more detail the individual steps outlined in the Introduction by means of which it is possible to solve Eq. (1).

A. Classification of elements of sp(2n,$\mathbb{R}$)

Matrices $X = KH$ with H as in Eq. (3) and K as in Eq. (4) can be written as

$$X = \begin{pmatrix} A & B \\ -C & -A^T \end{pmatrix}, \quad A \in \mathrm{gl}(n,\mathbb{R}),$$
$$B = B^T, \quad C = C^T, \quad B,C \in \mathbb{R}^{n\times n}. \tag{5}$$

They satisfy the equation

$$XK + KX^T = 0 \tag{6}$$

and hence $X \in \mathrm{sp}(2n,\mathbb{R})$ (the same would be true if K were some other real antisymmetric $2n \times 2n$ matrix and other choices of K are convenient in certain cases). We use sp(2n,$\mathbb{R}$) to denote the Lie algebra of the symplectic group Sp(2n,$\mathbb{R}$).

The classification of matrices X into conjugacy classes with respect to the group Sp(2n,$\mathbb{R}$) is based on several simple results of linear algebra.[14]

(i) Let V be a real $2n$ dimensional vector space, and let X be a real matrix acting on V. Let $\lambda_1,\ldots,\lambda_n$ and $V_1,\ldots,V_r$ be the (generalized) eigenvalues and eigenspaces of X, i.e.,

$$(X - \lambda_i I)^{m_i} v = 0, \quad v \in V_i, \quad m_i \geqslant 1 \tag{7}$$

(m_i are integers). Then,

$$V = \sum_{i=1}^{r} \oplus V_i, \tag{8}$$

each V_i is invariant under X, and

$$\det(\lambda I - X) = \prod_{i=1}^{r} (\lambda - \lambda_i)^{d_i}, \quad d_i = \dim V_i. \tag{9}$$

This result makes it possible to reduce any matrix to a block diagonal form, each block representing one indecomposable component.

(ii) Now let $X \in \mathrm{sp}(2n,\mathbb{R})$ and let us follow Ref. 14. If λ_i is an eigenvalue, so are $-\lambda_i, \lambda_i^*$, and $-\lambda_i^*$ and they all have the same multiplicity (the star denotes complex conjugation).

(iii) For $X \in \mathrm{sp}(2n,\mathbb{R})$ the invariant subspaces V_i are mutually orthogonal with respect to the invariant symplectic bilinear form

$$f^T K g = 0, \quad f \in V_i, \ g \in V_k, \ \lambda_i \neq \pm\lambda_k, \pm\lambda_k^*. \tag{10}$$

A consequence of these theorems is that a matrix $X \in \mathrm{sp}(2n,\mathbb{R})$ must belong to one of the following classes.[14]

1. Orthogonally indecomposable

Then one of four possibilities can occur:

(a) X has four distinct complex eigenvalues $\mu + i\nu$, $\mu - i\nu$, $-\mu - i\nu$, $-\mu + i\nu$, $\nu\mu \neq 0$, each with multiplicity $n/2$ (this can only happen for n even).

(b) X has two distinct real eigenvalues a, $-a$, $a > 0$, each with multiplicity n.

(c) X has two distinct mutually conjugate pure imaginary eigenvalues $\pm ib$, $b > 0$, each with multiplicity n.

(d) X has a single eigenvalue 0 with multiplicity $2n$.

In each of the above cases the matrix X can be reduced by a Sp(2n,$\mathbb{R}$) transformation to one of several standard forms, characterized by eigenvalues of X and in some cases by certain additional invariants. This has been studied extensively in the literature[11-14] and we shall not reproduce the details here.

2. Orthogonally decomposable

If the $2n$ dimensional real space V can be decomposed into two or more mutually orthogonal invariant subspaces V_i as in Eqs. (8) and (10), then the matrix X can be reduced to a block diagonal form, such that each block is one of the standard orthogonally indecomposable matrices discussed above. This is best done in a realization of sp(2n,$\mathbb{R}$) in which the matrix K of Eq. (6) is itself written in a decomposed form:

$$K=\begin{pmatrix} K_1 & & & \\ & K_2 & & \\ & & \ddots & \\ & & & K_p \end{pmatrix},\quad K_i=\begin{pmatrix} 0 & I_{n_i} \\ -I_{n_i} & 0 \end{pmatrix}, \tag{11}$$

where $1\leqslant n_1\leqslant n_2\leqslant\ldots\leqslant n_p$, $n_1+n_2+\ldots+n_p=n$, and all n_i are positive integers. We can then choose the representative

$$X=\begin{pmatrix} X_{n_1} & & & \\ & X_{n_2} & & \\ & & \ddots & \\ & & & X_{n_p} \end{pmatrix}, \tag{12}$$

with each X_{n_i} in a representative indecomposable form.

B. Classification of maximal Abelian subalgebras of sp(2n,$\mathbb{R}$)

The above classification of sp(2n,$\mathbb{R}$) matrices and Hamiltonians allows us to transform the Schrödinger equation into a simpler form, not however necessarily always simple enough to allow for an analytical solution. The problem can be further simplified by constructing a complete set of operators $\mathscr{R}_a$, commuting with the Hamiltonian $\mathscr{H}$ and among each other and such that each $\mathscr{R}_a$ is a quadratic homogeneous polynomial in the dynamical variables q_i and p_i. The Hamiltonian can then be expressed as

$$\mathscr{H}=\sum_{a=1}^{k} t_a\mathscr{R}_a, \tag{13}$$

where t_a are real constants and any solution of the Schrödinger equation can be written as a linear combination of the solutions of the set of equations

$$\mathscr{R}_a\psi(q_1,\ldots,q_n)=E_a\psi(q_1,\ldots,q_n),\quad 1\leqslant a\leqslant k. \tag{14}$$

Since the set of all operators that are homogeneous quadratic polynomials in q_i and p_i forms a basis of the Lie algebra sp(2n,$\mathbb{R}$), the problem of finding all possible commuting sets of operators $\{R_a\}$ is equivalent to classifying all MASA of sp(2n,$\mathbb{R}$) into conjugacy classes under Sp(2n,$\mathbb{R}$).

The problem of classifying commuting sets of matrices, in particular matrices of the form (5), is a complicated one that has received much attention in the literature.[15-19] It is important in this connection to realize that a maximal Abelian subalgebra of a semisimple Lie algebra need by no means necessarily be a Cartan subalgebra. A Cartan subalgebra of a semisimple Lie algebra, in addition to being maximal Abelian, contains no element that can be represented in a finite dimensional representation by a nilpotent matrix N (i.e., such that $N^P=0$ for some positive integer P).

It has recently been shown[18,19] that the algebra sp(2n,$\mathbb{R}$) can have the following types of MASA:

1. *Maximal Abelian nilpotent subalgebras* (MANS)

In any finite dimensional representation, in particular the defining representation of sp(2n,$\mathbb{R}$), a MANS is represented by nilpotent matrices. A MANS is indecomposable and in an appropriate realization can be reduced to the form[19]

$$X=\begin{pmatrix} 0_\lambda & B & D & C \\ 0 & R & S & J_\mu D^T \\ 0 & 0 & -JR^TJ & -J_\mu B^T \\ 0 & 0 & 0 & 0_\lambda \end{pmatrix},$$

$$J_\mu=\begin{pmatrix} & & & 1 \\ & & \iddots & \\ & \iddots & & \\ 1 & & & \end{pmatrix},\quad J_\mu\in R^{\mu\times\mu}, \tag{15}$$

where 0_λ is a $\lambda\times\lambda$ zero matrix, $C=C^T\in R^{\lambda\times\lambda}$, $S=JS^TJ\in\mathbb{R}^{\mu\times\mu}$, $R\in\mathbb{R}^{\mu\times\mu}$ is nilpotent, and $B,D\in\mathbb{R}^{\lambda\times\mu}$. Each MANS in Eq. (15) corresponds to a definite Kravchuk signature,[15,18,19] i.e., a triplet of nonnegative integers $(\lambda, 2\mu, \lambda)$ with $\lambda+\mu=n$, $1\leqslant\lambda\leqslant n$. Each Kravchuk signature must be considered separately and the further constraints on the entries in Eqs. (15), as well as the classification of these algebras into conjugacy classes, must be discussed for each allowed signature[18,19].

2. *Indecomposable nonnilpotent maximal Abelian subalgebras*

These only exist when n is even. They must contain precisely one nonnilpotent element which generates a compact one parameter subgroup of Sp(2n,$\mathbb{R}$) and can be written as

$$C=\begin{pmatrix} 0 & 1 & & & & \\ -1 & 0 & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & 0 & 1 \\ & & & & -1 & 0 \end{pmatrix} \tag{16}$$

in the realization (6) of sp(2n,$\mathbb{R}$). The matrices $X\in$sp(2n,$\mathbb{R}$). The matrices $X\in$sp(2n,$\mathbb{R}$) satisfying $[X,C]=0$ form an algebra isomorphic to su($n/2$,$n/2$) and the corresponding MASA are obtained by constructing all maximal Abelian nilpotent subalgebras of su($n/2$,$n/2$) (see Refs. 18 and 19).

3. *Orthogonally decomposable* MASA

These are best written in the realization (11) in which they have the form

$$S=\begin{pmatrix} A_{n_1} & & & & \\ & A_{n_2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & A_{n_p} \end{pmatrix}, \tag{17}$$

where each A_{n_i} is a maximal Abelian subalgebra of $\mathrm{sp}(2n_i,\mathbb{R})$ and at most one of the A_{n_i} is a maximal Abelian nilpotent algebra. Two algebras that differ by a permutation of the A_{n_i} only are equivalent and one of them must be eliminated from the list.

4. Decomposable but not orthogonally decomposable MASA

These are best written in the realization (5) in which they can be reduced to the form

$$X=\begin{pmatrix} A & 0 \\ 0 & -A^T \end{pmatrix}, \tag{18}$$

where A is an irreducible MASA of $\mathrm{gl}(n,\mathbb{R})$.

We see that the classification of MASA of $\mathrm{sp}(2n,\mathbb{R})$, $\mathrm{su}(n/2,n/2)$, and $\mathrm{gl}(n,\mathbb{R})$ are intimately related. The case of $\mathrm{sp}(4,\mathbb{R})$ will be treated in detail in Sec. 4 and that of $\mathrm{sp}(6,\mathbb{R})$ will be sketched there.

C. Separation of variables and solution of the Schrödinger equation

Given a Hamiltonian $\mathscr{H}$, we now write it in the form (13), where $R_a\,(1\leqslant a\leqslant k)$ is the basis for a MASA (k is the dimension of this MASA and need not in general be equal to the rank of the algebra). If the subalgebra $\{R_a\}$ is orthogonally decomposable into 2×2 blocks, then the equations (14) allow separation in Cartesian coordinates. Other cases can be associated with other types of separable coordinates, as will be seen below for $\mathrm{sp}(4,\mathbb{R})$. If more than one imbedding is possible, then more then one separable coordinate system exists.

D. Transformation of the solution of the representative problem into the solution of the original problem

In order to relate the solutions of Eq. (1) to the obtained solutions of the "representative" problem (14), we must find the transformation relating the two Hamiltonian matrices

$$gX_Rg^{-1}=X,\quad g\in \mathrm{Sp}(2n,\mathbb{R}), \tag{19}$$

and then construct the unitary operators T_g representing the group of real linear canonical transformations g in the space of wave functions.

In order to construct g, we can proceed in several steps:

(i) Establish which class X belongs to, i.e., determine whether it is orthogonally decomposable or indecomposable, what are its eigenvalues and eigenspaces, etc.

(ii) Write g in the form

$$g=\tilde{g}_1g_2g_1 \tag{20}$$

such that

$$g_1f_R=f_R,\ g_2f_R=f,\ \tilde{g}_1f=f. \tag{21}$$

Here f_R is a set of basis vectors for the eigenspaces of X_R, and f a set of basis vectors for the eigenspaces of X.

The subgroups g_1 and $\tilde{g}_1$ are mutually conjugate (and hence isomorphic) and are the little groups of f_R and f, respectively. This representation is not necessarily unique and different forms of g_2 may be needed in order to obtain all elements $g\in\mathrm{Sp}(2n,\mathbb{R})$ in the form (20).

The final step is to use Eq. (20) to calculate the representation T_g of g explicitly in terms of the representations of g_1, $\tilde{g}_1$, and g_2. In general, it may be quite difficult to construct the representation T_g and this is in itself an important problem. In the case of $n=1$ this has recently been solved completely[10,29,37] and we reproduce the results in Sec. 3. These $n=1$ results are also relevant for many body problems involving identical particles as we show in Sec. 6.

Actually, in many physical applications the explicit form of T_g is not needed. Indeed, if we are only interested in eigenvalues and matrix elements of relevant physical operators, then these can be calculated directly in the representative basis.

3. QUADRATIC HAMILTONIANS FOR ONE DEGREE OF FREEDOM

For $n=1$ the entire program simplifies considerably. The Hamiltonian (1) in this case is

$$\mathscr{H}=a(pq+qp)+bp^2+cq^2, \tag{22}$$

where a, b, and c are real constants. The matrices H and $X=KH$ are

$$H=\begin{pmatrix} c & a \\ a & b \end{pmatrix},\quad X=\begin{pmatrix} a & b \\ -c & -a \end{pmatrix}. \tag{23}$$

All maximal Abelian subalgebras of $\mathrm{sp}(2,\mathbb{R})$ are one dimensional, and hence the classifications of elements of $\mathrm{sp}(2,\mathbb{R})$ [step (1) of our program] and of MASA [step (2)] coincide. Elements of $\mathrm{sp}(2,\mathbb{R})$ are completely characterized by the eigenvalues of the corresponding matrix X, equal to $\lambda=\pm(a^2-bc)^{1/2}=\pm(\Delta)^{1/2}$ ($\Delta=-\det X$). For $\Delta>0$, X is decomposable (but not orthogonally); for $\Delta=0$, X is indecomposable and nilpotent; for $\Delta<0$, X is indecomposable nonnilpotent. Using a canonical transformation $g\in\mathrm{Sp}(2,\mathbb{R})$ (the form of g will be specified below), we can transform X to one of three canonical forms

$$X_{-1}=\begin{pmatrix} 0 & \sqrt{|\Delta|} \\ \sqrt{|\Delta|} & 0 \end{pmatrix},\quad X_0=\pm\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$
$$X_1=\pm\begin{pmatrix} 0 & \sqrt{|\Delta|} \\ -\sqrt{|\Delta|} & 0 \end{pmatrix} \tag{24}$$

for $\Delta>0$, $\Delta=0$, and $\Delta<0$, respectively. Let us comment here that the choice (24) of representatives of $\mathrm{Sp}(2,\mathbb{R})$ classes of elements is somewhat arbitrary. Alternative and useful choices of X_{-1} and X_0 would be, for example,

$$\tilde{X}_{-1}=\begin{pmatrix} \sqrt{|\Delta|} & 0 \\ 0 & -\sqrt{|\Delta|} \end{pmatrix},\quad \tilde{X}_0=\pm\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The Schrödinger equation for the three cases (24) can be

written as

$$\mathscr{H}_\tau\psi^\tau(q)=\kappa_\tau(p^2+\tau q^2)\psi^\tau(q)=\nu\psi^\tau(q),\quad \tau=-1,0,1, \tag{25}$$

where

$$\kappa_{-1}=|\Delta|^{1/2},\quad \kappa_0=\pm 1,\quad \kappa_{+1}=\pm|\Delta|^{1/2}. \tag{26}$$

Had we chosen $\tilde{X}_{-1}$ and $\tilde{X}_0$ instead of X_{-1} and X_0, respectively, the equations (25) would be replaced by[37]

$$|\Delta|^{1/2}(pq+qp)\tilde{\psi}^{-1}(q)=\nu\tilde{\psi}^{-1}(q),$$
$$\pm q^2\tilde{\psi}^0(q)=\nu\tilde{\psi}^0(q).$$

Step (3) consists of solving the representative equations (25). While the sign of κ_τ for $\tau=0$, 1 cannot be changed by a real linear canonical transformation, it makes sense, from the physical point of view, to assume that the kinetic energy is positive definite. This fixes the sign to be plus in both cases. The solutions of Eq. (25) are of course well known in all three cases, namely, the wave functions of a repulsive harmonic oscillator (r.h.o.), free particle (f.p.), and harmonic oscillator (h.o.) for $\tau=-1$, 0, and 1, respectively.[37] Step (4) consists of calculating the representations of the Sp(2,$\mathbb{R}$) canonical transformations in the r.h.o., f.p., and h.o. bases. These again are known.[10,29,37]

Instead of presenting the solutions and transformation matrices directly, we consider a somewhat more general problem. Indeed, let us replace the coordinate q and momentum p in Eq. (22) by the 3 vectors $\mathbf{q}$ and $\mathbf{p}$ and consider a quadratic Hamiltonian that is invariant under $O(3)$ rotations:

$$\tilde{\mathscr{H}}=a(\mathbf{qp}+\mathbf{pq})+b\mathbf{p}^2+c\mathbf{q}^2. \tag{27}$$

As above, this Hamiltonian can be transformed to one of three standard forms. Explicitly, this is achieved by a rotation in phase space

$$q_i=q_i'\cos\alpha+p_i'\sin\alpha,\quad i=1,2,3,$$
$$p_i=-q_i'\sin\alpha+p_i'\cos\alpha, \tag{28}$$

with $\tan 2\alpha=2a/(b-c)$, diagonalizing the quadratic form (27). The determinant $\Delta=-a^2+bc$ is invariant under this rotation. A further dilation belonging to Sp(2,$\mathbb{R}$), $q_i'=|\Delta|^{1/2}q_i''$, $p_i'=|\Delta|^{-1/2}p_i''$, accompanied, if necessary, by a rotation through $\pi/2$, will reduce $\tilde{\mathscr{H}}$ of Eq. (27) to one of the Hamiltonians

$$\tilde{\mathscr{H}}_\tau=\kappa_\tau(\mathbf{p}^2+\tau\mathbf{q}^2),\quad \tau=\pm 1,0, \tag{29}$$

with κ_τ as in Eq. (26). Thus, the spherically symmetric Hamiltonian (27) has been reduced to that of the harmonic oscillator, repulsive harmonic oscillator, or free particle in Euclidean three space.

Writing $\mathbf{q}$ in spherical coordinates (r,θ,ϕ), we can express the eigenfunctions of the operator (29) as

$$r^{-1}\phi_\nu^{l\sigma\tau}(r)Y_{lm}(\theta,\phi), \tag{30}$$

where $\phi_\nu^{l\sigma\tau}(r)$ satisfies

$$\left(-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+\tau r^2\right)\phi_\nu^{l\sigma\tau}(r)=\nu\phi_\nu^{l\sigma\tau}(r) \tag{31}$$

and we have added an index $\sigma=\pm$ to indicate parity with respect to the change $r\to -r$. For the three dimensional problem only the $\sigma=-$ is relevant, i.e., the wave function must vanish at $r=0$, but for the one dimensional problem, which corresponds to $l=0$, both values $\sigma=\pm$ are possible.

We take for the normalized $\phi_\nu^{l\sigma\tau}(r)$ the phases used in Ref. 37 and thus we have for $\tau=+$, $\sigma=-$ that $\nu=2n+l+3/2$ and replacing the index ν in $\phi_\nu^{l\sigma\tau}(r)$ of Eq. (31) by n, we obtain

$$\phi_n^{l-+}(r)=\left(\frac{2(n!)}{\Gamma(n+l+3/2)}\right)^{1/2}r^{l+1}\times\exp(-r^2/2)L_n^{l+1/2}(r^2), \tag{32a}$$

where L is a Laguerre polynomial[38]. For $\sigma=+$, we are only interested in $l=0$ and we have

$$\phi_n^{0++}(r)=\left(\frac{2(n!)}{\Gamma(n+1/2)}\right)^{1/2}\exp(-r^2/2)L_n^{-1/2}(r^2), \tag{32b}$$

where, as is well known[38], the Laguerre polynomials appearing in Eq. (32a) for $l=0$ and in Eq. (32b) are actually Hermite polynomials of odd and even order, respectively.

Passing now to $\tau=-$, $\sigma=-$, the eigenvalue ν can take all real values in the range $-\infty<\nu<\infty$ and then[37]

$$\phi_\nu^{l--}(r)=C_{\nu l}r^{-1/2}M_{i\nu/2,(1/2)[l+(1/2)]}(-ir^2), \tag{32c}$$

where M is a Whittaker function[38] and the normalization constant $C_{\nu l}$ takes the value

$$C_{\nu l}=(2\pi)^{-1/2}\exp i\{\tfrac{1}{2}\pi[(l+3/2)-(i\nu/2)]+(\nu/2)\ln 2\}\Gamma\left[\frac{1}{2}\left(l+\frac{3}{2}\right)+\frac{i\nu}{2}\right]\Big/\Gamma\left(l+\frac{3}{2}\right).$$

For $\sigma=+$, we are again only interested in $l=0$ and we get

$$\phi_\nu^{0+-}(r)=\frac{1}{\sqrt{2\pi}}\exp i\left[\frac{1}{2}\pi\left(\frac{1}{2}-\frac{i\nu}{2}\right)+\frac{\nu}{2}\ln 2\right]\times\Gamma\left(\frac{1}{4}+\frac{i\nu}{2}\right)M_{i\nu/2,-1/4}(-ir^2). \tag{32d}$$

Finally, for $\tau=0$, $\sigma=-$, the eigenvalue ν takes all real positive values $0\leqslant\nu\leqslant\infty$ and then[37]

$$\phi_\nu^{l-0}(r)=\exp\left[i\frac{\pi}{2}\left(l+\frac{3}{2}\right)\right](\nu r)^{1/2}J_{l+1/2}(\nu r), \tag{32e}$$

where J is a Bessel function. For $\sigma=+$, $l=0$, we get

$$\phi_\nu^{0+0}=\exp(i\pi/4)(\nu r)^{1/2}J_{-1/2}(\nu r)=\sqrt{\frac{2}{\pi}}\exp(i\pi/4)\cos\nu r. \tag{32f}$$

We need now the representation on the states (30) of a canonical transformation that affects in the same way all three components of the $\mathbf{q}$,$\mathbf{p}$ vectors, i.e.,

$$S=\begin{pmatrix}tI & uI\\ vI & wI\end{pmatrix},\quad tw-uv=1, \tag{33}$$

where I is a 3×3 matrix. Obviously, S commutes with the rotations $O(3)$ in this vector space and so it will not affect the angular part $Y_{lm}(\theta,\phi)$. Thus, what is required is

$$P_S\phi_\nu^{l\sigma\tau}(r)=\int\phi_{\nu'}^{l\sigma\tau}(r)\mathscr{D}_{\nu'\nu}^{l\sigma\tau}(S)d\nu', \tag{34}$$

where the representation is now also characterized by the

angular momentum l and the parity $\sigma = \pm$ and, when $\tau = +$, the integration becomes a sum over the discrete index n'.

The representations for $\tau = +, 0$ were derived in previous publications[10,29] and those for $\tau = -$ are given in Ref. 37, where also the results for $\tau = +, 0$ are summarized. From this last reference we get for $\tau = +$:

$$\mathscr{D}^{l\sigma+}_{n'n}(S) = 2^{1+\sigma[l+(1/2)]}\Gamma[n'+n+1+\sigma(l+\tfrac{1}{2})]$$
$$\times\Big(\{n'!n!\Gamma[n'+1+\sigma(l+1/2)]\Gamma[n+1+\sigma(l+1/2)]\}^{-1/2}$$
$$\times[(w-t)-i(u+v)]^{n'}[(t-w)-i(u+v)]^{n}[(t+w)+i(u-v)]^{-n'-n-1-\sigma[l+(1/2)]}$$
$$\times {}_2F_1\left[-n',-n,-n'-n-\sigma\left(l+\frac{1}{2}\right);\frac{t^2+u^2+v^2+w^2+1}{t^2+u^2+v^2+w^2-1}\right]\Big), \tag{35a}$$

where ${}_2F_1$ is a hypergeometric function[38] that reduces to a polynomial of the argument and whose degree is $\min(n,n')$.

For $\tau = -$, we in turn get

$$\mathscr{D}^{l\sigma-}_{\nu'\nu}(S) = \Gamma(\alpha)\Gamma(\alpha')\{4\pi\Gamma[1+\sigma(l+\tfrac{1}{2})]\}^{-1}$$
$$\times t^{-\alpha}w^{-\alpha'}(2iu)^{i(\nu-\nu')/2}$$
$$\times {}_2F_1[\alpha',\alpha;1+\sigma(l+\tfrac{1}{2});(tw)^{-1}], \tag{35b}$$

where

$$\alpha = \frac{1}{2} + \frac{\sigma}{2}\left(l+\frac{1}{2}\right) + \frac{i\nu}{2},$$
$$\alpha' = \frac{1}{2} + \frac{\sigma}{2}\left(l+\frac{1}{2}\right) - \frac{i\nu'}{2}.$$

The expression (35b) is valid for $tw > 1$ but its analytic continuation to all real values t, u, v, and w restricted by $tw - uv = 1$ presents no problem.[37]

Finally, for $\tau = 0$, we obtain

$$\mathscr{D}^{l\sigma 0}_{\nu'\nu} = \exp\left\{-\frac{i\pi}{2}[1+\sigma(l+\tfrac{1}{2})]\right\}u^{-1}(\nu'\nu)^{1/2}$$
$$\times\exp[(i/2u)(t\nu^2+w\nu'^2)]J_{\sigma(l+1/2)}\left(\frac{\nu'\nu}{u}\right), \tag{35c}$$

which is valid for $u \neq 0$, i.e., $v = u^{-1}(tw-1)$, but it has a well-defined limit[29,37] even when $u \to 0$.

Note that in all cases $\tau = \pm, 0$ we are interested in arbitrary integer l for $\sigma = -$, but only in $l = 0$ for $\sigma = +$.

We see that for quadratic Hamiltonians with one degree of freedom we are able to implement fully the program outlined in the previous sections. Though our results are valid only for one particle rotational invariant Hamiltonians in phase space, we shall show in Sec. 6 that they also apply to n identical particle Hamiltonians which are then invariant under permutations of the indices.

In the next section we implement most of our program for the case $n = 2$ and in the process face up to some of the difficulties that are also present in the problem for an arbitrary n.

4. QUADRATIC HAMILTONIANS IN PHASE SPACE FOR TWO DEGREES OF FREEDOM

For $n = 2$ we shall need several different realizations of the algebra sp(4,$\mathbb{R}$), corresponding to different choices of K in Eq. (6). We discuss these realizations in the Appendix where we also give a realization of the isomorphism[39] between the algebras sp(4,$\mathbb{R}$) and $o(3,2)$. This is useful in the present context, since all subalgebras of $o(3,2)$ have recently been classified[36] and the classification can be used here.

A convenient basis for sp(4,$\mathbb{R}$) is provided by a set of 10 matrices $A_1,\ldots,A_4;B_1,\ldots,B_3;C_1,\ldots,C_3$. In each realization discussed in the Appendix these basis matrices are obtained by setting all entries in the sp(4,$\mathbb{R}$) matrices equal to zero, except the one labeled by the same letter as the basis generator. This entry is set equal to 1. Thus, for example, A_1 is obtained by seting $a_1 = 1$ in Eq. (A3), (A5), (A7), or (A9) and all other entries equal to zero. These generators can be identified with dynamical variables, using the relations (2) and (3). Indeed, it is easy to check that the quadratic operators

$$A_1 = -\frac{i}{2}(q_1p_1 + p_1q_1), \quad A_3 = -iq_2p_1,$$
$$A_2 = -\frac{i}{2}(q_2p_2 + p_2q_2), \quad A_4 = -iq_1p_2,$$
$$B_1 = -\frac{i}{2}p_1^2, \quad C_1 = -\frac{i}{2}q_1^2, \tag{36}$$
$$B_2 = -\frac{i}{2}p_2^2, \quad C_2 = -\frac{i}{2}q_2^2,$$
$$B_3 = -ip_1p_2, \quad C_3 = -iq_1q_2$$

(where $p_k = -i\partial/\partial q_k$) satisfy the same commutation relations as the matrices A_i, B_i, and C_i.

Let us now follow the general procedure outlined in Sec. 2.

A. Classification of elements of sp(4,$\mathbb{R}$)

1. Orthogonally decomposable elements

The only allowed partition of 4 is $4 = 2 + 2$ so that Eq. (12) reduces to

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}, \quad X_i \in \mathrm{sp}(2,\mathbb{R}) \tag{37}$$

[see realization (A5)]. The one dimensional subalgebras of sp(2,$\mathbb{R}$) are well known and were discussed in Sec. 3. The six possible choices for the pair (X_1, X_2) lead to six possible types of Hamiltonians (in this section we lump together Hamiltonians differing by a nonzero multiplicative constant):

$$\mathscr{H}_1^\lambda = p_1^2 - q_1^2 + \lambda\,(p_2^2 - q_2^2), \quad 0 \leqslant \lambda \leqslant 1,$$
$$\mathscr{H}_2^\lambda = p_1^2 + q_1^2 + \lambda\,(p_2^2 - q_2^2), \quad 0 < \lambda < \infty,$$
$$\mathscr{H}_3^\lambda = p_1^2 + q_1^2 + \lambda\,(p_2^2 + q_2^2), \quad -1 \leqslant \lambda \leqslant 1,$$
$$\mathscr{H}_4 = p_1^2 + p_2^2 - q_2^2,$$
$$\mathscr{H}_5^\epsilon = p_1^2 + q_1^2 + \epsilon p_2^2, \quad \epsilon = \pm 1,$$
$$\mathscr{H}_6^\kappa = p_1^2 + \kappa p_2^2, \quad \kappa = \pm 1, 0. \tag{38}$$

Thus, orthogonally decomposable sp(2,ℝ) matrices correspond to Hamiltonians that can be written as

$$\mathscr{H} = \mathscr{H}_1 + \mathscr{H}_2, \tag{39}$$

where each $\mathscr{H}_i$ describes a one dimensional harmonic oscillator, repulsive harmonic oscillator, or free particle, moving in the ith direction.

2. Orthogonally indecomposable elements

In this case we use the realization (A3) of sp(4,ℝ). The following possibilities occur:

(i) two imaginary eigenvalues $\pm i\lambda$ ($\lambda > 0$, real):

$$X_1 = \begin{pmatrix} 0 & \lambda & 0 & 0 \\ -\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda \\ 0 & 0 & -\lambda & 0 \end{pmatrix},$$
$$X_2 = \begin{pmatrix} 0 & \lambda & 2\epsilon\lambda & 0 \\ -\lambda & 0 & 0 & 2\epsilon\lambda \\ 0 & 0 & 0 & \lambda \\ 0 & 0 & -\lambda & 0 \end{pmatrix}; \tag{40}$$

(ii) four complex eigenvalues ($\pm\mu, \pm i\nu$):

$$X_3 = \begin{pmatrix} 0 & \mu & \nu & 0 \\ -\mu & 0 & 0 & \nu \\ \nu & 0 & 0 & \mu \\ 0 & \nu & -\mu & 0 \end{pmatrix}, \quad \mu > 0, \nu > 0; \tag{41}$$

(iii) for two real eigenvalues $\pm\lambda$;

$$X_4 = \begin{pmatrix} \lambda & \lambda & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & -\lambda & 0 \\ 0 & 0 & -\lambda & -\lambda \end{pmatrix}; \tag{42}$$

(iv) one eigenvalue 0:

$$X_5 = \begin{pmatrix} 0 & 0 & 0 & \epsilon \\ 0 & 0 & \epsilon & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -\epsilon & 0 & 0 \end{pmatrix} \tag{43}$$

(notice that X_1 and X_4 are decomposable, and X_2, X_3, and X_5 indecomposable). The matrices $X_1, \ldots X_5$ give rise to five more types of representative Hamiltonians:

$$\mathscr{H}_7 = q_2p_1 - q_1p_2,$$
$$\mathscr{H}_8^\epsilon = p_1^2 + p_2^2 + \epsilon(q_2p_1 - q_1p_2), \quad \epsilon = \pm 1,$$
$$\mathscr{H}_9^\lambda = p_1^2 + p_2^2 - q_1^2 - q_2^2 + \lambda\,(q_2p_1 - q_1p_2), \quad \lambda > 0, \tag{44}$$
$$\mathscr{H}_{10} = q_1p_1 + p_1q_1 + q_2p_2 + p_2q_2 + 2q_2p_1,$$
$$\mathscr{H}_{11} = 2p_1p_2 + q_2^2.$$

Thus, in four dimensional phase space a general quadratic Hamiltonian depending on 10 parameters [a linear combination of all operators in Eq. (36)] can be reduced to one of the 11 representative Hamiltonians (38) and (44). Notice that some of these are actually infinite families, depending on a continuous parameter λ, others are pairs (depending on $\epsilon = \pm 1$), and the rest are individual Hamiltonians.

B. Maximal Abelian subalgebras of sp(4,ℝ)

Since in this case $n = 2$ is even, all types of MASA occur. Following the general theory outlined in Sec. 2, we construct them all.

1. Orthogonally decomposable MASA

We use the realization (A5) of the Appendix and write all orthogonally decomposable subalgebras of sp(4,ℝ) as

$$X = \begin{pmatrix} a_1 & b_1 & & \\ -c_1 & -a_1 & & \\ & & a_2 & b_2 \\ & & -c_2 & -a_2 \end{pmatrix} \tag{45}$$

This provides us with five MASA; they in turn correspond to the following pairs of dynamical observables:

$$D_1 = \{p_1^2 - q_1^2, p_2^2 - q_2^2\},$$
$$D_2 = \{p_1^2 + q_1^2, p_2^2 - q_2^2\},$$
$$D_3 = \{p_1^2, p_2^2 - q_2^2\}, \tag{46}$$
$$D_4 = \{p_1^2 + q_1^2, p_2^2 + q_2^2\},$$
$$D_5 = \{p_1^2 + q_1^2, p_2^2\}.$$

2. Maximal Abelian nilpotent subalgebras

The Kravchuk signature (2 0 2) provides us with one MANS which in the realization (A3) is

$$X = \begin{pmatrix} 0 & 0 & b_1 & b_3 \\ 0 & 0 & b_3 & b_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

corresponding to the observables

$$D_6 = \{p_1^2, p_2^2, p_1p_2\}. \tag{47}$$

The other Kravchuk signature (1 2 1) leads to one further MANS which in the realization (A7) can be written as

$$X = \begin{pmatrix} 0 & a_3 & 0 & b_1 \\ 0 & 0 & a_3 & 0 \\ 0 & 0 & 0 & -a_3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The corresponding observables would be $\{q_2p_1 + p_2^2, p_1^2\}$; however, we prefer to use a more convenient pair conjugate to this one, namely,

$$D_{10} = \{2p_1p_2 + q_2^2, p_1^2\}. \tag{48}$$

3. Indecomposable but not MANS

In realization (A3) these must be contained in the cen-

tralizer cent$C$ of C in sp(4,$\mathbb{R}$):

$$C=\begin{pmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & 0 & 1 \\ & & -1 & 0 \end{pmatrix},$$

$$\text{cent}C=\begin{pmatrix} a_1 & a_3 & b_1 & 0 \\ -a_3 & a_1 & 0 & b_1 \\ c_1 & 0 & -a_1 & a_3 \\ 0 & c_1 & -a_3 & -a_1 \end{pmatrix}.$$

Here, cent$C$ is the algebra u(1) $\oplus$ su(1,1) [u(1) is generated by C]. The only indecomposable MASA obtained in this manner is

$$X=\begin{pmatrix} 0 & a_3 & b_1 & 0 \\ -a_3 & 0 & 0 & b_1 \\ 0 & 0 & 0 & a_3 \\ 0 & 0 & -a_3 & 0 \end{pmatrix},$$

corresponding to

$$D_7=\{p_1^2+p_2^2, q_1p_2-q_2p_1\}. \tag{49}$$

4. Decomposable but not orthogonally decomposable subalgebras

In realization (A3) these can all be written as

$$X=\begin{pmatrix} A & 0 \\ 0 & -A_T \end{pmatrix},$$

where A is an indecomposable MASA of gl(2,$\mathbb{R}$). We thus obtain two more subalgebras

$$X=\begin{pmatrix} a_1 & a_3 & 0 & 0 \\ -a_3 & a_1 & 0 & 0 \\ 0 & 0 & -a_1 & a_3 \\ 0 & 0 & -a_3 & -a_1 \end{pmatrix},$$

$$X=\begin{pmatrix} a_1 & a_3 & 0 & 0 \\ 0 & a_1 & 0 & 0 \\ 0 & 0 & -a_1 & 0 \\ 0 & 0 & -a_3 & -a_1 \end{pmatrix}.$$

They correspond to the observables

$$D_8=\{q_1p_1+p_1q_1+q_2p_2+p_2q_2, q_1p_2-q_2p_1\}, \tag{50}$$

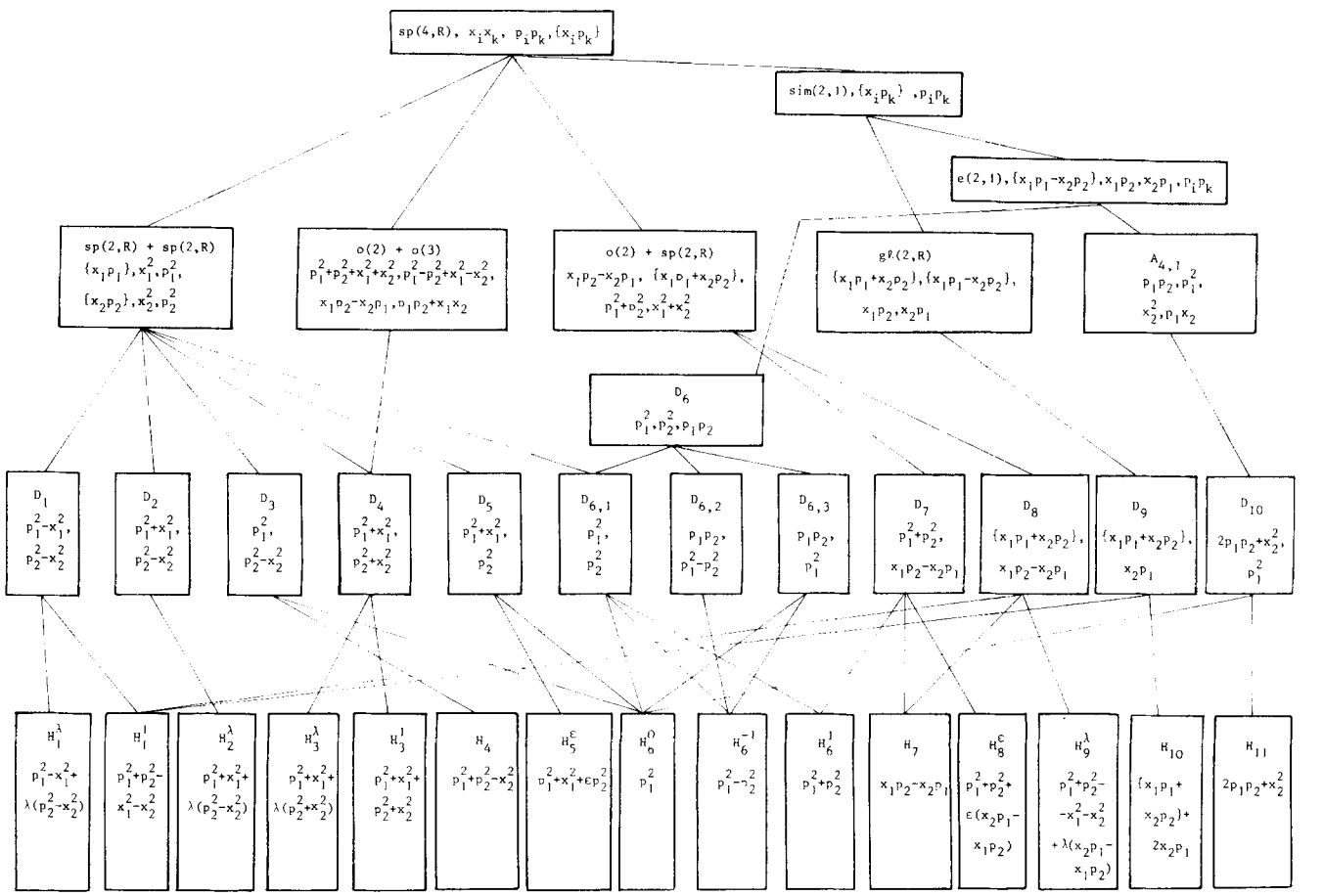


FIG. 1. Some subalgebras of sp(4,$\mathbb{R}$). Four of the seven maximal subalgebras are shown, as are all of the Abelian subalgebras. $D_1 \cdots D_{10}$ are maximal Abelian subalgebras ($D_{6,i}$ are not maximal). All one-dimensional subalgebras are indicated in the bottom row. The corresponding dynamical variables are given ($i,k=1,2$). Mutual inclusions are indicated. The algebra $A_{4,1}$ is four-dimensional nilpotent[35]. The ranges of parameters are as in formulas (38) and (44); however, in some cases specific values may be separated out (e.g., H_1^I is separate, and hence for H_1^λ we have $0 \leqslant \lambda < 1$). Curly brackets indicate anticommutators, e.g., $\{x_ip_k\} = x_ip_k + p_kx_i$. Note that the variable q_i in the text is replaced by x_i in the figure.

$$D_9 = \{q_1p_1 + p_1q_1 + q_2p_2 + p_2q_2, q_2p_1\}.$$

Several comments are appropriate at this stage.

(i) The number of conjugacy classes of one dimensional subalgebras of sp(4,$\mathbb{R}$) is infinite. We have however organized them into 11 types, many of which depend on a continuous parameter. These 11 types correspond to the Hamiltonians $\mathscr{H}_1,...,\mathscr{H}_{11}$ of Eqs. (38) and (44). For sp(2n,$\mathbb{R}$), $n\geqslant 3$, we would also obtain a parametrizable infinity of Hamiltonians.

(ii) The number of maximal Abelian subalgebras of sp(4,$\mathbb{R}$) is finite, namely, 10. They are represented by the dynamical systems $D_1,...,D_{10}$ of Eqs. (46)–(50). Nine of these algebras have dimension $d = 2$ [the rank of sp(4,$\mathbb{R}$) is also $r = 2$], and one has $d = 3$. The algebra sp(6,$\mathbb{R}$) also has a finite number of maximal Abelian subalgebras. For sp(2n,$\mathbb{R}$), $n\geqslant 4$, we would obtain a parametrizable infinity of maximal Abelian subalgebras and their dimensions d satisfy $d\leqslant[n(n+1)]/2$.

(iii) Each Hamiltonian $\mathscr{H}_\alpha$ ($\alpha = 1,...,11$) can be incorporated into at least one of the dynamical systems $D_1,...,D_{10}$, i.e., written as a linear combination of the basis elements of the corresponding D_k ($k = 1,...,10$). The mutual inclusions among some relevant subalgebras of sp(4,$\mathbb{R}$) are shown on Fig. 1.

(iv) Notice that while D_6 corresponds to an indecomposable MASA, each of its one dimensional subalgebras is orthogonally decomposable. In particular, $\mathscr{H}_6^\kappa = p_1^2 + \kappa p_2^2$ is contained in D_6.

C. Solutions of the representative Schrödinger equations

So far we have shown that any quadratic Hamiltonian in four dimensional phase space can be reduced by means of real linear canonical transformations into one of 11 simplified standard forms. In turn, each of the Hamiltonians $\mathscr{H}_1,...,\mathscr{H}_{11}$ can be written as a linear combination of two basis elements of some maximal Abelian subalgebra $D_1,...,D_{10}$ of sp(4,$\mathbb{R}$). Hence, all that we have to do at this stage is to find the common eigenfunctions of each dynamical system $D_1,...,D_{10}$.

For the decomposable case $D_1,...,D_5$ and also for D_6, this problem is trivial since in each case we get two equations of the type

$$\left(-\frac{\partial^2}{\partial q_1^2} + \tau_1 q_1^2\right)\psi(q_1,q_2) = \nu_1\psi(q_1,q_2),$$
$$\left(-\frac{\partial^2}{\partial q_2^2} + \tau_2 q_2^2\right)\psi(q_1,q_2) = \nu_2\psi(q_1,q_2),$$

where τ_1 and τ_2 are independently equal to 0 or ± 1. Thus, we have separation of variables in Cartesian coordinates and the wave function can be written as

$$\psi(q_1,q_2) = \phi_{\nu_1}^{0\sigma_1\tau_1}(q_1)\phi_{\nu_2}^{0\sigma_2\tau_2}(q_2), \tag{51}$$

where $\phi_\nu^{0\sigma\tau}(q_i)$ is the wave function (32) of a one dimensional harmonic oscillator, repulsive oscillator, or free particle, according to whether τ_i is equal to $+1$, -1, or 0, respectively.

Two of the remaining indecomposable systems are equally simple to treat, namely D_7 and D_8. Indeed, D_7 simply corresponds to a free particle with given angular momentum and D_8 is equivalent to a repulsive harmonic oscillator with given angular momentum ($q_1p_1 + p_1q_1 + q_2p_2 + p_2q_2$ can be transformed into $p_1^2 + p_2^2 - q_1^2 - q_2^2$). The corresponding equations can be solved by separating variables in polar coordinates

$$q_1 = r\cos\phi, \quad q_2 = r\sin\phi. \tag{52}$$

The wave functions satisfying

$$(p_1^2 + p_2^2)\psi(r,\phi) = k^2\psi(r,\phi),$$
$$(q_1p_2 - q_2p_1)\psi = m\psi, \tag{53}$$

are

$$\psi_{km}(r,\phi) = J_m(kr)e^{im\phi} \tag{54}$$

and describe the system D_7. The equations for the system D_8 are

$$\{q_1p_1 + p_1q_1 + p_2q_2 + q_2p_2\}\psi_{\lambda m}(r,\phi) = \lambda\psi_{\lambda m}(r,\phi),$$
$$(q_1p_2 - q_2p_1)\psi_{\lambda m}(r,\phi) = m\psi_{\lambda m}(r,\phi), \tag{55}$$

and their solution is

$$\psi_{\lambda m}(r,\phi) = r^{i\lambda/2-1}e^{im\phi}. \tag{56}$$

Now let us consider the system D_9:

$$(q_1p_1 + p_1q_1 + q_2p_2 + p_2q_2)\psi(q_1,q_2) = 2\lambda_1\psi(q_1,q_2),$$
$$q_2p_1\psi(q_1,q_2) = \lambda_2\psi(q_1,q_2). \tag{57}$$

In order to separate variables we introduce

$$\alpha = q_1/q_2, \quad \beta = q_1q_2, \tag{58}$$

and put $\psi(q_1,q_2) = f(\alpha)g(\beta)$. Solving Eq. (57), we obtain

$$\psi_{\lambda_1\lambda_2}(q_1,q_2) = \left(\frac{\beta}{\alpha}\right)^{(i\lambda_1-1)/2}e^{i\lambda_2\alpha} = q_2^{i\lambda_1-1}\exp\left(i\frac{\lambda_2q_1}{q_2}\right) \tag{59}$$

Finally, system D_{10} allows separation in Cartesian coordinates:

$$p_1^2\psi(q_1,q_2) = k_1^2\psi(q_1,q_2),$$
$$(2p_1p_2 + q_2^2)\psi(q_1,q_2) = \lambda\psi(q_1,q_2). \tag{60}$$

The solutions can be written as

$$\psi_{\lambda k_1}(q_1,q_2) = e^{ik_1q_1}\exp\left(i\frac{3\lambda q_2 - q_2^3}{6k_1}\right) \tag{61}$$

We can now summarize the results of this section. The most general quadratic Hamiltonian in four-dimensional phase space has been reduced to one of 11 forms. The wave functions of Hamiltonians $\mathscr{H}_1^\lambda,...,\mathscr{H}_6^\kappa$ can be expressed in terms of products of the wave functions of very elementary quantum mechanical systems, namely, one-dimensional free particles, harmonic oscillators, or repulsive harmonic oscillators (in Cartesian coordinates). The wave functions of $\mathscr{H}_7$, $\mathscr{H}_8^\epsilon$, and $\mathscr{H}_9^\lambda$ can be expressed in terms of similarly elementary two dimensional systems (in polar coordinates). The two individual remaining Hamiltonians $\mathscr{H}_{10}$ and $\mathscr{H}_{11}$ are

slightly more complicated but their wave functions can be expressed in terms of the functions (59) and (61), respectively.

D. Transformations relating the general Hamiltonians and their solutions to the representative ones

The final step in our procedure is to find a group element $g\in \mathrm{Sp}(4,\mathbb{R})$ realizing the transformation (19) and to construct the representation T_g in the appropriate bases. For each of the 10 systems $D_1,\ldots,D_{10}$, we have constructed the basis wave functions. In order to construct T_g we should represent the group $\mathcal{G}$ in the form of a double coset decomposition $\mathcal{G}=\mathcal{G}_3\mathcal{G}_2\mathcal{G}_1$, where $\mathcal{G}_3$ and $\mathcal{G}_1$ should preferably be the same group and the action of $\mathcal{G}_1$ and $\mathcal{G}_3$ on the basis functions should be as simple as possible.

On Fig. 1, we show some relevant subgroups of Sp(4,$\mathbb{R}$) (or rather their Lie algebras). In particular, we indicate all maximal Abelian subalgebras D_A ($A=1,\ldots,10$) and the three additional two dimensional nonmaximal Abelian subalgebras, of which one, namely $D_{6,1}=\{p_1^2,p_2^2\}$, is of special importance. We also indicate all one dimensional subalgebras H_a ($a=1,\ldots,11$) and their inclusions in D_A.

The subalgebras $D_1,\ldots,D_5$ and $D_{6,1}$ are orthogonally decomposable and they contain all the subalgebras $H_1,\ldots,H_6$. All orthogonally decomposable subalgebras are contained in one of the maximal subalgebras of sp(4,$\mathbb{R}$), namely, $\mathrm{sp}(2,\mathbb{R})\oplus\mathrm{sp}(2,\mathbb{R})$. In this case it is possible to prove that every element $g\in\mathrm{Sp}(4,\mathbb{R})$ can be written as

$$g=\tilde{g}_0 r g_0,\quad g_0,\tilde{g}_0\in\mathrm{Sp}(2,\mathbb{R})\otimes\mathrm{Sp}(2,\mathbb{R}),\tag{62}$$

where r, for different elements of g takes one of three possible forms

$$r=\exp\alpha(A_3-A_4),\quad r=\exp a(A_3+A_4),\quad\text{or}$$
$$r=\exp A_4,\tag{63}$$

with $0\leqslant\alpha<2\pi$ (the element A_3-A_4 generates rotations), $-\infty<a<\infty$ (A_3+A_4 generates hyperbolic transformations), and $\exp A_4$ has the character of a translation. The action of g_0 and $\tilde{g}_0$ on $D_1,\ldots,D_5$ and $D_{6,1}$ bases is known and was reviewed in Sec. 3. Thus, only the matrix elements of r need to be evaluated. The double coset decomposition (62) is thus very helpful.

The systems D_7 and D_8 (and also D_4) allow for a separation of variables in polar coordinates. In this case the group reduction $\mathrm{Sp}(4,\mathbb{R})\supset\mathrm{O}(2)\otimes\mathrm{Sp}(2,\mathbb{R})$ is more appropriate. A decomposition of the type $\mathcal{G}=\tilde{\mathcal{G}}_1A\mathcal{G}_1$ with $\mathcal{G}_1=\mathrm{O}(2)\otimes\mathrm{Sp}(2,\mathbb{R})$ would be desirable in this case. It does however not seem to be available in the literature and we have not determined all the forms of A that are necessary in order to represent every element $g\in\mathrm{Sp}(4,\mathbb{R})$ in terms of this double coset decomposition. (A must clearly depend on at least two parameters.)

The system D_9 is best treated in terms of a $\mathrm{Sp}(4,\mathbb{R})\supset\mathrm{Gl}(2,\mathbb{R})$ decomposition. Again, the exact forms needed for A in the double coset decomposition $\mathcal{G}=\tilde{\mathcal{G}}_2A\mathcal{G}_2$ [$\mathcal{G}_2\in\mathrm{Gl}(2,\mathbb{R})$] is not known.

Finally, the subgroup corresponding to D_{10} can be imbedded into a nilpotent group corresponding to the $A_{4,1}$ type algebra[35] $\{p_1^2,q_2^2;p_1p_2,p_1q_2\}$ which in turn is contained in the pseudo-Euclidean subgroup E(2,1) of SP(4,$\mathbb{R}$). In this case the decomposition $\mathcal{G}_3=\tilde{\mathcal{G}}_3A\mathcal{G}_3$ with $\mathcal{G}_3=\mathrm{E}(2,1)$ would be appropriate.

We plan to return to the problem of constructing the necessary decompositions of Sp(4,$\mathbb{R}$) and the representations T_g in the future. Let us just comment that the complete subalgebra structure of sp(4,$\mathbb{R}$) is quite complicated, but is known[36]. In particular, sp(4,$\mathbb{R}$) has seven maximal subalgebras, of which we only show four on Fig. 1.

E. Comments on six dimensional phase space

The algebra sp(6,$\mathbb{R}$) can again be realized in several different useful ways, depending on the choice of K in Eq. (6). Choosing K as in Eq. (4) with $N=3$, we have

$$X=\begin{pmatrix}A & B\\ -C & -A^T\end{pmatrix},\quad A\in\mathrm{gl}(3,\mathbb{R}),$$
$$B=B^T,\quad C=C^T,\quad B,C\in\mathbb{R}^{3\times3}\tag{64}$$

We can again choose a basis consisting of 21 matrices A_{ik},$B_{ik}=B_{ki}$, and $C_{ik}=C_{ki}$, having an entry 1 on the intersection of the ith row and k th column and 0 elsewhere. These operators can be identified with the dynamical variables as in Eq. (36), i.e., we put

$$A_{lk}=-\frac{i}{2}(q_kp_l+p_lq_k),\quad l,k=1,2,3,$$
$$B_{kk}=-\frac{i}{2}p_k^2,\quad B_{kl}=-ip_kp_l,\quad k\neq l,\tag{65}$$
$$C_{kk}=-\frac{i}{2}q_k^2,\quad C_{kl}=-iq_kq_l,\quad k\neq l.$$

A classification of the generators of sp(6,$\mathbb{R}$) and hence a classification of quadratic Hamiltonians can be extracted, e.g., from Ref. 14. We shall not reproduce it here. A classification of the maximal Abelian subalgebras of sp(6,$\mathbb{R}$) was obtained quite recently[19]. As in the case of sp(4,$\mathbb{R}$), there exists a finite number of MASA (not true for arbitrary n), namely, 29. Since $n=3$ is odd, only three types of MASA exist. The **orthogonally decomposable** ones correspond to the partitions $6=2+2+2$ (nine three-dimensional algebras) or $6=4+2$ (two four-dimensional subalgebras and 11 three-dimensional ones). The partition $2+2+2$ obviously corresponds to various combinations of harmonic oscillators, free particles, and repulsive oscillators in each of the dimensions. The partition $4+2$ corresponds to a h.o., f.p., or r.h.o. in one dimension and one of the systems $D_6,\ldots,D_{10}$ in the remaining two (we recall that at most one of the subalgebras in the decomposition is allowed to be nilpotent).

Decomposable but not orthogonally decomposable MANS are obtained by putting $B=C=0$ in Eq. (64) and taking A in the form of an indecomposable MASA of gl(3,$\mathbb{R}$). Three such MASA exist, all of dimension $d=3$.

Finally, we have **maximal Abelian nilpotent subalgebras**. The Kravchuk pattern (3 0 3) leads to one six-dimensional MANS (corresponding to the operators p_1^2, p_2^2, p_3^2, p_1p_2, p_2p_3, and p_3p_1). The pattern (2 2 2) leads to two four-dimensional MASA and (1 4 1) to three three-dimensional

ones. It would be a simple matter to write out the quadratic operators, corresponding to each MASA, but we shall not do this here.

5. QUADRATIC HAMILTONIANS IN $2n$-DIMENSIONAL PHASE SPACE CONTAINING NO BILINEAR TERMS IN q_ip_j, $(i,j = 1,...,n)$

As indicated in the title of this section we wish to deal with Hamiltonians $\mathscr{H}$ such that the corresponding matrix in the Lie algebra sp$(2n,\mathbb{R})$ is

$$X = \begin{bmatrix} 0 & B \\ -C & 0 \end{bmatrix}, \quad B = B^T, \quad C = C^T, \quad B,C \in \mathbb{R}^{n\times n}. \tag{66}$$

When, furthermore, both B and C are positive definite, we have the problem of small vibrations[2,3]. In what follows we shall assume B positive definite, as it is related to the momenta and, when diagonalized, its terms can be interpreted as recipocals of masses. On the other hand, we shall not impose the positive definite restrictions on C.

We first show that for the X in Eq. (66) the squares of the eigenvalues mentioned in Sec. 2 are all real and so λ is either real or pure imaginary and in both cases with each λ we can pair a $(-\lambda)$. For this purpose we note that as λ satisfies $\det(\lambda I_{2n} - X) = 0$ we have

$$\det(\lambda I_{2n} - X)\det(\lambda I_{2n} + X) = \det(\lambda^2 I_{2n} - X^2)$$
$$= \det\begin{bmatrix} \lambda^2 I_n + BC & 0 \\ 0 & \lambda^2 I_n + CB \end{bmatrix} = 0, \tag{67}$$

where I_{2n} and I_n are, respectively, $2n\times 2n$ and $n\times n$ unit matrices. As B is a positive definite real symmetric matrix, all its eigenvalues are positive and thus $B^{1/2}$ exists. Therefore, Eq. (67) leads to

$$\det[\lambda^2 I_n + B^{1/2} C B^{1/2}] = 0, \tag{68}$$

and as $B^{1/2}CB^{1/2}$ is real and symmetric, the λ^2 are all real.

How does one proceed to find the classes to which belong the matrices (66) of the Lie algebra? As usual,[2] one first carries out an $n\times n$ orthogonal transformation $\mathscr{O}$ in configuration space to diagonalize B, and then a dilation transformation D (given by a diagonal matrix whose eigenvalues are the inverse square roots of the eigenvalues of B) to reduce B to the unit matrix. Finally, one carries out an orthogonal transformation $\mathscr{O}'$ to convert $D^{-1}\mathscr{O}^T C \mathscr{O} D$ into a diagonal matrix whose eigenvalues, necessarily real, we denote by w_i, $i = 1,...,n$. The $X = KH$ is thus transformed to

$$\begin{bmatrix} 0 & \vdots & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -w_1 & & & & & \\ & -w_2 & & & 0 & \\ & & \ddots & & & \\ & & & -w_n & & \end{bmatrix}, \tag{69}$$

where from Eq. (68) we see that

$$\lambda_i^2 = -w_i, \quad i = 1,2,...,n. \tag{70}$$

The problem (69) is completely solvable in quantum mechanics as it corresponds to the Hamiltonian

$$\sum_{i=1}^{n} (p_i^2 + w_i q_i^2), \tag{71}$$

separable in the n coordinates, which means that we have a set of n quadratic observables $p_i^2 + w_i q_i^2$, $i = 1,2,...,n$ that commute among themselves and with the Hamiltonian. They are the orthogonally decomposable MASA's for the partition $2n = 2 + 2 + ... + 2$, which, incidentally, also appear in many problems that *do* contain bilinear terms q_ip_j, $i,j = 1,...,n$, as seen in the previous section.

Before discussing the set of quadratic observables mentioned, we carry out a further dilation D' given by $q_i \to |w_i|^{-1/4} q_i$, $p_i \to |w_i|^{1/4} p_i$ when $|w_i| \neq 0$, so that the terms in Eq. (71) become either

$$|w|^{1/2}(p^2 + q^2), \quad p^2, \quad \text{or} \quad |w|^{1/2}(p^2 - q^2), \tag{72}$$

depending on whether $w > 0$, $w = 0$, or $w < 0$, respectively.

As a last step we consider a permutation P of the indices $i = 1,2,...,n$ that enumerates first the w_i positive (n_+ of them), then $w_i = 0$ (n_0 of them), and finally w_i negative (n_- of them), where $n_+ + n_0 + n_- = n$. Thus, we can now write the following complete set of quadratic commuting observables in the normal coordinate system

$$\Omega_i^{1/2}(p_i^2 + \tau_i q_i^2), \quad i = 1,2,...,n, \tag{73a}$$

where

$$\begin{aligned} &\Omega_i = w_i, && \tau_i = 1, && \text{for } 1 \leqslant i \leqslant n_+, \\ &\Omega_i = 1, && \tau_i = 0, && \text{for } n_+ + 1 \leqslant i \leqslant n_+ + n_0, \\ &\Omega_i = -w_i, && \tau_i = -1, && \text{for } n_+ + n_0 + 1 \leqslant i \leqslant n, \end{aligned} \tag{73b}$$

and the Hamiltonian $\mathscr{H}$ in the same system is just the sum of the operators (73a).

For the problem discussed here, the elements of the Abelian subalgebra of sp$(2n,\mathbb{R})$ mentioned in Sec. 2 are given by Eqs. (73) when we replace the column vector $\mathbf{q}$ of components q_i $i = 1,2,...,n$ by

$$\mathbf{q} \to R^{-1}\mathbf{q} \equiv P^{-1} D'^{-1} \mathscr{O}'^T D^{-1} \mathscr{O}^T \mathbf{q} \tag{74}$$

and a similar substitution for the momenta. All the operators in Eq. (74) are $n\times n$ matrices associated with point transformations in configuration space.

All we have done so far is to discuss an elementary problem[2] in the language developed in the previous sections for quadratic Hamiltonians in phase space. There remains though the question of expressing the eigenstates of the observables (73) in the **original** configuration space. At first sight this seems only to imply that we take the wave function

$$\Psi^{\sigma_1\cdots\sigma_n}_{\nu_1\cdots\nu_n}(q_1,...,q_n) \equiv \prod_{i=1}^{n} \phi^{0\sigma_i\tau_i}_{\nu_i}(q_i), \tag{75}$$

where the ϕ's are given in Eq. (32), and replace the q_i, $i = 1,...,n$ of $\mathbf{q}$ by $R^{-1}\mathbf{q}$ as in Eq. (74).

The last solution would be very cumbersome to use in applications such as the calculation of matrix elements. Thus, we can either translate the observables we are interest-

ed in into the normal coordinate system and calculate their matrix elements with respect to the eigenstates (75) or expand $\Psi(R^{-1}\mathbf{q})$ back in terms of $\Psi(\mathbf{q})$. This last procedure implies finding the representation $\Delta(R)$—not necessarily irreducible—of the linear point transformation (74) on the basis (75).

Because of the explicit form (74) of R^{-1}, we see that $\Delta(R)$ decomposes into representations of dilations D affecting only one coordinate at a time and of $n\times n$ orthogonal transformation $\mathcal{O}$ as $P\subset\mathcal{O}$. For dilations the representation problem was solved in Eq. (35) of Sec. 3 if we take there $\delta=\alpha^{-1}, \beta=\gamma=0$. Thus, we are concerned only with $\Delta(\mathcal{O})$.

From the forms (73) and (75) of the states it appears convenient to consider the following subgroup of the n dimensional orthogonal group

$$\mathcal{O}(n\times n)\supset\begin{bmatrix}\mathcal{O}(n_+\times n_+) & & 0\\ & \mathcal{O}(n_0\times n_0) & \\ 0 & & \mathcal{O}(n_-\times n_-)\end{bmatrix}, \tag{76}$$

where the matrices have the dimensions indicated. The representations of $\mathcal{O}(n_\tau\times n_\tau)$, $\tau=+,0,-$ can be determined if we pass from the corresponding products $\Pi_i\phi_{\nu_i}^{0\sigma_i\tau}(q_i)$ in Cartesian coordinates to states in hyperspherical coordinates of the harmonic oscillator ($\tau=+$), free particle ($\tau=0$), and repulsive oscillator ($\tau=-$). Once we have these representations in hyperspherical coordinates, we can translate them to Cartesian ones with the help of the expansions of one type of solutions in terms of the other, e.g., the expansion of plane waves into spherical waves[40] when $\tau=0$, $n_0=3$. Thus, in principle, a general way is available to find the representation of $\mathcal{O}(n_\tau\times n_\tau)$, $\tau=+,0,-$ on the basis (75).

Our problem though is not finished. Clearly, the next step is to decompose an arbitrary $n\times n$ orthogonal matrix $\mathcal{O}$ into

$$\mathcal{O}=\tilde{\mathcal{O}}_1\mathcal{O}_2\mathcal{O}_1, \tag{77}$$

where $\mathcal{O}_1$ and $\tilde{\mathcal{O}}_1$ belong to the subgroup (76) while $\mathcal{O}_2$ is outside it, i.e., a double coset decomposition of the group $\mathcal{O}$ with respect to the subgroup (76). This in itself is no small undertaking, but even if we succeed there will still remain the problem of finding the representation of $\mathcal{O}_2$ in the basis (75). No general procedure seems to be available in this case though occasionally brute force methods (i.e., direct evaluation of the scalar product of the transformed state and the original one) may be successful.

While we cannot implement our program fully even in the elementary problem discussed in this section, awareness of its structure may permit evaluation of specific cases of physical interest. For example, it may happen that all the w_i are either positive, zero, or negative, in which case the subgroup (76) coincides with $\mathcal{O}(n\times n)$ and so the determination of the representation seems feasible, at least in principle.

As a last point we note, as was done in Sec. 3, that all our arguments can be extended from scalars q_i, p_i, $i=1,...,n$ to three dimensional vectors $\mathbf{q}_i, \mathbf{p}_i$, so long as the products p_ip_j and q_iq_j are replaced by the *scalar* products $\mathbf{p}_i\cdot\mathbf{p}_j$ and $\mathbf{q}_i\cdot\mathbf{q}_j$. We note that here R is a $3n\times 3n$ matrix, each of whose components is given by the corresponding element in the $n\times n$ matrix R but now multiplied by the unit three dimensional matrix. The representation of this $3n\times 3n$ R on the basis states corresponding to Eq. (75) will of course be a more complex affair, but already some examples of great physical interest were discussed long ago in the literature.[41]

6. QUADRATIC HAMILTONIANS IN PHASE SPACE INVARIANT UNDER ROTATIONS, TRANSLATIONS, AND PERMUTATIONS

In the previous sections we discussed general procedures for arriving at the eigenstates of quadratic Hamiltonians in phase space. In the present one we wish to consider these Hamiltonians as related with n identical particles moving in a three dimensional space. The Hamiltonians must then be invariant under rotations, which implies that they can only be linear combinations of the scalar products $\mathbf{q}_i\cdot\mathbf{q}_j$, $\mathbf{p}_i\cdot\mathbf{p}_j$, and $\frac{1}{2}(\mathbf{q}_i\cdot\mathbf{p}_j+\mathbf{p}_j\cdot\mathbf{q}_i)$. They also need to be invariant under translations which gives rise to some simple restrictions which we prefer to discuss later. Finally, as the particles are identical, they must be invariant under permutations of the indices $i=1,...,n$ simultaneously for coordinates and momenta. This implies that the matrix H of Eq. (3) satisfies

$$\begin{bmatrix}P & 0\\ 0 & P\end{bmatrix}\begin{bmatrix}H_{11} & H_{12}\\ H_{12}^T & H_{22}\end{bmatrix}\begin{bmatrix}P^{-1} & 0\\ 0 & P^{-1}\end{bmatrix}=\begin{bmatrix}H_{11} & H_{12}\\ H_{12}^T & H_{22}\end{bmatrix}, \tag{78}$$

where all submatrices in Eq. (78) are $n\times n$ and P is an arbitrary permutation matrix. Clearly, then we must first discuss the form of the matrix h of components

$$h=\|h_{ij}\|, \quad i,j=1,2,...,n, \tag{79}$$

if it satisfies for any P the relation

$$PhP^{-1}=h. \tag{80}$$

The h will stand for H_{11}, H_{12}, and H_{22} and that is why we prefer the notation (78) to that of Eq. (3).

An arbitrary permutation P can be formed from products and powers of just two generating ones[42]

$$P'=(1\,2\,3\,4\cdots n), \tag{81a}$$

$$P''=(1\,2), \tag{81b}$$

corresponding, respectively, to a cyclic permutation and to a transposition of the indices. If we apply P' to h, it transforms

$$h_{ij}\to h_{i+1,j+1} \text{ modulo } n, \tag{82}$$

and thus comparing with the original h we get

$$\begin{aligned}&h_{11}=h_{22}=h_{33}=\cdots=h_{nn},\\ &h_{12}=h_{23}=h_{34}=\cdots=h_{n-1\,n}=h_{n1},\\ &h_{13}=h_{24}=\cdots=h_{n-1\,1}=h_{n2},\\ &\quad\vdots\\ &h_{1n}=h_{21}=h_{32}=\cdots=h_{n\,n-1},\end{aligned} \tag{83}$$

where the arrows will be explained below. We note now that the application of P'' gives the transformation

$$h_{11}\to h_{22}, \quad h_{12}\to h_{21}, \quad h_{1i}\to h_{2i}, \quad h_{i1}\to h_{i2},$$

$$h_{ij} \to h_{ij}, \quad \text{for } i,j = 3,4,...,n, \tag{84a}$$

so that comparing with the original set h we get

$$h_{11} = h_{22}, \quad h_{12} = h_{21},$$
$$h_{1i} = h_{2i}, \quad h_{i1} = h_{i2}, \quad \text{for } i = 3,4,...,n. \tag{84b}$$

Looking back now at Eq. (83), we see that Eq. (84b) makes h_{23} in the second row equal to h_{13} in the third, h_{24} in the third equal to h_{14} in the fourth, and so on as indicated by the arrows. Thus, clearly, $h_{ii} = \beta$ for all i, $h_{ij} = \gamma$ for all i,j such that $i \neq j$, and thus the matrix h takes the form

$$h = \begin{pmatrix} \beta & \gamma & \gamma & \cdots & \gamma \\ \gamma & \beta & \gamma & \cdots & \gamma \\ \vdots & & & & \\ \gamma & \cdots & \cdots & \gamma & \beta \end{pmatrix}. \tag{85}$$

The Hamiltonian matrix (78) has then three submatrices of the type (85) associated with H_{11}, H_{12}, and H_{22} and we shall denote the corresponding parameters by β_{11}, γ_{11}, β_{12}, γ_{12}, β_{22}, and γ_{22}.

We wish now to reduce the Hamiltonian matrix to a canonical form in which, in particular, we can see the eigenvalues as well as the Abelian subalgebra of $sp(2n,\mathbb{R})$ with which it is associated. To achieve this purpose we first notice that the quadratic form $\mathbf{x}^T h \mathbf{x}$, where h is given by Eq. (85) and $\mathbf{x}$ is an n component column vector $\mathbf{x} = \{x_i, i = 1,2,...,n\}$, can be written as

$$\mathbf{x}^T h \mathbf{x} = (\beta - \gamma) \sum_{i=1}^{n} x_i^2 + n\gamma x_n'^2, \tag{86}$$

where

$$x_n' = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i. \tag{87a}$$

Clearly, then, an orthogonal transformation to center of mass coordinate x_n' and Jacobi coordinates x_j', $j = 1,2,...,n-1$, defined by

$$x_j' = [j(j+1)]^{-1/2} \sum_{i=1}^{j} x_i - [j/(j+1)]^{1/2} x_{j+1}, \tag{87b}$$

reduces the quadratic form to

$$\mathbf{x}^T h \mathbf{x} = (\beta - \gamma) \sum_{i=1}^{n-1} x_i'^2 + [\beta + (n-1)\gamma] x_n'^2. \tag{88}$$

The above analysis immediately suggests that a change from $\mathbf{q}_i$ to $\mathbf{q}_i'$ and $\mathbf{p}_i$ to $\mathbf{p}_i'$, where the primed variables are defined in terms of the original ones as in Eq. (87), will reduce the Hamiltonian to the form

$$\begin{aligned} \mathscr{H} = {} & (\beta_{11} - \gamma_{11}) \sum_{i=1}^{n-1} \mathbf{q}_i'^2 + (\beta_{12} - \gamma_{12}) \sum_{i=1}^{n-1} (\mathbf{q}_i' \cdot \mathbf{p}_i' + \mathbf{p}_i' \cdot \mathbf{q}_i') \\ & + (\beta_{22} - \gamma_{22}) \sum_{i=1}^{n-1} \mathbf{p}_i'^2 + [\beta_{11} + (n-1)\gamma_{11}] \mathbf{q}_n'^2 \\ & + [\beta_{12} + (n-1)\gamma_{12}] (\mathbf{q}_n' \cdot \mathbf{p}_n' + \mathbf{p}_n' \cdot \mathbf{q}_n') \\ & + [\beta_{22} + (n-1)\gamma_{22}] \mathbf{p}_n'^2. \end{aligned} \tag{89}$$

Before transforming further to a solvable Hamiltonian we note that in Eq. (89) the condition of translational invariance is very easy to implement. The Jacobi coordinates $\mathbf{q}_i'$, $i = 1,2,...,n-1$ do not change under translations and so only the center of mass changes to

$$\mathbf{q}_n' \to \mathbf{q}_n' + \mathbf{q}_0, \tag{90}$$

where $\mathbf{q}_0$ is some constant vector. Clearly, the Hamiltonian remains then invariant if and only if

$$\beta_{12} = (1-n)\gamma_{12}, \quad \beta_{11} = (1-n)\gamma_{11}, \tag{91}$$

so that only the kinetic energy term $p_n'^2$ remains associated with the center of mass. The matrix associated with each Jacobi degree of freedom $\mathbf{q}_i$ and $\mathbf{p}_i$, $i = 1,2,...,n-1$ takes then the form

$$\begin{bmatrix} -n\gamma_{11} & -n\gamma_{12} \\ -n\gamma_{12} & \beta_{22} - \gamma_{22} \end{bmatrix}, \tag{92}$$

and thus for quadratic Hamiltonians invariant under rotation, translation, and permutations there are $n-1$ equal eigenvalues

$$\lambda_i^2 = -\Delta = n^2\gamma_{12}^2 + n\gamma_{11}(\beta_{22} - \gamma_{22}), \quad i = 1,2,...,n-1, \tag{93a}$$

while for the center of mass degree of freedom we get of course

$$\lambda_n^2 = 0. \tag{93b}$$

The Hamiltonian (89) can be transformed through

$$\begin{aligned} \mathbf{q}_i'' &= \mathbf{q}_i' \cos\alpha + \mathbf{p}_i' \sin\alpha, \\ \mathbf{p}_i'' &= -\mathbf{q}_i' \sin\alpha + \mathbf{p}_i' \cos\alpha, \end{aligned} \quad i = 1,2,...,n-1, \tag{94}$$

where α, independent of i, is selected so that the symmetric matrix (92) becomes diagonal. Furthermore, we can consider the dilation transformation discussed in the paragraphs preceding Eqs. (29) and (72). Denoting by $\mathbf{q}_i'''$ and $\mathbf{p}_i'''$ our coordinates and momenta after all these operations, respectively, we see that the Abelian subalgebra discussed in Secs. 3,...,5 has as elements

$$\kappa_\tau (\mathbf{p}_i''^2 + \tau \mathbf{q}_i''^2), \quad i = 1,2,...,n-1, \tag{95a}$$

$$\mathbf{p}_n''^2, \tag{95b}$$

where $\tau = 1, 0$, and -1 if $\Delta > 0$, $\Delta = 0$, or $\Delta < 0$, respectively, and κ_τ is given by Eq. (26). The Hamiltonian in the coordinate system q_i''' and p_i''', $i = 1,2,...,n$ is the sum of all the terms in Eqs. (95).

The eigenstates of Eqs. (95) in the configuration space are given by Eqs. (32), and besides we can get the corresponding eigenstates in the Jacobi coordinates $\mathbf{q}_i'$ expressed as a linear combination of them. For this we would only need the representation (35) of the linear canonical transformation (33).

Our eigenstates are then linear combinations of products of the wave functions (30), i.e.,

$$\prod_{i=1}^{n-1} r_i^{-1} \phi_{\nu_i}^{l_i \sigma \tau}(r_i) Y_{l_i m_i}(\theta_i, \varphi_i) e^{i\mathbf{p}_n' \cdot \mathbf{q}_n'}, \tag{96}$$

where r_i, θ_i, and φ_i are spherical coordinates associated with the coordinates $\mathbf{q}_i'''$, $i = 1,2,...,n-1$, σ is restricted to negative parity (as indicated in Sec. 3), and $\tau = +, 0$, or $-$ depending on whether $\Delta > 0$, $\Delta = 0$, or $\Delta < 0$, respectively.

The states (96) do not correspond of course to definite total angular momentum or a given representation of the symmetric group $S(n)$ of permutations of n particles. The first point can be achieved by coupling successively the angular momenta while the second has been explicitly discussed elsewhere in the case $\tau = +$ by making full use of the con-

cept of transformation brackets[41] for harmonic oscillator functions as well as the chain of groups for n particles in the harmonic oscillator, i.e.,[5]

$$\begin{array}{ccc} U_{3n} \supset U_{3n-3} & \otimes & U_3 \\ U_{3n-3} \supset \quad U_3 & \times & U_{n-1} \\ \quad \cup & & \cup \\ \quad O_3 & & O_{n-1} \\ \quad \cup & & \cup \\ \quad O_2 & & S_n . \end{array} \tag{97}$$

For the case $\tau = 0$ this last part of the program can be achieved even more easily if we just discuss everything in terms of plane rather than spherical waves.

There remains the case of $\tau = -$ for which something equivalent to the concept of transformation brackets[5,41] for harmonic oscillator functions has to be developed for the repulsive oscillator. This though is a subject whose discussion we wish to leave to another publication.

Essentially, a full discussion has been carried out for the most general quadratic Hamiltonian in phase space invariant under rotations, translations, and permutations. This is a likely type of Hamiltonian to appear in many body problems involving identical particles. Furthermore, it is a genuine generalization of the small vibration problem as it also involves terms of the type $\mathbf{q}_i \cdot \mathbf{p}_j$.

7. CONCLUSIONS

The main result of this paper is the conclusion that physical problems leading to quadratic Hamiltonians can be explicitly solved. The solution makes use of real linear canonical transformations Sp($2n$,$\mathbb{R}$), of an imbedding of the Hamiltonian into a complete set of commuting quadratic integrals of motion, and of the resulting separation of variables in the Schrödinger equation. The canonical transformations thus play two roles in the present context. First, the entire group Sp($2n$,$\mathbb{R}$) is used to transform a general Hamiltonian of the considered type into a much simpler one, namely, one of the representative Hamiltonians $\mathscr{H}_R$. Next, a subgroup of Sp($2n$,$\mathbb{R}$) is used to help separate variables in the simplified equation, this subgroup being an appropriate maximal Abelian subgroup of Sp($2n$,$\mathbb{R}$). This subgroup of the group of canonical transformations leaves invariant the obtained Schrödinger equation.

We consider this to be an illustration of the dual role played by canonical transformations in quantum mechanics: to reduce complicated problems to simpler ones on the one hand and to help solve the simplified problems on the other.

A large part of this article was devoted to low dimensional cases, namely, to one and two degrees of freedom. Let us make two comments here. Firstly, when considering two or three body problems in Euclidean three space, described by quadratic Hamiltonians that are rotationally invariant, precisely these $n = 1$ and $n = 2$ cases occur (where each q_i and p_i is to be interpreted as a vector $\mathbf{q}_i$ or $\mathbf{p}_i$, respectively, and each product as a scalar product). Secondly, when considering systems with more degrees of freedom, as those treated in Secs. 5 and 6, many of the Hamiltonians and sets of integrals of motion will correspond to orthogonally decomposable subalgebras of sp($2n$,$\mathbb{R}$). These in turn lead to systems already treated in lower dimensions. In particular, the $n = 1$ and $n = 2$ cases will thus come up again for any $n \geqslant 3$.

ACKNOWLEDGMENTS

The authors are much indebted to C.P. Boyer, J. Patera, R.T. Sharp, and H. Zassenhaus for stimulating discussions.

APPENDIX: DIFFERENT REALIZATIONS OF THE ALGEBRA sp(4,$\mathbb{R}$) AND THE ISOMORPHISM WITH o(3,2)

In the text we make use of several realizations of the algebra sp(4,$\mathbb{R}$). In all cases these are 4×4 real matrices X, satisfying

$$X_\alpha K_\alpha + K_\alpha X_\alpha^T = 0. \tag{A1}$$

Several choices of the skew symmetric matrix K are convenient for various applications. Mainly we use

$$K_0 \equiv K = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}; \tag{A2}$$

then we have

$$X_0 \equiv X = \begin{pmatrix} A & B \\ -C & -A^T \end{pmatrix} = \begin{pmatrix} a_1 & a_3 & b_1 & b_3 \\ a_4 & a_2 & b_3 & b_2 \\ -c_1 & -c_3 & -a_1 & -a_4 \\ -c_3 & -c_2 & -a_3 & -a_2 \end{pmatrix}. \tag{A3}$$

Whenever decomposable subalgebras are involved it is more convenient to choose

$$K_1 = \begin{pmatrix} 0 & 1 & & \\ -1 & 0 & & \\ & & 0 & 1 \\ & & -1 & 0 \end{pmatrix} = RK_0R^{-1}, \tag{A4}$$

$$R = R^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We then obtain

$$X_1 = RXR^{-1} = \begin{pmatrix} a_1 & b_1 & a_3 & b_3 \\ -c_1 & -a_1 & -c_3 & -a_4 \\ a_4 & b_3 & a_2 & b_2 \\ -c_3 & -a_3 & -c_2 & -a_2 \end{pmatrix}. \tag{A5}$$

Finally, when considering some of the maximal Abelian subalgebras it is convenient to choose

$$K_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} = ZKZ^{-1}, \tag{A6}$$

$$Z = Z^{-1} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 0 & 1 \\ & & 1 & 0 \end{pmatrix}.$$

We then obtain

$$X_2 = ZXZ^{-1} = \begin{pmatrix} a_1 & a_3 & b_3 & b_1 \\ a_4 & a_2 & b_2 & b_3 \\ -c_3 & -c_2 & -a_2 & -a_3 \\ -c_1 & -c_3 & -a_4 & -a_1 \end{pmatrix}. \tag{A7}$$

It is well known[39] that the classical group Sp(4,ℝ) and SO(3,2) are locally isomorphic, i.e., that the algebras sp(4,ℝ) and o(3,2) are isomorphic. We can hence realize the algebra o(3,2) by 5×5 real matrices Y satisfying

$$YI_{3,2} + I_{3,2}Y^T = 0,$$

with

$$I_{3,2} = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & -1 & \\ & & & & -1 \end{pmatrix}. \tag{A8}$$

The o(3,2) matrix Y can then be directly related to the sp(4,ℝ) matrix X (and also X_1 and X_2) by identifying its elements as follows:

$$Y = \frac{1}{2}\begin{pmatrix} 0 & -b_1+b_2-c_1+c_2 & -a_3+a_4 & -b_1+b_2+c_1-c_2 & -a_1+a_2 \\ b_1-b_2+c_1-c_2 & 0 & -b_3-c_3 & -a_1-a_2 & b_1+b_2-c_1-c_2 \\ a_3-a_4 & b_3+c_3 & 0 & b_3-c_3 & a_3+a_4 \\ -b_1+b_2+c_1-c_2 & -a_1-a_2 & b_3-c_3 & 0 & -b_1-b_2-c_1-c_2 \\ -a_1+a_2 & b_1+b_2-c_1-c_2 & a_3+a_4 & b_1+b_2+c_1+c_2 & 0 \end{pmatrix}. \tag{A9}$$

The letters in the realizations (A3), (A5), (A7), and (A9) have exactly the same meaning, i.e., if we construct the basis matrices $A_1,\ldots,A_4$, $B_1,\ldots,B_3$, and $C_1,\ldots,C_3$ by setting, for example, for A_1, the entry $a_1 = 1$ and all others equal to zero, we obtain in each case 10 matrices that satisfy the same commutation relations as the dynamical quantities (36).

[1]M. Moshinsky, *The Harmonic Oscillator in Modern Physics. From Atoms to Quarks* (Gordon and Breach, New York, 1969).
[2]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950); E.T. Whittaker, *Analytical Dynamics* (Cambridge University, Cambridge, England, 1926).
[3]L.D. Landau and E.M. Lifshitz, *Quantum Mechanics. Nonrelativistic Theory* (Pergamon, London, 1958).
[4]E.K. Gora, Int. J. Quantum Chem. **6**, 681 (1972); R.C. Herman and W.H. Shaffer, J. Chem. Phys. **16**, 453 (1948); G. Amat and H.H. Nielsen, J. Chem. Phys. **27**, 834 (1957); **29**, 665 (1958); **36**, 1859 (1962).
[5]P. Kramer and M. Moshinsky, "Group Theory of Harmonic Oscillators and Nuclear Structure," in *Group Theory and Applications I*, edited by E.M. Loebl (Academic, New York, 1968); see also references therein.
[6]M.H. Boon and T.H. Seligman, J. Math. Phys. **14**, 1224 (1973).
[7]C.A. Hurst (private communication and to be published).
[8]V.I. Arnold, *Mathematical Methods in Classical Mechanics* (in Russian) (Nauka, Moscow, 1974).
[9]J.M. Souriau, *Structure des Systèmes Dynamiques* (Dunod, Paris, 1970).
[10]M.Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772, 1780 (1971).
[11]J. Williamson, Am. J. Math. **58**, 141 (1936); **59**, 599 (1937); **61**, 897 (1939).
[12]H. Zassenhaus, Can. Math. Bull. **1**, 31, 101, 183 (1958).
[13]G.E. Wall, J. Austr. Math. Soc. **3**, 1 (1963).
[14]A.J. Laub and K. Meyer, Celestial Mech. **9**, 213 (1974).
[15]D.A. Suprunenko and R.I. Tyshkevich, *Commutative Matrices* (Academic, New York, 1968).
[16]A.I. Maltsev, Izv. Akad. Nauk SSSR Ser. Mat. **9**, 291 (1943) [Ann. Math. Soc. **9**, 214 (1962)].
[17]M. Sugiura, J. Math. Soc. Jpn. **11**, 374 (1959).
[18]H. Zassenhaus, "On the Maximal Abelian Subalgebras of the Classical Lie Algebras," Preprint CRM-764, Montréal 1978.
[19]J. Patera, P. Winternitz, and H. Zassenhaus, "Maximal Abelian Subalgebras of Symplectic Lie Algebras," Preprint CRM-814, Montréal 1978.
[20]I. Friš and P. Winternitz, Yad. Fiz. **1**, 889 (1965) [Sov. J. Nucl. Phys. **1**, 636 (1965)]; P. Winternitz, I. Lukač, and Ya.A. Smorodinskiĭ, Yad. Fiz. **7**, 192 (1968) [Sov. J. Nucl. Phys. **7**, 139 (1968)]; A. Makarov, Ya.A. Smorodinskiĭ, Kh. Valiev, and P. Winternitz, Nuovo Cimento A **52**, 1061 (1967); P. Winternitz, Ya.A. Smorodinskiĭ, M. Uhliř, and I. Friš, Yad. Fiz. **4**, 625 (1966) [Sov. J. Nucl. Phys. **4**, 444 (1967)].
[21]J. Patera and P. Winternitz, J. Math. Phys. **14**, 1130 (1973); J. Chem. Phys. **65**, 2725 (1976).
[22]W. Miller, Jr., SIAM J. Math. Anal. **4**, 314 (1973); **5**, 626 (1974).
[23]E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **15**, 1025, 1263, 1728 (1974); **16**, 2507 (1975); **17**, 331, 356, 369 (1976); **18**, 1 (1977).
[24]C. Boyer, E.G. Kalnins, and W. Miller, Jr., J. Math. Phys. **16**, 499, 512 (1975); Nagoya Math. J. **60**, 35 (1976).
[25]E.G. Kalnins, W. Miller, Jr., and P. Winternitz, SIAM J. Appl. Math. **30**, 630 (1976).
[26]W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., 1977).
[27]W. Miller, Jr., J. Patera, and P. Winternitz, "Subgroups of Lie Groups and Separation of Variables," Preprint CRM-813, Montréal 1978.
[28]M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973).
[29]M. Moshinsky, T.H. Seligman, and K.B. Wolf, J. Math. Phys. **13**, 901 (1972).
[30]J.D. Louck, M. Moshinsky, and K.B. Wolf, J. Math. Phys. **14**, 692, 696 (1973).
[31]M. Moshinsky, J. Patera, and P. Winternitz, J. Math. Phys. **16**, 82 (1975); M. Moshinsky and J. Patera, J. Math. Phys. **16**, 1866 (1975).
[32]P. Kramer, M. Moshinsky, and T.H. Seligman, "Complex Extensions of Canonical Transformations and their Representations in Quantum Mechanics," in *Group Theory and Applications III*, edited by E.M. Loebl (Academic, New York, 1975).
[33]C. Itzykson, Commun. Math. Phys. **4**, 92 (1967); V. Bargmann, "Group Representations on Hilbert Spaces of Analytic Functions," in *Analytic Methods in Mathematical Physics*, edited by R.P. Gilbert and R.G. Newton (Gordon and Breach, New York, 1968).
[34]G. Burdet and M. Perrin, J. Math. Phys. **16**, 1692, 2172 (1975).
[35]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1378, 1932 (1974); **16**, 1597, 1615 (1975); **17**, 717 (1976); J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977, 986 (1976); Can. J. Phys. **54**, 950 (1976).
[36]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).

[37]C.P. Boyer and K.B. Wolf, Rev. Mex. Fis. **25**, 31 (1976).
[38]I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[39]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[40]P.M. Morse and H. Feshbach, *Methods of Mathematical Physics* (McGraw-Hill, New York, 1953).
[41]T.A. Brody and M. Moshinsky, *Tables of Transformation Brackets* (Gordon and Breach, New York, 1967); see also Rev. Mex. Fis. **9**, 181 (1960).
[42]M. Hamermesh, *Group Theory and its Applications to Physical Problems* (Addison-Wesley, Reading, Mass., 1964).

N. Fröman's phase-integrals *versus* JWKB: Documentation of the advantages of the former in higher-order approximations

Örjan Dammert and Per Olof Fröman
Institute of Theoretical Physics, University of Uppsala, Thunbergsvägen 3, S-752 38, Uppsala, Sweden

(Received 30 October 1979; accepted for publication 22 February 1980)

The purpose of the present paper is to clarify in a general way and to document the great advantage of using, instead of the higher-order JWKB-approximations, certain related but much simpler higher-order phase-integral approximations.

1. INTRODUCTION

If one wants to achieve a higher accuracy than that obtainable by means of the first-order JWKB-approximation, it may seem natural to turn the attention to the higher-order JWKB-approximations. There exist, however, certain related but much simpler higher-order phase-integral approximations, first systematically introduced by N. Fröman[1] and later generalized by N. Fröman and P. O. Fröman (Ref. 2, and pp. 126–31 in Ref. 3), which are preferable from several points of view. In Sec. 2 we shall make some general comments on and comparisons between the above-mentioned two kinds of approximations. In order to demonstrate in some detail how much simpler and more satisfactory the treatment of connection problems becomes if one uses the higher-order phase-integral approximations instead of the higher-order JWKB-approximations we shall in Sec. 3 analyze a treatment in which the higher-order JWKB-approximations were used.

2. GENERAL REMARKS ON THE HIGHER-ORDER JWKB-APPROXIMATIONS AND THE HIGHER-ORDER PHASE-INTEGRAL APPROXIMATIONS

The development of a rigorous and general theory for treating connection problems for higher-order phase-integral approximations, with complete control of the approximations involved and estimates of the errors, proceeded in three steps. The first step[4] was the development of a rigorous theory for mastering the connection problems of the first-order JWKB-approximation and certain modifications of it. The second and most decisive step was that N. Fröman[1] showed that in the infinite JWKB-series the sum of the odd-order terms can be exactly expressed in terms of the sum of the even-order terms. This yields a simplified expression for the wavefunction which, in a classically allowed region, displays a very simple relation between phase and amplitude. When the infinite series occurring in this expression for the wavefunction is truncated, one gets approximations with analytical properties quite different from those of the higher-order JWKB-approximations in the regions around the transition points. This is described in some detail on pp. 452–60 in Ref. 5. The analytical expressions for the higher-order approximations obtained through the above-mentioned transformation [cf. Eq. (7) in Ref. 5], followed by truncation, are considerably simpler than those of the higher-order JWKB-approximations [cf. Eqs. (5a–c) in Ref. 5], but, sufficiently far away from the transition points, the two kinds of approximations differ numerically very little from each other. The third step (Ref. 2, and pp. 126–31 in Ref. 3) was the construction of a general and consistent modification procedure for generalizing the actual phase-integral approximations of arbitrary order so that they can also be used in certain cases where the corresponding unmodified approximations would break down.

To distinguish these new approximations, obtained as described above, from the higher-order JWKB-approximations, we shall call them higher-order phase-integral approximations, or more precisely, in accordance with a terminology introduced by McHugh (cf. p. 280 in Ref. 6), higher-order phase-integral approximations of symmetric form. To retain the close correspondence between the higher-order JWKB-approximations and the higher-order phase-integral approximations, odd integers are used to designate the orders of the latter approximations (cf. p. 454 in Ref. 5).

It should be noted that, if actually pursued to higher orders, the procedure used by Messiah in Sec. 7 of Ch. VI of Ref. 7 would yield the unmodified version of this new type of higher-order phase-integral approximations instead of the higher-order JWKB-approximations. The distinction between these two kinds of approximations does not appear in Messiah's treatment, since he restricts himself to deriving the first-order approximation. Broer[8] used the same procedure as Messiah and pursued it by deriving the first two higher-order unmodified approximations, but he did not distinguish the resulting higher-order approximations from the higher-order JWKB-approximations. The unmodified version of the new kind of higher-order phase-integral approximations also appears implicitly in a paper by Bertocchi, Fubini, and Furlan,[9] but the possibility of using these approximations for solving connection problems is neither indicated nor made use of by these authors. The unmodified higher-order phase-integral approximations are actually obtained if the expansion of $a(x,\lambda)$ on p. 604 in Ref. 9 is truncated and inserted into (2.11) in Ref. 9, but instead of using these approximations in their reasoning the authors of Ref. 9 use the higher-order JWKB-approximations, which are obtained when the expansions for $a(x,\lambda)$ and $b(x,\lambda)$ on p. 604 in Ref. 9 are truncated and inserted into (2.5) with (2.6) and (2.9) in Ref. 9. That there is actually an essential difference between the two types of approximations (cf. pp. 453–54 in Ref. 5) has not been observed in the treatment in Ref. 9.

The higher-order phase-integral approximations have

 0022-2488/80/071683-05$1.00

certain very important properties which make them much more convenient and useful than the higher-order JWKB-approximations.[5] As already mentioned above, the higher-order phase-integral approximations have much simpler form than the corresponding higher-order JWKB-approximations. In a classically allowed region the higher-order phase-integral approximations have the same phase as the corresponding higher-order JWKB-approximations but differ in amplitude from the latter approximations; cf. pp. 453–54 in Ref. 5. Furthermore, in a classically allowed region the higher-order phase-integral approximations display the same relation between phase and amplitude as does the exact wavefunction (cf. pp. 1612–13 in Ref. 10, p. 1124 in Ref. 11, and p. 545 in Ref. 1) and the first-order JWKB-approximation. This relation between phase and amplitude grants to the higher-order phase-integral approximations the important property of having a Wronskian that is exactly constant. The higher-order JWKB-approximations, however, do not have this property of the exact solutions, in consequence of deviations from the above-mentioned relation between phase and amplitude. As described on p. 456 in Ref. 5, the greatest advantage of using the higher-order phase-integral approximations instead of the higher-order JWKB-approximations lies in the fact that for the former approximations the connection problems, i.e., the Stokes phenomenon ("shifting of the coefficients"), can be handled efficiently and rigorously by an available method, while there exists no such method for the higher-order JWKB-approximations because of their complicated analytical form. The method to be used was worked out for the first-order JWKB-approximation by N. Fröman and P. O. Fröman,[4] and the way to generalize several of the results derived in Ref. 4 for the first-order JWKB-approximation to apply as well to the higher-order phase-integral approximations was devised by N. Fröman.[1] Thus, the approach in Ref. 4 combined with the use of the arbitrary-order phase-integral approximations (Ref. 1, Ref. 2, pp. 126–31 in Ref. 3), yields a rigorous phase-integral method which permits exhaustive analysis and treatment of physical problems.

In several publications[1–5,12–43] this phase-integral method has been developed and used for the treatment of various physical problems. In these publications the authors have, however, only briefly entered upon the difficulties and complications which would in general appear if one instead tried to treat the same problems by means of the higher-order JWKB-approximations.

In a paper by Garola[44] the intention is "to discuss the general problem of the approximations of higher order in the WKB method, extending suitably Kemble's procedure." It had, however, escaped Garola's attention—which is not surprising in view of the evermore growing flow of scientific information—that the corresponding problem for the above-mentioned higher-order phase-integral approximations[1] had already been treated. The existence of Garola's investigation makes it possible to discuss in a concrete and detailed way a treatment performed by means of the higher-order JWKB-approximations,[44] for the purpose of making, at some crucial points, illuminating comparisons with a treatment performed by means of the higher-order phase-integral approximations. Our analysis of Garola's paper, which will be given in the next section, demonstrates the need for making a clear distinction between the higher-order JWKB-approximations and the higher-order phase-integral approximations, and the great advantage of the latter approximations.

3. COMMENTS ON GAROLA'S PAPER

In this section we shall make some comments on Garola's paper[44] and especially point out some equations which are approximate in his treatment but become either exact or considerably simplified in a treatment by means of the higher-order phase-integral approximations. We shall use the same notations as Garola,[44] and, when referring to certain equations, we always mean equations in Garola's paper, unless otherwise stated. We shall restrict ourselves to considering the principal features of Garola's paper and, in general, not bother about the rather large number of mistakes and misprints in the equations.

Garola's starting point is Kemble's[45,46] system of first-order differential equations for the JWKB-coefficients, extended to apply to the use of the higher-order JWKB-approximations. Following Kemble's approach for treating the first-order JWKB-approximation, Garola derives corresponding results pertaining to the higher-order JWKB-approximations. We remark that what we call the first-order JWKB-approximation is called the zeroth-order approximation by Garola; cf. p. 243 in Ref. 44.

In Sec. 1 on pp. 244–50 Garola starts the handling of the higher-order JWKB-approximations along the lines devised by Kemble.[45,46] According to Eqs. (1.2), (1.3), and (1.5), the solution ψ is a linear combination of the functions $\exp\{\int[\pm\lambda a(\xi)-b(\xi)]\,d\xi\}$. From the second one of Eqs. (1.6) it follows that $\exp\{-\int b(\xi)\,d\xi\}=\text{const}/\sqrt{a}$, but Garola does not make full use of this relation. If one does so, one obtains ψ as a linear combination of the functions $\exp\{\pm\lambda\int a(\xi)\,d\xi\}/\sqrt{a}$. If, in these functions, one substitutes for a an expansion in terms of inverse powers of λ^2 (the first one of Eqs. (1.7)) and truncates this series after λ^{-2n}, one obtains what we refer to as the unmodified phase-integral approximation of order $2n+1$. We remark that if the expression given in the second of Eqs. (1.6) is substituted for b in the first one of Eqs. (1.6) one obtains an equation from which the coefficients in the expansion of a, i.e., the first one of Eqs. (1.7), can be calculated. This equation is the same as Eq. (3.6) in Ref. 4 except for the notation. Garola, however, utilizes the higher-order JWKB-approximations. If one uses, instead, the higher-order phase-integral approximations, one has the *exact* relation $g_2=\frac{1}{2}g_1'/g_1$ [cf. the second Eq. (1.6)], i.e., $\exp\{-\int g_2\,d\xi\}=\text{const}/\sqrt{g_1}$. Hence one obtains instead of Garola's formulas (1.11) the formulas

$$F_u=\exp\{+\lambda\int g_1\,d\xi\}/\sqrt{g_1}$$

and

$$F_v=\exp\{-\lambda\int g_1\,d\xi\}/\sqrt{g_1}\,,$$

where

$$g_1=\sum_{k=0}^{n}a_k\lambda^{-2k}\,.$$

In Eqs. (1.12) and (1.13) the expression $(2\lambda g_1 g_2 - \lambda g_1')$ is then *exactly* equal to zero, and hence (1.16) is not an approximate but an *exact* equation. Thus, using the formula $g_2 = \frac{1}{2}g_1'/g_1$, one can, without introducing any approximations, write Eq. (1.16) in the form (1.17) with Q given, not by (1.18), but by

$$-\frac{Q}{(\lambda g_1)^2} = \frac{(-i\lambda\phi)^2 - (-i\lambda g_1)^2}{(-i\lambda g_1)^2} + (-i\lambda g_1)^{-3/2}\frac{d^2}{dx^2}(-i\lambda g_1)^{-1/2}.$$

Comparing this formula with formula (3.5a) in Ref. 4, we see that if Garola had used the higher-order phase-integral approximations instead of the higher-order JWKB-approximations, his quantity $-Q/(\lambda g_1)^2$ would have been the same as ϵ in Ref. 4. Furthermore, the Wronskian (1.22) would have been exactly constant ($= -2\lambda$) and Garola's approximate Eqs. (1.24) and (1.25) would (except for the notation) have been the same as the *exact* Eqs. (3.11) in Ref. 4. The transformation (1.26) which Garola uses for treating his Eqs. (1.24) and (1.25) is unnecessary and only complicates the treatment. It is preferable to solve the system of differential equations (1.24) and (1.25) in a quite straightforward way and obtain the solution in exact form in terms of convergent series. The result is given by Eqs. (3.26) and (3.22a–d) in Ref. 4.

In Sec. 2 on pp. 250–61 Garola continues the handling of the higher-order JWKB-approximations by means of Kemble's[45,46] method. If, instead of the higher-order JWKB-approximations, we use the higher-order phase-integral approximations, for which the relation $2g_1g_2 - g_1' = 0$ is exactly valid, the exact Eq. (2.1) for dA_u/dx and the analogous equation for dA_v/dx become identical to Eq. (1.24) and Eq. (1.25), respectively, without the introduction of any approximation. What has been said previously concerning the use of the transformation (1.26) also applies to the use of the transformation (2.2). Rigorous estimates corresponding to the estimates (2.18) and (2.23), which Garola obtains when $|\exp\{\lambda \int g_1\, d\xi\}|$ is monotonically increasing as one moves from X_0 to X_1 along a path Γ, can easily be obtained from Eqs. (3.26) and (4.3a–d) in Ref. 4. The same assertion is true for the estimates (2.29) and (2.30), which Garola obtains when $|\exp\{\lambda \int g_1\, d\xi\}|$ decreases monotonically as one moves from X_2 to X_3 along a path Δ. We also remark that the condition $(\nu_\Delta/2)(E_2/E_3) = 1 + O(\nu_\Delta)$ at the bottom of p. 256 in Garola's paper is not very realistic, since in general E_2/E_3 is very large compared to unity, and ν_Δ, although small, is not so very small compared to unity. When $|\exp\{\lambda \int g_1\, d\xi\}|$ is first monotonically increasing and then monotonically decreasing along the path considered, Garola obtains the estimates (2.32) and (2.33). These estimates are useful essentially only under the condition that $\nu_\Lambda E/2$ is not large compared to unity, which condition one must impose in order to be able to conclude from (2.32) and (2.33) that A_u and A_v have approximately the same values at the end point of the path as at its initial point. However, the condition just mentioned seems to be fulfilled only in very special cases, and it is probably difficult to find a realistic case of physical interest where it is actually fulfilled. For the treatment of typical physical problems, in which there appears a path on which $|\exp\{\lambda \int g_1\, d\xi\}|$ has a maximum, one needs estimates which are more refined than (2.32) and (2.33). For the phase-integral approximations of arbitrary order (Ref. 1, pp. 452–60 in Ref. 5, Ref. 2, and pp. 126–31 in Ref. 3) such estimates can be obtained by means of the methods developed by N. Fröman and P. O. Fröman.[4] Immediately after his Eq. (2.33) Garola says: "The procedures we have used so far can be easily generalized to demonstrate formulas analogous to (2.32) and (2.33) when we have along Λ many maxima and minima of $|\exp\{2\lambda \int_X^x g_1\, d\xi\}|$". It is, however, too optimistic to think that it is easy to obtain useful estimates of the behavior of A_u and A_v in such general cases. The simplest case of such a kind appears perhaps in the case of barrier transmission when the energy of the particle is close to the energy corresponding to the top of the barrier, but even in this case the procedure to be used is somewhat complicated; cf. pp. 46–51 in Ref. 4 and pp. 616–8 in Ref. 15. Limiting properties of A_u and A_v as $|x| \to \infty$, considered by Garola on pp. 259–60, have been rigorously discussed in Chapters 4 (pp. 27–29) and 10 (pp. 104–5) of Ref. 4. The treatment on pp. 104–5 in Ref. 4 can easily be generalized to apply to such a general path in the complex plane as Garola considers. When this is done, one obtains (from Eqs. (10.8a,b) and (4.3a–d) in Ref.4) estimates which correspond to a refinement of Garola's Eqs. (2.38) and (2.39). On p. 260 Garola says "we could extend our inequalities to Λ-paths along which $|\exp\{2\lambda \int_X^x g_1\, d\xi\}|$ is not monotonically decreasing when coming from infinity". Certainly such general estimates can be derived and in essentially the same way as Eqs. (2.32) and (2.33) are derived from the estimates (2.18), (2.23) and (2.29), (2.30) pertaining to the simple paths along which $|\exp\{2\lambda \int_X^x g_1\, d\xi\}|$ changes monotonically when x moves from one end point to the other. These general estimates, however, are not as useful as might be thought, not only because of the roughness of the estimates, but basically for the reason that the estimated quantities simply are not small in realistic situations. Even for the relatively simple case of a path Λ joining in the complex plane two real points ξ and η, lying on opposite sides of a system of two well separated potential barriers, it can be shown by means of methods developed in Ref. 4 and estimates given in Ref. 16 that, in general, neither of the quantities $|A_u(\xi) - A_u(\eta)|$ and $|A_v(\xi) - A_v(\eta)|$, which Garola intends to estimate, is small. If one wants to keep track of the sometimes drastically changing values of A_u and A_v (Stokes phenomenon) along a path of nonmonotonicity, it is therefore not enough to have estimates of the kind considered by Garola. The inability of such estimates to account for changes in A_u and A_v along a path of nonmonotonicity, together with Garola's erroneous belief in the smallness of his estimated quantities for more complicated paths of nonmonotonicity as well, explains why Garola's attempt to treat certain problems concerning a system of several potential barriers as well as a system of several potential wells is not successful and why his results actually are limited to a single potential barrier and a single potential well, respectively. We want to point out that the method in Ref. 4, when applied to the handling of the phase-integral approximations of arbitrary order (Ref. 1, pp. 452–60 in Ref. 5, Ref. 2, and pp. 126–

31 in Ref. 3), provides us with an effective means to treat these problems.

In Sec. 3, pp. 261–5, Garola continues the discussion of the connection problem for the higher-order JWKB-approximations, specializing to the case when the function $\lambda^2\phi^2$ in Eq. (1.1) is real on the real axis. If one uses, instead, the higher-order phase-integral approximations, Garola's Eqs. (3.5) and (3.6) are replaced by the previously mentioned exact relation $\exp\{-\int g_2\,d\xi\} = \text{const}/\sqrt{g_1}$. Garola's Eqs. (3.8) correspond to Eq. (3.26) in Ref. 4, and the matrix $\{g\}$ in Garola's Eq. (3.9) corresponds to the matrix F in Ref. 4, for which there are the explicit formulas (3.17) and (3.22a–d) in Ref. 4. When one uses the higher-order phase-integral approximations, Garola's Eq. (3.15) is replaced by the much simpler relation $\det\{g\} = 1$. Thus Garola's Eq. (3.17), which is approximate since he uses the higher-order JWKB-approximations, becomes *exact* if one uses instead the higher-order phase-integral approximations. On pp. 17–8 in Ref. 4 there is a very simple proof for this exact relation, which is given as Eq. (3.19). Since the function $\lambda^2\phi^2$ in Garola's Eq. (1.1) is assumed to be real on the real axis, one can choose ψ to be real on the real axis (Garola's Eqs. (3.18) and (3.19)), thus imposing a condition upon the matrix $\{g\}$, as Garola also mentions on p. 265. The resulting conditions on this matrix $\{g\}$, i.e., on the matrix F in Ref. 4, are called symmetry relations in Ref. 4 and are given there explicitly in very general form in Eqs. (5.7a,b), (5.8a–d) and (5.9a,b).

Section 4 on p. 266 is an introduction to the treatment of the radial wave equation. In (4.1) Garola gives the equation which results when one transforms the original radial wave equation by a Langer transformation. On pp. 126–31 in Ref. 3 it is shown that one need not perform such a transformation explicitly in every application but can, instead, directly use final expressions for modified phase-integral approximations.

In Sec. 5, on pp. 266–71, Garola treats a radial scattering problem. A formula identical (apart from the notation) to Garola's Eq. (5.14), when corrected for misprints, can be obtained directly by the use of modified higher-order phase-integral approximations (Ref. 2, pp. 126–31 in Ref. 3) together with the results in Chapter 4 of Ref. 4 and the connection formula (21) in Ref. 5. As on pp. 269–71 in Garola's paper, one can then obtain formulas identical (apart from the notation) to Garola's Eqs. (5.17) and (5.20), when corrected for misprints. In Eqs. (5.21) and (5.22) Garola gives approximate expressions for the g-matrix. Equation (5.22) is in agreement with the estimates on pp. 41–2 in Ref. 4 in connection with Fig. 6.3 in Ref. 4, if these estimates are generalized to the phase-integral approximations of arbitrary order, and if the point x'' in Ref. 4 is assumed to be a simple zero of $Q^2(z)$ instead of a simple zero of $q^2(z)$, and if the resulting estimates are changed to correspond to Garola's choice of phase for $-i\lambda\phi$. Estimates corresponding to Garola's Eq. (5.21) do not appear in Ref. 4, since there was no need for such estimates there.

In Sec. 6 on pp. 271–5 Garola treats a radial bound-state problem. The quantization condition (6.8), valid for a single-well potential, can be simplified into (cf. pp. 719–20 in Ref. 47, pp. 186–7 in Ref. 48, and pp. 98–9 in Ref. 26)

$$\tfrac{1}{2}\oint(-i\lambda g_1)\,dx = (n+\tfrac{1}{2})\pi\,.$$

This is also the form in which the quantization condition is directly obtained when one uses the higher-order phase-integral approximations (cf. Ref. 1) instead of the higher-order JWKB-approximations, as can easily be verified by using the exact relation $g_2 = \frac{1}{2}g_1'/g_1$ in Garola's Eq. (6.8). In connection with Garola's attempt on p. 274 to generalize (6.8) to the case of a system of potential wells, we remark that the quantization condition (6.8) cannot be valid for a multi-well potential. It is, for example, easily seen that from Garola's generalization of (6.8) one does not obtain any quantization condition for a symmetric double-oscillator. Garola's generalization, on p. 274, of the quantization condition (6.8) is thus erroneous. Bound states have, on the other hand, been treated correctly by means of phase-integral approximations of arbitrary order and with the possibility of error estimates for a single-well potential in Ref. 1, for a double-well potential in Refs. 14, 17, and 20, and for a more general system of potential wells in Ref. 19.

In Sec. 7, on pp. 275–80, Garola considers sub-barrier penetration through potential barriers. Eqs. (7.8), (7.23), and (7.24) become simplified when one replaces the higher-order JWKB-approximations by the higher-order phase-integral approximations, since $g_1' - 2g_1g_2$ then becomes exactly equal to zero. The right-hand members of Eqs. (7.10), (7.11), (7.18), and (7.24) should be multiplied by the factor ± 1, where the upper sign pertains to the case of an odd number of barriers and the lower sign to the case of an even number. With this correction introduced, Eqs. (7.11) correspond (except for the choice of sign for $-i\lambda\phi$) to the symmetry relations (22a,b) in Ref. 15 and (14a,b) in Ref. 16. Garola's Eq. (7.14), which is valid only for the case of a single potential barrier, corresponds to the estimate (6.31) in Ref. 4, the difference in appearance being due to the difference in sign for $-i\lambda\phi$ in Garola's paper and Q in Ref. 4. Garola's Eqs. (7.19), (7.27), (7.28), (7.30), and (7.31) are based on his Eq. (7.14) and are thus valid only for the case of a single potential barrier. When deriving (7.28) and (7.31), Garola puts $\exp\{R\int_{\eta}^{\infty}(g_1' - 2g_1g_2)/g_1\,d\xi\} = 1$. This relation, which is approximate in Garola's treatment by means of the higher-order JWKB-approximations, becomes exact when one uses instead the higher-order phase-integral approximations. Garola's intention to treat transmission through a system of n potential barriers (cf. p. 275 in Garola's paper) is too optimistic, and his treatment is, in fact, restricted to a single potential barrier. We find strong support for this opinion by comparing Garola's final formula (7.28) for the transmission coefficient, on the one hand with the formula (80b) in Ref. 15 for the transmission coefficient of a single potential barrier and, on the other hand with formula (41) in Ref. 16 for the transmission coefficient of a system of two potential barriers. Detailed, general treatments of transmission problems by means of phase-integral approximations of arbitrary order have been given in Ref. 15 for a single potential barrier, in Ref. 16 for a system of two well separated potential barriers, in Ref. 43 for a system of an arbitrary number of well separated potential barriers, and in

Ref. 39 for a periodic potential consisting of an infinite number of well separated simple barriers.

[1]N. Fröman, Ark. Fys. **32**, 541 (1966).
[2]N. Fröman and P. O. Fröman, Ann. Phys. (N.Y.) **83**, 103 (1974).
[3]N. Fröman and P. O. Fröman, Nuovo Cimento B **20**, 121 (1974).
[4]N. Fröman and P. O. Fröman, JWKB-*Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965). Russian translation: (MIR, Moscow, 1967).
[5]N. Fröman, Ann. Phys. (N.Y.) **61**, 451 (1970).
[6]J. A. M. McHugh, Archive for History of Exact Sciences **7**, 277 (1971).
[7]A. Messiah, *Mécanique Quantique* (Dunod, Paris, 1959), Tome I, p. 195. English translation: A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1967), Vol. I, p. 232.
[8]L. J. F. Broer, Appl. Sci. Res. B **10**, 110 (1963).
[9]L. Bertocchi, S. Fubini, and G. Furlan, Nuovo Cimento **35**, 599 (1965).
[10]L. A. Young, Phys. Rev. **38**, 1612 (1931).
[11]J. A. Wheeler, Phys. Rev. **52**, 1123 (1937).
[12]N. Fröman, Ark. Fys. **31**, 381 (1966).
[13]N. Fröman, Ark. Fys. **31**, 445 (1966).
[14]N. Fröman, Ark. Fys. **32**, 79 (1966).
[15]N. Fröman and P. O. Fröman, Nucl. Phys. A **147**, 606 (1970).
[16]N. Fröman and Ö. Dammert, Nucl. Phys. A **147**, 627 (1970).
[17]N. Fröman and U. Myhrman, Ark. Fys. **40**, 497 (1970).
[18]S. Yngve, J. Math. Phys. **12**, 114 (1971).
[19]S. Yngve, J. Math. Phys. **13**, 324 (1972).
[20]N. Fröman, P. O. Fröman, U. Myhrman, and R. Paulsson, Ann. Phys. (N.Y.) **74**, 314 (1972).
[21]N. Fröman and P. O. Fröman, Phys. Rev. A **6**, 2064 (1972).
[22]N. Fröman, Phys. Lett. A **48**, 137 (1974).
[23]P. O. Fröman, Ann. Phys. (N.Y.) **88**, 621 (1974).
[24]F. Karlsson, Phys. Rev. D **11**, 2120 (1975).
[25]B. Thidé, J. Math. Phys. **17**, 1222 (1976).
[26]N. Fröman and P. O. Fröman, J. Math. Phys. **18**, 96 (1977).
[27]B. Lundborg, Math. Proc. Cambridge Philos. Soc. **81**, 463 (1977).
[28]N. Fröman and P. O. Fröman, J. Math. Phys. **18**, 903 (1977).
[29]N. Fröman and W. Mrazek, J. Phys. A: Math., Gen. **10**, 1287 (1977).
[30]N. Fröman, Phys. Rev. A **17**, 493 (1978).
[31]N. Fröman, J. Math. Phys. **19**, 1141 (1978).
[32]N. Fröman and P. O. Fröman, J. Math. Phys. **19**, 1823 (1978).
[33]N. Fröman and P. O. Fröman, J. Math. Phys. **19**, 1830 (1978).
[34]N. Fröman and P. O. Fröman, Ann. Phys. (N.Y.) **115**, 269 (1978).
[35]G. Drukarev, N. Fröman, and P. O. Fröman, J. Phys. A: Math., Gen. **12**, 171 (1979).
[36]B. Lundborg, Math. Proc. Cambridge Philos. Soc. **85**, 493 (1979).
[37]N. Fröman and K. -E. Thylwe, J. Math. Phys. **20**, 1716 (1979).
[38]N. Fröman, P. O. Fröman, and F. Karlsson, Mol. Phys. **38**, 749 (1979).
[39]N. Fröman, J. Phys. A: Math., Gen. **12**, 2355 (1979).
[40]P. O. Fröman, F. Karlsson, and M. Lakshmanan, Phys. Rev. D **20**, 3435 (1979).
[41]N. Fröman, Semi-classical and higher-order approximations: Properties. Solution of connection problems, Proceedings of the Advanced Study Institute "*Semiclassical Methods in Molecular Scattering and Spectroscopy*," held in Cambridge, England, 9–22 September 1979, edited by M. S. Child (Reidel, Dordrecht, 1980).
[42]B. Thidé, J. Math. Phys. **21**, 1408 (1980).
[43]Ö. Dammert, "Transmission through a system of potential barriers" (to be published).
[44]C. Garola, Nuovo Cimento A **64**, 243 (1969).
[45]E. C. Kemble, Phys. Rev. **48**, 549 (1935).
[46]E. C. Kemble, *The Fundamental Principles of Quantum Mechanics*, Sec. 21 and Appendix D (McGraw–Hill, New York, 1937), (unabridged and corrected republication. Dover, New York, 1958).
[47]J. L. Dunham, Phys. Rev. **41**, 713 (1932).
[48]P. Urban, Acta Phys. Austriaca **6**, 181 (1952).

Semiclassical results on normalization of bound state wavefunctions

Peter Moxhay and Jonathan L. Rosner
School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455

(Received 25 February 1980; accepted for publication 20 March 1980)

Some methods are introduced and tested for approximating s-wave Schrödinger wavefunctions at the origin, and for examining their dependence on reduced mass of the bound system. Methods are discussed both for nonsingular and singular potentials. The approximations, previously tested for power-law potentials, are found to be exact for the Hulthén potential and excellent for a wide range of parameters in Coulomb + linear potentials: $V(r) = -A/r + Br$. General theorems for mass dependences are reviewed and extended; these theorems apply to potentials with a definite sign of d^2V/dr^2. Some specific instances in which d^2V/dr^2 does not have a unique sign and the theorems do not hold are also discussed.

I. INTRODUCTION

Recent attempts to interpret the mass spectra and leptonic-decay widths of the heavy mesons (ψ and Υ families) by treating these particles as bound states of a quark and an antiquark have met with considerable success.[1-3] These positronium-like models make use of nonrelativistic quantum mechanics (Schrödinger's equation) and a central potential $V(r)$. Since the leptonic-decay width of an s-wave state is proportional to $|\Psi(0)|^2$, the square of the wavefunction at the origin, it is of interest to determine how $|\Psi(0)|^2$ varies as a function of the reduced mass μ and parameters in the potential. The variation with μ is of particular importance, since there is some evidence[4] for universality of the interquark potential. By solving this problem we may hope to extrapolate present experimental information to systems of heavier quarks.

We have investigated a variety of monotonically increasing potentials in the semiclassical approximation to gain some insight into the behavior of $|\Psi(0)|^2$ as a function of μ. As a corollary to these efforts, we have obtained several other results:

(i) An expression for $|\Psi(0)|^2$ has been obtained for potentials with a singularity of the form r^{-s} near $r = 0$.[5] This expression is very accurate for power-law potentials, and is exact for both the Coulomb and Hulthén potentials.

(ii) The quantization condition and wavefunction normalization are obtained semiclassically for an arbitrary potential of the form

$$V(r) = -A/r + Br. \tag{1}$$

As this potential is the basis of much successful quarkonium phenomenology,[6,7,1] our results may be of some use as a ready reference. We find our approximations to be particularly good when the Coulomb term dominates. This is expected to be the situation for very heavy quarks.

Since[8]

$$|\Psi(0)|^2 = \frac{\mu}{2\pi\hbar^2}\left\langle\frac{dV}{dr}\right\rangle, \tag{2}$$

the quantity $|\Psi(0)|^2/\mu$ will be of particular interest to us. Many of the potentials used to describe the ψ and Υ families have $d^2V/dr^2 \leqslant 0$: They become steeper near the origin. Naively one would expect $|\Psi(0)|^2/\mu$ to increase for such potentials as μ is raised, since shorter Compton wavelengths should lead to wavefunctions that sample V at shorter distances. This expectation is borne out in the semiclassical limit. The degree to which it is true when one cannot neglect the quantum-mechanical oscillations of the wavefunction is still an open question, but certain results already have been obtained for nodeless wavefunctions or special potentials.[9]

In Sec. II we shall present semiclassical results for wavefunction normalization in nonsingular potentials. These results are extended to the case of a singularity at the origin in Sec. III, where a modification to the quantization condition is also presented. The results are illustrated for power-law potentials in Sec. IV, for the Hulthén potential in Sec. V, and for the Coulomb-plus-linear potential (1) in Sec. VI. We return to the problem of mass dependence in Sec. VII, quoting a result of Martin in the classical limit. Section VIII is devoted to a number of simple examples of mass dependences. Section IX contains our conclusions.

II. SEMICLASSICAL APPROXIMATION FOR NONSINGULAR POTENTIALS

For motion with angular momentum $l = 0$ in a centrally symmetric field, the reduced Schrödinger equation is

$$\frac{d^2u(r)}{dr^2} + \frac{2\mu}{\hbar^2}[E - V(r)]u(r) = 0. \tag{3}$$

The full wavefunction is

$$\Psi(r) = u(r)/\sqrt{4\pi}r, \tag{4}$$

and

$$u(0) = 0. \tag{5}$$

The semiclassical approximation to $u(r)$ for a nonsingular potential may be written

$$u(r) = \frac{N}{[p(r)]^{1/2}}\sin\left(\frac{1}{\hbar}\int_0^r p(r')dr'\right), \tag{6}$$

where

$$p(r) = \sqrt{2\mu[E - V(r)]} \tag{7}$$

is the classical momentum. Then

 0022-2488/80/071688-07$1.00

$$|\Psi(0)|^2 = [u'(0)]^2/4\pi = \sqrt{2\mu E}\, N^2/4\pi\hbar^2. \tag{8}$$

The constant N is determined by normalizing the radial wavefunction:

$$1 = \int_0^{r_0} dr[u(r)]^2 \simeq \frac{N^2}{2}\int_0^{r_0}\frac{dr}{p(r)}. \tag{9}$$

(The integral has been taken only out to the turning point, and the average of the $\sin^2$ factor has been approximated by $\frac{1}{2}$.) We then find

$$|\Psi(0)|^2 = \frac{\mu E^{1/2}}{\pi\hbar^2}\int_0^{r_0}\frac{dr}{[E-V(r)]^{1/2}}. \tag{10}$$

Recalling Eq. (2) and the subsequent discussion, we shall be interested in the sign of

$$\frac{\partial}{\partial\mu}[|\Psi(0)|^2/\mu] = \frac{\partial E}{\partial\mu}\frac{\partial}{\partial E}[|\Psi(0)|^2/\mu]. \tag{11}$$

We use a well-known theorem[10,11] on the expectation value of the derivative of the Hamiltonian $\hat{H}$ with respect to a parameter λ,

$$\left\langle\frac{\partial\hat{H}}{\partial\lambda}\right\rangle = \frac{\partial E}{\partial\lambda}, \tag{12}$$

which gives

$$\frac{\partial E}{\partial\mu} = -\frac{\langle\hat{T}\rangle}{\mu} < 0. \tag{13}$$

Here $\hat{T}$ is the kinetic-energy operator, whose expectation value is non-negative. Hence the problem reduces to that of finding $\partial(|\Psi(0)|^2/\mu)/\partial E$, which may easily be done with the help of Eq. (10). We shall consider several simple examples in Sec. VIII.

III. SINGULAR POTENTIALS

Let us begin by considering a simple class of potentials

$$V(r) = -A/r^s, \quad A>0, \quad 0<s<2. \tag{14}$$

We restrict attention to $s<2$, since for $s>2$ a particle "falls" to the origin; in some sense these potentials must be considered unphysical.[11] For the potentials (14), the preceding discussion is inapplicable, since the condition $V(0)=0$ is not satisfied. Moreover, the condition for applicability of the semiclassical approximation is[12]

$$\mu\hbar|F|/[p(r)]^3 \ll 1, \tag{15}$$

where $F = -dV/dr$ is the force acting on the particle. In the present case this becomes (with $p \approx \sqrt{2\mu|V|} \approx \sqrt{2\mu A/r^s}$):

$$r \gg (\hbar^2 s^2/8\mu A)^{1/(2-s)}, \tag{16}$$

so that the semiclassical approximation will not hold for small r. We can, however, determine the semiclassical $|\Psi(0)|^2$ by finding the exact solution near $r=0$ and matching it to the semiclassical solution valid for large r.[13] To obtain this exact solution, we neglect E in Schrödinger's equation (3):

$$\frac{d^2u}{dr^2} + \frac{2\mu}{\hbar^2}\frac{A}{r^s}u = 0. \tag{17}$$

By means of the substitutions

$$u(r) = z^q\rho(z), \tag{18}$$

$$q \equiv 1/(2-s), \tag{19}$$

$$r \equiv [\hbar z/2q(2\mu A)^{1/2}]^{2q}, \tag{20}$$

Eq. (17) is brought to the form of Bessel's equation,

$$\rho''(z) + \frac{1}{z}\rho'(z) + \left(1 - \frac{q^2}{z^2}\right)\rho(z) = 0. \tag{21}$$

The solution regular at $z=0$ is

$$\rho(z) = CJ_q(z) \tag{22}$$

or

$$u(r) = Cz^qJ_q(z). \tag{23}$$

This solution is applicable for $|E| \ll |V|$ or

$$r \ll (A/|E|)^{1/s}. \tag{24}$$

The large-r solution is

$$u(r) = \frac{N}{[p(r)]^{1/2}}\cos\left(\frac{1}{\hbar}\int_0^r p(r')dr' + \phi\right). \tag{25}$$

To determine the constant C in terms of N, which is given by the normalization condition (9), we demand that the solutions (23) and (25) be the same in the intermediate region:

$$(\hbar^2 s^2/8\mu A)^q \ll r \ll (A/|E|)^{1/s}. \tag{26}$$

This region must exist for sufficiently small $|E|$. With the asymptotic form of the Bessel function,

$$J_q(z) \simeq \left(\frac{2}{\pi z}\right)^{1/2}\cos\left(z - \frac{q\pi}{2} - \frac{\pi}{4}\right), \tag{27}$$

the solution (18) becomes

$$u(r) = Cz^{q-1/2}\left(\frac{2}{\pi}\right)^{1/2}\cos\left(z - \frac{q\pi}{2} - \frac{\pi}{4}\right). \tag{28}$$

This is to be compared with Eq. (25), which can be written as

$$u(r) \simeq N\left(\frac{r^s}{2\mu A}\right)^{1/4}\cos\left(\frac{2\sqrt{2\mu A r^{2-s}}}{(2-s)\hbar} + \phi\right), \tag{29}$$

and we find that

$$N = C\left(\frac{\hbar}{q\pi}\right)^{1/2}\left(\frac{2q\sqrt{2\mu A}}{\hbar}\right)^q. \tag{30}$$

Next we use the first term of the series expansion of the Bessel function

$$J_q(z) \simeq \frac{z^q}{2^q q!} \tag{31}$$

in order to evaluate $|\Psi(0)|^2 = [u'(0)]^2/4\pi$. The final result, analogous to Eq. (10) for nonsingular potentials, is

$$|\Psi(0)|^2 = q^{sq}\left(\frac{2\mu A}{\hbar^2}\right)^q\frac{\sqrt{2\mu}}{\hbar}\Big/2[\Gamma(q)]^2\int_0^{r_0}\frac{dr}{[E-V(r)]^{1/2}}. \tag{32}$$

An equivalent expression is

$$|\Psi(0)|^2 = \frac{N^2q^{sq}}{4\hbar[\Gamma(q)]^2}\left(\frac{2\mu A}{\hbar^2}\right)^q. \tag{33}$$

The results (32) and (33) may now be extended to apply to *any* potential that goes as r^{-s} $(0<s<2)$ near $r=0$, as

long as a suitable intermediate region exists in which both the solutions (23) and (25) are approximately valid. Examples are the Hulthén potential (Sec. V) and the Coulomb-plus-linear potential (Sec. VI), both of which behave as r^{-1} near the origin.

We may obtain a generalization of the Bohr-Sommerfeld quantization condition. For nonsingular potentials this is

$$\int_0^{r_0} [2\mu(E - V(r))]^{1/2} = (n - \tfrac{1}{4})\pi\hbar. \tag{34}$$

For singular potentials, by matching the solutions (23) and (25) in the region (26) and taking account of (27), we find

$$\int_0^{r_0} dr[2\mu(E - V(r))]^{1/2} = (n - \tfrac{1}{2} + \tfrac{1}{2}q)\pi\hbar = \left(n - \frac{1-s}{2(2-s)}\right)\pi\hbar. \tag{35}$$

Specifically, the integral in (35) should be proportional to n for potentials with a Coulomb singularity at the origin.

IV. POWER-LAW POTENTIALS

We shall be content with a few illustrations here. General expressions[1] are quoted at the end of this section.

For a nonsingular potential

$$V(r) = \lambda r^{\nu} \quad (0 < \nu < \infty), \tag{36}$$

the behavior

$$|\Psi(0)|^2 \sim E^{1-1/\nu} \tag{37}$$

is easily seen by rewriting the integral in Eq. (10) in the form of a variable $r' \equiv r/E^{\nu}$. For $\nu = 1$, $|\Psi(0)|^2$ is independent of E. The semiclassical result

$$|\Psi(0)|^2 = \frac{\lambda}{4\pi} \cdot \frac{2\mu}{\hbar^2} \tag{38}$$

is also the exact one, as may be seen from Eq. (2).

For singular potentials of the form (14), the result

$$|\Psi(0)|^2 \sim E^{1/2 + 1/s} \tag{39}$$

follows from Eq. (32). Thus, for a Coulomb potential, $|\Psi(0)|^2 \sim E^{3/2} \sim n^{-3}$. The exact coefficient is obtained from Eq. (32) in this case.

In Table I we present results for $|\Psi(0)|^2$ and E_n for nonsingular [Eq. (36)] and singular [Eq. (14)] power-law potentials, according to Eqs. (10), (32), (34), and (35). These results have also been quoted in Ref. 1, which may be consulted for further illustrations of their accuracy.

V. THE HULTHÉN POTENTIAL

The s-wave Schrödinger equation may be solved exactly (cf. Ref. 14) for the Hulthén potential

$$V(r) = -V_0 e^{-r/a}/(1 - e^{-r/a}). \tag{40}$$

Here we shall set $a = 2\mu/\hbar^2 = 1$. We shall show that Eqs. (32) and (35) give the exact results.

With

$$\alpha^2 \equiv -E, \quad \beta^2 \equiv V_0, \quad \gamma \equiv \sqrt{\alpha^2 + \beta^2}, \tag{41}$$

the solutions are

$$u(r) = Ne^{-\alpha r}(1 - e^{-r}) \times {}_2F_1(2\alpha + 1 + n, 1 - n; 2\alpha + 1; e^{-r}), \tag{42}$$

with eigenvalue condition

$$\alpha = \gamma - n = \frac{\beta^2 - n^2}{2n}. \tag{43}$$

The normalization condition

$$\int_0^{\infty} [u(r)]^2 dr = 1 \tag{44}$$

implies

$$N^2 = \frac{\Gamma(2\alpha + n)\Gamma(2\alpha + n + 1)(\alpha + n)}{\Gamma(2\alpha)\Gamma(2\alpha + 1)[\Gamma(n + 1)]^2}. \tag{45}$$

We seek

$$|\Psi(0)|^2 = \frac{N^2}{4\pi}[\,{}_2F_1(2\alpha + 1 + n, 1 - n; 2\alpha + 1; 1)\,]^2 = \frac{N^2}{4\pi}\left(\frac{\Gamma(n + 1)\Gamma(2\alpha + 1)}{\Gamma(2\alpha + n)}\right)^2 \tag{46}$$

or

$$|\Psi(0)|^2 = \frac{\alpha(\alpha + n)(2\alpha + n)}{2\pi}. \tag{47}$$

The semiclassical result, according to (32), is

$$|\Psi(0)|^2_{\mathrm{SC}} = \frac{V_0}{2} \Big/ \int_0^{r_0} \frac{dr}{[E + V_0 e^{-r}/(1 - e^{-r})]^{1/2}}. \tag{48}$$

With the substitutions

$$(1 - e^{-r})^{-1} = w^2 + \epsilon + 1, \quad \epsilon \equiv -E/V_0, \tag{49}$$

this becomes an elementary integral, whose value is

$$|\Psi(0)|^2_{\mathrm{SC}} = \frac{V_0^{3/2}}{2\pi} \frac{[\epsilon(\epsilon + 1)]^{1/2}}{(\epsilon + 1)^{1/2} - \epsilon^{1/2}} = |\Psi(0)|^2. \tag{50}$$

TABLE I. Summary of semiclassical results for power-law potentials.

	Nonsingular $V(r) = \lambda r^{\nu}$	Singular $V(r) = -Ar^{-s}$
$\lvert\Psi(0)\rvert^2$	$\dfrac{E^{1-1/\nu}\lambda^{1/\nu}\mu\Gamma(1/2 + 1/\nu)}{\pi^{3/2}\hbar^2\Gamma(1 + 1/\nu)}$	$\lvert E\rvert^{1/2 + 1/s} A^{2(s-1)/s(2-s)} \left(\dfrac{2\mu}{\hbar^2}\right)^{(4-s)/2(2-s)} \dfrac{\Gamma(1/s)}{2(\pi)^{1/2}[\Gamma(1/(2-s))]^2\Gamma(1/s + 1/2)}$
E_n	$\left[\lambda^2\left(\dfrac{\hbar^2}{2\mu}\right)^{\nu}\left(\dfrac{2\nu\sqrt{\pi}\,\Gamma(3/2 + 1/\nu)}{\Gamma(1/\nu)}(n - \tfrac{1}{4})\right)^{2\nu}\right]^{1/(2+\nu)}$	$\left\{A^2\left(\dfrac{\hbar^2}{2\mu}\right)^{-s}\left[\dfrac{2\sqrt{\pi}\,\Gamma(1/s)}{\Gamma(1/s - 1/2)}\left(n - \dfrac{1-s}{2(2-s)}\right)\right]^{-2s}\right\}^{1/(2-s)}$

In the last step we have made use of Eqs. (41) and (43):

$$\alpha = \sqrt{\epsilon V_0}, \quad n = \sqrt{V_0}(\sqrt{\epsilon+1} - \sqrt{\epsilon}). \tag{51}$$

The quantization condition (35) may be evaluated by means of a similar substitution:

$$n_{\rm SC}\pi = \int_0^{r_0} dr[V_0 e^{-r}/(1-e^{-r}) + E]^{1/2} \tag{52}$$

$$= \pi\sqrt{V_0}(\sqrt{\epsilon+1} - \sqrt{\epsilon}), \tag{53}$$

which indeed is the exact result (51).

VI. COULOMB-PLUS-LINEAR POTENTIAL

In this section we evaluate the semiclassical expressions for wavefunction normalizations and quantization condition for the Coulomb-plus-linear potential (1). We shall again set $2\mu = \hbar^2 = 1$.

It becomes convenient to define the variable

$$\xi \equiv E/(E^2 + 4AB)^{1/2}. \tag{54}$$

In the limit of a very Coulombic potential ($E \to -\infty$), the variable ξ tends to -1, while for highly excited levels ($E \to \infty$), whenever $B = 0$ we have $\xi \to 1$. The quantities of interest (32) and (35) are then

$$|\Psi(0)|^2_{\rm SC} = \frac{AB}{4}\left(\frac{\xi}{E}\right)^{1/2} \Big/ g(\xi), \tag{55}$$

$$n_{\rm SC}\pi = \frac{4\sqrt{2}A^{3/4}}{3B^{1/4}} f(\xi), \tag{56}$$

$$g(\xi) \equiv E\left(\frac{1+\xi}{2}\right) - \frac{1-\xi}{2} K\left(\frac{1+\xi}{2}\right), \tag{57}$$

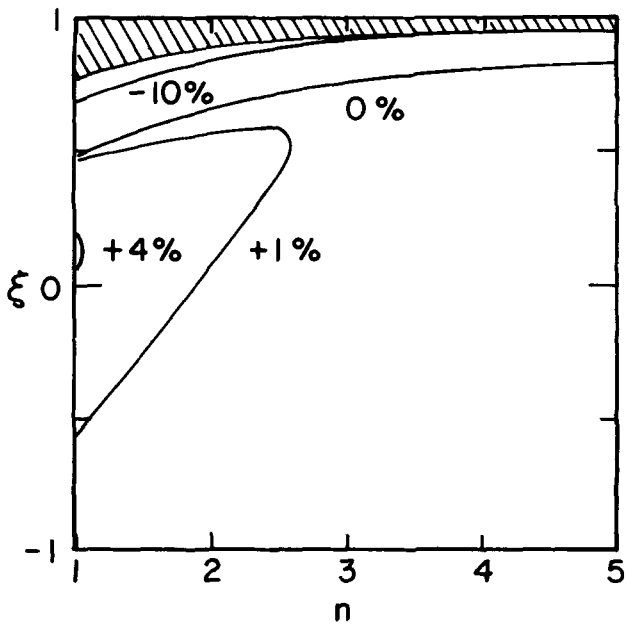


FIG. 1. Contours of $[|\Psi(0)|^2_{\rm SC} - |\Psi(0)|^2]/|\Psi(0)|^2$ as functions of $\xi \equiv E/\sqrt{E^2+4AB}$ and n for a Coulomb + linear potential $V(r) = -A/r + Br$, where $|\Psi(0)|^2_{\rm SC}$ is computed using the semiclassical formula (55). Shaded region denotes $\xi > n_0^{-1}(n)$ [Eq. (62)], for which the present approximation is unjustified.

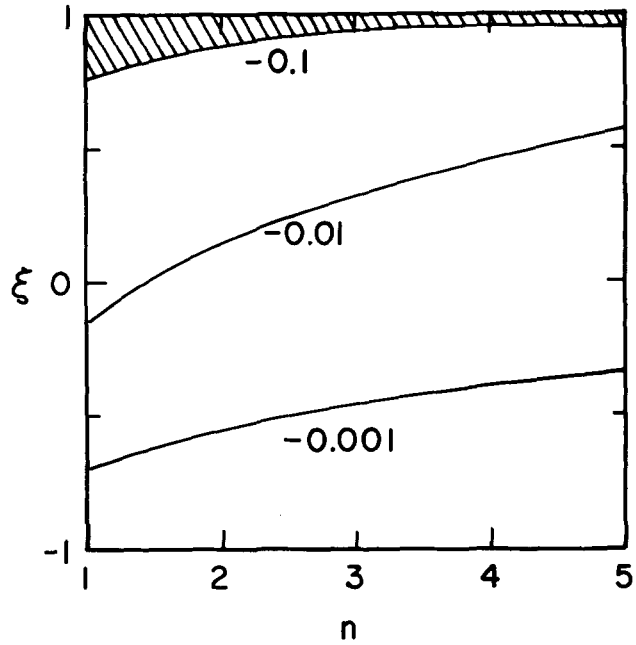


FIG. 2. Contours of $n_{\rm SC} - n$ as functions of ξ and n for a Coulomb + linear potential, where $n_{\rm SC}$ is computed using the semiclassical formula (56). Shaded region denotes $\xi > n_0^{-1}(n)$.

$$f(\xi) \equiv \left[\frac{1-\xi}{2} K\left(\frac{1+\xi}{2}\right) + \xi E\left(\frac{1+\xi}{2}\right)\right] \Big/ (1-\xi^2)^{3/4}, \tag{58}$$

where

$$E(m) \equiv \int_0^{\pi/2} (1 - m\sin^2\theta)^{1/2}\, d\theta \tag{59}$$

and

$$K(m) \equiv \int_0^{\pi/2} (1 - m\sin^2\theta)^{-1/2}\, d\theta \tag{60}$$

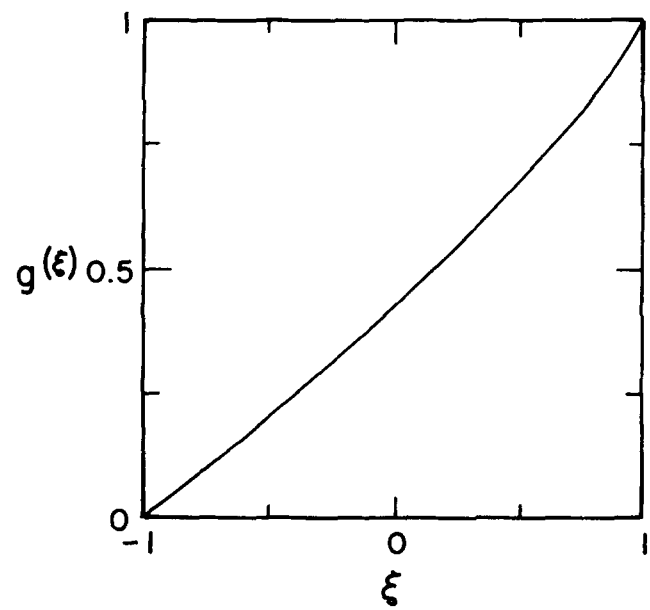


FIG. 3. Plot of the function $g(\xi)$ [Eq. (57)] used in computing $|\Psi(0)|^2_{\rm SC}$ for the Coulomb + linear potential via Eq. (55).

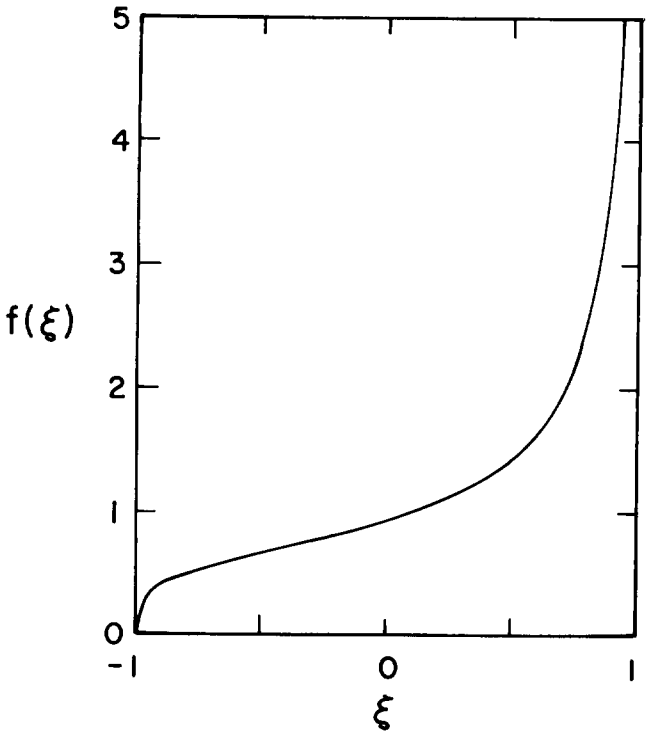


FIG. 4. Plot of the function $f(\xi)$ [Eq. (58)] used in computing n_{SC} for the Coulomb + linear potential via Eq. (56).

are elliptic integrals as defined and tabulated in Ref. 15. The subscript SC denotes the semiclassical approximation.

The expressions (55) and (56) approach the exact (Coulomb) values as $\xi \to -1$. They deteriorate in the limit $\xi \to 1$, since use has been made of formulas (32) and (35), which are appropriate only for singular potentials. In the limit $\xi \to 1$, no suitable matching region (26) exists. [For $\xi = 1$ we should have used instead Eqs. (10) and (34).] The two limits in (26) coalesce when

$$E = 4A^2,$$

$$\xi = (1 + B/4A^3)^{-1/2}, \tag{61}$$

or, with the help of (56),

$$n = n_0(\xi) \equiv \frac{4}{3\pi} f(\xi) \frac{\sqrt{\xi}}{(1-\xi^2)^{1/4}}. \tag{62}$$

The approximations (55) and (56) are expected to be good only for $\xi < n_0^{-1}(n)$. Here they are indeed excellent. Figures 1 and 2 show contours of errors in $|\Psi(0)|^2$ and n as function of n and ξ. The shaded region corresponds to $\xi > n_0^{-1}(n)$, for which no suitable matching region (26) exists.

For convenience we plot the function $g(\xi)$ and $f(\xi)$ in Figs. 3 and 4. Given A, B, and n, one may read off ξ from Eq. (59) and Fig. 4, thereby determining E through Eq. (54).

VII. CLASSICAL MASS DEPENDENCE

Here we investigate the behavior of $|\Psi(0)|^2/\mu$ as a function of μ, using the semiclassical approximations that have been shown to be quite good in the preceding sections. In this section we note theorems that are slight extensions of those quoted previously[1,16]; the next section is devoted to some simple examples.

For nonsingular potentials we may always take $V(0) = 0$. We shall show for such potentials that if $V'(r) > 0$,

$$\operatorname{sgn} \frac{\partial}{\partial \mu}\left(\frac{|\Psi(0)|^2}{\mu}\right) = -\operatorname{sgn} \int_0^E \frac{V'' V\, dV}{(V')^2[1-(V/E)]^{1/2}}. \tag{63}$$

In particular, if V'' always $\geqslant 0$ ($\leqslant 0$), then $|\Psi(0)|^2/\mu \sim \langle dV/dr \rangle$ always decreases (increases) with increasing μ. This corresponds to the intuitive expectation that wave functions for larger μ sample shorter distances in the potential.

To demonstrate Eq. (63) we write Eq. (10) as

$$\frac{|\Psi(0)|^2}{\mu} = \frac{E^{1/2}}{\pi \hbar^2 D}, \tag{64}$$

where

$$D \equiv \int_0^{r_0} \frac{dr}{[E - V(r)]^{1/2}}. \tag{65}$$

According to the discussion in Sec. II,

$$\begin{aligned} \operatorname{sgn}\left[\frac{\partial}{\partial \mu}\left(\frac{|\Psi(0)|^2}{\mu}\right)\right] &= -\operatorname{sgn}\left[\frac{\partial}{\partial E}\left(\frac{|\Psi(0)|^2}{\mu}\right)\right] \\ &= -\operatorname{sgn} \frac{\partial}{\partial E}\left(\frac{E^{1/2}}{D}\right) \\ &= \operatorname{sgn} \frac{\partial}{\partial E}\left(\frac{D}{E^{1/2}}\right). \end{aligned} \tag{66}$$

For $V'(r) > 0$, we may change variables in (65), obtaining[1,16]

$$\begin{aligned} D &= \int_{V(0)}^{E} \frac{dV}{V'(E-V)^{1/2}} \\ &= \frac{2\sqrt{E - V(0)}}{V'(0)} - 2\int_{V(0)}^{E} dV \frac{V''\sqrt{E-V}}{(V')^2}. \end{aligned} \tag{67}$$

If $V(0) = 0$, the first term does not contribute to $\partial(D/E^{1/2})/\partial E$, so that

$$\frac{\partial (D/E^{1/2})}{\partial E} = -\frac{1}{E^2}\int_0^E \frac{V'' V\, dV}{(V')^2[1-(V/E)]^{1/2}}, \tag{68}$$

verifying Eq. (63).

For singular potentials, we can only prove an inequality in one direction:

If $V'' < 0$ and if $V \sim -r^{-s}$ as $r \to 0$, $0 < s < 2$, then

$$\partial(|\Psi(0)|^2/\mu^{q+1/2})/\partial \mu > 0, \quad q \equiv 1/(2-s). \tag{69}$$

By examining Eq. (32), we see that

$$|\Psi(0)|^2/\mu^{q+1/2} = \text{const}/D, \tag{70}$$

so $\partial(|\Psi(0)|^2/\mu^{q+1/2})/\partial \mu > 0$ if $\partial D^{-1}/\partial \mu > 0$, or if $\partial D/\partial E > 0$. The first term of (67) may be seen to vanish if interpreted in the limit $r \to 0$, and

$$\frac{\partial D}{\partial E} = -\int_{-\infty}^{E} dV \frac{V''}{(V')^2(E-V)^{1/2}} \geqslant 0 \quad \text{for } V'' < 0. \tag{71}$$

Since many potentials of interest for charmonium [including Eq. (1)] have $V'' < 0$ and are singular at $r = 0$, Eq. (69) says that $|\Psi(0)|^2$ should increase at least as rapidly as μ. This is the assumption on the basis of which some bounds in Ref. 17 were established, permitting the conclusion [1,17] to be

drawn that the charge of the quarks in the Υ is $|e_Q| = \frac{1}{3}$. For potentials with a Coulomb singularity ($s = 1, q = 1$) at $r = 0$, and with $V'' < 0$, we find that $|\Psi(0)|^2$ should increase at least as rapidly as $\mu^{3/2}$.

VIII. ILLUSTRATIONS FOR NON-POWER-LAW POTENTIALS

A. $V(r) = -V_0/\cosh^2 \alpha r$

This potential has a point of inflection at $V = -\frac{2}{3}V_0$, so the theorems of the preceding section do not apply. For this potential, Eq. (10) gives

$$\frac{|\Psi(0)|^2}{\mu} = \frac{2\alpha^2}{\pi^3\hbar^2}\sqrt{|E|(V_0 - |E|)}\,, \tag{72}$$

which is an increasing (decreasing) function of E for $E < -V_0/2$ ($> -V_0/2$). The critical energy lies above the inflection point $-2V_0/3$ and thus the sign of $\partial[|\Psi(0)|^2/\mu]/\partial\mu$ cannot be directly associated with the value of V'' at the classical turning point, but must depend on the shape of the potential throughout the classical accessible region. This is further illustrated by the following example.

B. Nested square wells

In this case (Fig. 5) application of Eq. (10) leads to the expressions ($\epsilon \equiv E/V_0, \beta \equiv b/a$):

$$\frac{|\Psi(0)|^2}{\mu} = \frac{V_0}{\pi\hbar^2 a}h(\epsilon), \tag{73}$$

where

$$h(\epsilon) = \begin{cases} \epsilon, & \epsilon < 1, \qquad (74) \\ \dfrac{\epsilon\sqrt{\epsilon - 1}}{(\epsilon-1)^{1/2} + (\beta - 1)(\epsilon)^{1/2}}, & \epsilon > 1. \qquad (75) \end{cases}$$

The function $h(\epsilon)$ has the form shown in Fig. 6, which shows

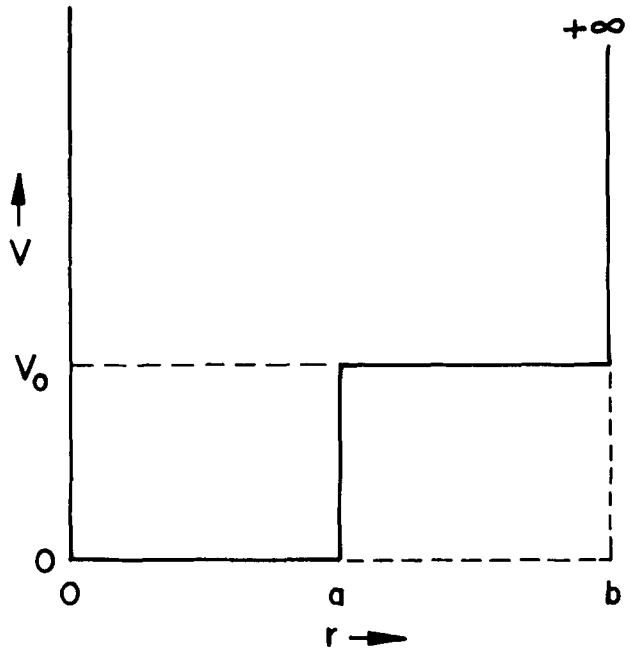


FIG. 5. A system of nested square wells: $V = 0$ for $0 \leqslant r \leqslant a$, $V = V_0$ for $a < r \leqslant b$, $V = \infty$ for $r > b$.

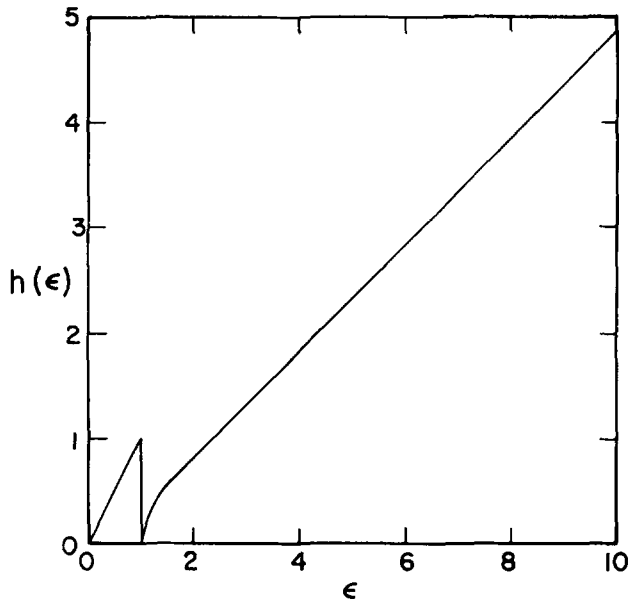


FIG. 6. The dimensionless function $h(\epsilon)$ [Eqs (73)–(75)] describing $|\Psi(0)|^2$ in the potential of Fig. 5. Here we have taken $b/a = 2$.

that $|\Psi(0)|^2/\mu$ is monotonically increasing with ϵ except for a jump. Such behavior is easily understood using the picture of a particle undergoing finite motion in a potential well in classical mechanics. When the particle has just enough energy to surmount the step in the nested square wells, it moves extremely slowly in the region $a < r < b$, and thus spends almost all of its time away from the origin (i.e., away from the region $r < a$).

The analogy between $|\Psi(0)|^2$ and the classical probability density at the origin can be made more explicit. The classical probability density is just proportional to the inverse of the particle velocity:

$$v(r) = \{2[E - V(r)]/\mu\}^{1/2}. \tag{76}$$

Therefore the classical probability that the particle is at the origin [in the purely one-dimensional problem with $V(-r) = V(r)$] is:

$$P(0) = \left(2\sqrt{E}\int_0^{r_0}\frac{dr}{[E - V(r)]^{1/2}}\right)^{-1}, \tag{77}$$

where the classical turning point r_0 is the root of the equation $V(r_0) = E$. The integral is related to the classical period for one-dimensional motion in a symmetric potential:

$$\tau = 4\int_0^{r_0}\frac{dr}{v(r)}. \tag{78}$$

Comparing Eq. (77) with the formula (10) for $|\Psi(0)|^2$, we find

$$\frac{|\Psi(0)|^2}{\mu} = \frac{2E}{\pi\hbar^2}P(0). \tag{79}$$

IX. CONCLUSIONS

We have been concerned in this paper with the behavior of the square of the s-wave wavefunction at $r = 0$ as a function of reduced mass μ in nonrelativistic bound systems.

This problem is of renewed interest because of its applicability to the physics of heavy quarks.

The investigation has been carried out in a semiclassical context. We have presented distinct approximations valid for potentials that are finite at $r = 0$ and for potentials behaving there as r^{-s}, $0 < s < 2$. The approximations have been previously shown[1] to be very good for power-law potentials. We have illustrated their accuracy for two more potentials behaving as r^{-1} at $r = 0$. The approximations have been shown to be exact for the Hulthén potential, and very good for Coulomb-plus-linear potentials as long as the effects of the Coulomb singularity are non-negligible. (This corresponds to the parameter ξ in Sec. VI being sufficiently below 1.)

Two classes of semiclassical results have been obtained for the behavior of $|\Psi(0)|^2/\mu$ as a function of μ. For nonsingular potentials, $\partial(|\Psi(0)|^2/\mu)/\partial\mu > 0\ (< 0)$ if $V'' < 0\ (> 0)$. For potentials behaving as r^{-s} at the origin, $\partial(|\Psi(0)|^2/\mu^{q+1/2})/\partial\mu > 0$ $[q \equiv 1/(2-s)]$ if $V'' < 0$. If the short-distance singularity of V is Coulombic, $|\Psi(0)|^2$ grows with μ at least as rapidly as $\mu^{3/2}$. Since leptonic widths of vector mesons of mass M behave as $|\Psi(0)|^2/M^2$, and $M \sim \mu$ (if we neglect binding effects), one expect these leptonic widths to decrease no more rapidly than $\mu^{-1/2}$.[18]

We have illustrated the mass dependence of $|\Psi(0)|^2$ when V'' does not have a definite sign by two examples. When $V(r) = -V_0/\cosh^2\alpha r$, $|\Psi(0)|^2/\mu$ decreases with increasing μ for $-V_0 < E < -V_0/2$, and it increases with increasing μ for $-V_0/2 < E < 0$. There is a correlation with the shape of the potential here but it is not directly associated with the value of V'' at the classical turning point. When $V(r)$ is a double (nested) well, we have arrived at a more classical understanding of why $|\Psi(0)|^2/\mu$ need not be monotonic in E: The particle can get "trapped" in the outer reaches of the potential if it moves very slowly there.

We recognize that some of our approximations are justified only by their apparent accuracy; in particular, we are suprised by the quality of the approximations in Figs. 1 and 2. A more rigorous discussion, and an extension of the theorems of Sec. VII, might form a fruitful area for further work.

ACKNOWLEDGMENTS

We wish to thank C. N. Leung for informative discussions and for his participation in parts of this work. J. L. R. also wishes to acknowledge the hospitality of the theory group at Fermilab and to thank J. D. Bjorken, R. P. Feynman, A. Martin, and C. Quigg for helpful comments. This work was performed in part under the sponshorship of the U.S. Department of Energy under Contract No. EY-76-C-02-1764.

[1] C. Quigg and Jonathan L. Rosner, Phys. Rep. C **56**, 167 (1979).
[2] C. Quigg, in *Proceedings of the 1979 International Symposium on Lepton and Photon Interactions at High Energy*, edited by T. B. W. Kirk and H. D. I. Abarbanel (Fermilab, Batavia, Ill., 1979), p. 239.
[3] Jonathan L. Rosner, in *Particles and Fields-1979* (American Institute of Physics Conference Proceedings No. 59, Particles and Fields subseries No. 19), edited by B. Margolis (American Institute of Physics, New York, 1980), p. 325.
[4] C. Quigg, H. B. Thacker, and Jonathan L. Rosner, Phys. Rev. D **21**, 234 (1980).
[5] Earlier results, including ones for potentials with a Coulomb singularity at the origin, are given by Nanny Fröman, J. Math. Phys. **19**, 1141 (1978); Phys. Rev. A **17**, 493 (1978); and Ref. 1.
[6] E. Eichten *et al.*, Phys. Rev. D **17**, 3090 (1978); *ibid.* **21**, 203 (1980).
[7] C. Quigg and J. Rosner, Phys. Lett. B **71**, 153 (1977).
[8] See, e.g., Ref. 1; J. Schwinger, Harvard University Lecture Notes (unpublished); E. Eichten *et al.*, Phys. Rev. Lett. **34**, 369 (1975); Nanny Fröman, Phys. Rev. A **17**, 493 (1978); D. M. Fradkin and F. Calogero, Nucl. Phys. **75**, 475 (1966); Nanny Fröman, J. Math. Phys. **19**, 1141 (1978). This last work also contains numerous results on an alternative method for obtaining $|\Psi(0)|^2$ semiclassically, both for nonsingular and singular potentials. The methods presented there appear to be distinct from the ones in the present work.
[9] C. N. Leung and Jonathan L. Rosner, J. Math. Phys. **20**, 1435 (1979).
[10] H. Hellmann, *Einführung in die Quantenchemie* (F. Deuticke, Leipzig and Vienna, 1937), p. 286; R. P. Feynman, Phys. Rev. **56**, 340 (1939). See also Ref. 11, p. 34.
[11] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Non-Relativistic Theory)*, 3rd ed., translated by J. B. Sykes and J. S. Bell (Pergamon, New York, 1977).
[12] See, e.g., Ref. 11, p. 165.
[13] This method is discussed, for example, in A. B. Migdal and V. Krainov, *Approximation Methods in Quantum Mechanics*, translated by A. J. Leggett (Benjamin, New York, 1969), pp. 129–31; A. B. Migdal, *Qualitative Methods in Quantum Theory*, translated by A. J. Leggett (Benjamin, New York, 1977); Ref. 1, Sec. 4.2.1.
[14] S. Flügge, *Practical Quantum Mechanics* (Springer, New York, 1974), pp. 175 ff.
[15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), Chap. 17.
[16] A. Martin, private communication.
[17] Jonathan L. Rosner, C. Quigg, and H. B. Thacker, Phys. Lett. B **74**, 350 (1978)
[18] For further discussion of the leptonic widths of vector mesons see C. Quigg, Ref. 2.

Construction of L^2 dependent potentials [a)]

M. A. Hooshyar
Physics Department, Shiraz University, Shiraz, Iran

(Received 17 August 1979; accepted for publication 11 January 1980)

Central and L^2-dependent potentials, acting among spinless particles, are constructed from the knowledge of the phase shifts for all angular momenta at two different fixed values of energy. It is shown that the information on phase shifts at two fixed energies does not lead to a unique set of central and L^2-dependent potentials and they will depend on an infinite number of parameters.

I. INTRODUCTION

In this work we are interested in constructing the central and L^2-dependent potentials from the knowledge of phase shifts for all angular momenta at two different fixed values of energy. Such a study has not only direct physical significance but also gives us a better understanding of what kind of information on the interaction we can obtain from scattering experiments.

An important tool in the construction of central potentials from phase shifts at a fixed energy is the Regge[1]–Newton[2] equation. Generalization of this equation for the case where an L^2-dependent potential is also present was found by the present author in a previous work.[3] In that work the problem of constructing the central potential from the phase shifts at a fixed energy was solved when the interaction also contained a known L^2-dependent potential. This work is then an extension of our previous results which enables one to construct not only the central potential but also the L^2-dependent potential from the knowledge of phase shifts for all angular momenta at two different values of energy.

The procedure is as follows: In Sec. 2 we first review the relevant parts of our previous results, but they are presented in a way that is more convenient for application to the present problem. Then using this modified version of that work, we arrive at two auxiliary central potentials which are related to the desired central and L^2-dependent potentials. The rest of that section is then devoted to finding the desired potentials from these auxiliary potentials. In Sec. 3 we have summarized the construction procedure.

The conclusion of this paper is as follows: Provided that $\delta_\lambda = o(\lambda^{-3})$ for large values of λ and a weak condition is satisfied by the auxiliary potentials, then a set of central and L^2-dependent potentials can be found which give the desired phase shifts. This set is not unique and it depends on an infinite number of parameters.

2. THE PROCEDURE

The radial Schrödinger equation for scattering of spinless particles by central and L^2-dependent potentials has the following form:

$$r^2\left[\frac{d^2}{dr^2}+k^2-V_1(r)-\lambda^2 V_2(r)\right]\psi_\lambda(r)=(\lambda^2-\tfrac{1}{4})\psi_\lambda(r), \tag{2.1}$$

where $V_2(r)$ is associated with the radial-dependent part of the L^2-dependent potential, $V_c(r)=(V_1+V_2/4)$ is the central potential, and $\lambda = l+\frac{1}{2}$.

In this work we assume that[4]:

$$0<1+r^2V_2(r)<\infty, \text{ for } r\geqslant 0,$$
$$\int_0^\infty dr\left|\frac{[1+r^2V_2(r)]^{1/2}-1}{r}\right|<\infty, \tag{2.2}$$
$$r^2V_2(r)\in C^2, \quad \text{for } r>0.$$

With the above conditions satisfied the following functions can be defiend:

$$\Theta(r)=\int_r^\infty ds\,\frac{[1+s^2V_2(s)]^{1/2}-1}{s},$$
$$R(r)=re^{-\Theta(r)}, \quad F(r)=\left[\frac{d}{dr}R(r)\right]^{-1/2},$$
$$f(R,R')=\sum_{\lambda\in\Omega} u_\lambda(kR)d_\lambda u_\lambda(kR'). \tag{2.3}$$

Where the constants d_λ are the potential coefficients, the set $\Omega=\{\frac{1}{2},1,\frac{3}{2}\cdots\}$, and $u_\lambda(kR)=(\frac{1}{2}\pi kR)^{1/2}J_\lambda(kR)$. Next, the Regge–Newton equation and an auxiliary wave function are defined in terms of the new variable R:

$$K(R,R')=f(R,R')-\int_0^R ds\,s^{-2}K(R,s)f(s,R'), \tag{2.4}$$

$$\phi_\lambda(R)=u_\lambda(kR)-\int_0^R ds\,s^{-2}K(R,s)u_\lambda(ks). \tag{2.5}$$

Then, the wave function ϕ_λ satisfies the following Schrödinger equation[5,6]:

$$R^2\left[\frac{d^2}{dR^2}+k^2-W(R)\right]\phi_\lambda(R)=(\lambda^2-\tfrac{1}{4})\phi_\lambda(R), \tag{2.6}$$

where

$$W(R)=-2R^{-1}\frac{d}{dR}[R^{-1}K(R,R)]. \tag{2.7}$$

The regular solution to Eq. (2.1) is connected to $\phi_\lambda(R)$ by the following relation:

$$\psi_\lambda(r)=F(r)\phi_\lambda(re^{-\Theta(r)}), \tag{2.8}$$

[a)]Supported in part by the Shiraz University Research Council, Shiraz, Iran.

 0022-2488/80/071695-03$1.00

if[7]:

$$W(R)=\dot{R}^{-2}[V_c(r)+k^2(\dot{R}^2-1)+N], \tag{2.9}$$

$$N=-\frac{3}{4}\left[\left(\frac{dR}{dr}\right)^{-1}\frac{d^2R}{dr^2}\right]^2+\frac{1}{2}\left(\frac{dR}{dr}\right)^{-1}\frac{d^3R}{dr^3}.$$

Since $(R/r)=F(r)=1$, for large values of r, it then follows from Eq. (2.8) that:

$$\lim_{r\to\infty}\psi_\lambda(r)=\lim_{R\to\infty}\phi_\lambda(R)$$
$$=A_\lambda\sin[kr-\tfrac{1}{2}\pi(\lambda-\tfrac{1}{2})+\delta_\lambda(k)]. \tag{2.10}$$

Following Newtons's method and taking the limit of Eq. (2.5) for large values of R, one arrives at a relation connecting the phase shifts with the potential coefficients:

$$A_\lambda e^{i\delta_\lambda(k)}=1-\sum_{\mu\in\Omega}A_\mu e^{i\delta_\mu(k)}d_\mu k e^{i\pi(\lambda-\mu)/2}L_\mu^\lambda(\infty), \tag{2.11}$$

$$L_\mu^\lambda(x)=\int_0^x ds\, s^{-2}u_\lambda(s)u_\mu(s).$$

Therefore the problem of constructing the central and L^2-dependent potentials reduces to the problem of finding the potential coefficients d_λ from Eq. (2.11). This problem has been studied in detail by Newton[2] and Sabatier.[8,9] They have shown that d_λ for physical values of angular momenta can be constructed from the phase shifts. But even for the case where we choose $d_\lambda=0$ for nonphysical values of λ, Eq. (2.11) still does not lead to a unique solution and d_λ will depend on one arbitrary parameter.

From the above consideration, it follows that if the phase shifts are given for all angular momenta, then a relation between the central and L^2-dependent potentials can be found [Eq. (2.9)]. Therefore, if a suitable central- or L^2-dependent potential is also given, then Eq. (2.9), together with Eq. (2.3), in general will give us the other unknown potential. But since in a scattering experiment usually only the phase shifts can be deduced, it is more desirable to construct a method for finding both the central and L^2- dependent potentials from the scattering information. Therefore, requiring the knowledge of all phase shifts for two different values of energy is the most feasible way for constructing both the central and L^2-dependent potentials from the scattering data. So, in this work we assume that the phase shifts $\delta_\lambda(k_1)$ and $\delta_\lambda(k_2)$ are given for all values of λ at two different values of energy, k_1^2 and k_2^2. Then we would like to construct a central potential and an L^2-dependent potential in such a way that the associated phase shifts at the two values of energy, k_1^2 and k_2^2, are the same as what we have been given by the scattering experiment. The first step in the construction procedure is clear. For each set of phase shifts, we find the associated auxiliary central potential $W(R)$ [Eq. (2.9)]. Having found $W_1(R)$ and $W_2(R)$, the auxiliary central potentials associated with $\delta_\lambda(k_1)$ and $\delta_\lambda(k_2)$ respectively, we then note that the central and L^2-dependent potentials will satisfy the following relations:

$$W_1(R)=\dot{R}^{-2}[V_c(r)+k_1^2(\dot{R}^2-1)+N], \tag{2.12}$$

$$W_2(R)=\dot{R}^{-2}[V_c(r)+k_2^2(\dot{R}^2-1)+N]. \tag{2.13}$$

Subtracting Eq. (2.13) from Eq. (2.12) we arrive at an ordinary differential equation for the variable $R(r)$:

$$\dot{R}(r)=G[R(r)], \tag{2.14}$$

where

$$G[R]=\{1-(k_1^2-k_2^2)^{-1}[W_1(R)-W_2(R)]\}^{-1/2}.$$

Looking into the definition of R, consistency of method requires the following relations to be valid:

$$\begin{aligned}&G[x]\text{ is positive and continuous for } x\geqslant 0,\\ &\lim_{x\to\infty}G[x]=1,\\ &\lim_{r\to\infty}(R/r)=1,\\ &R(0)=0.\end{aligned} \tag{2.15}$$

Using the Cauchy–Peano existence theorem,[10] we note that if $G[x]$ satisfies Eq. (2.15), then Eq. (2.14) has a solution for all finite values of r for which $R(0)=0$. Furthermore, if $G[x]$ satisfies a Lipschitz condition for all nonnegative real values of x, then the solution is unique. A solution to Eq. (2.14), satisfying the initial condition that $R(0)=0$, can easily be found by quadrature:

$$r=g(R)=\int_0^R dx/G[x]. \tag{2.16}$$

Since $G[x]$ satisfied Eq. (2.15), it follows that g is a one-to-one function,[11] and it has a unique inverse which is of class C^1 for $r\geqslant 0$. Hence R can be uniquely expressed as a function of r. Futhermore, it is easy to show that this solution will also have the property that $\lim_{r\to\infty}[R(r)/r]=1$. Therefore, R will have all the desired properties, and the L^2-dependent potential can be found by the following relation:

$$r^2V_2(r)=(r\dot{R}/R)^2-1. \tag{2.17}$$

Having found $V_2(r)$, then the central potential can be found from Eq. (2.12) or Eq. (2.13). By construction, this set of central and L^2-dependent potentials will be associated with the desired phase shifts at the two different values of energy.

The only point which needs to be considered is under what conditions on the phase shifts the function $G[R]$ will satisfy Eq. (2.15). It is shown by Sabatier[12] that if the phase shifts $\delta_\lambda=O(\lambda^{-3})$ as $\lambda\to\infty$, and also if the phase shifts do not belong to the exceptional set [which make the Fredholm determinants of Eq. (2.5) become zero], then the potential $W(R)$ is such that $ZW(Z)$ is a meromorphic function of Z with no poles on the nonnegative part of the real axis and $W(R)=O(R^{-3/2})$ as $R\to\infty$. From these properties it follows that the condition $G>0$ cannot be satisfied unless $\bar{W}(R)=[W_1(R)-W_2(R)]$ is finite at $R=0$. By making use of the nonunique nature of W_1 and W_2 we can always choose one of the arbitrary parameters in such a way that $\bar{W}(0)$ is finite. It is interesting to note that with this condition satisfied, $\bar{W}(Z)$ is a meromorphic function of Z and bounded for all values of $R\geqslant 0$. Therefore, in general, $G[R]$ will satisfy Eq. (2.15). Of course, if we do not choose the arbitrary parameters in $\bar{W}$ correctly, then the zeros of the denominator of G could be on the positive real axis and then Eq. (2.15) would

not be satisfied. But if the phase shifts are associated with an L^2-dependent potential which is in the class considered in this work, then, by choosing the arbitrary parameters appropriately, we can make $\bar{W}(R)$ satisfy the following inequality:

$$0 < 1 - [W_1(R) - W_2(R)]/(k_1^2 - k_2^2) < \infty, \quad \text{for } R \geqslant 0. \tag{2.18}$$

With the above condition satisfied, G will be bounded and will also satisfy a Lipschitz condition. Therefore, the solution to Eq. (2.14) will be unique and can be found from Eq. (2.16).

3. SUMMARY

The procedure for constructing the central and L^2-dependent potentials from the knowledge of phase shifts for all angular momenta at two different values of energy is as follows. For each set of phase shifts, we are to treat them as if we were solving a purely central potential. Using the method of Newton[2] and Sabatier,[8,9] the potential coefficients associated with each set of phase shifts are found [Eq. (2.11)]. Then, from Eq. (2.4) and Eq. (2.7) the associated auxiliary central potentials W_1 and W_2 are found. These potentials are not unique and depend on an infinite number of parameters. They are to be chosen in such a way that Eq. (2.18) is satisfied. Then Eq. (2.14) can be solved for $R(r)$. Having found $R(r)$, the L^2-dependent potential is found from Eq. (2.17). The central potential is then calculated from Eq. (2.12). By construction, the found set of central and L^2-dependent potentials is associated with the desired phase shifts at the two different values of energy.

ACKNOWLEDGMENT

The author would like to thank Professor R.G. Newton for making many helpful suggestions.

[1] T. Regge, Nuovo Cimento **14**, 951 (1959).
[2] R.G. Newton, J. Math. Phys. **3**, 75 (1962).
[3] M.A. Hooshyar, J. Math. Phys. **13**, 1931 (1972).
[4] The square root of a function is taken in the positive sense.
[5] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[6] K. Chadan and P.C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, New York, 1977).
[7] A dot on top of a function stands for the derivative of that function.
[8] P.C. Sabatier, J. Math. Phys. **7**, 1515 (1966).
[9] P.C. Sabatier, J. Math. Phys. **8**, 905 (1967).
[10] E.A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
[11] R.G. Buck, *Advanced Calculus* (McGraw-Hill, New York, 1956).
[12] P.C. Sabatier, J. Math. Phys. **7**, 2079 (1966).

Inverse scattering. II. Three dimensions

Roger G. Newton
Physics Department, Indiana University, Bloomington, Indiana 47405

(Received 27 December 1979; accepted for publication 28 March 1980)

Assuming that a scattering amplitude, given as a function of the energy and the directions of the incident and scattered particles, is associated with a local potential without spherical symmetry via the Schrödinger equation in three space dimensions, this potential is uniquely reconstructed by two methods. One is based on a generalization of the Marchenko equation; the other, on a generalization of the Gel'fand–Levitan equation.

1. INTRODUCTION

The three-dimensional inverse scattering problem for the Schrödinger equation with a local potential without spherical symmetry has been attacked several times over the last twenty-five years, but with only partial success. The two methods that led to beautiful results in one dimension and in the spherically symmetric case in three dimensions (which after separation of variables becomes one-dimensional, but on a half-line), that is, the Gel'fand–Levitan and the Marchenko methods, resisted generalization to higher dimensions. The "triangularity" of the integral kernels that play a crucial role in these procedures, was hard to generalize.

The first attempt at a solution of this problem was made by Kay and Moses.[1] They worked in a framework that admitted potentials which were local in the radial distance from the origin but nonlocal, i.e., not multiplicative operators, in the angles of rotation about the center. However, for a given scattering amplitude that is known to be associated with a local potential, this method was never shown to be able to reconstruct it. Later in the sixties, Faddeev introduced[2] a new Green's function for the three-dimensional Schrödinger equation specifically with the inverse scattering problem in mind, and both he[3] and independently I,[4] used his Green's function for this purpose. This method, however, is cumbersome and left some still unanswered questions concerning the possibility of exceptional points. Another attempt, which is restricted to weak potentials, is the generalization of an old method by Jost and Kohn to three dimensions by Prosser.[5]

The two methods presented in this paper are generalizations of those based on the Marchenko and the Gel'fand–Levitan equations in the radial (or one-dimensional half-line) case. An outline of the first was initially presented in Ref. 6. As a preliminary step, both were then applied to the (full-line) one-dimensional case in Ref. 7.

Perhaps the crucial step that was necessary to make the methods to be presented here possible was to divorce the dependence of the wave function on the magnitude of the momentum of a particle from that on its direction. In this paper the momentum direction will always be denoted by θ, a unit vector in $\mathbb{R}^3$, or a point on S^2; the wave number, or the square root of the energy, will be denoted by k. Whereas in one dimension the spectrum has the multiplicity two, in all higher dimensions it has infinite multiplicity. Correspondingly, the solutions of the Schrödinger equation were combined in Ref. 7 in two-component vectors, and the S matrix in a 2×2 matrix; in three dimensions the wave functions are vectors in $L^2(S^2)$ and the S matrix is an operator there. Undoubtedly the present methods can be generalized to any number of dimensions, though this will still require some nontrivial high-energy estimates.

The inverse-scattering problem has three separate aspects. One is the reconstruction of an underlying potential that is known to be associated with a given scattering amplitude; the second is the uniqueness of this solution; and the third is the characterization of the class of scattering amplitudes associated with local potentials (in a given class), or the question of the *existence* of an underlying local potential. In contrast to the one-dimensional case, the characterization or existence problem is well known to be difficult in three dimensions. As one easily learns by counting the number of parameters on which a scattering amplitude (five) and a potential (three) depend, the existence of an underlying local potential produces strong restrictions on the possible scattering amplitudes. The present paper has nothing to say about this aspect of the problem[8]; it deals with the reconstruction and uniqueness problems only.

If the characterization problem in higher dimensions is more difficult than in one dimension, the uniqueness problem is simpler. It is well known that at high energies the scattering amplitude approaches the Fourier transform of the potential. While at any fixed high energy, this leads to a knowledge of the Fourier transform at only one point in one dimension, it leads to the entire Fourier transform in three. Therefore, in three dimensions a knowledge of the whole scattering amplitude, as a function of the energy and the angles, uniquely determines the underlying potential if it exists. Though it is not new, this important statement is formulated as Lemma 3.1 in Sec. 3. As a result of this Lemma one might be tempted to regard the inverse problem as solved by the inverse Fourier transform of the scattering amplitude at large energy. Unfortunately, this solution is not practical and there is still a point in searching for other solutions. It does mean, however, that the lack of uniqueness that is associated with the bound states in one dimension does not arise in three dimensions.

 0022-2488/80/071698-18$1.00

The fundamental reason for this significant difference between one and higher dimensions is the existence of continuous transformations that connect the various elements of the S matrix. It is shown explicitly in Sec. 5 how the infinitesimal rotation generators lead to the determination of the analog of the constants that are associated with each of the bound states in one dimension, which we call their "characters" here.

In Sec. 2 we derive and discuss the needed properties of the wave function and we prove three lemmas that state conditions on which the wave function minus the free exponential, and the Fredholm determinant of the scattering integral equation minus one, are square integrable as functions of k.

Section 3 deals with the scattering amplitude A_k and the S matrix S_k. Lemma 3.2 gives conditions under which, first appearances to the contrary, $S_k - 1$, or kA_k, are square integrable, and in what sense. A simple by-product concerns the high-energy behavior of the average total cross section. Lemma 3.3 gives conditions under which the forward scattering amplitude minus its Born approximation is square integrable and has a meromorphic extension into the upper half-plane.

In Sec. 4 we derive and discuss a Marchenko-like equation, its solution, and its connection with the potential (Theorem 4.1). This is the equation first announced, in case there are no bound states, in Ref. 6. It has the remarkable property that, even if there are bound states, these need not be known or constructed. The analogous equation in one dimension does not have a unique solution, whereas in three dimensions Lemma 3.1 assures uniqueness.

We note that Eq. (3.22) shows that there is one parameter, the "volume" of the potential, that is needed in the construction and has to be obtained from the high-energy data. This formula, though not new, is derived in Appendix 5.

Section 5 deals with the bound states, their "characters," their multiplicities, and how to determine them from the scattering amplitude.

In Sec. 6 we pose a Hilbert problem whose solution is a generalization of the Jost function. If this Hilbert problem has a solution it is solved by a Marchenko-like equation. There is an interesting, and possibly significant, connection between this Hilbert problem for a *translated* potential, and the Hilbert problem discussed in Sec. 4, as was first pointed out for one dimension in Ref. 9.

The Schrödinger equation being an ordinary differential equation in one dimension, a "regular solution" that is an entire analytic function of k can be defined by boundary conditions at a point. There is no analog to this procedure in higher dimensions, where the Schrödinger equation is a partial differential equation. However, if a solution to the Hilbert problem of Sec. 6 exists, then the Jost function leads to a regular solution of the Schrödinger equation. This is shown in Sec. 7, and a Povsner–Levitan representation is derived for it. Thus we obtain the celebrated "triangular kernel," and a convenient alternative by an inverse Radon transform. We note that while the lemmas of Sec. 2 and 3 lead to all the other needed Fourier transforms by convergence in the mean, the auxiliary alternative here used for convenience may be a distribution. We also express the spectral function in terms of the Jost function.

In Sec. 8 we finally derive, in the standard manner from the completeness relation, a generalization of the Gel'fand–Levitan equation and we discuss the inversion procedure based on it. Its main result is stated as Theorem 8.1.

There are four Appendixes that contain the proofs of lemmas that would have been to cumbersome to prove in the main body of the paper.

As for notation, I want to emphasize that θ denotes a unit vector in $\mathbb{R}^3$. No special notation, such as bold face, will be used for points in $\mathbb{R}^3$. We also use the notation $\hat{x}$ for the unit vector $x/|x|$. In general, the notation is made as similar as possible to that in Ref. 7.

Finally, let us summarize the various conditions on the potential we shall use, though I have no doubt that it is possible to relax them.

The general conditions assumed will be the same as in Ref. 10, namely those stated in (2.1). Lemma 2.1 and later lemmas and theorems require, in addition, the existence of a monotonely decreasing function $M \in L^1(0,\infty) \cap L^2(0,\infty)$ such that for some x_0 and all x

$$|V(x + x_0)| \leqslant M(|x|).$$

Lemma 2.1 requires that for some $l > 0$

$$\int d^3y\, V^2(x + y)(1 + \theta \cdot \hat{y})^{-l} < C,$$

for all x and θ. Lemma 3.2 assumes that there exists another monotonely decreasing function M' such that

$$\int_0^\infty dt\,(1 + t)M'(t) < \infty$$

and

$$|\nabla V(x + x_0)| \leqslant M'(|x|)$$

for some x_0 and all x. All of these conditions are certainly satisfied if, for some C, $a > 0$, and $\epsilon > 0$,

$$|V(x)| < C(a + |x|)^{-3-\epsilon}$$

and

$$|\nabla V| < C(a + |x|)^{-2-\epsilon},$$

but a singularity at some point is not ruled out.

In order to be able to determine the multiplicities as well as the "characters" of the bound states from the derivatives of the forward scattering amplitude the existence of a sufficient number of absolute moments of the potential is needed in addition.

2. THE SCHRÖDINGER EQUATION

We start with the direct scattering problem. In order to be able to utilize the results obtained in Ref. 10 we shall generally assume that the potential satisfies the same conditions stated there: There exist two positive constants a and C such that for all $y \in \mathbb{R}$

$$\int d^3x\, |V(x)|^2 + \int d^3x\, |V(x)| \left(\frac{|x| + |y| + a}{|x - y|} \right)^2 < C. \tag{2.1}$$

This condition implies that $V\in L^1(\mathbb{R}^3)\cap L^2(\mathbb{R}^3)\cap R$, where R is the Rollnik class defined by $\|V\|_R<\infty$, and

$$\|V\|_R^2=\int d^3x\,d^3y\,\frac{|V(x)V(y)|}{|x-y|^2}$$

is the Rollnik norm. For specific results stated in the form of lemmas and theorems we shall require stronger conditions that will be explicitly stated. A summary of these conditions was given in the Introduction.

The solution $\psi_k(\theta,x)$ of the Schrödinger equation

$$-\Delta\psi_k(\theta,x)+V(x)\psi_k(\theta,x)=k^2\psi_k(\theta,x) \tag{2.2}$$

is defined by the integral equation

$$\psi_k(\theta,x)=\psi_k^0(\theta,x)+\int d^3y\,G_k^0(x,y)V(y)\psi_k(\theta,y), \tag{2.3}$$

where[11]

$$\psi_k^0(\theta,x)=e^{ik\theta\cdot x},$$
$$G_k^0(x,y)=-e^{ik|x-y|}/4\pi|x-y|.$$

We shall suppress the dependence on the unit vector θ that indicates the direction of the momentum, and regard $\psi_k(x)$ as a vector in $L^2(S^2)$. Multiplication of (2.3) by $|V|^{1/2}$ results in the equation

$$\xi_k(x)=\xi_k^0(x)+\int d^3y\,K_k(x,y)\xi_k(y), \tag{2.4}$$

where

$$\xi_k(x)=|V(x)|^{1/2}\psi_k(x),\quad \xi_k^0(x)=|V(x)|^{1/2}\psi_k^0(x), \tag{2.5}$$

$$K_k(x,y)=|V(x)|^{1/2}G_k^0(x,y)V^{1/2}(y), \tag{2.6}$$
$$V^{1/2}(x)=|V(x)|^{1/2}v(x),\quad v(x)=V(x)/|V(x)|.$$

If $V\in R$, $K_k(x,y)$ is an L^2-kernel and (2.4) is a Fredholm equation. For each fixed θ and each real $k\neq 0$ Eq. (2.4) is known[10] to have a unique solution in $L^2(\mathbb{R}^3\otimes S^2)$. Furthermore, for almost all x and all θ this solution is the continuous (except at $k=0$ if that is an exceptional point[10]) boundary value of an analytic function of k that is meromorphic in the upper half-plane. It then follows that the same is true of $\psi_k(\theta,x)$ for each θ and x. Furthermore, since by Schwarz's inequality

$$|\psi_k(\theta,x)-e^{ik\theta\cdot x}|^2\leqslant\int d^3y\,\frac{|V(y)|}{|x-y|^2}\int d^3z\,|\xi_k(\theta,z)|^2,$$

$\psi_k(\theta,x)$ is uniformly bounded for each real $k\neq 0$ because of (2.1).

We will need information about ψ_k for large k. As a preliminary we have

Lemma 2.1: If V satisfies (2.1) and there exists an x_0 and a monotone function[12] $M(t)\in L^1(0,\infty)\cap L^2(0,\infty)$ such that

$$M(t_1)\leqslant M(t_2)\quad \text{if } t_1\geqslant t_2,$$

and for all $x\in\mathbb{R}^3$

$$|V(x+x_0)|\leqslant M(|x|),$$

then for each x and $y\in\mathbb{R}^3$

$$I_k(x,y)=\int d^3z\,G_k^0(x,z)V(z)\,G_k^0(z,y)$$

is uniformly bounded as a function of k, $-\infty<k<\infty$,

$$\lim_{k\to\pm\infty}I_k(x,y)=0$$

uniformly in x and y, and $\exists C<\infty$ such that for all x and $y\in\mathbb{R}^3$

$$\int_{-\infty}^{\infty}dk\,|I_k(x,y)|^2\leqslant C.$$

We shall prove this lemma in Appendix 1.

If $K_k(x,y)$ is regarded as the kernel of an operator K_k on $L^2(\mathbb{R}^3)$ then the kernel of K_k^2 is

$$K_k^2(x,y)=|V(x)|^{1/2}I_k(x,y)V^{1/2}(y).$$

We then have the

Corollary: For almost all x and y, $K_k^2(x,y)$ is uniformly bounded for all k,

$$\lim_{k\to\pm\infty}K_k^2(x,y)=0,\quad \int_{-\infty}^{\infty}dk\,|K_k^2(x,y)|^2<\infty,$$

and

$$\int_{-\infty}^{\infty}dk\,\|K_k^2\|_2^2<\infty.$$

Here $\|\cdot\|_2$ denotes the Hilbert–Schmidt norm.

We also note that it is known[13] that if $V\in R$, then[14]

$$\lim_{k\to\pm\infty}\|K_k^\dagger K_k\|_2=0. \tag{2.7}$$

These results imply corresponding results for the complete Green's function defined by the resolvent equation

$$G_k=G_k^0+G_k^0VG_k \tag{2.8}$$

in operator form. Defining

$$\check{K}_k=|V|^{1/2}G_kV^{1/2}, \tag{2.9}$$

we have

$$\check{K}_k=(1-K_k)^{-1}K_k.$$

It is known that if V satisfies (2.1), then $(1-K_k)^{-1}$ is a bounded operator for each real $k\neq 0$. Equation (2.7) implies that, for each $k_0>0$, it is uniformly bounded for all real $|k|\geqslant|k_0|$, and

$$\lim_{k\to\pm\infty}\|\check{K}_k\|_2=0. \tag{2.10}$$

Let us write

$$\check{K}_k^2=|V|^{1/2}\check{I}_kV^{1/2},$$

so that

$$\check{I}_k=I_k+2G_k^0VI_k+2I_k+2I_kV^{1/2}(1-K_k)^{-1}|V|^{1/2}I_k$$
$$+I_kV^{1/2}(1-K_k)^{-2}|V|^{1/2}I_k.$$

It is easily seen from this, by Schwarz's inequality, the uniform boundedness of $(1-K_k)^{-1}$, and Lemma 2.1, that we have

Lemma 2.2: On the same hypotheses as in Lemma 2.1, for each x and $y\in\mathbb{R}^3$ the function $\check{I}_k(x,y)$ is bounded for all real $k\neq 0$,

$$\lim_{k\to\pm\infty}\check{I}_k(x,y)=0$$

uniformly in x and y, and for each $k_0 > 0$ $\exists C$ such that for all x and $y \in \mathbb{R}^3$

$$\int_{|k|>k_0} dk\ |\check{I}_k(x,y)|^2 < C.$$

It has the immediate

Corollary: For all x and y and $k > 0$, $\check{K}_k^2(x,y)$ is bounded,

$$\lim_{k \to \pm\infty} \check{K}_k^2(x,y) = 0,$$

for all $k_0 > 0$

$$\int_{|k|>k_0} dk\ |\check{K}_k(x,y)|^2 < \infty,$$

and

$$\int_{|k|>k_0} dk\ \|\check{K}_k\|_2^2 < \infty.$$

Let us now define the function

$$\gamma_k(\theta,x) = \psi_k(\theta,x)e^{-ik\theta\cdot x}. \tag{2.11}$$

Clearly γ_k shares the analytic properties of ψ_k. We need to examine its behavior for large $|k|$ in some detail. For that purpose we shall mean by $\hat{1}$ the vector in $L^2(S^2)$ that equals 1 for all θ: $\hat{1}(\theta) = 1$.

Defining

$$h_k(x) = |V(x)|^{1/2}\,[\gamma_k(x) - \hat{1}], \tag{2.12}$$

we have the integral equation

$$h_k(x) = \int d^3y\ \mathscr{K}_k(x,y)|V(y)|^{1/2} + \int d^3y\ \mathscr{K}_k(x,y)h_k(y), \tag{2.13}$$

where the kernel $\mathscr{K}_k(x,y)$ depends on θ,

$$\mathscr{K}_k(x,y) = -\frac{\exp\{ik\,[\,|x-y| - \theta\cdot(x-y)\,]\}}{4\pi|x-y|}\,|V(x)|^{1/2}\,V^{1/2}(y).$$

For real k, we have from (2.7)

$$\lim_{k \to \pm\infty} \|\mathscr{K}_k^\dagger \mathscr{K}_k\|_2 = 0 \tag{2.14}$$

if $\mathscr{K}_k(x,y)$ is considered as the kernel of an operator $\mathscr{K}_k$ on $L^2(\mathbb{R}^3)$. The same is easily seen to hold as $\mathrm{Re}k \to \pm\infty$ for any fixed $\mathrm{Im}k > 0$, as well as for $\mathrm{Im}k \to \infty$ by Lebesgue's dominated convergence theorem. Therefore, for each fixed θ and $\mathrm{Im}k \geqslant 0$

$$\lim_{|k|\to\infty} \|\mathscr{K}_k\| = 0, \tag{2.15}$$

where $\|\cdot\|$ denotes the operator norm on $L^2(\mathbb{R}^3)$. It then follows from (2.13) that similarly for each θ and $\mathrm{Im}k \geqslant 0$

$$\lim_{|k|\to\infty} \|h_k(\theta,\cdot)\| = 0, \tag{2.16}$$

where $\|h(\theta,\cdot)\|$ is the norm of $h(\theta,x)$ on $L^2(\mathbb{R}^3)$. We return to $\gamma_k(x)$ by means of the relation

$$\gamma_k(\theta,x) - 1 = -(1/4\pi)\alpha_k(\theta,x) - \int d^3y\ \frac{\exp\{ik\,[\,|x-y| - \theta\cdot(x-y)\,]\}V^{1/2}(y)h_k(\theta,y)}{4\pi|x-y|}, \tag{2.17}$$

where

$$\alpha_k(\theta,x) = \int d^3y\ \frac{V(y)}{|x-y|}\exp\{ik\,[\,|x-y| - \theta\cdot(x-y)\,]\}.$$

An easy result obtained form (2.16) is

Lemma 2.3: If V satisfies (2.1), then for $0 < \arg k < \pi$ and each $\theta \in S^2$ and $x \in \mathbb{R}^3$

$$\lim_{|k|\to\infty} \gamma_k(\theta,x) = 1.$$

Proof: Equation (2.16) together with (2.1) immediately shows that the second term on the right-hand side of (2.17) tends to zero. That α_k tends to zero follows from (2.1) and the dominated convergence theorem.

On the real axis we need a stronger statement that permits a Fourier transform to be defined. Appendix 2 contains a proof of

Lemma 2.4: Suppose that V satisfies (2.1) and $\exists l > 0$, $C_1 < \infty$, such that for all $\theta \in S^2$ and $x \in \mathbb{R}^3$

$$\int d^3y\ V^2(x+y)(1+\theta\cdot\hat{y})^{-l} < C_1,$$

(where $\hat{y} = y/|y|$) and, furthermore, $k = 0$ is not an exceptional point of (2.2). Then $\exists C_2 < \infty$ such that for each $\theta \in S^2$ and $x \in \mathbb{R}^3$

$$\int_{-\infty}^{\infty} dk\ |\gamma_k(\theta,x) - 1|^2 < C_2.$$

We remark that if $k = 0$ is an exceptional point,[10] then the same holds for the integral over the real axis with an arbitrarily small interval around $k = 0$ removed.

We next define the function

$$D(k) = \det{}_2(1 - \lambda K_k)|_{\lambda=1}, \tag{2.18}$$

where $\det_2$ is the modified Fredholm determinant[15] of (2.4). It is known[10] that if the potential satisfies (2.1), then $D(k)$ is the continuous boundary value of an analytic function regular in the upper half-plane $\mathrm{Im}k > 0$. If $-\kappa_n^2$, $n = 1,\ldots,n_{\max}$ [which is finite if V satisfies (2.1)] are the bound-state eigenvalues of (2.2), then $D(k)$ has zeros at $k = i\kappa_n$ whose multiplicities equal the multiplicities of the corresponding eigenvalues, and nowhere else in $\mathrm{Im}k \geqslant 0$, except possibly at $k = 0$. If $k = 0$ is an exceptional point, the nature of the zero of $D(k)$ at $k = 0$ is described in detail in Ref. 10. For $\mathrm{Im}k \geqslant 0$ we also have[10]

$$\lim_{|k|\to\infty} D(k) = 1. \tag{2.19}$$

Again we need more on the real axis, and we prove in Appendix 3

Lemma 2.5: If V satisfies the hypotheses of Lemma 2.1, then

$$\int_{-\infty}^{\infty} dk\ |D(k) - 1|^2 < \infty.$$

We may now define another solution of (2.2)

$$\chi_k(\theta, x) = \psi_k(\theta, x)D(k). \tag{2.20}$$

For each θ and x it is the continuous boundary value of an analytic function of k that is holomorphic in the upper half-plane. The same holds for

$$\beta_k(\theta, x) = \chi_k(\theta, x)e^{-ik\theta\cdot x} = \gamma_k(\theta, x)D(k). \tag{2.21}$$

What is more, we have

Lemma 2.6: If V satisfies (2.1), then for each $\theta\in S^2$ and $x\in\mathbb{R}^3$ and $0 < \arg k < \pi$

$$\lim_{|k|\to\infty} \beta_k(\theta, x) = 1.$$

If V satisfies the hypotheses of Lemma 2.4 then $\exists C < \infty$ such that for all $\theta\in S^2$ and $x\in\mathbb{R}^3$

$$\int_{-\infty}^{\infty} dk\ |\beta_k(\theta, x) - 1|^2 < C.$$

The proof follows immediately from Lemmas 2.3, 2.4, and 2.5, and the continuity of β_k.

Equation (2.3) shows that ψ_k has the symmetry property (for real k)

$$\psi_k^*(x) = \psi_{-k}(x). \tag{2.22}$$

Similarly,

$$D(k)^* = D(-k) \tag{2.23}$$

and

$$\gamma_k^*(x) = \gamma_{-k}(x), \tag{2.24}$$

$$\beta_k^*(x) = \beta_{-k}(x), \tag{2.24'}$$

as well as

$$\chi_k^*(x) = \chi_{-k}(x). \tag{2.22'}$$

3. THE SCATTERING AMPLITUDE AND THE S MATRIX

The scattering amplitude is defined as

$$A_k(\theta,\theta') = -(1/4\pi)\int d^3x\ \psi_k^0(\theta, x)^* V(x)\psi_k(\theta',x). \tag{3.1}$$

We shall regard it as the integral kernel of an operator on $L^2(S^2)$,

$$\begin{aligned} A_k &= -(1/4\pi)\int d^3x\ \psi_k^0(x)^* V(x)\psi_k(x) \\ &= -(1/4\pi)\int d^3x\ \xi_k^0(x)^* v(x)\,\xi_k(x) \\ &= -(1/4\pi)(B_k + C_k), \end{aligned} \tag{3.1'}$$

where

$$B_k = \int d^3x\ \xi_k^0(x)^* v(x)\,\xi_k^0(x), \tag{3.2}$$

$$\begin{aligned} C_k &= \int d^3x\, d^3y\ \xi_k^0(x)^* v(x) K_k(x, y)\,\xi_k(y) \\ &= \int d^3x\, d^3y\ \xi_k^0(x)^* v(x)\,\check{K}_k(x, y)\,\xi_k^0(y). \end{aligned} \tag{3.3}$$

One easily sees that A_k satisfies the *reciprocity* theorem

$$A_k(\theta,\theta') = A_k(-\theta', -\theta), \tag{3.4}$$

which we may write in the operator form

$$QA_kQ = \tilde{A}_k, \tag{3.4'}$$

where $\tilde{A}_k$ is the operator whose kernel is $A_k(\theta',\theta)$ if that of A_k is $A_k(\theta,\theta')$, and Q is the operator defined by

$$(QA)(\theta,\theta') = A(-\theta,\theta'). \tag{3.5}$$

It also follows from (2.22) that A_k satisfies

$$A_k^* = A_{-k}, \tag{3.6}$$

where A_k^* is the operator whose kernel is the complex conjugate of that of A_k.

Since $\psi_k(x)$ is continuous as a function of k, except possibly at $k = 0$, and it is uniformly (in x) bounded for each θ and $k \neq 0$, it follows from (3.1) and (2.1) that A_k is a continuous function of k for each θ and θ', except possibly at $k = 0$.

It follows directly from (3.3) and (2.10) that

$$\lim_{k\to\pm\infty} C_k(\theta,\theta') = 0 \tag{3.7}$$

uniformly in θ and θ'. Now we may parametrize the dependence of A_k on the two unit vectors θ and θ' (four variables) by using the vector

$$\tau = k(\theta - \theta')$$

and the unit vector

$$\hat{\pi} = (\theta + \theta')[2(1 + \theta\cdot\theta')]^{-1/2},$$

which is orthogonal to τ, writing

$$A_k(\theta,\theta') = \mathcal{A}_k(\tau,\hat{\pi}).$$

Then (3.7) implies the well-known fact that as $k\to\pm\infty$

$$\mathcal{A}_k(\tau,\hat{\pi}) = -(1/4\pi)\int d^3x\ e^{i\tau\cdot x}V(x) + o(1). \tag{3.8}$$

Note that $|\tau|^2 = 2k^2(1 - \theta\cdot\theta') \leqslant 4k^2$, so that at a fixed value of k the vector τ is restricted to lie within a ball of radius $2|k|$. As $|k|\to\infty$, however, this restriction becomes weaker and weaker and disappears. As a result, the leading term in (3.8) may take on all values of the Fourier transform of V. Since the Fourier transform uniquely determines a function almost everywhere, the implication is a uniqueness that we shall state in the form of

Lemma 3.1: Let $V^{(1)}$ and $V^{(2)}\in R$ and let $A_k^{(1)}(\theta,\theta')$ and $A_k^{(2)}(\theta,\theta')$ be the scattering amplitudes associated with them, respectively. If $A_k^{(1)}(\theta,\theta') = A_k^{(2)}(\theta,\theta')$ for all k, θ, and θ', then $V^{(1)}(x) = V^{(2)}(x)$ almost everywhere.

In other words, the scattering amplitude, known as a function of all k, θ, and θ', uniquely determines the potential. This result is not new, but it is worth recording because of its later implications.

We also obtain directly from (2.10) and (3.3) that

$$\lim_{k\to\pm\infty} k\,\|C_k\|_2 = 0. \tag{3.9}$$

Consequently, one easily finds that

$$\lim_{k\to\pm\infty} k^2\|A_k\|_2^2 = \tfrac{1}{2}V_R^2, \tag{3.10}$$

where $V_R^2 = \int d^3x\, d^3y\ V(x)V(y)|x - y|^{-2}$. We note that (3.10) is the physically interesting statement that for large energies $E = k^2$, the average total scattering cross section

$$\bar{\sigma}_E = (1/4\pi)\int d\theta\, d\theta'\ |A_k(\theta,\theta')|^2 = (1/4\pi)\|A_k\|_2^2 \tag{3.11}$$

goes as[16]

$$\bar{\sigma}_E = (1/8\pi)V_R^2 E^{-1} + o(E^{-1}). \tag{3.12}$$

Nevertheless, we have the following

Lemma 3.2: Let V satisfy the hypotheses of Lemma 2.1 as well as (2.1). Furthermore, assume that $\exists M'(t)$ that is monotonely decreasing,

$$\int_0^\infty dt\,(1+t)M'(t) < \infty,$$

and $\exists x_0$ such that for each x

$$|\nabla V(x+x_0)| \leqslant M'(|x|).$$

Then for each k_0 $\exists C < \infty$ such that for all $f \in L^2(S^2)$

$$\int_{-\infty}^{\infty} dk\, k^2 \|A_k f\|^2 \leqslant C\|f\|^2.$$

This lemma is proved in Appendix 4.

In view of (3.7) and (3.12), the result stated in this lemma is at first sight somewhat surprising. It is also likely to be the strongest general result of its kind that can be obtained.[17] On the other hand, its hypotheses can no doubt be weakened.

If f is choosen to be a characteristic function of a small angular region of solid angle Ω centered at θ then one might expect the result of Lemma 3.2 to read

$$\int_{-\infty}^{\infty} dk\, k^2 \sigma_k(\theta) \leqslant C/\Omega,$$

where

$$\sigma_k(\theta) = \int d\theta'\ |A_k(\theta',\theta)|^2$$

is the total cross section for scattering from the direction θ. This inequality, however, can be obtained only by assuming that for all energies the amplitude $A_k(\theta',\theta)$ is essentially constant over the region Ω, and in view of the existence of a sharp diffraction peak in the forward direction at high energies, such an assumption cannot be regarded as reliable.

For the forward scattering amplitude we have

Lemma 3.3: Suppose that the potential satisfies the hypotheses of Lemma (2.1). Then for each θ the forward scattering amplitude $A_k(\theta,\theta)$ has an analytic continuation into the upper half-plane, $\mathrm{Im}k > 0$, that is a meromorphic function of k there, with simple poles at $i\kappa_n$ if the bound state eigenvalues are $-\kappa_n^2$, and such that

$$\lim_{|k|\to\infty} A_k(\theta,\theta) = -(1/4\pi)\bar{V},$$

where

$$\bar{V} = \int d^3x\, V(x). \tag{3.13}$$

If $k = 0$ is not an exceptional point of the second kind, then $A_k(\theta,\theta)$ is continuous on the real axis and

$$\int_{-\infty}^{\infty} dk\ |A_k(\theta,\theta) + (1/4\pi)\bar{V}|^2 < \infty.$$

Proof: The first part of this lemma is well known. Let us write $A_k(\theta,\theta)$ in the form

$$A_k(\theta,\theta) = -(1/4\pi)\int d^3x\, V(x)\gamma_k(\theta,x) = -(1/4\pi)\left[\bar{V} + \int d^3x V^{1/2}(x)\, h_k(\theta,x)\right]. \tag{3.14}$$

If V satisfies (2.1), then, for each θ, $\mathscr{K}_k$ is an $L^2(\mathbb{R}^3)$-operator-valued analytic function of k that is holomorphic in the upper half-plane and continuous on the real axis; hence it follows from (2.13) that h_k is, for each θ, and L^2-vector-valued analytic function that is meromorphic in the upper half-plane and continuous on the real axis, except possibly at $k = 0$. Because of (2.1) the same analyticity consequently holds for $A_k(\theta,\theta)$. That $A_k(\theta,\theta)$ is continuous at $k = 0$ even if $k = 0$ is an exceptional point of the first kind, so long as it is not one of the second kind, was proved in Ref. 10.

Since $\gamma_k(\theta,x)$ is uniformly bounded for all $|k| \geqslant k_0 > 0$ it follows from (2.1) and (2.12) that $\|h_k(\theta,\cdot)\| < \infty$ for each θ. Therefore (2.1), (2.13), the corollary to Lemma 2.1, and the fact that $\mathscr{K}_k$ differs from K_k by a unitary transformation, imply that

$$\int_{-\infty}^{\infty} dk\ \|h_k(\theta,\cdot)\|^2 < \infty.$$

The square integrability stated in the lemma then follows from (3.14) and (2.1) by Schwarz's inequality. This proves the lemma.

The S matrix[18] is defined as the operator on $L^2(S^2)$

$$S_k = \mathbf{1} - (k/2\pi i)A_k\,. \tag{3.15}$$

In order to conform to the notation of Ref. 7 we shall allow operators on $L^2(S^2)$ to act to the left, in the sense that $\psi S = \psi'$ is the operator version of the integral transformation

$$\int d\theta'\ \psi(\theta')S(\theta',\theta) = \psi'(\theta),$$

or $\psi S = \tilde{S}\psi$.

If the "incoming wave" solution $\psi_k^-(x)$ of (2.2) is defined by the integral equation

$$\psi_k^-(x) = \psi_k^0(x) + \int d^3y\, G^0{}_{-k}(x,y)V(y)\psi_k^-(y), \tag{2.3'}$$

then S_k maps ψ_k^- onto ψ_k in the sense

$$\psi_k(x) = \psi_k^-(x)S_k. \tag{3.16}$$

It is well known that on the condition (2.1) S_k is unitary,[19]

$$S_k S_k^\dagger = S_k^\dagger S_k = \mathbf{1}. \tag{3.17}$$

Because of (3.4) and (3.6) we have

$$\tilde{S}_k = QS_kQ \tag{3.18}$$

and

$$S_{-k} = S_k^*\,. \tag{3.19}$$

Therefore (3.17) may be written

$$S_k Q S_{-k} Q = \mathbf{1}. \tag{3.17'}$$

Comparison of (2.3) with (2.3′) shows that

$$\psi_k^-(\theta, x) = \psi_{-k}(-\theta, x) \tag{3.20}$$

or

$$\psi_k^-(x) = \psi_{-k}(x)Q. \tag{3.20′}$$

Therefore (3.16) may be written

$$\psi_k = \psi_{-k}QS_k \tag{3.16′}$$

or

$$\psi_{-k} = \psi_k QS_k^*. \tag{3.16″}$$

Since for each real $k \neq 0$ the kernel $A_k(\theta,\theta')$ of the operator A_k exists for all θ and θ' and S^2 is compact, the operator A_k is in the Hilbert–Schmidt class and has a finite trace. Therefore the Fredholm determinant of the operator S_k is well defined and a continuous function of k

$$\det S_k = \det\left(1 - \frac{k}{2\pi i}A_k\right). \tag{3.21}$$

It is well known that $\det S_k$ is connected with the $D(k)$ by the equation

$$R_k \equiv D(-k)/D(k) = \det S_k \exp[i(k/2\pi)\bar{V}], \tag{3.22}$$

the exponential factor having the effect of reconciling (3.7) with (2.19). [$\bar{V}$ is defined by (3.13).] We derive (3.22) formally in Appendix 5. We remark that in view of (3.15) we have the

Corollary to Lemma 3.2: For all $f \in L^2(S^2)$

$$\int_{-\infty}^{\infty} dk\, \|(S_k - 1)f\|^2 \leqslant C\|f\|^2.$$

One may therefore regard $g_k \equiv (S_k - 1)f$ as a vector-valued function of k that is square integrable. Therefore it has a Fourier transform that converges in the mean to a vector-valued function h, and h is a linear functional of f. Consequently, it defines an operator W such that

$$\int_{-\infty}^{\infty} dk\, e^{ik\alpha}(S_k - 1)f = W(\alpha)f$$

and

$$\int_{-\infty}^{\infty} d\alpha\, \|W(\alpha)f\|^2 = 2\pi \int_{-\infty}^{\infty} dk\, \|(S_k - 1)f\|^2$$

is the corresponding Parseval relation.

4. A MARCHENKO EQUATION

The main result of this section is contained in

Theorem 4.1: Suppose that $V(x)$ satisfies the hypotheses of Lemma 3.2 and that $S_k(\theta,\theta')$ is the corresponding S matrix. Let $\mathfrak{S}_k$ be the operator on $L^2(S^2)$ whose kernel is

$$\mathfrak{S}_k(\theta,\theta') = e^{ik\theta\cdot x}S_k(\theta,\theta')e^{-ik\theta'\cdot x} \tag{4.1}$$

and let $G(\alpha)$ be the operator

$$G(\alpha) = (1/2\pi)\int_0^{\infty} dk\, Q(\mathfrak{S}_k^* R_k - 1)e^{ik\alpha}, \tag{4.2}$$

where R_k is defined by (3.22). Then $V(x)$ has the representation

$$V(x) = -2\theta\cdot\nabla\eta(\theta, x, 0), \tag{4.3}$$

where $\eta(\theta, x, \alpha)$ is the only solution of the integral equation

$$\eta(\alpha) = \hat{1}G(\alpha) + \int_0^{\infty} d\beta\, \eta(\beta)G(\alpha + \beta) \tag{4.4}$$

for which the right-hand side of (4.3) is independent of θ and satisfies the hypotheses of Lemma 3.2.

Proof: Lemma 2.6 implies that the Fourier transform of $\beta_k - \hat{1}$ exists in the mean

$$\eta(\alpha) = (1/2\pi)\int_{-\infty}^{\infty} dk\,(\beta_k - \hat{1})e^{-ik\alpha} \tag{4.5}$$

and, because of the analyticity of β_k, it vanishes for $\alpha < 0$

$$\beta_k = \hat{1} + \int_0^{\infty} d\alpha\, \eta(\alpha)e^{ik\alpha}. \tag{4.6}$$

Let us write (3.16″) in the form

$$\beta_{-k} = \beta_k Q\mathfrak{S}_k^* R_k. \tag{4.7}$$

Therefore, for $\alpha > 0$

$$\begin{aligned}\eta(\alpha) &= (1/2\pi)\int_{-\infty}^{\infty} dk\,(\beta_{-k} - \hat{1})e^{ik\alpha} \\ &= (1/2\pi)\int_{-\infty}^{\infty} dk\beta_k Q(\mathfrak{S}_k^* R_k - 1)e^{ik\alpha} \\ &\quad + (1/2\pi)\int_{-\infty}^{\infty} dk\,(\beta_k Q - \hat{1})e^{ik\alpha},\end{aligned}$$

and the last integral vanishes because of the analyticity of β_k and Lemma 2.6. As a result,

$$\eta(\alpha) = (1/2\pi)\int_{-\infty}^{\infty} dk\, \beta_k Q(\mathfrak{S}_k^* R_k - 1)e^{ik\alpha} \tag{4.8}$$

for $\alpha > 0$. Lemmas 2.5 and 3.2, and the boundedness of β_k and $D(k)$, imply that for each x the integrand of this Fourier integral is a square integrable vector-valued function of k; therefore the integral converges in the mean to the vector-valued function $\eta(\alpha)$.

By the convolution theorem, the Fourier transform of the product $(\beta_k - \hat{1})[Q(\mathfrak{S}_k^* R_k - 1)]$ equals the convolution of the Fourier transforms of the two terms. Therefore, by (4.6), we get (4.4) for $\alpha \geqslant 0$.

The Schrödinger equation (2.2) for β_k reads

$$(\Delta + 2ik\theta\cdot\nabla)\beta_k = V\beta_k. \tag{4.9}$$

If we insert the Fourier transform (4.6) in this equation we obtain, after an integration by parts, that $\eta(\alpha)$ must satisfy[20] the partial differential equation (we now write out the θ and x dependence)

$$\left(\Delta - 2\frac{\partial}{\partial\alpha}\theta\cdot\nabla\right)\eta(\theta, x, \alpha) = V(x)\eta(\theta, x, \alpha) \tag{4.10}$$

for $\alpha > 0$, and the boundary condition (4.3). The requirement that (4.3) has to be independent of θ, in spite of the appearance of the right-hand side, is the analog of the "miracle" of Ref. 7.

Since $\eta(\alpha)$ of (4.5) solves (4.4), we know that a solution of the integral equation exists. If it is not the only solution, then Lemma 3.1 ensures that no other solution can lead to a $V(x)$ in (4.3) that is independent of θ and satisfies the hypotheses of Lemma 3.2. This completes the proof.

Using (4.5) in (4.3), we obtain the following representation for the potential

$$V(x) = -(1/\pi)\theta\cdot\nabla \int_{-\infty}^{\infty} dk\ [D(k)\psi_k(\theta, x)e^{-ik\theta\cdot x} - 1]. \tag{4.11}$$

Equation (4.4) is a generalized Marchenko equation for $\eta(\alpha)$. Written out more explicitly for the function

$$\check{\eta}(\theta, x,\alpha) = \eta(\theta, x,\alpha - x\cdot\theta), \tag{4.12}$$

it reads for $\alpha \geqslant \theta\cdot x$,

$$\check{\eta}(\theta, x,\alpha) = \int d\theta' H(\theta',\theta,\alpha + x\cdot\theta') + \int d\theta' \times \int_{x\cdot\theta'}^{\infty} d\beta\, \check{\eta}(\theta', x, \beta)H(\theta',\theta,\alpha+\beta), \tag{4.4'}$$

where

$$H(\theta',\theta,\alpha) = (1/2\pi)\int_{-\infty}^{\infty} dk\, e^{ik\alpha}\left[(R_k - 1)\delta(-\theta',\theta) + \frac{k}{2\pi i}R_k A_k^*(-\theta',\theta)\right]. \tag{4.2'}$$

Because of Lemma 2.5 and 3.2, these Fourier integrals converge in the mean (in the sense of the remark after the corollary to Lemma 3.2). In terms of $\check{\eta}$ Eqs. (4.10) and (4.3) take on the form

$$\left(\Delta - \frac{\partial^2}{\partial\alpha^2}\right)\check{\eta}(\theta, x,\alpha) = V(x)\check{\eta}(\theta, x,\alpha), \tag{4.10'}$$

$$V(x) = -2\theta\cdot\nabla\check{\eta}(\theta, x, \theta\cdot x). \tag{4.3'}$$

We note that Eq. (4.4) leads to a solution of the following Hilbert problem: To find a function β_k that satisfies (4.7), is analytic in the upper half-plane, and tends to $\hat{1}$ there. While we cannot conclude from the uniqueness of $V(x)$ that this Hilbert problem has a unique solution, Lemma 6.2, below, will establish a connection between it and the existence of a solution to another Hilbert problem (6.2).

The reconstruction of the potential now proceeds by first constructing the function R_k from (3.22). This requires that $\bar{V}$ be extracted from the high-energy data. Then the function H of (4.2') or G of (4.2) is calculated and the Fredholm equation (4.4') or (4.4) is solved; $V(x)$ is obtained from (4.3') or (4.3).

This completes the reconstruction procedure via a Marchenko-like equation. We note that this procedure does not require a knowledge of the bound states at any stage of the construction.

5. THE BOUND STATES[21]

The L^2-eigenvalues of (2.2) are well known[10] to be confined to $k^2 \leqslant 0$ if V satisfies (2.1). The number $k^2 = -\kappa^2$, $\kappa \geqslant 0$, is an eigenvalue if and only if $D(i\kappa) = 0$. For $\kappa > 0$, the multiplicity of the zero of $D(k)$ at $i\kappa$ equals the degeneracy of the eigenvalue. For $\kappa = 0$ there are three possibilities, depending upon whether the origin is an exceptional point of the first kind or the second kind, or both. These are discussed in detail in Ref. 10.

If we define

$$\delta_k = \tfrac{1}{2}\arg S_k\,, \tag{5.1}$$

then, by (2.23) and (3.22),

$$\delta_k = -\arg D(k) - (k/4\pi)\bar{V}. \tag{5.2}$$

Equation (3.22) and the analyticity of $D(k)$, together with (2.19), then lead to the Levinson theorem[10]

$$\delta_{k=0} - \lim_{k\to\infty}[\delta_k + (k/4\pi)\bar{V}] = \pi(n + \tfrac{1}{2}q), \tag{5.3}$$

where n is the number of linearly independent L^2-eigenvectors of (2.2), $q = 1$ if $k = 0$ is an exceptional point of the second kind, and $q = 0$ otherwise.

In the spherically symmetric case and a fixed angular momentum, the Levinson theorem constitutes the only general connection between the bound states and the scattering data. In the present case, as in the (full-line) one-dimensional case, the bound-state eigenvalues can be determined from the forward scattering. What is more, in contrast to the one-dimensional case, their "character" (in the sense to be explained below) can also be determined from the S-matrix.

Suppose that $-\kappa^2 < 0$ is an eigenvalue of multiplicity N and that $u_\kappa^b(x)$, $b = 1,\ldots,N$, are a set of corresponding orthonormal eigenfunctions

$$\int d^3x\, u_\kappa^b(x)^* u_\kappa^{b'}(x) = \delta_{bb'}\,. \tag{5.4}$$

The asymptotic form of $u_\kappa^b(x)$ for large $|x|$ can be obtained by assuming an asymptotic expansion

$$u_\kappa^b(x)e^{\kappa|x|} = f(|x|)Y_\kappa^b(\hat{x}) + \cdots,$$

where $Y_\kappa^b(\hat{x}) \not\equiv 0$. The differential equation (2.2) leads to the result that $f = \text{const}\ |x|^{-1}$, and we may normalize $Y_\kappa^b(\hat{x})$ so that

$$u_\kappa^b(x) = -(e^{-\kappa|x|}/4\pi|x|)Y_\kappa^b(\hat{x}) + o(e^{-\kappa|x|}|x|^{-1}). \tag{5.5}$$

The homogeneous form of (2.3) shows that

$$Y_\kappa^b(\theta) = \int d^3x\, V(x)\, e^{\kappa\theta\cdot x}u_\kappa^b(x). \tag{5.6}$$

Because the asymptotic expansion implies that $Y_\kappa^b(\theta)$ cannot vanish identically as a function of θ, it follows that if two of the functions Y_κ^b with the same value of κ are linearly dependent, then so must be the corresponding functions u_κ^b. Since linear independence of two Y_κ^b's obviously implies linear independence of the corresponding u_κ^b's, the functions Y_κ^b, $b = 1,\ldots,N$, uniquely characterize the N bound-state eigenfunctions u_κ^b.

The resolvent $(k^2 + \Delta - V)^{-1}$, whose kernel is the Green's function $G_k(x, y)$, has a simple pole at $k = i\kappa$ and the principal part of the Green's function there is

$$G_k(x, y) = \frac{\Sigma_{b=1}^N u_\kappa^b(x)u_\kappa^b(y)^*}{2i\kappa(k - i\kappa)} + \cdots. \tag{5.7}$$

Consequently, the principal part of $\psi_k(x)$ there is

$$\psi_k(\theta,x) = [1/2i\kappa(k - i\kappa)]\sum_{b=1}^{N} Y_\kappa^b(-\theta)^* u_\kappa^b(x) + \cdots \tag{5.8}$$

and that of the forward scattering amplitude

$$A_k(\theta,\theta) = R_\kappa(\theta)/(k - i\kappa) + \cdots, \tag{5.9}$$

where

$$R_\kappa(\theta) = (i/8\pi\kappa) \sum_{b=1}^{N} Y_\kappa^b(\theta) Y_\kappa^b(-\theta)^*. \tag{5.10}$$

If $k = 0$ is not an exceptional point of the second kind, the positions of the poles of $A_k(\theta,\theta)$ in the upper half-plane, and hence the negative eigenvalues, may be determined by Fourier transformation

$$\int_{-\infty}^{\infty} dk\, e^{ikt} [A_k(\theta,\theta) + (1/4\pi)\bar{V}] = 2\pi i \sum_n e^{-\kappa_n t} R_{\kappa_n}(\theta) \tag{5.11}$$

for $t > 0$. Lemma 3.3 implies this result and the fact that the integral on the left converges in the mean for each θ. The value of $\bar{V}$, defined by (3.13) has to be considered as a parameter to be determined from $A_k(\theta,\theta)$. Alternatively, we have for $t > 0$

$$\int_{-\infty}^{\infty} dk\, e^{ikt} \partial A_k(\theta,\theta)/\partial k = -2\pi t \sum_n e^{-\kappa_n t} R_\kappa(\theta). \tag{5.12}$$

In order to obtain the functions Y_κ^b from the residues R_κ, one may use Faddeev's method and introduce the differential rotation operator

$$M_\theta^\eta = -(\eta\theta\nabla_\theta), \tag{5.13}$$

where η is a unit vector in the direction of the axis of rotation, ∇_θ is the gradient operator with respect to θ, and (abc) is the triple vector product. When M acts on any function $f(a\cdot\theta)$ of the scalar product $a\cdot\theta$ it becomes

$$M_\theta^\eta f(a\cdot\theta) = f'(a\cdot\theta)(a\theta\eta),$$

f' denoting the derivative of f. If V decreases sufficiently rapidly at $|x| \to \infty$, then

$$(M_\theta^\eta)^m A_k(\theta,\theta')|_{\theta'=\theta} \equiv A_k^{(m)}(\theta,\theta) \tag{5.14}$$

is an analytic function of k that is meromorphic in the upper half-plane, with simple poles at $k = i\kappa$, where its residues are

$$R_\kappa^{(m)}(\theta) = (i/8\pi\kappa) \sum_{b=1}^{N} Y_\kappa^{b(m)}(\theta) Y_\kappa^b(-\theta)^*, \tag{5.15}$$

with

$$Y_\kappa^{b(m)}(\theta) = (M_\theta^\eta)^m Y_\kappa^b(\theta). \tag{5.16}$$

Setting

$$R_\kappa^{(0)} = R_\kappa, \quad Y_\kappa^{b(0)} = Y_\kappa^b,$$

and considering $Y_\kappa^{b(m)}$ as a matrix

$$\mathscr{Y}_\kappa^{mb} \equiv Y_\kappa^{b(m)}, \quad b = 1,...,N, \ M = 0,...,N-1,$$

we have

$$(i/8\pi\kappa) Y_\kappa^b(-\theta)^* = \sum_{m=0}^{N-1} (\mathscr{Y}_\kappa^{-1})_{bm} R_\kappa^{(m)},$$

and hence we obtain a system of N ordinary differential equations of order $2N-1$

$$R_\kappa^{(m)} = \sum_{b=1}^{N} \sum_{n=0}^{N-1} Y_\kappa^{b(m)} (\mathscr{Y}_\kappa^{-1})_{bn} R_\kappa^{(n)}, \quad m = N,...,2N-1, \tag{5.17}$$

for the N functions Y_κ^b, $b = 1,...,N$, considered as functions of the azimuthal angle α of rotation about the axis η, in which the angle of inclination with respect to η appears as a parameter. If $\det(\mathscr{Y}_\kappa^{mb}) \equiv 0$ identically in θ, then the functions Y_κ^b form N solutions of an ordinary linear differential equation of order $N-1$, and hence they must be linearly dependent as functions of α. If this were true for all choices of η, the Y_κ^b would be linearly dependent as functions of θ on the unit sphere, which we know to be incorrect. Therefore, if $\det(\mathscr{Y}_\kappa^{mb}) \equiv 0$ for one choice of η, then a different choice of η will make $\det(\mathscr{Y}_\kappa^{mb}) \not\equiv 0$. The system (5.17) can thus be integrated, and the functions $Y_\kappa^b(\theta)$, $b = 1,...,N$, may be expressed in terms of $N(2N-1)$ arbitrary functions of the angle of inclination with respect to η [subject to the N constraints (5.15)]. These may be taken to be the values and the $2N-2$ first derivatives with respect to α of the N functions Y_κ^b on half of the great circle through η and orthogonal to $\eta' \perp \eta$. The procedure must now be repeated with operators $M_\theta^{\eta'}$, and the unknown functions are thereby determined.

In this manner it is possible, in principle, to determine the functions $Y_\kappa^b(\theta)$, $b = 1,...,N$, from the residues $R_\kappa(\theta)$. Since, *a priori*, the multiplicity N is unknown, it has to be determined at the same time, with (5.15) and the requirement that Y_κ^b be unchanged by any 2π-rotation serving as consistency checks.

Let us now anticipate from Secs. 6 and 7 that the solution $\phi_k(x)$ of (2.2) is an entire function of k and related to $\psi_k(x)$ by the Jost function J_k

$$\phi_k(x) = \psi_k(x) J_k, \tag{7.1}$$

where J_k is an analytic operator [on $L^2(S^2)$] valued function regular in the upper half-plane. Its Fredholm determinant is well defined and

$$\det J_{i\kappa} = 0$$

for all eigenvalues $-\kappa^2$, and nowhere else. The nullspace of $J_{i\kappa}$ is N-dimensional. Therefore there exist N linearly independent vectors X_κ^b, $b = 1,...,N$, such that

$$J_{i\kappa} X_\kappa^b = 0. \tag{5.18}$$

J_k being analytic, we have near $k = i\kappa$

$$J_k X_\kappa^b = (k - i\kappa) \dot{J}_{i\kappa} X_\kappa^b + \cdots,$$

where $\dot{J}_{i\kappa} = \partial J_k/\partial k$ at $k = i\kappa$.

Now it follows from (5.8) and (5.18) that $\phi_{i\kappa} X_\kappa^b$ is an eigenfunction and we may choose X_κ^b to be such that

$$\phi_{i\kappa} X_\kappa^b = u_\kappa^b. \tag{5.19}$$

Furthermore, we get from (7.1) and (5.8)

$$\lim_{k \to i\kappa} \psi_k(x) J_k X_\kappa^b = u_\kappa^b(x)$$
$$= (1/2i\kappa) \sum_{b'} (Y_\kappa^{b'*} Q \dot{J}_{i\kappa} X_\kappa^b) u_\kappa^{b'}(x).$$

The linear independence of the function $u_\kappa^b(x)$ implies that therefore

$$(1/2i\kappa) Y_\kappa^{b'*} Q \dot{J}_{i\kappa} X_\kappa^b = \delta_{bb'}. \tag{5.20}$$

Consequently, once the functions Y_κ^b have been obtained from the forward scattering amplitude, and J_k is known, then the functions X_κ^b are uniquely determined by the requirement that, together with the set $(1/2i\kappa) Y_\kappa^{b*} Q \dot{J}_{i\kappa}$, they form a bi-orthonormal system.

The functions X_κ^b are such that the projection onto the eigenspace at $-\kappa^2$ is given by

$$\sum_{b=1}^{N} u_\kappa^b(x)u_\kappa^b(y)^* = \int d\theta\, d\theta'\, \phi_{i\kappa}(\theta,x)M_\kappa(\theta,\theta')\phi_{i\kappa}(\theta',y)^* = \phi_{i\kappa}(x)M_\kappa\phi_{i\kappa}(y)^*, \tag{5.21}$$

where

$$M_\kappa(\theta,\theta') = \sum_{b=1}^{N} X_\kappa^b(\theta)X_\kappa^b(\theta')^*. \tag{5.22}$$

Let us also form the function

$$\Pi_k = \prod_{n=1}^{l}\left(\mathbb{1} + B_n\frac{2i\kappa_n}{k-i\kappa_n}\right), \tag{5.23}$$

where the B_n are orthogonal projections that are uniquely determined by the requirements

$$Y_{\kappa_n}^b \Pi_{-i\kappa_n} = 0,$$

which may also be written

$$Y_{\kappa_n}^b \prod_{m<n}\left(\mathbb{1} - B_m\frac{\kappa_m}{\kappa_m+\kappa_n}\right)(\mathbb{1}-B_n) = 0. \tag{5.24}$$

Since

$$B_n^2 = B_n = B_n^\dagger,$$

we have

$$\Pi_k = \Pi_k^\dagger = \Pi_k^{-1}. \tag{5.25}$$

We also note that

$$\det\Pi_k = \prod_{n=1}^{l}\left(\frac{k+i\kappa_n}{k-i\kappa_n}\right). \tag{5.26}$$

6. THE JOST FUNCTION

Let us pose the following *Hilbert problem*: Find an operator-valued function J_k that is the boundary value of an analytic function of k, holomorphic in the upper half-plane, such that for $\mathrm{Im}k > 0$

$$\lim_{|k|\to\infty} J_k = \mathbb{1} \tag{6.1}$$

in some appropriate sense, while on the real axis it is continuous and $J_k - \mathbb{1}$ is strongly square integrable as defined by the corollary to Lemma 3.2. Furthermore, we require that $(\det J_k)/D_k$ be continuous on the real axis, regular in the upper half-plane, and without zeros for $\mathrm{Im}k \geqslant 0$. Finally, J_k is to be such that for all real k

$$J_{-k} = QS_kJ_kQ, \tag{6.2}$$

where S_k is given. We shall call such a function a 1-solution of the Hilbert problem (6.2).

If we define

$$J_k^{\mathrm{red}} = \Pi_k J_k, \tag{6.3}$$

then J_k^{red} is the 1-solution of the reduced Hilbert problem

$$J_{-k}^{\mathrm{red}} = QS_k^{\mathrm{red}}J_k^{\mathrm{red}}Q, \tag{6.2}^{\mathrm{red}}$$

in which

$$S_k^{\mathrm{red}} = Q\Pi_{-k}QS_k\Pi_{-k} \tag{6.4}$$

and

$$D^{\mathrm{red}}(k) = D(k)\prod_{n=1}^{l}\left(\frac{k+i\kappa_n}{k-i\kappa_n}\right)^{N_n}. \tag{6.5}$$

Because of (5.25), S_k^{red} is unitary. [It does not, however, necessarily satisfy the reciprocity theorem (3.18).] Since

$$\det S_k^{\mathrm{red}} = \prod_{n=1}^{l}\left(\frac{k-i\kappa_n}{k+i\kappa_n}\right)^{N_n}\det S_k, \tag{6.6}$$

S_k^{red} satisfies the Levinson theorem appropriate to no negative eigenvalues. The function $D^{\mathrm{red}}(k)$ shares the analytic properties of $D(k)$ and has no zero for $\mathrm{Im}k > 0$; hence $\det J_k^{\mathrm{red}}$ has no zeros there. Once the reduced Hilbert problem $(6.2)^{\mathrm{red}}$ has been solved, the bound states are reintroduced via (6.3). We shall therefore assume for the remainder of this section that there are no bound states, but we shall drop the cumbersome superscript "red."

If J_k is a solution of (6.2), then (3.19) and the unitarity of S_k imply that it satisfies the relations

$$J_{-k} = J_k^*, \tag{6.7}$$

$$J_k^\dagger J_k Q = Q(J_k^\dagger J_k)^*. \tag{6.8}$$

Suppose now that a 1-solution of (6.2) exists. This solution must then be *unique*. If there were two solutions, $J_k^{(1)}$ and $J_k^{(2)}$, then (6.2) would imply that

$$J_{-k}^{(1)-1}J_{-k}^{(2)} = QJ_k^{(1)-1}J_k^{(2)}Q,$$

from which it would follow that $J_k^{(1)-1}J_k^{(2)}$ is an entire function of k. What is more, because of (6.1) it would tend to 1 as $|k|\to\infty$ everywhere, except possibly on the real axis. But the continuity of J_{-k}^{-1}, together with the strong square integrability of $J_k - \mathbb{1}$, would imply that $J_k^{(1)-1}J_k^{(2)} - \mathbb{1}$ is strongly square integrable. Hence we may conclude, by means of a generalization of Liouville's theorem that $J_k^{(1)} = J_k^{(2)}$. Consequently, *if* (6.2) *has a* 1*-solution, it cannot have a* 0*-solution* (i.e., one with same analyticity properties, but which approaches 0 and $|k|\to\infty$ for $\mathrm{Im}k > 0$ and which is strongly square integrable on the real axis). [21a]

To construct a 1-solution of (6.2) we proceed as in Sec. 4. If a 1-solution exists, its analyticity implies that the Fourier transform of $J_k - \mathbb{1}$ (which strongly converges in the mean) has support only on the positive half-axis:

$$J_k = \mathbb{1} + \int_0^\infty d\alpha\, L(\alpha)e^{ik\alpha}, \tag{6.9}$$

where for $\alpha > 0$

$$\begin{aligned} L(\alpha) &= (1/2\pi)\int_{-\infty}^{\infty} dk\,(J_k - \mathbb{1})e^{-ik\alpha} \\ &= (1/2\pi)\int_{-\infty}^{\infty} dk\,(J_{-k} - \mathbb{1})e^{ik\alpha} \\ &= (1/2\pi)\Big[\int_{-\infty}^{\infty} dk\,Q(S_k - \mathbb{1})J_kQe^{ik\alpha} \\ &\quad + \int_{-\infty}^{\infty} dk\,Q(J_k - \mathbb{1})Qe^{ik\alpha}\Big]. \end{aligned} \tag{6.10}$$

The last integral vanishes, and we have for $\alpha > 0$

$$L(\alpha) = (1/2\pi)Q\int_{-\infty}^{\infty} dk\,(S_k - \mathbb{1})J_ke^{ik\alpha}Q. \tag{6.11}$$

(All of these Fourier integrals converge in the strong sense in the mean.) Finally, we use the convolution theorem to obtain

$$L(\alpha) = G(\alpha)Q + \int_0^\infty d\beta\, G(\alpha+\beta)L(\beta)Q, \quad (6.12)$$

where

$$G(\alpha) = (1/2\pi)Q\int_{-\infty}^{\infty} dk\,(S_k - \mathbb{1})e^{ik\alpha}. \quad (6.13)$$

Thus the Fourier transform of $J_k - \mathbb{1}$ must satisfy the Marchenko-like equation (6.12). We note that for each $f \in L^2(S^2)$, the integral in (6.13), with its integrand acting of f, converges in the mean to the vector $G(\alpha)f$, which has finite norm.

Suppose a 1-solution of (6.2) exists and that the homogeneous version (6.12)° of (6.12) has an L^2-solution $L^\circ(\alpha)$. Then its Fourier transform (over $\alpha > 0$) leads, as in (6.9) to a function

$$J_k^\circ = \int_0^\infty d\alpha\, L^\circ(\alpha)e^{ik\alpha} \quad (6.14)$$

that is analytic in the upper half-plane, and tends to *zero* as $|k| \to \infty$. Furthermore, the function

$$F_k^\circ = J_{-k}^\circ - QS_kJ_k^\circ Q$$

is the boundary value of a function that is analytic in the upper half-plane and vanishes there as $|k| \to \infty$, because

$$\begin{aligned} L^\circ(\alpha) &= (1/2\pi)\int_{-\infty}^{\infty} dk\, J_{-k}^\circ e^{ik\alpha} \\ &= (1/2\pi)\int_{-\infty}^{\infty} dk\, QS_kJ_k^\circ Q_k e^{ik\alpha} \end{aligned}$$

vanishes for $\alpha < 0$. If P is an operator that anticommutes with Q, i.e., whose kernel is such that[22]

$$P(-\theta,\theta') = -P(\theta,-\theta'),$$

then one easily sees that

$$(F_{-k}^\circ P) = QS_k^*(F_k^\circ P)Q,$$

so that $F_k^\circ P$ is a 0-solution of the accompanying problem (6.2)* in which S_k is replaced by S_k^*. However, if J_k is a 1-solution of (6.2), then

$$\tilde{J}_{-k}^{-1} = QS_k^*\tilde{J}_k^{-1}Q,$$

and therefore, $\tilde{J}_k^{-1}$ is a 1-solution of (6.2)*. Consequently, if (6.2) has a 1-solution then (6.2)* also has one, and hence (6.2)* cannot have a 0-solution. If follows that $F_k = 0$, and thus J_k° is a 0-solution of (6.2). But that implies $J_k^\circ = 0$. We conclude that if (6.2) has a 1-solution (6.12)° cannot have a nontrivial L^2-solution.

Suppose next that (6.12) has an L^2-solution $L(\alpha)$. Then it defines a function J_k, by (6.9), that is analytic in the upper half-plane and which is such that $J_k - \mathbb{1}$ is strongly square integrable. Furthermore, since

$$\begin{aligned} 2\pi L(\alpha) &= \int_{-\infty}^{\infty} dk\,(J_{-k} - \mathbb{1})e^{ik\alpha} \\ &= \int_{-\infty}^{\infty} dk\,(QS_kJ_kQ - \mathbb{1})e^{ik\alpha}, \end{aligned}$$

the function

$$F_k = J_{-k} - QS_kJ_kQ$$

has all the properties of F_k° discussed above. Therefore, either $F_k = 0$, in which case J_k is the 1-solution of (6.2), or else F_kP is a 0-solution of (6.2)*. If the latter is the case then neither (6.2)* nor (6.2) can have a 1-solution. But if (6.2)* has a 0-solution, it must be associated, as in (6.14), with an L^2-solution of (6.12)*°, the homogeneous form of (6.12), in which S_k is replaced by S_k^*. We therefore conclude

Lemma 6.1: Suppose that S_k is unitary, that it satisfies the Levinson theorem

$$\log\det S_{k=0} = \lim_{k\to\infty} \log\det S_k,$$

and that $S_k - \mathbb{1}$ is strongly square integrable as a function of k. Then (6.2) has a 1-solution if and only if (6.12)*° has no nontrivial L^2-solution and (6.12) has a unique L^2-solution. The latter is associated with the 1-solution of (6.2) by (6.9).

We now want to establish a remarkable connection between the Hilbert problem (6.2) and that discussed in Sec.. 4. First of all, we note that the potential

$$V^y(x) = V(x+y),$$

whose center is shifted by y, leads to the solution

$$\psi_k^y(\theta,x) = e^{-ik\theta\cdot y}\psi_k(\theta,x+y), \quad (6.15)$$

and hence to the scattering amplitude

$$A_k^y(\theta,\theta') = e^{ik\theta\cdot y}A_k(\theta,\theta')e^{-ik\theta'\cdot y}. \quad (6.16)$$

Therefore the S-matrix associated with a potential that has been translated by x is equal to the function defined in (4.1)

$$S_k^x = \mathfrak{S}_k(x). \quad (6.17)$$

The associated equation for the Jost function is

$$J_{-k}^x = QS_k^xJ_k^xQ = Q\mathfrak{S}_kJ_k^xQ, \quad (6.18)$$

which implies that[22a]

$$(J_{-k}^x)^{-1} = Q(J_k^x)^{-1}\mathfrak{S}_k^\dagger Q. \quad (6.19)$$

Therefore if S_k is the full S matrix rather than a reduced one, so that it obeys the reciprocity law (3.18), then

$$\hat{1}(J_{-k}^x)^{-1} = \hat{1}(J_k^x)^{-1}Q\mathfrak{S}_k^*. \quad (6.20)$$

Comparison with (4.7) shows that the vector $\hat{1}(J_k^x)^{-1}D(k)$ satisfies the same equation as β_k and has the same properties. So the Hilbert problem of Sec. 4 is directly connected to the one for the Jost function associated with a potential that is *translated by* x. It is worth emphasizing that this is so despite the fact that the x-dependence of β_k has an entirely different origin; $\beta_k(x)$ is a solution of the differential equation (4.9).

We may use the connection between the two Hilbert problems as a handle on the uniqueness question in Sec. 4 as follows.

Suppose that (6.2) has a 1-solution. Then the equation

$$\Gamma_{-k} = \Gamma_kQ\mathfrak{S}_k^*R_k \quad (6.21)$$

has the 1-solution

$$\Gamma_k = (J_k^x)^{-1}D(k).$$

It follows that this solution is unique and (6.21) cannot have a 0-solution. On the other hand, assume that (4.7) has a

0-solution $\xi_k(\theta,x)$. We may consider $\xi_k(\theta',x)$ as the kernel $\xi_k(\theta,\theta',x)$ of an operator ξ_k that is independent of θ. Then ξ_k is a 0-solution of (6.21). Therefore we conclude

Lemma 6.2: If the Hilbert problem (6.2) has a 1-solution then the solution of the Hilbert problem (4.7) is unique.

7. THE REGULAR SOLUTION

In one dimension it is possible to define a "regular" solution that is an entire analytic function of k by means of a boundary condition at a finite point. This method does not work for the partial differential equation (2.2). However, suppose that a 1-solution J_k of (6.2) exists. Then it follows from (3.16a) that the vector-valued function

$$\phi_k(x) = \psi_k(x)J_k \tag{7.1}$$

is a solution of (2.2) with the property

$$\phi_{-k}(x) = \phi_k(x)Q. \tag{7.2}$$

Therefore, for all x and almost all θ, $\phi_k(\theta,x)$ is an entire analytic function of k of exponential order $\theta\cdot x$. Furthermore, because of Lemma 2.4 and the properties of J_k, $\phi_k(\theta,x) - \exp(ik\theta\cdot x)$is square integrable as a function of k for all x and almost all θ. It follows from the Paley–Wiener theorem[23] that $\phi_k(\theta,x)$ has a Povsner–Levitan representation

$$\phi_k(\theta,x) = e^{ik\theta\cdot x} - \int_{-\theta\cdot x}^{\theta\cdot x} d\alpha\, h(x,\theta,\alpha)e^{ik\alpha}. \tag{7.3}$$

Use of this representation in the Schrödinger equation and integration by parts leads to the conclusion that $h(x,\theta,\alpha)$ must be a solution of the partial differential equation[24]

$$\left(\frac{\partial^2}{\partial\alpha^2} + V(x) - \Delta\right)h(x,\theta,\alpha) = 0 \tag{7.4}$$

and satisfy the boundary conditions

$$-2\theta\cdot\nabla h(x,\theta,\theta\cdot x) = V(x), \tag{7.5}$$

$$\theta\cdot\nabla h(x,\theta,-\theta\cdot x) = 0. \tag{7.6}$$

We also note that (7.2) and (7.3) lead to the symmetry requirement

$$h(x,-\theta,-\alpha) = -h(x,\theta,\alpha) \tag{7.7}$$

Let us then write (7.3) in the form

$$\phi_k(\theta,x) = e^{ik\theta\cdot x} - \operatorname{sgn}(\theta\cdot x)\int_{-|\theta\cdot x|}^{|\theta\cdot x|} d\alpha\, h(x,\theta,\alpha)e^{ik\alpha}. \tag{7.3'}$$

On the other hand, $\phi_k(\theta,x)$ may also be regarded as a function of the vector $k\theta = p$ in $\mathbb{R}^3$ and we may take the three-dimensional Fourier transform of $\phi_k(\theta,x) - e^{ip\cdot x}$

$$\phi_k(\theta,x) = e^{ik\theta\cdot x} - \int d^3y\, e^{ik\theta\cdot y}\check{h}(x,y), \tag{7.8}$$

though $\check{h}(x,y)$ may be a distribution. Comparison of (7.8) with (7.3') shows that $\operatorname{sgn}(\theta\cdot x)h(x,\theta,\alpha)$ is the Radon transform of $\check{h}(x,y)$

$$\operatorname{sgn}(\theta\cdot x)h(x,\theta,\alpha) = \int d^3y\, \delta(\alpha - \theta\cdot y)\check{h}(x,y).$$

The support of $h(x,\theta,\alpha)$ as a function of α being confined to $[-\theta\cdot x,\theta\cdot x]$, that of $\check{h}(x,y)$ as a function of y must be contained in the ball $|y|\leqslant|x|$.

Equation (7.8) may be written in the form

$$\phi_k(\theta,x) = \int d^3y\, \omega(x,y)\psi_k^0(\theta,y), \tag{7.8'}$$

where

$$\omega(x,y) = \delta^3(x-y) - \check{h}(x,y), \tag{7.9}$$

or in operator notation,

$$\omega = 1 - \check{h}. \tag{7.9'}$$

If we write the inverse in the form

$$\omega^{-1} = 1 + \check{l},$$

then the kernel of $\check{l}$ satisfies the integral equation

$$\check{l}(x,y) = \check{h}(x,y) + \int_{|z|\leqslant|x|} d^3z\check{h}(x,z)\check{l}(z,y), \tag{7.10}$$

which has the Volterra property in the magnitude $|x|$.Therefore the support of $\check{l}(x,y)$, regarded as a function of y, is contained in the ball $|y|\leqslant|x|$. Consequently, for $|x|\geqslant|y|$,

$$(\omega^{\dagger-1} - \omega)(x,y) = \check{h}(x,y). \tag{7.11}$$

Note that (7.3) holds for complex values of k as well as for real, and so does (7.8). Therefore (7.8') is valid for the bound state functions $\phi_{i\kappa}(\theta,y)$ as well, even though the function $\psi_{i\kappa}^0(\theta,y) = e^{-\kappa\theta\cdot y}$ tends to infinity as $\theta\cdot y\to-\infty$. Since $\check{h}(x,y)$ has compact support as a function of y, there are no convergence problems in (7.8).

Consider now the well-known completeness relation for $\psi_k(x)$, say, in the form of the Parseval relation. For any $f\in L^2(\mathbb{R}^3)$

$$\begin{aligned}\int d^3x\,|f(x)|^2 &= (1/2\pi)^3\int_0^\infty dk\,k^2\int d\theta\left|\int d^3x\,\psi_k(\theta,x)^*f(x)\right|^2 \\ &\quad + \sum_{n,b}\left|\int d^3x\,u_{\kappa_n}^b(x)^*f(x)\right|^2.\end{aligned} \tag{7.12}$$

We may write the integral on the right-hand side

$$\int_0^\infty dk\,k^2\int d^3x\,d^3y\,f^*(x)f(y)\psi_k(x)\psi_k^\dagger(y),$$

in which the inner product in $L^2(S^2)$ is understood

$$\begin{aligned}\psi_k(x)\psi_k^\dagger(y) &= \int d\theta\,\psi_k(\theta,x)\psi_k(\theta,y)^* \\ &= \phi_k(x)(J_k^\dagger J_k)^{-1}\phi_k^\dagger(y).\end{aligned}$$

Furthermore, by (5.21), the sum on the right-hand side of (7.12) may be written

$$\begin{aligned}&\sum_n\sum_b\int d^3x\,d^3y\,f^*(x)f(y)u_{\kappa_n}^b(x)u_{\kappa_n}^b(y)^* \\ &\quad = \sum_n\int d^3x\,d^3y\,f^*(x)f(y)\phi_{i\kappa_n}(x)M_{\kappa_n}\phi_{i\kappa_n}(y)^*,\end{aligned}$$

where M_κ is defined by (5.22).

Thus the Parseval relation may be expressed in the form

$$\begin{aligned}&\int d^3x\,|f(x)|^2 \\ &\quad = \int d^3x\,d^3y\,f^*(x)f(y)\int\phi_k(x)d\rho_E\,\phi_k^\dagger(y),\end{aligned}$$

where $E = k^2$ and the spectral function ρ_E is given by

$$d\rho_E/dE = \begin{cases} (1/2\pi)^3 \tfrac{1}{2} k\,(J_k^\dagger J_k)^{-1}, & E>0, \\ \sum_{n=1}^{I} M_{\kappa_n}\delta(E+\kappa_n^2), & E<0. \end{cases} \tag{7.13}$$

The completeness relation for ϕ_k may also be written in the shorthand symbolic form

$$\int \phi_k\, d\rho_E\, \phi_k^\dagger = \delta, \tag{7.14}$$

and the corresponding one for ψ_k^0 reads

$$\int \psi_k^0\, d\rho_E^0\, \psi_k^{0\dagger} = \delta, \tag{7.15}$$

where

$$d\rho_E^0/dE = \begin{cases} (1/2\pi)^3 \tfrac{1}{2} k & \text{for } E>0, \\ 0 & \text{for } E<0. \end{cases} \tag{7.16}$$

In this language (7.8) has the form

$$\phi_k = \omega \psi_k^0, \tag{7.8''}$$

where ω is an operator whose kernel is $\omega(x,y)$ and which is independent of θ.

8. A GEL'FAND–LEVITAN EQUATION

We present the main result of this section in the form of

Theorem 8.1: Suppose that $V(x)$ satisfies the hypotheses of Lemma 3.2 and that ρ_E is the corresponding spectral function, represented as in (7.13), where $M_{\kappa_n}(\theta',\theta)$ is defined by (5.22), and we define $M_k(\theta',\theta)$ to be the kernel of the operator

$$M_k = (J_k^\dagger J_k)^{-1} - \mathbb{1} = QM_{-k}Q \tag{8.1}$$

on $L^2(S^2)$. Let $\mathscr{M}(\theta',\theta,\alpha)$ be defined by

$$\begin{aligned} &\mathscr{M}(\theta',\theta,\alpha) \\ &\quad = (1/2\pi)\int_{-\infty}^{\infty} dk\, M_k(\theta',-\theta)e^{ik\alpha} - 4\pi^2 \\ &\qquad \times \sum_n \kappa_n^{-2}\left[M_{\kappa_n}(\theta',\theta)e^{-\kappa_n\alpha} + M_{\kappa_n}(-\theta',-\theta)e^{\kappa_n\alpha}\right]. \end{aligned} \tag{8.2}$$

Then $V(x)$ has the representation (7.5), where $h(x,\theta,\alpha)$ is a solution of the integral equation

$$\begin{aligned} &\operatorname{sgn}(\theta\cdot x)h(x,\theta,\alpha) \\ &\quad = \int d\theta'\,\mathscr{M}(\theta',\theta,\alpha+\theta'\cdot x) - \int d\theta' \\ &\quad - \int d\theta' \int_{-|\theta'\cdot x|}^{|\theta'\cdot x|} d\beta\, \operatorname{sgn}(\theta'\cdot x)h(x,\theta',\beta) \\ &\qquad \times \mathscr{M}(\theta',\theta,\alpha+\beta). \end{aligned} \tag{8.3}$$

The corresponding solution of the Schrödinger equation is given by (7.3).

Proof: Using (7.8b) and the fact that ω is independent of θ, we may write (7.14) in the form

$$\omega^{\dagger -1} = \omega \int \psi_k^0\, d\rho_E\, \psi_k^{0\dagger},$$

and (7.15) as

$$\omega = \omega \int \psi_k^0 d\rho_E^0\, \psi_k^{0\dagger}.$$

Substracting these two equations one finds

$$\omega^{\dagger -1} - \omega = \omega \int \psi_k^0 d(\rho_E - \rho_E^0)\, \psi_k^{0\dagger}.$$

If this equation is written out for the corresponding kernels and (7.11) is used, one obtains for $|x|\geqslant|y|$

$$\check{h}(x,y) = \check{h}_0(x,y) - \int_{|z|<|x|} d^3z\, \check{h}(x,z)\check{h}_0(z,y), \tag{8.4}$$

where

$$\check{h}_0(z,y) = \int \psi_k^0(z) d(\rho_E - \rho_E^0)\, \psi_k^0(y)^\dagger, \tag{8.5}$$

or, more explicitly,

$$\begin{aligned} \check{h}_0(z,y) &= (2\pi)^{-3}\int_0^\infty dk\, k^2 \int d\theta\, d\theta'\, e^{ik(\theta'\cdot z - \theta\cdot y)}M_k(\theta',\theta) \\ &\quad + \sum_{n=1}^{I}\int d\theta\, d\theta'\, e^{-\kappa_n(\theta'\cdot z + \theta\cdot y)}M_{\kappa_n}(\theta',\theta) \\ &= -(1/8\pi^2)\int d\theta \frac{\partial^2}{\partial\alpha^2}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty} dk \right. \\ &\quad \times \int d\theta'\, e^{ik(\theta'\cdot z-\alpha)}M_k(\theta',\theta) \\ &\quad - 4\pi^2 \int d\theta' \sum_n \kappa_n^{-2}\left[M_{\kappa_n}(\theta',\theta)e^{-\kappa_n(\theta'\cdot z+\alpha)}\right. \\ &\quad \left.\left. + M_{\kappa_n}(\theta',-\theta)e^{-\kappa_n(\theta'\cdot z-\alpha)}\right]\right\}_{\alpha=\theta\cdot y}. \end{aligned}$$

The inverse of the Radon transform[25] being

$$\check{f}(y) = -(1/8\pi^2)\int d\theta \frac{\partial^2}{\partial\alpha^2} f(\theta,\alpha)\Big|_{\alpha=\theta\cdot y},$$

we recognize the bracket as the Radon transform

$$\begin{aligned} h_0(z,\theta,\alpha) &= (2\pi)^{-1}\int_{-\infty}^{\infty} dk \int d\theta'\, M_k(\theta',-\theta)e^{ik(\theta'\cdot z+\alpha)} \\ &\quad - (2\pi)^2 \int d\theta' \sum_n \kappa_n^{-2}\left[M_{\kappa_n}(\theta',\theta)e^{-\kappa_n\alpha}\right. \\ &\quad \left. + M_{\kappa_n}(\theta',-\theta)e^{\kappa_n\alpha}\right]e^{-\kappa_n\theta'\cdot z}, \end{aligned} \tag{8.6}$$

and (8.4) leads to

$$\begin{aligned} &\operatorname{sgn}(x\cdot\theta)h(x,\theta,\alpha) \\ &\quad = h_0(x,\theta,\alpha) - \int_{|z|<|x|} d^3z\, \check{h}(x,z)h_0(z,\theta,\alpha). \end{aligned}$$

But since

$$\int_{|z|<|x|} d^3z\, \check{h}(x,z)e^{ik\theta'\cdot z} = \int_{-\theta'\cdot x}^{\theta'\cdot x} d\beta\, h(x,\theta',\beta)e^{ik\beta},$$

this becomes a three-dimensional analog of the Gel'fand–Levitan equation, (8.3). This completes the proof.

Using (8.2) in (8.3), we obtain

$$\begin{aligned} &\operatorname{sgn}(\theta\cdot x)h(x,\theta,\alpha) \\ &\quad = (1/2\pi)\int d\theta' \int_{-\infty}^{\infty} dk\, \phi_k(\theta',x)M_k(\theta',-\theta)e^{ik\alpha} \\ &\qquad -4\pi^2 \sum_{n,b} \kappa_n^{-2} u_{\kappa_n}^b(x) \\ &\qquad \times \left[e^{-\kappa_n\alpha}X_{\kappa_n}^b(\theta)^* + e^{\kappa_n\alpha}X_{\kappa_n}^b(-\theta)^*\right]. \end{aligned}$$

By (7.5), therefore, the potential has the representation

$$V(x) = -2\theta\cdot\nabla\left\{\operatorname{sgn}(\theta\cdot x)\left[(1/2\pi)\int d\theta'\right.\right.$$
$$\times\int_{-\infty}^{\infty} dk\,\phi_k(\theta',x)M_k(\theta',-\theta)e^{ik\theta\cdot x}$$
$$-4\pi^2\sum_{n,b}\kappa_n^{-2}u^b_{\kappa_n}(x)$$
$$\left.\left.\times\left[e^{-\kappa_n\theta\cdot x}X^b_{\kappa_n}(\theta)^* + e^{\kappa_n\theta\cdot x}X^b_{\kappa_n}(-\theta)^*\right]\right]\right\}. \quad (8.7)$$

Again one encounters the "miracle" that, the appearance of the right-hand side not withstanding, this must be independent of θ.

One proves the uniqueness of a solution of Eq. (8.4), and hence of (8.3), in the standard way. Suppose

$$\bar{\xi}(y) = \begin{cases}\xi(y) & |y|\leqslant|x|,\\ 0 & |y|>|x|,\end{cases}$$

is an L^2-solution of the homogeneous form of (8.4)

$$\xi(y) = -\int d^3z\,\xi(z)\check{h}_0(z,y).$$

Then

$$\int d^3y\,|\xi(y)|^2 = -\int d^3y\,d^3z\,\xi(z)\check{h}_0(z,y)\xi^*(y).$$

Inserting the form (8.5) of $\check{h}_0$ and using the completeness (7.14), one obtains

$$\int \mathring{\xi}_k\,d\rho_E\,\mathring{\xi}^\dagger_k = 0,$$

where

$$\mathring{\xi}_k(\theta) = \int_{|y|\leqslant|x|} d^3y\,\xi(y)e^{ik\theta\cdot y}$$

and

$$\mathring{\xi}_{i\kappa}(\theta) = \int_{|y|\leqslant|x|} d^3y\,\xi(y)e^{-\kappa\theta\cdot y}.$$

$d\rho_E/dE$ being positive definite, it follows that $\mathring{\xi}_k(\theta) = 0$, and hence $\xi(y) = 0$.[26]

The reconstruction of the underlying potential via the Gel'fand–Levitan equation now proceeds as follows. As a first step, it is necessary to find the bound-state energies, degeneracies, and characters (i.e., the functions Y^b_κ) from the forward scattering amplitude by the method of Sec. 5. Next the projections B_n are constructed via (5.24), the functions Π_k by (5.23), and the reduced S-matrix by (6.4). After calculating $G(\alpha)$ by (6.13) one must now solve the Hilbert problem (6.2)$^{\text{red}}$ by means of (6.12). If (6.12)$^{*\,\circ}$ has only the trivial solution and (6.12) has a unique L^2-solution then it yields the 1-solution of (6.2)$^{\text{red}}$ by (6.9). The full Jost function J_k is then obtained from (6.3), and the functions X^b_κ are calculated from (5.20). One then has all the ingredients needed for the calculation of the kernel $\mathcal{M}(\theta',\theta,\alpha)$ by (8.2), with (8.1) and (5.22). The final step is to solve the Gel'fand–Levitan equation (8.3). The potential is then obtained from (7.5), and the wavefunction from (7.3). If a Jost function exists then it is obtained in this way, (8.3) is known to have a solution and the potential is uniquely reconstructed.

While this method is considerably more involved than that of Sec. 4, it does have the advantage that the region of integration in (8.3) is *finite*. The structure of (6.12) is similar to that of (4.4), but in the Marchenko method (4.4) has to be solved for each x, whereas (6.12) needs to be solved only once. In constrast to the method of Sec. 4, the Gel'fand–Levitan method requires a knowledge of the bound states, of their eigenvalues, their degeneracies, and their "characters." However, whereas in one dimension the "characters" are independent data, in three dimensions they are implied by the scattering amplitude.

APPENDIX 1

Proof of Lemma 2.1:

$$(4\pi)^2 I_k(x,y) = \int d^3z\,V(z)\frac{e^{ik(|x-z|+|y-z|)}}{|x-z|\,|y-z|}.$$

Since by Schwarz's inequality

$$\left(\int d^3z\frac{|V(z)|}{|x-z|\,|y-z|}\right)^2$$
$$\leqslant\int d^3z\frac{|V(z)|}{|x-z|^2}\int d^3t\frac{|V(t)|}{|y-t|^2},$$

the integrand of I_k is in $L^1(\mathbb{R}^3)$ if V satisfies (2.1). It follows that $I_k(x,y)$ is uniformly bounded for all k, x, and y, and, by the same argument as in Simon,[13] p. 23, that by the Riemann–Lebesgue lemma

$$\lim_{k\to\pm\infty} I_k(x,y) = 0$$

uniformly in x and y. The square integrability, however, requires stronger assumptions.

Shifting variables and setting $\eta = y - x$, we obtain

$$(4\pi)^2 I_k(x,y) = \int d^3z\,V(z+x)\frac{e^{ikt}}{|z-\eta|\,|z|},$$

where $t = r + |r\hat{z} - \eta|$, $r = |z|$, $\hat{z} = z/|z|$. Thus

$$(4\pi)^2 I_k(x,y) = \tfrac{1}{2}\int_{|\eta|}^{\infty} dt\left|\int d\hat{z}\frac{t^2-\eta^2}{(t-\eta\cdot\hat{z})^2}V(x+\hat{z}r)\right|^2,$$

where $r = (t^2-\eta^2)/2(t-\eta\cdot\hat{z})$. Now

$$r \geqslant \frac{t^2-\eta^2}{2(t+|\eta|)} = \tfrac{1}{2}(t-|\eta|),$$
$$|x+\hat{z}r| \geqslant |r-|x||$$
$$\geqslant \tfrac{1}{2}|t-|\eta|-2|x|\,| \geqslant \tfrac{1}{2}|t-3|x|-3|y|\,|,$$

and

$$\int d\hat{z}\frac{t^2-\eta^2}{(t-\eta\cdot z)^2} = 4\pi;$$

consequently

$$\int dk\,|I_k|^2 \leqslant \tfrac{1}{2}\pi\int_{|\eta|}^{\infty} dt\,M^2(t-3|x|-3|y|) \leqslant C < \infty$$

if V satisfies the hypotheses.

APPENDIX 2

Proof of Lemma 2.4: We use spherical polar coordinates for the integration and set $r = |y|$, $u = \theta\cdot\hat{y}$, $V(x+y) = U(\hat{y},r) = W(\hat{y},t)$, $t = r(1+u)$, and $\epsilon > 0$. Then

$$\alpha_k^\epsilon(\theta,x) = \int_0^\infty dr\, r \int_{u>\epsilon-1} d\hat{y}\, Ue^{ikt}$$
$$= \int_0^\infty dt\, te^{ikt} \int_{u>\epsilon-1} d\hat{y}(1+u)^{-2}W,$$

so that, twice using Schwarz's inequality,

$$(2\pi)^{-2}\int_{-\infty}^{\infty} dk\, |\alpha_k^\epsilon(\theta,x)|^2$$
$$= \int_0^\infty dt\, t^2\Big[\int_{u>\epsilon-1} d\hat{y}(1+u)^{-2}W\Big]^2$$
$$\leqslant \int_{u>\epsilon-1} d\hat{y}(1+u)^{-2}\int_{u'>\epsilon-1} d\hat{y}'(1+u')^{-2}$$
$$\times \int_0^\infty dt\, t^2 |W(\hat{y},t)W(\hat{y}',t)|$$
$$\leqslant \Big[\int_{u>\epsilon-1} d\hat{y}(1+u)^{-2}\Big(\int_0^\infty dt\, t^2W^2\Big)^{1/2}\Big]^2$$
$$= \Big[\int_{u>\epsilon-1} d\hat{y}(1+u)^{-1/2}\Big(\int_0^\infty dr\, r^2U^2\Big)^{1/2}\Big]^2$$
$$\leqslant \int_{u'>\epsilon-1} d\hat{y}'(1+u')^{l-1}$$
$$\times \int_{u>\epsilon-1} d\hat{y}(1+u)^{-l}\int_0^\infty dr\, r^2U^2$$

for all l. If we choose $l>0$ such that the hypotheses are satisfied, then we may allow ϵ to approach zero and we have

$$\int_{-\infty}^{\infty} dk\, |\alpha_k(\theta,x)|^2 \leqslant C < \infty,$$

where C is independent of x and θ.

Now from (2.17)

$$|\gamma_k(\theta,x) - 1|$$
$$\leqslant (1/4\pi)|\alpha_k(\theta,x)| + \int d^3y\, \frac{|V(y)|^{1/2}}{4\pi|x-y|}\,|h_k(\theta,y)|,$$

and by Schwarz's inequality,

$$\int dk\Big(\int d^3y\, \frac{|V|^{1/2}}{|x-y|}\,|h_k|\Big)^2$$
$$= \int d^3y\, d^3z\, \frac{|V(y)V(z)|^{1/2}}{|x-y|\,|x-z|}$$
$$\times \int dk\, |h_k(\theta,y)h_k(\theta,z)|$$
$$\leqslant \Big\{\int d^3y\, \frac{|V|^{1/2}}{|x-y|}\Big[\int dk\, |h_k(\theta,y)|^2\Big]^{1/2}\Big\}^2$$
$$\leqslant \int d^3y\, \frac{|V(y)|}{|x-y|^2}\int d^3z$$
$$\times \int dk\, |h_k(\theta,z)|^2 \leqslant C\int dk\, \|h_k(\theta,\cdot)\|^2$$

because of (2.1). Therefore, for all θ and x

$$\int_{-\infty}^{\infty} dk\, |\gamma_k(\theta,x) - 1|^2 \leqslant C + C'\int_{-\infty}^{\infty} dk\, \|h_k(\theta,\cdot)\|^2.$$

We must now get an estimate for $\|h_k(\theta,\cdot)\|$. From (2.13),

$$h_k(x) = -(1/4\pi)\int d^3y(1-\mathscr{K}_k)^{-1}(x,y)|V(y)|^{1/2}$$
$$\times \alpha_k(y),$$

so that

$$\|h_k(\theta,\cdot)\|^2 \leqslant (1/4\pi)^2\|(1-\mathscr{K}_k)^{-1}\|^2\,\|V^{1/2}\alpha_k(\theta,\cdot)\|^2.$$

But if $k=0$ is not a singular point, then $\|(1-\mathscr{K}_k)^{-1}\|$ is bounded for each real k, and by (2.7) it tends to one as $k\to\pm\infty$. Consequently, by (2.1)

$$\int_{-\infty}^{\infty} dk\, \|h_k(\theta,\cdot)\|^2$$
$$\leqslant C\int_{-\infty}^{\infty} dk\, \|V^{1/2}\alpha_k(\theta,\cdot)\|^2$$
$$= C\int d^3x\,|V|\int_{-\infty}^{\infty} dk\, |\alpha_k(\theta,x)|^2 < \infty.$$

This completes the proof.

APPENDIX 3

Proof of Lemma 2.5. Define

$$\tau_k^{(n)} = \operatorname{tr} K_k^n.$$

Then

$$\int_{-\infty}^{\infty} dk\, |\tau_k^{(2)}|^2$$
$$\leqslant \int d^3x\, d^3y\, |V(x)V(y)|\int_{-\infty}^{\infty} dk\, I_k(x,x)I_k(y,y)$$
$$\leqslant \Big\{\int d^3x\, |V(x)|\Big[\int_{-\infty}^{\infty} dk\, |I_k(x,x)|^2\Big]^{1/2}\Big\}^2 < \infty$$

by (2.1) and Lemma 2.1. By Schwarz's inequality

$$|\tau_k^{(3)}| = \Big|\int d^3x\, d^3y\, K_k(y,x)K_k^2(x,y)\Big|$$
$$\leqslant \Big[\int d^3x\, d^3y\, |K_k(x,y)|^2\int d^3z\, d^3t\, |K_k^2(z,t)|^2\Big]^{1/2}$$
$$= \Big[\int d^3x\, d^3y\, \frac{|V(x)V(y)|}{|x-y|^2}\int d^3z\, d^3t\, |K_k^2(z,t)|^2\Big]^{1/2}.$$

Hence, by (2.1) and the corollary of Lemma 2.1,

$$\int dk\, |\tau_k^{(3)}|^2 < \infty.$$

Also,

$$|\tau_k^{(4)}| = |\operatorname{tr}(K_k^2K_k^2)| \leqslant \operatorname{tr}(K_k^2K_k^{2\dagger})$$

and hence, by the corollary to Lemma 2.1,

$$\int dk\, |\tau_k^{(4)}|^2 \leqslant C\int dk\, |\tau_k^{(4)}| < \infty,$$

since Lemma 2.1 and (2.1) imply that $\tau_k^{(4)}$ is uniformly bounded.

Finally, for $n>4$

$$|\tau_k^{(n)}| = |\operatorname{tr}(K_k^{n-4}K_k^4)| \leqslant \|K_k^{n-4}\|\operatorname{tr}(K_k^2K_k^{2\dagger})$$
$$\leqslant \|K_k\|^{n-4}\operatorname{tr}(K_k^2k_k^{2\dagger}).$$

But

$$\|K_k\|^2 \leqslant \operatorname{tr}(K_kK_k^\dagger) \leqslant C < \infty,$$

and therefore, for $n\geqslant 4$

$$\int dk\ |\tau_k^{(n)}|^2 < \infty.$$

In fact, by (2.7) $\exists k_0$ such that for all $n \geqslant 4$ and $|k| > k_0$

$$|\tau_k^{(n)}| \leqslant C'C^n \operatorname{tr}(K_k^2 K_k^{2\dagger}),$$

where $C < 1$. It follows that for sufficiently large k_0 the series

$$\log D(k) = -\sum_2^\infty (1/n)\tau_k^{(n)}$$

converges absolutely for all $k \geqslant k_0$, and

$$\int_{|k|>k_0} dk\ |\log D(k)|^2$$
$$\leqslant C'\left[\sum_2^\infty \frac{1}{n}\left(\int_{|k|>k_0} dk\ |\tau_k^{(n)}|^2\right)^{1/2}\right]^2$$
$$\leqslant C'\left(\sum_2^\infty \frac{1}{n} C^n\right)^2 < \infty.$$

Since, furthermore, (2.19) implies that $\exists C$ such that for $k > k_0$

$$|D(k) - 1| \leqslant C\,|\log D(k)|,$$

it follows that

$$\int_{|k|>k_0} dk\ |D(k) - 1|^2 < \infty.$$

But since $D(k)$ is continuous, we also have

$$\int_{-k_0}^{k_0} dk\ |D(k) - 1|^2 < \infty.$$

The lemma follows.

APPENDIX 4

Proof of Lemma 3.2: Employing spherical polar coordinates, writing $u = \hat{y}\cdot\theta$, and integrating by parts

$$\int d\hat{y}\ V(x+y)e^{ik\theta\cdot y}$$
$$= \int_0^{2\pi} d\phi \int_{-1}^1 du\ V(x+y)e^{ik|y|u}$$
$$= \frac{1}{ik|y|}\int_0^{2\pi} d\phi \int_{-1}^1 du\ V(x+y)\frac{\partial}{\partial u} e^{ik|y|u}$$
$$= \frac{2\pi i}{k|y|}\left[e^{ik|y|}V(x-\theta|y|) - e^{ik|y|}V(x+\theta|y|)\right]$$
$$+ \frac{i}{k|y|}\int_0^{2\pi} d\phi \int_{-1}^1 du\ e^{ik|y|u}\frac{\partial}{\partial u} V(x+y)$$
$$= \frac{2\pi i}{k|y|}\left[e^{-ik|y|}V(x-\theta|y|) - e^{ik|y|}V(x+\theta|y|)\right]$$
$$+ \frac{i}{k}\int d\hat{y}\ V'(x,y,\theta)e^{ik\theta\cdot y} \equiv (1/k)\mathring{V}_k(x,|y|,\theta),$$

with

$$V'(x,y,\theta) = \frac{(\theta - \hat{y}\theta\cdot\hat{y})\cdot\nabla V(x+y)}{[1-(\theta\cdot\hat{y})^2]^{1/2}}.$$

Therefore

$$|\mathring{V}_k(x,|y|,\theta)|$$
$$\leqslant (2\pi/|y|)[\,|V(x-\theta|y|)| + |V(x+\theta|y|)|\,]$$
$$+ \int d\hat{y}\,|\nabla V(x+y)|$$
$$\leqslant (4\pi/|y|)M(|\,|x| - |y|\,|) + \int d\hat{y}\,M'(|x+y|)$$
$$\equiv \bar{V}(x,|y|).$$

Now, with B_k defined by (3.2) and shifting variables of integration,

$$(B_k^\dagger B_k)(\theta,\theta')$$
$$= (4\pi/k)\int d^3x\, d^3y\, \frac{V(x)V(y)}{|x-y|}$$
$$\times \sin(k|x-y|)e^{ik(y\cdot\theta' - x\cdot\theta)}$$
$$= (4\pi/k^2)\int d^3x \int_0^\infty d|y|\ |y|$$
$$\times \sin(k|y|)e^{ikx\cdot(\theta'-\theta)}\mathring{V}_k(x,|y|,\theta')V(x).$$

Consequently,

$$k^2\|B_k f\|^2 = 4\pi\int d^3x \int_0^\infty d|y|\ |y|$$
$$\times \sin(k|y|)\tilde{f}_k^*(x)\mathring{f}_k(x,|y|)V(x),$$

where

$$\tilde{f}_k(x) = \int d\theta f(\theta)e^{ikx\cdot\theta},$$

$$\mathring{f}_k(x,|y|) = \int d\theta f(\theta)e^{ikx\cdot\theta}\mathring{V}_k(x,|y|,\theta).$$

Therefore, by Schwarz's inequality

$$\int_{-\infty}^\infty dk\, k^2\|B_k f\|^2$$
$$\leqslant 4\pi\int d^3x\,|V(x)|\int_0^\infty dt\, t\left[\int_{-\infty}^\infty dk\,|\tilde{f}_k(x)|^2\right]^{1/2}$$
$$\times\left[\int_{-\infty}^\infty dk\,|\mathring{f}_k(x,t)|^2\right]^{1/2}.$$

Now

$$\tilde{f}_k(x) = \int_{-1}^1 du \int_0^{2\pi} d\phi\, f(u,\phi)e^{iku|x|},$$

$$\int_{-\infty}^\infty dk\,|\tilde{f}_k(x)|^2$$
$$= \frac{2\pi}{|x|}\int_{-1}^1 du\left|\int_0^{2\pi} d\phi\, f(\theta)\right|^2$$
$$\leqslant \frac{2\pi}{|x|}\left\{\int_0^{2\pi} d\phi\left[\int_{-1}^1 du\,|f(\theta)|^2\right]^{1/2}\right\}^2$$
$$\leqslant \frac{(2\pi)^2}{|x|}\int d\theta\,|f(\theta)|^2 = \frac{(2\pi)^2}{|x|}\|f\|^2;$$

$$\mathring{f}_k = \mathring{f}_k^{(1)} + \mathring{f}_k^{(2)} + \mathring{f}_k^{(3)},$$

$$\mathring{f}_k^{(1)} = \frac{2\pi i}{|y|}e^{-ik|y|}\int d\theta f(\theta)e^{ikx\cdot\theta}V(x-\theta|y|),$$

$$\mathring{f}_k^{(2)} = -\frac{2\pi i}{|y|}e^{ik|y|}\int d\theta f(\theta)e^{ikx\cdot\theta}V(x+\theta|y|),$$

$$\mathring{f}_k^{(3)} = i\int d\theta\int d\hat{y}\, f(\theta)V'(x,y,\theta)e^{ik\theta\cdot(x+y)}.$$

By the same argument as above we obtain

$$\int_{-\infty}^{\infty} dk\ |\mathring{f}_k^{(1)}|^2$$
$$\leqslant (2\pi)^4 |y|^{-2}|x|^{-1}\int d\theta\, |f(\theta)|^2 |V(x-\theta\,|y|)|^2$$
$$\leqslant (2\pi)^4 |y|^{-2}|x|^{-1}\|f\|^2 M^2(|\,|x|-|y|\,|).$$

The same inequality holds for $\mathring{f}_k^{(2)}$.

$$\int_{-\infty}^{\infty} dk\ |\mathring{f}_k^{(3)}|^2$$
$$= 2\pi\int_{-1}^{1} du \left|\int_0^{2\pi} d\phi \int d\hat{y} f \frac{V'}{|x+y|^{1/2}}\right|^2$$
$$\leqslant 2\pi\left\{\int_0^{2\pi} d\phi \int d\hat{y}\left[\int_{-1}^{1} du \left|\frac{fV'}{|x+y|^{1/2}}\right|^2\right]^{1/2}\right\}^2$$
$$\leqslant 16\pi^3 \int d\theta \int d\hat{y}\, |f|^2 \frac{|V'|^2}{|x+y|}$$
$$\leqslant 16\pi^3 \int d\hat{y} \frac{|\nabla V(x+y)|^2}{|x+y|}$$
$$\leqslant 16\pi^3 \|f\|^2 \int d\hat{y} \frac{[M'(|\,|x|-|y|\,|)]^2}{|x+y|}$$
$$\leqslant 64\pi^4 \|f\|^2 [M'(|\,|x|-|y|\,|)]^2 |y|^{-1}$$

by Schwarz's inequality and the hypotheses. Therefore

$$\left[\int_{-\infty}^{\infty} dk\ |\mathring{f}_k|^2\right]^{1/2}$$
$$\leqslant \sum_{i=1}^{3}\left[\int_{-\infty}^{\infty} dk\ |\mathring{f}_k^{(i)}|^2\right]^{1/2}$$
$$\leqslant 8\pi^2\|f\|[\,|y|^{-1}|x|^{-1/2}M(|\,|x|-|y|\,|) + |x|^{-1/2}M'(|\,|x|-|y|\,|)].$$

As a result we have

$$\int_{-\infty}^{\infty} dk\, k^2 \|B_k f\|^2$$
$$\leqslant 64\pi^4\|f\|^2 \int d^3x\, |x|^{-1/2}|V(x)|$$
$$\times \int_0^{\infty} dt\, t\,[\,|x|^{-1/2}t^{-1}M(|\,|x|-t\,|) + |x|^{-1/2}M'(|\,|x|-t\,|)]$$
$$\leqslant 64\pi^4\|f\|^2 \int d^3x |x|^{-1}|V(x)|$$
$$\times\left[\int_0^{\infty} dt\, M(t) + \int_0^{\infty} dt\, tM'(|t-|x|\,|)\right]$$
$$\leqslant C\|f\|^2$$

if V satisfies the hypotheses.

Next consider

$$C_k'(\theta,\theta') = \int d^3x\, d^3y\, V(x)V(y)\frac{e^{ik|x-y|}}{|x-y|} e^{ik(\theta'\cdot y - \theta\cdot x)}$$
$$= \int d^4x\, d^3y\, V(x)\frac{e^{ik|y|}}{|y|} \times V(x+y)e^{ik[\theta\cdot y + (\theta'-\theta)\cdot x]}$$
$$= k\int d^3x \int d\,|y|\ |y|V(x) \times e^{ik|y|}\mathring{V}_k(x,|y|,\theta)e^{ikx\cdot(\theta'-\theta)},$$

from which we obtain

$$k^2\|C_k' f\|^2$$
$$= \int d^3x\, d^3x'\, d\,|y|\, d\,|y'|\ |y|\ |y'|\, V(x)V(x')e^{ik(|y|-|y'|)} \times \int d\theta''\ \mathring{V}_k(x,|y|,\theta'')\mathring{V}_k^*(x',|y'|,\theta'')\tilde{f}_k(x)\tilde{f}_k^*(x').$$

Therefore

$$\int_{-\infty}^{\infty} dk\, k^2\|C_k' f\|^2$$
$$\leqslant 4\pi\int d^3x\, d^3x' \int_0^{\infty} dt\, t\, dt'\, t'\,|V(x)V(x')| \times |\bar{\bar{V}}(x,t)\bar{\bar{V}}(x',t')|\int dk\ |\tilde{f}_k(x)\tilde{f}_k(x')|$$
$$\leqslant 16\pi^3\|f\|^2\left[\int d^3x \int_0^{\infty} dt\, t\,|V(x)|\ |\bar{\bar{V}}(x,t)|\right]^2 |x|^{-1/2}$$
$$\leqslant C\|f\|^2$$

if V satisfies the hypotheses. (This is where $M\in L^1$ is needed.)

The third term we need is

$$D_k(\theta,\theta') = \int d^3x\, d^3y V^{1/2}(x)\mathring{K}_k(x,y) \times |V|^{1/2}(y)e^{ik(y\cdot\theta' - x\cdot\theta)},$$

where $\mathring{K}_k(x,y)$ is the kernel of the operator

$$\mathring{K}_k = (1-K_k)^{-1}K_k^2.$$

Now we have

$$\|D_k\|_2^2 = \mathrm{tr}D_k^{\dagger}D_k$$
$$= (4\pi)^2 k^{-2}\int d^3x\, d^3x'\, d^3y\, d^3y'\, V^{1/2}(x)V^{1/2}(x') \times |V|^{1/2}(y)|V|^{1/2}(y')\frac{\sin(k\,|x-x'|)}{|x-x'|} \times \frac{\sin(k\,|y-y'|)}{|y-y'|}\mathring{K}_k(x,y)\mathring{K}_k(x',y'),$$

so that for every $k_0>0$

$$\int_{|k|>k_0} dk\, \|D_k\|_2^2 k^2$$
$$\leqslant (4\pi)^2\int d^3x\, d^3x'\, d^3y\, d^3y' \frac{|V|^{1/2}(x)|V|^{1/2}(x')}{|x-x'|} \times \frac{|V|^{1/2}(y)|V|^{1/2}(y')}{|y-y'|} \times \int_{|k|>k_0} dk\ |\mathring{K}_k(x,y)\mathring{K}_k(x',y')|.$$

Since (2.1) implies that for every x the function $|V|^{1/2}(y)|x-y|^{-1}$ may be regarded as a vector in $L^2(\mathbb{R}^3)$ as a function of y, with a norm that is uniformly bounded with respect to x, we have

$$\int_{|k|>k_0} dk\, \|D_k\|_2^2 k^2 \leqslant C\int_{|k|>k_0} dk\, \|\mathring{K}_k\|^2.$$

But since $\|(1-K_k)^{-1}\|$ is bounded for all $|k|\geqslant k_0$ and tends to 1 as $|k|\to\infty$, by (2.7), for all $k\geqslant k_0$

$$\|\mathring{K}_k\|^2 \leqslant C\|K_k^2\|^2 \leqslant C\|K_k^2\|_2^2 < C' < \infty$$

by the corollary to Lemma 2.1. As a result

$$\int_{|k|>k_0} dk\, k^2 \|D_k f\|^2$$
$$\leqslant \int_{|k|>k_0} dk\, k^2 \|D_k\|^2 \|f\|^2 \leqslant \int_{|k|>k_0} dk\, k^2 \|D_k\|_2^2 \|f\|^2$$
$$\leqslant C\|f\|^2.$$

It was proved in Ref. 10 that under the hypothesis (2.1) on the potential, the Green's function G_k is $O(1/k)$ at $k=0$ if there is an exceptional point of the second kind at $k=0$. Hence in that case also, $A_k = O(1/k)$. The integral over $k^2\|D_k f\|^2$ may therefore be extended to zero

$$\int_{-\infty}^{\infty} dk\, k^2 \|D_k f\|^2 \leqslant C\|f\|^2.$$

We now assemble our results

$$-4\pi A_k = B_k + C_k' + D_k,$$

$$\int_{-\infty}^{\infty} dk\, k^2 \|A_k f\|^2 \leqslant C\|f\|^2.$$

APPENDIX 5

We shall give here a brief, formal proof of (3.22). We have

$$D^*(k) = \det_2(1 - G_k^{0\dagger} V)$$
$$= \det_2\{(1 - G_k^0 V)$$
$$\times [1 - (1 - G_k^0 V)^{-1}(G_k^{0\dagger} - G_k^0)V]\}.$$

But

$$G_k^0(y,x)^* - G_k^0(x,y) = \frac{ik}{8\pi^2}\int d\theta\, \psi_k^0(\theta,x)\psi_k^0(\theta,y)^*$$

and hence

$$H(x,y) = [(1 - G_k^0 V)^{-1}(G_k^{0\dagger} - G_k^0)](x,y)$$
$$= \frac{ik}{8\pi^2}\int d\theta\, \psi_k(\theta,x)\psi_k^0(\theta,y)^*.$$

Now we easily see from (3.1) that

$$\mathrm{tr}(HV)^n = \left(\frac{k}{2\pi i}\right)^n \mathrm{tr} A_k^n,$$

where the trace on the left means $\mathrm{tr}O = \int d^3x\, O(x,x)$, and the trace on the right, $\mathrm{tr}P = \int d\theta P(\theta,\theta)$. Therefore, by (3.15)

$$\det(1 - HV) = \det\left(1 - \frac{k}{2\pi i} A_k\right) = \det S_k.$$

We now use

$$\det_2[(1 - A)(1 - B)]$$
$$= \det_2(1 - A)\det(1 - B)e^{\mathrm{tr}[(1-A)B]}$$

and

$$\mathrm{tr}[(1 - A)B]$$
$$= \frac{ik}{8\pi^2}\int d\theta \int d^3x\, \psi_k^0(\theta,x)\psi_k^0(\theta,x)^* V(x).$$
$$= \frac{ik}{2\pi}\bar{V}.$$

Therefore (3.22) follows.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant No. PHY77-25337. I am indebted to Professor A. Ramm for pointing out a minor error.

[1] I. Kay and H. E. Moses, Nuovo Cimento **22**, 689 (1961); Commun. Pure Appl. Math. **14**, 435 (1961).
[2] L. D. Faddeev, Dokl. Akad. Nauk SSSR **165**, 514 (1965); **167**, 69 (1966) [Sov. Phys. Dok. **10**, 1033 (1966), **11**, 209 (1966)].
[3] L. D. Faddeev, Report No. ITP-71-106E, Kiev, 1971; and Itogi Nauk. Tekh. Sov. Probl. Mat. **3**, 93 (1974) [J. Sov. Math. **5**, 334 (1976)].
[4] R. G. Newton, in *Scattering Theory in Mathematical Physics*, edited by J. A. Lavita and J.-P. Marchand (Reidel, Dordrecht, 1974), pp. 193–235; also lectures at the Am. Math. Soc. 1974 Summer Seminar on Inverse Problems, University of California (unpublished).
[5] R. T. Prosser, J. Math Phys. **10**, 1819 (1969); **17**, 1775 (1976). Also invited talk at the International Symposium on Ill-Posed Problems: Theory and Practice, October 2–6, 1979, University of Delaware (to be published).
[6] R. G. Newton, Phys. Rev. Lett. **43**, 541 (1979).
[7] R. G. Newton, J. Math. Phys. **21**, 493 (1980).
[8] A solution to this problem was given by Faddeev in Ref. 3, but it is of a nature that makes a practical decision for a given amplitude very difficult.
[9] R. G. Newton, "Inverse Scattering as a Hilbert Problem," invited talk at the International Symposium on Ill-Posed Problems: Theory and Practice, Oct. 2–6, 1979, University of Delaware (to be published).
[10] R. G. Newton, J. Math. Phys. **18**, 1348 (1977).
[11] The inner product of two vectors in $\mathbf{R}^3$ will be denoted by $x\cdot y$.
[12] $M \in L^1$ will actually not be needed except for Lemma 3.2.
[13] C. Zemach and A. Klein, Nuovo Cimento **10**, 1078 (1958); B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U.P., Princeton, New Jersey, 1971), p. 23.
[14] $\dagger$ denotes the Hermitian adjoint, and $*$ the complex conjugate.
[15] See, for example, R. G. Newton, *Scattering Theory of Waves and Particles*, (McGraw-Hill, New York, 1966) p. 259.
[16] This result may be compared to those of W. O. Amrein and D. B. Pearson, J. Phys. A **12**, 1469 (1979) and A. Martin, Commun. Math. Phys. **69**, 89 (1979).
[17] We note that Lemma 3.2 does not imply that the operator norm $\|kA_k\|$ is square integrable.
[18] If S is the abstract S-matrix as defined by time-dependent scattering theory, then $S - 1$ is decomposable in the sense defined by W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, Massachusetts, 1977), pp. 215ff, and S_k is its spectral representation.
[19] B. Simon, *op.cit.*, Ref. 13, p. 106.
[20] This may be true only in the weak sense, as the Fourier transform of $k\theta\cdot\nabla\beta_k$ may not converge.
[21] Some of the arguments presented in this section will be of a rather heuristic nature. We believe that they can be made rigorous for a class of potentials that is more restricted than assumed up to this point.
[21a] As in Sec. 1 of Ref. 7 we conclude that the requirement of an absence of zeros of $\det_k^j$ in $\mathrm{Im}k > 0$ is redundant. Therefore, if a 0-solution J_k° existed then $J_k + J_k^\circ$ should be a 1-solution different from J_k.
[22] For example, $P(\theta,\theta') = a\cdot\theta$, where a is an arbitrary vector in $\mathbf{R}^3$.
[22a] We cannot directly infer from this how J_k^x is related to J_k without a detailed analysis of the asymptotics. The corresponding conclusion in Appendix B of Ref. 7 is based on a faulty equation for φ_k^y and is in error.
[23] See, for example, K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory*, (Springer, New York, 1977), p. 36.
[24] Again this equation may hold only in a weak sense.
[25] See, for example, I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[26] We note, however, that one may need distributional solutions of (8.1) and the uniqueness proof does not apply to them.

The Gel'fand–Levitan equation can give simple examples of non-self-adjoint operators with complete eigenfunctions and spectral representations. I. Ghosts and resonances

B. DeFacio [a)]
Ames Laboratory-USDOE and Department of Mathematics, Iowa State University, Ames, Iowa 50011

H. E. Moses
Center for Atmospheric Research, University of Lowell, Lowell, Massachusetts 01854

(Received 27 December 1979; accepted for publication 20 March 1980)

It is shown that the Gel'fand–Levitan equation for the radial and one-dimensional Schrödinger equations yield complex and singular potentials which support complete sets of eigenfunctions. Explicit examples are given involving complex point eigenvalues and/or complex or negative normalization for the corresponding eigenfunctions. Ghosts and other resonance phenomena thus appear as special cases of this more general inverse spectral theory. It is hoped that these explicit examples for simple situations will clarify ideas relating to non-self-adjoint operators occurring in field theory, scattering theory, and the use of inverse spectral transforms to solve certain classes of nonlinear equations. The treatment given here indicates that the Gel'fand–Levitan equation is a powerful tool in constructing operators with spectral decompositions, even when the operator is not self-adjoint.

1. INTRODUCTION

In recent years there has been a considerable amount of work on the properties of non-self-adjoint operators.[1-8] One central problem for non-self-adjoint operators, which remains open, is the construction of a spectral theory. This not-yet-discovered spectral theory must be applicable to a broad class of non-self-adjoint operators and should be a generalization of the reduction of matrices to Jordan form by a similarity transformation. Much of this non-self-adjoint work has been carried out in the Soviet union and many of these references are available in the monographs by Gohberg and Krein,[1,2] Gohberg and Feldman,[3] and Ramm.[4] Also, additional discussions and more references are available in the books by Dunford and Schwartz,[5,6] Sz-Nagy and Fois,[7] and Reed and Simon.[8]

The spectral theory of self-adjoint operators was synthesized and extended in the work of von Neumann[9] and Stone.[10] J. von Neumann[11] used these ideas to give a mathematically correct formulation of quantum mechanics. The central role of self-adjoint operators in quantum mechanics can be seen from Sec. VIII. 11 of Reed and Simon,[12] which is entitled "Three Mathematical Problems in Quantum Mechanics." These problems are

(1) *self-adjointness* of formally symmetric operators on dense domains,

(2) *spectral analysis*, including position and multiplicity of the point eigenvalues, and

(3) *scattering theory*, the large time-parameter asymtotics. Some of the important physical properties which are guaranteed for all self:adjoint operators include

(i) real eigenvalues,

(ii) orthogonal eigenvectors for nondegenerate eigenvalues, and

(iii) complete sets of eigenvectors.

However, there are physically interesting objects in quantum mechanics, *resonant states*, which have complex energy eigenvalues. The physical picture of resonances is that they arise from perturbations of spectra through "Fermi's second golden rule." Friedrichs[13] initiated a careful mathematical study of these questions and Keldys[14] proved completeness for general non-self-adjoint operators. Livcik[15] studied the dispersion theoretic structure of non-self-adjoint operators and Schwartz[16] showed that the perturbation expansions to certain wave operators for non-self-adjoint Hamiltonians converge. Howlands[17] has extended Friedrichs' method to include models with behavior closely related to the autoionization phenomenon. Herbst and Simon[18] have presented some remarkable examples of the perturbation of spectra and Avron, Herbst, and Simon[19] have developed a spectral theory for self-adjoint Schrödinger operators in magnetic fields. Saénz and Zachary[20] proved the equality of S-matrices for single channel scattering whenever their Hamiltonians were unitarily equivalent, and by treating time reversal invariance, also treated those cases in which an anti-unitary map exists between "equivalent" Hamiltonians. Barut *et al.*[21] have studied a more general non-self-adjoint theory of resonances due to magnetic interactions. Their study is based on the demonstration by Barut and Kraus[21] that two-spin-1/2 particles have deep narrow effective potentials, which binds narrow resonances just like the surprising ψ/J particles. Thus, Ref. 21 provides a non-self-adjoint generalization of the work of Simon *et al.*,[18,19] although the full spectral theory has not yet been developed. Finally, Weber and Hammer[22] have given a contour integration formulation of completeness for symmetric (Hermitian) operators,

[a)]Permanent address: Department of Physics, University of Missouri-Columbia, Columbia, Missouri 65211.

 0022-2488/80/071716-08$1.00

which are not necessarily self-adjoint, on some Hilbert space. Their construction also seems to be valid in a Banach space. Since their approach requires all of the bound states and continuum states as input, to determine a set of expansion coefficients, it is complementary to our study, in some sense.

The papers by Ljance[23] are the most similar in approach to this paper. He uses the Marchenko equation[24] of inverse scattering theory to generate a non-self-adjoint spectral theory, whereas we use the Gel'fand–Levitan[25-28] inverse scattering equation. Ljance gave no specific examples of his theory, where, in contrast, we give both a radial and a one-dimensional example. Also, the Gel'fand–Levitan equation can handle any finite number of bound states whereas the Marchenko equation usually doesn't have point eigenvalues.

Most of these discussions are highly abstract. Examples of the non-self-adjoint operators and their spectra are seldom given, with the important exception of the paper by Zakharov and Shabat,[29] who introduced a system of two one-dimensional equations which solve a nonlinear equation by an inverse spectral transform technique. They gave the explicit non-self-adjoint operators in spectral representations which solved their nonlinear equations. Among the nonlinear inverse problems for which one would like to have a Gel'fand–Levitan algorithm, there are some which are intrinsically non-self-adjoint, e.g., the nonlinear Schrödinger equation with the wrong sign of coupling constant. It is hoped that the present study will aid such studies by clarifying non-self-adjoint linear inverse problems.

The object of the present paper is to give simple explicit examples of non-self-adjoint operators with spectral decompositions. Our examples will be simpler and more systematic than those of Ref. 29 because the equations which we consider—the radial and one-dimensional Schrödinger equations—are simpler than those in Ref. 29. We shall use the Gel'fand–Levitan[25-28] equations of inverse spectral theory to obtain these examples. The fact that we can use the Gel'fand–Levitan equation to construct such non-self-adjoint operators explicitly indicates a new kind of application of the inverse method. Much remains to be learned about the Gel'fand–Levitan equation and its relation to operators on a Hilbert space. We will present some simple examples which show that the Gel'fand–Levitan equation can be used to generate a more general spectral theory than the self-adjoint spectral theory from which it was derived.

In some of our examples the point eigenvalues are complex, in others the normalizations of the eigenfunctions associated with the point eigenvalues are complex (resonant states) or negative (ghosts). Ghosts occur in regularized field theory (Refs. 30 and 31, for example). A nonfield theory example will, hopefully, help to understand the role which they play. The non-self-adjoint inverse methods may prove useful for some important nonlinear problems.

2. THE GEL'FAND–LEVITAN EQUATION FOR THE RADIAL SCHRODINGER EQUATION. THE EXAMPLES

We consider a Hilbert space $\mathscr{H} = L^2\{[0,\infty], d\mu\}$ of complex square integrable functions $f(r)$ with $0 < r < \infty$, the usual inner product

$$(f,g) = \int_0^\infty f^*(r)g(r)\,dr,$$

norm

$$\|f\| = +\sqrt{(f,f)},$$

where $d\mu$ is Lebesgue measure. We will study operators H of the form

$$H = H_0 + V,$$

$$H_0 = -\frac{d^2}{dr^2}, \tag{1}$$

which operate on a dense domain $\mathscr{D}_H$ of $\mathscr{H}$. This domain $\mathscr{D}_H$ consists of doubly differentiable functions which satisfy

$$f(0) = 0.$$

Let us define $\psi_0(r\,|k)$ by

$$\psi_0(r|k) = (2/\pi)^{1/2}\sin kr. \tag{2}$$

Clearly

$$H_0\psi_0(r\,|k) = k^2\psi_0(r\,|k), \tag{3}$$

so that $\psi_0(r\,|k)$ are eigenfunctions of H_0 corresponding to the eigenvalue k^2 in the continuous spectrum. They satisfy the completeness and orthonormality relations

$$\int_0^\infty \psi_0(r\,|k)\psi_0(r'|k)\,dk = \delta(r - r'),$$

$$\int_0^\infty \psi_0(r\,|k)\psi_0(r\,|k')\,dr = \delta(k - k'), \tag{4}$$

For any set of complex numbers $\{\kappa_i\}$, we introduce "pseudo-eigenfunctions" $\psi_{0i}(r)$

$$\psi_{0i}(r) = \sin\kappa_i r \quad (\kappa_i \neq 0), \tag{5}$$
$$= r \quad (\kappa_i = 0).$$

They satisfy the formal eigenfunction equations for H_0

$$H_0\psi_{0i}(r) = E_i\psi_{0i}(r),$$

$$E_i = \kappa_i^2. \tag{6}$$

Since these functions are not square integrable, they are not true eigenfunctions of H_0; they are simply calculational tools.

To complete the spectral data, we introduce, for each i for which κ_i has been given, nonzero complex numbers C_i and also prescribe a (generally complex) function $\omega(k)$ for $-\infty < k < \infty$.

The Gel'fand–Levitan equation for the kernel $K(r\,|r')$ is

$$K(r\,|r') = -\Omega(r\,|r') - \int_0^r K(r\,|r'')\Omega(r''|r')\,dr'', \tag{7}$$

where $\Omega(r\,|r')$ is defined by

$$\Omega(r\,|r') = \int_0^r \psi_0(r\,|k)[\omega(k) - 1]\psi_0(r'|k)\,dk + \sum_i^N \frac{\psi_{0i}(r)\psi_{0i}(r')}{C_i}. \tag{8}$$

Finally, let us define $\psi(r\,|k)$ and $\psi_i(r)$ by

$$\psi(r\,|k) = \psi_0(r\,|k) + \int_0^r K(r\,|r')\psi_0(r'|k)\,dr', \tag{9}$$

$$\psi_i(r) = \psi_{0i}(r) + \int_0^r K(r\,|r')\psi_{0i}(r')\,dr'. \tag{10}$$

Thus $\Omega(0|r') = \Omega(r\,|0) = \Omega(0|0) = 0$. As can be seen, this leads to the equations $K(0|r') = K(r\,|0) = 0$. We shall now prove the following theorem.

Theorem 1: If Eq. (7) has a *unique* solution for $K(r\,|r')$ in $\mathscr{S}$, then

$$\begin{aligned} H\psi(r\,|k) &= k^2\psi(r\,|k), \\ H\psi_i(r) &= E_i\psi_i(r), \end{aligned} \tag{11}$$

where the operator H is given by Eq. (1) and $V(r)$ is given by

$$V(r) = 2\frac{d}{dr}K(r\,|r). \tag{12}$$

Moreover, these eigenfunctions of H satisfy the completeness relation in $\mathscr{S}'$

$$\int_0^r \psi(r\,|k)\omega(k)\psi(r'|k)\,dk + \sum_i \frac{\psi_i(r)\psi_i(r')}{C_i} = \delta(r-r'). \tag{13}$$

Discussion of the Theorem: The original Gel'fand–Levitan equation for the radial Schrödinger equation (Ref. 25) is a special case of our treatment in which the constants κ_i are real and the function $\omega(k)$ and constants C_i are real and positive. In Ref. 25 interest is in self-adjoint operators H. Furthermore, if Eq. (7) has a solution, it is unique under the conditions of Ref. 25. In our generalization uniqueness must be *assumed* for the theorem to hold.

Secondly, we note that the potential $V(r)$ and the wavefunctions $\psi(r\,|k)$ and $\psi_i(r)$ are complex in general, in contrast to the situation of Ref. 31, in which all these quantities are real. Thus the Hamiltonian H is not self-adjoint. Equation (13) is still the spectral resolution of H and provides an expansion theorem as follows: Let $f(r)$ be any complex function in the Hilbert space. Then, in the sense of convergence in the mean,

$$f(r) = \int_0^\infty \psi(r\,|k)\omega(k)a(k)\,dk + \sum_i \psi_i(r)(C_i)^{-1}a_i(k), \tag{14}$$

where in a distributional sense

$$\begin{aligned} a(k) &= \int_0^\infty \psi(r\,|k)f(r)\,dr, \\ a_i &= \int_0^r \psi_i(r)f(r)\,dr, \end{aligned} \tag{15}$$

with the integral signs meaning continuous linear functionals rather than integrals, in general. The space $\mathscr{S}$ is discussed in the Appendix.

One might find Eq. (13) more appealing if one used the notion of biorthogonal functions. Thus we introduce functions $\chi(r\,|k)$ and $\chi_i(r)$ which are eigenfunctions of the adjoint of H, which is its complex conjugate H^*, and write the resolution of the identity Eq. (13) as

$$\int_0^r \psi(r\,|k)\omega(k)\chi^*(r'|k)\,dk + \sum_i \frac{\psi_i(r)\chi_i^*(r')}{C_i} = \delta(r-r'). \tag{16}$$

From Eq. (13) we have

$$\begin{aligned} \chi(r\,|k) &= \psi^*(r\,|k), \\ \chi_i(r) &= \psi_i^*(r), \end{aligned} \tag{17}$$

and hence the biorthogonal functions are indeed eigenfunctions of H^*.

The eigenfunctions $\psi(r\,|k)$ and $\psi_i(r)$ provide a spectral representation of the Hilbert space in the usual sense, i.e., if $f(r)$ is represented by $[a(k),a_i]$ by means of the expansions Eqs. (14) and (15), then $Hf(r)$ is represented by $[k^2a(k), E_ia_i]$, as follows from the eigenfunction equations (11).

Proof: Our method is a generalization of a method used in Ref. 31 and 32. Recall that Eq. (9) gave

$$\begin{aligned} \frac{d}{dr}\psi(r\,|k) = \frac{d}{dr}\psi_0(r\,|k) &+ K(r\,|r)\psi_0(r\,|k) \\ &+ \int_0^r \frac{\partial}{\partial r}K(r\,|r')\psi_0(r'|k)\,dr'. \end{aligned}$$

Thus

$$\begin{aligned} \frac{d^2}{dr^2}\psi(r\,|k) = \frac{d^2}{dr^2}\psi_0(r\,|k) &+ \left[\frac{d}{dr}K(r\,|r)\right]\psi_0(r\,|k) \\ &+ K(r\,|r)\frac{d}{dr}\psi_0(r\,|k) \\ &+ \left[\frac{\partial}{\partial r}K(r\,|r)\right]\psi_0(r\,|k) \\ &+ \int_0^r \frac{\partial^2}{\partial r^2}K(r\,|r')\psi_0(r'|k)\,dr'. \end{aligned} \tag{18}$$

In Eq. (18) and later we use the notation $(\partial/\partial r)K(r\,|r)$ to mean $(\partial/\partial r)K(r\,|r')$ with $r' = r$, and $(\partial/\partial r')K(r\,|r)$ to mean $(\partial/\partial r')K(r\,|r')$ with $r' = r$. We now use $(d^2/dr^2)\psi_0(r\,|k) = -k^2\psi_0(r\,|k)$ to obtain

$$\begin{aligned} \left(\frac{d^2}{dr^2} + k^2\right)\psi(r\,|k) &= \left[\frac{d}{dr}K(r\,|r)\right]\psi_0(r\,|k) \\ &+ K(r\,|r)\frac{d}{dr}\psi_0(r\,|k) + \left[\frac{\partial}{\partial r}K(r\,|r)\right]\psi_0(r\,|k) \\ &+ \int_0^r \frac{\partial^2}{\partial r^2}K(r\,|r')\psi_0(r'|k)\,dr' \\ &- \int_0^r K(r\,|r')\left[\frac{d^2}{dr'^2}\psi_0(r'|k)\right]dr'. \end{aligned} \tag{19}$$

We now consider the last term on the right of Eq. (19) and integrate by parts. It is here that we explicitly make use of the fact that $K(r\,|0)\equiv 0$ and $\psi_0(0|k) = 0$. Thus

$$\begin{aligned} &\int_0^r K(r\,|r')\left[\frac{d^2}{dr'^2}\psi_0(r'|k)\right]dr' \\ &\quad = K(r\,|r)\left[\frac{d}{dr}\psi_0(r\,|k)\right] - \int_0^r \left[\frac{\partial}{\partial r'}K(r\,|r')\right], \\ &\left[\frac{d}{dr'}\psi_0(r'|k)\right]dr' \\ &\quad = K(r\,|r)\left[\frac{d}{dr}\psi_0(r\,|k)\right] - \left[\frac{\partial}{\partial r'}K(r\,|r)\right]\psi_0(r\,|k) \end{aligned}$$

$$+\int_0^r \left[\frac{\partial^2}{\partial r'^2}K(r\,|r')\right]\psi_0(r'|k)\,dr'. \tag{20}$$

On substituting Eq. (20) in Eq. (21), and noting that

$$\frac{\partial}{\partial r}K(r\,|r)+\frac{\partial}{\partial r'}K(r\,|r)=\frac{d}{dr}K(r\,|r), \tag{21}$$

we have

$$\left(\frac{d^2}{dr^2}+k^2\right)\psi(r\,|k)$$
$$=V(r)\psi_0(r\,|k)$$
$$+\int_0^r\left\{\left[\frac{\partial^2}{\partial r^2}-\frac{\partial}{\partial r'^2}\right]K(r\,|r')\right\}\psi_0(r'|k)\,dr'. \tag{22}$$

In Eq. (22), $V(r)$ is given by Eq. (12).

Our next objective is to find $[(\partial^2/\partial r^2)-(\partial^2/\partial r'^2)]\times K(r\,|r')$, which appears in Eq. (22). It is at this point that the requirement that the Gel'fand–Levitan equation (7) have a unique solution is used. Since much of the derivation for the expression follows closely that for obtaining Eq. (22), we shall simply summarize the result. We want to obtain an integral equation for $[(\partial^2/\partial r^2)-(\partial^2/\partial r'^2)]K(r\,|r')$. In a manner similar to that for the derivation of Eq. (19), we obtain from Eq. (7)

$$\frac{\partial^2}{\partial r^2}K(r\,|r')=\frac{\partial^2}{\partial r^2}\Omega(r\,|r')-\left[\frac{d}{dr}K(r\,|r)\right]\Omega(r\,|r')$$
$$-K(r\,|r)\frac{\partial}{\partial r}\Omega(r\,|r')$$
$$-\left[\frac{\partial}{\partial r}K(r\,|r)\right]\Omega(r\,|r')$$
$$-\int_0^r\frac{\partial^2}{\partial r^2}K(r\,|r'')\Omega(r''\,|r')\,dr''. \tag{23}$$

To find $(\partial^2/\partial r'^2)K(r\,|r')$ we apply the operator $(\partial^2/\partial r'^2)$ to both sides of the Gel'fand–Levitan equation (7), use the fact that

$$\frac{\partial^2}{\partial r'^2}\Omega(r\,|r')=\frac{\partial^2}{\partial r^2}\Omega(r\,|r'), \tag{24}$$

integrate by parts, and use $K(r\,|0)\equiv 0$, $\Omega(0|r')\equiv 0$, to obtain

$$\frac{\partial^2}{\partial r'^2}K(r\,|r')=-\frac{\partial^2}{\partial r^2}\Omega(r\,|r')$$
$$-K(r\,|r)\frac{\partial}{\partial r}\Omega(r\,|r')$$
$$+\left[\frac{\partial}{\partial r'}K(r\,|r)\right]\Omega(r\,|r')$$
$$-\int_0^r\left[\frac{\partial^2}{\partial r'^2}K(r\,|r'')\right]\Omega(r''|r')\,dr''. \tag{25}$$

Hence, from Eqs. (23) and (24) we have the following integral equation for $[(\partial^2/\partial r^2)-(\partial^2/\partial r'^2)]K(r\,|r')$

$$\left[\frac{\partial^2}{\partial r^2}-\frac{\partial^2}{\partial r'^2}\right]K(r\,|r')$$
$$=-V(r)\Omega(r\,|r')$$
$$-\int_0^r\left[\frac{\partial^2}{\partial r^2}-\frac{\partial^2}{\partial r'^2}\right]K(r\,|r'')\Omega(r''|r')\,dr''. \tag{26}$$

Let us defind $F(r\,|r')$ by

$$\left[\frac{\partial^2}{\partial r^2}-\frac{\partial^2}{\partial r'^2}\right]K(r\,|r')=V(r)F(r\,|r'). \tag{27}$$

On substituting Eq. (27) in to Eq. (26) it is seen that $F(r\,|r')$ satisfies the Gel'fand–Levitan equation (7). Since the solution of this equation is unique,

$$F(r\,|r')=K(r\,|r'), \tag{28}$$

so that

$$\left[\frac{\partial^2}{\partial r^2}-\frac{\partial^2}{\partial r'^2}\right]K(r\,|r')=V(r)K(r\,|r'). \tag{29}$$

Equation (29) is sometimes used as the starting point of the discussion of the Gel'fand–Levitan equation for the self-adjoint case. For the non-self-adjoint case this result follows from the uniqueness of the Gel'fand–Levitan equation.

Using Eq. (29) in Eq. (22) and using Eq. (9), it follows that $\psi(r\,|k)$ satisfies the Schrödinger equation

$$\left(\frac{d^2}{dr^2}+k^2\right)\psi(r\,|k)=V(r)\psi(r\,|k). \tag{30}$$

The second of Eq. (11) for the eigenfunctions corresponding to the point eigenvalues is proved similarly.

We shall now prove the completeness relation Eq. (16).

In Eq. (7) we replace $\Omega(r''|r)$ by its definition Eq. (8). Thus for $r\geqslant r'$, on interchanging orders of integration and summation,

$$K(r\,|r')=-\Omega(r\,|r')-\int_0^{r\psi_0(r|k)}[\omega(k)-1]\psi_0(r'|k)\,dk$$
$$\times\int_0^r K(r\,|r'')\psi_0(r''|k)\,dr''$$
$$-\sum_i^N\frac{\psi_{0i}(r')}{C_i}\int_0^r K(r\,|r'')\psi_{0i}(r'')\,dr''. \tag{31}$$

But on using Eqs. (9) and (10) and then Eq. (8), we have

$$K(r\,|r')=\int_0^r\psi(r\,|k)[\omega(k)-1]\psi_0(r'|k)\,dk$$
$$+\sum_i^N\frac{\psi_i(r)\psi_{0i}(r')}{C_i}\quad(r\geqslant r'). \tag{32}$$

Now in Eq. (32) use Eqs. (9) and (10), to replace $\psi_0(r\,|k)$ and $\psi_{0i}(r')$ by $\psi(r'|k)$ and $\psi_i(r')$, and integrals involving $K(r\,|r')$ and the unperturbed eigenfunctions. On using Eq. (32) after the substitutions have been made one has

$$K(r\,|r')=-\int_0^r\psi(r\,|k)[\omega(k)-1]\psi_0(r'|k)\,dk$$
$$-\sum_i^N\frac{\psi_i(r)\psi_i(r')}{C_i}$$
$$+\int_0^r K(r\,|r'')K(r'|r'')\,dr''\quad(r\geqslant r'). \tag{33}$$

We can find the expression for $K(r'|r)$ for $r'\geqslant r$ by reversing r and r' in Eq. (33). On introducing the Heavyside function $\eta(x)$ defined by $\eta(x)=1$ for $x>0$ and $\eta(x)=0$ for $x<0$, the results can be combined by writing

$$\int_0^\infty\psi(r\,|k)[\omega(k)-1]\psi(r'|k)\,dk+\sum_i^N\frac{\psi_i(r)\psi_i(r')}{C_i}$$

$$= -\eta(r-r')\int_0^r K(r\,|r'')K(r'|r'')\,dr''$$
$$-\eta(r'-r)\int_0^r K(r\,|r'')K(r'|r'')\,dr''$$
$$-\eta(r-r')K(r\,|r') - \eta(r'-r)K(r'|r). \quad (34)$$

Now from Eqs. (9) and (10) on interchanging orders of integration,

$$\int_0^\infty \psi(r\,|k)\psi(r'|k)\,dk$$
$$= \int_0^\infty \psi_0(r\,|k)\psi_0(r'|k)\,dk + \int_0^r K(r\,|r'')\,dr''$$
$$\times \int_0^\infty \psi_0(r''|k)\psi_0(r'|k)\,dk + \int_0^r K(r'|r'')\,dr''$$
$$\times \int_0^\infty \psi_0(r\,|k)\psi_0(r''|k)\,dk + \int_0^r K(r\,|r'')\,dr''$$
$$+ \otimes \int_0^r K(r'|r''')\,dr \int_0^\infty \psi_0(r''|k)\psi_0(r'''|k)\,dk. \quad (35)$$

We now use the completeness relations for the unperturbed eigenfunctions, which is the first of Eq. (4), to obtain

$$\int_0^\infty \psi(r\,|k)\psi(r'|k)\,dk$$
$$= \delta(r-r')$$
$$+ \eta(r-r')K(r\,|r') + \eta(r'-r)K(r'|r)$$
$$+ \eta(r-r')\int_0^r K(r\,|r'')K(r'|r'')\,dr''$$
$$+ \eta(r'-r)\int_0^r K(r\,|r'')K(r'|r'')\,dr''. \quad (36)$$

On adding Eq. (36) to Eq. (34) we have the completeness relation Eq. (13). Thus the proof of the fundamental theorem has been proved for operators more general than self-adjoint ones. ■

We now come to the examples. They are, in fact, obtainable by analytic continuation of the results of Refs. 33 and 34. Our proofs show that analytic continuations are valid.

In the first example we take $\omega(E) = 1$, and assume a point eigenvalue such that $\kappa_1 = 0$. Unlike the example of Ref. 32, however, we take the normalization constant $C_1 = C$ to be complex. From Eqs. (5) and (8)

$$\Omega(r\,|r') = rr'/C. \quad (37)$$

On substituting into Eq. (7), one sees that $K(r\,|r')$ can be written

$$K(r\,|r') = F(r)r', \quad (38)$$

which, on substituting into Eq. (7), leads to an equation for $F(r)$ which has the unique solution

$$F(r) = -\frac{3r}{r^3+3C}. \quad (39)$$

Hence, the Gel'fand–Levitan kernel is unique and is given by

$$K(r\,|r') = -\frac{3rr'}{r^3+3C}. \quad (40)$$

Thus the scattering potential, which is now complex, is

$$V(r) = \frac{6r(r^3-3C)}{(r^3+3C)^2}. \quad (41)$$

The eigenfunction corresponding to the point eigenvalue is

$$\psi_1(r) = 3C\frac{r}{r^3+3C}. \quad (42)$$

The eigenfunctions $\psi(r\,|k)$ for the continuous spectrum are

$$\psi(r\,|k) = (2/\pi)^{1/2}\left[\sin kr + \frac{3r^2}{r^3+3C}\left(\frac{\cos kr}{k} - \frac{\sin kr}{rk^2}\right)\right]. \quad (43)$$

These eigenfunctions satisfy the completeness relations (13) for any complex value of C. If C is *any* complex number which is not negative and real, the eigenfunction $\psi_1(x)$ is quadratically integrable and has the norm C:

$$\int_0^\infty [\psi_1(r)]^2\,dr = C. \quad (44)$$

If C is a positive real number, we have the usual situation for self-adjoint operators, i.e., for $V(r)$ real. For complex C, however, we have obtained a spectral representation for a non-self-adjoint operator. In particular, the norm of the eigenfunction ψ_1 is complex. We shall refer to such states as resonant states. If C approached a negative real value through a set of complex values, the norms of the sequence of eigenfunctions $\psi_1(r)$ would exist for C arbitrarily close to the negative value. When C is a negative real number, it is easy to see that the left-hand side of Eq. (44) diverges and the eigenfunction is not normalizable. Moreover, the potential has so severe a singularity that H is not a self-adjoint operator. The continuous spectrum eigenfunctions $\psi(r\,|k)$ also have this singularity. Nevertheless, our proof shows that we still have a representation of the Hamiltonian and a corresponding expansion theorem. When C is negative, we shall call the eigenstate $\psi_1(r)$ corresponding to the point eigenvalue a ghost. The notion of ghost states first appeared in field theory and this simple example relates it to non-self-adjoint operators. A different aspect of singular potentials was lucidly discussed by Aly and Müller.[35]

A second example, which also may be considered as an analytic continuation of an example of Ref. 31, is given by the following: We take, as before, $\omega(E) = 1$. We assume that there is a single point eigenvalue $E_1 = \kappa_1^2$ where $\kappa_1 = \kappa$ is *any* complex number. The corresponding normalization constant $C_1 = C$ is also *any* complex number. Under these circumstances,

$$\Omega(r\,|r') = \frac{\sin\kappa r \sin\kappa r'}{C}. \quad (45)$$

The corresponding Gel'fand–Levitan equation can again be solved uniquely by separation of variables. The potential is

$$V(r) = \frac{2\sin\kappa r}{[C+\frac{1}{2}(r-(\sin 2\kappa r)/2\kappa)]^2} \times [\sin\kappa r - 2\kappa(C+\tfrac{1}{2}r)\cos\kappa r]. \quad (46)$$

The eigenfunctions of the continuous spectrum are

$$\psi(r\,|k) = (2/\pi)^{1/2}\{\sin kr - \frac{\sin\kappa r(\kappa\sin kr\cos\kappa r - k\sin\kappa r\cos kr)}{[C+\frac{1}{2}(r-(\sin 2\kappa r)/2\kappa)](k^2-\kappa^2)}\}, \quad (47)$$

The eigenfunction corresponding to the point eigenvalue is

$$\psi_1(r) = C\frac{\sin \kappa r}{C + \tfrac{1}{2}(r - (\sin 2\kappa r)/2\kappa)}. \tag{48}$$

For real values of κ and complex values of C other than real negative values, the potential is everywhere finite and generally complex. Furthermore, the eigenfunction ψ_1 is quadratically integrable with value C. For complex values of C this eigenfunction is a resonant state. As in the previous example, when C approaches the negative real axis the normalization remains finite and equal to C. However, when C is negative, the state no longer has a norm. Moreover, the potential has severe singularities. Nevertheless, the expansion theorem Eqs. (13) and (14) remain valid.

For complex values of κ and C more general statements about the singularities of $V(x)$, ψ_1, and ψ can be made. However, such statements seem too lengthy for the purposes of the present paper. In all cases Eqs. (13) and (14) hold.

3. THE ONE-DIMENSIONAL SCHRÖDINGER EQUATION: AN EXAMPLE

The Gel'fand–Levitan equation for the one-dimensional problem ($-\infty < x < +\infty$) also provides non-self-adjoint operators with complete sets of eigenfunctions.

Let us define $\psi_0(x|E,a)$ and $\psi_{0i}(x)$ by

$$\psi_0(x|E,a) = [2|k|]^{1/2}\psi_0(x|k),$$
$$\psi_0(x|k) = (1/2\pi)^{1/2}e^{ikx}, \tag{49}$$
$$E = k^2, \quad a = (k/|k|),$$

$$\psi_{0i}(x) = \exp[\kappa_i x]. \tag{50}$$

The functions $\psi_0(x|k)$ are, once again, the eigenfunctions of $H_0 = d^2/dx^2$ with the eigenvalues k^2 as given in the momentum representation. The eigenfunctions $\psi_0(x|E,a)$ are the same eigenfunctions expressed in terms of the eigenvalue E and direction of momentum $a = \pm 1$. They satisfy the orthonormality and completeness relation

$$\int_{-\infty}^{+\infty} \psi_0^*(x|E,a)\psi_0(x|E',a')\,dx = \delta(E - E')\delta_{a,a'},$$
$$\sum_a \int_0^\infty \psi_0(x|E,a)\psi_i^*(x'|E,a)\,dE = \delta(x - x'). \tag{51}$$

In Eq. (51) $\delta_{a,a'}$ is a Kronecker delta. The functions $\psi_{0i}(x)$ are formal eigenfunctions of H_0 with the eigenvalues $E_i = -(\kappa_i)^2$, where κ_i is a complex number. The values E_i will, of course, ultimately be identified as a point eigenvalue of a perturbed operator.

Let us introduce functions $\omega_{ab}(E)$ for $a, b = \pm 1$ which are generally complex and complex constants C_i, κ_i for $i = 1,\ldots,N$. We also define

$$\Omega(x|x') = \sum_{a,b}\int_0^\infty \psi_0(x|E,a)[\omega_{ab}(E) - \delta_{a,b}]\psi_0^*(x'|E,b)\,dE + \sum_i \frac{\psi_{0i}(x)\psi_{0i}(x')}{C_i}. \tag{52}$$

If $K(x|x')$ is the unique solution of the Gel'fand–Levitan equation

$$K(x|x') = -\Omega(x|x') - \int_{-\infty}^{x} K(x|z)\Omega(z|x')\,dz, \tag{53}$$

the following theorem holds.

Theorem 2: The functions $\psi(x|E,a)$ defined by

$$\psi(x|E,a) = \psi_0(x|E,a) + \int_{-\infty}^{x} K(x|x')\psi_0(x'|E,a)\,dx', \tag{54}$$

and the functions $\psi_i(x)$ defined by

$$\psi_i(x) = \psi_{0i}(x) + \int_{-\infty}^{x} K(x|x')\psi_{0i}(x')\,dx' \tag{55}$$

satisfy the eigenfunctions equations

$$H\psi(x|E,a) \equiv \left[-\frac{d^2}{dx^2} + V(x)\right]\psi(x|E,a) = E\psi(x|E,a),$$
$$H\psi_i(x) = E_i\psi_i(x), \tag{56}$$
$$(E_i = -\kappa_i^2),$$

with the potential $V(x)$ being given by

$$V(x) = 2\frac{d}{dx}K(x|x). \tag{57}$$

Moreover, these eigenfunctions of H satisfy the completeness relation in $\mathscr{S}$ of

$$\sum_{a,b}\int_0^\infty \psi(x|E,a)\omega_{ab}(E)\psi(x'|E,-b)\,dE + \sum_i^N \frac{\psi_i(x)\psi_i(x')}{C_i} = \delta(x - x'). \tag{58}$$

Proof: This theorem is proved by modifying the corresponding proof for the radial equation, Theorem 1, in an obvious manner.

Like the corresponding result for the radial equation, Eq. (58) is equivalent to an expansion theorem. Similarly, biorthogonal eigenfunctions may be introduced in a manner close to that for the radial equation.

Perhaps the simplest example is the case for which $\omega_{ab}(E) = \delta_{ab}$ and for which there is only one point eigenvalue for which $\kappa \equiv \kappa_1$ has a positive real part. The corresponding normalization constant C is any nonzero complex number. In the case that κ is positive and real and C is positive, we are led to the simplest of reflectionless potentials for a self-adjoint H. In face, the derivation for the potential and eigenfunctions are identical, as are the expressions for them

$$V(x) = -2\kappa^2\operatorname{sech}^2\kappa(x - x_0),$$
$$\psi_1(x) = \tfrac{1}{2}\exp(\kappa x_0)\operatorname{sech}\kappa(x - x_0), \tag{59}$$
$$\psi(x|E,a) = [2|k|]^{1/2}\psi(x|k) \quad \text{[see Eq. (49)]},$$

where

$$x_0 = (1/2\kappa)\ln(2\kappa C), \tag{60}$$

and

$$\psi(x|k) = (1/2\pi)^{1/2}e^{ikx}\left[1 - \frac{\kappa e^{\kappa(x - x_0)}\operatorname{sech}\kappa(x - x_0)}{\kappa + ik}\right]. \tag{61}$$

For any value of C, other than negative real values, and for real κ, there are no singularities in either the wave functions or the potential. Moreover,

$$\int_{-\infty}^{+\infty} [\psi_1(x)]^2\,dx = C. \tag{62}$$

Hence the eigenfunction $\psi_1(x)$ is a resonant function. As in

the example for the radial equation, the eigenfunctions remain square integrable as in Eq. (62) as C approaches the negative real axis. However, if we set C equal to a negative value and thus have a ghost, from the fact that

$$\operatorname{sech}\kappa(x - x_0) = i\,\operatorname{csch}\kappa(x - y_0), \qquad (63)$$
$$y_0 = (1/2\kappa)\ln(2\kappa|C|),$$

singularities do appear at $x = y_0$ in the wavefunctions and potentials. In particular, $\psi_1(x)$ is no longer square integrable, though the expansion theorem implied by the relation Eq. (59) is still valid and in a certain symbolic sense C is still the norm of ψ_1. For complex κ with positive real part and for real positive C there are no singularities and the norm still has meaning. Such cases correspond to resonances. As in the radial equation, the general situation of complex κ with a positive real part and complex C can also be studied. Though for the one-dimensional and radial equation cases we have discussed particularly simple examples, other explicit examples coming from the Gel'fand-Levitan equation abound.

ACKNOWLEDGMENTS

The work of B. DeFacio was supported by the United States department of Energy, contract No.W-7405-eng-82, Division of Engineering, Mathematical and Geosciences, budget code AK-01-04-02 and the work of H. E. Moses was supported by the Air Force Office of Scientific Research. Air Force Systems Command USAF under Contract No. AFOSR-77-3169. The authors wish to thank Professors Justin Peters and J. P. Corones, of the ISU Mathematics Department for useful conversations, and Professor Roger Newton for reading our preprint and making a very useful suggestion.

APPENDIX: THE SPACE $\mathcal{T}$

As stated in Theorems 1 and 2, the Generalized Gel'fand–Levitan kernel $K(x, y)$ exists as as operator on a space $\mathcal{T}$ which we describe here. Unbounded self-adjoint operators, such as H, often do not have enough eigenvectors to span a Hilbert space $\mathcal{H}$.[23,25,36] The unbounded operators of interest have a dense domain $\mathcal{D}_H \subset \mathcal{H}$ and $\mathcal{H} \subset \mathcal{S}$, where $\mathcal{H}$ is spanned by the generalized eigenvectors of H, giving the inclusion

$$\mathcal{D}_H \subset \mathcal{H} \subset \mathcal{S}. \qquad (A1)$$

Choose $\mathcal{S}$ to be "near" $\mathcal{H}$ by constructing a maximal $\mathcal{D}_H$. For example, $\mathcal{S}$ often includes the plane waves of scattering theory as generalized eigenvectors. The space $\mathcal{T}$ which K "lives in" is $\mathcal{S}' \otimes \mathcal{S}'$, where $\mathcal{S}'$ is a space related to, but not equal to, $\mathcal{S}$ as follows. Let $\mathcal{H}'$ be a Hilbert space $L_2(0, \infty)$ for a non-self-adjoint operator.[23] If H is a normal operator $HH^* = H^*H$ or if H is a scalar spectral operator[6] then the construction of the resolution of the identity is possible and then

$$\mathcal{D}_H \subset \mathcal{H}' \subset \mathcal{S}' \qquad (A2)$$

is the inclusion. One has a spectral theory in $\mathcal{H}'$, and $\mathcal{S}'$ which generalizes the self-adjoint spectral theory. This theory is discussed in detail in Refs. (1–8).

Our point is that an inverse spectral theory *also* works for generalized eigenvectors in $\mathcal{S}'$ of (A2) whenever the Gel'fand–Levitan kernel K has a unique solution in $\mathcal{T} = \mathcal{S}' \otimes \mathcal{S}'$ even if H is non-self-adjoint. Furthermore, this inverse spectral theory may be more general than the normal operator or scalar spectral operator spectral theories since it is not based on commuting, unlike normal operators and, its complex poles need not be in a strip.

Notes added in proof:

(1) Three references to non-self-adjoint operator theories were missed in our introduction. Two of these are by Ramm[37,38] and one by Coz.[39] The comments that Ramm[38] showed how to compute complex resonant poles using his earlier work[37] and that he proved the convergence of his method, should appear right after mention of Howland's work in Ref. 18.

(2) Two papers by R. G. Newton on inverse scattering theory, now Refs. 40 and 41, were inadvertently missed. Reference 40 gives a new approach to one-dimensional inverse problems, and Ref. 41 proves existence and uniqueness for a class of three-dimensional inverse problems, including both Gel'fand-Levitan and Marchenko equations without spherical symmetry.

(3) Theorems 1 and 2 can be stated in stronger form. Since $\Omega(r|r')$ is jointly continuous in r and r' in all cases, the Gel'fand–Levitan kernel $K(r|r')$ exists and is unique in $\mathcal{T}$ by standard theorems on Volterra integral equations.[4] Thus, our statement that, "if it exists ..." can be strengthened to a statement of existence and uniqueness of $K(r|r')$ in $\mathcal{T}$ provided that $\Omega(r|r')$ is jointly continuous in r and r'.

(4) We thank Professor A. G. Ramm of the University of Michigan's Mathematics Department for pointing out his Refs. 37 and 38 and for suggesting the strengthened form of our theorems.

[1] I. C. Gohberg and M. G. Krein, *Introduction to the Theory of Linear Non-self-adjoint Operators*, Am. Math. Soc. Transl. Vol. 18 (Russian original 1965) (Am. Math. Soc., Providence, Rhode Island, 1969).
[2] I. C. Gohberg and M. G. Krein, *Theory and Applications of Volterra Operators in Hilbert Space*, Am. Math. Soc. Transl. Vol. 24, (Am. Math. Soc. Providence, Rhode Island, 1970).
[3] I. C. Gohberg and I. A. Feldman, *Convolution Equations and Projective Methods for Their Solution*, Am. Math. Soc. Transl. Vol. 40 (Am. Math. Soc., Providence, Rhode Island, 1974).
[4] A. G. Ramm, *Investigation of Some Classes of Integral Equations and Their Applications* (Springer, New York, in press).
[5] N. Dunford and J. T. Schwartz, *Linear Operators*, Vol. II (Wiley–Interscience, New York, 1963); see especially Sec. XI. 6, 9, and 10.
[6] N. Dunford and J. T. Schwartz, *Linear Operators*, Vol. III (Wiley–Interscience, New York, 1971).
[7] B. Sz-Nagy and C. Fois, *Harmonic Analysis of Operators on Hilbert Space* (North-Holland, Amsterdam, 1970).
[8] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV. Analysis of Operators* (Academic, New York, 1978); See especially Sec. XIII. 16.
[9] J. von Neumann, Math. Ann. **102**, 49 (1929); Ann. Math. **33**, 294 (1936).
[10] M. Stone, *Linear Transformations in Hilbert Spaces and Their Application to Analysis*, Am. Math. Soc. Colloq. Pub. 15 (Am. Math. Soc., New York, 1932).
[11] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton U. P., Princeton, New Jersey, 1955).
[12] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I, Functional Analysis* (Academic, New York, 1972); see especially pp. 302 *et seq.*

[13]K. O. Friedrichs, Math. Ann. **115**, 249 (1938); Commun. Pure Appl. Math. **1**, 361 (1948).
[14]M. V. Keldys, Dokl. Akad. Nauk SSSR **77**, 11 (1951); In Russian, Math. Rev. **12**, 835 has an English summary.
[15]M. Livcik, Usp. Mat. Nauk **12** (1957); In Russian, a full English translation appears in Amer. Math. Soc. Transl. Ser. 2, vol. 16 (Am. Math. Soc., Providence, Rhode Island, 1960), p. 422.
[16]J. T. Schwartz, Commun. Pure Appl. Math. **13**, 609 (1960).
[17]J. S. Howlands, Bull, Am. Math. Soc. **78**, 380 (1972).
[18]I. W. Herbst and B. Simon, Phys. Lett. **78** B, 304 (1978).
[19]J. Avron, I. Herbst, and B. Simon, Phys. Rev. Lett. **39**, 1068 (1977); J. Avron, I. Herbst, and B. Simon, Phys. Lett. A **62**, 214 (1977); and J. Avron, I. Herbst, and B. Simon, Ann. Phys. (N. Y.) **114**, 431 (1978).
[20]A. W. Saénz and W. W. Zachary, J. Math. Phys. **17**, 409 (1976).
[21]A. O. Barut and J. Kraus, Phys. Rev. D **16**, 161 (1977); and A. O. Barut, M. Berrondo, and G. Garcia-Calderón, J. Math. Phys. **21**, xxxx (1980).
[22]T. A. Weber and C. L. Hammer, J. Math. Phys. **21**, 24 (1980).
[23]V. E. Ljance, Mat. Sb. **64**, 521 (1964) and **65**, 47 (1964). In Russian; English translation available in Amer. Math. Soc. Transl. vol. 60 (Am. Math. Soc., Providence, Rhode Island, 1967, p. 60.)
[24]Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory*, translated from the Russian by B. D. Seckler (Gordon and Breach, New York, 1963).
[25]I. M. Gel'fand and B. M. Levitan, Isv. Akad. Nauk SSSR Math. Series **15**, 309 (1951). In Russian; English translation available in Amer. Math. Soc. Transl. Ser. 2, vol. 1 (Am. Math. Soc., Providence, Rhode Island, 1955), p. 253.
[26]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw–Hill, New York, 1966).
[27]K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, New York, 1977).
[28]R. G. Newton in, *Scattering Theory in Mathematical Physics*, edited by J. A. LaVita and J.-P. Marchand (Reidel, Boston, 1974).
[29]V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972).
[30]K. Nagy and J. Rzewuski, Bull. Acad. Pol. Sci. Co. III **7**, 93 (1958).
[31]N. Nakanishi, Phys. Rev. D **5**, 1968 (1972).
[32]H. E. Moses and C. M. deRidder, MIT Lincoln Laboratory Technical Report No. 322, 1963 (unpublished).
[33]I. Kay and H. E. Moses, Nuovo Cimento **22**, 689 (1961).
[34]H. E. Moses and S. F. Tuan, Nuovo Cimento **13**, 197 (1959).
[35]H. H. Aly and H. W. J. Müller, J. Math. Phys. **7**, 1 (1965).
[36]Ju. M. Berezanskii, *Expansions in Eigenfunctions of Self-adjoint Operators*, Vol. 17 Transl. Am. Math. Soc. (Am. Math. Soc. Providence, Rhode Island, 1968).
[37]A. G. Ramm, Dokl. **191**, 50 (1970).
[38]A. G. Ramm, Dokl. **204**, 1071 (1972).
[39]M. Coz, University of Kentucky Preprint (to be published).
[40]R. G. Newton, J. Math. Phys. **21**, 493 (1980).
[41]R. G. Newton, J. Math. Phys. **21**, 1698 (1980).

The Jost solutions for Yukawa potentials

L. Trlifaj
Institute of Physics of the Czechoslovak Acad. Sci, Prague, Czechoslovakia

(Received 9 January 1979; accepted for publication 20 August 1979)

Using a suitable Laplace transformation of the regular solution and the expansion of its Laplace transform in partial fractions we found new explicit expressions for the Jost solutions of the radial Schrödinger equation with complex wave numbers and angular momenta, in general, and for potentials of the exponential and, especially, Yukawa type, which were described by means of the Stiltjes integral with respect to a function $\mu(t)$ of bounded variation. Many other relations including those for the Jost functions were derived or rederived, which have their counterpart among relations for the Bessel functions.

1. INTRODUCTION

The physical theory of the nonrelativistic scattering of spinless particles by a central potential is a mathematically elaborated theory nowadays. This concerns not only the direct and inverse problems of the scattering theory but also problems arising from making the angular momentum as well as the energy of particles complex. The corresponding analytic continuation of the wave functions is governed by the Schrödinger equation, in which the interaction potential is usually taken real. In studying these problems a local potential of the Yukawa type that is given by a Stiltjes integral with respect to a function $\mu(t)$ of bounded variation with a real "interaction" constant g

$$V(r) = -\frac{g}{r}\int_{\tau_2}^{\tau_1} d\mu(t)\exp(-tr) \quad (0<\tau_1<\tau_2\leqslant\infty) \qquad (1)$$

played the important role. It is simple and general enough and possesses some justification from the side of particle physics. The first property was fully made use of in the investigation of the S waves for which there only exist relatively simple mathematical expressions or procedures. This is not the case for higher waves where such an approach is more complicated and does not lead to closed and explicit formulas.[1,2] The reason consists in the presence of a centrifugal potential, which is zero for the S wave.

We overcame this difficulty by using a suitable form of the Laplace transformation and derived explicit formulas for wave functions. They are solutions of the complicated recurrence relations we have mentioned. At the same time we opened a way towards the analytic continuation of wave functions to complex values of angular momenta and simplified the investigation of some questions, such as the convergence of Born series and the determination of their regions of analyticity and of their asymptotic behavior.

Problems of scattering processes are solved on the background and in the relation to the purely kinematic, forceless processes which are described by a different kind of Bessel function and by the Legendre polynomials. We therefore expect that there is an analogy between the radial scattering wave functions and the Bessel functions. We succeeded, indeed, in finding similar expressions and relations for them except the single but essential distinction given by the possible occurence of the natural Yukawa cuts. This was also the standpoint which determined the choice and the extent of the problems solved. On the other hand such a restriction was necessary because the subject had already been worked out in various aspects in the literature. As for the particular papers we refer to the monographs,[1-4] where a complete information is collected. We mention only an early paper[5], where some allied formulas can be found.

As the potential (1) is spherically symmetric, the scattering matrix S is diagonal in the quantum numbers of angular momentum and can be determined by solving the radial Schrödinger equation for partial waves. We made use of the exceptionality of the S waves in order to describe in more detail the approach, that is then applied to the case of general angular momenta. According to this approach, the Laplace transform of the radial wave functions with the integral angular momenta is expanded in partial fractions, which enables us to pass from the regular to the irregular (Jost) solutions and to continue them analytically to complex values of angular momenta. In the case of the S waves a slightly different class of potentials instead of (1) was used, i.e., potentials of the exponential type

$$V(r) = -g\int_{\tau_1}^{\tau_2} d\mu_e(t)\exp(-tr) \quad (0<\tau_1<\tau_2\leqslant\infty). \qquad (2)$$

For some classes of functions $\mu(t)$ the potentials (1) and (2) can be identified.

2. PRELIMINARIES

A regular solution $\varphi(\lambda,k,r)$ of the radial Schrödinger equation

$$\left(-\frac{d^2}{dr^2}+\frac{\lambda^2-\frac{1}{4}}{r^2}+V(r)-k^2\right)\psi(\lambda,k,r)=0 \qquad (3)$$

as a function of complex values of the angular momenta $l=\lambda-\frac{1}{2}$ and of the wave numbers k is defined for $\mathrm{Re}\lambda\geqslant 0$ by the boundary condition

$$\varphi(\lambda,k,r) = r^{\lambda+1/2} \left[\approx (k/2)^{-\lambda}\Gamma(\lambda+1)r^{1/2}J_\lambda(kr)\right] \text{ as } r\to 0^+. \qquad (4a)$$

The corresponding irregular solution $\varphi(-\lambda,k,r)$ is uniquely determined as the analytic continuation of $\varphi(\lambda,k,r)$ only when it exists. Therefore, two other linearly independent irregular solutions of Eq. (3) are used, namely the Jost func-

 0022-2488/80/071724-08$1.00

tions $f_+(\lambda,k,r)$ and $f_-(\lambda,k,r)$, which are defined by the boundary conditions

$$f_\pm(\lambda,k,r) = \exp(\mp ikr)$$
$$\{\sim f_0(\lambda,\pm k,r)$$
$$\equiv \exp\left[-i\left(\frac{\pi}{2}\right)(\lambda+\tfrac{1}{2})\right](\pm kr\pi/2)^{1/2}H_\lambda^{(2)}(\pm kr),$$
$$-k = e^{-i\pi}k, \quad -\pi < \arg k < \pi\} \quad \text{as } r\to\infty. \tag{4b}$$

Expressions written in the curly brackets of Eqs. (4) with the Bessel and Hankel functions (of the second kind) were added in order to show which of the kinematic solutions fulfill the conditions (4). Moreover, in the case (4b), along with a convention $-k = e^{-i\pi}k$, they make possible a unique determination of the Jost solutions based now on the analytic continuation in the complex k plane. Here, we remark, that our Jost functions are the traditional ones.[1,3] They are used in order to make easier comparison of the obtained formulas. In the monographs,[2,4] a slightly different definition of the Jost functions is given. The regular solutions as well as the irregular ones can be derived by iterating the Fredholm integral equations,[1,3] which are equivalent to the differential equations (3) and to one of the boundary conditions (4). If the expressions in the brackets (4) are iterated, the well-known Born series for functions $\varphi(\lambda,k,r)$ and $f_\pm(\lambda,k,r)$ are obtained.

The forms (1) and (2) of the potential $V(r)$ suggest the application of the Laplace transformation in finding the regular solution. The difficulty with the centrifugal term in Eq. (3) prefers the following form of the Laplace transformation:

$$\omega(\lambda,k,s) = \int_0^\infty dr\, r^{\lambda-1/2}e^{-sr}\varphi(\lambda,k,r)$$
$$(\mathrm{Re}s > |\mathrm{Im}k|,\ \mathrm{Re}\lambda > 0), \tag{5a}$$
$$r^{\lambda-1/2}\varphi(\lambda,k,r) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} ds\, e^{sr}\omega(\lambda,k,s)(c > |\mathrm{Im}k|), \tag{5b}$$

where, according to Eq. (4a), asymptotically $\omega(\lambda,k,s) \sim \int_0^\infty dr\, r^{2\lambda}e^{-sr} = s^{-2\lambda-1}\Gamma(2\lambda+1)$ as $s\to\infty$. Applying the operation $\int_0^\infty dr\, r^{\lambda+1/2}e^{-sr}\cdots$ to Eq. (3) and making some elementary manipulations we transform Eq. (3) into a first-order differential equation for the function $\omega(\lambda,k,s)$ with one argument s shifted. It can easily be integrated. In view of the prescribed asymptotic behavior of $\omega(\lambda,k,s)$, we have

$$\omega(\lambda,k,s) = \Gamma(2\lambda+1)(s^2+k^2)^{-\lambda-1/2} - g(s^2+k^2)^{-\lambda-1/2}\int_{c_\infty}^{s} dz\,(z^2+k^2)^{\lambda-1/2} \times \int_{\tau_1}^{\tau_2} d\mu_e(t)\frac{d}{dz}\omega(\lambda,k,z+t) \tag{6}$$

for the potential (2). Since the function $\omega(\lambda,k,s)$ is analytic in the half-plane $\mathrm{Re}s > |\mathrm{Im}k|$, we can take any point $c_\infty = c + \infty e^{i\varphi}$ with $|\varphi| < \frac{1}{2}\pi$ for the point c_∞.

The performed transformations of the potential term in Eq. (3) are correct for $\mathrm{Re}\lambda > 0$, if $\left|\int_0^\infty dr\, rV(r)\right| < \infty$ according to Eq. (4a). This condition defines the allowed classes of functions $\mu(t)$ in the potential (1) and (2). We remember that any function $\mu(t)$ of bounded variation can be written as a difference of two (bounded) nondecreasing positive functions $\mu_1(t)$ and $\mu_2(t)$ [$\mu(t) = \mu_1(t) - \mu_2(t)$]. Each of these functions determines by itself the Stiltjes integral with respect to the function $\mu(t)$. This is important to keep in mind, especially when $\tau_2 = \infty$ in Eqs. (1) and (2). Always denoting $\mu_1(t) + \mu_2(t)$ by $\nu(t)$ in what follows, we satisfy condition $\left|\int_0^\infty dr\, rV(r)\right| < \infty$ by demanding that

$$\nu_p(\tau_2) = \int_{\tau_1}^{\tau_2} d\nu(t)t^p < \infty, \tag{7a}$$

where $p = -1$ for the potential (1) and $p = -2$ for the potential (2). As $\int_{\tau_1}^{\tau_2} d\nu(t)\, e^{-tr} = (-d/dr)^{-p}\int_{\tau_1}^{\tau_2} d\nu(t) \times t^p e^{-tr}$ these potentials behave as $r^{-2+\epsilon}(\epsilon > 0)$ at $r = 0^+$. Equation (7a) is satisfied for any $p < \kappa$, where $\kappa \geqslant 0$, by a class of functions $\mu(t)$, which is characterized by such a maximum number κ. In the case of the potential (1) then, it implies that $rV(r)$ and its first $[\kappa]$ derivatives are continuous for $r \geqslant 0$.

Subclasses of functions $\mu(t)$ characterized by maximum nonnegative integers $q \leqslant \kappa$ besides κ fulfill the requirement

$$0 = \mu_p(\tau_2) = \int_{\tau_1}^{\tau_2} d\mu(t)\, t^p \text{ for any integer } 0 \leqslant p \leqslant q \leqslant \kappa. \tag{7b}$$

It implies for the potentials of the Yukawa type that $rV(r)$ and its first q derivatives are zero at $r = 0^+$. Concluding these consideration we emphasize that the introduction of classes κ is useful and nontrivial only if the functions $\mu(t)$ fully realize themselves in the whole interval $[\tau_1, \infty)$, i.e., if $\tau_2 = \infty$.

In the intersection of the analyticity domains of all the functions considered the following well-known relation[1-4] holds:

$$\varphi(\lambda,k,r) = (1/2ik)[f_+(\lambda,k)f_-(\lambda,k,r) - f_-(\lambda,k)f_+(\lambda,k,r)]. \tag{8}$$

According to (8) the diagonal elements $S(\lambda,k)$ of the scattering matrix S are defined by the Jost functions $f_\pm(\lambda,k)$ as follows:

$$S(\lambda,k) = [f_+(\lambda,k)/f_-(\lambda,k)]\exp[i\pi(\lambda-\tfrac{1}{2})] = \exp[2i\delta(\lambda,k)], \tag{9}$$

where the phase shifts $\delta(\lambda,k)$ are real for physical values of k and λ. The behavior of the function $S(\lambda,k)$ plays an important role in the realization of the Sommerfeld–Watson transformation of the scattering amplitude, which eventually leads to the concept of the Regge poles.

If we succeed in decomposing the Laplace transform $\omega(\lambda,k,s)$ in an analogous way to the relation (8) we obtain both the Jost functions and the Jost solutions.

3. POTENTIALS OF THE EXPONENTIAL TYPE, $\lambda = \frac{1}{2}$

If we put $\lambda = \frac{1}{2}$ in Eq. (6), its second term on the right-hand side can be integrated and the whole equation can be solved by iterations. We get

$$\omega(\tfrac{1}{2},k,s) = A^{-1}(s,k)\left(1 + \sum_{n=1}^{\infty}(-g)^n\int d\mu_e(t_1)\cdots \times \int d\mu_e(t_n)A^{-1}(s+v_1,k)\cdots A^{-1}(s+v_n,k)\right), \tag{10a}$$

where

$$A(s,k)=s^2+k^2,\quad v_0=0,$$
$$v_1=t_1,\ldots,\quad v_m=t_1+\cdots+t_m. \tag{10b}$$

As

$$|A(s+v_m,k)|^{-1}<v_m^{-2}\quad \text{for } \mathrm{Re}s>|\mathrm{Im}k| \tag{11a}$$

and

$$\left|\int d\mu_e(t_1)\cdots\int d\mu_e(t_n)\times A^{-1}(s+v_1,k)\cdots A^{-1}(s+v_n,k)\right|$$
$$<\int t_1^{-2}\,dv_e(t_1)\cdots\int t_n^{-2}\,dv_e(t_n)\left(\frac{t_1\cdots t_n}{v_1\cdots v_n}\right)^2$$
$$<\tfrac{1}{2}\int t_1^{-2}\,dv_e(t_1)\cdots\int t_n^{-2}\,dv_e(t_n)\times\frac{t_3\cdots t_n}{(t_1+t_2+t_3)\cdots(t_1+\cdots+t_n)}$$
$$=\frac{1}{n!}\left(\int_{\tau_1}^{\tau_2}t^{-2}\,dv_e(t)\right)^n, \tag{11b}$$

the series (10a) converges absolutely and uniformly if $\nu_{-2}(\tau_2)<\infty$ in accordance with the condition (7a). If the qth interation of Eq.(6) is carried out, the function $\omega(\frac{1}{2},k,s+v_{q+1})$ appears in its last term. According to the foregoing statement it is an analytic function of s for $\mathrm{Re}s>|\mathrm{Im}k|-(q+1)\tau_1$. The other terms and factors of this iteration are also analytic functions of s except for poles or cuts

$$\pm ik,\quad \pm ik-\tau_1-t\quad (t\in[0,\infty)), \tag{11c}$$

depending on the nature of the function $\mu_e(t)$. Hence the analyticity region of the function (10a) can be extended to the whole complex s-plane except for the points (11c).

We expand the fractions contained in each term of series (10a) into a sum of partial fractions:

$$A^{-1}(s+v_0,k)\cdots A^{-1}(s+v_n,k)$$
$$=(1/2ik)(s-ik)^{-1}A^{-1}(ik+v_1,k)\cdots A^{-1}(ik+v_n,k)$$
$$+\sum_{m=1}^{n-1}A^{-1}(ik-v_m+v_0,k)\cdots A^{-1}(ik-v_m+v_{m-1},k)(1/2ik)(s+v_m-ik)^{-1}A^{-1}(ik+v_{m+1}-v_m,k)\cdots A^{-1}(ik-v_m+v_n,k)+A^{-1}(ik-v_n,k)\cdots A^{-1}(ik-v_n+v_{n-1},k)(1/2ik)(s+v_n-ik)^{-1}$$
$$+(\text{the same expression with } k \text{ and } -k \text{ interchanged}). \tag{12}$$

This expansion is valid unless $s+v_p\pm ik=s+v_q\mp ik$ for $p\neq q$. We evade this possibility by taking out the cuts on poles

$$\pm i(\tfrac{1}{2}\tau_1+x),\quad x\in[0,\infty) \tag{13}$$

from the complex k plane.

Each term of the expansion (12) consists of the product of two independent factors, one being a function of variables $(t_1,\ldots,t_m)$ only and the other one of the complementary variables $(t_{m+1},\ldots,t_n)$ only. This makes it possible to rearrange the series (10a), so that

$$\omega(\tfrac{1}{2},k,s)=(1/2ik)[F(\tfrac{1}{2},k)\Omega(-k,s)-F(\tfrac{1}{2},-k)\Omega(k,s)], \tag{14a}$$

where the series

$$F(\tfrac{1}{2},k)=1+\sum_{n=1}^{\infty}(-g)^n\int d\mu_e(t_1)\cdots\int d\mu_e(t_n)\times A^{-1}(ik+v_1,k)\cdots A^{-1}(ik+v_n,k)\equiv F_+(\tfrac{1}{2},k)$$
$$[\text{and } F_-(\tfrac{1}{2},k)=F(\tfrac{1}{2},-k)], \tag{15a}$$

$$\Omega(k,s)=(s+ik)^{-1}+\sum_{n=1}^{\infty}(-g)^n\int d\mu_e(t_1)\cdots\int d\mu_e(t_n)\times A^{-1}(ik+v_1,k)\cdots A^{-1}(ik+v_n,k)\times(s+v_n+ik)^{-1}, \tag{14b}$$

converges absolutely and uniformly except for zeros of the functions A and $s+v+ik$ provided $\nu_{-2}(\tau_2)<\infty$. The inverse Laplace transformation (5b) gives

$$\varphi(\tfrac{1}{2},k,r)=(1/2ik)[F(\tfrac{1}{2},k)f(\tfrac{1}{2},-k,r)-F(\tfrac{1}{2},-k)f(\tfrac{1}{2},k,r)], \tag{15b}$$

where the Jost solution $f_+(\frac{1}{2},k,r)=f(\frac{1}{2},k,r)$ and $f_-(\frac{1}{2},k,r)=f(\frac{1}{2},-k,r)$.

$$f(\tfrac{1}{2},k,r)=e^{-ikr}\left\{1+\sum_{n=1}^{\infty}(-g)^n\int d\mu_e(t_1)\cdots\int d\mu_e(t_n)\times e^{-v_nr}A^{-1}(ik+v_1,k)\cdots A^{-1}(ik+v_n,k)\right\}, \tag{15c}$$

as well as the (normalized) Jost function $F(\frac{1}{2},k)=f(\frac{1}{2},k,0)$ is given by a Born series, which is an analytic function of k except for the natural upper cut (poles) (13) provided $\nu_{-2}(\tau_2)<\infty$. We can easily verify that the series (15c) satisfies Eq. (3).

If a formal variable $x\geqslant 0$ and a function $G(k,r,x)$, which satisfies the integral equation $[G(k,r,\infty)=1]$:

$$G(k,r,x)=1-g\int_{\tau_1}^{\tau_2}d\mu_e(t)\,A^{-1}(ik+t+x,k)\times e^{-tr}G(k,r,t+x), \tag{16}$$

are introduced, then the solution of the equation determines the Jost solution and the Jost function $f(\frac{1}{2},k,r)=e^{-ikr}G(k,r,0)$ and $F(\frac{1}{2},k)=G(k,0,0)$, respectively. It is worth remarking that there exists the solution $G(k,0,x)=[x/(x+2ik)]^{g/2ik}$ of Eq. (16) for the coulomb potential, which does not belong to the class of the allowed potentials because of $\tau_1=0$ and $\tau_2=\infty$, $[d\mu_e(t)=dt]$. In the case of the potentials like the Hulthén, Woods–Saxon, Eckhart, and Morse potentials

$$g\,d\mu_e(t)=\sum_{n=1}^{\infty}g_n\,d\Theta(t-n)$$
$$[\Theta(t)=1 \text{ for } t>0 \text{ and } \Theta(t)=0 \text{ for } t<0], \tag{17}$$

Eq. (16) converts into a difference equation of the infinite order, in general, which, however, can be transformed into a first- or second-order difference equation and solved by ordinary methods. They are not always too simple even if bound-state problems only are solved.

The eigenvalues $E=-b^2\,(b>0)$ of the bound S states are solutions of the equation $F(\frac{1}{2},-ib)=0$ and the corresponding regular solution $\varphi(\frac{1}{2},+ib,r)=-(1/2b)$

$\times F(\frac{1}{2},ib)f(\frac{1}{2},-ib,r)$. According to Eq. (15c) it formally resembles the potential (2). This suggests the idea of solving the simple Hartree–Fock problem $V(r)=\text{const}[r^{-1}\varphi(\frac{1}{2},+ib,r)]^2$ by finding the appropriate function $\mu_e(t)$ of the potential (2).

4. POTENTIALS OF THE YUKAWA TYPE, λ GENERAL

The potentials of the type (1) are simpler for $\lambda\neq\frac{1}{2}$ in the technical aspect, for the equation corresponding to Eq. (6) now reads:

$$\omega(\lambda,k,s)=\Gamma(2\lambda+1)A^{-\lambda-1/2}(s,k)-gA^{-\lambda-1/2}(s,k)\times\int_0^\infty dz\,A^{\lambda-1/2}(s+z,k)\times\int_{\tau_1}^{\tau_2}d\mu(t)\,\omega(\lambda,k,s+z+t),\quad \mathrm{Re}\lambda>0,\quad (18)$$

in view of analytic properties of the function $\omega(\lambda,k,s)$. Its iterative solution

$$\omega(\lambda,k,s)=\Gamma(2\lambda+1)A^{-\lambda-1/2}(s,k)\Big\{1+\sum_{n=1}^{\infty}(-g)^n\times\int_0^\infty dz_1\int d\mu(t_1)\cdots\int_0^\infty dz_n\int d\mu(t_n)\times\prod_{p=1}^{n}[A^{\lambda-1/2}(s+u_p+v_{p-1},k)\times A^{-\lambda-1/2}(s+u_p+v_p,k)]\Big\},\quad (19a)$$

where

$$u_0=v_0=0,\quad u_m=\sum_{p=1}^{m}z_p,\quad v_m=\sum_{p=1}^{m}t_p,$$

$$A(s,k)=s^2+k^2,\quad (19b)$$

represents again an analytic function of s in the whole complex plane except for cuts (11c) under the assumption $\nu_{-1}(\tau_2)<\infty$ for $\mathrm{Re}\lambda\geqslant\frac{1}{2}$ in accordance with the condition (7a). This follows from the evident inequalities

$$\left|A(s+u_p+v_{p-1},k)A^{-1}(s+u_p+v_p,k)\right|^{\mathrm{Re}\lambda-1/2}<1,\quad \text{for } \mathrm{Re}\lambda\geqslant\tfrac{1}{2}\text{ and }\mathrm{Re}s>|\mathrm{Im}k|,\quad (20a)$$

$$\int_0^\infty dz_p\left|(s+u_p+v_p)^2+k^2\right|^{-1}<\int_0^\infty dz_p\,(u_p+v_p)^{-2}<v_p^{-1},\quad \text{for } \mathrm{Re}s>|\mathrm{Im}k|,\quad (20b)$$

and from the further considerations, which go along the same lines as in Sec. 3 [Eqs. (11)]. If $-\frac{1}{2}\leqslant\mathrm{Re}\lambda<\frac{1}{2}$, the stronger condition upon functions $\mu(t)$ coming from the requirement of the convergence of series (19a) must be taken $\nu_0(\tau_2)<\infty$ and so on.

Let us assume for the time being that λ are positive half-integers ($\lambda=\frac{1}{2},\frac{3}{2},\cdots$). The expansion in partial fractions differs, now, from that one in Eq. (12) by the presence of powers $(s+u_m+v_m-ik)^{-p}, p=1,\ldots,\lambda+\frac{1}{2}$. They are multiplied by the coefficients, which are the $(\lambda+\frac{1}{2}-p)$th derivatives of the product of A and A^{-1} factors and of $(s+u_m+v_m-ik)^{\lambda+1/2}$ at the point $s=-u_m-v_m+ik$ divided by $(\lambda+\frac{1}{2}-p)!$. They factorize again, one factor depending on the group of variables $(z_1,\ldots,z_m;t_1,\ldots,t_m)$ only, the other on the variables $(z_{m+1},\ldots,z_n;t_{m+1},\ldots,t_n)$ only. In the latter factor, nonzero derivatives do not contribute in the resulting series (19a). One can easily show this by integrating by parts with respect to z_{m+1}. The expansion is valid except for the cuts (13), so that the series (19a) can be rearranged and both its terms dependent on s and those independent of s can be summed separately. The latter sum gives the normalized Jost function

$$F_+(\lambda,k)\equiv F(\lambda,k)=1+\sum_{n=0}^{\infty}(-g)^n\int_0^\infty dz_1\int d\mu(t_1)\cdots\int_0^\infty dz_n\int d\mu(t_n)\times\prod_{m=1}^{n}[A^{\lambda-1/2}(u_m+v_{m-1}+ik,k)\times A^{-\lambda-1/2}(u_m+v_m+ik,k)]$$

$$=\lim_{s\to ik}\Gamma^{-1}(2\lambda+1)A^{\lambda+1/2}(s,k)\omega(\lambda,k,s),$$

$$F_-(\lambda,k)=F(\lambda,-k),\quad (21)$$

which is an analytic function of k everywhere except for the cut (13), provided $\nu_{-1}(\tau_2)<\infty$. The sum dependent on s is

$$\Omega(\lambda,k,s)=\Gamma(\lambda+\tfrac{1}{2})(-2ik)^{\lambda-1/2}\sum_{l=1}^{\lambda+1/2}\frac{1}{(\lambda+\frac{1}{2}-l)!}\times\left\{\frac{d^{\lambda+1/2-l}}{dz^{\lambda+1/2-l}}(z-2ik)^{-\lambda-1/2}\Big[(s+ik)^{-l}+\sum_{n=1}^{\infty}(-g)^n\int_0^\infty dz_1\int d\mu(t_1)\cdots\int_0^\infty dz_n\times\int d\mu(t_n)(s+u_n+v_n+ik)^{-l}\times\prod_{p=1}^{n}A^{\lambda-1/2}(z+u_p+v_{p-1}-u_n-v_n-ik,k)\times A^{-\lambda-1/2}(z+u_{p-1}+v_{p-1}-u_n-v_n-ik,k)\Big]\right\}_{z=0},\quad (22a)$$

so that

$$\omega(\lambda,k,s)=\frac{\Gamma(2\lambda+1)}{\Gamma(\lambda+1/2)}[(-2ik)^{-\lambda+1/2}F(\lambda,-k)\times\Omega(\lambda,k,s)+(2ik)^{-\lambda+1/2}F(\lambda,k)\times\Omega(\lambda,-k,s)].\quad (23a)$$

Inverting the Laplace transformation (5b) and expressing the derivatives at $z=0$ by means of the Cauchy integral taken around a closed contour C in a tight neighborhood of $z=0$, we have

$$\varphi(\lambda,k,r)=\frac{\Gamma(2\lambda+1)}{2ik\Gamma(\lambda+1/2)}[(2ik)^{-\lambda+1/2}F(\lambda,k)\times f(\lambda,-k,r)-(-2ik)^{-\lambda+1/2}F(\lambda,-k)\times f(\lambda,k,r)],\quad (23b)$$

where the Jost solution is given by the Born series

$$f_+(\lambda,k,r)\equiv f(\lambda,k,r)=(-2ik)^{\lambda+1/2}r^{-\lambda+1/2}\frac{\Gamma(\lambda+1/2)}{2\pi i}\int_C ds e^{(s-ik)r}\times A^{-\lambda-1/2}(s-ik,k)\Big\{1+\sum_{n=1}^{\infty}(-g)^n$$

$$\times \int_0^\infty dz_1 \int d\mu(t_1) \cdots \int_0^\infty dz_n \int d\mu(t_n)$$

$$\times \exp[-(u_n + v_n)r] \prod_{p=1}^{n} A^{\lambda - 1/2}(s - u_{p-1} - v_p - ik,k) A^{-\lambda - 1/2}(s - u_p - v_p - ik,k)\Big\}$$

$$[f_-(\lambda,k,r) = f(\lambda, -k,r)]. \tag{22b}$$

It converges absolutely and uniformly except for the upper cut (13) under the assumption $\nu_{-1}(\tau_2) < \infty$. Integrating by parts with respect to z_n or s we easily verify again that (22b) satisfies Eq. (3).

We stress, in this place, that the Jost solutions $f_\pm(\lambda,k,r)$ as well as the functions $F_\pm(\lambda,k)$ are determined only by the functions $f(\lambda,k,r)$ and $F(\lambda,k)$, respectively, taken in the points $\pm k$, where arg$k$ is not specified in any way. This statement applies to the Jost solutions and the Jost functions of Sec. 3 too. (This follows naturally from the derivation, which makes use of the fact that λ are half-integers. For complex λ, the point $k = 0$ becomes an additional singular point of branch point type of the Jost solution and of the Jost function, so that their sheets in the k plane and, especially, the quantity $-k$, must be properly defined.)

A. The Jost solutions

Substituting the expression (22b) for $\psi(\lambda,k,r)$ into Eq. (3), and carrying out some elementary transformations, we easily find that the Born series (22b) is a solution of Eq. (3) not only for a half-integer λ but also for a general complex λ of either sign. At the same time, owing to the complex power $s^{-\lambda - 1/2}$ in the factor $A^{-\lambda - 1/2}(s - ik,k)$, the closed contour C must be replaced by such an open one that the integrands in (22b) vanish in its initial and final point and that it encircles the point $s = 0$ in the anticlockwise direction 0^+. The asymptotic behavior of (22b) for $r \to \infty$ is given by the first term of the Born series ($g = 0$). Thus this term must conform with the function $f_0(\lambda,k,r)$ defined in the boundary condition (4b), if the expression (22b) is to represent the Jost solution. However, according to the integral representation of Hankel functions,[6]

$$f_0(\lambda,k,r) = \exp\left(-i\frac{\pi}{2}(\lambda + \tfrac{1}{2})\right)\left(\frac{\pi kr}{2}\right)^{1/2} H_\lambda^{(2)}(kr)$$

$$= \pi^{-1} \exp\left(i\frac{\pi}{2}(\lambda - \tfrac{3}{2})\right)\Gamma(\tfrac{1}{2} - \lambda)\left(\frac{kr}{2}\right)^{1/2}\left(\frac{r}{2k}\right)^{\lambda} \times \int_{\infty \exp(-i\pi + i\beta)}^{0} ds\, e^{(s - ik)r} s^{\lambda - 1/2}(s - 2ik)^{\lambda - 1/2}$$

$$= f_0(-\lambda,k,r)$$

$$= \pi^{-1} \exp\left(-i\frac{\pi}{2}(\lambda + \tfrac{3}{2})\right)\Gamma(\tfrac{1}{2} + \lambda)\left(\frac{kr}{r}\right)^{1/2}\left(\frac{2k}{r}\right)^{\lambda} \times \int_{\infty \exp(-i\pi + i\beta)}^{0} ds\, e^{(s - ik)r} s^{-\lambda - 1/2}(s - 2ik)^{-\lambda - 1/2}$$

$$\text{for } |\beta| < \pi/2 \text{ and } -\tfrac{3}{2}\pi + \beta < \arg k < \tfrac{1}{2}\pi + \beta, \tag{24}$$

so that the new contour must be identified with that given in Eqs. (24). In what follows we shall denote it by C_β. It goes from $\infty \exp(-i\pi + i\beta)$ along a line of constant phase to 0, encircles 0 in an anticlockwise direction and goes back along a line of constant phase to $\infty \exp(i\pi + i\beta)$. Changing β we achieve an extension of the analyticity region of k. By replacing the contour C in Eq. (22b) by C_β we get an analytic continuation of the Jost solution $f_+(\lambda,k,r) = f(\lambda,k,r) = f(-\lambda,k,r)$ in the complex λ plane based, as usual, on the analytic continuation of the Schrödinger equation (3) and given by an explicit expression. Its virtue consists in the possibility of the alternative sign choice of the complex variable λ.

In view of the exponential factors $\exp(-(u_n + v_n)r)$ this expression converges absolutely and uniformly for $r > 0$ and any complex λ provided $\nu_{-1}(\tau_2) < \infty$. Therefore, it represents an analytic function of k everywhere except for possible poles or branch points at $k = 0$ and except for the regions of k defined by zeros of A factors in (22b). These regions change by changing β in the allowed bounds ($|\beta| < \pi/2$), but they intersect in the upper cut (13), so that formula (22b) defines the "abstract" Jost solution as an analytic function of k except for the natural upper cut (13) and, possibly, $k = 0$. According to the relation (24) the integral representation of the function $f_0(\lambda,e^{-i\pi}k,r)$ with $-\pi < \arg k < 2\pi$ is identical with that for the second function in the boundary condition (4b).

Hence, we define the Jost solution $f_-(\lambda,k,r)$ as $f_-(\lambda,k,r) = f(\lambda,e^{-i\pi}k,r)$ [with the induced lower cut (13)]. We see at the same time, that the quantity $-k$ is defined as $e^{-i\pi}k$ in the case of a general λ.

In some cases the integration contour C_β can be replaced by the line $[0, \infty e^{i\beta})$, of constant phase

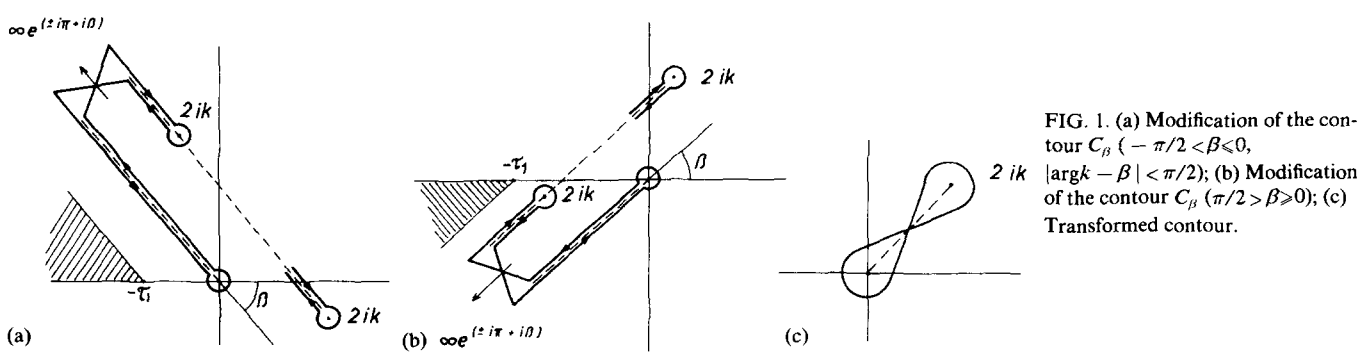


FIG. 1. (a) Modification of the contour C_β ($-\pi/2 < \beta \leqslant 0$, $|\arg k - \beta| < \pi/2$); (b) Modification of the contour C_β ($\pi/2 > \beta \geqslant 0$); (c) Transformed contour.

$$\int_{\infty \exp(-i\pi + i\beta)}^{0^+} ds\, s^{-A} f(s) = 2i \sin(\pi A) \int_0^{\infty \exp(i\beta)} dt\, t^{-A} f(-t) \quad \text{for } \mathrm{Re}(1-A) > 0, \tag{25}$$

and for suitable functions $f(s)$. It is the standard modification used in the integral representations of the Hankel as well as the Γ functions.[7]

For any $k \neq 0$ with $|\arg k| < \pi$ under the suitable choice of β, i.e., $-\pi/2 + \arg k < \beta < \arg k + \pi/2$, one can always close the integration contour C_β in the expression (22b) into the closed loop depicted for the four cases in Figs. 1(a) and 1(b) and replace it by that depicted in Fig. 1(c), which depends on the location of k only. The dashed regions are those ones of $2ik$, where possible singularities of the used integral representation of the Jost solution may exist. It is, again, the same modification, which is carried out in the integral representations of Bessel and Hankel functions.[6] Since Jost solutions $f(\lambda,k,r)$ are even functions of λ, it results in

$$\begin{aligned} f(\lambda,k,r) - \exp[i\pi(\pm\lambda + \tfrac{1}{2})] f(\lambda, e^{-i\pi}k, r) &= \chi(\pm\lambda,k,r) \\ \equiv \exp\left(-i\frac{\pi}{2}(\pm\lambda + \tfrac{1}{2})\right)(2k)^{\pm\lambda + 1/2} r^{\mp\lambda + 1/2} &\times \frac{\Gamma(\pm\lambda + 1/2)}{2\pi i} \\ \times \int_L ds\, e^{sr} A^{\mp\lambda - 1/2}(s,k) \Big(1 + \sum_{n=1}^{\infty} (-g)^n &\times \int_0^\infty dz_1 \int d\mu(t_1) \cdots \int_0^\infty dz_n \int d\mu(t_n) \\ \times \exp[-(u_n + v_n)r] \prod_{p=1}^{n} A^{\pm\lambda - 1/2}(s - u_{p-1} &- v_p, k) A^{\mp\lambda - 1/2}(s - u_p - v_p, k)\Big), \end{aligned} \tag{26}$$

where the integration loop L has the form given on Fig. 1(c), i.e., it encircles the point $-ik$ in an anticlockwise direction and the point ik in a clockwise direction. The relation (26) becomes the well-known relation between Bessel and Hankel functions[6] for $g = 0$, considering that the first term of the right-hand side (26) equals $\exp[i(\pi/2)(\pm\lambda - 1/2)] \times (2\pi kr)^{1/2} J_{\mp\lambda}(kr)$, according to Ref. 6. As the loop L is closed, we easily find out by direct substitution in Eq. (3) too, that the functions $\chi(\pm\lambda,k,r)$ defined by Eq. (26) are its solutions. Taking $s = ikx$ in their integral representation (26) we see that

$$\chi(\pm\lambda, e^{-i\pi}k, r) = -\exp[i\pi(\pm\lambda + \tfrac{1}{2})]\chi(\pm\lambda,k,r) \quad \text{for } |k| < \tau_1/2. \tag{27}$$

The double application of this relation to the inverted relation (26) (with a noninteger λ)

$$f(\lambda,k,r) = (2i \sin\pi\lambda)^{-1}[e^{i\pi\lambda}\chi(-\lambda,k,r) - e^{-i\pi\lambda}\chi(\lambda,k,r)] \tag{28}$$

gives the well-known circuit relation for $f(\lambda,k,r)$.[3,1] We shall not, however, give it nor will we discuss the many other known and elementary relations and properties of Jost solutions,[1-4] which follow from the explicit expressions here. We remark only that the integration contour C_β can also be extended to an open and nonintersecting loop unlike that depicted in Figs. 1(a) – (c) and thus obtain other relations.

B. The Jost functions

When the analytic continuation of the normalized Jost function (21) to the complex region of λ is to be found, it is better to proceed from its alternative definition[1-4]

$$\begin{aligned} f_0(\lambda,k)F(\lambda,k) &= \lim_{r\to 0^+} 2\lambda r^{\lambda - 1/2} f(\lambda,k,r) \\ &= \lim_{r\to 0^+} 2\lambda r^{\lambda - 1/2} f(-\lambda,k,r), \end{aligned} \tag{29a}$$

where the kinematic factor

$$\begin{aligned} f_0(\lambda,k) &= \lim_{r\to 0^+} 2\lambda r^{\lambda - 1/2} f_0(\pm\lambda,k,r) \\ &= \pi^{-1/2} \exp[-i(\pi/2)(\lambda - \tfrac{1}{2})] \\ &\quad \times 2^{\lambda + 1/2} \Gamma(1+\lambda) k^{-\lambda + 1/2} \\ &= \frac{\Gamma(2\lambda+1)}{\Gamma(\lambda + \frac{1}{2})} (2ik)^{-\lambda + 1/2}. \end{aligned} \tag{29b}$$

We shall see that it is indeed identical to the expression (21) for a half-integer λ. In view of this definition we are also not obliged to distinguish between functions $F_+(\lambda,k)$ and $F_-(\lambda,k)$, since evidently $F_+(\lambda,k) = F(\lambda,k)$ and $F_-(\lambda,k) = F(\lambda, e^{-i\pi}k)$ with the similar context as for the Jost solutions.

When the expression (22b) is substituted for $f(\lambda,k,r)$ in (29a) the first limit is easily obtained. The corresponding Born series represents an (multivalued) analytic function of k except for $k = 0$ and the upper cut (13) and a function of λ, which is analytic everywhere (for any regular k) provided $\nu_{-1}(\tau_2) < \infty$. The modification (25) can be used for $\frac{1}{2} > \mathrm{Re}\lambda > 0$. Before carrying out the second limit we note that the simultaneous sign change of the first arguments of A factors does not change the expression in the brace of Eq. (22b). However,

$$\lim_{r\to 0^+} r^{2\lambda} \int_0^\infty dz_n\, A^{\lambda - 1/2}(-s + v_n + u_{n-1} + z_n + ik,k)\, e^{-z_n r} = \Gamma(2\lambda), \quad \text{for } \mathrm{Re}\lambda > 0, \tag{30}$$

so that successively under the use of Eq. (19a)

$$\begin{aligned} F(\lambda,k) &= 1 + (2e^{-i\pi/2}k)^{-\lambda + 1/2} \Gamma(-\lambda + \tfrac{1}{2}) f_0^{-1}(\lambda,k) \\ &\quad \times \frac{1}{2\pi i} \int_{C_\beta} ds - g \int_{\tau_1}^{\tau_2} d\mu(t)\, A^{\lambda - 1/2}(s - ik,k) \\ &\quad \times \omega(\lambda,k,-s+t+ik) \\ &= 1 + e^{i\pi(\lambda - 1/2)}(2i\cos\pi\lambda)^{-1} \\ &\quad \times \int_{C_\beta} ds A^{\lambda - 1/2}(s - ik,k) \\ &\quad \times \sum_{n=1}^{\infty} (-g)^n \int_0^\infty dz_1 \int d\mu(t_1) \cdots \\ &\quad \times \int_0^\infty dz_{n-1} \int d\mu(t_{n-1}) \int d\mu(t_n) \\ &\quad \times A^{-\lambda - 1/2}(-s + v_1 + ik,k) \end{aligned}$$

$$\times \prod_{p=1}^{n-1} A^{\lambda-1/2}(-s+u_p+v_p+ik,k)$$
$$\times A^{-\lambda-1/2}(-s+u_p+v_{p+1}+ik,k)$$
$$= 1 - i\frac{\pi}{2}\left(\frac{k}{2}\right)^{\lambda}\Gamma^{-1}(\lambda+1)$$
$$\times \int_0^{\infty} dr\, r^{1/2}V(r)H_{\lambda}^{(2)}(kr)\varphi(\lambda,k,r), \qquad (31)$$

where passing from the first expression to the last one we apply relations (5a), (24), (29b), and (1). It is the well-known formula[1,3] valid for general potentials $V(r)$ and applicable as soon as the written integral exists. According to the transformation (25) the second expression with $\beta = 0$ turns into the expression (21), which holds now for any complex λ with $\mathrm{Re}\lambda > -\frac{1}{2}$ and k with $-\frac{3}{2}\pi < \arg k < \pi/2$. The series in Eq. (31) converges absolutely and uniformly in its regular points k, if $\nu_{-1}(\tau_2) < \infty$ for $\mathrm{Re}\lambda \geqslant \frac{1}{2}$ and $\nu_0(\tau_2) < \infty$ for $-\frac{1}{2} < \mathrm{Re}\lambda < \frac{1}{2}$. It can be analytically continued even to the regions with $\mathrm{Re}\lambda < 0$, when the function $\mu(t)$ belongs to the one of the classes κ introduced in Sec. 2. Then the series in Eq. (31) converges absolutely and uniformly in its regular points k, as soon as $\mathrm{Re}\lambda > -\frac{1}{2}(\kappa+1)$. Thus it represents an analytic function of k except for the upper cuts (13) and $k = 0$, respectively, and an analytic function of λ in the half-plane $\mathrm{Re}\lambda > -\frac{1}{2}(\kappa+1)$ except for the possible simple poles at $\lambda = -\frac{1}{2}, -\frac{3}{2}, \ldots$. Using l'Hospital's rule we easily find out that poles $-\frac{1}{2}, -\frac{3}{2}, \ldots, -\frac{1}{2} - [q/2]$ are not realized, as soon as the function $\mu(t)$ belongs to the subclass q of Eq. (7b) as well. The detailed proofs of these particular statements are based on the inequalities of the type (20) and are carried out along the same lines as in Sec. 3 or as in the beginning of this section.

If we know, now, the analytic continuation of both the Jost solutions and the Jost functions, we can construct the analytic continuation of the regular solution $\varphi(\lambda,k,r)$ to the region of the complex values of λ. Additional poles, however, appear then at the negative integers owing to the factor $\Gamma(\lambda+1)$ in Eq. (23b).

Limit (29a) can be carried out in the relations (26) and (28) too. In Eq. (28) it leads to the circuit relation of the Jost function under the use of formula (27). In the minus alternative of Eq. (26) we get

$$F(\lambda,k) - F(\lambda,e^{-i\pi}k)$$
$$= -g\frac{\exp[i\pi(\lambda-\frac{1}{2})]}{2i\Gamma(2\lambda+1)\cos\pi\lambda}\int_{\tau_1}^{\tau_2} d\mu(t)$$
$$\times \int_L ds\, A^{\lambda-1/2}(s,k)\omega(\lambda,k,-s+t)$$
$$= -\frac{i\pi}{\Gamma(\lambda+1)}\left(\frac{k}{2}\right)^{\lambda}$$
$$\times \int_0^{\infty} dr\, r^{1/2}V(r)J_{\lambda}(kr)\varphi(\lambda,k,r), \qquad (32)$$

according to the formulas (29), (30), (19a), (5a), and (1).

C. Behavior of the Jost functions as $|\lambda| \to \infty$ ($\mathrm{Im}k = 0$)

Since the exponential factor $\exp(sr)$ is absent in the first and second expressions (31) for the normalized Jost function the integration contour C_{β} can be deformed rather arbitrarily. As k is taken real and positive, the choice $\beta = \frac{1}{2}\pi$ simplifies the investigation of the asymptotic behavior of $F(\lambda,k) = F^*(\lambda^*,-k)$ with respect to λ. Its dependence on λ is given by means of the factors

$$\left|A(s-u_p-v_p-ik,k)A^{-1}(s-u_p-v_{p+1}-ik,k)\right|^{\lambda}, \qquad (33a)$$

$$\exp\{i\lambda[\arg A(s-u_p-v_p-ik,k) - \arg A(s-u_p-v_{p+1}-ik,k)]\} \qquad (p=0,\ldots n). \qquad (33b)$$

The absolute value in (33a) is less than unity if the contour $C_{(1/2)\pi}$ encircles the origin $s = 0$ in the close neighborhood ($|\mathrm{Re}s| < \frac{1}{2}\tau_1$). Hence $F(\lambda,k) = 1$ as $\mathrm{Re}\lambda \to \infty$ and $\mathrm{Im}\lambda = \mathrm{const}$, according to the (second) expression (31) and (33a).

The phases of the expression (33b) satisfy the relations

$$0 \leqslant \arg A(s-u_p-v_{p+1}-ik,k)$$
$$\leqslant \arg A(s-u_p-v_p-ik,k)$$
$$\leqslant \arg A(s-u_{p-1}-v_p-ik,k)$$
$$\leqslant \pi \quad \text{for } p = 1,\ldots,n,$$
$$\arg A(s-v_1-ik,k) \leqslant \arg A(s-ik,k) + 2\pi,$$
$$|\arg A(s-ik,k)| \leqslant \pi, \qquad (34)$$

where the equality applies in the isolated integration points only. Thus $F(\lambda,k) = 1$ as $\mathrm{Re}\lambda = \mathrm{const}$ and $\mathrm{Im}\lambda \to \infty$ according to Eq. (31) too.

The sum (series) in the second expression (31) converges absolutely and each of its terms tends to zero according to the relations (34) as $\mathrm{Re}\lambda = \mathrm{const}$ and $\mathrm{Im}\lambda \to -\infty$. Because of the inequalities

$$0 \leqslant \arg A(s-u_{n-1}-v_n-ik,k)$$
$$+ \sum_{p=1}^{n-1}[\arg A(s-u_{p-1}-v_p-ik,k)$$
$$- \arg A(s-u_p-v_p-ik,k)]$$
$$\leqslant \arg A(s-v_1-ik,k), \qquad (35)$$

its order with respect to $\mathrm{Im}\lambda \to -\infty$ can be determined by its first term only, so that

$$F(\lambda,k) = 1 + O\left(\left|\int_0^{\infty} dr\, rV(r)H_{\lambda}^{(2)}(kr)J_{\lambda}(kr)\right|\right), \qquad (36)$$

according to Eqs. (31) and (23b) for this limit. But this relation, in which $p(\lambda) = O(q(\lambda))$ means that $|p(\lambda)| = \mathrm{const}\,q(\lambda)$ as $|\lambda| \to \infty$, determines the asymptotic behavior of $F(\lambda,k)$ in all the (three) cases investigated as follows from the preceding considerations.

Replacing the Hankel function by the known combination of the Bessel functions and using the integration formulas of Ref. 7 for products of the Bessel and exponential functions we have

$$\int_0^{\infty} dr \int_{\tau_1}^{\tau_2} d\mu(t)e^{-itr}(i\sin\pi\lambda)^{-1}[e^{i\pi\lambda}J_{\lambda}(kr)$$
$$- J_{-\lambda}(kr)]J_{\lambda}(kr)$$
$$= \frac{\exp(i\pi\lambda)}{i\pi k\sin\pi\lambda}\int_{\tau_1}^{\tau_2} d\mu(t)\left[Q_{\lambda-1/2}\left(1+\frac{t^2}{2k^2}\right)\right.$$

$$- k \int_0^{\pi} d\varphi \frac{\cos\lambda\varphi}{(t^2 + 2k^2 + 2k^2\cos\varphi)^{1/2}} \Big], \quad (37)$$

where $Q_{\lambda - 1/2}(z)$ is the Legendre function of the second kind. Its asymptotic behavior as $|\lambda| \to \infty$ follows from the formulas given in Ref. 7. The second term in (37) is asymptotically constλ^{-1}. In such a way the order terms in Eq. (36) can be estimated (see also Ref. 3) for any asymptotic values of λ, where either $|\arg\lambda| < \frac{1}{2}\pi$ or $\mathrm{Re}\lambda > -\frac{1}{2}$ and $\mathrm{Im}\lambda \to \pm\infty$. As $F(\lambda,k) = 1$ in these cases, the Sommerfeld–Watson transformation can be performed. The integral along the imaginary λ axis in this transformation can be pushed further to the left[3] if the function $\mu(t)$ belongs to the subclass (7b) as the relation (37) then holds for $\mathrm{Re}\lambda > -(q+1)/2$ too.

ACKNOWLEDGMENTS

I would like to thank Dr. P. Kolář and Dr. J. Fischer for helpful discussions and for reading the manuscript.

[1] V.De Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965).
[2] K. Chadan and P.C. Sabatier, *Inverse Problems in Quantum Scattering* (Springer, New York, Berlin, 1977).
[3] R.G. Newton, *The Complex j-Plane* (Benjamin, New York, 1964).
[4] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[5] D.I. Fivel and A. Klein, J. Math. Phys. **1**, 274 (1960).
[6] G.N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., New York, 1945).
[7] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vol. 1.

Exact eigenvalues, eigenfunctions, matrix elements and phase-shifts for a particle of angular momentum l in a particular screened potential

Ulla Myhrman
Institute of Theoretical Physics, University of Uppsala, Thunbergsvägen 3, S-752 38 Uppsala, Sweden

(Received 28 December 1979; accepted for publication 22 February 1980)

The radial Schrödinger equation is solved for an effective potential that can be considered as a generalization of a potential suggested by Hylleraas and Risberg and by Hulthén. For the scattering problem the phase-shifts are deduced. For the bound state problem the energy eigenvalues and the normalized eigenfunctions are derived. Closed-form expressions for certain matrix elements are calculated, and for the particular case of expectation values a recurrence formula is derived. In an appendix Levinson's theorem is discussed for the particular potential under consideration.

1. INTRODUCTION

In studying the nature of the interaction of elementary particles, Yukawa[1] suggested as an interaction potential between a neutron and a proton the expression $C\exp(-\nu r)/r$, where C and ν are constants. The same potential was used by Möller and Rosenfeld[2] in their investigation of the stationary states of a deuteron. Also, N. Bohr[3] used a potential of this form when examining the scattering of charged particles in atomic fields. The corresponding radial Schrödinger equation for s states, i.e., for angular momentum $l = 0$, was solved numerically by Wilson[4] and by Sachs and Goeppert-Mayer.[5] As an approximation of the interaction potential, Hylleraas and Risberg[6] and Hulthén[7] suggested the expression $C\exp(-\nu r)/[1-\exp(-\nu r)]$, for which the Schrödinger equation was solved analytically for the ground state by Hylleraas and Risberg and for any s state by Hulthén. The usefulness of this potential for calculations of scattering and bound states in atomic potentials was emphasized by Lindhard and Winther[8]. A modification of the above-mentioned approximate potential was used by Gustavi.[9] As an approximation of an alternative two-nucleon interaction potential $C\exp(-\nu r)/(\nu r)^2$, Gustavi used the expression $C\exp(-\nu r)/[1-\exp(-\nu r)]^2$, for which he discussed the solution of the radial Schrödinger equation for s states.

A linear combination of the previously mentioned approximate potentials, $-\lambda\exp(-\nu r)/[1-\exp(-\nu r)] + \mu\exp(-\nu r)/[1-\exp(-\nu r)]^2$, where λ and ν are positive constants and $\mu \geqslant 0$ [cf. Eqs. (4a), (4b)], was treated as an effective potential by Rosenzweig and Krieger[10] in their investigation of exact Jeffreys–Wentzel–Kramers–Brillouin (JWKB) quantization conditions by means of a method developed by N. Fröman and P. O. Fröman.[11] With this method the exact energy eigenvalues were calculated in Ref. 10. By the use of the factorization method, Infeld and Hull[12] have calculated the same energy eigenvalues.

The present author will use the last-mentioned potential as a effective potential in the radial Schrödinger equation. This potential can, for small νr and for fixed angular momentum l, be considered as a generalization of the potential suggested in Refs. 6 and 7, from the case of angular momentum $l = 0$ to the case of a general angular momentum l. It should, however, be emphasized that this interpretation implies an l dependence of the physical potential. This will be further discussed in Sec. 2. The phase shifts for this potential are calculated. The energy eigenvalues and the normalized eigenfunctions are derived, as well as closed-form expressions for matrix elements of products of powers of $\exp(-\nu r)$ and $1-\exp(-\nu r)$. For the particular case of the expectation values, a derivation is given of a recurrence formula which, in the limit when the parameter ν tends to zero, transforms into Kramers' recurrence formula (see Sec. 7) for the expectation values of powers of r for the hydrogen atom.

In an abstract of a paper contributed to an American Physical Society Meeting but apparently never published, Manning and Rosen[13] gave the eigenvalues and outlined expressions for the nonnormalized eigenfunctions for a potential equivalent to the one treated in the present paper. Since Manning and Rosen published their results without any derivation, and since Hulthén treated only a special case of the potential and also used a different method, the author thought it worthwhile to publish the present derivation of the eigenvalues together with the phase shifts, the normalized eigenfunctions, and the results concerning the above-mentioned matrix elements.

2. THE POTENTIAL

Consider the radial Schrödinger equation

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + V(r) + \frac{l(l+1)\hbar^2}{2mr^2}\right)u = Eu, \tag{1}$$

where we use standard notations. We shall solve this equation exactly for the case when the effective potential is (cf. Ref. 13 and 14)

$$V_{\text{eff}}(r) = V(r) + \frac{l(l+1)\hbar^2}{2mr^2} = -\frac{\lambda}{e^{\nu r}-1} + \frac{\mu e^{\nu r}}{(e^{\nu r}-1)^2} = -\frac{\lambda-\mu}{e^{\nu r}-1} + \frac{\mu}{(e^{\nu r}-1)^2}. \tag{2}$$

In Fig. 1 the effective potential (2) is plotted for two different choices of the real constants λ, μ, and ν.

When $\nu r \ll 1$, Eq. (2) can be written

 0022-2488/80/071732-08$1.00

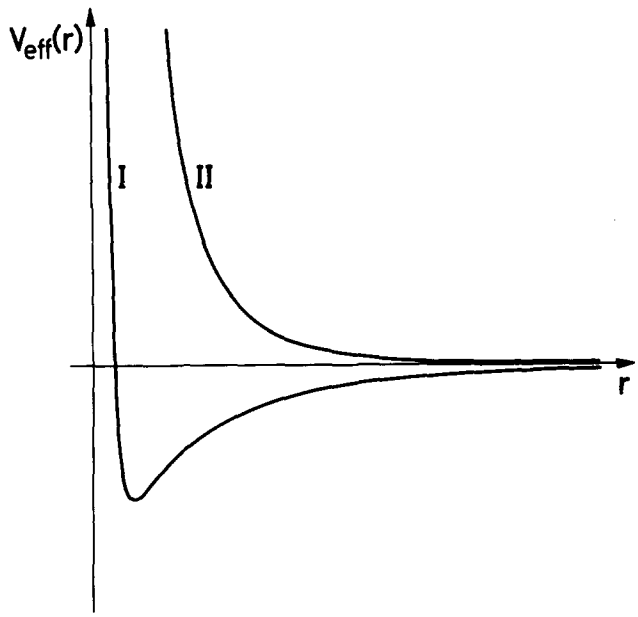


FIG. 1. Effective potential $V_{\text{eff}}(r) = -(\lambda-\mu)/[\exp(\nu r)-1] + \mu/[\exp(\nu r)-1]^2$ plotted for two different choices of the real constants λ, μ, and ν. In both cases $\nu>0$. The curve indicated by I is valid for $\lambda>\mu\geqslant 0$ and curve II is valid for $\mu>\lambda>0$.

$$V_{\text{eff}}(r) = V(r) + \frac{l(l+1)\hbar^2}{2mr^2} = \frac{\mu}{\nu^2 r^2} - \frac{\lambda}{\nu r} + \frac{6\lambda-\mu}{12} + \cdots. \tag{3}$$

If $V(r)$ represents an attractive screened Coulomb potential, which tends to Ze^2/r as $\nu\to 0$, Eq. (3) gives

$$\lambda = \nu Ze^2, \tag{4a}$$

$$\mu = \nu^2 \frac{l(l+1)\hbar^2}{2m}, \tag{4b}$$

where Ze is the charge of the nucleon. For physical reasons we choose the constants ν and l such that $\nu>0$ and $l\geqslant 0$ which, according to (4a) and (4b), implies that $\lambda>0$ and $\mu\geqslant 0$.

Using (4a) and (4b), we can write (2) as

$$V(r) = -\frac{\nu Ze^2}{e^{\nu r}-1} + \frac{\nu^2 l(l+1)\hbar^2 e^{\nu r}}{2m(e^{\nu r}-1)^2} - \frac{l(l+1)\hbar^2}{2mr^2} \tag{2'}$$

and (3) as

$$V(r) = -\frac{Ze^2}{r} + \frac{1}{12}\left(6\nu Ze^2 - \frac{\nu^2 l(l+1)\hbar^2}{2m}\right) + \cdots. \tag{3'}$$

Thus it is seen that the physical potential $V(r)$ is l dependent. If r is finite and the parameter $\nu\to 0$, we get from (3') the potential of a hydrogenlike ion.

In the following, the constants λ and μ are chosen according to (4a) and (4b).

3. SOLUTION OF THE RADIAL SCHRÖDINGER EQUATION

We introduce the new constants

$$a = 2mE/\hbar^2\nu^2, \tag{5a}$$

$$b = 2m(\lambda-\mu)/\hbar^2\nu^2 = 2Z/\nu a_0 - l(l+1), \tag{5b}$$

where $a_0 = \hbar^2/(me^2)$ is the Bohr radius, and the new variable

$$z = \nu r. \tag{6}$$

The radial Schrödinger equation (1), with $V(r) + l(l+1)/(2mr^2)$ given by (2), can, by means of (4b), (5a), (5b), and (6), be written

$$\frac{d^2u}{dz^2} + \left(a + \frac{b}{e^z-1} - \frac{l(l+1)}{(e^z-1)^2}\right)u = 0. \tag{7}$$

For $l=0$ this equation is simplified into Eq. (5.4) of Ref. 6 and Eq. (6) of Ref. 7.

Defining

$$\xi = e^{-z}, \tag{8}$$

we can write (7) as

$$\frac{d^2u}{d\xi^2} + \frac{1}{\xi}\frac{du}{d\xi} + \left(\frac{a}{\xi^2} + \frac{b}{\xi(1-\xi)} - \frac{l(l+1)}{(1-\xi)^2}\right)u = 0. \tag{9}$$

To solve this differential equation we shall use the method described in Ref. 14 (cf. also Sec. 10.2 of Ref. 15). Thus we insert

$$u = \xi^{-1/2}v \tag{10}$$

into (9), getting

$$\frac{d^2v}{d\xi^2} + \frac{1}{4}\left(\frac{1+4a}{\xi^2} + \frac{4b}{\xi(1-\xi)} - \frac{4l(l+1)}{(1-\xi)^2}\right)v = 0. \tag{11}$$

Defining new constants p, q, and s as

$$p = -2\sqrt{-a} = -2(-2mE)^{1/2}/\hbar\nu, \tag{12a}$$

$$q = -(2l+1), \tag{12b}$$

$$s = 2[-a+b+l(l+1)]^{1/2} = 2[2m(\lambda-E)]^{1/2}/\hbar\nu, \tag{12c}$$

we can write (11) in the form

$$\frac{d^2v}{d\xi^2} + \frac{1}{4}\left(\frac{1-p^2}{\xi^2} + \frac{1+s^2-p^2-q^2}{\xi(1-\xi)} + \frac{1-q^2}{(1-\xi)^2}\right)v = 0. \tag{13}$$

The solutions of (13) are generalized hypergeometric functions. In order to obtain these solutions in the form of Gauss' hypergeometric functions, we introduce a new dependent variable y defined by

$$v = \xi^{(1-p)/2}(\xi-1)^{(1-q)/2}y. \tag{14}$$

Inserting (14) into (13), we get

$$\xi(1-\xi)\frac{d^2y}{d\xi^2} + [(1-p) - (1-p-q+1)\xi]\frac{dy}{d\xi} - [\tfrac{1}{2}(1-p)(1-q) - \tfrac{1}{4}(1+s^2-p^2-q^2)]y = 0. \tag{15}$$

Defining

$$\alpha = \tfrac{1}{2}(1-p-q) + \tfrac{1}{2}s, \tag{16a}$$

$$\beta = \tfrac{1}{2}(1-p-q) - \tfrac{1}{2}s, \tag{16b}$$

$$\gamma = 1-p, \tag{16c}$$

we can write Eq. (15) as

$$\xi(1-\xi)\frac{d^2y}{d\xi^2} + [\gamma - (\alpha+\beta+1)\xi]\frac{dy}{d\xi} - \alpha\beta y = 0. \tag{17}$$

This is the hypergeometric differential equation, the solu-

tions of which are given for the nondegenerate case in Sec. 2.3.1 of Ref. 16 and for the degenerate case in Sec. 2.2.2 of Ref. 16.

Thus, according to (10), (14), and (8), the solution of the radial Schrödinger equation (1), with (2) inserted, can be written

$$u = e^{pz/2}(e^{-z}-1)^{(1-q)/2}y$$
$$= e^{-z(-a)^{1/2}}(e^{-z}-1)^{l+1}y, \tag{18}$$

where (12a) and (12b) have been used to obtain the last member. When y is a general solution to Eq. (17), we get the general solution of (1) from Eq. (18). To find the physically acceptable solution of (1) we have to impose the condition

$$u(0) = 0. \tag{19}$$

The solution of (17), for which (18) fulfills the condition (19), can be written

$$y = NF(\alpha,\beta;\ \alpha+\beta+1-\gamma;\ 1-e^{-z}), \tag{20}$$

where N is a constant factor and $F(\alpha,\beta;\ \alpha+\beta+1-\gamma;\ 1-e^{-z})$is the hypergeometric function defined in Sec. 2.1 of Ref. 16. The wavefunction satisfying Eqs. (1), (2), and (19) is thus given by (18) and (20) as

$$u = Ne^{-z(-a)^{1/2}}(e^{-z}-1)^{l+1}$$
$$\times F(\alpha,\beta;\ \alpha+\beta+1-\gamma;\ 1-e^{-z}), \tag{21}$$

where N is to be chosen conveniently. Using relations between Kummer's solutions of the hyprgeometric equation given by Eqs. (35), (1), (5), and (17) in Sec. 2.9 of Ref. 16, we can rewrite (21) as

$$u = N\frac{\Gamma(\alpha+\beta+1-\gamma)\Gamma(1-\gamma)}{\Gamma(\beta+1-\gamma)\Gamma(\alpha+1-\gamma)}e^{-z(-a)^{1/2}}$$
$$\times(e^{-z}-1)^{l+1}F(\alpha,\beta;\gamma;e^{-z})$$
$$+N\frac{\Gamma(\alpha+\beta+1-\gamma)\Gamma(\gamma-1)}{\Gamma(\alpha)\Gamma(\beta)}e^{z(-a)^{1/2}}$$
$$\times(e^{-z}-1)^{l+1}F(\alpha+1-\gamma,\beta+1-\gamma;2-\gamma;e^{-z}). \tag{22}$$

4. THE SCATTERING PROBLEM

In this section we shall be dealing with energies such that $E>0$, which according to (5a) implies that $a>0$. Choosing $\sqrt{-a} = i\sqrt{a}$, letting $z\to+\infty$ in (22), and using the fact that both the hypergeometric functions in Eq. (22) tend to unity when $z\to+\infty$, we obtain

$$u \sim N\frac{\Gamma(\alpha+\beta+1-\gamma)\Gamma(1-\gamma)}{\Gamma(\beta+1-\gamma)\Gamma(\alpha+1-\gamma)}(-1)^{l+1}e^{-iz(a)^{1/2}}$$
$$+N\frac{\Gamma(\alpha+\beta+1-\gamma)\Gamma(\gamma-1)}{\Gamma(\alpha)\Gamma(\beta)}(-1)^{l+1}e^{iz(a)^{1/2}}$$
$$=2N(-1)^{l+1}\Gamma(\alpha+\beta+1-\gamma)\left|\frac{\Gamma(\gamma-1)}{\Gamma(\alpha)\Gamma(\beta)}\right|$$
$$\times\cos\left(z\sqrt{a}+\arg\frac{\Gamma(\gamma-1)}{\Gamma(\alpha)\Gamma\beta)}\right),\quad z\to+\infty. \tag{23}$$

Thus we have obtained the correct asymptotic behavior for large positive values of z. From (23) we can deduce the phase shift δ_l as

$$\delta_l = \tfrac{1}{2}(l+1)\pi + \arg\frac{\Gamma(\gamma-1)}{\Gamma(\alpha)\Gamma(\beta)}, \tag{24}$$

where the argument shall be chosen such that $\delta_l\to0$ as $V(r)\to0$. For the special case $E=\lambda$, we get from (12c), (16a) and (16b) that $s=0$ and $\alpha=\beta$, and (24) can thus be simplified to

$$\delta_l = \tfrac{1}{2}(l+1)\pi + \arg\frac{\Gamma(\gamma-1)}{[\Gamma(\alpha)]^2}. \tag{24'}$$

The phase shift is further discussed in the Appendix.

We recall that, according to Sec. 2, different values of l correspond to different physical potentials.

The behavior of u for small z is easily obtained from (21) as

$$u\sim N(-1)^{l+1}z^{l+1},\quad z\to0. \tag{25}$$

5. THE BOUND-STATE PROBLEM

In the present section we are concerned with energies such that $E<0$, which according to (5a) implies that a is negative and hence that $\sqrt{-a}$ is a real parameter. This parameter shall be chosen to be positive.

In order to find the physically acceptable solution for the bound-state problem, we have to impose the boundary condition

$$u(r)\to0,\quad r\to\infty. \tag{26}$$

Letting $z\to+\infty$ in (22) and recalling that we have chosen $\sqrt{-a}$ to be positive, we realize that the solution fulfills the condition (26) when

$$\alpha = -(k-1), \tag{27}$$

where k is a positive integer. By the use of (16a) and (12a), (12b), and (12c), the quantization condition (27) can be written

$$\sqrt{-a} = \frac{1}{2}\left(\frac{b+l(l+1)}{k+l} - (k+l)\right), \tag{28}$$

i.e. [cf. Eqs. (4b) and (5b)],

$$\sqrt{-a} = \frac{1}{2}\left(\frac{2m\lambda}{(k+l)(\hbar\nu)^2} - (k+l)\right). \tag{28'}$$

An investigation of (21) for the special value $E=0$ shows that, when $z\to+\infty$, the condition (26) is not fulfilled, i.e., that $E=0$ is not a bound state. Hence $\sqrt{-a}$ shall be positive, and therefore it is seen from (28') that k can assume only those positive integers that fulfill the inequality

$$0<k<[2m\lambda]^{1/2}/\hbar\nu - l. \tag{29}$$

From (5a) and (28') we get the possible energies for the bound states:

$$E = -\frac{\hbar^2\nu^2}{8m}\left(\frac{2m\lambda}{(k+l)(\hbar\nu)^2} - (k+l)\right)^2, \tag{30}$$

where k is a positive integer fulfilling (29).

Letting the parameter $\nu\to0$ in (30), putting $Z=1$, and noting that $k+l=n$ is the principal quantum number, we arrive at the energy eigenvalues of the hydrogen atom.

The result in formulas (29) and (30) is in agreement

with the result obtained by Infeld and Hull.[12] Manning and Rosen[13] gave the eigenvalues (30) but without the condition (29).

The eigenvalues (30) have also been obtained by Rosenzweig and Krieger[10] by means of a phase-integral technique,[11] but the condition corresponding to (29) has not been given correctly by the authors of Ref. 10. The condition in the last column of their Table I for the case under consideration should, with their notations, be

$$n<\sqrt{\lambda}\,/\alpha-\tfrac{1}{2}-\tfrac{1}{2}(1+4b/\alpha^2)^{1/2}. \tag{31}$$

Inserting (27), (16a), (16b), (16c), (12a), (12b), and (12c) into (21), we get

$$u_{k-1}=Ne^{-z(-a)^{1/2}}(e^{-z}-1)^{l+1}F(-(k-1), 2l+1+k+2\sqrt{-a};2l+2;1-e^{-z}), \tag{32}$$

where the index $k-1$ of u enumerates the quantum states in such a way that the ground state has the index 0. Formula (32) was given by Manning and Rosen,[13] but only with the specification that the function F should be a terminated hypergeometric series.

We shall now determine the normalization factor N in (32) from the normalization condition

$$\int_0^\infty |u_{k-1}|^2\,dr=1. \tag{33}$$

Using formulas 15.4.6 of Ref. 17 and 10.8 (13) of Ref. 18, we can express u_{k-1} given by (32) in terms of Jacobi polynomials instead of hypergeometric functions, getting

$$u_{k-1}=N(-1)^{l+k}2^{-[l+1+(-a)^{1/2}]}\frac{\Gamma(k)\Gamma(2l+2)}{\Gamma(2l+1+k)}\times(1-x)^{(-a)^{1/2}}(1+x)^{l+1}P_{k-1}^{2(-a)^{1/2},2l+1}(x), \tag{34}$$

where

$$x=1-2e^{-z}. \tag{35}$$

Inserting (34) into (33) and using (35) and (6), we get

$$\frac{|N|^2}{\nu}2^{-2[l+1+(-a)^{1/2}]}\left(\frac{\Gamma(k)\Gamma(2l+2)}{\Gamma(2l+1+k)}\right)^2\times\int_{-1}^{1}(1-x)^{-1+2(-a)^{1/2}}(1+x)^{2l+2}\times(P_{k-1}^{2(-a)^{1/2},2l+1}(x))^2\,dx=1. \tag{36}$$

Writing one of the $2l+2$ factors $1+x$ in the integrand of (36) as

$$1+x=2-(1-x), \tag{37}$$

and using formulas (5) and (6) in Sec. 16.4 of Ref. 19, which are valid since the conditions

$$\mathrm{Re}(2\sqrt{-a})>0. \tag{38a}$$

and

$$\mathrm{Re}(2l+1)>-1 \tag{38b}$$

are fulfilled, we get

$$\frac{|N|^2}{\nu}\frac{(l+k)}{2(-a)^{1/2}[l+k+(-a)^{1/2}]}\times\frac{\Gamma(2l+1+k)\Gamma(k+(-a)^{1/2})}{\Gamma(2l+1+k+2(-a)^{1/2})\Gamma(k)}\times\left(\frac{\Gamma(k)\Gamma(2l+2)}{\Gamma(2l+1+k)}\right)^2=1. \tag{39}$$

Choosing the phase of N conveniently, we obtain

$$N=(-1)^{l+1}\frac{1}{\Gamma(2l+2)}\left(\nu\frac{2(-a)^{1/2}[l+k+(-a)^{1/2}]}{(l+k)}\times\frac{\Gamma(2l+1+k)\Gamma(2l+1+k+2(-a)^{1/2})}{\Gamma(k)\Gamma(k+2(-a)^{1/2})}\right)^{1/2}. \tag{40}$$

Using (40), we can write the normalized wavefunction (34) as

$$u_{k-1}=(-1)^{k-1}2^{-[l+1+(-a)^{1/2}]}\times\left(\nu\frac{2(-a)^{1/2}[l+k+(-a)^{1/2}]}{(l+k)}\times\frac{\Gamma(k)\Gamma(2l+1+k+2(-a)^{1/2})}{\Gamma(2l+1+k)\Gamma(k+2(-a)^{1/2})}\right)^{1/2}\times(1-x)^{(-a)^{1/2}}(1+x)^{l+1}P_{k-1}^{2(-a)^{1/2},2l+1}(x). \tag{41}$$

Using (32) and (40) instead, we get an alternative expression for the normalized wavefunction:

$$u_{k-1}=\left(\nu\frac{2(-a)^{1/2}[l+k+(-a)^{1/2}]}{(l+k)}\times\frac{\Gamma(2l+1+k)\Gamma(2l+1+k+2(-a)^{1/2})}{\Gamma(k)\Gamma(k+2(-a)^{1/2})}\right)^{1/2}\times\frac{1}{\Gamma(2l+2)}e^{-z(-a)^{1/2}}(1-e^{-z})^{l+1}\times F(-(k-1),2l+1+k+2\sqrt{-a};2l+2;1-e^{-z}). \tag{42a}$$

For small values of z, Eq. (42a) behaves like

$$u_{k-1}\sim\left(\nu\frac{2(-a)^{1/2}[l+k+(-a)^{1/2}]}{(l+k)}\times\frac{\Gamma(2l+1+k)\Gamma(2l+1+k+2(-a)^{1/2})}{\Gamma(k)\Gamma(k+2(-a)^{1/2})}\right)^{1/2}\times\frac{1}{\Gamma(2l+2)}z^{l+1},\quad z\to 0. \tag{42b}$$

Inserting (6) and (28′) into (42b), this equation can after simplifications be written

$$u_{k-1}\sim\frac{(2Z/a_0)^{l+3/2}}{(2l+1)!2^{1/2}(l+k)^{3/2}}\times\left[1-\left(\frac{\nu a_0}{2Z}\right)^2(l+k)^4\right]^{1/2}\prod_{j=0}^{l}\left\{\left(1-\frac{j^2}{(l+k)^2}\right)\times\left[1-\left(\frac{\nu a_0}{2Z}\right)^2(l+k)^2j^2\right]\right\}^{1/2}r^{l+1},\quad r\to 0, \tag{42c}$$

where $a_0=\hbar^2/me^2$ is the Bohr radius.

When we particularize (29), (30), and (42a) to $l=0$, we obtain the formulas given by Hulthén in Sec. 1 of Ref. 7.

6. DEFINITION AND CALCULATION OF CERTAIN MATRIX ELEMENTS $\mathcal{M}^{A,B}_{k-1,k'-1}$

Assuming that

$$\mathrm{Re}(A+\sqrt{-a_k}+\sqrt{-a_{k'}})>0 \tag{43a}$$

and

$$\mathrm{Re}(B+2l+3)>0, \tag{43b}$$

A and B being integers, we define matrix elements of products of the A th power of $e^{-\nu r}$ and the B th power of $1-e^{-\nu r}$ as

$$\mathcal{M}^{A,B}_{k-1,k'-1} = \langle u_{k-1}|(e^{-\nu r})^A(1-e^{-\nu r})^B|u_{k'-1}\rangle = \int_0^\infty u^*_{k-1}(e^{-\nu r})^A(1-e^{-\nu r})^B u_{k'-1}\,dr. \tag{44}$$

The conditions (43a) and (43b) are to be fulfilled in order to make the integral in (44) convergent. Inserting (6), (35), and (41) into (44), we get

$$\begin{aligned}\mathcal{M}^{A,B}_{k-1,k'-1} &= (-1)^{k+k'}2^{-[2l+2+A+B+(-a_k)^{1/2}+(-a_{k'})^{1/2}]}\left(\frac{2(-a_k)^{1/2}[l+k+(-a_k)^{1/2}]\Gamma(k)\Gamma(2l+1+k+2(-a_k)^{1/2})}{(l+k)\Gamma(2l+1+k)\Gamma(k+2(-a_k)^{1/2})}\right.\\ &\times\left.\frac{2(-a_{k'})^{1/2}[l+k'+(-a_{k'})^{1/2}]\Gamma(k')\Gamma(2l+1+k'+2(-a_{k'})^{1/2})}{(l+k')\Gamma(2l+1+k')\Gamma(k'+2(-a_{k'})^{1/2})}\right)^{1/2}\\ &\times\int_{-1}^{1}(1-x)^{A-1+(-a_k)^{1/2}+(-a_{k'})^{1/2}}(1+x)^{B+2l+2}P^{2(-a_k)^{1/2},2l+1}_{k-1}(x)\,P^{2(-a_{k'})^{1/2},2l+1}_{k'-1}(x)\,dx.\end{aligned} \tag{45}$$

For some diagonal matrix elements, the integral in (45) can be evaluated in a fairly direct way. If we put $k=k'$, $A=0$, and $B=-1$, the conditions (43a) and (43b) are fullfilled and (45) can be written

$$\begin{aligned}\mathcal{M}^{0,-1}_{k-1,k-1} &= 2^{-[2l+1+2(-a_k)^{1/2}]}\frac{2(-a_k)^{1/2}[l+k+(-a_k)^{1/2}]\Gamma(k)\Gamma(2l+1+k+2(-a_k)^{1/2})}{(l+k)\Gamma(2l+1+k)\Gamma(k+2(-a_k)^{1/2})}\\ &\times\int_{-1}^{1}(1-x)^{-1+2(-a_k)^{1/2}}(1+x)^{2l+1}\left[P^{2(-a_k)^{1/2},2l+1}_{k-1}(x)\right]^2dx.\end{aligned} \tag{46}$$

Using formula (6) in Sec. 16.4 of Ref. 19 in (46), we obtain

$$\mathcal{M}^{0,-1}_{k-1,k-1} = (l+k+\sqrt{-a_k})/(l+k). \tag{47}$$

If instead we put $k=k'$, $A=0$, and $B=-2$, the conditions (43a), (43b) are still fulfilled and (45) can be written

$$\begin{aligned}\mathcal{M}^{0,-2}_{k-1,k-1} &= 2^{-[2l+2(-a_k)^{1/2}]}\frac{2(-a_k)^{1/2}[l+k+(-a_k)^{1/2}]\Gamma(k)\Gamma(2l+1+k+2(-a_k)^{1/2})}{(l+k)\Gamma(2l+1+k)\Gamma(k+2(-a_k)^{1/2})}\\ &\times\int_{-1}^{1}(1-x)^{-1+2(-a_k)^{1/2}}(1+x)^{2l}\left[P^{2(-a_k)^{1/2},2l+1}_{k-1}(x)\right]^2dx.\end{aligned} \tag{48}$$

Inserting in the integrand of (48) a factor 1 as

$$1=\tfrac{1}{2}[(1+x)+(1-x)] \tag{49}$$

and using formulas (6) and (15) of Sec. 16.4 of Ref. 19, we obtain

$$\mathcal{M}^{0,-2}_{k-1,k-1} = \frac{[l+k+(-a_k)^{1/2}][2l+1+2(-a_k)^{1/2}]}{(l+k)(2l+1)}. \tag{50}$$

The matrix elements given by (47) and (50) will be used in the next section.

We shall now return to the calculation of the general matrix element $\mathcal{M}^{A,B}_{k-1,k'-1}$. Inserting formula 22.3.1 of Ref. 17 into (45), we get

$$\begin{aligned}\mathcal{M}^{A,B}_{k-1,k'-1} &= 2^{-[2l+A+B+k+k'+(-a_k)^{1/2}+(-a_{k'})^{1/2}]}\\ &\times\left(\frac{2(-a_k)^{1/2}[l+k+(-a_k)^{1/2}]\Gamma(k)\Gamma(2l+1+k+2(-a_k)^{1/2})}{(l+k)\Gamma(2l+1+k)\Gamma(k+2(-a_k)^{1/2})}\right.\\ &\times\left.\frac{2(-a_{k'})^{1/2}[l+k'+(-a_{k'})^{1/2}]\Gamma(k')\Gamma(2l+1+k'+2(-a_{k'})^{1/2})}{(l+k')\Gamma(2l+1+k')\Gamma(k'+2(-a_{k'})^{1/2})}\right)^{1/2}\\ &\times\sum_{i=0}^{k-1}\sum_{j=0}^{k'-1}(-1)^{i+j}\binom{k-1+2(-a_k)^{1/2}}{i}\binom{k+2l}{k-1-i}\binom{k'-1+2(-a_{k'})^{1/2}}{j}\binom{k'+2l}{k'-1-j}\\ &\times\int_{-1}^{1}(1+x)^{B+2l+2+i+j}(1-x)^{k+k'+A-3-i-j+(-a_k)^{1/2}+(-a_{k'})^{1/2}}dx.\end{aligned} \tag{51}$$

Putting

$$x = 2X - 1 \tag{52}$$

in the integrand of (51) and using the definition of the β function and the formula which expresses the β function in terms of the γ functions, i.e., formulas 6.2.1 and 6.2.2 of Ref. 17, we get

$$\begin{aligned}\mathscr{M}^{A,B}_{k-1,k'-1} = &\left(\frac{2(-a_k)^{1/2}[l+k+(-a_k)^{1/2}]\Gamma(k)\Gamma(2l+1+k+2(-a_k)^{1/2})}{(l+k)\Gamma(2l+1+k)\Gamma(k+2(-a_k)^{1/2})}\right.\\ &\left.\times\frac{2(-a_{k'})^{1/2}[l+k'+(-a_{k'})^{1/2}]\Gamma(k')\Gamma(2l+1+k'+2(-a_{k'})^{1/2})}{(l+k')\Gamma(2l+1+k')\Gamma(k'+2(-a_{k'})^{1/2})}\right)^{1/2}\\ &\times\sum_{i=0}^{k-1}\sum_{j=0}^{k'-1}(-1)^{i+j}\binom{k-1+2(-a_k)^{1/2}}{i}\binom{k+2l}{k-1-i}\binom{k'-1+2(-a_{k'})^{1/2}}{j}\binom{k'+2l}{k'-1-j}\\ &\times\frac{\Gamma(B+2l+3+i+j)\Gamma(k+k'+A-2-i-j+(-a_k)^{1/2}+(-a_{k'})^{1/2})}{\Gamma(2l+1+k+k'+A+B+(-a_k)^{1/2}+(-a_{k'})^{1/2})}.\end{aligned} \tag{53}$$

We have thus obtained an explicit formula for the general matrix element $\mathscr{M}^{A,B}_{k-1,k'-1}$ defined by (44).

As mentioned in Sec. 2, we get the potential of a hydrogenlike ion when $\nu\to 0$. Letting the parameter $\nu\to 0$ in (53) and putting at the same time $A = 0$ and $Z = 1$, we arrive, except for the factor $(-1)^{k+k'}$, at formula (21) of Ref. 20, which is valid for the hydrogen atom. This last-mentioned factor is due to a different choice of phase in the normalization of the hydrogen wavefunction.

In particular, for the diagonal matrix elements $\mathscr{M}^{A,B}_{k-1,k-1}$, we get from (53) the formula

$$\begin{aligned}\mathscr{M}^{A,B}_{k-1,k-1} = &\frac{2(-a_k)^{1/2}[l+k+(-a_k)^{1/2}]\Gamma(k)\Gamma(2l+1+k+2(-a_k)^{1/2})}{(l+k)\Gamma(2l+1+k)\Gamma(k+2(-a_k)^{1/2})}\\ &\times\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}(-1)^{i+j}\binom{k-1+2(-a_k)^{1/2}}{i}\binom{k+2l}{k-1-i}\binom{k-1+2(-a_k)^{1/2}}{j}\binom{k+2l}{k-1-j}\\ &\times\frac{\Gamma(B+2l+3+i+j)\Gamma(2k+A-2-i-j+2(-a_k)^{1/2})}{\Gamma(2l+1+2k+A+B+2(-a_k)^{1/2})}.\end{aligned} \tag{54}$$

7. A RECURRENCE FORMULA FOR THE EXPECTATION VALUES $\mathscr{M}^{A,B}_{k-1,k-1}$

As an alternative method for obtaining the expectation values $\mathscr{M}^{A,B}_{k-1,k-1}$, we shall in this section derive a recurrence formula by using a similar method as the one devised by Kramers in Sec. 59 of Ref. 21. Multiplying both members of the differential equation (9) by

$$\begin{aligned}&\xi^{A+2}(1-\xi)^{B+3}\frac{du}{d\xi} - \frac{A}{2}\xi^{A+1}(1-\xi)^{B+3}u\\ &+\frac{B+3}{2}\xi^{A+2}(1-\xi)^{B+2}u\end{aligned} \tag{55}$$

and integrating the resulting formula with respect to ξ, we get

$$\begin{aligned}\int_0^1\Big(&\xi^{A+2}(1-\xi)^{B+3}\frac{du}{d\xi} - \frac{A}{2}\xi^{A+1}(1-\xi)^{B+3}u\\ &+\frac{B+3}{2}\xi^{A+2}(1-\xi)^{B+2}u\Big)\Big(\frac{d^2u}{d\xi^2}+\frac{1}{\xi}\frac{du}{d\xi}+\frac{a}{\xi^2}u\\ &+\frac{b}{\xi(1-\xi)}u - \frac{l(l+1)}{(1-\xi)^2}u\Big)d\xi = 0.\end{aligned} \tag{56}$$

Assuming, in order that the integral in (56) be convergent, that [cf. (43a)]

$$\mathrm{Re}\big(A+2\sqrt{-a}\big) > 0 \tag{57a}$$

and [cf. (43b)]

$$\mathrm{Re}(B+2l+3) > 0, \tag{57b}$$

and integrating (56) by parts, we get when u is normalized according to (33),

$$\begin{aligned}&c_0\langle\xi^A(1-\xi)^{B+3}\rangle + c_1\langle\xi^{A+1}(1-\xi)^{B+2}\rangle\\ &\quad + c_2\langle\xi^{A+2}(1-\xi)^{B+1}\rangle + c_3\langle\xi^{A+3}(1-\xi)^B\rangle = 0,\end{aligned} \tag{58}$$

where

$$c_0 = A\,[4a + A^2], \tag{59a}$$

$$c_1 = 2b\,(2A+1) - (B+3)[4a + A^2 + (A+1)(2A+1)], \tag{59b}$$

$$c_2 = 3(A+1)(B+2)(B+3) - 4(A+1)l\,(l+1) - 2b\,(2B+5), \tag{59c}$$

$$c_3 = 4(B+2)(l+\tfrac{3}{2}+\tfrac{1}{2}B)(l-\tfrac{1}{2}-\tfrac{1}{2}B). \tag{59d}$$

Choosing $A = 0$, we get $c_0 = 0$ from (59a), and (58) becomes

$$\begin{aligned}&(c_2-c_1-c_3)\langle(1-\xi)^{B+3}\rangle + (c_1-2c_2+3c_3)\\ &\quad\times\langle(1-\xi)^{B+2}\rangle\\ &\quad + (c_2-3c_3)\langle(1-\xi)^{B+1}\rangle + c_3\langle(1-\xi)^B\rangle = 0.\end{aligned} \tag{60}$$

Inserting (59b), (59c), (59d), (5a), (5b), (4a), (4b), and (30) into (60), and letting $\nu\to 0$, we arrive at Kramers' recur-

rence formula for expectation values of powers of r for the hydrogen atom (Sec. 59 of Ref. 21).

The formulas (33), (47), and (50) can be written

$$\langle 1\rangle = 1, \tag{33'}$$

$$\langle (1-\xi)^{-1}\rangle = [l+k+(-a_k)^{1/2}]/(l+k), \tag{47'}$$

$$\langle (1-\xi)^{-2}\rangle = [l+k+(-a_k)^{1/2}][2l+1+2(-a_k)^{1/2}]/(l+k)(2l+1). \tag{50'}$$

By means of these expectation values together with the recurrence relation (60), one can obtain the expectation value of any power B of $1-\xi$ fulfilling the condition (57b). Examples of expectation values thus calculated by means of the recurrence formula (60) are

$$\langle 1-\xi\rangle = 1 - \frac{2(-a)^{1/2}[b+2l(l+1)]}{\{[2k+2l+2(-a)^{1/2}]^2-1\}(k+l)} \tag{61}$$

and

$$\langle (1-\xi)^2\rangle = 1 - \frac{2(-a)^{1/2}}{8\{[k+l+(-a)^{1/2}]^2-1\}(k+l)} \times\left(\frac{[b+2l(l+1)][-8a+14b-14+20l(l+1)]}{\{[2k+2l+2(-a)^{1/2}]^2-1\}} - 2l(l+1)\right). \tag{62}$$

APPENDIX

In this Appendix we shall investigate the validity of Levinson's theorem for the particular potential under consideration. The theorem can be formulated as follows: Provided the physical potential satisfies the conditions

$$\int_0^\infty |V(r)|r\,dr < \infty, \tag{63a}$$

$$\int_0^\infty |V(r)|r^2\,dr < \infty, \tag{63b}$$

the phase shift $\delta_l(E)$ and the number N_l of bound states with angular momentum l are connected by the relation

$$\delta_l(+0) - \delta_l(+\infty) = (N_l + \tfrac{1}{2})\pi \quad \text{if } l=0 \text{ and } f_0(0)=0, \tag{64a}$$

$$= N_l\pi \quad \text{otherwise}, \tag{64b}$$

where $f_l(E)$ is the Jost function defined according to Chap. 12 in Ref. 22.

For the physical potential given in (2') the conditions (63a) and (63b) are satisfied if $l=0$ but violated if $l\neq 0$. From (22) and Eq. (3.10) in Ref. 23 it follows that

$$f_0(E) = \lim_{z\to 0}[e^{iz(a)^{1/2}}(1-e^{-z}) \times F(\alpha+1-\gamma,\beta+1-\gamma;2-\gamma;e^{-z})]. \tag{65a}$$

Inserting (12a), (12b), (12c), and (16a), (16b), (16c) into (65a) and using Example 18 in Chap. 14 of Ref. 15, we obtain

$$f_0(E) = \Gamma(1-2i(2mE)^{1/2}/\hbar\nu)/\Gamma(1+[2m(\lambda-E)]^{1/2}/\hbar\nu - i(2mE)^{1/2}/\hbar\nu)\Gamma(1-[2m(\lambda-E)]^{1/2}/\hbar\nu - i(2mE)^{1/2}/\hbar\nu). \tag{65b}$$

Since $\lambda > 0$, it easily follows from (65b) that

$$f_0(0) = 0 \text{ if } (2m\lambda)^{1/2}/\hbar\nu \text{ is a positive integer}, \tag{66a}$$

$$f_0(0) \neq 0 \text{ if } (2m\lambda)^{1/2}/\hbar\nu \text{ is not a positive integer}. \tag{66b}$$

Consequently, Levinson's theorem states that for $l=0$ the relation (64a) is fulfilled if $(2m\lambda)^{1/2}/\hbar\nu$ is a positive integer, while (64b) is fulfilled if $(2m\lambda)^{1/2}/\hbar\nu$ is not a positive integer. For $l\neq 0$ the relation (64b) is expected to be violated, since the conditions (63a) and (63b) are then not fulfilled. We shall now verify the correctness of this in a direct way by means of the expression (24) for δ_l.

Inserting (16a), (16b), (16c), and (12a), (12b), (12c) into (24), we get

$$\delta_l = \tfrac{1}{2}(l+1)\pi + \arg\frac{\Gamma(2i(2mE)^{1/2}/\hbar\nu)}{\Gamma(l+1+i(2mE)^{1/2}/\hbar\nu + i[2m(E-\lambda)]^{1/2}/\hbar\nu)\Gamma(l+1+i(2mE)^{1/2}/\hbar\nu - i[2m(E-\lambda)]^{1/2}/\hbar\nu)}$$

$$= \tfrac{1}{2}(l+1)\pi - \arg\prod_{n=0}^{2l}[2l+2i(2mE)^{1/2}/\hbar\nu - n] - \arg\frac{\Gamma(l+1+i(2mE)^{1/2}/\hbar\nu + i[2m(E-\lambda)]^{1/2}/\hbar\nu)\Gamma(l+1+i(2mE)^{1/2}/\hbar\nu - i[2m(E-\lambda)]^{1/2}/\hbar\nu)}{\Gamma(2l+1+2i(2mE)^{1/2}/\hbar\nu)}. \tag{67}$$

Using the formula (cf. p. 2 in Ref. 24)

$$\frac{\Gamma(x+1)\Gamma(y+1)}{\Gamma(x+y+1)} = \prod_{n=1}^{\infty}\frac{n(x+y+n)}{(x+n)(y+n)}, \quad x,y\neq -1,-2,-3,\cdots,$$

we can write (67) as

$$\delta_l = \tfrac{1}{2}(l+1)\pi - \arg\prod_{n=0}^{2l}[2l+2i(2mE)^{1/2}/\hbar\nu - n] - \arg\prod_{n=1}^{\infty}\frac{n[2l+2i(2mE)^{1/2}/\hbar\nu + n]}{\{l+n+i(2mE)^{1/2}/\hbar\nu + i[2m(E-\lambda)]^{1/2}/\hbar\nu\}\{l+n+i(2mE)^{1/2}/\hbar\nu - i[2m(E-\lambda)]^{1/2}/\hbar\nu\}}$$

$$= \tfrac{1}{2}(l+1)\pi - \sum_{n=0}^{2l} \arg[2l + 2i(2mE)^{1/2}/\hbar v - n] - \sum_{n=1}^{\infty} \arg \frac{n[2l + 2i(2mE)^{1/2}/\hbar v + n]}{(l+n)^2 - 2m\lambda/(\hbar v)^2 + 2i(l+n)(2mE)^{1/2}/\hbar v}. \quad (68)$$

From (68) it follows that

$$\delta_l(+0) - \delta_l(+\infty) = \begin{cases} (l+\tfrac{1}{2})\pi + [(2m\lambda)^{1/2}/\hbar v - l - 1]\pi & \text{if } (2m\lambda)^{1/2}/\hbar v \text{ is a positive integer,} \quad (69a) \\ l\pi + [(2m\lambda)^{1/2}/\hbar v - l]\pi & \text{if } (2m\lambda)^{1/2}/\hbar v \text{ is not a positive integer,} \quad (69b) \end{cases}$$

where $(2m\lambda)^{1/2}/\hbar v$ is equal to the integer part of $(2m\lambda)^{1/2}/\hbar v$. Recalling (29), we can write (69a) and (69b) as

$$\delta_l(+0) - \delta_l(+\infty) = \begin{cases} \pi + (N_l + \tfrac{1}{2})\pi & \text{if } (2m\lambda)^{1/2}/\hbar v \text{ is a positive integer,} \quad (70a) \\ \pi + N_l\pi & \text{if } (2m\lambda)^{1/2}/\hbar v \text{ is not a positive integer,} \quad (70b) \end{cases}$$

where N_l is the number of bound states in the potential under consideration. A comparison between (70a), (70b), on the one hand, and (64a), (64b) and (66a), (66b), on the other hand, shows that Levinson's theorem is fulfilled when $l = 0$ but violated when $l \neq 0$. This fact is consistent with the conditions for the validity of Levinson's theorem.

ACKNOWLEDGMENTS

I am most grateful to Dr. N. Fröman and Dr. P. O. Fröman for suggesting the problem and for their encouraging interest and help during my work.

[1]H. Yukawa, Proc. Phys. Math. Soc. Jpn. **17**, 48 (1935).
[2]C. Möller and L. Rosenfeld, Mat. Fys. Medd. Dan. Vid. Selsk. XVII, no. 8 (1940)
[3]N. Bohr, Mat. Fys. Medd. Dan. Vid. Selsk. XVIII, no. 8 (1948).
[4]A. H. Wilson, Proc. Cambridge Philos. Soc. **34**, 365 (1938).
[5]R. Sachs and M. Goeppert-Mayer, Phys. Rev. **53**, 991 (1938).
[6]E. A. Hylleraas and V. Risberg, Avh. Nor. Vidensk. Akad. Oslo Mat. Naturvidersk. Kl. I, No. 3 (1941).
[7]L. Hultén, Ark. Mat. Astron. Fys. A **28**, No. 5 (1942)
[8]J. Lindhard and A. Winther, Nucl. Phys. A **166**, 413 (1971).
[9]S. G. M. Gustavi, Ark. Fys. **11**, 437 (1957).
[10]C. Rosenzweig and J. B. Krieger, J. Math. Phys. **9**, 849 (1968).
[11]N. Fröman and P. O. Fröman, JWKB *Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965) (MIR, Moscow, 1967).
[12]L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).
[13]M. Manning and N. Rosen, Phys. Rev. **44**, 953 (1933).
[14]A. K. Bose, Nuovo Cimento **32**, 679 (1964).
[15]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University, Cambridge, New York, 1969).
[16]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions. Bateman Manuscript Project. Vol.* 1 (McGraw-Hill, New York, 1953).
[17]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions. National Bureau of Standards Applied Mathematical Series.* 55, Fourth Printing, Washington, December 1965.
[18]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Bateman Manuscript Project, Vol. 2 (McGraw-Hill, New York, 1953).
[19]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms*, Bateman Manuscript Project, Vol. 2 (McGraw-Hill, New York, 1954).
[20]N. Bessis, G. Bessis, and G. Hadinger, Phys. Rev. A **8**, 2246 (1973).
[21]H. A. Kramers, *Die Grundlagen der Quantentheorie. Quantentheorie des Elektrons und der Strahlung* (Akademische Verlagsgesellschaft, Leipzig, 1938).
[22]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[23]G. Drukarev, N. Fröman, and P. O. Fröman, J. Phys. A Math. Nucl. Gen. **12**, 171 (1979).
[24]W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics* (Chelsea, New York, 1954).

Analysis of interacting quantum field theory in curved spacetime

N. D. Birrell[a] and J. G. Taylor
Department of Mathematics, University of London, King's College, Strand, London WC2R 2LS, United Kingdom

(Received 17 May 1979; accepted for publication 11 October 1979)

A detailed analysis of interacting quantized fields propagating in a curved background spacetime is given. Reduction formulas for S-matrix elements in terms of vacuum Green's functions are derived, special attention being paid to the possibility that the "in" and "out" vacuum states may not be equivalent. Green's functions equations are obtained and a diagrammatic representation for them given, allowing a formal, diagrammatic renormalization to be effected. Coordinate space techniques for showing renormalizability are developed in Minkowski space, for $\lambda\phi^3_{(4,6)}$ field theories. The extension of these techniques to curved spacetimes is considered. It is shown that the possibility of field theories becoming nonrenormalizable there cannot be ruled out, although, allowing certain modifications to the theory, $\phi^3_{(4)}$ is proven renormalizable in a large class of spacetimes. Finally particle production from the vacuum by the gravitational field is discussed with particular reference to Schwarzschild spacetime. We shed some light on the nonlocalizability of the production process and on the definition of the S matrix for such processes.

1. INTRODUCTION

Despite the fact that in recent years much work has been done on the study of free quantized fields in curved background spacetimes, there has been very little study of self, or mutually interacting fields in such backgrounds. The work that has been done on interacting fields has tended to be on very simple special cases. In particular the first substantial calculations to be performed in this area, as far as we know, are those of Drummond,[1] who considers the case of a massless self-interacting field in a spherical (Euclideanized De Sitter) universe in perturbation theory. Subsequently, Drummond and Shore[2] performed perturbation calculations for massless electrodynamics in a spherical background, a situation first studied by Adler.[3] More recently Birrell and Ford[4-6] and Bunch, Panangaden, and Parker[7,8] have carried out perturbation calculations in Robertson Walker spacetimes. On the other hand studies of the exactly solvable interacting Thirring model in a curved background have been made by Scarf[9] and Birrell and Davies[10]; however, this is a rather unrealistic theory of massless fermions in a two-dimensional universe.

The main reason that so little work has been done in this area is that it is very difficult to obtain exact results in perturbation theory. As a start it is generally a most difficult task to merely find the free Feynman propagator on which perturbation theory is based. However there are a number of investigations which can be carried out without resorting to actual calculation, and it is to some of these that we devote our attention in this paper.

The most obvious question to ask ourselves is whether a curved background spacetime (that is the presence of an external gravitational field) can destroy the renormalizability of an otherwise renormalizable field theory. One intuitively expects that it cannot, because the ultraviolet divergences are, in coordinate space, short distance divergences, and by the principle of equivalence, the short distance behavior should be the same as in flat space. However it is not clear "how far" about a point the divergences probe and whether the part of the free propagator $D_F(x, x')$ which is finite when $x \to x'$ (and different from that part in flat space) will play a part in the divergence structure. Indeed a naive examination of say the second-order perturbation contribution to the complete $\lambda\phi^4$ propagator using the De Witt Schwinger expansion for the bare propagator[11,12] suggests that there will be divergent terms that do not appear in flat space. It is thus of some interest to make a more rigorous study of the renormalization process in curved spacetimes. This we do in Sec. 5 for $\lambda\phi^3$ theory in four and six dimensions and find that we are not, in general, able to eliminate the possibility that the theories might be nonrenormalizable in certain curved spacetimes. The reason for this is that the divergences which occur in the theory do indeed probe more than just the local neighborhood of a point, preventing us from making general statements about the structure of the counterterms which will be needed to remove certain of the infinities. However $\lambda\phi^3$ in four dimensions does not have this dangerous possibility to such a degree; its renormalization proceeds as in flat spacetimes.

Before coming to this conclusion we carefully develop in Secs. 2 and 3 reduction formulas and Green's function equations for the self-interacting fields in a curved background. In so doing we give particular attention to the possibility that the initial and final vacuum states may not be the same; a subject already much discussed for free fields in curved spacetimes. The Green's function equations are represented in a compact diagrammatic form identical to the flat spacetime case.[13] in writing equations for connected Green's functions, however, we must contend with "tadpole" diagrams which represent absorption or emission of one particle by the vacuum, and which do not contribute in flat spacetime. The diagrammatic Green's function approach allows us to formally renormalize the equations just as in flat spacetime. That is, all bare vertices and propagators

[a]Supported by a Flinders University of South Australia Overseas Scholarship.

 0022-2488/80/071740-21$1.00

are replaced by their clothed counterparts in a consistent way. Since this procedure is to a certain extent modified by the occurrence of tadpole diagrams it is described in some detail in Sec. 3.

In general it is not possible to perform momentum space calculations of amplitudes in curved spacetimes, unless they satisfy very restrictive requirements. For this reason we work in coordinate space, a technique not too familiar in flat-space-field theory. To demonstrate our strategy for considering the renormalization of $\lambda\phi^3$ in curved spacetime, we show in Sec. 4 how it applies in flat spacetime. This is of interest in its own right, as we use the Green's function equations along with the theory of products of distributions in an iterative procedure which makes short work of the proof of renormalizability. In extending the procedure to curved spacetime we come upon new problems which cannot be completely solved for the reasons mentioned briefly above.

The spacetimes to which the developments of this paper apply are globally hyperbolic with Cauchy hypersurfaces which are either noncompact, or compact without boundary but asymptotically having infinite volume in the far past and future. These restrictions ensure that the usual procedure of defining as S matrix and setting up Green's function equation works almost as in flat spacetime. The difficult problems of defining interacting quantum field theories in curved spacetimes which are not globally hyperbolic or which are globally hyperbolic with arbitrary compact Cauchy surfaces is not considered here. In the case of a spacetime which has asymptotically compact Cauchy surfaces, there is a serious difficulty in deciding on initial values for fields which are never free. If the Cauchy surfaces are also bounded there are additional problems in determining the boundary values of the fields at all times. These are problems which should be given consideration in the future, and in particular are of importance in discussing radiative corrections to closed cosmologies where the gravitational field is also quantized.

In Sec. 6 we make some comments on how our formalism can be applied to particle production calculations, with particular reference to Schwarzschild spacetime. The final section concludes by considering the main features of our work. In particular the question of whether renormalizability can be firmly proved or disproved is discussed. In this connection several second-order perturbation theory calculations in $\lambda\phi^4$ theory[4-8] are mentioned.

2. REDUCTION FORMULAS

If we wish to consider the renormalization of the S-matrix for scattering from an initial "in" state to a final "out" state in terms of Green's functions it is essential that we have some means of calculating the former quantity in terms of the latter. The derivation of such reduction formulas for interacting fields in flat space is well known (see for example Ref 14, Chap. 16), however in curved space the discussion is complicated by the possiblity of the final and initial vacuum states being different. The effect of having different final and initial vacua has been much discussed in the case of free fields in curved backgrounds (for a review see Ref. 15); in particular DeWitt[11] derives equations for the scattering amplitudes for free scalar fields in such spacetimes. We shall now obtain equations which in the limit of flat space give the usual reduction formulas, and in the limit of free fields (zero coupling constant) in curved space give DeWitt's equations.

For our discussion we consider a self-interacting scalar field with Lagrangiam density[16]

$$\mathscr{L} = \tfrac{1}{2}\sqrt{-g}\,[g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - (m^2 + \xi R)\phi^2] + \mathscr{L}_I, \tag{2.1}$$

where the exact nature of the interaction Lagrangian does not enter into the derivation of this section, but might as well be assumed to be

$$\mathscr{L}_I = -\sqrt{-g}\,[\tfrac{1}{3}\lambda\phi^3 + a\phi]\,. \tag{2.2}$$

(The term $a\phi$, where a is a function of spacetime, is included for use later.)

The field equation resulting from (2.1) and (2.2) is

$$K_x\phi(x) + a(x) + \lambda\phi(x)^2 = 0, \tag{2.3}$$

where we have defined

$$K_x \equiv \Box_x + m^2 + \xi R_x, \tag{2.4}$$

and

$$\Box_x = (1/\sqrt{-g_x})\partial_\mu\,[g^{\mu\nu}(x)\sqrt{-g_x}\,\partial_\nu\phi\,] = \nabla_\mu\nabla^\mu, \tag{2.5}$$

∇_μ being the usual covariant derivative (also denoted by a semi-colon).

We assume that the spacetime is globally hyperbolic, and hence[17] homeomorphic to $R^1\times\Sigma$, where for each $x^0\in R^1$, $\{x^0\}\times\Sigma$ is a Cauchy surface for the spacetime.[18] We can now define for any two solutions u_1, u_2 of the free field equation,

$$K_x u(x) = 0, \tag{2.6}$$

a conserved scalar product

$$(u_1,u_2) \equiv -i\int_\Sigma (u_1\partial_\mu u_2^* - u_2^*\partial_\mu u_1)\sqrt{-g}\,d\Sigma^\mu = -i\int_\Sigma u_1\overleftrightarrow{\partial}_\mu u_2^*\sqrt{-g}\,d\Sigma^\mu, \tag{2.7}$$

where $d\Sigma^\mu$ is the future directed surface element to Σ.

Let f_n, f_n^* and g_n, g_n^* be two sets of complete solutions of (2.6), satisfying

$$(f_n,f_m) = \delta_{nm}, \quad (f_n^*f_m^*) = -\delta_{nm}, \quad (f_nf_m^*) = 0, \tag{2.8}$$

and similarly for g_n. These will be chosen in some way from the infinite number of such sets. We assume that as $x^0\to\pm\infty$ the interaction is switched off adiabatically and the solution of (2.3) becomes asymptotically free (in the sense of weak operator convergence):

$$\lim_{x^0\to+\infty}\phi(x) = \phi_{\text{out}}(x), \tag{2.9}$$

$$\lim_{x^0\to-\infty}\phi(x) = \phi_{\text{in}}(x), \tag{2.10}$$

where

$$K_x \phi_{\text{out(in)}}(x) = 0. \tag{2.11}$$

With these asymptotic conditions, which are only physically tenable in spacetimes which asymptotically have spatial sections of unbounded volume, we can expand ϕ_{in} and ϕ_{out} in terms of either of the complete sets of modes, thus

$$\phi_{\text{in}}(x) = \sum_n [A^{\text{in}}_n f_n(x) + A^{\text{in}\,\dagger}_n f^*_n(x)] \tag{2.12a}$$

$$= \sum_n [B^{\text{in}}_n g_n(x) + B^{\text{in}\,\dagger}_n g^*_n(x)], \tag{2.12b}$$

$$\phi_{\text{out}}(x) = \sum_n [A^{\text{out}}_n f_n(x) + A^{\text{out}\,\dagger}_n f^*_n(x)] \tag{2.13a}$$

$$= \sum_n [B^{\text{out}}_n g_n(x) + B^{\text{out}\,\dagger}_n g^*_n(x)]. \tag{2.13b}$$

We may now quantize the system and construct "in" and "out" Fock spaces as in flat space. However while in flat space the definition of the "in" and "out" vacua (denoted $|\text{in}\rangle$ and $|\text{out}\rangle$, respectively) is clearly established, in curved space there can exist an infinite number of equally acceptable inequivalent vacua. In particular we could define $|\text{in}\rangle$ by $A^{\text{in}}_n |\text{in}\rangle = 0$ or $B^{\text{in}}_n |\text{in}\rangle = 0$, and similarly $|\text{out}\rangle$ could be defined by $A^{\text{out}}_n |\text{out}\rangle = 0$ or $B^{\text{out}}_n |\text{out}\rangle = 0$. In certain spacetimes there will be a "natural" definition of positive frequency solution f_n to (2.6) and these can be used to give a "natural" definition of the vacuum. For example for universes which are asymptotically slowly expanding Parker's[5,19] extended WKB definition of positive frequency modes is appropriate.

Let us suppose that the modes f_n are in some sense positive frequency as $x^0 \to -\infty$, and, as there is no reason why f_n should necessarily also be positive frequency as $x^0 \to +\infty$,[5] let us assume that for $x^0 \to +\infty$ the g_n are positive frequency. We thus define the "in" and "out" vacua by

$$A^{\text{in}}_n |\text{in}\rangle = 0, \tag{2.14}$$

$$B^{\text{out}}_n |\text{out}\rangle = 0. \tag{2.15}$$

This then gives asymptotic definitions of particles. How "natural" these definitions will be will depend on the particular spacetime under consideraion, and the choice of positive frequency modes.

Since the modes f_n, f^*_n are complete we can relate them to g_n (or g^*_n) by a Bogolubov transformation:

$$g_n = \sum_m (\alpha_{nm} f_m + \beta_{nm} f^*_m), \tag{2.16}$$

and conversely one can show that

$$f_n = \sum_m (\alpha^*_{mn} g_m - \beta_{mn} g^*_m). \tag{2.17}$$

Using (2.8) we find the following relation between the Bogolubov coefficients which will be used in the ensuing discussion:

$$\sum_r (\alpha_{mr} \alpha^*_{nr} - \beta_{mr} \beta^*_{nr}) = \delta_{mn}. \tag{2.18}$$

From (2.12), (2.13) and (2.16), (2.17) we obtain

$$A^{\text{in}}_n = \sum_m (\alpha_{mn} B^{\text{in}}_m + \beta^*_{mn} B^{\text{in}\,\dagger}_m), \tag{2.19}$$

$$B^{\text{in}}_n = \sum_m (\alpha^*_{nm} A^{\text{in}}_m - \beta^*_{nm} A^{\text{in}\,\dagger}_m), \tag{2.20}$$

$$A^{\text{out}}_n = \sum_m (\alpha_{mn} B^{\text{out}}_m + \beta^*_{mn} B^{\text{out}\,\dagger}_m), \tag{2.21}$$

$$B^{\text{out}}_n = \sum_m (\alpha^*_{nm} A^{\text{out}}_m - \beta^*_{nm} A^{\text{out}\,\dagger}_m). \tag{2.22}$$

Now for free fields we have $A^{\text{in}}_n = A^{\text{out}}_n$ and $B^{\text{in}}_n = B^{\text{out}}_n$, in which case the S matrix elements can be determined from a knowledge of the Bogolubov coefficients.[11] For interacting fields, even in flat space, these equalities will not necessarily hold, and additional information (the Green's functions) is needed to construct the S matrix. We now consider such a construction.

Define an s-particle out state by

$$|\underline{n}^{(s)}, \text{ out}\rangle = \prod_{i=1}^{s} B^{\text{out}\,\dagger}_{n_i} |\text{out}\rangle, \tag{2.23}$$

where $\underline{n}_{(s)} = \{n_1,\ldots,n_s\}$, and an r-particle in state by

$$|\underline{m}^{(r)}, \text{ in}\rangle = \prod_{i=1}^{r} A^{\text{in}\,\dagger}_{m_i} |\text{in}\rangle, \tag{2.24}$$

with $\underline{m}^{(r)} = \{m_1,\ldots,m_r\}$. Our aim is to evaluate the amplitude $\langle \text{out}, \underline{n} \mid \underline{m}, \text{in}\rangle$ in terms of Green's functions, that is, in terms of vacuum expectation values of time (x^0) ordered products of field operator solutions of (2.3).

Assume for the sake of argument that $s \geqslant r$, the case $r < s$ being very similar, and consider

$$\langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)}, \text{in}\rangle = \langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | A^{\text{in}\,\dagger}_{m_i} | \underline{m}^{(r)} - m_i, \text{in}\rangle. \tag{2.25}$$

The time-ordered product in (2.25) is of an arbitrary collection of field operators, and $\underline{m}^{(r)} - m_i$ denotes $\{m_1,\ldots,m_{i-1}, m_{i+1},\ldots,m_r\}$. Using (2.10), (2.12a), and (2.8) we can write (2.25) as

$$i \int_{x^0 = \text{const}} \langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)\phi(x)) | \underline{m}^{(r)} - m_i, \text{in}\rangle \overleftrightarrow{\partial}_\mu f_{m_i}(x) \sqrt{-g}\, d\Sigma^\mu_x,$$

which upon using Gauss' theorem, (2.9), (2.13a), (2.8), and the fact that $f_{m_i}(x)$ satisfies (2.6) gives

$$\begin{aligned} &\langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)}, \text{in}\rangle \\ &\quad = i \int f_{m_i}(x) K_x \langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)\phi(x)) | \underline{m}^{(r)} \\ &\qquad - m_i, \text{in}\rangle \sqrt{-g}\, d^4x \\ &\qquad + \langle \text{out}, \underline{n}^{(s)} | A^{\text{out}\,\dagger}_{m_i} T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)} \\ &\qquad - m_i, \text{in}\rangle. \end{aligned} \tag{2.26}$$

In the derivation of the flat space reduction formulas one has $A^{\text{out}\,\dagger}_{m_i} = B^{\text{out}\,\dagger}_{m_i}$, and thus if one assumes that $\underline{n}^{(s)} \cap \underline{m}^{(r)} = \{\ \}$, then $\langle \text{out}, \underline{n}^{(s)} | A^{\text{out}\,\dagger}_{m_i} = 0$ and the second term in (2.26) vanishes. However in curved space the "out" Fock space is based on the $B^{\text{out}\,\dagger}_m$ operators which are not necessarily the same as the $A^{\text{out}\,\dagger}_m$, and for this reason the second term in (2.26) will not in general vanish, but requires further treatment.

From (2.18) one can deduce that α, considered as a matrix, will have an inverse α^{-1}. This allows us to obtain

from (2.22) the relation

$$A^{\text{out}\,\dagger}_{m_i} = \sum_q \left(\alpha^{-1}_{m_i p} B^{\text{out}\,\dagger}_p + \sum_q \alpha^{-1}_{m_i p} \beta_{pq} A^{\text{out}}_q \right), \tag{2.27}$$

which when used in the second term of (2.26) gives

$$\langle \text{out}, \underline{n}^{(s)} | A^{\text{out}\,\dagger}_{m_i} T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)} - m_i, \text{in} \rangle$$

$$= \sum_{p,q} \alpha^{-1}_{m_i p} \beta_{pq} i \int f^*_q(x) K_x \langle \text{out}, \underline{n}^{(s)}$$

$$| T(\phi(x)\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)} - m_i, \text{in} \rangle \sqrt{-g}\, d^4x$$

$$+ \sum_j \alpha^{-1}_{m_i n_j} \langle \text{out}, \underline{n}^{(s)} - n_j | T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)} - m_i, \text{in} \rangle$$

$$+ \sum_{p,j} \alpha^{-1}_{m_i p} \beta_{p m_j} \langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)}$$

$$- m_i - m_j, \text{in} \rangle, \tag{2.28}$$

where we have also used (2.9), (2.13a), (2.8), Gauss' theorem, (2.10), (2.12a), and the fact that f_q satisfies (2.6). If we now eliminate f_q from (2.28) using (2.16) and substitute the result into (2.26) we find

$$\langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)}, \text{in} \rangle$$

$$= O_{z_i} \langle \text{out}, \underline{n}^{(s)} | T(\phi(z_i)\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)} - m_i, \text{in} \rangle$$

$$+ \sum_j \alpha^{-1}_{m_i n_j} \langle \text{out}, \underline{n}^{(s)} - n_j | T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)} - m_i, \text{in} \rangle$$

$$+ i \sum_j \Lambda_{ij} \langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \underline{m}^{(r)} - m_i - m_j, \text{in} \rangle. \tag{2.29}$$

In writing (2.29) we have introduced the integral operator

$$O_{z_i} \equiv i \sum_p \alpha^{-1}_{m_i p} \int \sqrt{-g_{z_i}}\, d^4 z_i g_p(z_i) K_{z_i}, \tag{2.30}$$

and have defined[20]

$$\Lambda_{ij} \equiv -i \sum_p \beta_{p m_j} \alpha^{-1}_{m_i p}, \tag{2.31}$$

which is in fact symmetric in i, j.[11]

By repeated use of (2.29) we arrive at

$$\langle \text{out}, \underline{n}^{(s)} | \underline{m}^{(r)}, \text{in} \rangle$$

$$= \sum_p i^{(r-l)/2} O_{z_{\rho(1)}} \cdots O_{z_{\rho(k)}} \alpha^{-1}_{m_{\rho(k+1)} n_{\rho(k+1)}} \cdots \alpha^{-1}_{m_{\rho(l)} n_{\rho(l)}}$$

$$\times \Lambda_{\rho(l+1)\rho(l+2)} \cdots \Lambda_{\rho(r-1)\rho(r)}$$

$$\times \langle \text{out}, \underline{n}^{(s)} - n_{\rho(k+1)} - \cdots$$

$$- n_{\rho(l)} | T(\phi(z_{\rho(1)}\cdots\phi(z_{\rho(k)})) | \text{in} \rangle, \tag{2.32}$$

where the sum is over all distinct ways of attaching the indices $\rho(i) = 1, 2., \ldots, r$, to combination of the operators O, α^{-1}, and Λ.

If we assume that there is no interaction, by putting $\lambda = 0, a = 0$ in Eq. (3.15) of the next section and noting (2.6) we easily see that there will be no O_z terms in (2.32). If we further take $s = 0$ there will be no α^{-1} term, and (2.32) reduces to

$$\langle \text{out} | \underline{m}^{(r)}, \text{in} \rangle = i^{r/2} \langle \text{out} | \text{in} \rangle \sum_p \Lambda_{\rho_1\rho_2} \cdots \Lambda_{\rho_{r-1}\rho_r}, \quad r \text{ even}$$

$$= 0, \quad r \text{ odd}, \tag{2.33}$$

which is precisely DeWitt's[11] Eq. (164).

We now only need to know how to reduce amplitudes of the form

$$\langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \text{in} \rangle$$

$$= \langle \text{out}, \underline{n}^{(s)} - n_i | B^{\text{out}}_{n_i} T(\phi(x_1)\cdots\phi(x_n)) | \text{in} \rangle.$$

By arguments very similar to those used in arriving at (2.29) we can show that such amplitudes may be written as

$$\langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \text{in} \rangle$$

$$= Q_{z_i} \langle \text{out}, \underline{n}^{(s)} - n_i | T(\phi(z_i)\phi(x_1)\cdots\phi(x_n)) | \text{in} \rangle$$

$$+ i \sum_j V_{ij} \langle \text{out}, \underline{n}^{(s)} - n_i - n_j | T(\phi(x_1)\cdots\phi(x_n)) | \text{in} \rangle, \tag{2.34}$$

where we have defined

$$Q_{z_i} = i \sum_p \alpha^{-1}_{p n_i} \int \sqrt{-g_{z_i}}\, d^4 z_i f_p(z_i) K_{z_i}, \tag{2.35}$$

$$V_{ij} = i \sum_p \beta_{n_j p} \alpha^{-1}_{p n_i} \tag{2.36}$$

(see Ref. 20). Repeated application of this equation gives

$$\langle \text{out}, \underline{n}^{(s)} | T(\phi(x_1)\cdots\phi(x_n)) | \text{in} \rangle$$

$$= \sum_\sigma i^{(s-k)/2} Q_{z_{\sigma(1)}} \cdots Q_{z_{\sigma(k)}} V_{\sigma(k+1)\sigma(k+2)} \cdots V_{\sigma(s-1)\sigma(s)}$$

$$\times \langle \text{out} | T(\phi(z_{\sigma(1)})\cdots\phi(z_{\sigma(k)})\phi(x_1)\cdots\phi(x_n)) | \text{in} \rangle, \tag{2.37}$$

which in the limit of no interaction (and no initial time-ordered product) gives DeWitt's Eq. (163).

If we now use (2.37) in (2.32) we finally arrive at the complete reduction formula:

$$\langle \text{out}, \underline{n}^{(s)} | \underline{m}^{(r)}, \text{in} \rangle$$

$$= \sum_p i^{(r-l)/2} O_{z_{\rho(1)}} \cdots O_{z_{\rho(k)}} \alpha^{-1}_{m_{\rho(k+1)} n_{\rho(k+1)}} \cdots \alpha^{-1}_{m_{\rho(l)} n_{\rho(l)}}$$

$$\times \Lambda_{\rho(l+1)\rho(l+2)} \cdots \Lambda_{\rho(r-1)\rho(r)} \sum_\sigma i^{(s-l-w+k)/2} Q_{z_{\sigma(1)}} \cdots Q_{z_{\sigma(w)}}$$

$$\times V_{\sigma(w+1)\sigma(w+2)} \cdots V_{\sigma(s-l+k-1)\sigma(s-l+k)}$$

$$\times \langle \text{out} | T(\phi(z_{\sigma(1)})\cdots\phi(z_{\sigma(w)})\phi(z_{\rho(1)})\cdots\phi(z_{\rho(k)})) | \text{in} \rangle. \tag{2.38}$$

As before the ρ sum is over all distinct ways of attaching the indices $\rho(i) = 1, 2, \ldots, r$ to combinations of the operations O, α^{-1} and Λ. The σ sum is then over all distinct ways of attaching the indices $\sigma(i) = \{ \underline{n}^{(s)} - n_{\rho(k+1)} - \cdots - n_{\rho(l)} \}$ to combinations of the operators Q and V.

Finally, if we go to the limit of flat space we have $\beta_{ij} = 0$, $\alpha_{ij} = \delta_{ij}$ and hence $\Lambda = V = 0$. We also have $f_p = g_p$, and, assuming as is usual in flat space derivations that $\underline{n}^{(s)} \cap \underline{m}^{(r)} = \{\ \}$, then (2.38) reduces to the usual flat space reduction formula [see, for example, Ref. 1, Eq. (16.81)].

The discussion given here applies equally well to globally hyperbolic spacetimes with or without horizons, however in the former case great care is needed in the choice of the modes f_n and g_n, as has been discussed, for example by Unruh[21] in the case of free fields. The interpretation of the S

matrix as far as scattering of particles across the horizon is concerned also needs care.[22] We shall return to those topics in Sec. 6.

3. GREEN'S FUNCTION EQUATIONS IN CURVED SPACETIME

In the last section we showed how the determination of scattering amplitudes could be reduced to a calculation of the Green's functions.

$$\tau(x,x_1\cdots x_n)\equiv\frac{\langle \text{out}|T(\phi(x)\phi(x_1)\cdots\phi(x_n))|\text{in}\rangle}{\langle \text{out}|\text{in}\rangle}. \tag{3.1}$$

We now wish to derive equations from which the Green's functions can in principle be determined.

To obtain such equations we need only assume (in addition to the assumptions of Sec. 2) that the interacting field is quantized in the usual (canonical) fashion, satisfying commutation relations[5]

$$\begin{aligned}&[\phi(x^0,\mathbf{x}),\phi(x^0,\mathbf{x}')]=0,\\&[\pi(x^0,\mathbf{x}),\pi(x^0,\mathbf{x}')]=0,\\&[\phi(x^0,\mathbf{x}),\pi(x^0,\mathbf{x}')]=i\delta(\mathbf{x}-\mathbf{x}'),\end{aligned} \tag{3.2}$$

where the canonical momentum conjugate to ϕ is

$$\pi=\partial\mathscr{L}_0/\partial(\partial_0\phi)=\sqrt{-g}g^{0\mu}\partial_\mu\phi, \tag{3.3}$$

with $\mathscr{L}_0$ given by (2.1) without the interaction term.

We shall also need to use the Feynman propagator (for the free fields) which satisfies

$$K_xD_F(x,y)=-(1/\sqrt{-g})\delta(x-y), \tag{3.4}$$

and is given in terms of free-field solutions ϕ_f of (2.6) by

$$iD_F(x,y)={}_f\langle \text{out}|T(\phi_f(x)\phi_f(y))|\text{in}\rangle_f/{}_f\langle \text{out}|\text{in}\rangle_f. \tag{3.5}$$

If ϕ_f has a mode decomposition in terms of the f_n or g_n of the last section given by

$$\phi_f(x)=\sum_n[A_nf_n(x)+A_n^\dagger f_n^*(x)] \tag{3.6a}$$

$$=\sum_n[B_ng_n(x)+B_n^\dagger g_n^*(x)], \tag{3.6b}$$

then the vacua $|\text{out}\rangle_f$ and $|\text{in}\rangle_f$ are defined by

$$B_n|\text{out}\rangle_f=0 \tag{3.7}$$

and

$$A_n|\text{in}\rangle_f=0,$$

respectively. Note that A_n need not necessarily be equal to A_n^{in} of Sec. 2 and hence $|\text{in}\rangle_f$ will not in general be the same as $|\text{in}\rangle$. Similarly $|\text{out}\rangle_f$ will not necessarily be equal to $|\text{out}\rangle$.

If in (3.5) we use (3.6a) for $x^0>y^0$ and (3.6b) for $x^0<y^0$ we obtain

$$\begin{aligned}iD_F(x,y)=\theta(x^0-y^0)[D_{\text{in}}^{(-)}(x,y)+d_{\text{in}}(x,y)]\\+\theta(y^0-x^0)[D_{\text{out}}^{(+)}(x,y)+d_{\text{out}}(x,y)],\end{aligned} \tag{3.8}$$

where we have defined

$$D_{\text{in}}^{(-)}(x,y)=\sum_n f_n(x)f_n^*(y), \tag{3.9}$$

$$D_{\text{out}}^{(+)}(x,y)=\sum_n g_n(y)g_n^*(x), \tag{3.10}$$

$$d_{\text{in}}(x,y)=\sum_{n,m}\frac{{}_f\langle \text{out}|n,m,\text{in}\rangle_f}{{}_f\langle \text{out}|\text{in}\rangle_f}f_n^*(x)f_m^*(y)$$

$$=i\sum_{n,m}\Lambda_{n,m}f_n^*(x)f_m^*(y), \tag{3.11}$$

[see Eq. (2.33)],

$$d_{\text{out}}(x,y)=\sum_{n,m}\frac{{}_f\langle \text{out},n,m|\text{in}\rangle_f}{{}_f\langle \text{out}|\text{in}\rangle_f}g_n(x)g_m(y)$$

$$=i\sum_{n,m}V_{n,m}g_n(x)g_m(y). \tag{3.12}$$

We are now in position to derive Green's function equations.

Let us consider the action of K_x [defined in (2.4)] on the time-ordered product appearing in (3.1). If we write

$$\begin{aligned}&\Box T(\phi(x)\phi(x_1)\cdots\phi(x_n))\\&=(1/\sqrt{-g})\partial_0\{g^{0\mu}\sqrt{-g}\partial_\mu T(\phi(x)\phi(x_1)\cdots\phi(x_n))\}\\&+(1/\sqrt{-g})\partial_i\{g^{ij}\partial_jT(\phi(x)\phi(x_1)\cdots\phi(x_n))\},\end{aligned} \tag{3.13}$$

we can evaluate this by expanding the time-ordered products in terms of step functions to obtain

$$\begin{aligned}&\Box T(\phi(x)\phi(x_1)\cdots\phi(x_n)\\&\quad=T(\Box\phi(x)\phi(x_1)\cdots(x_n))\\&\qquad+\sum_{j=1}^n\sum_{\text{perms}}\theta(x_{\rho_1}^0-x_{\rho_2}^0)\cdots\delta(x^0-x_{\rho_j}^0)\\&\qquad\times\theta(x_{\rho_2}^0-x_{\rho_{j+1}}^0)\cdots\theta(x_{\rho_{n-1}}^0-x_{\rho_n}^0)\\&\quad\times\phi(x_{\rho_1})\cdots\{(1/\sqrt{-g})\partial_0[g^{00}\sqrt{-g}[\phi(x),\phi(x_{\rho_j})]]\\&\quad+(1/\sqrt{-g})[\pi(x),\phi(x_{\rho_j})]\}\phi(x_{\rho_{j+1}})\cdots\phi(x_{\rho_n}),\end{aligned} \tag{3.14}$$

where the second term arises from the first term in (3.13). Noting that the delta function makes the commutators in (3.14) equal-time commutators, we may use (3.2) and definition (3.1) to obtain

$$\begin{aligned}&K_x\tau(x,x_1,\ldots,x_n)\\&=\langle \text{out}|T(K_x\phi(x)\phi(x_1)\cdots\phi(x_n))|\text{in}\rangle/\langle \text{out}|\text{in}\rangle\\&\quad-\frac{i}{\sqrt{-g_x}}\sum_{j=1}^n\delta(x-x_j)\tau(x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n).\end{aligned}$$

If we now use (2.3) in the first term of this expression we obtain

$$\begin{aligned}K_x\tau(x,x_1,\ldots x_n)=&-\lambda\tau(x,x,x_1\cdots x_n)-a(x)\tau(x_1,..,x_n)\\&-\frac{i}{\sqrt{-g_x}}\sum_{j=1}^n\delta(x-x_j)\tau_j(x_1\cdots x_n),\end{aligned} \tag{3.15}$$

where we have introduced the notation

$$\tau_j(x_1,\ldots x_n)\equiv\tau(x_1\cdots x_{j-1},x_{j+1}\cdots x_n). \tag{3.16}$$

Intergral equations. Our task now is to invert (3.15) to obtain integral equations for the Green's functions. Using Gauss' theorem, and, as usual, assuming the spacetime to be globally hyperbolic with the fields vanishing at spacelike infinity we have

$$\int d^4y\sqrt{-g_y}D_F(x,y)K_yT(\phi(y)\phi(x_1)\cdots\phi(x_n))$$

$$-\int d^4y\sqrt{-g_y}\,[K_yD_F(x,y)]T(\phi(y)\phi(x_1)\cdots\phi(x_n))$$
$$=\int_{\substack{\Sigma\\ y^0\to+\infty}} d\Sigma_y{}^\mu\sqrt{-g_y}\,D_F(x,y)\overset{\leftrightarrow}{\partial}{}^y_\mu \times T(\phi(y)\phi(x_1)\cdots\phi(x_n))$$
$$-\int_{\substack{\Sigma\\ y^0\to-\infty}} d\Sigma_y{}^\mu\sqrt{-g_y}\,D_F(x,y)\overset{\leftrightarrow}{\partial}{}^y_\mu \times T(\phi(y)\phi(x_1)\cdots\phi(x_n)),$$

where we have used the fact that the spacetime is globally hyperbolic with Cauchy surfaces which are either compact without boundaries or noncompact with the fields vanishing at spacelike infinity. The second term on the left-hand side of this expression can be simplified using (3.4), while the right-hand side can be evaluated using Eq. (3.8)–(3.12) for D_F, along with the asymptotic conditions (2.9), (2.10), and the orthonormality conditions for the modes (2.8). We find

$$T(\phi(y)\phi(x_1)\cdots\phi(x_n))$$
$$=-\int d^4y\sqrt{-g_y}\,D_F(x,y)K_yT(\phi(y)\phi(x_1)\cdots\phi(x_n))$$
$$+\left[\sum_n g_n^*(x)B_n^{\text{out}\,\dagger}+i\sum_{n,m}V_{n,m}g_m(x)B_n^{\text{out}\,\dagger}\right]\times T(\phi(x_1)\cdots\phi(x_n))+T(\phi(x_1)\cdots\phi(x_n))$$
$$\times\left[\sum_n f_n(x)A_n^{\text{in}}+i\sum_{n,m}A_{n,m}f_n^*(x)A_m^{\text{in}}\right]. \tag{3.17}$$

Taking the vacuum expectation value of this equation between $\langle\text{out}|\cdots|\text{in}\rangle$ vacua, we see from (2.14) and (2.15) that the second and third terms from (3.17) vanish, leaving

$$\tau(x,x_1\cdots x_n)=-\int d^4y\sqrt{-g_y}\,D_F(x,y)K_y\tau(y,x_1,..,x_n). \tag{3.18}$$

Finally applying (3.15) to the right-hand side of (3.18) we obtain the inverted Green's function equations

$$\tau(x,x_1\cdots x_n)=\lambda\int d^4y\sqrt{-g_y}\,D_F(x,y)\tau(y,y,x_1\cdots x_n)+\tau(x_1\cdots x_n)\int d^4y\sqrt{-g_y}\,D_F(x,y)a(y)+i\sum_{j=1}^n D_F(x,x_j)\tau_j(x_1\cdots x_n). \tag{3.19}$$

The equivalent of this equation for $\lambda\phi^4$ theory is obvious.

It is interesting to note that the disappearance of all but the first term on right-hand side of (3.17), when the vacuum expectation value is taken, is not merely fortuitous, but rather due to the fact that we calculated the Feynman propagator (3.5) using ${}_f\langle\text{out}|$, $|\text{in}\rangle_f$ vacua defined with respect to the same modes as the $\langle\text{out}|$, $|\text{in}\rangle$ vacua, even though the "free" vacua need not necessarily be the same as the "asymptotic interacting" vacua. This simply means that we are being consistent with our definition of positive frequency in the far past and future. If we had not, these extra terms would not have disappeared.

Since the reduction formula (2.38) involves

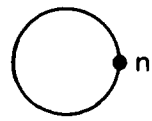
$$\gamma(x_1,...,x_n)\equiv\left[\prod_{j=1}^n iK_{x_j}\right]\tau(x_1\cdots x_n), \tag{3.20}$$

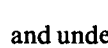
it is worthwhile obtaining equations which can, in principle, be solved for the γ's directly. Such equations are derived in an obvious way using (3.20), (3.19), and (3.4) to give

$$\gamma(x,x_1,...,x_n)$$
$$=-\lambda\int d^4y_1\,d^4y_2\sqrt{-g_{y_1}}\sqrt{-g_{y_2}}\,iD_F(x,y_1)\times iD_F(x,y_2)\gamma(y_1,y_2,x_1,..,x_n)-ia(x)\gamma(x_1,..,x_n)$$
$$+i\sum_{j=1}^n\frac{1}{\sqrt{-g_{x_j}}}K_x\delta(x-x_j)\gamma_j(x_1,...,x_n), \tag{3.21}$$

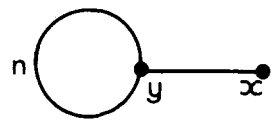
where γ_j is defined in analogy to τ_j [Eq. (3.16)].

For the purposes of the formal renormalization of these equations it is convenient to give them a diagrammatic representation. (See Ref. 13 for the corresponding procedure in flat space.) To this end we let

$\gamma(x_1,...,x_n) =$  , (3.22)

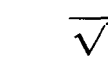
$iD_F(x,y) =$  , (3.23)

and understand that internal vertices are integrated over; for example

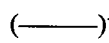
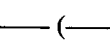


$$=i\int d^4y\sqrt{-g_y}\,D_F(x,y)\gamma(y,x_1,...,x_n). \tag{3.24}$$

If we also denote

$$\frac{i}{\sqrt{-g_y}}K_x\delta(x-y)=(\text{  })^{-1}, \tag{3.25}$$

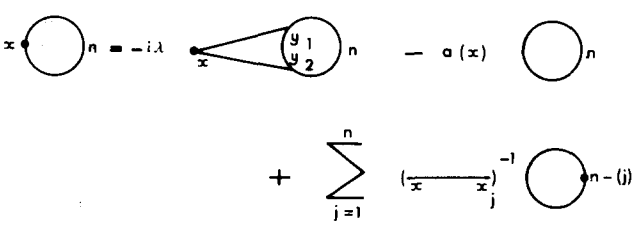
then, with the internal vertex y being integrated over, we have

$$(\text{  })^{-1}\ \text{}=\text{  }\ (\text{  })^{-1}$$

$$=\frac{1}{\sqrt{-g_x}}\delta(x-z)\ , \tag{3.26}$$

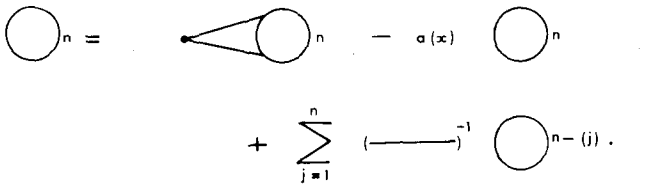
which is the identity (the spacetime volume element being $(-g_x)^{1/2}d^4x$).

With these conventions established we can represent Eq. (3.21) as



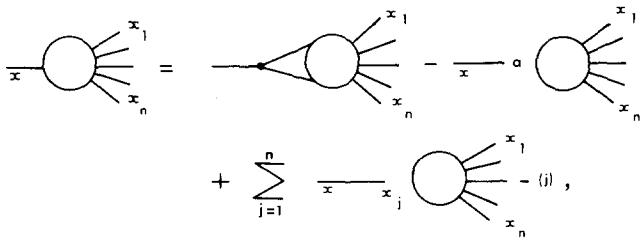
(3.27)

or, suppressing the labeling of vertices and letting the spot where three lines come together represent $-i\lambda$ we have simply



(3.28)

Alternatively it may be easily verified from (3.21) that attaching external propagators to (3.28) gives the expected result

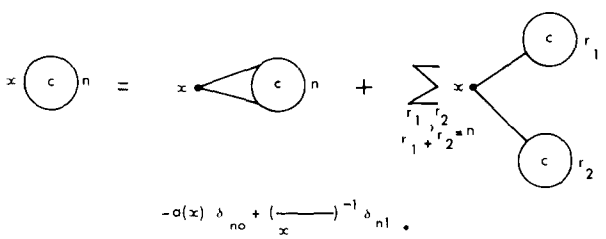


(3.29)

where

$$\text{---}_{x\ a} \equiv i\int \sqrt{-g_y}\, D_F(x,y)a(y)\, d^4y.$$

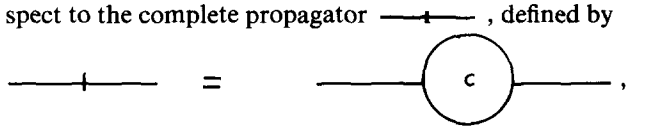
Apart from the term involving $a(y)$, Eq. (3.29) has exactly the same diagrammatic structure as did that for $\lambda\phi^3$ theory in a flat background spacetime.[13] We may thus apply much of the development given in that reference. There is however one major difference in curved spacetime which requires consideration. After application of the connectivity anzatz[13] to (3.29), we obtain equations for the connected Green's functions:



(3.30)

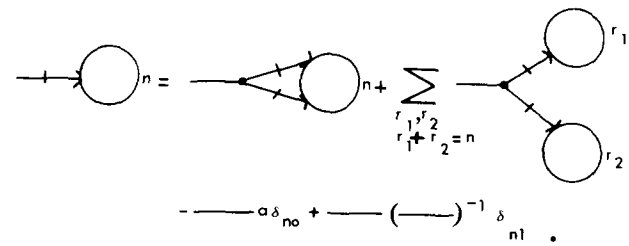
Now r_1 and r_2 in the second term in (3.30) run from zero to n, but in flat spacetime the terms involving r_1 or r_2 zero are either ignored on the grounds of energy momentum conservation, or are removed by normal ordering. Similarly the $n = 0$ Green's function equation is not considered. In curved spacetime we can no longer ignore these contributions as they can possibly give significant physical effects (such as production of a particle from the vacuum). The inclusion of these "tadpole" diagrams alters slightly the formal renormalization[13] of (3.30). We shall thus briefly review this procedure paying particular attention to the differences introduced by the Green's functions with $n = 0$.

As in Ref. 13 we write Eq. (3.30) amputated with respect to the complete propagator ——+—— , defined by



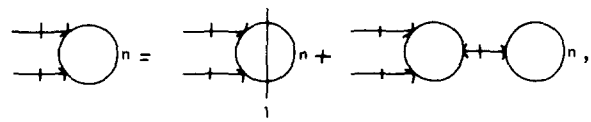
(3.31)

which (henceforth dropping the "c" from connected Green's functions) gives



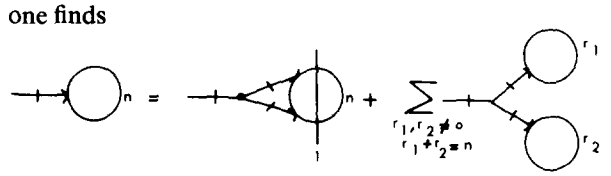
(3.32)

Introducing the one particle irreducible function defined by



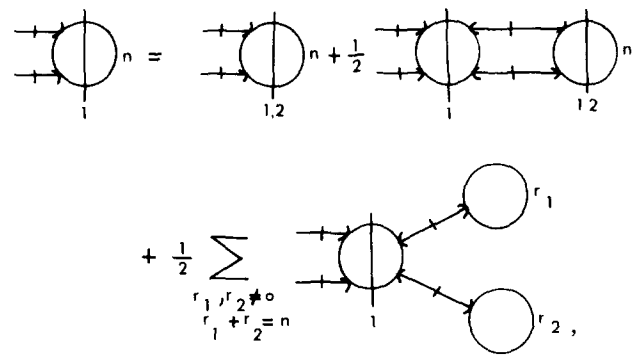
(3.33)

one finds



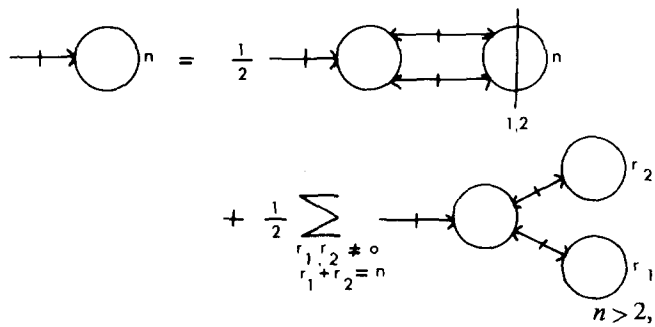
$n > 1,$ (3.34)

where (3.32) for $n = 1$ has been used in the removal of the bare propagators. Notice that this equation contains no explicit "tadpole" terms. Next, introducing the two particle irreducible function by



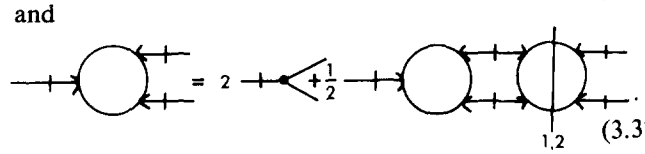
(3.35)

the final term only being included for $n > 2$, one finds



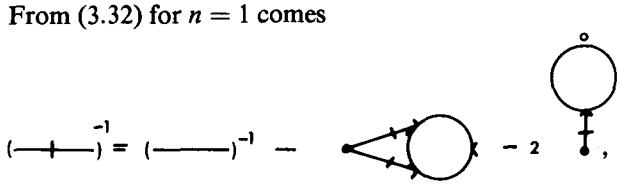
$n > 2,$ (3.36)

and



(3.37)

From (3.32) for $n = 1$ comes

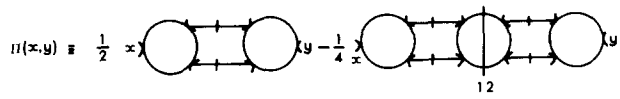


(3.38)

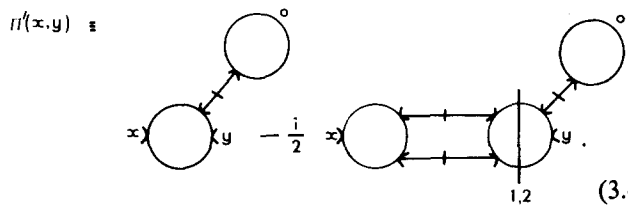
which upon insertion of the bare vertex • from (3.37) gives

$$(\underset{x}{\bullet}\!\!-\!\!\!-\!\!\underset{y}{\bullet})^{-1} = (\underset{x}{\bullet}\!\!-\!\!\!-\!\!\underset{y}{\bullet})^{-1} - \Pi(x,y) - \Pi'(x,y), \quad (3.39)$$

where

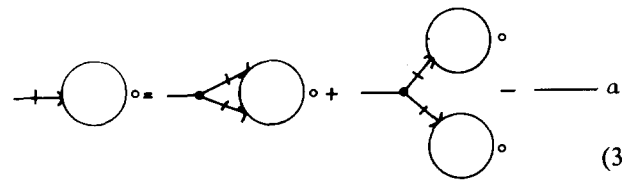


(3.40)



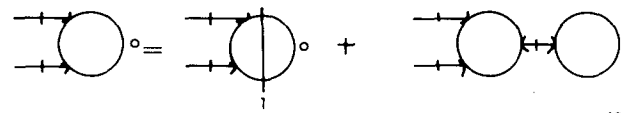
(3.41)

The self-energy, $\Pi + \Pi'$, contains only renormalized objects; that is it does not depend on ——— or •, and in this sense is formally renormalized. Π is the same as the self-energy dealt with in flat spacetime,[13] while Π' is an additional contribution containing $n = 0$ Green's functions (i.e., "tadpoles"). These tadpoles are given by (3.32) for $n = 0$:

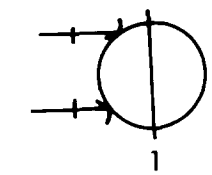


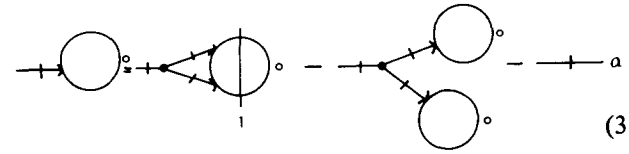
(3.42)

In analogy to (3.33) let



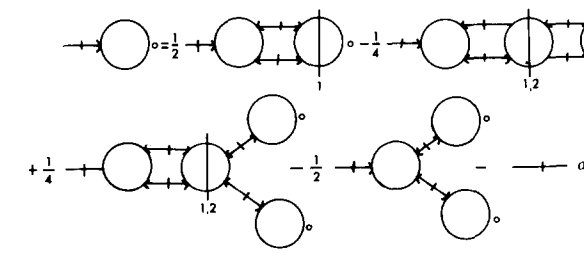
(3.43)

define the object  . Then one finds



(3.44)

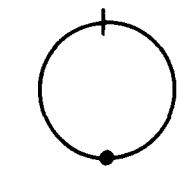
Replacement of the bare vertex in (3.44) using (3.37) gives



(3.45)

which is the formally renormalized equation for the $n = 0$ Green's function. Iteration of (3.36), (3.37), (3.39), and (3.45) allows one to build up perturbation theory in the manner described in Ref. 13, complete with the combinatorial factors obtained in other methods.

Similar renormalizations can be applied to other interactions as was described in Ref. 13, however in most cases there will be additional terms arising in curved spacetime. For example in $\lambda\phi^4$ theory all the renormalized Green's function equations are the same as in Ref. 13 except that in the equation for the inverse propagator there are additional terms arising from a term of the form



(3.46)

in the ϕ^4 equivalent of (3.38). Replacement of the bare vertex in (3.46) results in several new terms in the renormalized equation for the inverse propagator.

The general idea of constructing Green's function equations by the principle of "complete unitarity" (Sec. 5 of paper III of Ref. 13) may also be extended to curved background situations by means of the interpretation of the lines and bubbles given above. We may thus generalize beyond field theories or consider approximations to them.

We have used the adjecting "renormalized" to the equations obtained from (3.29) by exposing certain one and two particle intermediate states. As was stated in Ref. 13 there is no assurance that the resulting equations have finite solutions. We have, therefore, to extend the analysis of Sec. 3 in paper II of Ref. 13 in order to show that such finiteness does indeed result. We will turn to that question now.

4. FINITENESS OF INTERACTING FIELD THEORIES IN FLAT SPACETIME

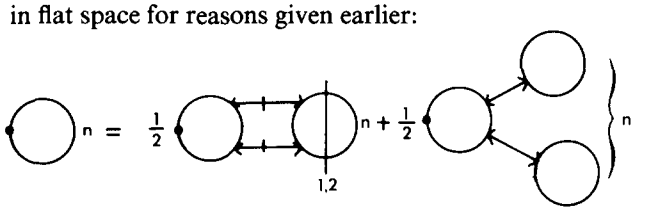
We wish to prove that the renormalized Green's function equations for interacting fields in a given background spacetime have finite solutions. The method used to prove this result in the case of a flat background was by means of iteration of the equations in momentum space.[13] The iteration technique may still be used in the nontrivial background spacetime case but we cannot assume that Fourier transform techniques are applicable without further restrictions (such as asymptotic flatness) on the background. In lieu of making such restrictions we will develop the renormalization argument in position spacetime. Since such an

analysis has not been effected even for flat spacetime at the same level of completeness as for momentum space we will attempt it in this section before extending our results to a curved background spacetime in the next section.

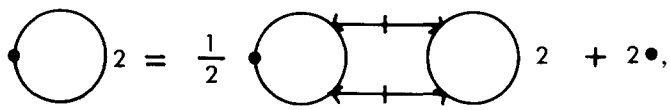
Renormalization of perturbation theory in coordinate spacetime has received considerable attention, most recently by Collecott[23] and earlier by Caianello and collaborators.[24] All of these and earlier discussions in position space analyze Feynman graphs directly, and do not utilize the Green's function equations in the manner of Ref. 13.

We will attempt to adopt this latter method so that it applies in position space. We will do that in detail only for the topologically simplest example of the cubic self-interaction $\lambda\phi^3$, though we consider this theory in both four and six dimensions so as to mimic the divergence content of Q.E.D. and other nontrivial theories in four dimensions. The method in flat space goes through for these latter theories but their added complexity does not seem to add anything useful to our discussion, and we will not consider them further here.

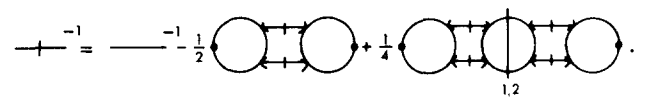
The equations to be discussed are (3.36), (3.37), and (3.39) with the $n = 0$, "tadpole," Green's functions omitted in flat space for reasons given earlier:



$$n > 2, \tag{4.1}$$



(4.2)



(4.3)

In order to construct the most general renormalization approach we will use the distribution techniques associated with the theory of products of distributions[25] as related to products of free propagators.[26] Thus we wish to define the products of distributions on the right-hand side of (4.1), (4.2), and (4.3) initially on a subspace of the space of all test functions. We will then extend this definition to the whole space by the Hahn-Banach theorem in the usual fashion. We hope to show that the ambiguities which arise in such an extension correspond exactly to mass and charge renormalization effects.

We will begin by considering the simplest divergence arising in (4.3), the self-energy bubble $[D_F(x-y)]^2$, where $D_F(x) = (x^2 - iO^+)^{-1}$ in four dimensions (the massive propagator has the same light-cone singularity). We consider the space of test functions $\phi(x)$ for which $D_f(x)\phi(x)$ is suitably smooth to act as a test function for $D_F(x)$.

We will simplify our analysis even further by making a formal analytic continuation to Euclidean spacetime. We thus reduce the singularities of $D_F(x)$ from the light cone to the origin. We appeal to general theorems[27] that indicate that analytic continuation of Green's functions back to the Lorentz metric should be possible at least in principle. Thus our renormalization will explicitly be of the Euclideanized field theory.

We may write formally

$$\langle D_F^2,\phi\rangle = \int [D_F(x)]^2\phi(x)\,d^4x = \int_0^\infty \frac{dr}{r}\int d\Omega\phi(r,\Omega), \tag{4.4}$$

where the usual spherical polar coordinates have been introduced in right-hand side of (4.4). Thus the left-hand side of (4.4) is finite if $|\phi(r,\Omega)| = O(r)$ as $r\to 0$. We may write any function $\phi(x)$ in terms of a function $\alpha(x)$, equal to 1 at the origin and suitably decreasing for large r, as

$$\phi(x) = \alpha(x)\phi(0) + \psi(x),$$

where $|\psi(r,\Omega)| = O(r)$ as $r\to 0$. Then

$$\langle D_F^2,\phi\rangle = \phi(0)\langle D_F^2,\alpha\rangle + \langle D_F^2,\psi\rangle. \tag{4.5}$$

The second term on the right-hand side or (4.5) is finite, as has already been shown. We take $\langle D_F^2,\alpha\rangle$ to be an arbitrary constant c. The Hahn–Banach extension of D_F^2 from the subspace $S = \{\phi : |\phi(r,\Omega)| = O(r) \text{ as } r\to 0\}$ to all (suitably differentiable) functions will thus be

$$D_F^2 = c\delta(x) + (D_F^2)_{\text{finite}}, \tag{4.6}$$

where

$$\langle (D_F^2)_{\text{finite}},\phi\rangle = \langle D_F^2,(\phi - \phi(0)\alpha)\rangle.$$

Clearly (4.6) is the position-space analog of the usual p-space representation.[13]

Before we turn to other Green's functions let us consider the problem of defining D_F^2 in six dimensions, so that in the massless case, $D_F(x) = (x^2)^{-2}$. Then

$$\langle D_F^2,\phi\rangle = \int_0^\infty r^{-3}\,dr\int d\Omega\,\phi(r,\Omega). \tag{4.7}$$

In order for the right-hand side of (4.7) to be finite we require

$$\phi(0) = \phi'(0) = \phi''(0) = 0,$$

where the prime on ϕ denotes any of the partial derivatives. We extend the expansion of a test function to be

$$\phi(x) = \alpha(x)\phi(0) + \beta^\mu(x)\partial_\mu\phi(0) + \gamma^{\mu\nu}(x)\partial_\mu\partial_\nu\phi(0) + \psi(x),$$

where

$$\alpha(0) = 1, \quad \partial_\mu\alpha(0) = \partial_\mu\partial_\nu\alpha(0) = 0,$$
$$\beta^\mu(0) = 0, \quad \partial_\nu\beta^\mu(0) = \delta^\mu_\nu, \quad \partial_\lambda\partial_\sigma\beta^\mu(0) = 0,$$
$$\lambda^{\mu\nu}(0) = 0, \quad \partial_\lambda\lambda^{\mu\nu}(0) = 0,$$
$$\partial_\lambda\partial_\sigma\lambda^{\mu\nu}(0) = \tfrac{1}{2}(\delta^\mu_\lambda\delta^\nu_\sigma + \delta^\nu_\lambda\delta^\mu_\sigma),$$

so that

$$\psi(0) = \partial_\mu\psi(0) = \partial_\mu\partial_\nu\psi(0) = 0.$$

Then

$$\langle D_F^2,\phi\rangle = \phi(0)\langle D_F^2,\alpha\rangle + \partial_\mu\phi(0)\langle D_F^2,\beta^\mu\rangle + \partial_\mu\partial_\nu\phi(0)\langle D_F^2,\gamma^{\mu\nu}\rangle + \langle D_F^2,\psi\rangle. \tag{4.8}$$

By similar arguments to those used in deriving (4.6) from (4.5), together with the absence of any constant vector for $\langle D_F^2, \beta_\mu \rangle$ and the only possibility for $\langle D_F^2, \gamma_{\mu\nu} \rangle = \eta_{\mu\nu}$, we have

$$D_F^2 = c\delta(x) + d\,\Box\delta^4(x) + (D_F^2)_{\text{finite}}, \tag{4.9}$$

where

$$\langle (D_F^2)_{\text{finite}}, \phi \rangle = \langle D_F^2, \phi - \alpha\phi(0) - \beta^\mu \partial_\mu \phi(0) - \gamma^{\mu\nu}\partial_\mu\partial_\nu\phi(0) \rangle.$$

Let us next look at the vertex function, which in lowest order is $D_F(x-y)D_F(y-z)D_F(z-x)$. In four dimensions this is finite, but in six dimensions we find

$$\int d^6(x-y)\, d^6(y-z)\, D_F(x-y)D_F(y-z) \times D_F(z-x)\phi(x-y, y-z) = \int_0^\infty dr_1\, dr_2\, r_1 r_2 (r_1 + r_2)^{-4} \times \int d\Omega_1\, d\Omega_2\, \phi(r_1, \Omega_1, r_2, \Omega_2). \tag{4.10}$$

The possible logarithmic divergence in (4.10) can be prevented by working with test functions $\phi(x-y, y-z)$ for which $\phi(0,0) = 0$. We may again write in the general case

$$\phi(x_1, x_2) = \alpha(x_1 x_2)\phi(0,0) + \psi(x_1 x_2),$$

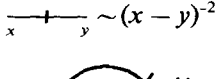
where $\alpha(0,0) = 1$, so that $\psi(0,0) = 0$. Thus by a similar argument as before,

$$D_F(x-y)D_F(y-z)D_F(z-x) = c'\delta(x-z)\delta(x-y) + [D_F(x-y)D_F(y-z)D_F(z-x)]_{\text{finite}}, \tag{4.11}$$

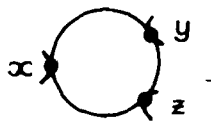
which is again an x-space version of a well-known p-space result.

We now turn to analyze the clothed Green's function equations (4.1), (4.2), and (4.3). We do that by choosing a trial position space behaviour for the Green's functions on the right-hand side of these equations and show that this behavior is possessed by left-hand side.

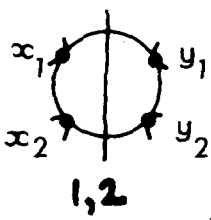
Let us consider the trial singularity behavior for the various Green's functions and one and two particle irreducible functions, the later being given by (3.33) and (3.35). For four dimensions we have



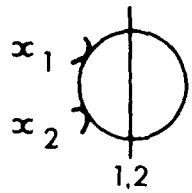
$$\sim (x-y)^{-2} + O(1), \tag{4.12}$$



$$\sim \delta(x-y)\delta(x-z) + O([((x-y)(y-z)(z-x))^{-2}]_f), \tag{4.13}$$



$$\sim \delta(x_1 - y_1)\delta(x_2 - y_2)(x_1 - x_2)^{-2} + (y_1 \longleftrightarrow y_2) + O(1), \tag{4.14}$$

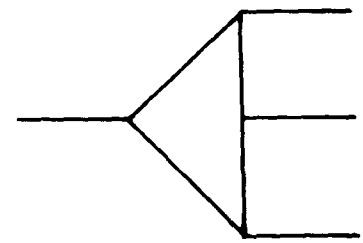


$$\sim \sum_{\substack{(i,j,k)\ \text{dinstinct} \\ (i,j,l)\ \text{distinct}}} \delta(x_1 - y_i)\delta(x_2 - y_j) \times (x_1 - y_k)^{-2}(x_2 - y_l)^{-2} + O(1), \tag{4.15}$$

with the extension to six dimensions being obvious. In (4.15) the singular behavior has been exposed in its dependence on the variables x_1 and x_2. There will be singular behavior in the other variables, as given by the one and two particle structures exposed on the right-hand side of (3.33) and (3.35), but that is not relevant to the renormalization problem of (4.1)–(4.3). We may then use (3.33), (3.35) to generate the singular behavior of the remaining Green's functions not specified in (4.12)–(4.15). In particular for $= 2$, (3.35) is to be regarded as an integral equation in  , while the right-hand side of (4.33) and (4.35) for $n > 2$ then describes the resulting singular behavior of  and  in coinciding variables.

We will consider the removal of coincident position-space singularities from (4.1)–(4.3) exactly as described in Refs. 26 and 13. Thus each of the equations is considered as a distribution on that subspace of test functions on which the right-hand sides of the equations are well defined. The extension to the space of all test functions will then involve the introduction of ambiguities along the lines of those in (4.6), (4.9), and (4.11). Since the distribution-theoretic analysis has already been performed for the momentum space variable equations in Ref. 13 we will dispense with all similar discussions in position space except for the details referring to the position-space singularity structures. We turn to that now.

The singularity structures of (4.12)–(4.15) have to be substituted into the right-hand side of (4.1)–(4.3). Equations (4.12), (4.13), and (4.15) ensure that the integral in the first term on the right-hand side of (4.1) is finite, for the relevant singularity structure is that arising at most from the product of free propagators described by the box diagram

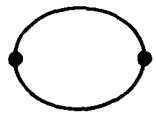


By methods similar to those used in discussing (4.10) it is easily seen that this is finite for all test functions, even in six dimensions. We have thus removed all of the coincident singularities in the propagator and vertex function equations, (4.3) and (4.2), respectively; this is identical to the situation arising in momentum space.[13]

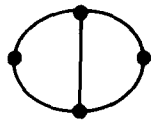
Use of (4.13) and (4.14) in the vertex function equation (4.2), together with our earlier discussion of (4.10), indicates that in four dimensions the right-hand side of (4.2) is finite for a test function. In six dimensions there will be an ambiguity arising from the product of (4.10), so giving an arbitrary addition to the bare coupling constant term [the second term

on the right-hand side of (4.2)]. The behavior of (4.13) is preserved by the right-hand side of (4.2), thus ensuring consistency for that case.

The propagator equation (4.3) has singular products arising from the assumed behaviors (4.12), (4.13), and (4.17) corresponding to the graphs

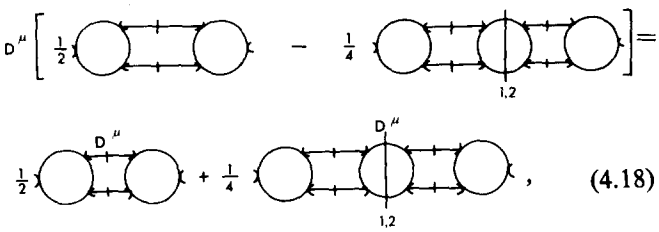


(4.16)



(4.17)

Since only the first of these is singular in four dimensions, in that case we only have the ambiguity of the form (4.6) in (4.3). In the case of six dimensions it is necessary to remove the overlapping divergences present in the last two terms of (4.3). We do that essentially by transforming the momentum space methods developed in Ref. 13, p. 916 to position space. To that we denote multiplication by $(x-t)^{\mu}$ of a function of $(x-t)$ by D^{μ}. Then we can write, by use of D^{μ} applied to (4.2),



(4.18)

where the product has been effected only on the internal lines of the irreducible function in the appropriate diagrams. To obtain (4.18) D_{μ} applied to (4.2) has been used. Then the expression corresponding to (4.17) is

$$\int d^6y\, d^6z\, d^6(x-t)\,[(x-y)(x-z)(y-z)(y-t)\times(z-t)]^{-4}\phi(x-t). \tag{4.19}$$

If we take $\phi(0)=\phi'(0)=\phi''(0)=0$, and let one power of $(x-t)$ acting on the right-hand side of (4.3) be split as in (4.18) we find that (4.19) reduces to expressions of the form

$$\int d^6y\, d^6z\, d^6(x-t)\,[(x-y)(x-z)(y-z)(y-t)(z-t)]^{-4}\times(x-y)^l(y-z)^m(z-t)^n\psi(x-t), \tag{4.20}$$

where $m\geqslant 1$, $l+m+n=3$, and ψ is also a test function. All terms of the form (4.20) are finite, as may be seen by inspection. Extension of (4.19), and also of (4.18), to the space of all test functions gives the ambiguities described earlier. Our method also ensures that (4.12) is preserved by the right-hand side of (4.3) [modified by (4.18)].

Finally we note that (4.14)(and (4.15) are also preserved by (4.1)–(4.3) and (3.33), (3.35), (4.18), since the former are the relevant one-particle singularities of the left-hand sides, which are generated by right-hand side of (3.33) and (3.35).

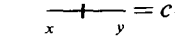
We have thus shown that field equations for $\lambda\phi^3$ are renormalizable, with only mass and charge effects in a spacetime of six dimensions, there only being mass effects in four dimensions.

5. CONSIDERATION OF THE FINITENESS OF $\lambda\phi^3$ IN CURVED SPACETIME

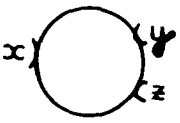
In this section we wish to extend the methods of the previous section to a curved background spacetime. The only obvious point of difference from flat spacetime is that we must include the "tadpole" Green's functions. However we shall see that there are other differences, brought about by the loss of Poincaré invariance, which can have a possibly disastrous effect on the finiteness of the theory. To see where this effect arises, let us closely follow the flat spacetime procedure and see how far we can get.

As in the previous section we consider $\lambda\phi^3$ theory in four and six dimensions, although we shall concentrate on the six dimensional case, which has singularity structure similar to four dimensional Q.E.D., and gives rise to the effect mentioned above. We shall also consider the spacetime to be analytically continued so as to have Euclidean signatured metric $g_{\mu\nu}$. Consideration of Euclidean spacetimes rather than their Minkowskian counterparts should not affect the renormalization of the field theory. However, since the topological properties of the spacetime might be different in its Euclideanized form[28] even this is not guaranteed.[6]

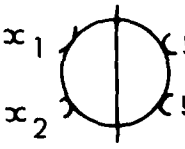
As in flat spacetime we postulate trial singularity behavior for the various Green's functions and one and two particle irreducible functions, and show that this behavior is preserved under iteration of the equations. In six dimensions this behavior is



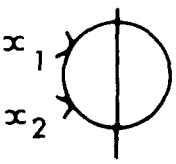
$$= c_1\sigma^{-2}(x,y) + O(\sigma^{-1}), \tag{5.1}$$



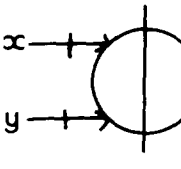
$$= c_2 \frac{\delta(x-y)}{\sqrt{g_y}} \frac{\delta(x-z)}{\sqrt{g_z}} + O([\sigma(x,y)\sigma(x,z)\sigma(y,z)]_f^{-2}), \tag{5.2}$$



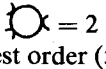
$$= c_3 \frac{\delta(x_1-y_1)}{\sqrt{g_{y_1}}} \frac{\delta(x_2-y_2)}{\sqrt{g_{y_2}}} \times[\sigma^{-2}(x_1,x_2) + O(\sigma^{-1})] + (y_1 \longleftrightarrow y_2), \tag{5.3}$$



$$= c_4 \sum_{\substack{(i,j,k)\text{ distinct} \\ (i,j,l)\text{ distinct}}} \frac{\delta(x_1-y_i)}{\sqrt{g_{y_i}}} \frac{\delta(x_2-y_j)}{\sqrt{g_{y_j}}} \times\sigma^{-2}(x_1,y_k)\sigma^{-2}(x_2,y_l) + \text{less singular}, \tag{5.4}$$



$$= c_5\sigma^{-2}(x,y) + O(\sigma^{-1}), \tag{5.5}$$

where $\sigma(x, y)$ is one half of the square of the geodesic distance between x and y. The trial behaviors (5.1)–(5.5) are seen to be correct in lowest order, if one notes that in that case —•— = ——— $= iD_F(x, y)$, which has the behavior (5.1) as can be seen from the DeWitt–Schwinger expansion[11,12] (in six dimensions), and  $= 2 \cdot$, which certainly has behavior (5.2). Then in lowest order (5.3)–(5.5) are formed from these lowest-order quantities according to their definitions, and are seen to have the behaviors given above. It remains to be shown that these trial behaviors are preserved under iteration. We also have not included any mention of the behavior of the "tadpole" Green's functions in the above. These are obviously a special case and will be returned to later.

As in flat spacetime, substitution of (5.1), (5.2), and (5.4) into the right-hand side of (3.36) shows that the Green's functions for $n > 2$ are finite, and preserve the singular behavior (5.4). We thus turn to the case $n = 2$ and so to Eq. (3.37). Writing the second term on the right-hand side of (3.37) as $G(x, y, z)$, we see that substitution of the singular parts of (5.1), (5.2), (5.3) gives the only part of G to give a divergence. This has the structure

$$c_1^2 c_2 c_3 [\sigma(x, y)\sigma(x,z)\sigma(y,z)]^{-2}, \tag{5.6}$$

in six dimensions. In four dimensions the vertex function is finite, and will not be considered further.

As in flat spacetime we wish to define G on suitable test functions. It is convenient though not essential, to define a test function "about x" by

$$\psi_x(y^\mu, z^\mu) = \Psi(\sigma^\mu(x, y), \sigma^\nu(x,z)), \tag{5.7}$$

where Ψ is a test function in eight variables and

$$\sigma^\mu(x, y) = \sigma^{;\mu}(x, y) \tag{5.8}$$

is the tangent vector at x to the geodesic from x to y, and has length equal to the geodesic distance from x to y.[11,12] Then G is defined on such test functions about x which satisfy

$$\psi_x(x,x) = 0, \tag{5.9}$$

where the action of G on the test function is defined by

$$\langle G, \psi_x \rangle = \int \sqrt{g_x}\sqrt{g_y}\, G(x, y,z)\psi_x(y,z)\, d^6y\, d^6z. \tag{5.10}$$

We extend the definition to arbitrary test functions $\phi_x(y, z)$ about x by writing (as in flat spacetime)

$$\phi_x(y,z) = \alpha_x(y,z)\phi_x(x,x) + \psi_x(y,z), \tag{5.11}$$

where

$$\alpha_x(x,x) = 1, \tag{5.12}$$

which implies that ψ_x satisfies (5.9). Then

$$\langle G, \phi_x \rangle = \langle G, \alpha_x \rangle \phi_x(x,x) + \langle G, \psi_x \rangle, \tag{5.13}$$

where $\langle G, \psi_x \rangle$ is finite and $\langle G, \alpha_x \rangle$ is an arbitrary dimensionless scalar function. Since ϕ_x is arbitrary we have

$$G(x, y,z) = \langle G, \alpha_x \rangle \frac{\delta(x - y)}{\sqrt{g_y}} \frac{\delta(x - z)}{\sqrt{g_z}} + G^{\text{finite}}(x, y,z). \tag{5.14}$$

It is at this stage that we strike the first major difference from the flat spacetime analysis; for while in flat spacetime Poincaré invariance implies that $\langle G, \alpha_x \rangle$ is a constant, allowing it to be absorbed into coupling constant renormalization in (3.37), in curved spacetime there may be an enormous range of dimensionless, scalar functions all of which must be considered as candidates for (5.14). To make further progress we must decide which of these functions can actually appear in perturbation theory. In order to achieve this we must not only specify trial singularity behavior in (5.1)–(5.3), but also trial metric dependence of the singular terms. Then, an well as showing that the singularity behavior is preserved under iteration, we must also show that the metric dependence is preserved. In fact in writing c_1–c_3 as constants in (5.1)–(5.3) we have presupposed the metric dependence of the singular parts of these functions, and this is indeed the metric dependence at lowest order of iteration.

Equation (5.14) establishes that the singularity structure in (5.3) is maintained under iteration, so we must next decide whether the metric dependence is also preserved. That is, we must determine whether $\langle G, \alpha_x \rangle$ is a constant. To this end we consider the singular part (5.6) of G acting on a test function about x satisfying (5.9):

$$c_1^2 c_2 c_3 \int \sqrt{g_x}\sqrt{g_y}\, [\sigma(x, y)\sigma(x,z)\sigma(y,z)]^{-2}\psi_x(y,z) \times d^6y\, d^6z. \tag{5.15}$$

We consider the evaluation of (5.15) in normal coordinates y^μ, z^μ at x,[29,30] in which case

$$\begin{aligned} &\sigma^\mu(x, y) = y^\mu, \quad \sigma^\mu(x,z) = z^\mu, \\ &\sigma(x,z) = \tfrac{1}{2}\eta_{\mu\nu}z^\mu z^\nu \equiv \tfrac{1}{2}z^2, \\ &\sigma(y,z) = \tfrac{1}{2}\eta_{\mu\nu}(y^\mu - z^\mu)(y^\nu - z^\nu) + O((y - z)^4), \\ &\sqrt{g_y} = 1 + O(y^2), \quad \sqrt{g_z} = 1 + O(z^2) \end{aligned} \tag{5.16}$$

($\eta_{\mu\nu} = -\delta_{\mu\nu}$ in the Euclideanized spacetime), and noting (5.7), (5.15) is

$$16c_1^2 c_2 c_3 \int y^{-4}z^{-4}(y - z)^{-4}\Psi(y,z)\, d^6y\, d^6z + \text{finite terms}. \tag{5.17}$$

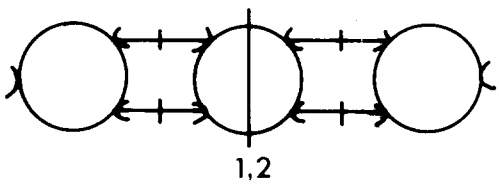
In the extension to arbitrary test functions it is only the integral shown in (5.17) which contributes to $\langle G, \alpha_x \rangle$, it being the only "infinite" piece. However this piece is not explicitly dependent on the metric, and in fact is exactly the same as in flat spacetime. We may thus take $\langle G, \alpha_x \rangle$ to be a constant, allowing its absorption into coupling constant renormalization, and hence showing that the metric dependence (5.3) is preserved under iteration.

Thus, in the vein of the discussion in the Introduction, the divergences in the vertex function only probe the local neighborhood of the point x, and do not sense the less local properties of the spacetime geometry. This is just as one would expect. We shall now see that this is not necessarily the case with the self-energy given by (3.40) and (3.41).

Let us suppose that we have already dealt with the $n = 0$ Green's function equations, so that those appearing in (3.41) are finite. There seems to be an immediate problem with the second term on the right-hand side of (3.41) which

is the divergent G, just considered, combined with the finite part of a "tadpole." We have just seen that the divergent part of G is constant, but the finite part of the tadpole can have very complicated dependence on the metric, even in the first iteration. Thus, at first sight it seems that (3.41) is going to produce divergences which cannot be absorbed into renormalization of the constants m and ξ which appear in (3.39) [see Eqs. (3.25) and (2.4)]. However we show in the Appendix that the divergences involving the "tadpoles" which arise from the second term in (3.41) exactly cancel similar ones which arise from the first term in (3.40) under iteration. Since the first term in (3.41) is finite we now need only consider the divergences arising in Π, given by (3.40).

As in flat spacetime the second term in (3.40) contains overlapping divergences, which can have a possibly ruinous effect. This potential threat arises from the product of the logarithmic divergences in "one half" of



(5.18)

multiplying the finite part of the "other half," which can have very complicated metric dependence. If such a product is not canceled, then once again we have the possibility of infinities which cannot be removed by mass or ξ renormalization. However, this is no more than the overlapping divergence problem already dealt with in flat spacetime, and can be treated in a similar way in curved spacetimes by multiplying (3.41) by $(x-y)^{\mu}$. We prefer to work in a slightly more obviously covariant fashion by multiplying (3.41) by $\sigma(x,y)$. If we let

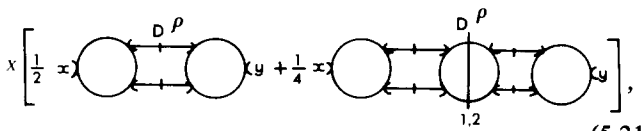
$$\sigma^{\mu}(x,y) = (x-y)^{\nu}\Sigma^{\mu}_{\nu}(x,y), \tag{5.19}$$

where

$$\Sigma^{\mu}_{\nu}(x,y) = \delta^{\mu}_{\nu} - \tfrac{1}{2}\Gamma^{\mu}_{\rho\nu}(x-y)^{\rho} + O\left((x^{\rho}-y^{\rho})(x^{\omega}-y^{\omega})\right), \tag{5.20}$$

and allow $(x-y)^{\nu}$ to act on the diagrams in Π as in flat space we obtain

$$\sigma(x,y)\Pi(x,y) = \tfrac{1}{2}g_{\mu\nu}(x)\sigma^{\mu}(x,y)\Sigma^{\nu}_{\rho}(x,y)$$



(5.21)

where, as in flat spacetime $D^{\rho}f(x,y) = (x-y)^{\rho}f(x,y)$. Substitution of the singular behaviors (5.1)–(5.3) into (5.21) shows that $\Pi(x,y)$ is defined on test functions $\psi_x(y)$ about x (i.e., $\psi_x(y) = \Psi(\sigma^{\mu}(y))$; Ψ is a test function) satisfying

$$\begin{aligned} &\psi_x(x) = 0, \\ &[\nabla^{\mu}_y\psi_x(y)]_{y=x} = 0, \\ &[\nabla^{\mu}_y\nabla^{\nu}_y\psi_x(y)]_{y=x} = 0. \end{aligned} \tag{5.22}$$

In fact we can represent such a test function as

$$\psi_x(y) = \sigma(x,y)\bar{\psi}_x(y),$$

where $\bar{\psi}_x$ is a test function about x satisfying $\bar{\psi}_x(x) = 0$. Then letting σ act on Π as in (5.21) we see that

$$\langle \Pi,\psi_x\rangle \equiv \int \sqrt{g_y}\,\Pi(x,y)\psi_x(y)\,d^6y \tag{5.23}$$

will be finite.

We extend to arbitrary test functions $\phi_x(y)$, about x by writing

$$\begin{aligned} \phi_x(y) = \alpha_x(y)\phi_x(x) &+ \beta^{\mu}_x(y)[\nabla^z_{\mu}\phi_x(z)]_{z=x} \\ &+ \gamma^{\mu\nu}_x(y)[\nabla^z_{\mu}\nabla^z_{\nu}\phi_x(z)]_{z=x} + \psi_x(y), \end{aligned} \tag{5.24}$$

where

$$\alpha_x(x) = 1, \quad [\nabla^{\mu}_z\alpha_x(z)]_{z=x} = [\nabla^{\mu}_z\nabla^{\nu}_z\alpha_x(z)]_{z=x} = 0, \tag{5.25}$$

$$\begin{aligned} &\beta^{\mu}_x(x) = 0, \quad [\nabla^{\mu}_z\beta^{\nu}_x(z)]_{z=x} = g^{\mu\nu}(x), \\ &[\nabla^{\lambda}_z\nabla^{\rho}_z\beta^{\mu}_x(z)]_{z=x} = 0, \end{aligned} \tag{5.26}$$

$$\begin{aligned} &\gamma^{\mu\nu}_x(x) = 0 = [\nabla^{\lambda}\gamma^{\mu\nu}_x(z)]_{z=x}, \\ &[\nabla^{\lambda}\nabla^{\rho}\gamma^{\mu\nu}_x(z)]_{z=x} = \tfrac{1}{2}[g^{\lambda\mu}(x)g^{\rho\nu}(x) + g^{\lambda\nu}(x)g^{\rho\mu}(x)], \end{aligned} \tag{5.27}$$

which imply that $\psi_x(y)$ satisfies (5.21). We then obtain

$$\begin{aligned} \Pi(x,y) = &[\langle \Pi,\alpha_x\rangle + \langle \Pi,\beta^{\mu}_x\rangle\nabla^x_{\mu} + \langle \Pi,\gamma^{\mu\nu}_x\rangle\nabla^x_{\mu}\nabla^x_{\nu}] \\ &\times [\delta(x-y)/\sqrt{g_y}] + \Pi_f(x,y). \end{aligned} \tag{5.28}$$

All the quantities in the square brackets are arbitrary apart from the requirements that

$\langle \Pi,\alpha_x\rangle$ is a scalar field of dimension (length)$^{-2}$,

$\langle \Pi,\beta^{\mu}_x\rangle$ is a vector field of dimension (length)$^{-1}$,

and

$\langle \Pi,\gamma^{\mu\nu}_x\rangle$ is a rank-two, dimensionless tensor field.

For the field theory to be renormalizable without the addition of extra counterterms we need

$$\langle \Pi,\alpha_x\rangle = a + bR(x), \tag{5.29a}$$

$$\langle \Pi,\beta^{\mu}_x\rangle = 0, \tag{5.29b}$$

$$\langle \Pi,\gamma^{\mu\nu}_x\rangle = c\,g^{\mu\nu}(x), \tag{5.29c}$$

where a, b, c are constants. That this is the case in flat spacetime is guaranteed by Poincaré invariance, but in curved spacetime we do not have such a restrictive covariance as to eliminate all but the possibilities (5.29), and must proceed as in the case of the vertex function to determine what the metric dependence of these functions might be in perturbation theory. To obtain the metric dependence of $\langle \Pi,\alpha_x\rangle$, which receives contributions from the logarithmic divergences in Π, we must postulate trial metric behavior of the σ^{-1} terms in (5.1) and (5.3). In lowest order the metric dependence of the σ^{-1} term in (5.1) is $[a + bR(x)]\sigma^{-1}(x,y)$, while in (5.3) it is obtained by inserting such a term into the square bracket. Unfortunately it is not possible, for reasons to be discussed shortly, to show that this metric dependence of the σ^{-1} terms is preserved under iteration. In fact, in general it will not be preserved. This means that we are not in a position to determine whether $\langle \Pi,\alpha_x\rangle$ has metric dependence as in (5.29a). We shall return to this problem after we have gleaned all the information that we can from the substitution of the leading singular terms in (5.1)–(5.3) into Π, as given in (5.21).

Such a substitution gives all the quadratically divergent behavior of Π. Let us consider only the quadratic contribution to Π from the second diagram in (5.21), that from the first being simpler. Substitution of (5.1)–(5.3) gives this contribution to Π acting on a test function ψ_x of the type in (5.22) to be

$$\tfrac{1}{8}c_1^4c_2^2c_3\int \frac{\sqrt{g_y}\sqrt{g_z}\sqrt{g_w}\,g_{\mu\nu}(x)\Sigma^{\nu}_{\rho}(x,y)\sigma^{\mu}(x,y)(z-w)^{\rho}\psi_x(y)\,d^6y\,d^6w\,d^6z}{\sigma(x,y)\sigma^2(x,z)\sigma^2(x,w)\sigma^2(y,w)\sigma^2(z,w)\sigma^2(y,z)}. \tag{5.30}$$

Now expanding everything in normal coordinates about x using (5.16), and the fact that $\Sigma^{\nu}_{\rho}(x,y)=\delta^{\nu}_{\rho}$ in these coordinates, we have, retaining only the quadratically divergent term (but noting that there arise no linearly divergent pieces) that

$$\text{Eq. (5.30)} = 256c_1^4c_2^2c_3\int \frac{\eta_{\mu\rho}(x-y)^{\mu}(z-w)^{\rho}\Psi(y)\,d^6y\,d^6z\,d^6w}{(x-y)^2(x-z)^4(x-w)^4(y-w)^4(z-w)^4(y-z)^4} + \text{log divergent terms.} \tag{5.31}$$

This integral contains no explicit metric dependence, and is the same as occurs in flat spacetime. We can extend to arbitrary test functions $\Phi(y)$, writing

$$\Phi(y)=\alpha(y)\Phi(0)+\beta^{\mu}(y)(\partial_{\mu}\Phi)(0) + \gamma^{\mu\nu}(y)(\partial_{\mu}\partial_{\nu}\Phi)(0)+\Psi(y), \tag{5.32}$$

where α,β,γ satisfy relations similar to α_x, β_x,γ_x in (5.25)–(5.27). Then we can write for the kernel, I, of the integral in (5.31), acting on Φ

$$\langle I,\Phi\rangle = \langle I,\alpha\rangle\Phi(0)+\langle I,\beta^{\mu}\rangle\partial_{\mu}\Phi(0) + \langle I,\gamma^{\mu\nu}\rangle\partial_{\mu}\partial_{\nu}\Phi(0)+\langle I,\Psi\rangle, \tag{5.33}$$

where $\langle I,\Psi\rangle$ is finite. Since I is independent of the metric we can consistently take $\langle I,\alpha\rangle=\text{constant}=c'$, $\langle I,\beta^{\mu}\rangle=0$, and $\langle I,\gamma^{\mu\nu}\rangle = c\,\eta^{\mu\nu}$, just as in flat spacetime. The treatment of the quadratically divergent part of the first diagram in (5.21) is similar, and from (5.33) we obtain for the total quadratically divergent part of Π in normal coordinates

$$\Pi^{\text{quad}}(x,y)=(d\,\Box+d')\left[\delta(x-y)/\sqrt{g_y}\,\right]+\Pi_f^{\text{quad}}(x,y). \tag{5.34}$$

Since this is a covariant expression it holds true in all coordinates [the $\sqrt{g_y}$, which, when dividing $\delta(x-y)$, is one in normal coordinates at x has been inserted to give the right-hand side the same coordinate transormation properties as the left-hand side.].

As the quadratically divergent part of Π is the only piece which can give rise to a $\Box\,\delta(x-y)$ term, we have from comparison of (5.28) and (5.34) that (5.29c) does indeed hold. Further, Π^{quad} is the only part which could give rise to a term of the form $\langle\Pi,\beta^{\mu}\rangle$, but we have seen above that this vanishes, so (5.29b) also holds. We also note that modifying Π as in (5.21) has shown the cancellation of all the infinities resulting from the product of a logarithmic divergence in one side of (5.18) with a finite contribution from the other half. This may be seen by considering the replacement of the singular part, σ^{-2}, one of the propagators in (5.30) by a finite part. Then one sees that rather than obtaining an infinity of the type just mentioned, as would have been the case without the modification (5.21), the integral now is finite for any test function about x. However if the σ^{-2} part of one of the propagators in (5.30) is replaced by a σ^{-1} part, then there is still an overall logarithmic divergence which gives a contribution to $\langle\Pi,\alpha_x\rangle$. It is for this reason that we need the metric dependence of the σ^{-1} term in (5.1) and (5.3).

If these σ^{-1} terms maintained their lowest-order metric dependence under iteration it is possible to show by a fairly lengthy calculation in normal coordinates that $\langle\Pi,\alpha_x\rangle$ would satisfy (5.29a), allowing all the divergences in the self-energy to be removed in mass, ξ, and "wave function" renormalization. Before we indicate why in general the σ^{-1} metric behavior will not be preserved under iteration, we show how the σ^{-2} metric dependence is preserved.

Multiplying (3.39) by $(z\text{—}\!\!+\!\!\text{—}x)$ $(y\text{———}w)$ and integrating we have

$$\tau(z,w)=z\text{—}\!\!+\!\!\text{—}w = iD_F(z,w)+\int\sqrt{-g_x}\sqrt{-g_y}\,\tau(z,x)[\Pi_f(x,y) + \Pi_f'(x,y)]iD_F(y,w)\,d^6x\,d^6y, \tag{5.35}$$

where the subscript "f" has been attached to the self-energy to show that it is the finite part. That is, we assume that the metric dependence of the singular parts in (5.1)–(5.3) is such that Π can be made finite by renormalization, and see whether this metric dependence is preserved by (5.35). The lowest-order singularity behavior is given by D_F, which, as already stated has σ^{-2} metric dependence as in (5.1), and σ^{-1} dependence of the form

$$[a+bR(z)]\sigma^{-1}(z,w). \tag{5.36}$$

At higher orders of iteration there will be new σ^{-2} and σ^{-1} terms coming from the integral in (5.35). Such σ^{-2} terms will only arise from the combination of the most singular part of $\tau(z,x)$, which goes like $c_1\,\sigma^{-2}$, with a term containing $\delta(y-w)$ coming from the action of $\Box_y$ in $\Pi_f(x,y)$ $+\Pi_f'(x,y)$ on $D_F(y,w)$. However the only part of Π_f or Π_f' which can involve terms proportional to $\Box$ is Π_f^{quad} given in (5.34), which has such terms just as in flat space. Since Π_f^{quad} does not depend explicitly on the metric, and c_1 is a constant, the most singular part of the integral in (5.35) will go like $c_1'\,\sigma^{-2}$, thus preserving the metric dependence of the σ^{-2} part of (5.1) and, with (3.35), (3.36), and (5.3), the metric dependence of (5.4).

If we now turn to the contribution to the σ^{-1} term in (5.1) from the integral in (5.35), we see, in particular, that there will be such terms arising from the σ^{-2} parts of τ and D_F combined with the parts of Π_f and Π_f' not proportional to $\Box$. Such finite parts of Π can in general have complex metric dependence which can change the metric behavior of the σ^{-1} term from that of the lowest order (5.36).

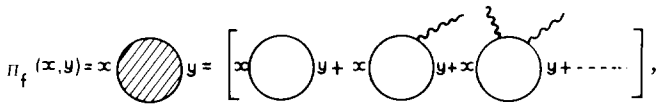
The most obvious modification to the σ^{-1} term in iteration stems from the occurrence of "tadpoles" in Π'. From these, via (5.35), there will arise σ^{-1} terms with metric dependence modified by the complicated metric behavior of the "tadpole". We have, however, shown in the Appendix that the logarithmic infinities caused by these "tadpole"-modified σ^{-1} terms are expected to cancel.

To acquire a feeling for where else modified σ^{-1} behav-

ior can come from, it is convenient to consider the calculation of (5.35) in momentum space. As was stated earlier, momentum space is not wholly appropriate for curved-spacetime quantum-field-theory calculations, however, in particular spacetimes-momentum-space methods have been devised,[31,5] and more recently a means of calculating the divergent parts of Feynman graphs in curved spacetime using momentum space has been proposed.[8,32] For our present purposes momentum space has the advantage of making more transparent the reason for the modification of the σ^{-1} term. Let us write (for justification see Ref. 5)

$$\Pi_f(x,y) = \int \exp[i(p\cdot x - q\cdot y)]\Pi_f(p,q)\, d^6p\, d^6q, \tag{5.37}$$

where, unlike flat space time, $\Pi_f(p,q)$ is a function of two independent momenta as translation invariance does not necessarily apply. Another way of seeing this is to write symbolically



(5.38)

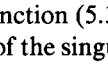
where the quantities in the square brackets are flat space objects with varying numbers of couplings to the external gravitational field, which feeds extra momentum into the system. We can consider a similar expansion for the propagators, which will have a contribution from the zero coupling (flat space) propagators going like p^{-2}. Then the contribution of order σ^{-1} currently under discussion comes from parts of

$$\int \exp[i(p\cdot w - q\cdot z)]p^{-2}q^{-2}\Pi_f(p,q)\, d^6p\, d^6q$$
$$= \int \exp\left[i(p+q)\cdot\frac{(w-z)}{2}\right]p^{-2}q^{-2}\Pi_f(p,q)$$
$$\times\exp\left[i(p-q)\cdot\frac{(w+z)}{2}\right]d^6p\, d^6q, \tag{5.39}$$

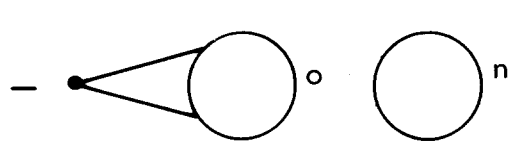
which have $\Pi_f(p,q)$ going like a constant or $\ln(p+q)$ for large $|p+q|$. Such contributions come not only from the first term on the right-hand side of (5.38) but also other terms with couplings to the gravitational field. It is such terms which we are not able to adequately examine within the framework of this paper. On the other hand the σ^{-2} contributions arise from parts of Π_f which go like $(p+q)^2$ for large $|p+q|$, and these arise only from the first term of (5.38) as indicated previously. It seems that the momentum space methods mentioned above could provide a useful tool for investigating this point. In particular, in the Bunch Panangaden method, it would be necessary to ensure that the metric dependence of the p^{-4} term in the momentum space propagator (corresponding to the σ^{-1} term in six dimensional coordinate space) is preserved under iteration. We shall return to discussion of this question in Sec. 7.

We finally turn to renormalization of Eq. (3.45), for $n=0$, "tadpole," Green's functions. We have left this until last because it is markedly different from the $n=1$ and $n=2$ Green's function equations just dealt with. The reason for this is that it cannot be defined on a suitable class of test functions in the same way as the $n=1,2$ cases, but has to be defined by some regularization scheme. We first observe that one can show, by means similar to those used in the Appendix, that the divergence in the third term on the right-hand side of (3.45) cancels against similar divergencies from the first term. We need thus concern ourselves only with the first two terms on the right-hand side of (3.45). These terms confront us with two problems to which we cannot give definitive solutions within the present framework.

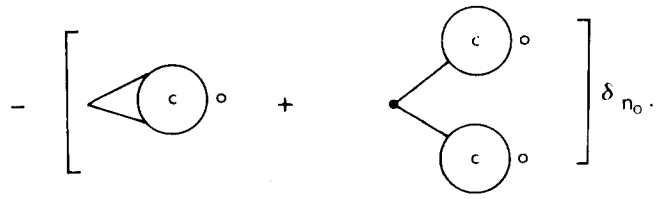
First we note that the singular behavior (5.5) is obviously preserved under iteration, as follows from (3.43) and the fact that (5.1) is preserved under iteration. Then we see that the second term in (3.45) has a severe overlapping divergence problem: The "right-hand half of the diagram" is quadratically divergent, while the "left-hand half" is logarithmically divergent. We cannot treat this overlapping divergence in the same way as we did that in the second term in (3.40), and it is by no means obvious that it will not have the effect of preventing us from absorbing the infinities in (3.45) into a local, geometrical counterterm $a(x)$.

Even if this first problem is overcome there remains a second stumbling block: If we imagine calculating the divergence in the first term of (3.45), there will be a contribution from the delta function (5.3), in the complete vertex, combined with each of the singular parts of  , which, just like the complete propagator, will have σ^{-2}, σ^{-1}, and $\ln\sigma$ singularities. We have already discussed how the metric dependence of the σ^{-1} term senses the finite part of the self-energy, but the $\ln\sigma$ singularity is even more dependent on the metric dependence of this finite part, and there is no reason at all to expect that it will have simple metric structure to all order of iteration. Thus, unless there is a great deal of cancellation between the diagrams in (3.45) it is most unlikely that the infinities can be removed by a geometrical counterterm $a(x)$.

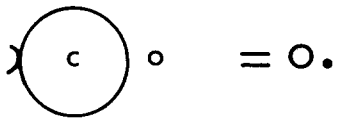
If it happens that (3.45) cannot be made finite by renormalization of a finite number of constants multiplying geometrical objects in $a(x)$, we can ask whether there is any reasonable way of overcoming the problem. The most obvious thing to try is normal ordering, as in flat spacetime. The equivalent of this is to take $a(x) = -\lambda\tau(x,x)$. Then it is easily shown that the term involving $a(x)$ in (3.28) is now



which in terms of connected Green's functions in (3.30) gives the $a(x)$ term to be



Then the $n=0$ Green's function equation is simply



This procedure quite clearly obliterates any problems with the "tadpole" Green's functions, however it means introducing into the Lagrangian a term $\tau(x, x)$, which, from its definition (3.1), depends on the "in" and "out" vacuum states, neither of which is uniquely defined in a general curved spacetime. This would not be a very desirable step to have to take, and the full implications of such a move still have to be analyzed; however should our expectation that the nonlocal infinities do not cancel be confirmed, then this action does offer a means of escape.

We must then conclude this section with a question mark hanging over the finiteness of $\lambda\phi^3$ in curved spacetime. We have not been able to show that the theory is definitely nonrenormalizable, since there will undoubtedly be spacetimes other than Minkowski space in which it is renormalizable. It would be necessary to find one particular spacetime or class of spacetimes in which explicit calculation shows nonrenormalizablility. What we have shown is that the divergences are sufficiently severe as to probe more than just the local structure of spacetime. This is the case with the self-energy in six dimensions, but not in four, where it is only the leading σ^{-1} behavior of the complete propagator which produces divergences; in four dimensions the $n = 1$ Green's function equation is renormalizable. In both four and six dimensions the $n = 0$ Green's function divergences depend on more than the leading divergence; however, as a last resort, such a problem can be cured by the normal ordering mentioned above, making $\lambda\phi^3$ in four dimensions completely renormalizable. It is this dependence of the divergences on more than the leading singularity of the propagator which has prevented us from giving a general proof of renormalizability along the lines proposed. We shall make further remarks on this result, and comment on similar difficulties in $\lambda\phi^4$ theory in Sec. 7.

6. PARTICLE PRODUCTION

Particle production by gravitational fields has received a great deal of attention for free-field theories, but little in the case where interactions are present (see, however, Ref. 4 where low-order perturbation calculations of particle production in the latter case are performed).

Although formulas for the calculation of particle production amplitudes are all contained in Sec. 2, it is worth considering in more detail the production of particles from the vacuum, this being a uniquely curved spacetime (or external field) phenomenon. As we merely wish to demonstrate the principles of particle production, and not perform detailed calculations, we shall restrict our attention to the production of two particles in the final state, one in mode m and one in mode n from out of the initial vacuum. We thus wish to look at the amplitude

$$I_{mn} \equiv \langle \text{out}, \mathbf{m}, \mathbf{n}|\text{in}\rangle / \langle \text{out}|\text{in}\rangle. \tag{6.1}$$

Applying the reduction formula (2.37) we have

$$\begin{aligned} I_{mn} &= iV_{mn} + Q_z Q_x \langle \text{out}|T(\phi(z)\phi(x))|\text{in}\rangle \\ &= iV_{mn} - \int d^3k\, d^3l\, \alpha^{-1}_{km}\alpha^{-1}_{ln} \int \sqrt{-g_x}\sqrt{-g_z} f^*_k(x) \\ &\quad \times f^*_l(z) K_x K_z \tau(z,x)\, d^4x\, d^4z \\ &= iV_{mn} + \int d^3k\, d^3l\, \alpha^{-1}_{km}\alpha^{-1}_{ln} \int \sqrt{-g_x}\sqrt{-g_z} f^*_k(x) \\ &\quad \times f^*_l(z) \gamma(z,x)\, d^4x\, d^4z, \end{aligned} \tag{6.2}$$

in the notation of Eq. (3.20). In writing (6.2) we have adopted continuum rather than discrete normalization as previously.

It is perhaps worth reminding ourselves why I_{mn} vanishes in the usual Minkowski space-scalar-field theory. In this case there is no mode mixing, $\alpha_{ij} = \delta^{(3)}(\mathbf{i} - \mathbf{j}), \beta_{ij} = 0$ and hence from (2.36) $V_{m,n} = 0$. We also have

$$f_k(x) = [(2\pi)^3 2\omega_k]^{-1/2} \exp(-i\omega_k x^0 + i\mathbf{k}\cdot\mathbf{x}), \tag{6.3}$$

where $\omega_k^2 = |\mathbf{k}|^2 + m^2$, and because of Poincaré invariance

$$\tau(z,x) = (2\pi)^{-4} \int d^4p \exp[ip\cdot(z-x)]\tau(p^2),$$

implying

$$\gamma(z,x) = (2\pi)^{-4} \int d^4p \exp[ip\cdot(z-x)]\gamma(p^2), \tag{6.4}$$

with

$$\gamma(p^2) = (p^2 - m^2)^2 \tau(p^2).$$

Substituting (6.3) and (6.4) into (6.2) and performing all but the p integral we have

$$\begin{aligned} I_{mn} = \pi(\omega_m\omega_n)^{-1/2} \int \delta^{(3)}(\mathbf{n} - \mathbf{p})\delta^{(3)}(\mathbf{m} + \mathbf{p}) \\ \times \delta(\omega_n + p^0)\delta(\omega_m - p^0)\gamma(p^2)\, d^4p, \end{aligned} \tag{6.5}$$

which is clearly zero because of the p^0 delta functions (unless the mass is zero, in which case there is an infrared divergence problem).

It is quite obviously time-translation invariance which prohibits the production of particles due to the interaction in Minkowski space. In a spacetime with a time-varying metric however, we cannot rely on such an invariance, and hence the modes (6.3) and the amputated Green's function (6.4) will not have such simple time dependence. In such situations I_{mn} can be nonzero even in the case where there is no mode mixing; $\beta_{ij} = 0$. Several calcuations of particle production in time dependent Robertson–Walker Universes,

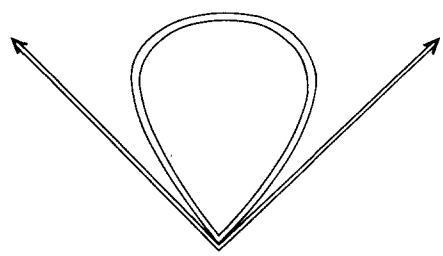


FIG. 1. A first-order particle production diagram.

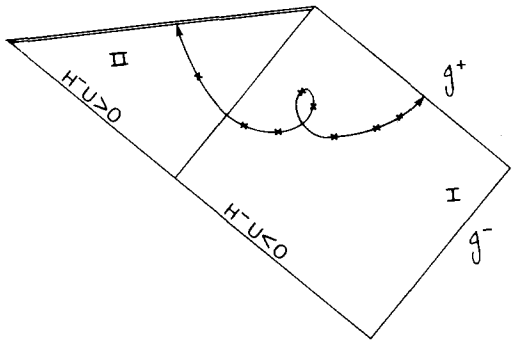


FIG. 2. The Penrose diagram for part of the extended Schwarzschild manifold, with a particle production diagram superimposed.

with and without mode mixing, have been performing by Birrell and Ford[4] in lowest, nontrivial order perturbation theory. It should be noted that while perturbation theory based on (6.2) would use the Feynman propagator (3.5), the calculations in Ref. 4 used

$$iD_F(x,y) = {}_f\langle \text{in}|T(\phi_f(x)\phi_f(y))|\text{in}\rangle_f / {}_f\langle \text{in}|\text{in}\rangle_f. \quad (6.6)$$

Thus, while outgoing lines in Feynman diagrams (e.g., Fig. 1, where the double line is used to denote the propagator (3.5) for later convenience) resulting from (6.2) would represent physical "out" particles, those in Ref. 4 represent "in" particles. To ensure that the correct "out" observables are finally computed Birrell and Ford make use of the Bogolubov transformations (2.19). To a given order in perturbation theory the two approaches will give the same results.

It is interesting to observe that even in the conformally trivial case of a conformally invariant field theory in a conformally flat spacetime, in which case the integral in (6.2) will formally vanish, there can still be particle production because renormalization can break the invariance.

This claim can be substantiated by combining two very elegant pieces of work. The first is by Parker[30] who shows from the conformal trace anomaly that there is no *free* particle production in the conformally trivial case. If one examines how this argument proceeds if instead of the free anomaly one uses the interacting field anomaly as calculated for arbitrary spacetimes (in particular Robertson Walker spacetimes) by Drummond and Shore,[33] one easily sees that Parker's argument will not now go through, and particles may be produced in the interacting case.

Particle production by black holes is one area which, for free fields, has received considerable attention since Hawking's pioneering work.[34] Based on the general consideration of Gibbons and Perry,[35] and the treatment of the particular case of the Thirring model by Birrell and Davies,[10] we can safely assert that the particles produced by interacting fields in a black-hole background will have the spectrum of thermal radiation. We shall thus not attempt to further verify this assertion, or perform actual calculations (which in any case would be most difficult in this case), but rather discuss the way in which the formalism of Sec. 2 can be adapted to a spacetime with a horizon. In particular we shall confine our attention to the examination of (6.2) for a scalar field in a Schwarzschild spacetime.

To this end we use the method originated by Unruh,[21] and subsequently discussed by Hawking[22] of treating the analytically extended Schwarzschild manifold rather than the spacetime of a collapsing star (for a review of this and other relevant topics see Ref. 36). We shall follow as closely as possible the notation of Hawking,[22] and refer the reader to that reference, or any of Refs. 21, 35, or 36, for the physical motivation for the choice of "in" and "out" modes and hence vacua. Thus, corresponding to Hawking's Eq. (4.7), we have Eq. (2.12a), where now we switch to continuum normalization and suppress angular summations to write

$$\phi_{\text{in}} = \int_0^\infty d\omega \sum_i (a_\omega^{(i)\text{in}} f_\omega^{(i)} + \text{H.c.}). \quad (6.7)$$

The "in" modes have been divided into three orthogonal families, $i = 1, 3, 4$ in Hawking's notation. We shall not repeat the detailed specification of these modes, as it is mainly their relation to the decomposition of ϕ_{out} which is important and consequently we write (2.13b) as

$$\phi_{\text{out}} = \int_0^\infty \sum_i (B_\omega^{(i)\text{out}} g_\omega^{(i)} + \text{H.c.}). \quad (6.8)$$

Here we have deviated from Hawking's notation, with his being related to ours by $B_\omega^{(1)} = j_\omega$, $B_\omega^{(3)} = h_\omega$, $B_\omega^{(4)} = g_\omega$, $g_\omega^{(1)} = z_\omega$, $g_\omega^{(3)} = y_\omega$, and $g_\omega^{(4)} = \omega_\omega$. Of the three surfaces, $\mathcal{I}^-$ (past-spacelike infinity), H^- for retarded Krushkal coordinate $U>0$, and for $U<0$, the modes $g_\omega^{(1)}$ have nonzero Cauchy data only on $\mathcal{I}^-$; $g_\omega^{(3)}$ have nonzero Cauchy data only on H^- for $U<0$, while $g_\omega^{(4)}$ have nonzero Cauchy data only on H^- for $U>0$ (see Fig. 2).

From Hawkings Eqs. (4.10), (4.13) (noting that the complex conjugate on the left-hand side of this equation is in error, as is obvious from the remarks preceding it) and (4.14), we can read off the Bogolubov coefficients as defined in our Eq. (2.17) to be

$$\begin{aligned} &\alpha_{\omega,\omega'}^{(1,1)} = \delta(\omega-\omega'), \\ &\alpha_{\omega,\omega'}^{(3,3)} = \alpha_{\omega,\omega'}^{(4,4)} = (1-x)^{-1/2}\delta(\omega-\omega'), \\ &\beta_{\omega,\omega'}^{(3,4)} = \beta_{\omega,\omega'}^{(4,3)} = -(1-x)^{-1/2}x^{1/2}\delta(\omega-\omega'), \end{aligned} \quad (6.9)$$

and all others are zero ($x \equiv e^{-2\pi\omega\kappa^{-1}}$[22]). Then from (2.36) we easily find

$$V_{\omega,\omega'}^{(4,3)} = V_{\omega,\omega'}^{(3,4)} = -i\delta(\omega-\omega')x \quad (6.10)$$

and all other $V^{(ij)}$ are zero. Thus in the case of a free scalar field, in which only the first term in (6.2) contributes, we see that a pair of particles will only be produced if one is created by $B_\omega^{(3)\text{out}\,\dagger}$, thus appearing in region I of the Penrose diagram (Fig. 2), and the other is created by $B_\omega^{(4)\text{out}\,\dagger}$, appearing in region II. This is by now a well known result.[21,22,34,36,37]

Let us now turn to the inclusion of interactions, for which we must examine the integral in (6.2). If we perform the integral using Schwarzschild coordinates then all of the quantities in the integrand have simple Schwarzschild time (t) dependence. In particular

$$\begin{aligned} f_\omega^{(1)} &= e^{-i\omega t}F_\omega^{(1)}(r,\theta,\phi), \\ f_\omega^{(3)} &= e^{-i\omega t}F_\omega^{(3)}(r,\theta,\phi), \\ f_\omega^{(4)} &= e^{+i\omega t}F_\omega^{(4)}(r,\theta,\phi), \end{aligned}$$

and

$$\gamma(z,x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dp \exp[-ip(z^0 - x^0)]\gamma(\mathbf{z},\mathbf{x};p)$$

[see Ref. 38 Eq. (3.11) and thereunder for an explanation of the time dependence of τ and thus γ]. We can now carry out the t integrals in (6.2), and for the combination $f_k^{(i)*}(x)f_l^{(j)*}(z)$ appearing in the integrand we obtain a factor $\delta(k+p)\delta(l-p)$ for $i=j$ or neither of i,j equal to four, and $\delta(k+p)\delta(l+p)$ if one of i,j is equal to four and the other is not. Since $k, l \geqslant 0$ the integral will necessarily be zero unless the latter situation ($i=4, j\neq 4$ or vice versa) arises (in fact with nonzero mass the modes, and hence the integral will vanish for $k=l=0$, avoiding an infrared divergence). Noting that α^{-1} is diagonal, we see that, as in the free-field case, the interaction will only give rise to production of a pair of particles if one occurs in region I and the other across the future horizon (H^+) in region II.

It is clear that the amplitudes for the production of a larger number of particles will be nonzero provided that at least one particle is produced in region II to "pay for" the energy of the particles which escape to $\mathscr{I}^+$.

A number of comments on these results are in order. First we note that we have been careful not to localize the place of production of the pair of particles. This seems to run against the usual understanding of Green's function techniques, which, in perturbation theory, give rise to diagrams such as Fig. 1 showing production as a point. We have not been able to specify this point, only where the external lines "end up." This seems to require a reinterpretation of diagrams like Fig. 1. We can easily give such a reinterpretation if we consider the curved spacetime propagator, (denoted in this section by a double line) as being made up of flat spacetime propagators (denoted by a single line) with varying numbers of interactions with the external gravitational field (denoted by crosses). Such a notion can be made precise by considering the expansion of (3.5) in a Volterra series in the metric about $\eta^{\mu\nu}$, but this will not be necessary for our present purposes. Viewed this way we see that the vertex in Fig. 1 simply represents the point at which the $\lambda\phi^4$ interaction acts, and it is only when this is combined with the infinite number of interactions (at various spacetime points) with the external gravitational field, represented in the external lines of Fig. 1, that the production of two particles in the final state is ensured. Thus, one contribution to Fig. 1 is a diagram which might (as well as any other way) be drawn as in Fig. 2 (its superimposition on the Penrose diagram illustrating its contribution to the Hawking radiation) clearly demonstrating the lack of localizability in the production process.

A second point requiring comment is the interpretation of the S matrix which the reduction formulas and the discussion of this section tell us how to calculate. We have shown that we are able to calculate S matrix elements describing the complete scattering situation, involving regions I and II; however the existence of the event horizon H^+ between the two regions means that an observer can never have complete knowledge of the final state in both of these regions. Thus an observer on $\mathscr{I}^+$ cannot determine the amplitude for the occurrence of n particles in the final state on $\mathscr{I}^+$, but only the probability of such an occurrence, once all the possible accompanying outcomes in region II have been summed over. It was with this in mind that Hawking[22] introduced the superscattering operator, which maps between density matrices rather than pure states, as the operator which determines observations on $\mathscr{I}^+$. We note that an S-matrix formulation of quantum mechanics has not failed in this case, contrary to Ref. 22; the S-matrix elements we predict could, however, only be utilized in region II by the supreme sacrifice.

7. DISCUSSION

In this section we wish to discuss the achievements and shortcomings of this work. While much has been achieved in constructing a rigorous framework for interacting quantum field theory in curved spacetime, we have not been able to achieve our main aim of establishing whether or not $\lambda\phi^3$ in curved spacetime is renormalizable. Let us retrace our steps and see where the failure occurred.

Once formally renormalized Green's function equations had been developed it was an easy matter to show that all the infinities in the vertex could be expressed in a quantity which had a form allowing for its absorption into coupling constant renormalization, apart from the fact that the unknown quantity [see (5.14)] was possibly spacetime dependent. This led us to decide that we must not only show that the singular structure of various Green's functions is preserved under iteration, but also that the metric behavior of these singularities is preserved. The vertex function renormalization in six dimensions depended only on the leading singularity structures of the complete propagator, and we were able to show that the infinity could indeed be absorbed into coupling constant renormalization. There was no renormalization of the vertex in four dimensions. On turning to the self-energy and the renormalization of (3.39) we could once again show that the infinities in six dimensions could be absorbed into three different types of renormalization as in (5.28). It was shown that the metric dependence of two of these quantities depended only on the leading singularity structure of the Green's functions and could be determined to vanish in one case and give wave function renormalization in the other [see (5.29b, c)]. In trying to establish whether the third quantity would give mass and ξ renormalization as in (5.29a), we found that we needed to show that the σ^{-1} dependence of the metric is preserved at each order of iteration. It was pointed out that the "tadpoles," which do not play a role in flat spacetime, immediately change the σ^{-1} dependence, even at the first iteration. However we were able to show that the effect of such σ^{-1} dependence was canceled in the divergences. We showed that there is other modification of the σ^{-1} term coming from the finite part of the self-energy, which is coupled to the external gravitational field. Our method was unable to determine the nature of this modification and so we were unable to verify that the remaining infinities could indeed be absorbed into mass and ξ renormalization, as is the case if the σ^{-1} dependence is unchanged from its lowest order.

We must ask whether this problem is only an inadequacy of our method, or whether it points to the possible nonrenormalizability of $\lambda\phi^3$ theory in curved spacetime.

One can advance arguments in support of the first of these points of view; for even though in Eq. (5.38) the σ^{-1} modification arises from couplings to the external metric, a more detailed study suggests that in fact they only arise from terms with up to second derivatives of the external metric. This strongly suggests that the form (5.36) is preserved, but it would be foolish to make a stronger statement without careful analysis, which we cannot do in the present framework.

Even if the σ^{-1} metric dependence of the complete propagator is preserved it is much less likely that the $\ln \sigma$ metric behavior will retain its lowest-order form. This we saw would be necessary for the renormalization of the $n = 0$ Green's function equation (3.45). We were also unable to show that overlapping divergences in these equations would cancel, thus leaving the possibility of divergences with very complicated, nonlocal, metric dependence which might also lead to the breakdown of the theory. However if we allowed the removal of these terms by normal ordering then in four dimensions ϕ^3 theory was shown to be renormalizable.

What then can be done to firmly establish whether interacting field theories in curved spacetime can become nonrenormalizable? An obvious approach is to find, by explicit calculation, a case in which the renormalization procedure breaks down. As was mentioned in the Introduction such explicit calculations are of great difficulty in curved spacetime. However, since the initiation of the writing of this paper, methods have been developed which bring such calculations from the realm of the near impossible into the domain of laborious computation. These methods make use of momentum space representations of the curved-space Feynman propagator. That of Birrell[31,5] allows the representation of the entire propagator of spacetimes which are asymptotically flat, while that of Bunch and Panangaden,[8] and Bunch and Parker[32] represents only the singular part of the propagator. Such momentum space representations have recently been used to consider the renormalization of $\lambda\phi^4$ theory in curved spacetime.

In this paper we have considered ϕ^3 rather than ϕ^4 theory mainly because of the reduction in the number of diagrams which appear in the Green's function equations. There is, however, a difference between the theories which make $\lambda\phi^4$ an even more likely candidate for nonrenormalizability in curved spacetime. Just as in ϕ^3 theory we had to contend with the occurrence of "tadpole" Green's functions appearing in the self-energy, in ϕ^4 theory we have the appearance of extra terms in the self-energy resulting from (3.46). While in ϕ^3 theory it was possible to show that infinities with complicated metric and state dependence resulting from terms with "tadpoles" canceled between themselves, it is not possible to give a similar proof for the contributions from (3.46). Indeed it can be shown that such infinities will not cancel amongst themselves, but, if they are to cancel, must cancel infinites from diagrams with quite different topology. That this would have to be the case for renormalization to be successfully effected at second-order perturbation theory was pointed out by Bunch, Panangaden, and Parker.[7] It was then shown independently by Birrell,[5] and Bunch and Panangaden,[8] using their respective momentum space methods, that such a cancellation does in fact occur at second order in spacetimes with metric

$$ds_a^2 = \Omega^2(x)\eta_{\mu\nu}. \tag{7.1}$$

That this is the case appears to us to be merely fortuitous, not necessarily continuing to occur at higher orders of perturbation theory and/or in other spacetimes. The reason we cannot be sure of cancellation at higher order is that the metric dependence, or even the numerical coefficient, of the nonleading singular term in the complete propagator, which contributes to the divergence under consideration, can change at each order of iteration. Since the removal of the state dependent infinities depends on exact numerical cancellation between terms, showing that such cancellation occurs at second order, in no way guarantees that it will occur at higher orders. A proof of renormalization to all orders thus remains as work for the future.

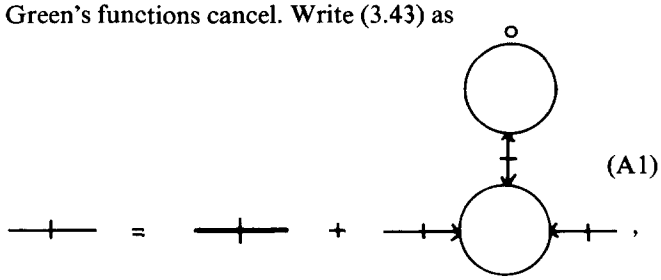
Finally we note that there are numerous other investigations involving interacting quantum fields in curved spacetime which must still be made. Examples of these are studies of the stress tensor, currents and associated anomalies, the renormalization group, operator products, and the spectral representation of propagator. These investigations are best left until after the renormalizability of the underlying theory is firmly established.

Note added in proof: In arriving at the allowed forms of the quantities $\langle \Pi, \alpha_x \rangle < \langle \Pi, \beta_x{}^{\mu} \rangle$ and $\langle \Pi, \gamma_x{}^{\mu\nu} \rangle$ appearing in Eq. (5.28), we have imposed only two constraints, namely: (i) correct dimensionality, and (ii) general covariance. Only one further constraint is needed to ensure renormalizability, this being (iii) locality of these terms. Locality means that the quantities listed above can only depend on the spacetime structure at the point x, and not elsewhere. The only possible local, generally covariant expressions with the correct dimensions are those exhibited in Eq. (5.29). In considering the dimensions of the quantities involved, one must bear in mind that the only dimensional constant that appears in the theory is m, and that only positive powers of m, or $\ln m$ can arise in the self energy, because there is at worst a logarithmic infrared divergence in the terms which are also ultraviolet divergent. We would like to point out that our three conditions above arise naturally from the initial Lagrangian structure, and are to be regarded as analogous to the imposition of Poincaré or gauge invariance in flat spacetime theories.

We hope to discuss this point in more detail elsewhere.

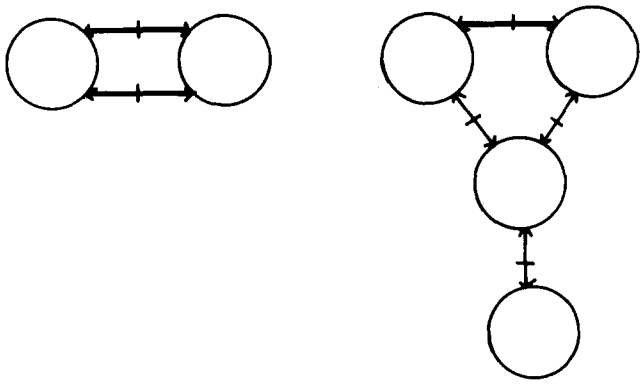
APPENDIX A

It is shown that all the logarithmic divergences in (3.40) and (3.41) containing dependence on the $n = 0$, "tadpole," Green's functions cancel. Write (3.43) as



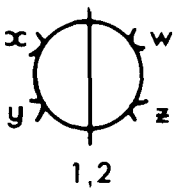
(A1)

then substitution of (A1) into the first term of (3.40) gives



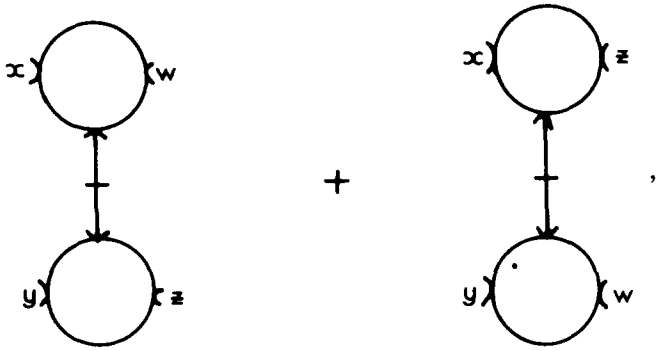
+ finite term.
(A2)

From (3.35) and (3.36) we have that



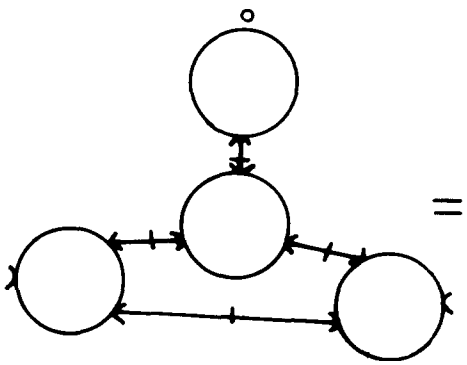
1,2

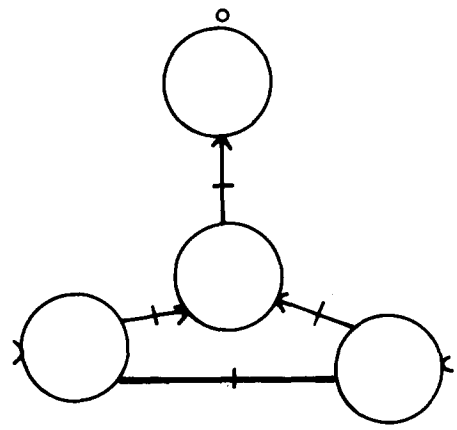
contains a term



(A3)

which, when substituted along with (A1) into the second term in (3.4) gives





(A4)

+ finite term.

This exactly cancels the second term in (A2). Provided that —+— has no single "tadpole" contributions we have shown that the logarithmic divergences arising from insertions of a single "tadpole" in a complete propagator in the first term of (3.40) cancel terms with no "tadpole" insertions in the propagators of the seond term of (3.41). Terms with more than one "tadpole" insertion in the first diagram of (3.40) are finite as are those with any insertions of tadpoles in the second diagram of (3.41). That —+— does in fact have no single "tadpole" insertions can be checked by induction on the iteration procedure.

We are still left with logarithmic infinites arising from insertions of tadpoles in the vertex functions of the diagrams in (3.40), or into the propagators in one side of the second diagram of (3.40) (insertion into propagators on both sides giving a finite diagram). However these infinites are some of those removed by the overlapping divergence treatment leading to (5.21). In fact, since separate multiplication of either of the two terms in (3.40) by $\sigma^{\mu}(x, y)$ does not remove these infinities, the fact that they are gone in (5.21) after manipulation using D^{μ} of (3.37) and executing various cancellations between the two diagrams, shows that the logarithmic infinities of the type under discussion must cancel between the two diagrams.

[1]I.T. Drummond, Nucl. Phys. B **94**, 115 (1975).
[2]I.T. Drummond and G.M. Shore, Ann. Phys. (in press).
[3]S.L. Adler, Phys. Rev. D **6**, 3445 (1972).
[4]N.D. Birrell and L.H. Ford, Ann. Phys. (N.Y.) **122**, 1 (1979).
[5]N.D. Birrell, J. Phys. A **13**, 569 (1980).
[6]N.D. Birrell and L.H. Ford, Phys. Rev. D (in press).
[7]T.S. Bunch, P. Panangaden, and L. Parker, J. Phys. A **13**, 901 (1980).
[8]T.S. Bunch and P. Panangaden, J. Phys. A **13**, 919 (1980).
[9]L. Scarf, "A soluble quantum field theory in curved space," in *Let Theóries Relativistes de la Gravitation* (Centre National De Recherche Scientifique, 1962).
[10]N.D. Birrell and P.C.W. Davies, Phys. Rev. D **18**, 4408 (1978).
[11]B.S. DeWitt, Phys. Rep. **19**, 295 (1975).
[12]S.M. Christensen, Phys. Rev. D **15**, 2490 (1976).
[13]J.G. Taylor, Nuovo Cimento Suppl. **1**, 857 (1963).
[14]J.D. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).
[15]L. Parker, "The production of elementary particles by strong gravita-

tionsl fields," in *Asymptotic Structure of Space–Time*, F.D. Esposito and L. Witten (Plenum, New York, 1977).
[16]We use units in which $\hbar = c = 1$, and sign conventions for the metric and curvature tensors which are $(- - -)$ in the terminology of C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[17]S.W. Hawking and G.F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University, Cambridge, Mass., 1973).
[18]For a more extensive discussion of the important question of being able to choose a time parametrization see B.S. Kay, Commun. Math. Phys. **71**, 29 (1980).
[19]L. Parker, Phys. Rev. Lett. **21**, 562 (1968).
[20]This is the same notation as used by DeWitt in Ref. 11, except that the indices on α^{-1} in his Eq. (162) are the other way around. It is clear from examination of his (161) that this is in fact an error. These is a similar error in his equation for V_{ij} [see our Eq. (2.36)].
[21]W.G. Unruh, Phys. Rev. D **14**, 870 (1976).
[22]S.W. Hawking, Phys. Rev. D **14**, 2460 (1976).
[23]P.S. Collecot, Ann. Phys. (in press).
[24]E.R. Caianello, "Combinatorics and Renormalization in Quantum Field Theory," *Frontiers in Physics* (Benjamin, New York, 1973).
[25]H. König, Math. Ann. **128**, 420 (1954).
[26]J.G. Taylor, Nuovo Cimento **17**, 695 (1960).
[27]K. Osterwolder and R. Schroder, Commun. Math. Phys. **31**, 83 (1973).
[28]N.K. Nielson, H. Romer, and B. Schroer, Phys. Letts. B **70**, 445 (1977).
[29]E. Kreyszig, *Introduction to differential geometry and Riemannian Geometry* (University of Toronto, 1968), Sec. 46.
[30]L. Parker, "Aspects of Quantum Field Theory in Curved Sapcetime, Effective Action and Energy-Momentum Tensor," in *Proceedings of the NATO Advanced Study Institute on Gravitation: Recent Developments*," edited by M. Levy and S. Deser (Plenum, New York, in press).
[31]N.D. Birrell, Proc. R. Soc. A **367**, 123 (1979).
[32]T.S. Bunch and L. Parker, Phys. Rev. D **20**, 2499 (1980).
[33]I.T. Drummond and G.M. Shore, Phys. Rev. D **19**, 1134 (1979).
[34]S.W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[35]G.W. Gibbons and M.J. Perry, Proc. R. Soc. **A358**, 467 (1978).
[36]P.C.W. Davies, Rep. Prog. Phys. **41**, 1313 (1978).
[37]L. Parker, Phys. Rev. D **12**, 1519, (1975).
[38]D.G. Boulware, Phys. Rev. D **11**, 1404 (1975).

Extended objects in quantum field theory

H. Matsumoto, G. Semenoff, H. Umezawa, and M. Umezawa [a)]
Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2J1

(Received 12 October 1979; accepted for publication 4 March 1980)

A general formulation for treatment of extended objects in a quantum field system is presented. It is shown how to construct the solutions of the Heisenberg equation when static extended objects are present. A detailed study of the quantum coordinate is presented. The Hamiltonian for the quantum mechanical object, which comes out of the quantum field theory, is obtained. It is shown that creation of extended objects does not upset the renormalizability.

1. INTRODUCTION

This paper is aimed at a general formalism for the treatment of extended objects which are created in quantum field systems by the condensation of certain bosons. There has been an active interest in the analysis of quantum effects in solition solutions[1,2] among physicists of many fields. The interest of the high energy physicists lies in their search for extended models of elementary particles, while many mathematical scientists are attracted to this subject purely by their mathematical curiosity. In solid state physics this study has a more realistic purpose; there we find a variety of phenomena in which microscopic objects (i.e., quanta) and extended (or macroscopic) objects coexist and interact with each other. Consider a crystal as an example. A perfect crystal without boundaries is a quantum system consisting of phonons and other quantum modes. However, perfect crystals are rarely found in reality; in almost all practical cases crystals contain many sorts of macroscopic objects such as dislocations, grain boundaries and point defects.[3] Vortices in superconductors[4] and magnetic domains in ferromagnets are also well known examples. The study of these extended objects in quantum many-body systems signals the opening of a new development in quantum field theory. At the beginning of the history of quantum field theory, the problem was to quantize the classical field equations in order to derive the quantum field theory. Our problem now is to start with the quantum field theory and to derive quantum mechanical or classical objects out of quantum field theory.

As a preliminary formulation of this kind, we have presented the boson transformation method.[5] A simple presentation of the boson transformation method was given in Refs. 6 and 7. In this method, one first solves the Heisenberg equations for the homogeneous system without any extended object and then certain extended objects are created in this quantum system by means of the boson transformation, which corresponds to a boson condensation. The boson transformed Heisenberg field describes both extended objects and quanta and also their mutual interaction.

It has been shown[7] that, as soon as an extended object is created by the boson condensation, the "quantum coordinate" **Q**, which is the quantum mechanical coordinate of the extended object, appears naturally. It was shown[7] that the operator **Q** appears only through the combination $(\mathbf{x}+\mathbf{Q})$, where **x** is the space coordinate. This quantum coordinate is the collective coordinate introduced by Gervais and Sakita.[8] When the size of the extended object is much larger than the quantum fluctuation of **Q**, the object behaves as a classical object; otherwise it behaves as a quantum mechanical object. In this way, both classical and quantum mechanical objects can result from quantum field theory. In Ref. 9 we applied the boson transformation method to the calculation of the quantum effects in the static soliton solution of the $(1+1)$-dimensional $\lambda\phi^4$-model, and the results were compared with those[2,10] of other methods. The purpose of the present paper is to study a general formalism for extended objects in quantum field systems. In particular, we make a detailed analysis of the problems of the quantum correction, the renormalization and the **Q**- and **P**-dependence of the Heisenberg operators. Here **P** is the canonical momentum of **Q**.

As a continuation of our study in Ref. 7, we consider a Heisenberg equation of the simple form

$$\Lambda(\partial)\psi(x) = F[\psi(x)], \tag{1.1}$$

where ψ is a scalar Heisenberg field. We assume that this model is renormalizable. Equation (1.1) leads to the Yang–Feldman equation

$$\psi(x) = \varphi_0(x) + \Lambda(\partial)^{-1}F[\psi(x)], \tag{1.2}$$

where $\varphi_0(x)$ is a renormalized free boson field satisfying

$$\Lambda(\partial)\varphi_0(x) = 0. \tag{1.3}$$

Note that our choice of φ_0 in (1.2) is not necessarily limited to the infield. As a matter of fact, in solid state physics one frequently uses φ_0 which is different from the in-field. For simplicity, we assumed that only one kind of free field appears in the theory. When we solve (1.2) by successive iteration we are led to an expression for ψ in terms of φ_0

$$\psi(x) = \psi(x;\varphi_0) = \sum_{n=0}^{\infty}\frac{1}{n!}\int d\sigma_1\cdots d\sigma_n\, c(x;\sigma_1\cdots\sigma_n)\ :\varphi_0(\sigma_1)\cdots\varphi_0(\sigma_n):, \tag{1.4}$$

which is called the dynamical map.

We now introduce a c-number function $f(x)$ which satisfies

$$\Lambda(\partial)f(x) = 0. \tag{1.5}$$

[a)]Centre de Recherches Nucleaires, Université Louis Pasteur, 67037 Strasbourg, France

 0022-2488/80/071761-09$1.00

Then we can generalize the Yang–Feldman equation as

$$\psi(x) = \varphi_0(x) + f(x) + [\Lambda(\partial)]^{-1} F[\psi(x)]. \tag{1.6}$$

Solving this equation successively, we obtain a new solution of (1.1)

$$\psi^f(x) = \psi(x;\varphi_0 + f). \tag{1.7}$$

Note that ψ^f is related to ψ through the substitution (the boson transformation) $\varphi_0 \to \varphi_0 + f$. The fact that both ψ and ψ^f satisfy the same Heisenberg equation (1.1) is the content of the boson transformation theorem

$$\Lambda(\partial)\psi^f = F[\psi^f]. \tag{1.8}$$

It is obvious that, when $f(x)$ is Fourier transformable, the result of the boson transformation is a particular case of the coherent representation[11] in which the c-number $f(x)$ is required to satisfy the free boson equation (1.5). The boson transformation method admits also those $f(x)$ which are not Fourier transformable. This is the origin of topological objects.[12,13]

The boson transformation theorem (and, therefore, the boson transformation method) has been formulated[5] even in those general cases in which there appear many free bosons, some of which are composite. However, in this paper we confine our consideration to the above simple case (1.1). We further simplify the situation by assuming that the extended objects are static [i.e., $f(x) = f(\mathbf{x})$].

Let us now write the dynamical map of the boson-transformed Heisenberg field ψ^f as

$$\psi^f(x) = \sum_{n=0}^{\infty} \frac{1}{n!} :\psi_f^{(n)}:, \tag{1.9}$$

where $\psi_f^{(n)}$ has the form

$$\psi_f^{(n)} = \int d\sigma_1 \cdots d\sigma_n \, c_f(x;\sigma_1 \cdots \sigma_n)\varphi_0(\sigma_1)\cdots\varphi_0(\sigma_n). \tag{1.10}$$

The leading term $\psi_f^{(0)}$ is equal to

$$\phi_f(\mathbf{x}) \equiv \langle 0|\psi^f(x)|0\rangle. \tag{1.11}$$

This is the classical field which describes the classical behavior of the extended object created by the condensation of φ_0. The function ϕ_f depends on $\mathbf{x}$ through $f(\mathbf{x})$. The linear term in the dynamical map of ψ^f is

$$\psi_f^{(1)} = \int d\sigma \, c_f(x;\sigma)\varphi_0(\sigma). \tag{1.12}$$

The matrix elements of $\psi_f^{(1)}$ describe the behavior of a single-quantum under the influence of the extended object. The higher order terms describe the quantum reactions in the presence of the extended objects. In this way the ψ^f describes the system of quanta and extended objects. The $\psi_f^{(n)}$ with $n > 0$ contains up to the nth order products of the quantum coordinate $\mathbf{Q}$.[7] When we use the Taylor expansion, we have[7]

$$\psi_f^{(n)}(x) = (\delta_f)^n \phi_f(\mathbf{x}), \tag{1.13}$$

with

$$\delta_f \equiv \int d\sigma \, \varphi_0(\sigma) \frac{\delta}{\delta f(\sigma)}. \tag{1.14}$$

In the next section it will be shown that, when the Heisenberg equation (1.1) is renormalizable in the homogeneous system (i.e., the system without any extended object), the creation of extended objects does not upset the renormalizability.[9] It has been proved[6] that, when the three approximation is used, the classical field ϕ_f satisfies the Euler equation. It was briefly shown in Ref. 7 how each term in the dynamical map (1.9) in the tree approximation can be determined from the knowledge of ϕ_f. Thus, when we specify a solution ϕ_f, the boson-transformed Heisenberg field ψ^f in the tree approximation is determined. In Sec. 3 a short résumé of the content of the Ref. 7 will be presented. In Sec. 4 it will be shown that all of the quantum corrections in ψ^f can be calculated from the dynamical map of ψ^f in the tree approximation. Thus, we are led to the remarkable conclusion that the knowledge of the classical field ϕ_f in the tree approximation determines the boson-transformed Heisenberg field ψ^f with full quantum-corrections. Hereafter we will use the hat notation $\hat{\psi}^f$ to distinguish operators in the tree approximation from exact operators. Those without a hat are exact. Section 4 will be devoted to a detailed study of the quantum coordinate.

2. RENORMALIZATION AND EXTENDED OBJECTS

Let us assume that the Heisenberg equation (1.1) is renormalizable when it is applied to the homogeneous system (i.e., the system without any extended object). Then the coefficients $c(x;\sigma_1 \cdots \sigma_n)$ in (1.4) are finite functions when they are expressed in terms of the renormalized quantities. Equations (1.7) and (1.10) now lead to

$$c_f(x;\sigma_1 \cdots \sigma_n) = \sum_{l=0}^{\infty} \frac{1}{l!} \int d^4\xi_1 \cdots d^4\xi_l \, c(x;\sigma_1 \cdots \sigma_n \xi_1 \cdots \xi_l) \times f(\xi_1)\cdots f(\xi_l). \tag{2.1}$$

Thus, the coefficients $c_f(x;\sigma_1 \cdots \sigma_n)$ do not contain any divergences when they are put in this expansion form. Recalling the fact that the statement of renormalizability has always been associated with certain expansions (such as perturbative expansion, loop expansions, etc.), we may express the above result by the statement that, when $c(x;\sigma_1 \cdots \sigma_n)$ are renormalized, so are $c_f(x;\sigma_1 \cdots \sigma_n)$. In other words, the creation of extended objects does not upset the renormalizability. There still remains the question of how the summation in (2.1) is defined.

3. THE DYNAMICAL MAP IN THE TREE APPROXIMATION

In the tree approximation the classical field satisfies the Euler equation

$$\Lambda(\partial)\phi_f = F[\phi_f]. \tag{3.1}$$

We now show that a solution of this classical equation determines uniquely the dynamical map of the Heisenberg operator in the tree approximation. This is a preparation for the proof of the statement that the classical field ϕ_f in the tree approximation determines the boson-transformed Heisenberg field ψ^f with full quantum corrections.

Recall that operators in the tree approximation are denoted by the corresponding symbols with a hat. The Heisenberg equation (1.1) is read in the tree approximation as

$$\Lambda(\partial)\Big(\sum_{n=0}^{\infty}\frac{1}{n!}\hat{\psi}_f^{(n)}\Big)=F\Big[\sum_{n=0}^{\infty}\frac{1}{n!}\hat{\psi}_f^{(n)}\Big]. \tag{3.2}$$

In the above, noncommutativity among $\psi_f^{(n)}$ can be disregarded in the tree approximation. This equation leads to the recurrence formula

$$\Lambda(\partial)\hat{\psi}_f^{(n)}/n!=\sum_l F^{(l)}[\hat{\phi}_f]\prod_i\frac{1}{\alpha_i!}(\hat{\psi}_f^{(n_i)}/n_i!)^{\alpha_i}, \tag{3.3}$$

with

$$\sum\alpha_i=l,\ \sum\alpha_i n_i=n, \tag{3.4a}$$

$$0<n_1<n_2\cdots<n. \tag{3.4b}$$

In (3.3), $F^{(l)}[\hat{\phi}_f]$ is the nth derivative of $F[\hat{\phi}_f]$ with respect to $\hat{\phi}_f$. The summation in (3.3) runs over all the possible choices of l, n_i and α_i satisfying the conditions (3.4a) and (3.4b). In particular, (3.3) with $n=1$ and 2 leads to

$$(\Lambda(\partial)-F^{(1)}[\hat{\phi}_f(x)])\hat{\psi}^{(1)}(x)=0 \tag{3.5}$$

and

$$(\Lambda(\partial)-F^{(1)}[\hat{\phi}_f(x)])\hat{\psi}^{(2)}(x)=F^{(2)}[\hat{\phi}_f(x)](\hat{\psi}_f^{(1)}(x))^2, \tag{3.6}$$

respectively. Another derivation of (3.3) is given by a successive operation of δ_f on both sides of (3.1).[7]

Let us now note that the equations for $\hat{\psi}_f^{(n)}$ with an $n\geqslant 2$ are inhomogeneous. Thus, $\hat{\psi}_f^{(n)}$ with $n\geqslant 2$ are determined by $\hat{\psi}_f^{(1)}$. In other words, $\hat{\psi}_f^{(n)}$ with $n\geqslant 2$ can be expressed in terms of linear combinations of products of $\hat{\psi}_f^{(1)}$. Since (1.12) shows that $\hat{\psi}_f^{(1)}$ is the physical field modified by the influence of the extended object, we see that $\hat{\psi}_f^{(n)}$ with $n\geqslant 2$ are expressed in terms of physical fields. Equation (3.5) shows[7] that the influence of the extended object on the physical field appears through the self-consistent potential $F^{(1)}[\hat{\phi}_f(x)]$.

Since Eq. (3.5) is homogeneous, it cannot determine the structure of $\psi_f^{(1)}$. This structure is determined by the equal time canonical commutation relation.[7] Treatment of (3.5) requires particular care, because this equation admits, not only solutions for scattering waves, but also bound states; we meet the well-known intricate relation between the iterative method and the appearance of bound states. When we rewrite (3.5) as

$$\hat{\psi}_f^{(1)}(x)=\varphi^0(x)+\Lambda^{-1}(\partial)F^{(1)}[\hat{\phi}_f(x)]\hat{\psi}_f^{(1)}(x), \tag{3.7}$$

the iterative method gives the solution for the scattering waves only. This solution will be denoted by $\hat{\chi}_0^s(x)$. Equation (3.5) admits also some solutions which have discrete energies. Among these solutions, there appear[7] three zero-energy modes when $f(\mathbf{x})$ depends on the three components of $\mathbf{x}$. These zero-energy modes originate from the translational invariance of the Heisenberg equation (1.1), and therefore they are called the translation modes. Since they are not observed as excitation modes, we associate[7] the quantum coordinate $\mathbf{Q}$ (rather than the creation or annihilation operators) with them. We denote the canonical conjugates of $\mathbf{Q}$ by $\mathbf{P}$. Then our Hilbert space is the product of the Hilbert space $\mathscr{H}_Q=\{|\Psi_Q\rangle\}$ of the quantum coordinate and the Fock space $\mathscr{H}_F=\{|\Psi_F\rangle\}$ of the excitation modes

$$|\Psi\rangle=|\Psi_F\rangle\times|\Psi_Q\rangle. \tag{3.8}$$

We use the notation $|0\rangle$ to denote the vacuum state in $\mathscr{H}_F$. Thus, the definition (1.11) of the classical field should be modified as

$$\phi_f(\mathbf{x})=\langle 0|\psi_f(x)|0\rangle\,|_{P=Q=0}. \tag{3.9}$$

As was shown in Ref. 7, the canonical commutator requires that all the members of the complete set of solutions of (3.5) should appear in $\hat{\psi}_f^{(1)}$. Thus, $\hat{\psi}_f^{(1)}$ has the form

$$\hat{\psi}_f^{(1)}(x)=(\mathbf{Q}\cdot\nabla)\hat{\phi}_f(\mathbf{x})+\hat{\chi}_0(x), \tag{3.10a}$$

with

$$\hat{\chi}_0(x)=\hat{\chi}_0^s(x)+\hat{\chi}_0^\alpha(x), \tag{3.10b}$$

where $\hat{\chi}_0^\alpha(x)$ is the physical field associated with nonzero discrete energy solution of (3.5) in the tree approximation. The fields $\hat{\chi}_0^s$ and $\hat{\chi}_0^\alpha$ have the forms

$$\hat{\chi}_0^s(x)=\int d^3k\,(1/\sqrt{2\omega_k})\{u(\mathbf{x},\mathbf{k})\alpha(\mathbf{k})e^{-i\omega_k t}+u^*(\mathbf{x},\mathbf{k})\alpha\dagger(\mathbf{k})e^{i\omega_k t}\}, \tag{3.11a}$$

$$\hat{\chi}_0^\alpha(x)=\sum_i(1/\sqrt{2\omega_i})\{u_i(\mathbf{x})\alpha_i e^{-i\omega_i t}+u_i^*(\mathbf{x})\alpha_i^\dagger e^{i\omega_i t}\}, \tag{3.11b}$$

where $\alpha(\mathbf{k})$ and α_i are the annihilation operators of quanta with continuous energy and nonzero discrete energies respectively, and $u(\mathbf{x},\mathbf{k})$ and $u_i(\mathbf{x})$ are their orthonormalized wavefunctions. In writing (3.11a) and (3.11b), we assumed that the canonical conjugate of ψ is $(\partial/\partial t)\psi$. When this assumption does not hold, (3.11a) and (3.11b) require a suitable modification. The appearance of the bound states through the boson transformation may depend on the way of defining the limit of the iterative method applied to the Schrödinger equation (3.5). This may depend also on the way of defining the summation of the power expansion (2.1) with $n=1$, i.e.,

$$c_f(x,\sigma)=\sum_{l=0}^{\infty}\frac{1}{l!}\int d^4\xi_1\cdots d^4\xi_l\,c(x;\sigma\xi_1\cdots\xi_l)\times f(\xi_1)\cdots f(\xi_l). \tag{3.12}$$

The summation should be defined in such a way that the equal-time canonical commutation relation for ψ^f holds.

Summarizing, the rule of calculation of $\hat{\psi}_f^{(1)}$ is a follows: calculate $\hat{\chi}_0^s$ by the iterative method and then replace this by $\hat{\chi}_0^s+\hat{\chi}_0^\alpha+(\mathbf{Q}\cdot\nabla)\hat{\phi}_f$. The operator $\hat{\psi}_f^{(1)}$ is then given by (3.10a). Using $\hat{\psi}_f^{(1)}$ thus determined, $\hat{\psi}_f^{(n)}$ with $n\geqslant 2$ are obtained from the equation (3.3) with $n\geqslant 2$. The dynamical map of ψ^f in the tree approximation is now determined.

It was shown in Ref. 7 that $\mathbf{x}$ and $\mathbf{Q}$ appear only through the combination $(\mathbf{x}+\mathbf{Q})$. Thus the spatial translation $\mathbf{x}\to\mathbf{x}+\mathbf{a}$ is induced by the transformation $\mathbf{Q}\to\mathbf{Q}+\mathbf{a}$. Considering this, we can simplify the above-mentioned rule of calculation of ψ^f in the tree approximation as follows: we use $\hat{\psi}_f^{(1)}=\hat{\chi}_0$(rather than $\hat{\psi}_f^{(1)}=(\mathbf{Q}\cdot\nabla)\hat{\phi}_f+\hat{\chi}_0$) and determine $\hat{\psi}_f^{(n)}$ by means of Eq. (3.2) with $n\geqslant 2$. We then replace $\mathbf{x}$ by $\mathbf{x}+\mathbf{Q}$. This point will be further elaborated in Sec. 5.

In the above argument the tree approximation is applied to the unrenormalized Heisenberg equation (1.1). In practice, the renormalization process is performed before

the boson transformation is made. In renormalization theory, $F[\psi]$ has the form

$$F[\psi] = F_0[\psi] + F_c[\psi], \tag{3.13}$$

where $F_c[\psi]$ contains the counter terms which compensate the differences between the renormalized parameters and the bare parameters. In renormalization theory, the tree approximation is applied to

$$\Lambda(\partial)\hat{\psi} = F_0[\hat{\psi}], \tag{3.14}$$

which leads to the Euler equation

$$\Lambda(\partial)\hat{\phi}_f^0 = F_0[\hat{\phi}_f^0]. \tag{3.15}$$

When our consideration begins with a solution of this equation, $\hat{\phi}_f$ in all the formulas in this section should be expressed in terms of $\hat{\phi}_f^0$. To do this we use the equation

$$\Lambda(\partial)\Delta\hat{\phi}_f = F_0[\hat{\phi}_f^0 + \Delta\hat{\phi}_f] - F_0[\hat{\phi}_f^0] + F_c[\hat{\phi}_f^0 + \Delta\hat{\phi}_f], \tag{3.16}$$

where

$$\Delta\hat{\phi}_f \equiv \hat{\phi}_f - \hat{\phi}_f^0. \tag{3.17}$$

Using the Taylor expansion

$$F^{(n)}[\hat{\phi}_f] = \sum_{l=0}^{\infty} \frac{1}{l!} F^{(n+l)}[\hat{\phi}_f^0](\Delta\hat{\phi}^f)^l, \tag{3.18}$$

we can express Eq. (3.3) in terms of $\hat{\phi}_f^0$. Solving this equation, we can express the dynamical map of $\hat{\psi}^f$ [satisfying (1.1) in the tree approximation] in terms of $\hat{\phi}_f^0$.

In the above consideration, we ignored the quantum coordinate. The latter coordinate will be recovered in Sec. 5.

4. THE QUANTUM EFFECTS

In the last section we constructed the dynamical map of ψ^f in the tree approximation. In this section we show how to take into account the quantum effects. As was pointed out previously, operators with a hat indicate those in the tree approximation.

It should be noted that $\hat{\psi}^f$ in the tree approximation considered in the last section contains the renormalization effect $\Delta\hat{\phi}_f$ which is created by the counter term $F_c[\psi]$. This renormalization effect $\Delta\hat{\phi}_f$ is a part of the quantum effect. In the following we consider the other part of the quantum corrections.

Let us first calculate the quantum correction in the classical field $\phi_f(x)$ defined by

$$\phi_f(x) = \langle 0|\psi^f(x)|0\rangle|_{P=Q=0}. \tag{4.1}$$

The $\phi_f(x)$ can be obtained from $\hat{\psi}_f(x)$ in the tree approximation by contracting the physical field operators in all the possible ways. Therefore, we have

$$\phi_f(\mathbf{x}) = \hat{\phi}_f^0(\mathbf{x}) + \Delta\hat{\phi}_f(\mathbf{x}) + \overline{\Delta}\hat{\phi}_f(\mathbf{x}), \tag{4.2}$$

with

$$\overline{\Delta}\hat{\phi}_f(\mathbf{x}) = \sum_{n=1}^{\infty} \frac{1}{(2n)!} \langle 0|\hat{\psi}_f^{(2n)}(\mathbf{x})|0\rangle. \tag{4.3}$$

Here the $\hat{\psi}_f^{(n)}(x)$ are those which were calculated in the tree approximation. In deriving (4.2) we considered the fact that all possible contractions among the physical fields φ_0 appearing in the dynamical map of $\hat{\psi}^f$ in the tree approximation contribute to the classical field ϕ_f.

The quantum effects in the linear term in the dynamical map of ψ^f can be calculated in a similar manner. In this case we consider $\hat{\psi}_f^{(2l+1)}$ of the tree approximation and perform all possible contractions of $2l$ physical fields. The result is the l-line corrections to $\hat{\psi}_f^{(1)}$. In this way, we can calculate the quantum effects in any term in the dynamical map of ψ^f.

As an example we consider the one-quantum correction to the classical field ϕ_f. Then, (4.3) is approximated by

$$\overline{\Delta}\hat{\phi}_f \simeq \tfrac{1}{2}\langle 0|\hat{\psi}_f^{(2)}|0\rangle. \tag{4.4}$$

Equation (3.4) with $\hat{\phi}_f = \hat{\phi}_f^0 + \Delta\hat{\phi}_f$ gives

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f^0 + \Delta\hat{\phi}_f]\}\langle 0|\hat{\psi}_f^{(2)}|0\rangle = F^{(2)}[\hat{\phi}_f^0 + \Delta\hat{\phi}_f]\langle 0|(\hat{\psi}_f^{(1)})^2|0\rangle. \tag{4.5}$$

Following the computational rule presented in the last section, we replace $\hat{\psi}_f^{(1)}$ by $\hat{\chi}_0$, disregarding the quantum coordinate. Then (4.5), together with (4.4), gives

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f^0 + \Delta\hat{\phi}_f]\}\overline{\Delta}\hat{\phi}_f = \tfrac{1}{2}F^{(2)}[\hat{\phi}_f^0 + \Delta\hat{\phi}_f]\langle 0|(\hat{\chi}_0)^2|0\rangle. \tag{4.6}$$

Since we are considering the one-quantum correction only, we can rewrite (4.6) as

$$\{\Lambda(\partial) - F_0^{(1)}[\hat{\phi}_f^0]\}\overline{\Delta}\hat{\phi}_f = \tfrac{1}{2}F_0^{(2)}[\hat{\phi}_f^0]\langle 0|(\hat{\chi}_0)^2|0\rangle. \tag{4.7}$$

Here we considered the fact that $F_c[\hat{\phi}_f^0]$ is a quantum correction. The renormalization term $\Delta\hat{\phi}_f$ is a function of renormalized constants and the Planck constant; the latter constant results from the quantum corrections in the counter term F_c. Expanding this function in powers of the Planck constant as

$$\Delta\hat{\phi}_f = \sum_{l=1}^{\infty} \Delta\hat{\phi}_f^{(l)}, \tag{4.8}$$

we approximately put $\Delta\hat{\phi}_f \simeq \Delta\hat{\phi}_f^{(1)}$. Then Eq. (3.16) for $\Delta\hat{\phi}_f$ becomes

$$\{\Lambda(\partial) - F_0^{(1)}[\hat{\phi}_f^0]\}\Delta\hat{\phi}_f^{(1)} = F_c[\hat{\phi}_f^0]. \tag{4.9}$$

Combining (4.7) and (4.9), we obtain

$$\{\Lambda(\partial) - F_0^{(1)}[\hat{\phi}_f^0]\}(\Delta\hat{\phi}_f^{(1)} + \overline{\Delta}\hat{\phi}_f) = \tfrac{1}{2}F_0^{(2)}[\hat{\phi}_f^{(0)}]\langle 0|(\hat{\chi}_0)^2|0\rangle + F_c[\hat{\phi}_f^0]. \tag{4.10}$$

This is the equation which, together with (4.2), determines the one-quantum correction of the classical field ϕ_f. As it was proved in Sec. 2, the divergence which appears in $\langle 0|(\hat{\chi}_0)^2|0\rangle$ should be cancelled by the counter term $F_c[\hat{\phi}_f^0]$.

When we are interested in the one-quantum effect in $\langle 0|Q[\psi^f]|0\rangle$ where $Q[\psi^f]$ is a linear combination of products of ψ^f, we can simply make the replacement

$$\psi^f \to \phi_f + \chi_0, \tag{4.11}$$

where ϕ_f is given by (4.2). For example, the energy of the vacuum state with the extended object in this approximation is given by $\langle 0|H[\psi_f]|0\rangle$ with the replacement (4.11). Here H is the Hamiltonian.

Let us now illustrate the one-quantum correction in ϕ_f by means of the simple models. First we start from the $\lambda\rho^4$-model

$$(-\partial^2 + \mu^2)\rho(x) = \lambda\rho^3(x). \tag{4.12}$$

When we put

$$\rho(x) = Z^{1/2}(m/g + \psi(x)), \tag{4.13a}$$

$$\lambda = \tfrac{1}{2}g^2 Z_1 Z^{-2}, \tag{4.13b}$$

$$\mu^2 = \tfrac{1}{2}m^2 Z_2 Z^{-1}, \tag{4.13c}$$

we have

$$(-\partial^2 - m^2)\psi(x) = F_0[\psi] + F_c[\psi], \tag{4.14}$$

where

$$F_0[\psi] = \tfrac{3}{2}gm\psi^2 + \tfrac{1}{2}g^2\psi^3, \tag{4.15a}$$

$$\begin{aligned} F_c[\psi] &= -\tfrac{1}{2}\frac{m^3}{g}Z^{-1}(Z_2 - Z_1) \\ &\quad + m^2\left[\tfrac{3}{2}Z_1 Z^{-1} - (1 + \tfrac{1}{2}Z_2 Z^{-1})\right]\psi \\ &\quad + \tfrac{3}{2}(Z_1 Z^{-1} - 1)gm\psi^2 + \tfrac{1}{2}(Z_1 Z^{-1} - 1)g^2\psi^3 \\ &= -\tfrac{1}{2}m^2(Z^{-1}Z_2 - 1)(m/g + \psi) \\ &\quad + \tfrac{1}{2}g^2(Z^{-1}Z_1 - 1)(m/g + \psi)^3. \end{aligned} \tag{4.15b}$$

Here, m is the renormalized mass, g is the renormalized coupling constant, and Z, Z_1, and Z_2 are the renormalization factors. The renormalization constants in the $(1+1)$-dimensional case were calculated in Ref. 9. Deviations of the constants Z,Z_1, and Z_2 from 1 are caused by the quantum effects.

The Euler equation for $\hat{\phi}_f^0$ is given by

$$(-\partial^2 - m^2)\hat{\phi}_f^0 = F_0[\hat{\phi}_f^0], \tag{4.16}$$

which reads as

$$(-\partial^2 + \tfrac{1}{2}m^2)(m/g + \hat{\phi}_f^0) = \tfrac{1}{2}g^2(m/g + \hat{\phi}_f^0)^3. \tag{4.17}$$

This leads to the well-known soliton solution for ϕ_f^0. Equation (4.15a) gives

$$F_0^{(1)}[\hat{\psi}] = 3g\, m\hat{\psi} + \tfrac{3}{2}g^2\hat{\psi}^2, \tag{4.18}$$

$$F_0^{(2)}[\hat{\psi}] = 3g\, m + 3g^2\hat{\psi}. \tag{4.19}$$

Since we have $(Z_i Z^{-1} - 1) \simeq (Z_i - Z)(i = 1,2)$ in the one-loop approximation (4.15b) gives

$$\begin{aligned} F_c[\hat{\phi}_f^0] &= -\tfrac{1}{2}m^2(Z_2 - Z)(m/g + \hat{\phi}_f^0) \\ &\quad + \tfrac{1}{2}g^2(Z_1 - Z)(m/g + \hat{\phi}_f^0)^3, \end{aligned} \tag{4.20}$$

and Eq. (4.10) for $(\Delta\hat{\phi}_f + \bar{\Delta}\hat{\phi}_f)$ becomes

$$\begin{aligned} &\{-\partial^2 - m^2 - 3g\, m\hat{\phi}_f^0 - \tfrac{3}{2}g^2(\hat{\phi}_f^0)^2\}(\Delta\hat{\phi}_f + \bar{\Delta}\hat{\phi}_f) \\ &= \tfrac{3}{2}g^2(m/g + \hat{\phi}_f^0)\langle 0|\hat{\chi}_0^2|0\rangle + \tfrac{1}{2}g^2(Z_1 - Z)(m/g + \hat{\phi}_f^0)^3 \\ &\quad - \tfrac{1}{2}m^2(Z_2 - Z)(m/g + \hat{\phi}_f^0). \end{aligned} \tag{4.21}$$

This agrees with the result in Ref. 9. As was shown there, the divergencies in Z_1, $Z\,Z$, and $\langle 0|\hat{\chi}_0^2|0\rangle$ cancel out among themselves even in the $(1+3)$-dimension, implying that creation of soliton does not upset the renormalization. Calculation of the one-quantum correction of the energy for the model under consideration was calculated also in Ref. 9 and 10.

In the case of the sine-Gordon equation

$$\partial^2\rho = \alpha/\beta \sin\beta\rho, \tag{4.22}$$

the renormalized field ψ and constants α_r, β_r are defined, respectively, as

$$\rho = Z^{1/2}\psi, \tag{4.23a}$$

$$\alpha = Z^{-1}Z_\alpha\alpha_r, \tag{4.23b}$$

$$\beta = Z^{-1/2}\beta_r. \tag{4.23c}$$

The sine-Gordon model is renormalizable only in the $(1+1)$-dimension, so that the divergence appears only in the mass renormalization Z_α. The β_r is determined in the course of defining the renormalized coupling constant. Then Eq. (4.22) is rewritten as

$$\partial^2\psi = F_0[\psi] + F_c[\psi], \tag{4.24}$$

where

$$F_0[\psi] = \alpha_r/\beta_r \sin\beta_r\psi, \tag{4.25}$$

$$F_c[\psi] = (1 - Z)\partial^2\psi + \alpha_r/\beta_r(Z_\alpha - 1)\sin\beta_r\psi. \tag{4.26}$$

Notice that the one-quantum correction of the mass renormalization in the homogeneous case leads to the relation

$$(Z_\alpha - Z) = \tfrac{1}{2}\beta_r^2\langle 0|\varphi_0^2(x)|0\rangle, \tag{4.27}$$

with

$$(\partial^2 + \alpha_r)\varphi_0(x) = 0. \tag{4.28}$$

The equation which determines the one-quantum correction of the classical fields is given by (4.10)

$$\begin{aligned} &\{\partial^2 - \alpha_r\cos\beta_r\hat{\phi}_f^0\}(\Delta\hat{\phi}_f^{(1)} + \bar{\Delta}\hat{\phi}_f) \\ &= -\tfrac{1}{2}\alpha_r\beta_r \sin\beta_r\hat{\phi}_f^0[\langle 0|\hat{\chi}_0^2|0\rangle - 2\beta_r^{-2}(Z_\alpha - Z)]. \end{aligned} \tag{4.29}$$

To derive (4.29), the Euler equation

$$\partial^2\hat{\phi}_f^0 = \alpha_r/\beta_r \sin\beta_r\hat{\phi}_f^0 \tag{4.30}$$

was used. Since $\{\langle 0|\hat{\chi}_0^2(x)|0\rangle - \langle 0|\varphi_0^2(x)|0\rangle\}$ is finite, one gets the finite value for the one-quantum correction of ϕ_f.

We have so far disregarded the quantum coordinate, which will be recovered in the next section.

5. THE QUANTUM COORDINATES

In the calculations of the dynamical map of $\hat{\psi}^f$ in the last two sections we ignored the quantum coordinate $\mathbf{Q}$ and its canonical conjugate $\mathbf{P}$. In this section we study how ψ^f depends on $\mathbf{Q}$ and $\mathbf{P}$. Since each term $\psi_f^{(n)}$ in the dynamical map depends on $\mathbf{Q}$ only through the combination $(\mathbf{x} + \mathbf{Q})$, we have

$$\psi_f^{(n)}(x) = \psi_f^{(n)}(\mathbf{x} + \mathbf{Q}, \mathbf{P}, t). \tag{5.1}$$

For the sake of simplicity, we assume

$$\Lambda(\partial) = -\partial^2 - m^2. \tag{5.2}$$

Note also that, throughout this paper, we are considering static extended objects.

Since the total Hamiltonian H does not depend on $\mathbf{x}$, it does not contain $\mathbf{Q}$; in other words, H contains $\mathbf{P}$ only. Then the relation

$$\dot{\mathbf{Q}} = i[H, \mathbf{Q}] \tag{5.3}$$

shows that $\dot{\mathbf{Q}}$ depends on $\mathbf{P}$ only, implying

$$\ddot{\mathbf{Q}} = i[H, \dot{\mathbf{Q}}] = 0. \tag{5.4}$$

Furthermore,

$$\dot{\mathbf{P}} = i[H, \mathbf{P}] = 0. \tag{5.5}$$

Thus, $\dot{\mathbf{Q}}$ and $\mathbf{P}$ are independent of time. We can thus write

$$\mathbf{Q}(t) = \mathbf{Q}(0) + t\dot{\mathbf{Q}}, \tag{5.6}$$

which leads to

$$\mathbf{Q}(t') = \mathbf{Q}(t) + (t' - t)\dot{\mathbf{Q}}. \tag{5.7}$$

Making use of (5.6) and (5.7), we can calculate $\psi_f^{(n)}(\mathbf{x} + \mathbf{Q}, \mathbf{P}, t)$.

Let us begin with $\hat{\psi}_f^{(2)}$ in the tree approximation. Equation (3.6) gives

$$\hat{\psi}_f^{(2)}(x) = \int d^4\sigma\, G(x,\sigma) F^{(2)}[\hat{\phi}_f(\sigma)](\hat{\psi}_f^{(1)}(\sigma))^2, \tag{5.8}$$

where

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f(\mathbf{x})]\} G(x,\sigma) = \delta(x - \sigma). \tag{5.9}$$

According to (3.10a), we have

$$\hat{\psi}_f^{(1)}(\sigma) = (\mathbf{Q}(t_\sigma)\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma})) + \hat{\chi}_0(\sigma) \tag{5.10}$$

$$= (\mathbf{Q}(t)\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma})) + (t_\sigma - t)(\dot{\mathbf{Q}}\cdot\nabla)\hat{\phi}_f(\boldsymbol{\sigma}) + \hat{\chi}_0(\sigma). \tag{5.11}$$

Here t means t_x. Now (5.8) gives

$$\begin{aligned}\hat{\psi}_f^{(2)}(x) = \int d^4\sigma\, G(x,\sigma) F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})][(\mathbf{Q}(t)\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma}))^2 \\ + 2(\mathbf{Q}(t)\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma}))\hat{\chi}_0(\sigma) + (\hat{\chi}_0(\sigma))^2 \\ + (t_\sigma - t)\{(\mathbf{Q}(t)\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma}))(\dot{\mathbf{Q}}\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma})) \\ + (\dot{\mathbf{Q}}\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma}))(\mathbf{Q}(t)\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma}))\} + (t_\sigma - t)^2 \\ \times(\dot{\mathbf{Q}}\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma}))^2 + 2(t_\sigma - t)(\dot{\mathbf{Q}}\cdot\nabla\hat{\phi}_f(\boldsymbol{\sigma}))\hat{\chi}_0(\sigma)].\end{aligned} \tag{5.12}$$

On the on the hand, operating on both sides of the Euler equation (3.1) with derivatives we obtain

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\}\nabla\hat{\phi}_f(\mathbf{x}) = 0, \tag{5.13a}$$

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\}\partial_i\partial_j\hat{\phi}_f(\mathbf{x}) = F^{(2)}[\hat{\phi}_f]\partial_i\hat{\phi}_f\cdot\partial_j\hat{\phi}_f. \tag{5.13b}$$

The latter equation gives

$$\int d^4\sigma\, G(x,\sigma) F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\partial_j\hat{\phi}_f(\boldsymbol{\sigma}) = \partial_i\partial_j\hat{\phi}_f(\mathbf{x}). \tag{5.14}$$

This determines the first term on the right-hand side of (5.12). To calculate the second term, we recall (3.5), which gives

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\}\hat{\chi}_0 = 0. \tag{5.15}$$

Operating on both sides of this with ∂_i, we obtain

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\}\partial_i\hat{\chi}_0 = F^{(2)}[\hat{\phi}_f]\partial_i\hat{\phi}_f\hat{\chi}_0, \tag{5.16}$$

which gives:

$$\int d^4\sigma\, G(x,\sigma) F^{(2)}[\hat{\phi}_f(\sigma)]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\hat{\chi}_0(\sigma) = \partial_i\hat{\chi}_0(x). \tag{5.17}$$

This determines the second term. These calculations of the first and second terms were already presented in Ref. 7.

Let us now calculate other terms. We have

$$\begin{aligned}\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\} \int d^4\sigma\, G(x,\sigma)(t_\sigma - t) \\ \times F^{(2)}[\hat{\phi}_f(\sigma)]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\partial_j\hat{\phi}_f(\boldsymbol{\sigma}) \\ = 2\frac{\partial}{\partial t}\int d^4\sigma G(x,\sigma) F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\partial_j\hat{\phi}_f(\boldsymbol{\sigma}) \\ = 2\frac{\partial}{\partial t}(\partial_i\partial_j\hat{\phi}_f(\mathbf{x})) = 0,\end{aligned} \tag{5.18}$$

$$\begin{aligned}\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\} \int d^4\sigma\, G(x,\sigma)(t_\sigma - t)^2 \\ \times F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\partial_j\hat{\phi}_f(\boldsymbol{\sigma}) \\ = 4\frac{\partial}{\partial t}\int d^4\sigma G(x,\sigma)(t_\sigma - t) \\ \times F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\partial_j\hat{\phi}_f(\boldsymbol{\sigma}) \\ + 2\int d^4\sigma\, G(x,\sigma) F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\partial_j\hat{\phi}_f(\boldsymbol{\sigma}) \\ = 2\partial_i\partial_j\hat{\phi}_f(\mathbf{x}).\end{aligned} \tag{5.19}$$

On the other hand, it is easy to find that

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\}(x_i\partial_j\hat{\phi}_f(\mathbf{x})) = 2\partial_i\partial_j\hat{\phi}_f. \tag{5.20}$$

Thus, (5.19) gives

$$\begin{aligned}\int d^4\sigma\, G(x,\sigma)(t_\sigma - t) F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\partial_j\hat{\phi}_f(\boldsymbol{\sigma}) \\ = x_i\partial_j\hat{\phi}_f(\mathbf{x}).\end{aligned} \tag{5.21}$$

Finally, we note that

$$\begin{aligned}\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\} \int d^4\sigma\, G(x,\sigma)(t_\sigma - t) \\ \times F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma}]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\hat{\chi}_0(\boldsymbol{\sigma}) \\ = 2\frac{\partial}{\partial t}\int d^4\sigma\, G(x,\sigma) F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\hat{\chi}_0(\sigma) \\ = 2\partial_i\dot{\hat{\chi}}_0(x),\end{aligned} \tag{5.22}$$

where (5.17) has considered. Since

$$\{\Lambda(\partial) - F^{(1)}[\hat{\phi}_f]\}(x_i\hat{\chi}_0) = 2\partial_i\hat{\chi}_0, \tag{5.23}$$

(5.22) leads to

$$\begin{aligned}\int d^4\sigma\, G(x,\sigma)(t_\sigma - t) F^{(2)}[\hat{\phi}_f(\boldsymbol{\sigma})]\partial_i\hat{\phi}_f(\boldsymbol{\sigma})\cdot\hat{\chi}_0(\sigma) \\ = x_i\dot{\hat{\chi}}_0(x).\end{aligned} \tag{5.24}$$

When (5.14), (5.17), (5.18), (5.21), and (5.24) are used, (5.12) gives

$$\begin{aligned}\hat{\psi}_f^{(2)}(x) = (\mathbf{Q}(t)\cdot\nabla)^2\hat{\phi}_f(\mathbf{x}) + 2(\mathbf{Q}(t)\cdot\nabla)\hat{\chi}_0(x) \\ + (\dot{\mathbf{Q}}\cdot\mathbf{x})(\dot{\mathbf{Q}}\cdot\nabla)\hat{\phi}_f(\mathbf{x}) + 2(\dot{\mathbf{Q}}\cdot\mathbf{x})\dot{\hat{\chi}}_0(x) \\ + \int d^4\sigma\, G(x,\sigma) F^{(2)}[\hat{\phi}_f(\sigma)](\hat{\chi}_0(\sigma))^2.\end{aligned} \tag{5.25}$$

Repeating the same calculation process, we can in principle determine how $\mathbf{Q}$ and $\dot{\mathbf{Q}}$ appear in all of the $\hat{\psi}_f^{(n)}$. However, since such a calculation is very tedious, we use another, shorter route.

Since we are studying the tree approximation, we can ignore the contributions of commutator $[\mathbf{Q}(t), \dot{\mathbf{Q}}]$. Then, since the quantum coordinate $\mathbf{Q}$ appears only through the combination $(\mathbf{x} + \mathbf{Q})$ and since $\dot{\mathbf{Q}}$ is independent of time, we

have

$$\frac{\partial}{\partial t}\hat{\psi}^f(x) = \bar{\partial}_t \hat{\psi}^f(x) + \dot{Q}_i \frac{\partial}{\partial Q_i}\hat{\psi}^f(x), \tag{5.26}$$

where the symbol $\bar{\partial}_t$ means that the time derivative operates, not on $\mathbf{Q}$, but on the time coordinate of x only. Let us use the symbol $\bar{\partial}_\mu \equiv (\partial_i, \bar{\partial}_t)$. Then the Heisenberg equation (1.1), with $\Lambda(\partial) = -\partial^2 - m^2$, becomes

$$\left[-\bar{\partial}^2 - m^2 - 2\bar{\partial}_t \dot{Q}_i \frac{\partial}{\partial Q_i} - \left(\dot{Q}_i \frac{\partial}{\partial Q_i}\right)^2\right]\hat{\psi}^f = F[\hat{\psi}^f]. \tag{5.27}$$

Since $\mathbf{x}$ and $\mathbf{Q}$ appear through the combination $\mathbf{x}_Q \equiv (\mathbf{x} + \mathbf{Q})$, (5.27) becomes

$$[-\bar{\partial}_t^2 + \nabla^2 - m^2 - 2(\dot{\mathbf{Q}}\cdot\nabla)\bar{\partial}_t - (\dot{\mathbf{Q}}\cdot\nabla)^2]\hat{\psi}^f = F[\hat{\psi}^f]. \tag{5.28}$$

For any vector $\mathbf{V}$, we define the transverse part $\mathbf{V}_T$ and longitudinal part $\mathbf{V}_L$ with respect to $\dot{\mathbf{Q}}$ as follows,

$$\mathbf{V}_T = \mathbf{V} - \frac{1}{\dot{Q}^2}\dot{\mathbf{Q}}(\dot{\mathbf{Q}}\cdot\mathbf{V}), \tag{5.29a}$$

$$\mathbf{V}_L = \frac{1}{\dot{Q}^2}\dot{\mathbf{Q}}(\dot{\mathbf{Q}}\cdot\mathbf{V}), \tag{5.29b}$$

where

$$\dot{Q}^2 = (\dot{\mathbf{Q}}\cdot\dot{\mathbf{Q}}). \tag{5.30}$$

When we introduce

$$\mathbf{x}' = \mathbf{x}_T + 1/(1-\dot{Q}^2)^{1/2}\mathbf{x}_L, \tag{5.31a}$$

$$t' = (1-\dot{Q}^2)^{1/2}t + 1/(1-\dot{Q}^2)^{1/2}(\dot{\mathbf{Q}}\cdot\mathbf{x}), \tag{5.31b}$$

we have

$$\nabla^2 - (\dot{\mathbf{Q}}\cdot\nabla)^2 = \nabla'^2 + \dot{Q}^2(\partial_t')^2 + 2(\dot{\mathbf{Q}}\cdot\nabla')\partial_t', \tag{5.32a}$$

$$\partial_t^2 + 2(\dot{\mathbf{Q}}\cdot\nabla)\partial_t = (\partial_t')^2 + \dot{Q}^2(\partial_t')^2 + 2(\dot{\mathbf{Q}}\cdot\nabla')\partial_t'. \tag{5.32b}$$

Then (5.28) becomes

$$[-\partial'^2 - m^2]\hat{\psi}^f = F[\hat{\psi}^f]. \tag{5.33}$$

It is not surprising that (5.31) is not a Lorentz transformation, because we are studying a static extended object so that the motion of the object is only Brownian motion caused by quantum fluctuations.

As has been pointed out, when we replace $\hat{\psi}_f^{(1)}$ in the dynamical map of $\hat{\psi}^f$ by $\hat{\chi}_0$ [cf. (5.10)], then $\mathbf{x}$ should be replaced by $\mathbf{x}_Q = \mathbf{x} + \mathbf{Q}$, including the replacement of $\mathbf{x}'$ and t' by $\mathbf{X}$ and T, respectively

$$\mathbf{X} = \mathbf{x}_{QT} + 1/(1-\dot{Q}^2)^{1/2}\mathbf{x}_{QL}, \tag{5.34a}$$

$$T = (1-\dot{Q}^2)^{1/2}t + 1/(1-\dot{Q}^2)^{1/2}(\dot{\mathbf{Q}}\cdot\mathbf{x}_Q). \tag{5.34b}$$

The rule for calculation of the dynamical map $\psi^f(x)$ in the tree approximation is now clear; first, follow the steps in Sec. 3 to obtain $\hat{\psi}^f(x)$ with $\mathbf{Q} = \mathbf{P} = 0$ (and therefore, with $\hat{\psi}_f^{(1)} \to \hat{\chi}_0$), and then replace $(\mathbf{x},t)$ by (X,T) in order to take into account $\mathbf{Q}$ and $\dot{\mathbf{Q}}$. The quantum coordinate appears only through $X_\mu = (\mathbf{X},T)$.

As a check of consistency of the calculation, we now show that the expansion of $\hat{\psi}^f$ in powers of $\mathbf{Q}$ and $\dot{\mathbf{Q}}$ leads to (5.25). To do this, we note that

$$\mathbf{X} = \mathbf{x} + \mathbf{Q} + \tfrac{1}{2}\dot{\mathbf{Q}}(\dot{\mathbf{Q}}\cdot\mathbf{x}) + \cdots, \tag{5.35}$$

$$T = t + (\dot{\mathbf{Q}}\cdot\mathbf{x}) + \cdots. \tag{5.36}$$

We thus have

$$\begin{aligned}\hat{\psi}^f &= \hat{\phi}_f(\mathbf{X}) + \hat{\chi}_0(X) + \cdots \\ &= \hat{\phi}_f(\mathbf{x}) + \hat{\chi}_0(x) + 2(\mathbf{Q}\cdot\nabla)\hat{\phi}_f(x) \\ &\quad + \tfrac{1}{2}[(\dot{\mathbf{Q}}\cdot\mathbf{x})(\dot{\mathbf{Q}}\cdot\nabla)\hat{\phi}_f(\mathbf{x}) \\ &\quad + (\mathbf{Q}\cdot\nabla)^2\hat{\phi}_f(\mathbf{x}) + 2(\mathbf{Q}\cdot\nabla)\hat{\chi}_0(x) + 2(\dot{\mathbf{Q}}\cdot\mathbf{x})\dot{\hat{\chi}}_0(x)] + \cdots.\end{aligned} \tag{5.37}$$

The terms in the square brackets are the contributions of the quantum coordinate to $\hat{\psi}_f^{(2)}$ when we use $\hat{\psi}_f^{(1)} = (\mathbf{Q}\cdot\nabla)\hat{\phi}_f + \hat{\chi}_0$ [cf. (5.10)] rather than $\hat{\psi}_f^{(1)} = \hat{\chi}_0$. Thus (5.37) agrees with the result, (5.25). This explicitly shows how $\mathbf{X}$ and T take care of the quantum coordinate when $\hat{\psi}_f^{(1)} = \hat{\chi}_0$ is used.

Note that $\phi_f(\mathbf{X}) = \langle 0|\psi^f(x)|0\rangle$ is not a c-number, but an operator, because it contains the quantum coordinate.

To consider the quantum corrections, we need to know how $\mathbf{Q}$ and $\dot{\mathbf{Q}}$ are ordered, because $Q_i\dot{Q}_j$ is not equal to $\dot{Q}_j Q_i$. For this purpose, let us note that the dynamical map of ψ^f has the form

$$\psi_f = \sum_n \int d\sigma_1 \cdots d\sigma_n \, g(x;\sigma_1\cdots\sigma_n) :\psi_f^{(1)}(\sigma_1)\cdots\psi_f^{(1)}(\sigma_n):, \tag{5.38}$$

where $\psi_f^{(1)} = (\mathbf{Q}\cdot\nabla)\phi_f + \chi_0$. Since $g(x;\sigma_1\cdots\sigma_n)$ is symmetric in $\sigma_1\cdots\sigma_n$, we see that the ordering among $\mathbf{Q}$ and $\dot{\mathbf{Q}}$ are symmetrized. In other words, any power consisting of $\mathbf{Q}$ and $\dot{\mathbf{Q}}$ in ψ^f contains all of the possible permutations among $\mathbf{Q}$ and $\dot{\mathbf{Q}}$. This will be called the symmetrization of the quantum coordinates.

The ψ^f with full quantum corrections can be obtained by the following method: first calculate ψ^f without the quantum coordinates by following the method in Sec. 4, replace $\mathbf{x}$ and t by $\mathbf{X}$ and T, and then symmetrize the ordering among $\mathbf{Q}$ and $\dot{\mathbf{Q}}$.

For example, the calculation of ϕ^f with quantum corrections proceeds as follows: first calculate $\hat{\psi}_f^{(n)}$ by solving the equations in (3.8) with $\hat{\psi}_f^{(1)} = \hat{\chi}_0(x)$. Then, the dynamical map of $\hat{\psi}^f$ in the tree approximation is given by

$$\hat{\psi}^f(x) = \sum_n^\infty \frac{1}{n!} :\hat{\psi}_f^{(n)}(x):, \tag{5.39}$$

in which the quantum coordinate is ignored. The classical field $\phi_f(\mathbf{x})$ with quantum correction is given by

$$\phi_f(x) = \sum_n^\infty \frac{1}{n!}\langle 0|\hat{\psi}_f^{(n)}(x)|0\rangle. \tag{5.40}$$

To take into account the quantum coordinate, we replace $\mathbf{x}$ by $\mathbf{X}$ and symmetrize the order among $\mathbf{Q}$ and $\dot{\mathbf{Q}}$. The result $\phi_f(\mathbf{X})$ describes both the classical and quantum *mechanical* properties of the extended object; when we ignore $\mathbf{Q}$ and $\dot{\mathbf{Q}}$ we obtain a classical object, while a quantum mechanical object appears when $\mathbf{x}$ is disregarded.

Let us now calculate the ground-state energy in presence of the extended object in the tree approximation. When we assume $\Lambda(\partial) = -\partial^2 - m^2$, we have

$$\begin{aligned}H[Q,\dot{Q}] &= \langle 0|H|0\rangle \\ &= \int d^3x \, \{\tfrac{1}{2}[\dot{\hat{\phi}}_f^2(\mathbf{X}) + (\nabla\hat{\phi}_f(\mathbf{X}))^2\end{aligned}$$

$$+ m^2\hat{\phi}_f^2(\mathbf{X})] + V[\hat{\phi}_f(\mathbf{X})]\}, \tag{5.41}$$

where ∇ is the derivative with respect to $\mathbf{x}$, and

$$\frac{\delta V}{\delta\hat{\phi}_f} = F[\hat{\phi}_f]. \tag{5.42}$$

The $H(Q,\dot{Q})$ is the quantum mechanical Hamiltonian of the extended object. We have

$$\dot{\hat{\phi}}_f(\mathbf{X}) = (\dot{X}\cdot\nabla_X)\hat{\phi}_f(\mathbf{X})$$
$$= 1/(1-\dot{Q}^2)^{1/2}(\dot{Q}\cdot\nabla_X)\hat{\phi}_f(\mathbf{X}), \tag{5.43}$$

where ∇_X is the derivative with respect to $\mathbf{X}$. In deriving (5.43) we used (5.34a) and the relations $\dot{\mathbf{Q}}_T = 0$ and $\dot{\mathbf{Q}}_L = \dot{\mathbf{Q}}$. Since

$$[(\mathbf{Q}\cdot\nabla_X)\hat{\phi}_f(\mathbf{X})]^2 = \dot{Q}^2(\nabla_{XL}\hat{\phi}_f(\mathbf{X}))^2, \tag{5.44}$$

we have

$$(\dot{\hat{\phi}}_f(\mathbf{X}))^2 = \dot{Q}^2/(1-\dot{Q}^2)(\nabla_{XL}\hat{\phi}_f(\mathbf{X}))^2. \tag{5.45}$$

Similarly, we have

$$(\nabla\hat{\phi}_f(\mathbf{X}))^2 = (\nabla_X\hat{\phi}_f(\mathbf{X}))^2 + \dot{Q}^2/(1-\dot{Q}^2)(\nabla_{XL}\hat{\phi}_f(\mathbf{X}))^2. \tag{5.46}$$

From the variable change of the integration in (5.41),

$$d^3x = (1-\dot{Q}^2)^{1/2}d^3X, \tag{5.47}$$

we obtain

$$H(Q,\dot{Q} = (1-\dot{Q}^2)^{1/2}M_0 + \dot{Q}^2/(1-\dot{Q}^2)^{1/2}M_1(\dot{Q}), \tag{5.48}$$

where M_0 is the constant defined by

$$M_0 = \int d^3x\,\{\tfrac{1}{2}[(\nabla\hat{\phi}_f(\mathbf{x}))^2 + m^2\hat{\phi}_f^2(\mathbf{x})] + V[\hat{\phi}_f(\mathbf{x})]\} \tag{5.49}$$

and M_1 is given by

$$M_1(\dot{Q}) = \int d^3x\,(\nabla_L\cdot\hat{\phi}_f(\mathbf{x}))^2. \tag{5.50}$$

In particular, when $M_0 = M_1$ we have

$$H(Q,\dot{Q}) = M_0/(1-\dot{Q}^2)^{1/2}, \tag{5.51}$$

which has the relativistic energy expression and leads to

$$H = (p^2 + M_0^2)^{1/2} \tag{5.52}$$

and

$$\dot{\mathbf{Q}} = \mathbf{p}/(p^2 + M_0^2)^{1/2}. \tag{5.53}$$

In one-dimensional models, $\hat{\phi}_f(x)$ satisfies

$$\frac{d^2}{dx^2}\hat{\phi}_f(x) = m^2\hat{\phi}_f(x) + \delta V[\hat{\phi}_f(x)]/\delta\hat{\phi}_f(x). \tag{5.54}$$

Multiplying $(d/dx)\hat{\phi}_f(x)$ on both sides of (5.54), we have

$$\frac{1}{2}\frac{d}{dx}\left(\frac{d}{dx}\hat{\phi}_f(x)\right)^2 = \frac{d}{dx}[\tfrac{1}{2}m^2\phi_f^2(x) + V[\hat{\phi}_f(x)]], \tag{5.55}$$

which leads to $M_0 = M_1$. Therefore any relativistic one-dimensional model leads to (5.51) in the tree approximation. The result (5.52) has been conjectured in a $\lambda\phi^4$-model and was shown in the lower order expansion of p^2 in Ref. 8. In general cases, M_1 depends on $\dot{\mathbf{Q}}$, and therefore $M_0 = M_1$ is not always expected to hold. This is not surprising because we are considering the static extended objects.

6. SUMMARY

A general formulation of quantum field theory for a system with extended objects was presented. It was shown that creation of extended objects does not upset the renormalization. We also showed that when we specify the solution of the classical Euler equation, the entire form of the dynamical map of ψ^f in the tree approximation can be determined. Furthermore, this dynamical map in the tree approximation determines the dynamical map of ψ^f with quantum corrections. A general method for the treatment of the quantum coordinate Q was given. It was shown that a static extended object exhibits Brownian motion due to the quantum fluctuation effect. When this fluctuation effect is negligibly small, the object behaves as a classical object. We found that, in the dynamical map of ψ^f (and therefore, of the Hamiltonian) appearance of Q and $\dot{Q}$ is symmetrical. This leads us to the commonly used quantization method in which the arrangement of Q and $\dot{Q}$ is symmetrized [e.g., $Q\dot{Q}\to(1/2)\times(Q\dot{Q}+\dot{Q}Q)$]. This traditional method is now justified by our study in which the classical theory comes out of the quantum field theory.

Let us close this section by making a comment on the quantum coordinates of many extended objects. When $f(x)$ has the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} f^{(i)}(\mathbf{x} - a^{(i)}), \tag{6.1}$$

in which each $f^{(i)}$ satisfies (1.5), i.e.,

$$\Lambda(\partial)f^{(i)}(\mathbf{x}) = 0, \tag{6.2}$$

the boson transformation with this $f(\mathbf{x})$ creates n extended objects, the positions of which are given by $\mathbf{a}_i$. When we make the transformation $(\mathbf{a}^{(1)}\to\mathbf{a}^{(1)} + \mathbf{a}, \mathbf{a}^{(i)}\to\mathbf{a}^{(i)}; i\neq 1)$, the new $f(\mathbf{x})$ also satisfies (1.5) because $f^{(1)}(\mathbf{x} - \mathbf{a}_1 - \mathbf{a})$ satisfies (6.2). The ϕ^f depends on $\mathbf{a}_i$ and the Euler equation for ϕ^f is invariant under the above transformation, implying that $(\partial/\partial a^{(1)}j)\phi^f (j = 1,2,3)$ appear to be wavefunctions of zero-energy quantum modes. This mode is called the translation modes of the first extended object. Following the consideration in the previous sections, we find the quantum coordinate $\mathbf{Q}^{(1)}$ which is associated with the translation modes of the first object. In a similar manner, we find the quantum coordinate for each object. The $\mathbf{Q}^{(i)}$ is the quantum coordinate of the ith object and it appears in ψ^f through the combination $(\mathbf{a}^{(i)} + \mathbf{Q}^{(i)})$. When the quantum fluctuation of $\mathbf{Q}^{(i)}$ is observable, the ith object behaves as a quantum-mechanical object.

Note added in proof: The analysis of the quantum coordinates in multi-soliton system (such as multi-vortices system) is more involved than the one illustrated in the summary. The linear independence of wavefunctions must be considered and independent canonical variables (and their canonical momenta) must be identified. The study of this subject is in progress.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and

Engineering Research Council of Canada and the Faculty of Science of the University of Alberta. One of us (M.U.) would like to express his gratitude for the hospitality of the Theoretical Physics Institute, University of Alberta, where this work was done. Two of the authors (H.U. and M.U.) acknowledge the Max Planck Institute at München, where a part of this work was done.

[1]For the review of the soliton, see, for example, A. C. Scott, F. Chu, and D. McLaughlin, Proc. IEEE **61**, 1443 (1973).

[2]There are many works on the quantum effects in solitons. Many references are found, for example, in the following review articles: R. Rajaraman, Phys. Rep. C **21**, 227 (1975); J. L. Gervais and A. Neveu, Phys. Rep. C **23**, 237 (1976); R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977).

[3]See, for example, J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw–Hill, New York, 1968).

[4]See, for example, R. D. Parks, *Superconductivity* (Dekker, New York, 1969).

[5]L. Leplae, F. Mancini, and H. Umezawa, Phys. Rev. B **2**, 3594 (1970); H. Matsumoto, N.Y. Papastamatiou, and H. Umezawa, Nucl. Phys. B **82**, 45 (1974); *ibid*, B **97**, 90 (1975).

[6]H. Matsumoto, P. Sodano, and H. Umezawa, Phys. Rev. D **19**, 511 (1979).

[7]H. Matsumoto, G. Oberlechner, M. Umezawa, and H. Umezawa, J. Math. Phys. **20**, 2088 (1979).

[8]J. L. Gervais and B. Sakita, Phys. Rev. D **11**, 2943 (1975); J. L. Gervais, A. Jevicki, and B. Sakita, Phys. Rev. D **12**, 1038 (1975); C. G. Callan, Jr. and D. J. Gross, Nucl. Phys. B **93**, 29 (1975);

[9]G. Semenoff, H. Matsumoto, and H. Umezawa, (to be published in Prog. Theor. Phys.)

[10]R. F. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **10**, 4114, 4136 (1974).

[11]R. J. Glauber, Phys. Rev. Lett. **10**, 84 (1963); Phys. Rev. **130**, 2529 (1963); J. R. Klauder, J. Math. Phys. **4**, 1055, 1058 (1963); **5**, 177 (1963). E. C. G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963). See also J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York 1968).

[12]M. Wadati, H. Matsumoto, and H. Umezawa, Phys. Lett. B **73**, 448 (1978); Phys. Rev. D **18**, 520 (1978).

[13]H. Umezawa and H. Matsumoto, (the Symposium on *Symmetries in Science* held at Southern Illinois University at Carbondale, Ill., 1979, Albert Einstein Centennial Week).

Nondiffractive scattering: Scattering from kaleidoscopes [a)]

Bill Sutherland
Department of Physics, University of Utah, Salt Lake City, Utah 84112

(Received 11 September 1979; accepted for publication 28 December 1979)

We introduce and exactly solve a series of multidimensional scattering problems. The potential is a generalization of the delta function potential and has the symmetry of a kaleidoscope. The solution demonstrates the property of nondiffractive scattering. The bound state energies are calculated exactly for all cases. We prove a remarkable identity for certain finite groups.

I. INTRODUCTION

This paper generalizes existing exactly soluble models for nondiffractive scattering. Such phenomena has been of interest for years; we trace our own line of investigation back to an early paper of H. Bethe.[1] In this paper, Bethe investigated a model of magnetism by means of a particular form for the wavefunction—a form now called "Bethe's ansatz"—which embodies the assumption of nondiffractive scattering.

Subsequent developments of this concept have been made by an investigators; we mention the work of McGuire,[2] Lieb,[3] Yang,[4] Sutherland,[5] and Baxter[6] as indicating the range of development. A brief summary of work to date is contained in a recent review by the author.[7]

An alternate line of investigation of nondiffractive phenomena is through the study of "soliton behavior." Increasingly these two areas are seen to be essentially the same.

This paper will introduce and solve a multidimensional scattering problem, which may be considered as a generalization of the delta function potential.

II. THE PROBLEM AND A GUESS AT THE SOLUTION

(A) We are in d-dimensional space and consider a discrete set H of hyperplanes h such that the set is invariant under reflection about any one of the hyperplanes; such a set we call a Coxeter set. If the set is finite, we will use the more descriptive term kaleidoscope in recognition of the most common example. Let $\boldsymbol{\alpha}_h$ be a unit normal to the hyperplane h. We then wish to solve the equation:

$$-\nabla^2\psi = k^2\psi, \tag{1}$$

subject to the homogeneous boundary condition that the discontinuity of the normal derivative of ψ across the hyperplanes is proportional to the value of ψ on the hyperplane. Further, the constant of proportionality is the same for all hyperplanes.

The equation of hyperplane h we write as:

$$\mathbf{x}\cdot\boldsymbol{\alpha}_h = h. \tag{2}$$

Let $\mathbf{x}_h = \boldsymbol{\alpha}_h h + \mathbf{x}_\perp$ be a point on the hyperplane h. Then our boundary conditions say:

$$\lim_{\epsilon \to 0+} [\boldsymbol{\alpha}_h\cdot\nabla\psi(\mathbf{x}_h + \epsilon\boldsymbol{\alpha}_h) - \boldsymbol{\alpha}_h\cdot\nabla\psi(\mathbf{x}_h - \epsilon\boldsymbol{\alpha}_h)] = 2c\psi(\mathbf{x}_h). \tag{3}$$

(B) We now look more closely at the action of the reflections in the hyperplanes. These reflections we denote by α_h, and they generate a group G' called a Coxeter group, defined by the relations (and these relations alone):

$$\alpha_h^2 = I, \quad (\alpha_h\alpha_{h'})^{p(h,h')} = I, \tag{4}$$

where $p(h,h')$ denotes a positive integer. The action of α_h is given by

$$\alpha_h(\mathbf{x}) = \mathbf{x}' = \mathbf{x} - 2\boldsymbol{\alpha}_h(\boldsymbol{\alpha}_h\cdot\mathbf{x}) + 2h\boldsymbol{\alpha}_h. \tag{5}$$

For a vector (translation) however,

$$\boldsymbol{\alpha}_h(\mathbf{k}) = \mathbf{k}' = \mathbf{k} - 2\boldsymbol{\alpha}_h(\boldsymbol{\alpha}_h\cdot\mathbf{k}). \tag{6}$$

Thus parallel hyperplanes have the same effect on a vector $\mathbf{k}$, and this smaller group we denote by G; it is a finite Coxeter group. It will be generated by a representative set of nonparallel hyperplanes which we assume to meet at the origin. If T is the normal subgroup of translations of G', then $G = G'/T$. In general, G' will be infinite while G will be finite. In case $G = G'$, we have a finite Coxeter group, and it is this situation we largely deal with in this paper, beginning with case (E) of Sec. III. The geometry of these transformations is indicated in Fig. 1.

(C) the hyperplanes partition space into cells. Let $g\in G$ and $g'\in G'$. We choose one cell to be the "fundamental cell" and let the normals be directed outward from the cell. Let $\mathbf{x}_0$ be a point in the fundamental cell, and then label the cells so that $g'(\mathbf{x}_0)$ is contained in cell g'.

Our basic guess for a solution to the problem as posed is that $\psi(\mathbf{x})$ in cell g' is given by

$$\psi_{g'}(\mathbf{x}) = \sum_{g\in G} A(g',g)\exp(i\mathbf{x}\cdot\mathbf{k}_g). \tag{7}$$

$A(g',g)$ are numbers, dependent on $\mathbf{k}$, to be determined by the boundary conditions. The symbol $\mathbf{k}_g$ denotes the image of the vector $\mathbf{k}$ under the group element g.

This form may be considered as a generalization of the method of images. It is also a statement of the nondiffractive property.

If $c = 0$,

$$A(g',g) = A(g). \tag{8a}$$

If $c = +\infty$,

$$A(g',g) = A(g')(-1)^g. \tag{8b}$$

(D) Suppose we are near the hyperplane h which divides cell g_1' from $g_2' = \alpha_h g_1'$. We suppress the subscript h for now. Let α be directed from 1 to 2. Then

[a)]Work supported in part by the National Science Foundation under grant DMR77-08474.

 0022-2488/80/071770-06$1.00

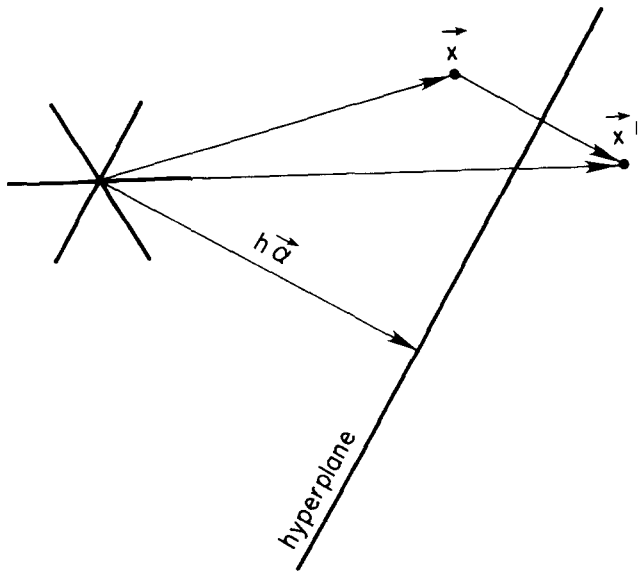


FIG. 1. Geometry of a reflection about a hyperplane.

$$\nabla\psi = i\sum_g \mathbf{k}_g A(g)e^{i\mathbf{x}\cdot\mathbf{k}_g}. \tag{9}$$

Let $\mathbf{x}_0$ be a point on h. Then continuity requires

$$\sum_g [A_2(g) - A_1(g)]\, e^{i\mathbf{k}_g\cdot\mathbf{x}_0} = 0, \tag{10a}$$

while the boundary condition requires that

$$\sum_g i\alpha\cdot\mathbf{k}_g\, [A_2(g) - A_1(g)]\, e^{i\mathbf{k}_g\cdot\mathbf{x}_0} = 2c\psi(\mathbf{x}_0). \tag{10b}$$

Let us consider two particular k's in this sum:

$$\mathbf{k} = \mathbf{k}_g,$$
$$\mathbf{k}' = \mathbf{k}_{\alpha g} = \mathbf{k}_\alpha = \mathbf{k} - 2\alpha(\alpha\cdot\mathbf{k}). \tag{11}$$

We now seek cancellations of such corresponding pairs of terms in Eqs. (10a) and (10b).

First, let

$$\mathbf{x}_0 = h\alpha + \mathbf{x}_\perp, \quad \mathbf{x}_\perp\cdot\alpha = 0. \tag{12}$$

Then in the exponentials,

$$\mathbf{k}'\cdot\mathbf{x}_0 = [\mathbf{k} - 2\alpha(\alpha\cdot\mathbf{k})]\cdot\mathbf{x}_0 = \mathbf{k}\cdot\mathbf{x}_0 - 2h(\alpha\cdot\mathbf{k}). \tag{13}$$

Factoring out the common exponential factors of pairs of terms in (10a) and (10b) we find

$$A_1(k) - A_2(k) + e^{-2ih(\alpha\cdot\mathbf{k})}[A_1(k') - A_2(k')] = 0, \tag{14a}$$

and

$$i(\alpha\cdot\mathbf{k})(A_2(k) - A_1(k) - e^{-2ih(\alpha\cdot\mathbf{k})}[A_2(k') - A_1(k')]) = 2c(A_2(k) + e^{-2ih(\alpha\cdot\mathbf{k})}A_2(k')). \tag{14b}$$

III. CONDITIONS FOR CONSISTENCY

(E) We now assume that G' is finite so that $G' = G$, and $h = 0$; we call this situation a kaleidoscope.

Conditions (14a) and (14b) may then be put into the matrix form:

$$\begin{vmatrix} A_1(k') \\ A_2(k) \end{vmatrix} = \frac{1}{-1+ix}\begin{vmatrix} 1 & ix \\ ix & 1 \end{vmatrix}\begin{vmatrix} A_1(k) \\ A_2(k') \end{vmatrix}, \tag{15}$$

with

$$x \equiv (\alpha\cdot\mathbf{k})/c. \tag{16}$$

The picture of such a scattering is indicated in Fig. 2. The matrix is unitary, provided $\mathbf{k}$ is real, as is appropriate for a scattering problem.

(F) Let us define the following object belonging to the group ring:

$$A(g) = \sum_{g_1\in G} g_1^{-1}A(g_1, g\, g_1^{-1}). \tag{17}$$

The action of α on this object is

$$\begin{aligned}\alpha A(g) &= \sum_{g_1\in g} \alpha g_1^{-1}A(g_1, g\, g_1^{-1}) \\ &= \sum_{g_1\in g} (g_1\alpha)^{-1}A(g_1, g\, g_1^{-1}) \\ &= \sum_{g_1\in g} g_1^{-1}A(g_1\alpha, g\alpha\, g_1^{-1}).\end{aligned} \tag{18}$$

Then our pairwise connections may be put in the concise form

$$A(g\alpha) = \frac{1+ix_{g,\alpha}\alpha}{-1+ix_{g,\alpha}}A(g) \equiv T_\alpha(g)A(g), \tag{19}$$

where

$$x_{g,\alpha} = (\alpha\cdot\mathbf{k}_g)/c. \tag{20}$$

(G) We may finally state the condition for consistency in the determination of the A's:

If $\alpha_\tau (\tau = 1,...,d)$ are the generators of G, let $\sigma_j (j = 1,...,N)$ be a sequence of generators $\alpha_{\tau_1}, \alpha_{\tau_2},..., \alpha_{\tau_N}$ such that

$$\sigma_N\sigma_{N-1},...,\sigma_2\sigma_1 = I. \tag{21}$$

Define

$$g_j = \sigma_{j-1}\sigma_{j-2}\cdots\sigma_2\sigma_1, \quad g_1 = I. \tag{22}$$

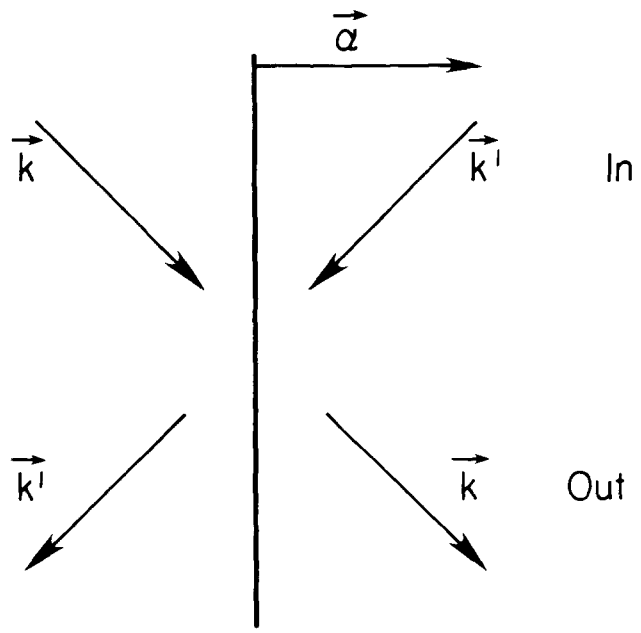


FIG. 2. A binary scattering from a hyperplane.

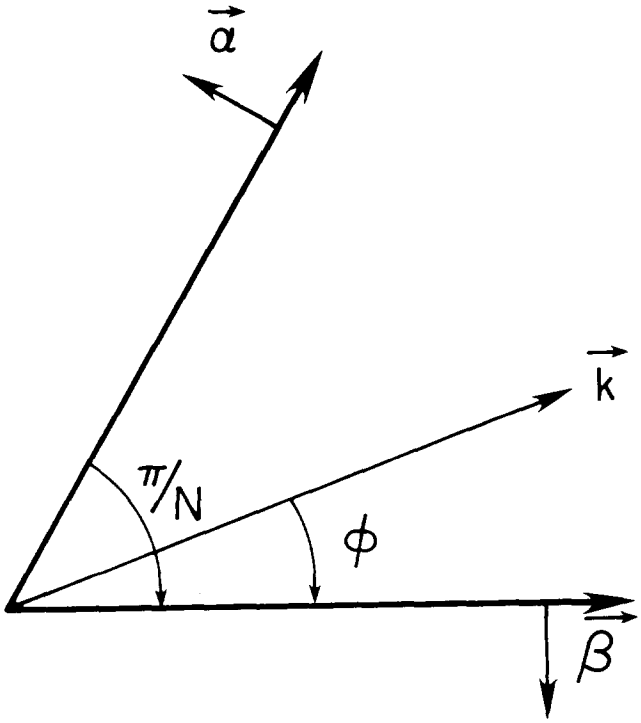


FIG. 3. Geometry and coordinates of a dihedral kaleidoscope.

Then the following is an identity for all $\mathbf{k}$:

$$T_{\sigma_N}(g_N)\,T_{\sigma_{N-1}}(g_{N-1})\cdots T_{\sigma_2}(g_2)T_{\sigma_1}(g_1) = \frac{1+ix_N\sigma_N}{-1+ix_N}\cdots\frac{1+ix_2\sigma_2}{-1+ix_2}\cdot\frac{1+ix_1\sigma_1}{-1+ix_1} = I, \quad (23)$$

where

$$x_j = (\boldsymbol{\sigma}_j\cdot\mathbf{k}_{g_j})/c. \quad (24)$$

(H) The problem of determining all finite G—very similar to determining all regular polytopes in d dimensions or all compact Lie groups—has been solved by Coxeter.[8] The results are most easily expressed in graphical notations as follows:

Let a point represent each of the d generators α_τ. If $p(\tau,\tau')$ in the defining relation

$$(\alpha_\tau\alpha_{\tau'})^{p(\tau,\tau')} = I, \quad (25)$$

(i) is equal to 3 we place a line between points τ and τ'; (ii) is equal to $p>3$, we place a line with a f over it between points τ and τ'; (iii) is equal to 2, we do nothing, and note that the two generators commute.

With this notation, we list in Table I all finite Coxeter groups or kaleidoscopes, eliminating unconnected graphs.

(I) To say the group G is generated by the generators $\alpha_1,\ldots,\alpha_d$ and defined by the relations of Eq. (4), implies that all such "words" as Eq. (21) are a consequence of the defining relations Eq. (4). Thus we need only prove case (G) Sec. III for the defining relations.

First, let us consider $\alpha^2 = 1$.

$$g_1 = I, \quad x_1 = (\boldsymbol{\alpha}\cdot\mathbf{k})/c, \quad (26a)$$

$$g_2 = \alpha, \quad x_2 = (\boldsymbol{\alpha}\cdot\mathbf{k}_\alpha)/c = \{\boldsymbol{\alpha}\cdot[\mathbf{k}-2(\mathbf{k}\cdot\boldsymbol{\alpha})\boldsymbol{\alpha}]\}/c = -(\boldsymbol{\alpha}\cdot\mathbf{k})/c = -x_1. \quad (26b)$$

Thus, we need only verify that

$$\frac{1-ix_1\alpha}{-1-ix_1}\times\frac{1+ix_1\alpha}{-1+ix_1} = T_\alpha(\alpha)T_\alpha(I) = \frac{1+x_1^2\alpha^2}{1+x_1^2} = I. \quad (27)$$

IV. PROOF OF A THEOREM

(J) We now need to verify the dihedral relation: If

$$\alpha^2 = \beta^2 = (\beta\alpha)^N = I, \quad (28)$$

then

$$I = \frac{1+ix_{2N}\beta}{-1+ix_{2N}}\cdots\frac{1+ix_2\beta}{-1+ix_2}\,\frac{1+ix_1\alpha}{-1+ix_1}, \quad (29)$$

where

$$\alpha_j = \begin{cases}\alpha, & j \text{ odd},\\ \beta, & j \text{ even},\end{cases} \quad (30)$$

and x_j, g_j are as previously defined by Eqs. (22) and (24).

We first make a parameterization of the x_j as follows: Let α and β reflection planes be separated by an angle π/N, the normals $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ be directed outward from this wedge, and the positive sense of rotation be from α to β. We then choose ϕ to be the angle of $\mathbf{k}$ with respect to the β plane bounding the wedge. The length of $\mathbf{k}$ is k. This geometry is shown in Fig. 3.

With this notation,

$$\mathbf{k}_1 = \mathbf{k}_\alpha, \quad x_1 = (\boldsymbol{\alpha}\cdot\mathbf{k}_1)/c = (k/c)\sin(\phi+\pi/N),$$
$$\mathbf{k}_2 = \mathbf{k}_{\beta\alpha}, \quad x_2 = (\boldsymbol{\alpha}\cdot\mathbf{k}_2)/c = (k/c)\sin(\phi+2\pi/N) \quad (31)$$

Subsequent x_j's are obtained from these by rotation by $2\pi/N$. Thus we have the general formula

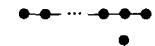
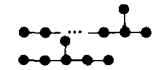
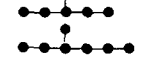
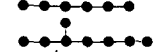
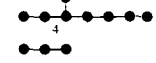
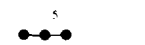
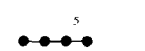
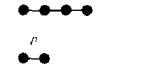
$$x_j = (k/c)\sin(\phi+j\pi/N) \equiv (x/i)\phi_j, \quad (32)$$

with $x = ik/c$, and

$$\phi_j(\phi) = \sin(\phi+j\pi/N). \quad (33)$$

We remark the following properties of the functions $\phi_j(\phi)$:

TABLE I. Coxeter graphs and binding energies for all the finite Coxeter groups. The symbol n designates the number of vertices of the graph and the dimensionality of the scattering problem.

Symbol	Graph	Binding Energy $= -E_0$
A_n		$n(n+1)(n+2)/6$
B_n		$n[2n^2-3n+4+3\sqrt{2}(n-1)]$
D_n		$n(n-1)(2n-1)/3$
E_6		156
E_7		399
E_8		1240
F_4		$4(14+9\sqrt{2})$
H_3		$31+12\sqrt{5}$
H_4		$4(83+36\sqrt{5})$
$I_2(p)$		$\sin^{-2}(\pi/2p) = 1,2,4,2(2+\sqrt{2}),$ $2(3+\sqrt{5}), 4(2+\sqrt{3}),\cdots$

$$\text{(a) } \phi_{j+N} = -\phi_j,$$
$$\text{(b) } \phi_j(\phi + \pi/N) = \phi_{j+1}(\phi). \tag{34}$$

These will be very important in our subsequent discussion.

(K) By multiplying Eq. (29) on the left by the inverse of the left-hand factor, then again on the left by the inverse of the next factor, etc., until we have eliminated N factors, we may put our dihedral relation into the equivalent form:

$$\left[1 + x\phi_N\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right]\cdots(1 + x\phi_2\beta)(1 + x\phi_1\alpha)$$
$$= \left[1 + x\phi_1\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right]\cdots(1 + x\phi_{N-1}\alpha)(1 + x\phi_N\beta). \tag{35}$$

In fact, we will prove the more general result: If $\alpha^2 = \beta^2 = I$, then

$$\left[1 + x\phi_N\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right]\cdots(1 + x\phi_3\alpha)(1 + x\phi_2\beta)(1 + x\phi_1\alpha)$$
$$- \left[1 + x\phi_1\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right]\cdots(1 + x\phi_{N-2}\beta)(1 + x\phi_{N-1}\alpha)$$
$$\times(1 + x\phi_N\beta) = x^N\sum_{j=1}^{N}\phi_j\left[\underbrace{\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\cdots\alpha\beta\alpha}_{N\text{ factors}} - \underbrace{\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\cdots\beta\alpha\beta}_{N\text{ factors}}\right]. \tag{36}$$

The upper choice is for N odd, the lower choice is for N even.

We may simplify the final form by

$$\sum_{j=1}^{N}\phi_j = -\sin N\phi/2^{N-1}. \tag{37}$$

We write our expression as $G(x,\phi)$ and say it is a sum of a first term and a second term.

Let us pick a particular product of K factors:

$$\Phi^K = \phi_{j1}\phi_{j2}\cdots\phi_{jK} = \prod_{l=1}^{K}\phi_{jl} \equiv \Phi^K(\phi), \tag{38}$$

with $1 \leqslant j1 < j2 < \cdots < jK \leqslant N$. This multiplies a particular product of K α's and β's in the first term which we denote by

$$\sigma(\Phi^K) = \sigma_{jk}\cdots\sigma_{j2}\sigma_{j1}, \tag{39}$$

and another product of K α's and β's in the first term which we denote by

$$\phi(-\Phi^K) = \sigma_{j1+N}\sigma_{j2+N}\cdots\sigma_{JK+N}, \tag{40}$$

where as before,

$$\sigma_j = \begin{cases}\alpha & \text{if } j \text{ odd},\\ \beta & \text{if } j \text{ even}.\end{cases} \tag{41}$$

Then we may write G in the concise form:

$$G = \sum_{K=0}^{N} x^K \sum_{\Phi^K}\Phi^K[\sigma(\Phi^K) - \sigma(-\Phi^K)]$$
$$= \sum_{K=0}^{N} x^K \sum_{\pm\Phi^K}\Phi^K\sigma(\Phi^K). \tag{42}$$

(L) We first prove the following:

Lemma 1: For N even, odd powers of x vanish; while for N odd, even powers of x vanish.

Proof: Upon repeated use of $\alpha^2 = \beta^2 = I$, we see that $\sigma(\Phi^K)$ will be of the form:

$$\sigma(\Phi^K) = \begin{cases}\alpha\beta\alpha\cdots\beta\alpha\\ \beta\alpha\beta\cdots\alpha\beta\end{cases} K \text{ odd, or } \begin{cases}\alpha\beta\alpha\cdots\alpha\beta\\ \beta\alpha\beta\cdots\beta\alpha\end{cases} K \text{ even}. \tag{43}$$

In all cases, the number of factors will be $\leqslant K$.

However, for N even, K odd, we find

$$\sigma(\Phi^K) = \sigma(-\Phi^K), \tag{44}$$

while when N odd, K even, Eq. (44) is also true.

Inserting these results into the concise form for G, we prove the Lemma.

(M) We now wish to prove a second Lemma

Lemma 2:

$$G(\phi + \pi/N) = -G(\phi). \tag{45}$$

Proof: We need only to prove the statement for even K, even N or odd K, odd N; we henceforth restrict ourselves to these terms.

Let us begin with a particular $\Phi(\phi)$ (K suppressed): $\Phi = \pm\Pi_{l=1}^{K}\phi_{jl}$, and transform it by

$$\Phi \to \Phi_L(\phi) \equiv (-1)^L\prod_{l=1}^{K}\phi_{jl+L}(\phi)$$
$$= (-1)^L\prod_{l=1}^{K}\phi_{jl}(\phi + L\pi/N)$$
$$= (-1)^L\Phi(\phi + L\pi/N). \tag{46}$$

This relationship, for arbitrary L, defines an equivalence relation and thus partitions the Φ's into classes.

Let us consider

$$\Psi(\phi) \equiv \sum_{L}\Phi_L(\phi), \tag{47}$$

where the sum is over all members of a class. We have the relation

$$\Psi(\phi + \pi/N) = -\Psi(\phi). \tag{48}$$

For $L =$ even, $\sigma(\Phi_L) = \sigma(\Phi)$; while for $L =$ odd, $\sigma(\Phi_L) = \sigma(-\Phi_1) = \sigma(\Phi)$. Thus $\sigma(\Phi_L)$ is the same for all nembers of a class, and we thus write $\sigma(\Psi)$. (Remember K, $N =$ even throughout this Lemma.)

This allows us to write G in the form:

$$G(x,\phi) = \sum_{\text{classes}} x^K\Psi(\phi)\sigma(\psi). \tag{49}$$

Using the periodicity properties of Ψ we then have our Lemma.

We are now ready to prove our Theorem.

Theorem:

$$G(x,\phi) = x^N\left[\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\cdots\alpha\beta\alpha - \begin{pmatrix}\beta\\ \alpha\end{pmatrix}\cdots\beta\alpha\beta\right]\left(-\frac{\sin N\phi}{2^{N-1}}\right). \tag{50}$$

Proof: Since $G(\phi + \pi/N) = -G(\phi)$, we have the Fourier expansion,

$$G = \sum_{l=-\infty}^{+\infty} e^{i(2l+1)N\phi}A_l, \tag{51}$$

with $A_l^* = A_{-(l+1)}$. But G is a polynomial in $e^{i\phi}$ with no power greater than N nor less than $-N$.

Thus, we need only evaluate the coefficients of $e^{iN\phi}$ and $e^{-iN\phi}$ in

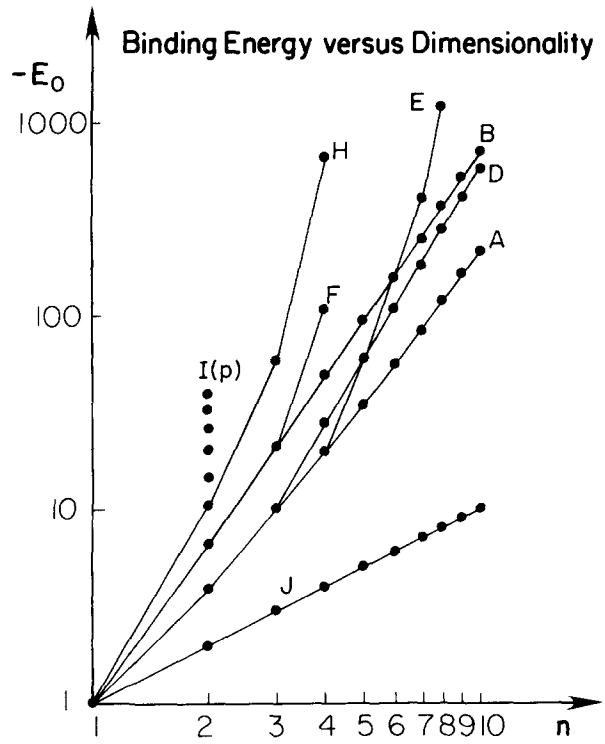


FIG. 4. Binding energy as a function of dimensionality in a log-log plot. The values are listed in Table I. We have included the family J of n unconnected points as a reference. Points of a family are connected by line segments.

$$x^K\left[\binom{\alpha}{\beta}\cdots\alpha\beta\alpha-\binom{\beta}{\alpha}\cdots\beta\alpha\beta\right]\prod_{j=1}^{N}\phi_j. \tag{52}$$

But

$$\prod_{j=1}^{N}\phi_j = e^{iN\phi}\frac{e^{(i\pi/N)N(N+1)/2}}{(2i)^N}+e^{-iN\phi}\frac{e^{(-i\pi/N)N(N+1)/2}}{(-2i)^N} + \cdots = -(1/2i)(e^{iN\phi}-e^{-iN\phi})+\cdots. \tag{53}$$

Thus, the theorem.

V. BOUND STATES

(O) If the strength of the interaction c is negative, then there is a unique solution which is bound at the origin. That is, we seek the energy E_0 of the state for which

$$\psi(\mathbf{x})\to 0 \quad \text{as} \quad |\mathbf{x}|\to\infty. \tag{54}$$

The solution given in Eq. (7) will only satisfy this condition provided the $\mathbf{k}$ is pure imaginary

$$\mathbf{k}_g = i\boldsymbol{\kappa}_g, \quad E_0 = -\kappa^2; \tag{55}$$

and only exponentials occur for which

$$\boldsymbol{\kappa}_g\cdot\mathbf{x}>0. \tag{56}$$

Within each cell there will be only one $\boldsymbol{\kappa}_g$ which satisfies this inequality.

The lowest energy state is always symmetric, and thus we may suppose that the generators acting on the A's of the ground state belong to the symmetric representation $\alpha = 1$. The coefficients in the connection formulas are then simply numbers,

$$\frac{1+ix_g}{-1+ix_g} = T(g). \tag{57}$$

In order that our boundary conditions not generate additional exponentials which will increase at infinity, the $\mathbf{k}$ must then be a zero of the coefficients in the connection equations. Thus, we have the d equations for the vector $\boldsymbol{\kappa}$:

$$\boldsymbol{\kappa}\cdot\boldsymbol{\alpha}_j = -1, \quad j=1,2,\ldots,d. \tag{58}$$

We have taken $c=-1$, since the energies are trivially normalized by c^2.

(P) It is convenient to seek a solution for $\boldsymbol{\kappa}$ as an expansion in the $\boldsymbol{\alpha}_j$,

$$\boldsymbol{\kappa} = -\kappa_j\boldsymbol{\alpha}_j. \tag{59}$$

The scalar products of $\boldsymbol{\alpha}_j$'s are elements of the Coxeter matrix M

$$\boldsymbol{\alpha}_j\cdot\boldsymbol{\alpha}_l = -\cos\pi/p_{jl} \equiv M_{jl} = M_{lj}. \tag{60}$$

The integer p_{jl} is of course, the integer in the defining relation

$$(\alpha_j\alpha_l)^{p_{jl}} = 1. \tag{61}$$

Values of $-\cos\pi/p$ for $p=(1,2,3,4,5,6)$ are $(1,0,-1/2,-1/\sqrt{2},-(1+\sqrt{5})/4,-\sqrt{3}/2)$.

Thus, we seek a solution to the matrix equation

$$M_{jl}\kappa_l = 1, \tag{62}$$

in order to calculate the bound state energy

$$-E_0 = (\boldsymbol{\kappa}_l\boldsymbol{\alpha}_l)^2 = \kappa_l M_{lj}\kappa_l = \sum_{l=1}^{d}\kappa_l. \tag{63}$$

As the Coxeter matrices are very simple in structure, the easiest procedure for solving this equation is probably to successively solve for the $\kappa_1,\kappa_2,\cdots$, and then sum the results for the bound state energy. We illustrate this for the most important case of A_n.

(Q) For the case of A_n, which has the Coxeter graph ●—●—● ··· —●—●, the Coxeter matrix is

$$M = \begin{vmatrix} 1 & -\frac{1}{2} & 0 & & & & \\ -\frac{1}{2} & 1 & -\frac{1}{2} & & & & \\ 0 & -\frac{1}{2} & 1 & & & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & & \ddots & \\ & & & & & 1 & -\frac{1}{2} \\ & & & & & -\frac{1}{2} & 1 \end{vmatrix}. \tag{64}$$

The matrix equation is then equivalent to

$$\begin{aligned} &\kappa_2-\tfrac{1}{2}\kappa_2 = 1 \quad \text{or} \quad \kappa_2 = 2(\kappa_1-1),\\ &\kappa_2-\tfrac{1}{2}(\kappa_1+\kappa_3) = 1 \quad \text{or} \quad \kappa_3 = 3(\kappa_1-2),\\ &\kappa_3-\tfrac{1}{2}(\kappa_2+\kappa_4) = 1 \quad \text{or} \quad \kappa_4 = 4(\kappa_1-3),\\ &\vdots\\ &\kappa_{j-1}-\tfrac{1}{2}(\kappa_j+\kappa_{j-2}) = 1 \quad \text{or} \quad \kappa_j = j(\kappa_1-j+1),\\ &\vdots\\ &\kappa_{n-1}-\tfrac{1}{2}(\kappa_n+\kappa_{n-2}) = 1 \quad \text{or} \quad \kappa_n = n(\kappa_1-n+1). \end{aligned} \tag{65}$$

We then have one equation remaining:

$$\kappa_n - \tfrac{1}{2}\kappa_{n-1} = 1 \quad \text{or} \quad 0 = (n+1)(\kappa_1 - n). \tag{66}$$

This equation then determines $\kappa_1 = n$, and

$$-E_0 = \sum_{j=1}^{n} j(n-j+1) = \frac{n(n+1)(n+2)}{3!} = \binom{n+2}{3}. \tag{67}$$

Further results are only slightly more complicated, and the method is essentially the same, so we do not reproduce the calculations. The results are tabulated in Table I and graphed in Fig. 4.

There are states bound in some directions, but not all; these cases reduce to Coxeter groups of lower order. In the corresponding scattering problems involving the bound states one upon another, there is no production nor disassociation.

ACKNOWLEDGMENTS

The author would like to thank Eugene Gutkin for very useful conversations and much help in the proof of our theorem.

[1]H. Bethe, Z. Phys. **71**, 205 (1931).
[2]J.B. McGuire, J. Math. Phys. **5**, 622 (1964).
[3]E.H. Lieb, Phys. Rev. **130**, 1616 (1963).
[4]C.N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[5]B. Sutherland, Phys. Rev. Lett. **20**, 98 (1968).
[6]R.J. Baxter, Ann. Phys. (N.Y.) **70**, 193 (1972).
[7]B. Sutherland, Rocky Mount. J. Math. **8**, 413 (1978).
[8]H.S.M. Coxeter, *Regular Polytopes* (Dover, New York, 1973).

The anholonomic Cauchy problem in general relativity

John Stachel[a)]
Einstein Project, Institute for Advanced Study, Princeton, New Jersey 08540
and Deparment of Physics, Princeton University, New Jersey 08540

(Received 18 July 1979; accepted for publication 26 September 1979)

The Lie derivative approach to the Cauchy problem in general relativity is applied to the evolution along an arbitrary timelike vector field for the case where the dynamical degrees of freedom are chosen as the (generally anholonomic) metric of the hypersurface elements orthogonal to the vector field. Generalizations of the shear, rotation, and acceleration are given for a nonunit timelike vector field, and applied to the three-plus-one breakup of the Riemann tensor into components parallel and orthogonal to the vector field, resulting in the anholonomic Gauss–Codazzi equations. A similar breakup of the Einstein field equations results in the form of the constraint and evolution equations for the anholonomic case. The results are applied to the case of a space–time with a timelike Killing vector field (stationary field) to demonstrate their utility. Other possible applications, such as in the numerical integration of the field equations, are mentioned. Definitions are given of three-index shear, rotation, and acceleration tensors, and their use in a two-plus-two decomposition of the Riemann tensor and field equations is indicated.

I. INTRODUCTION

It has been known for some time that the Cauchy problem for various field theories could be formulated covariantly in terms of the evolution of some initial data, given on an initial spacelike hypersurface of space–time, along an arbitrary contravariant vector field. The problem thus falls into two parts:

(1) determination of the necessary and sufficient initial data to fix a unique solution to the field equations;

(2) calculation of that solution in terms of the evolution of the initial data along the vector field.

While the problem for field theories in a given "background" space–time could always be formulated in terms of covariant derivatives of the field quantities along the vector field, in general-relativistic field theories this was impossible since the metric affine connection was also to be determined by the field equations. In Refs. 1 and 2, it was shown how to formulate the Cauchy problem for general-relativistic field theories in terms of Lie derivatives along the vector field.[1,2] The choice of an initial hypersurface and vector field results, by dragging of the initial surface along the vector field, in a foliation of the space–time together with the fibration given by the vector field; this yields a correspondence between points on different slices of the foliation but on the same fiber (i.e., trajectory of the vector field). In Ref. 1, the analysis was first carried out for the case when the foliation resulted in a geodesically parallel family of spacelike hypersurfaces and the vector field was the timelike geodesic normal field. This was then generalized to the case of an arbitrary foliation and vector field. But then in both cases the field variables in terms of which the initial data was given and the evolution analysis made were the induced metric on the hypersurface (first fundamental form or intrinsic metric of the hypersurface), and the extrinsic metric or second fundamental form of the hypersurface (which is—up to a numerical factor—the Lie derivative in the unit normal direction of the first fundamental form).

More recently O'Murchadha,[3] basing himself upon earlier unpublished work of York, and using the Lie derivative technique; and Kulhanek,[4] using a more coordinate-dependent approach, have pointed out that, instead of the induced metric on the family of hypersurfaces one could use the projection of the metric tensor onto the hypersurface-element orthogonal to the given field, even though these projections do not in general fit together holonomically to form the induced metric on any hypersurface. Of course, if the vector field is hypersurface orthogonal, and one chooses to use this family of orthogonal hypersurfaces for the foliation, the two approaches coincide. But in the general case, where the vector field is not hypersurface orthogonal—or, even if it is, we choose another foliation—the two approaches are distinct. We shall refer to them as the *holonomic* and *anholonomic* cases, for short.

In coordinate language, if we adapt a coordinate system in which the vector field takes the form δ^μ_o and the hypersurfaces of the foliation are given by $x^o = \text{const.}$ (always possible locally) this is the difference between using $g_{ab}(a,b = 1,2,3)$ on the hypersurface (induced metric tensor of the hypersurface), and using ${}^3g_{ab} = g_{ab} - g_{oa}g_{ob}/g_{oo}$ ($d\sigma^2 = {}^3g_{ab}\,dx^a\,dx^b$ being the infinitesimal interval on the hypersurface element orthogonal to the vector field). Møller,[5] among others, has given an extensive discussion of the possible physical interpretation of the latter choice in terms of measurement with rods and clocks.

O'Murchadha,[3] in particular, carried out an extensive analysis of the Cauchy problem in general relativity for a unit vector field, discussing (following Schouten[6]) the induced connection and covariant differentiation on the orthogonal hypersurface element, the Riemann tensor and the

[a)]On leave from Department of Physics, Boston University

 0022-2488/80/071776-07$1.00

Gauss–Codazzi equations; as well as the breakup of the Einstein equation into constraint and evolution equations.

However, he (as well as Kulhanek[4]) confined himself to unit vector fields. While in a sense this constitutes no real restriction, since the evolution of any gravitational field satisfying the Einstein equations may be analyzed along such a vector field; in another sense, it does constitute a restriction, analogous to that resulting in the holonomic case if we were to confine ourselves to the case of a unit vector field.

In the latter case, as Smarr and York[7] have recently reemphasized, this is often not only a theoretical inconvenience, but a practical hazard in numerical computation of the evolution of the gravitational field, since the focusing effect of gravitation on the timelike trajectories of the congruence must be compensated for by moving different distances along the trajectories from hypersurface to hypersurface of the foliation to avoid coordinate singularities. In this paper, we shall remove the restriction to unit vector fields. Of course, there is an additional restriction implied by the choice of *orthogonal* hypersurface elements in our analysis. We could drop this restriction as well; but (we are not demanding that the hypersurface elements fit together holonomically into a hypersurface, anyway) the simplification resulting from the use of orthogonal hypersurface elements outweighs the generality that could be introduced by dropping this restriction. In the language of the holonomic case, we can say that we are going to allow arbitrary lapse functions, but no shift functions.[8]

Our interest in this generalization is twofold. First of all, as indicated above, this generalization provides an interesting and potentially useful alternate approach to the Cauchy problem in general relativity (or any other field theory, for that matter). Secondly, it provides a useful guide to generalization of the work on the Cauchy problem in general relativity, based on a three-plus-one breakup of tensors into components parallel and orthogonal to a timelike vector field, to a two-plus-two breakup. This type of breakup has recently proved its value in the analysis of a number of other types of initial value problem in general relativity.[9]

In the next section we shall discuss a generalization of the standard breakup[10] of the covariant derivative of a unit timelike vector field to an arbitrary timelike vector field. In Sec. 3, we shall discuss the three-plus-one parallel-orthogonal breakup of the metric, affine connection and Riemann tensor with respect to a timelike vector field. In Sec. 4, we shall discuss the breakup of the Einstein equations and the Cauchy problem for the empty-space gravitational field. Section 5 will consider a simple application of the methods developed: The analysis of a space–time with a timelike Killing vector. We then conclude with a brief discussion of other possible applications, and of the generalization to the two-plus-two breakup.

II. DECOMPOSITION OF THE COVARIANT DERIVATIVE OF A VECTOR FIELD

Our aim is to parallel as closely as possible the standard decomposition[10] of the covariant derivative of a unit timelike ("velocity") congruence into rotation, shear (including expansion) and acceleration terms. Let v^μ, represent an arbitrary contravariant vector field. Instead of its covariant components v_μ, we shall use the parallel covariant vector V_μ, normalized so that

$$v^\mu V_\mu = 1. \tag{2.1}$$

Thus, if $v^\mu = \rho n^\mu$, where n^μ is the unit vector along the congruence, $V_\mu = (1/\rho)n_\mu$, so that $v^\mu = \rho^2 V^\mu$. The advantage of this normalization is apparent when we project along and normal to the vector field. This is done with the aid of the breakup of the identity matrix:

$$\delta^\mu_\nu = B^\mu_\nu + C^\mu_\nu, \tag{2.2}$$

where $C^\mu_\nu = v^\mu V_\nu$. Evidently, C^μ_ν projects onto the direction of v^μ or V_ν for contra- or covariant indices, respectively; while B^μ_ν projects out any components along the vector field, and thus, is the orthogonal projection operator.

Before we begin our decomposition of $\nabla_\mu V_\nu$, it is useful to compute the Lie derivative of V_μ with respect to v^κ:

$$L_v V_\mu = v^\kappa \partial_\kappa V_\mu + V_\kappa \partial_\mu v^\kappa = v^\kappa(\partial_\kappa V_\mu - \partial_\mu V_\kappa), \tag{2.3}$$

using (2.1), where $L_v \Phi$ denotes the Lie derivative of the geometrical object Φ with respect to v^μ. If V_μ were the unit vector field n_μ (i.e., $\rho = 1$) this would be the acceleration a_μ of the unit congruence. This is easily seen by replacing ordinary by covariant derivatives in (2.3), and noting that $n_\kappa \nabla_\mu n^\kappa = 0$. So $L_v V_\mu$ is a generalized acceleration of the congruence, and has the same property of orthogonality

$$v^\mu L_v V_\mu = 0. \tag{2.4}$$

as the acceleration vector a_μ.

We could also introduce another generalization of the ordinary acceleration, which we denote by A_μ:

$$A_\mu = v^\kappa \nabla_\kappa V_\mu. \tag{2.5}$$

This is related to $L_v V_\mu$ by the following equation:

$$L_v V_\mu = A_\mu + (1/\rho)\partial_\mu \rho. \tag{2.6}$$

This shows that A_μ is not orthogonal to v^μ; indeed,

$$A_\mu v^\mu = -(1/\rho) L_v \rho; \tag{2.7}$$

in this sense, $L_v \mathbf{V}$ has a better claim to be called the generalized acceleration of the congruence, and we shall so denote it.

We are now ready to proceed with our decomposition of $\nabla_\mu V_\nu$. First we break it up into symmetric and antisymmetric parts, and then project each of them into the parallel and orthogonal directions as often as we get a nonvanishing result. In this way, we are led to define

$$\Omega_{\mu\nu} = \tfrac{1}{2} B^{\alpha\beta}_{\mu\nu}(\partial_\alpha V_\beta - \partial_\beta V_\alpha) \tag{2.8}$$

(here $B^{\alpha\beta}_{\mu\nu}$ is just shorthand for $B^\alpha_\mu B^\beta_\nu$; similar shorthand will be used, following Schouten,[6] throughout for projection operators). $\Omega_{\mu\nu}$ is the generalized rotation of the vector field; it is easily shown that

$$\Omega_{\alpha\beta} = (1/\rho)\omega_{\alpha\beta}, \tag{2.9}$$

when $\omega_{\alpha\beta}$ is the rotation of the unit congruence. Thus, the vanishing of $\Omega_{\alpha\beta}$ is the necessary and sufficient condition for the congruence to be hypersurface orthogonal. Similarly, we define

$$H_{\mu\nu} = \tfrac{1}{2} B^{\alpha\beta}_{\mu\nu}(\nabla_\alpha V_\beta + \nabla_\beta V_\mu). \tag{2.10}$$

This is the generalized shear tensor of the vector field; again

$$H_{\mu\nu} = (1/\rho)h_{\mu\nu}, \tag{2.11}$$

where $h_{\mu\nu}$ is the shear of the unit conguence (including the expansion, its trace in this term. More properly, it should be call the generalized rate of deformation tensor). It is now easy to show that

$$\nabla_\mu V_\nu = \Omega_{\mu\nu} + H_{\mu\nu} + (L_v V_\nu)V_\mu - (1/\rho)L_v\rho V_\mu V_\nu - V_\mu \frac{'\partial_\nu \rho}{\rho} - V_\nu \frac{'\partial_\mu \rho}{\rho}, \tag{2.12}$$

where $'\partial_\mu = B^\alpha_\mu \partial_\alpha$.

Note that this decomposition is much like the usual one for a unit vector field, to which it reduces, of course, if $\rho = 1$. Then the last three terms vanish and the other terms reduce to the corresponding expressions for the unit congruence, as is evident from (2.6), (2.9), and (2.11). In Sec. 5, we shall further illustrate the usefulness of this decomposition in analyzing the case of a timelike Killing vector field.

For our later work, it is helpful to discuss what properties of the congruence could be given on an arbitrary cross section of the congruence (i.e., a hypersurface transvecting the streamlines of the congruence), assuming that we were in a bare manifold without metric or connection: Suppose we are given the vector field v^ν in such a bare manifold, together with a cross section. What can we specify about the covariant vector field V_μ on that cross section without implying any restriction on a possible metric? Clearly, we can pick V_μ itself on the cross section, subject only to the condition (2.1) which involves no metrical relations. We can also clearly give $L_v V_\mu$ on the cross section, subject to (2.4); and its Lie derivatives to all orders, since the definition of the Lie derivative involves no metrical relations. Thus, the generalized acceleration can be arbitrarily specified along the entire congruence (since this is equivalent to giving the Lie derivative to all orders on a cross section).[1,2,6] Similarly, ρ and its Lie derivatives with respect to v^μ to all orders can be given arbitrarily on the cross section: Although their ultimate interpretation may be metrical, they merely serve here to define the n^μ field; to say that n^μ is a unit field has no meaning until a metric is introduced. Alternatively, specification of ρ along a trajectory of the congruence can be considered to give an intrinsic metric along each fiber; but all metrizations of a one-dimensional manifold are trivial, so this constitutes no restriction on the possible metrical properties of the manifold in which the congruence is embedded. Given V_μ and $L_v V_\mu$ on the cross section, then $\partial_\mu V_\nu$ is determined on the cross section; Eq. (2.8) then shows that $\Omega_{\mu\nu}$ is thereby fixed. Indeed, taking the Lie derivative of (2.8), we find

$$L_v \Omega_{\mu\nu} = B^\alpha_\mu \partial_\alpha L_v V_\nu - B^\alpha_\nu \partial_\alpha L_v V_\mu. \tag{2.13}$$

Thus, additionally giving the Lie derivative of the generalized acceleration on a cross section determines the Lie derivative of the generalized rotation, and so on to higher orders of Lie derivatives.

Thus, if we give V_μ and all its Lie derivative on a cross section, and ρ and all its Lie derivatives, we can determine the generalized rotation $\Omega_{\mu\nu}$ [and from (2.9) even $\omega_{\mu\nu}$] everywhere along the congruence. Even though its definition involves the covariant derivative, (2.6) shows that this fixes A_μ everywhere along the congruence: The projection of the covariant derivative along v^μ is fixed by nonmetrical quantities in our sense. $H_{\mu\nu}$, however, cannot be similarly reduced; and in that sense constitutes the only genuinely metrical term in the decomposition (2.12) of the covariant derivative of V_μ. We shall not be surprised, therefore, to see the fundamental role that it plays in the Cauchy problem for the Einstein field equations. Indeed, our results demonstrates that, without even looking at any field equations, we can decide which parts of the covariant derivative of the vector field could *possibly* carry dynamical information about the metrical field. Our result is then seen to be quite independent of the exact form of the field equations. For example, it would hold for field equations derived from any Langrangian formed from invariants of the Riemann tensor. Another, and perhaps simpler way to convince oneself of this result is to consider the contravariant vector field v^μ in a bare manifold as one vector of a tetrad, whose other three vectors we shall label B^μ_a ($a = 1,2,3$ indexes the three vectors). Any such tetrad field will have a dual tetrad field, which we shall denote by (V_μ, B^a_μ). Then $v^\mu V_\mu = 1$ will be automatically satisfied, as one of the defining relations of the dual tetrad. We can now compute the curl of the V_μ field, $\partial_\nu V_\mu - \partial_\mu V_\nu$, and project it onto the contravariant tetrad vectors. In this way we get $B^\nu_a v^\mu(\partial_\mu V_\nu - \partial_\nu V_\mu)$ and $B^{\mu\nu}_{ab}(\partial_\mu V_\nu - \partial_\nu V_\mu)$. All of these are perfectly nonmetrical operations—indeed, the projections listed are just some components of the object of anholonomicity[6] for the tetrad fields. But once we introduce any metric in such a way to make the B^μ_a orthogonal to v^μ, they become the generalized acceleration and rotation of the V_μ field, as reference to Eqs. (2.3) and (2.8) shows at once. But in this way we have exhausted all the nonmetrical information about V_μ. As for v^μ, the comparison of its length with that of a parallel vector field n^μ is given by ρ. Once we metrize in any way that makes n^μ a unit vector field, the rest of our results follow.

III. DECOMPOSITION OF METRIC, AFFINE CONNECTION, RIEMANN TENSOR

Next we shall consider the three-plus-one decomposition of the metric tensor. It only has double projections onto and double projections orthogonal to v^μ, since we originally chose to project orthogonally to the vector field. Thus

$$g_{\mu\nu} = {}'g_{\mu\nu} + \rho^2 V_\mu V_\nu, \tag{3.1}$$

where $'g_{\mu\nu}$, given by

$${}'g_{\mu\nu} = B^{\alpha\beta}_{\mu\nu} g_{\alpha\beta} \tag{3.2}$$

is not only the double orthogonal projection of the metric, but also the induced metric, or first fundamental form, on the orthogonal surface element. This surface element, rigged with the vector field v^μ, also has an affine connection induced on it, which is the "natural" projection of the metric affine connection of the 4-space onto this rigged hypersurface[6]: The anholonomic components of this connection are numerically equal to the corresponding anholonomic components of the four-dimensional connection in an anholonomic coordinate system adapted to the rigging field v^κ. This means

$'\nabla_\mu w^\nu = B^{\alpha\nu}_{\mu\beta}\nabla_\alpha w^\beta$, where $'\nabla$ stands for the surface element covariant derivative, and w^β is a vector in the surface element ($w^\nu V_\nu = 0$).

It is also necessary to define the extrinsic curvature of the surface element. The usual extrinsic curvature, or second fundamental form of a surface element is defined by projecting the symmetrized covariant derivative of the unit normal vector twice onto the surface element. But in our case, since we are not using the unit normal, we can form two such tensors,[6] depending on whether we use the covariant derivative of v^κ or V_κ. Of course, the two will be related to each other by the factor ρ^2, so it is merely a question of picking one of them as our definition of the second fundamental form. Since we have decomposed the covariant V_μ in the last section, we choose to define the second fundamental form by

$$H_{\mu\nu} = \tfrac{1}{2}B^{\alpha\beta}_{\mu\nu}(\nabla_\alpha V_\beta + \nabla_\beta V_\alpha), \tag{3.3}$$

so that it coincides with our previous definition of the generalized shear tensor (note that we here reverse the sign of Schouten's definition,[6] which we used in our previous papers[1-2]).

This definition of the second fundamental form has one inconvenience, which will not bother us much in this work, but is important for the generalization to the two-plus-two decomposition. Namely, while $'g_{\mu\nu}$ is invariant under a change of length of v^μ (so long as we preserve the normalization $v^\mu V_\mu = 1$, the projection operators are invariant), $H_{\mu\nu}$ obviously is not [see Eq. (2.11)]. This can be remedied, following Schouten,[6] by introducing a three-index tensor $H_{\mu\nu}{}^\kappa$, defined in our case by

$$H_{\mu\nu}{}^\kappa = H_{\mu\nu}v^\kappa = h_{\mu\nu}n^\kappa,$$

which is clearly invariant under such a change. A similar three-index rotation tensor $\Omega_{\mu\nu}{}^\kappa = \Omega_{\mu\nu}v^\kappa$ could also be introduced. Schouten calls $H_{\mu\nu}{}^\kappa + \Omega_{\mu\nu}{}^\kappa$ the curvature tensor of valence three. We shall call them the three-index shear and three-index rotation and we shall use them occasionally in this paper, even though they could be eliminated, because this will facilitate comparison of our work here with the two-plus-two decomposition.

Schouten shows how to define the Riemann tensor of the surface element, in such a way that it depends only on the rigging of the surface element and the displacement within the surface element of vectors lying in that element, and not on any other features of the connection in the enveloping manifold. Following O'Murchadha,[3] the expression for the three-Riemann tensor is most easily found by noting that the commutator of the second order three-covariant derivatives of a 3-vector w_λ (such that $w_\lambda v^\lambda = 0$) is not linear in w_μ unless a term is subtracted from it. This leads to the following definition of the Riemann tensor $'R_{\mu\nu\kappa}{}^\lambda$ of the orthogonal surface element [10a]:

$$-{}'R_{\mu\nu\kappa}{}^\lambda w_\lambda = 2'\nabla_{[\mu}{}'\nabla_{\nu]}w_\kappa + 2\Omega_{\mu\nu}L_v w_\kappa. \tag{3.4}$$

Direct computation of the commutator then gives Gauss' equation for the components of the Riemann tensor projected four times into the surface element:

$$\begin{aligned}{}'R_{\mu\nu\kappa}{}^\lambda &= B^{\alpha\beta\sigma\lambda}_{\mu\nu\kappa\tau}R_{\alpha\beta\sigma}{}^\tau - \rho^2(\Omega_{\mu\kappa} + H_{\mu\kappa})(\Omega_\nu{}^\lambda + H_\nu{}^\lambda)\\ &\quad + \rho^2(\Omega_{\nu\kappa} + H_{\nu\kappa})(\Omega_\mu{}^\lambda + H_\mu{}^\lambda)\\ &\quad - 2\rho^2\Omega_{\mu\nu}(\Omega_\kappa{}^\lambda + H_\kappa{}^\lambda).\end{aligned} \tag{3.5}$$

Note that (3.4) and (3.5) reduce to the more familiar holonomic forms when $\Omega_{\mu\nu} = 0$. We have here used the nonvariant two-index rotation and shear in Gauss' equation; the invariant three-index rotation and shear could be used just as easily:

$$\begin{aligned}{}'R_{\mu\nu\kappa}{}^\lambda &= B^{\alpha\beta\sigma\lambda}_{\mu\nu\kappa\tau}R_{\alpha\beta\sigma}{}^\tau - (\Omega_{\mu\kappa}{}^\alpha + H_{\mu\kappa}{}^\alpha)(\Omega_{\nu\cdot\alpha}^{\ \ \lambda} + H_{\nu\cdot\alpha}^{\ \ \lambda})\\ &\quad + (\Omega_{\nu\kappa}{}^\alpha + H_{\nu\kappa}{}^\alpha)(\Omega_{\mu\cdot\alpha}^{\ \ \lambda} + H_{\mu\cdot\alpha}^{\ \ \lambda})\\ &\quad - 2\Omega_{\mu\nu}{}^\alpha(\Omega_{\kappa\cdot\alpha}^{\ \ \lambda} + H_{\kappa\cdot\alpha}^{\ \ \lambda}).\end{aligned} \tag{3.5a}$$

In this form it is evident that all terms in the right-hand side are invariant under a change in the length of v^μ (i.e., a change of ρ), so that the left-hand side must be too.

The components of the Riemann tensor projected three times onto the surface element and once onto the vector field are given by the Codazzi equations. These are found most easily by computing the commutator of the second 3-surface covariant derivatives of the V_μ, vector, giving

$$\begin{aligned}&{}'\nabla_\mu(\Omega_{\nu\kappa} + H_{\nu\kappa}) - {}'\nabla_\nu(\Omega_{\mu\kappa} + H_{\mu\kappa})\\ &\quad = -B^{\alpha\beta\sigma}_{\mu\nu\kappa}R_{\alpha\beta\sigma}{}^\lambda V_\lambda - 2\Omega_{\mu\nu}B^\sigma_\kappa[L_v V_\sigma - (1/\rho)\partial_\sigma\rho]\\ &\qquad + (\Omega_{\mu\kappa} + H_{\mu\kappa})(1/\rho)B^\sigma_\nu\partial_\sigma\rho\\ &\qquad - (\Omega_{\nu\kappa} + H_{\nu\kappa})(1/\rho)B^\sigma_\mu\partial_\sigma\rho.\end{aligned} \tag{3.6}$$

Again, this could be expressed interms of the three-index rotation and shear quite simply.

The only remaining nonvanishing projected components of the Riemann tensor are the components projected twice onto the surface element and twice along the vector field. But these are expressible in terms of the orthogonal components of the Ricci tensor:

$$\begin{aligned}B^{\mu\lambda}_{\alpha\beta}R_{\mu\lambda} &= B^{\mu\lambda}_{\alpha\beta}g^{\rho\sigma}R_{\rho\mu\lambda\sigma}\\ &= B^{\mu\lambda\tau\sigma}_{\alpha\beta\kappa\nu}R_{\tau\mu\lambda\sigma}{}'g^{\kappa\nu} + (1/\rho^2)B^{\mu\lambda}_{\alpha\beta}v^\tau v^\sigma R_{\tau\mu\lambda\sigma}.\end{aligned} \tag{3.7}$$

As we shall see, in the next section, it is just these components of the Riemann tensor which occur when we try to calculate the evolution of the $H_{\mu\nu}$—the dynamical terms in the decomposition of $\nabla_\mu V_\nu$.

We shall now discuss an alternate way of arriving at some of our previous results, which proves to be of great importance because it can be generalized, leading to definitions of the shear, rotation, and acceleration of "congruences" of two and higher-dimensional subspaces. Equation (3.1) is essentially a breakup of the metric tensor into the metric of a three-dimensional subspace $'g_{\mu\nu}$, and the metric of an orthogonal one-dimensional subspace $''g_{\mu\nu} = \rho^2 V_\mu V_\nu$:

$$g_{\mu\nu} = {}'g_{\mu\nu} + {}''g_{\mu\nu}, \quad {}'g_{\mu\nu}{}''g^{\nu\kappa} = 0. \tag{3.8}$$

We shall show that it is possible to carry out the analysis entirely in terms of these two orthogonal subspaces without any need to further break up one of them, as we previously broke up $''g_{\mu\nu}$ into $\rho^2 V_\mu V_\nu$.

To do this we need a decomposition of the covariant derivative of $''g_{\mu\nu}$, (which, of course, is equal to minus the covariant derivative of $'g_{\mu\nu}$) rather than of $\nabla_\mu V_\nu$. It turns out that it is simplest to use that particular linear combination of covariant derivatives which has the formal structure of the Christoffel symbol of the first kind. To denote this, we introduce[11] the symbol

$$A_{\{\lambda\kappa\mu\}} = \tfrac{1}{2}(-A_{\lambda\kappa\mu} + A_{\mu\lambda\kappa} + A_{\kappa\mu\lambda}). \tag{3.9}$$

With its aid, we can write the combination we want as

$$\begin{aligned} {}''g_{\{\mu\nu;\kappa\}} &= \tfrac{1}{2}(-{}''g_{\mu\nu;\kappa} + {}''g_{\kappa\mu;\nu} + {}''g_{\nu\kappa;\mu}) \\ &= \dot{n}_\kappa n_\mu n_\nu + n_\nu\omega_{\mu\kappa} + n_\mu\omega_{\nu\kappa} + n_\kappa h_{\mu\nu}, \end{aligned} \tag{3.10}$$

as is easily checked. Now the important thing is that we can extract the three-index generalized shear and rotation from ${}''g_{\{\kappa\mu;\nu\}}$ by operating on it with ${}'g^{\mu\nu}$ and ${}''g^{\mu\nu}$ alone. Namely

$$\begin{aligned} &{}'g^{\kappa}{}_{\alpha}\,{}'g^{\mu}{}_{\beta}\,{}''g^{\nu\lambda}\,{}''g_{\{\kappa\mu;\nu\}} = H_{\alpha\beta}{}^{\lambda}, \\ &{}''g^{\kappa\lambda}\,{}'g^{\mu}{}_{\alpha}\,{}'g^{\nu}_{\beta}\,{}''g_{\{\kappa\mu;\nu\}} = \Omega_{\alpha\beta}{}^{\lambda}, \end{aligned} \tag{3.11}$$

where $H_{\mu\nu}{}^{\lambda}$ and $\Omega_{\mu\nu}{}^{\lambda}$ have just the meanings defined earlier in this section.

We can also define an acceleration 3-tensor:

$$A_{\alpha\beta}{}^{\lambda} = {}''g^{\kappa}{}_{\alpha}\,{}''g^{\mu}{}_{\beta}\,{}'g^{\nu\lambda}\,{}''g_{\{\kappa\mu;\nu\}}. \tag{3.12}$$

The most noteworthy feature of the definitions of H, Ω, and A is that they can be applied to a subspace of *any* number of dimensions. Elsewhere[12] we have shown that the vanishing of the trace of the acceleration tensor ${}''g^{\alpha\beta}A_{\alpha\beta}{}^{\lambda}$ is equivalent to the condition that the 2-surfaces formed by holonomic congruences of two-dimensional subspaces obey the string equations of motion. In space–time of course, only one-, two-, and three-dimensional subspaces are possible; but the mathematical generalization to m-dimensional subspaces of an n-dimensional Riemannian space is quite possible, and will be discussed in a joint paper with J. Plebanski now in preparation. Applications to the two-plus-two breakup of the Einstein equations will also be discussed elsewhere.

IV. DECOMPOSITION OF THE FIELD EQUATIONS AND THE CAUCHY PROBLEM

As mentioned in the Introduction, the field equations $G_{\mu\nu} = 0$ break up into two sets: The four projections onto the vector field, which as we shall see can be expressed in terms of initial data and are therefore called the constraint equations; and the six projections onto the orthogonal 3-surface elements, which as we shall see contain the information about evolution of the dynamical degrees of freedom of the metrical field.

In order to compute $G_{\mu\lambda}v^{\nu}v^{\lambda}$ most simply, it is useful to have the following expression for R, which is also needed for a Lagrangian or Hamiltonian formulation of the field equations in three-plus-one breakup.[1,2]

$$\begin{aligned} R &= g^{\mu\lambda}R_{\mu\lambda} = g^{\mu\lambda}g^{\kappa\nu}R_{\kappa\mu\lambda\nu} \\ &= [{}'g^{\mu\lambda} + (1/\rho^2)v^{\mu}v^{\lambda}]\,[{}'g^{\kappa\nu} + (1/\rho^2)v^{\kappa}v^{\nu}]R_{\kappa\mu\lambda\nu} \\ &= {}'g^{\alpha\beta}\,{}'g^{\tau\sigma}B^{\mu\lambda\kappa\nu}_{\alpha\beta\tau\sigma}R_{\kappa\mu\lambda\nu} + (2/\rho^2)R_{\kappa\nu}v^{\kappa}v^{\nu}. \end{aligned} \tag{4.1}$$

With this result it is easy to show that

$$G_{\mu\lambda}v^{\mu}v^{\lambda} = -\tfrac{1}{2}\rho^2 B^{\mu\lambda\kappa\nu}_{\alpha\beta\tau\sigma}R_{\kappa\mu\lambda\nu}\,{}'g^{\alpha\beta}\,{}'g^{\tau\sigma}. \tag{4.2}$$

A similar elementary calculation gives

$$G_{\mu\lambda}B^{\mu}_{\alpha}v^{\lambda} = B^{\mu\kappa\nu}_{\alpha\tau\sigma}R_{\kappa\mu\lambda\nu}v^{\lambda}\,{}'g^{\tau\sigma}. \tag{4.3}$$

Reference to the Gauss and Codazzi equations (3.5) and (3.6) shows that they enable the right-hand sides of (4.2) and (4.3), respectively, to be expressed entirely in terms of the 3-surface Riemann tensor and the elements of the covariant derivative of the vector field V_{μ},

$$\begin{aligned} G_{\mu\lambda}v^{\mu}v^{\lambda} &= \tfrac{1}{2}\rho^2[-{}'R + \rho^2(H^2 - H_{\alpha\beta}H^{\alpha\beta}) \\ &\quad + 3\rho^2\Omega_{\alpha\beta}\Omega^{\alpha\beta}], \end{aligned} \tag{4.4}$$

$$\begin{aligned} G_{\mu\lambda}B^{\mu}_{\alpha}v_{\lambda} &= \rho^2[{}'\nabla_{\beta}(H_{\alpha}{}^{\beta} + \Omega_{\alpha}{}^{\beta}) - {}'\partial_{\alpha}H - H\,{}'\partial_{\alpha}(\ln\rho) \\ &\quad + (3\Omega_{\alpha}{}^{\beta} + H_{\alpha}{}^{\beta})\,{}'\partial_{\beta}(\ln\rho) - 2\Omega_{\alpha}{}^{\beta}L_v V_{\beta}]. \end{aligned} \tag{4.5}$$

We have seen in the last section, Eq. (3.7), that the remaining projections of the Einstein tensor—or rather the equivalent projections of the Ricci tensor—are equivalent, modulo terms given by the Gauss equation, to the components of the Riemann tensor projected twice onto the vector field and twice onto the orthogonal three-surface element. As we shall now show, these are just the components that occur when we compute the evolution of the metric along the vector field v^{μ}.

To do this, we need to compute the Lie derivatives of $g_{\mu\nu}$ with respect to v^{μ}. The breakup (3.1) of the metric tensor shows that we shall need the Lie derivatives of ρ and V_{μ} to compute this, as well as the Lie derivatives of ${}'g_{\mu\nu}$. But, as we discussed in Sec. 2, the former are arbitrarily specifiable, since they involve only nonmetrical information. Thus, only the evolution of ${}'g_{\mu\nu}$ remains to be determined.

Now

$$L_v\,{}'g_{\mu\nu} = B^{\alpha\beta}_{\mu\nu}L_v g_{\alpha\beta}, \tag{4.6}$$

since all other terms in the difference between the Lie derivatives of $g_{\mu\nu}$ and ${}'g_{\mu\nu}$ vanish. Thus,

$$\begin{aligned} L_v\,{}'g_{\mu\nu} &= B^{\alpha\beta}_{\mu\nu}(\nabla_{\alpha}v_{\beta} + \nabla_{\beta}v_{\alpha}) \\ &= 2\rho^2 H_{\mu\nu}. \end{aligned} \tag{4.7}$$

Thus, we see that $L_v\,{}'g_{\mu\nu}$ is essentially the generalized shear of the vector field, up to the factor $2\rho^2$. By using the three-index shear, we can transform this equation into a form invariant under changes of length of the vector field:

$$V^{\kappa}L_v\,{}'g_{\mu\nu} = 2H_{\mu\nu}{}^{\kappa}. \tag{4.8}$$

Now we take the crucial step, and compute the second Lie derivative of the 3-metric. The Lie derivation of Eq. (4.6) shows that

$$L^{2}_{v}\,{}'g_{\alpha\beta} = B^{\mu\nu}_{\alpha\beta}L^{2}_{v}g_{\mu\nu} + L_v B^{\mu\nu}_{\alpha\beta}L_v g_{\mu\nu}. \tag{4.9}$$

A brief calculation shows that second term on the right of (4.9) reduces to

$$L_v B^{\mu\nu}_{\alpha\beta}L_v g_{\mu\nu} = -2(L_v V_{\alpha})(L_v V_{\beta}). \tag{4.10}$$

Thus, only the first term on the right of (4.9) remains to be determined. By writing it as

$$\begin{aligned} B^{\mu\nu}_{\alpha\beta}L^{2}_{v}g_{\mu\nu} &= B^{\mu\nu}_{\alpha\beta}(v^{\kappa}\nabla_{\kappa}L_v g_{\mu\nu} + L_v g_{\kappa\nu}\nabla_{\mu}v^{\kappa} \\ &\quad + L_v g_{\mu\kappa}\nabla_{\nu}v^{\kappa}), \end{aligned} \tag{4.11}$$

and using the commutation relations between covariant and Lie differentiation (see Schouten,[6] p. 152), the right-hand side of (4.11) is easily reduced to

$$\begin{aligned} B^{\mu\nu}_{\alpha\beta}L^{2}_{v}g_{\mu\nu} &= 2B^{\mu\nu}_{\alpha\beta}v^{\kappa}v^{\lambda}R_{\lambda\mu\kappa\nu} + 2B^{\mu\nu}_{\alpha\beta}(\nabla_{\nu}v^{\kappa}) \\ &\quad \times(\nabla_{\mu}v_{\kappa}) + B^{\mu\nu}_{\alpha\beta}[\nabla_{\mu}(v^{\kappa}\nabla_{\kappa}v_{\nu}) \\ &\quad + \nabla_{\nu}(v^{\kappa}\nabla_{\kappa}v_{\mu})]. \end{aligned} \tag{4.12}$$

The last two terms on the right of (4.12) may now be readily evaluated using the decomposition of $\nabla_{\mu}V_{\nu}$ given in Eq. (2.12). Combining this result with (3.10) we finally get

$$L_v^2 {}'g_{\alpha\beta} = 2B^{\mu\nu}_{\alpha\beta} v^\kappa v^\lambda R_{\lambda\mu\kappa\nu} + 2\rho^4(H_{\beta\kappa} + \Omega_{\beta\kappa})(H_\alpha{}^\kappa + \Omega_\alpha{}^\kappa) + \rho^2({}'\nabla_\alpha L_v V_\beta + {}'\nabla_\beta L_v V_\alpha) + 2\rho^2({}'\partial_\alpha \rho L_v V_\beta + {}'\partial_\beta \rho L_v V_\alpha) - 2\rho'\nabla_{\alpha\beta}\rho - (2L_v V_\alpha)(L_v V_\beta) + (L_v\rho^2)H_{\alpha\beta}. \quad (4.13)$$

Thus, the evolution of the metric along the vector field v^ν is essentially governed by the projections of the Riemann tensor twice onto the vector field and twice onto the orthogonal surface element. But, as shown at the end of the last section, these are equivalent, modulo the six terms $B^{\mu\nu}_{\alpha\beta}R_{\mu\nu}$, to the contraction of the components of the Riemann tensor given by Gauss' Eq. (3.5). So if, we assume these six field equations $B^{\mu\nu}_{\alpha\beta}R_{\mu\nu} = 0$ to hold, it follows from (3.7) that

$$B^{\mu\nu}_{\alpha\beta} v^\kappa v^\lambda R_{\lambda\mu\kappa\nu} = \rho^2[{}'R_{\alpha\beta} + 3\rho^2\Omega_{\kappa\alpha}(\Omega_\beta{}^\kappa + H_\beta{}^\kappa) + \rho^2 H_{\alpha\kappa}(\Omega^\kappa{}_\beta + H^\kappa{}_\beta) - \rho^2 H(\Omega_{\alpha\beta} + H_{\alpha\beta})]. \quad (4.14)$$

Inserting (4.14) into (4.13), we see that the six field equations $B^{\mu\nu}_{\alpha\beta}R_{\mu\nu} = 0$ are equivalent to the six evolution equations for ${}'g_{\alpha\beta}$:

$$L_v^2 {}'g_{\alpha\beta} = 2\rho^2 {}'R_{\alpha\beta} + 4\rho^4(H_\alpha{}^\kappa H_{\kappa\beta} + \Omega_\alpha{}^\kappa\Omega_{\kappa\beta}) - 2\rho^4(\Omega_{\alpha\beta} + H_{\alpha\beta})H + \rho^2({}'\nabla_\alpha L_v V_\beta + {}'\nabla_\beta L_v V_\alpha) + 2\rho^2({}'\partial_\alpha\rho L_v V_\beta + {}'\partial_\beta\rho L_v V_\alpha) - 2\rho({}'\nabla_{\alpha\beta}\rho) - 2(L_v V_\alpha)(L_v V_\beta) + (L_v\rho^2)H_{\alpha\beta}. \quad (4.15)$$

The remaining four field equations $G_{\mu\nu}v^\mu v^\nu = 0$ and $G_{\mu\nu}B^\mu_\alpha v^\nu = 0$ are equivalent to

$$-{}'R + \rho^2(H^2 - H_{\alpha\beta}H^{\alpha\beta}) + 3\rho^2\Omega_{\alpha\beta}\Omega^{\alpha\beta} = 0, \quad (4.16)$$

and

$${}'\nabla_\beta(H_\alpha{}^\beta + \Omega_\alpha{}^\beta) - {}'\partial_\alpha H - H\,{}'\partial_\alpha(\ln\rho) + (3\Omega_\alpha{}^\beta + H_\alpha{}^\beta)\partial_\beta(\ln\rho) - 2\Omega_\alpha{}^\beta L_v V_\beta = 0, \quad (4.17)$$

by Eqs. (4.4) and (4.5), respectively.

Thus, looking at (4.15)–(4.17), we can see that if, on the initial cross section of the vector field v^μ, we are given the nonmetrical data V_α, ρ, $L_v V_\alpha$, $L_v\rho$ (from which $\Omega_{\alpha\beta}$ follows, as pointed out in Sec. 2), plus the metrical data ${}'g_{\alpha\beta}$ and $H_{\alpha\beta}$ (which, together with ρ is equivalent to $L_v {}'g_{\alpha\beta}$), all this data being subject to the constraints (4.16)–(4.17) on the initial cross section, then (4.15) enables us to compute $L_v^2 {}'g_{\alpha\beta}$, and thus $L_v^2 g_{\alpha\beta}$. Iteration of this procedure, at least in the analytic case, enables us to calculate the metric tensor on each cross section of the vector field resulting from dragging the initial cross section some parameter distance. Moreover, the well-known consequence of the Bianchi identities that the system is in involution in the sense of Cartan[13] means that once the constraints (4.16)–(4.17) are satisfied on the initial cross section, they will be satisfied on any other cross section if the six evolution equations (4.15) are satisfied on the initial and all intervening cross sections.

Thus we have found the required anholonomic generalization of the Cauchy problem for the orthogonal surface elements of an arbitrary vector field.

V. A SIMPLE APPLICATION: TIMELIKE KILLING VECTOR FIELDS

As a simple illustration of how these results may be used to build into a space-time satisfying the Einstein equations a timelike vector field with given properties, let us consider the case of a stationary space–time, i.e., one with a timelike Killing vector field. That is, we specify that the vector field v^μ is to be a Killing vector when the metric $g_{\alpha\beta}$ has been constructed by our prescription. The first thing to do is to see how to express the fact that v^μ is a Killing vector in terms of the decomposition of $\nabla_\mu V_\mu$, Eq. (2.12). A short calculation shows that

$$B^{\alpha\beta}_{\mu\nu}(\nabla_\alpha v_\beta + \nabla_\beta v_\alpha) = 2\rho^2 H_{\mu\nu}, \quad (5.1a)$$

$$v^\alpha v^\beta(\nabla_\alpha v_\beta + \nabla_\beta v_\alpha) = L_v\rho^2, \quad (5.1b)$$

$$B^\alpha_\mu v^\beta(\nabla_\alpha v_\beta + \nabla_\beta v_\alpha) = \rho^2 L_v V_\mu. \quad (5.1c)$$

(ρ is, of course, the length of the Killing vector.) Thus, each component of Killing's equations finds a simple expression: The vanishing of the orthogonal components is equivalent to $H_{\mu\nu} = 0$ (the condition for a Born-rigid congruence), while the vanishing of the other components requires that $L_v\rho = L_v V_\mu = 0$. Thus, a Killing congruence can only have $\Omega_{\mu\nu}$ and ${}'\partial_\mu\rho \neq 0$. If $\Omega_{\mu\nu} = 0$ in addition, the congruence is hypersurface orthogonal, as we have seen; so the metric is static, of course. In addition, of course $L_v^2 {}'g_{\mu\nu}$ will vanish for a Killing vector. We thus get the following system of equations for a stationary space–time:

$$\rho\,{}'R_{\alpha\beta} - {}'\nabla_{\alpha\beta}\rho + 2\rho^3\Omega_{\kappa\alpha}\Omega_\beta{}^\kappa = 0; \quad (5.2a)$$

$$-{}'R + 3\rho^2\Omega_{\alpha\beta}\Omega^{\alpha\beta} = 0; \quad (5.2b)$$

$${}'\nabla_\beta\Omega_\alpha{}^\beta + 3\Omega_\alpha{}^\beta\partial_\beta(\ln\rho) = 0. \quad (5.2c)$$

These are equivalent to the form of the field equations for a stationary metric given by Lichnerowicz.[14]

If one confined oneself to the case of a unit vector field, these equations would assume a much more complicated form. Indeed, the fact that the generalized acceleration, defined as $L_v V_\mu$, vanishes for any vector field whose Killing form has vanishing projections once along the vector field and once orthogonal to it, provides a strong argument for the utility of this definition. We can see that in this sense a Killing vector field is a generalization of a shear-free geodesic field.

VI. CONCLUSION

As the last section shows, the approach of this paper may facilitate the construction of space–times possessing timelike vector fields with preassigned properties, either such properties as limit the class of space–times having such vector fields; or properties which constitute no limitation on the space–time (other than satisfying the field equations), but which may simplify the form of the field equations. In either case, the use of this method may result in forms of the field equations useful for their actual integration in the evolution form into which they have been cast. Smarr and York[7] have discussed this question for the holonomic case, where ${}'g_{\mu\beta}$ is actually the induced metric on a family of spacelike hypersurfaces.

Finally, as mentioned in the Introduction, we are particularly interested in the generalization of the techniques of this paper to the case where two commuting vectors fields, spanning a family of timelike 2-surfaces are introduced and

the evolution of the metric tensor along both fields is to be determined. This problem provides the nonholonomic generalization of the double-null and Bondi–Sachs type of initial value problem, as well as several other possible types.[9] Consideration of the spacelike 2-surface element spanned by the two vector fields leads to a two-plus-two breakup of the covariant derivatives of the vector fields, of the metric and Riemann tensor and of the field equations. However, it is obvious that we cannot, without loss of generality, both demand that the two vector fields commute and that they be unit vectors. Since commutation of the two vector fields enables us to materially simplify the analysis, we prefer to keep this condition, and are led to the study of nonunit vectors. Thus, the analysis in this paper provides the background for the step from the three-plus-one to the two-plus-two breakup. This problem will be discussed in a paper with Ben Rosen, based on his Ph.D. thesis. The discussion at the end of Sec. 3 also suggests that the two-plus-two breakup of the field equations can be done entirely in terms of orthogonal 2-spaces, without explicit introduction of the two commuting vector fields which span the timelike 2-space. This problem will also be discussed in a separate paper.

[1]J. Stachel, "Lie Derivatives and the Cauchy Problem in the General Theory of Relativity," Ph.D. dissertation (Department of Physics, Stevens Institute of Technology, 1962)(unpublished).

[2]J. Stachel, Acta Phys. Polon. **35**, 689 (1969).

[3]N. O'Murchadha, "Existence and Uniqueness of Solutions to the Hamiltonian Constraint of General Relativity," Ph.D. dissertation (Department of Physics, Princeton University, 1973)(unpublished).

[4]J.D. Kulhanek, Lett. Nuovo Cimento **10**, 389 (1974).

[5]C. Møller, *The Theory of Relativity* (Oxford, Clarendon, 1972), 2nd ed.

[6]J.A. Schouten, *Ricci Calculus* (Springer, Berlin, 1954).

[7]L. Smarr and J.W. York, Phys. Rev. D **17**, 2529 (1978).

[8]For definitions of the lapse and shift functions, see C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[9]R.A. D'Inverno and J. Stachel, J. Math. Phys. **19**, 2447 (1978).

[10]For the definitions of the shear, expansion, rotation, and acceleration of a unit timelike congruence, see J. Ehlers and W. Kundt, "Exact Solutions of the Gravitational Field Equations," in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), p. 58.

[10a]Note that as a result of this definition ${}'R_{\mu\nu\kappa\lambda}$ does not have all of the symmetries of a metric Riemann tensor when $\Omega_{\mu\nu}$ does not vanish. It is not antisymmetric in the last pair of indices, nor symmetric under interchange of the first and second pair of indices. As a consequence, ${}'R_{\mu\nu} = {}'g^{\kappa\lambda}\,{}'R_{\mu\kappa\lambda\nu}$ is not symmetric unless $\Omega_{\mu\nu} = 0$.

[11]We follow the idea of Schouten in Ref. 6, p. 132; but note that we adopt a different definition. The semicolon used hereafter stands for covariant differentiation, and is thus equivalent to ∇.

[12]J. Stachel, "String Dusts, Fluids and Subspaces," to appear in *Proceedings of the Third Latin American Symposium on Relativity and Gravitation*.

[13]See for example, A. Lichnerowicz, *Théories Relativistes de la Gravitation et de l'Electromagnétisme* (Masson, Paris, 1955), p. 32.

[14]A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson et Cie, Paris, 1955), p. 120. Note his tensor H is one-half of Ω.

On the local simultaneity in general relativity

J. Olivert
Departamento de Mecánica y Astronomia, Facultad de Matemáticas, Burjasot (Valencia), Spain

(Received 17 January 1980; accepted for publication 20 March 1980)

In this paper the existence of simultaneous points, in the Landau sense, in the neighborhood of an observer P is studied. It is shown that the set of such points has the structure of a regular submanifold of the space–time manifold.

1. INTRODUCTION

Let us consider the space–time as a connected pseudo-Riemannian manifold V, of Hausdorff type and four dimensions, in which a metric tensor field of hyperbolic signature (3,1) has been defined. The coordinate x^4 of a point, belonging to a coordinate neighborhood of a given local chart F, coincides with the coordinate time in the system described by F. It is assumed that the measure units have been chosen such that the constant light velocity in the vacuum has the value 1.

In these measure units, the velocity 4-vector u of a moving point P in V has the contravariant components, in the proper system of P, given by

$$u^i = (0,0,0,1). \tag{1.1}$$

We notice that the indices represented by latin letters take values from 1 to 4 and the greek indices are restricted to the values 1, 2, 3.

The components of the Minkowski tensor are represented by η_{ij}, given by

$$\eta_{ij} = \mathrm{diag}(1,1,1,-1), \tag{1.2}$$

leaving the notation g to denote the metric tensor field of V.

Landau and Lifshitz[1] established a method to synchronize nearest points with P. This nearness is understood in the sense that the coordinate differences between P and one of its nearest points Q are infinitesimals $\{\Delta x^i\}$. The method to synchronize such points consists in that the time for a light signal to travel from P to Q is determined in the proper system of P by

$$\Delta x^4 = -\frac{g_{4\alpha}dx^\alpha}{g_{44}}; \tag{1.3a}$$

another, equivalent, relation is

$$g_{4i}dx^i = 0. \tag{1.3b}$$

Eq. (1.3) will be taken by us as the *Landau–Lifshitz simultaneity condition.*

Finally, we represent by T_PV the tangent space of V at P, and by $\phi_{\bullet m}$ the derived linear function at $m \in V$ of a differentiable function defined in V. The concepts of differential geometry necessary to develop the following calculations may be found in Brickell and Clark.[2]

2. PHYSICAL SPACE AND LANDAU MANIFOLD OF AN OBSERVER

Definition: The physical space E_P of an observer P is the hyperplane orthogonal to its velocity 4-vector u. From its own definition, as u is a time-like vector

$$g(u,u) < 0,$$

the vectors of E_p are space-like,

$$g(v,v) > 0,$$

because a time-like vector is orthogonal to neither another time-like vector nor to a light-like vector (Synge).[3]

Theorem 2.1: Let P be a moving point, u its velocity 4-vector and E_P the physical space of P. There is a unique submanifold L of dimension 3 such that E_P is tangent to L at P and its points are simultaneous with P in the local inertial proper system of P.

Proof: As the space–time V is a linear connected manifold, at each point $p \in V$ there exists a neighborhood W diffeomorphic to a neighborhood of $0 \in T_PV$ by the function $\exp_P$. Taking W open, with the velocity 4-vector u of P and the metric tensor field g, we define the submersion

$$\phi\colon\ W \to R,\ m \to g(\exp_P^{-1} m, u).$$

As

$$L = \phi^{-1}(0) \tag{2.1}$$

is fiber of ϕ, it is a regular submanifold of V of dimension 3.

We prove that $P \in L$. Take c as a geodesic starting from P

$$c(0) = P.$$

There exists a $v \in T_pV$ such that

$$\exp_P sv = c(s), \quad \forall s \in R,$$

and therefore

$$\exp_P(0) = c(0) = P.$$

As $\exp_P$ is a diffeomorphism in W,

$$\exp_P^{-1} P = 0,$$

with the result that

$$\phi(P) = g[\exp_P^{-1}(P), u] = g(0,u) = 0,$$

leading to

$$P \in \phi^{-1}(0) = L.$$

We prove that E_P *is tangent to* L. Let us consider the foliation

$$\Omega(m) = \ker\phi_{\bullet m}, \quad \forall m \in W, \tag{2.2}$$

in which L is an integral manifold. The natural injection

$$i\colon\ L \to W$$

is, therefore, an immersion which satisfies

$$i_\bullet(T_PL) = \ker\phi_{\bullet P}. \tag{2.3}$$

 0022-2488/80/071783-03$1.00

On the other hand, let us consider the geodesic local chart

$$G = h\cdot\exp_P^{-1} \tag{2.4}$$

of coordinate neighborhood W, with h the isomorphism of T_PV in R^4, such that the components of the metric tensor are

$$g_{ij} = \eta_{ij}, \tag{2.5}$$

and such that a vector $v\in E_P$ would be expressed in this local chart by

$$v = v^\alpha\left(\frac{\partial}{\partial x^\alpha}\right)_P, \tag{2.6}$$

and the velocity 4-vector by

$$u = \frac{\partial}{\partial x^4}. \tag{2.7}$$

Equation (2.7) gives to G the property of being the proper system of P.

To an element $x\in W$ there corresponds the vector of T_PV

$$\exp^{-1}x = x^i\left(\frac{\partial}{\partial x^i}\right)_P,$$

expressed in the local chart G. Then

$$\phi(x) = g(\exp^{-1}x,u) = \eta_{i4}x^i = -x^4. \tag{2.8}$$

If we call $\partial/\partial t$ the corresponding vector field on R of the identity chart on R, one has

$$\phi_{*P}(v) = v^\alpha\left(\frac{\partial\phi(x)}{\partial x^\alpha}\right)_P\left(\frac{\partial}{\partial t}\right)_{\phi(P)} = v^\alpha\left(\frac{\partial(-x^4)}{\partial x^\alpha}\right)_P\left(\frac{\partial}{\partial t}\right)_0 = 0;$$

that is,

$$v\in \ker\phi_{*P},$$

and therefore

$$E\subset \ker\phi_{*P},$$

As

$$\dim E = \dim\ker\phi_{*P},$$

we have proved that

$$E = \ker\phi_{*P},$$

and, using (2.3),

$$i_*(T_PL) = E_P; \tag{2.9}$$

i.e., E_P is tangent to L.

We prove that the points of L are simultaneous with P. We have seen that, for a geodesic coordinate system

$$\phi(x) = -x^4, \quad \forall x\in W,$$

x^4 being the fourth coordinate of x in that system. Let us take $m\in L$; using (2.1)

$$x^4 = -\phi(m) = 0,$$

that is,

$$G(m) = (x^1,x^2,x^3,0); \quad \forall m\in L. \tag{2.10}$$

We prove the uniqueness. Let us assume that, apart from L, there is another submanifold L_1, which satisfies the conditions of the theorem. The uniqueness will be proved if one shows that L_1 is an open set of L.

Take $m_1\in L_1$; as m_1 is simultaneous with P in the geodesic coordinate system the use of (2.10) says that

$$G(m_1) = (x_1^1,x_1^2,x_1^3,0).$$

From the definition (2.4) of G we obtain

$$\exp_P^{-1}(m_1) = h^{-1}(x_1^1,x_1^2,x_1^3,0)\in E_P,$$

hence

$$\phi(m_1) = g(\exp_P^{-1}m_1,u) = 0,$$

and therefore

$$m_1\in\phi^{-1}(0) = L \Rightarrow L_1\subset L. \tag{2.11}$$

On the other hand, as L_1 is a submanifold of W, the natural injection

$$i_1\colon L_1\to W$$

is an immersion, therefore a differentiable map. From (2.11) we know

$$\mathrm{Im}i_1\equiv L_1\subset L,$$

and L is a regular submanifold of W. The function j induced by i_1

$$j\colon L_1\to L,$$

is differentiable and with rank

$$\mathrm{rank}\,j = \mathrm{rank}\,i_1 = \dim L_1 = 3;$$

therefore j is an immersion. With this result, one has that L_1 is a submanifold of L. Furthermore, as

$$\dim L_1 = \dim L,$$

one sees that L_1 is an open set of L. This completes the proof of our theorem.

For reasons to be justified below and which are the central subject of this work, we introduce the following

Definition: We call the Landau manifold of a point P to the submanifold L tangent to E_P, in agreement with the above theorem (2.1).

Theorem 2.2: The connected Landau manifold of a point P can be implemented with a Riemannian structure.

Proof: Given a Landau manifold L, one can extract an open submanifold of L with the property of being connected and containing P. It is enough to take the connected component L_c which contains P. This connected component L_c is also a submanifold of V tangent to E_P at P.

From the natural immersion

$$i\colon L_c\to V,$$

we define in L_c the metric γ as

$$\gamma = i^*g.$$

We must prove that γ is a positive definite metric for each point of L_c. Let us consider two arbitrary vectors v, $w\in T_PL_c$. Using (2.9), i_*v, $i_*w\in E_P$, and

$$\gamma(v,w) = i^*g(v,w) = g(i_*v,i_*w) > 0; \quad v,w\neq 0,$$

γ is a positive definite metric at P. Therefore, as L_c is connected, the signature is conserved at all points of the manifold.[4]

3. STUDY OF THE POINTS RELATED WITH THE LANDAU–LIFSHITZ SIMULTANEITY CONDITION

With the objective of proving that the points joined by the Landau–Lifshitz simultaneity condition belong to the Landau manifold, we now show the following proposition.

Theorem 3.1: Let F be a local chart of P with the condition $F(P) = 0$. For each $v \in T_p V$, there is an $\epsilon > 0$ such that for all $s \in R$ with $|s| \leqslant \epsilon$, it satisfies p.p. $F \cdot \exp_P sv = F_{*P} sv$, with p.p. the first order approximation of the corresponding function.

Proof: We know that

$$\exp_P sv = c(s) \tag{3.1}$$

is a geodesic starting from P such that

$$\dot{c}(0) = v,$$

with $\dot{c}$ denoting the canonical lift of c.

The curve of R^4 $F \cdot c(s)$ starts from $0 \in R^4$, and its canonical lift at 0 is

$$\dot{F \cdot c}\,(0) = F_{*P} \dot{c}(0) = F_{*P} v. \tag{3.2}$$

Following (3.1) we develop the expression $F \cdot \exp_P sv$ to first order in s

$$F \cdot \exp_P sv = F \cdot c(s) = F \cdot c(0) + \left. \frac{d\, F \cdot c}{ds} \right|_{s=0} s + O_2(s)$$
$$= \left. \frac{d\, F \cdot c}{ds} \right|_{s=0} s + O_2(s),$$

and, using (3.2)

$$\left. \frac{d\, F \cdot c}{d \cdot s} \right|_{s=0} = F_{*P} \cdot \dot{c}(0) = F_{*P} v,$$

we obtain

$$F \cdot \exp_P sv = sF_* v + O_2(s) = F_* sv + O_2(s)$$
$$\Rightarrow \text{p.p. } F \cdot \exp_P sv = F_* sv \tag{3.3}$$

as the result of the Theorem.

In agreement with the "nearness" discussed in the Introduction, a "nearest" point Q to P has the coordinates

$$F(Q) = (\Delta x^1,\ \Delta x^2,\ \Delta x^3,\ \Delta x^4)$$

and therefore

$$d\mathbf{x} = \text{p.p. } F(Q). \tag{3.4}$$

On the other hand, we can consider $Q \in W$, so there is an $sv \in T_P V$ such that

$$Q = \exp_P sv, \quad |s| \leqslant \epsilon.$$

Using (3.4) and (3.3)

$$d\mathbf{x} = \text{p.p. } F(Q) = \text{p.p. } F \cdot \exp_P sv = F_{*P} sv.$$

Due to the isomorphism F_{*P}, we identify $d\mathbf{x}$ and sv. This identification allows us to write

$$Q = \exp_P d\mathbf{x}$$
$$\Rightarrow d\mathbf{x} = \exp_P^{-1} Q. \tag{3.5}$$

From Eq. (3.5) we get

$$g(\exp_P^{-1} Q, u) = g(d\mathbf{x}, u) = g_{ij} dx^i u^j,$$

and (1.1) is verified, since the Landau–Lifshitz simultaneity condition is expressed in the proper system of P. Using (1.3) and the definition of ϕ

$$\phi(Q) = g(\exp_P^{-1} Q, u) = g_{4i}\, dx^i = 0,$$

from which together with Eq. (2.1), $Q \in L$.

With this result we show that the points related with P by the Landau–Lifshitz condition belong to the Landau manifold of P. Finally, in the literature of Relativity the standard time is sometimes mentioned, as it is, for example, by Tonnelat,[5] and defined by

$$dT = -g(u, d\mathbf{x}).$$

It indicates the time of an event as measured by an observer, who corresponds to P in our case. With this new concept, we can say that the points of the Landau manifold are not only simultaneous with P in the proper inertial local system of P but also simultaneous in any system with the condition to measure the simultaneity with the standard time.

4. DISCUSSION

First we must notice that it is usual to express the components of the velocity 4-vector in the proper system as $u^i = (0, 0, 0, u^4)$ instead of (1.1).

This does not change anything in the procedure of the proofs and conclusions of the above sections, because the metrics of both are related by the factor u^4. This corresponds in physics to choosing the measure units.

The existence of submanifolds tangent to hyperplanes of the tangent space at a point can be proven for pseudo-Riemannian manifolds following a similar procedure to the one used in the Theorem 2.1. The uniqueness of these submanifolds has been obtained by imposing the simultaneity condition among its points.

On the other hand, we have seen that the synchronization established by Landau–Lifshitz is equivalent to the simultaneity of the synchronized points if the standard time is taken as the time measure.

The description of the synchronization process leads us to say that the synchronized points are arcwise connected. This implies that such points belong to the connected Landau manifold, because any arcwise connected manifold is connected.

To conclude, by virtue of our Theorem 2.2, the local simultaneity in Relativity, with respect to a point P, remains restricted to a Riemannian submanifold L_c tangent at P to its physical space.

ACKNOWLEDGMENTS

I would like to thank Dr. A. M. Naveira and Dr. V. Miquel for the suggestions related to the mathematical content of this work and, especially, Dr. J. Bernabéu for his care in reading the manuscript and suggestions about its physical content.

[1] Landau and Lifshitz, *Classical Field Theory*, edited by Reverté (S. A., Barcelona, Spain, 1973).

[2] F. Brickell and R. S. Clark, *Differentiable Manifolds* (Van Nostrand–Reinhold, London, England, 1970).

[3] J. L. Synge, *Relativity: The special theory* (North Holland, Amsterdam, 1965).

[4] Y. Choquet-Bruhat, *Géometrie Differentielle et Systèmes extérieurs* (Dunod, Paris, 1968).

[5] M. A. Tonnelat, *Vérifications expèrimentales de la Relativité Générale* (Masson, Paris, 1964).

The asymptotic behavior of the generalized Einstein–Maxwell field theory

Gregory Walter Horndeski
Department of Applied Mathematics, [a] *University of Waterloo, Waterloo, Ontario, Canada*

(Received 24 December 1979; accepted for publication 22 February 1980)

It is shown that the Einstein–Maxwell and generalized Einstein–Maxwell field theories essentially agree to first order at null and spatial infinity.

1. INTRODUCTION

Let $(\tilde{M},\tilde{g}_{ab})$ be a spacetime; i.e., a four-dimensional manifold $\tilde{M}$ with a Lorentzian metric tensor $\tilde{g}_{ab}$ of signature $+2$. In the generalized Einstein–Maxwell theory[1] of gravitation and electromagnetism the source-free equations governing $\tilde{g}_{ab}$ and the electromagnetic field tensor $\tilde{F}_{ab}$ are[2]

$$\tilde{G}^{ab} = 8\pi\tilde{\mathcal{T}}^{ab}, \tag{1.1}$$

$$\tilde{\nabla}_b\tilde{F}^{ab} + \tfrac{1}{2}k\tilde{\nabla}_b\tilde{F}_{cd}{}^*\tilde{R}^{*abcd} = 0, \tag{1.2}$$

and

$$\tilde{\nabla}_{[a}\tilde{F}_{bc]} = 0, \tag{1.3}$$

where

$$\tilde{\mathcal{T}}^{ab} := \tilde{T}^{ab} + k\tilde{A}^{ab},$$

$$\tilde{T}^{ab} := \frac{1}{4\pi}(\tilde{F}^{ac}\tilde{F}^{b}{}_{c} - \tfrac{1}{4}\tilde{g}^{ab}\tilde{F}^{cd}\tilde{F}_{cd}),$$

$$\tilde{A}^{ab} := \frac{1}{8\pi}(\tilde{F}_{ce}\tilde{F}_{d}{}^{e}{}^*\tilde{R}^{*acbd} + \tilde{\nabla}_c{}^*\tilde{F}^{ad}\tilde{\nabla}_d{}^*\tilde{F}^{bc}),$$

and k is a constant with units of length squared. Following Ref. 3, $\tilde{\mathcal{T}}^{ab}$ will be interpreted as the energy–momentum tensor of the electromagnetic field in the generalized Einstein–Maxwell theory.

When $(\tilde{M},\tilde{g}_{ab})$ is Minkowski space, Eqs. (1.2) and (1.3) reduce to Maxwell's equations, while $\tilde{\mathcal{T}}^{ab}$ does *not* reduce to $\tilde{T}^{ab}$, which is Maxwell's energy–momentum tensor for the electromagnetic field. However, in Ref. 3 it is demonstrated that near spatial infinity in Minkowski space $\tilde{\mathcal{T}}^{ab}$ is dominated by $\tilde{T}^{ab}$ when evaluated for either static or outgoing multipole radiation solutions to Maxwell's equations which originate from spatially bounded sources. (By $\tilde{\mathcal{T}}^{ab}$ being dominated by $\tilde{T}^{ab}$ near spatial infinity, I mean that with respect to a Lorentzian coordinate system for Minkowski space the leading term in an r^{-1} series expansion of $\tilde{\mathcal{T}}^{ab}$ comes from $\tilde{T}^{ab}$.) The results presented in Ref. 3 can also be employed to show that $\tilde{\mathcal{T}}^{ab}$ is *not* dominated by $\tilde{T}^{ab}$ near null infinity when evaluated for outgoing multipole radiation solutions to Maxwell's equations in Minkowski space. In fact, in this case $\tilde{A}^{ab}$ and $\tilde{T}^{ab}$ have the same order near null infinity.

The purpose of this paper is to extend the investigation of the asymptotic behavior of $\tilde{\mathcal{T}}^{ab}$ in Minkowski space to a study of the asymptotic structure of the asymptotically flat solutions of the source-free generalized Einstein–Maxwell field equations. Now there exist two regimes in which a spacetime representing an isolated system can be regarded as being asymptotically flat; viz., at null and at spatial infinity. We shall discuss the properties of the asymptotically flat solutions to the source-free generalized Einstein–Maxwell field equations in both of these regimes. The main conclusion of our analysis will be that asymptotically, the Einstein–Maxwell and generalized Einstein–Maxwell field theories are essentially indistinguishable to first order. We shall also find that the behavior of $\tilde{\mathcal{T}}^{ab}$ for outgoing multiple solutions to Maxwell's equations in Minkowski space extends to the asymptotically flat generalized Einstein–Maxwell spacetimes, in the sense that $\tilde{\mathcal{T}}^{ab}$ agrees with $\tilde{T}^{ab}$ to first order at spatial infinity while $\tilde{\mathcal{T}}^{ab}$ differs from $\tilde{T}^{ab}$ to first order at null infinity, due to a contribution from $\tilde{A}^{ab}$.

Our treatment of null infinity will be based upon Geroch's notion of an asymptote[4] for a spacetime, and our treatment of spatial infinity will be based upon Ashtekar and Hansen's work.[5] We assume that the physical spacetime metric and electromagnetic field are smooth when discussing null infinity, and of class C^4, C^3, and C^2 respectively, when discussing spatial infinity.

2. BEHAVIOR AT NULL INFINITY

By an *asymptote* for the physical spacetime $(\tilde{M},\tilde{g}_{ab})$ we shall mean a triple, (M,g_{ab},Ω), consisting of a four-dimensional C^∞ manifold M, with boundary I, a smooth Lorentzian metric g_{ab} on M, and a smooth real-valued function Ω on M, which is such that[6]:

(i) There exists a diffeomorphism ψ of $\tilde{M}$ onto the interior of M (by means of which we shall identify $\tilde{M}$ with $M\setminus I$);

(ii) $g_{ab} = \Omega^2\tilde{g}_{ab}$ on $\tilde{M}$; and

(iii) $I = \Omega^{-1}(0)$ with $d\Omega \neq 0$, at any point of I.

Suppose that (M,g_{ab},Ω) is an asymptote for the physical spacetime $(\tilde{M},\tilde{g}_{ab})$, and $\tilde{F}_{ab}$ is a *generalized Maxwell field* on $(\tilde{M},\tilde{g}_{ab})$, i.e., a solution to the source-free generalized Einstein–Maxwell field equations. We say that $\tilde{F}_{ab}$ is *asymptotically regular* [with respect to the asymptote (M,g_{ab},Ω)] if $F_{ab} := \tilde{F}_{ab}$ admits a smooth extention to I. Ideally, one would like to be able to show that if $(\tilde{M},\tilde{g}_{ab})$ admits an asymptote, I is a null hypersurface, and the source-free generalized Einstein–Maxwell field equations are satisfied, then $\tilde{F}_{ab}$ must be asymptotically regular. Unfortunately, such a result has not even been established for the Einstein–Maxwell field theory, although comparable re-

[a] Work completed while visiting the Enrico Fermi Institute at the University of Chicago, September, 1979–May, 1980.

 0022-2488/80/071786-04$1.00

sults can be proved for this theory if a few more assumptions are added to the hypothesis.[7] At present it is not clear whether these weakened results concerning the asymptotic regularity of the electromagnetic field in the Einstein–Maxwell theory can be extended to the generalized theory since their proofs seem to hinge upon properties of the Einstein–Maxwell equations which are not possessed by the generalized Einstein–Maxwell equations. For example, Geroch's result[7] depends heavily upon the deterministic nature of Maxwell's equations in a curved space, and it appears as though the generalized Maxwell equations (1.2) and (1.3) will not enjoy this property.[8] Thus, the only motivation I can offer in defense of the assumption that asymptotic regularity is the "correct" boundary condition to be imposed upon $\tilde{F}_{ab}$ is that the electromagnetic field is asymptotically regular for the "asymptotically flat" solutions to the generalized Einstein–Maxwell field equations presented in Refs. 9 and 10.

In passing, I would like to point out that the assumption of asymptotic regularity implies that $\tilde{F}_{ab}$ has the usual peeling-off behavior (see pp. 42–46 of Ref. 4, or Sec. 13 of Penrose's article cited in Ref. 7).

We shall now demonstrate that if the generalized Maxwell field $\tilde{F}_{ab}$ is asymptotically regular with respect to the asymptote (M,g_{ab},Ω), then $\Omega^{-4}\tilde{\mathscr{T}}_a{}^b$ admits a smooth extension to I. This, in turn, will imply that I must be a null hypersurface.

In order to simplify the form of the ensuing expressions we let

$$\tilde{B}_a{}^b := \frac{1}{8\pi}\tilde{F}^{ce}\tilde{F}_{de}{}^*\tilde{R}^*{}_{ac}{}^{bd},$$
$$\tilde{C}_a{}^b := \frac{1}{8\pi}(\tilde{\nabla}_c{}^*\tilde{F}_{ad})(\tilde{\nabla}^d{}^*\tilde{F}^{bc}). \tag{2.1}$$

We also adopt the convention that indices on tilded quantities (untilded quantities, respectively) will be raised and lowered by $\tilde{g}^{ab}$ and $\tilde{g}_{ab}$ (by g^{ab} and g_{ab}, respectively).

Employing the results presented in Sec. 28 of Ref. 11, one easily finds that

$$\tilde{T}_a{}^b = \Omega^4 T_a{}^b, \tag{2.2}$$

$$\tilde{B}_a{}^b = \Omega^6 B_a{}^b - \frac{\Omega^5}{8\pi}\delta^{bdm}_{acl}F^{ce}F_{de}\nabla^l n_m - \frac{\Omega^4}{8\pi}n_m n^m\left[F_{ac}F^{bc} - \delta^b_a F_{cd}F^{cd}\right], \tag{2.3}$$

and

$$\begin{aligned}\tilde{C}_a{}^b = {}&\Omega^6 C_a{}^b + \frac{\Omega^5}{8\pi}\left[2{}^*F_{ac}n_d\nabla^c{}^*F^{bd}\right.\\ &- {}^*F_{cd}n_a\nabla^c{}^*F^{bd} + {}^*F_{ac}n_d\nabla^d{}^*F^{bc}\\ &+ 2{}^*F^{bd}n^c\nabla_d{}^*F_{ac} - {}^*F^{cd}n^b\nabla_c{}^*F_{ad}\\ &+ {}^*F^{bd}n^c\nabla_c{}^*F_{ad} - {}^*F_{ac}n^c\nabla_d{}^*F^{bd}\\ &\left.- {}^*F^{bd}n_d\nabla^c{}^*F_{ac}\right] + \frac{\Omega^4}{8\pi}\left[2{}^*F_{ac}n^c{}^*F^{bd}n_d\right.\\ &+ n_a{}^*F^{bc}{}^*F_{cd}n^d + n^b{}^*F_{ac}{}^*F^{cd}n_d\\ &\left.- n_a n^b{}^*F_{cd}{}^*F^{cd} + 4n_c n^c{}^*F_{ad}{}^*F^{bd}\right],\end{aligned} \tag{2.4}$$

where $n_a := \nabla_a\Omega$, and no use has been made of the field equations. In view of Eqs. (2.2)–(2.4), it is clear that $\Omega^{-4}\tilde{T}_a{}^b$, $\Omega^{-4}\tilde{B}_a{}^b$, and $\Omega^{-4}\tilde{C}_a{}^b$ admit smooth extensions to I. Since $\tilde{\mathscr{T}}_a{}^b = \tilde{T}_a{}^b + k(\tilde{B}_a{}^b + \tilde{C}_a{}^b)$, we must have $\Omega^{-4}\tilde{\mathscr{T}}_a{}^b$ admitting a smooth extension to I.

Following Geroch[4] we define $S_{ab} := R_{ab} - \frac{1}{6}g_{ab}R$ and $L_a{}^b := \tilde{R}_a{}^b - \frac{1}{6}\delta^b_a\tilde{R}$. Equation (1.1) implies that on $\tilde{M}$

$$L_a{}^b = 8\pi(\tilde{\mathscr{T}}_a{}^b - \tfrac{1}{3}\delta^b_a\tilde{\mathscr{T}}_c{}^c). \tag{2.5}$$

Thus, due to the above work, we know that $\Omega^{-4}L_{ab}$ admits a smooth extention to I. Under the conformal transformation $g_{ab} = \Omega^2\tilde{g}_{ab}$ we find that S_{ab} and L_{ab} are related by

$$\Omega S_{ab} + 2\nabla_a n_b - \Omega^{-1}n_m n^m g_{ab} = \Omega^{-1}L_{ab}. \tag{2.6}$$

Upon multiplying this equation by Ω and taking the limit as we approach I we see that $n_m n^m = 0$ on I. Thus I must be a null hypersurface when the generalized Maxwell field $\tilde{F}_{ab}$ is asymptotically regular, and hence we can think of I as representing "null infinity" in this case.

In passing we note that Eq. (2.6) can be used to conclude that

$$\nabla_a n_b = \tfrac{1}{2}\Omega^{-1}n_m n^m g_{ab}, \tag{2.7}$$

on I. This result will prove to be quite useful when we examine Eqs. (1.2) and (1.3) on I.

Since $\Omega^{-4}\tilde{\mathscr{T}}_a{}^b$ admits a smooth extension to I (when $\tilde{F}_{ab}$ is asymptotically regular) all of the results concerning the asymptotic geometry of I presented in Secs. 3, 4, and 5 of Ref. 4 apply to I in the present situation. In fact, these results of Geroch only require that $\Omega^{-2}\tilde{\mathscr{T}}_a{}^b$ admits a smooth extension to I.

Remark: An asymptote (M,g_{ab},Ω) for the spacetime $(\tilde{M},\tilde{g}_{ab})$ is said to be *Minkowskian* if it satisfies the following four conditions:

(i) I is a null hypersurface;

(ii) The vector field $\mathbf{n}^\alpha$ induced on I by n^a is complete;

(iii) Each component of I is a principal fibre bundle over S^2 with structure group $\mathbb{R}$ and group action given by the flow of $\mathbf{n}^\alpha$; and

(iv) $\pounds_{\mathbf{n}}\,\mathbf{g}_{\alpha\beta} = 0$, where $\mathbf{g}_{\alpha\beta}$ is the pull-back of g_{ab} to I.

It turns out that the "asymptotically flat" solutions to the generalized Einstein–Maxwell field equations presented in Refs. 9 and 10 admit Minkowskian asymptotes. The reason why Minkowskian asymptotes are important is due to the fact that if $(\tilde{M},\tilde{g}_{ab})$ admits a Minkowskian asymptote (M,g_{ab},Ω), and the generalized Maxwell field $\tilde{F}_{ab}$ is asymptotically regular, then the Weyl tensor of (M,g_{ab}) vanishes on I. (This fact follows immediately from Theorem 11 on p. 47 of Ref. 4.) Thus under these assumptions we see that it is not unreasonable to say that "the physical spacetime is asymptotically flat at null infinity". □

We shall now turn our attention to the asymptotic field equations and the behavior of $\Omega^{-4}\tilde{\mathscr{T}}_a{}^b$ on I. To this end suppose that the generalized Maxwell field $\tilde{F}_{ab}$ is asymptotically regular. We let $\mathbf{F}_{\alpha\beta}$ and ${}^*\mathbf{F}_{\alpha\beta}$ denote the pullback to I of F_{ab} and ${}^*F_{ab}$ with the natural injection $j:I\to M$. $\mathbf{F}_{\alpha\beta}$ and ${}^*\mathbf{F}_{\alpha\beta}$ are regarded as representing the asymptotic electromagnetic field at null infinity.

Since j commutes with exterior differentiation it is clear that Eq. (1.3) implies that $d\,\mathbf{F} = 0$ on I. We shall presently demonstrate that Eq. (1.2) implies that $d{}^*\mathbf{F} = 0$ on I, and

hence $\mathbf{F}_{\alpha\beta}$ and $*\mathbf{F}_{\alpha\beta}$ satisfy the same equations on I as do asymptotically regular Maxwell fields. As a result we arrive at the usual expression for the total electric and magnetic charges for $(\tilde{M},\tilde{g}_{ab})$ (see p. 58 of Ref. 4).

A straightforward calculation shows that

$$\begin{aligned}\tilde{\nabla}_b\tilde{F}^{ab} &+ \tfrac{1}{2}k\tilde{\nabla}_b\tilde{F}_{cd}*\tilde{R}*^{abcd}\\ &= \Omega^4[\nabla_b F^{ab} + \tfrac{1}{2}k(\Omega\nabla_b F_{cd} + 2F_{cd}n_b\\ &\quad + 2F_{bd}n_c - 2g_{bc}n^e F_{ed})(\Omega*R*^{abcd}\\ &\quad - g^{ae}g^{bf}\delta^{cdh}_{efl}\nabla^l n_h + \Omega^{-1}n_m n^m g^{ae}g^{bf}\delta^{cd}_{ef})]. \qquad (2.8)\end{aligned}$$

Due to Eq. (2.7) we see that Eqs. (1.2) and (2.8) can be used to conclude that on I, $\nabla_b F^{ab} = 0$, or equivalently, $\nabla_{[a}*F_{bc]} = 0$ on I. Upon pulling this equation back to I with j we find that $d*\mathbf{F} = 0$, as claimed.

Consequently on I we cannot distinguish $(\mathbf{F},*\mathbf{F})$ pairs which arise from asymptotically regular Maxwell fields from such pairs which arise from asymptotically regular generalized Maxwell fields.

Using Eqs. (2.2)–(2.5) we see that when the generalized Maxwell field $\tilde{F}_{ab}$ is asymptotically regular[12]

$$\begin{aligned}\lim_{\to I}\Omega^{-4}L_{ab} &= 2(F_{ac}F_b{}^c - \tfrac{1}{4}g_{ab}F_{cd}F^{cd})\\ &\quad + k(2*F_{ac}n^c*F_{bd}n^d + n_a*F_{bc}*F^{cd}n_d\\ &\quad + n_b*F_{ac}*F^{cd}n_d - n_a n_b*F_{cd}*F^{cd}). \qquad (2.9)\end{aligned}$$

Since $j^* d\Omega = 0$, the pullback of the second term on the right-hand side of (2.9) is

$$2k*\mathbf{F}_{\alpha\gamma}\mathbf{n}^\gamma*\mathbf{F}_{\beta\delta}\mathbf{n}^\delta, \qquad (2.10)$$

which is nonzero (in general) and vanishes if there is no radiation incident upon null infinity. Thus $\mathbf{L} := j^*(\lim_{\to I}\Omega^{-4}L_{ab}dx^a\otimes dx^b)$ will differ from its counterpart in the Einstein–Maxwell theory due to the presence of the term given in (2.10). However, it is conventional to regard $\mathbf{L}_{\alpha\beta}\mathbf{n}^\beta$ as the flux of the electromagnetic field through I (see p. 56 of Ref. 4), and this quantity is given by the usual expression since $*\mathbf{F}_{\beta\delta}\mathbf{n}^\beta\mathbf{n}^\delta = 0$. Consequently the new term in $\mathbf{L}_{\alpha\beta}$ which arises in the generalized Einstein–Maxwell theory leaves no impression upon the usual quantities of physical significance on I, such as the flux of the electromagnetic field, the News tensor, or the Bondi energy–momentum vector.[13]

Thus, in summary, if $(\tilde{M},\tilde{g}_{\alpha\beta})$ admits an asymptote, and the electromagnetic field $\tilde{F}_{ab}$ is asymptotically regular, then the geometry and physics of I are essentially the same irrespective of whether the Einstein–Maxwell or generalized Einstein–Maxwell field equations hold on $\tilde{M}$.

3. BEHAVIOR AT SPATIAL INFINITY

Following Ashtekar and Hansen,[5] we say that the physical spacetime $(\tilde{M},\tilde{g}_{ab})$ is *asymptotically flat at spatial infinity* (abbreviated by AFASI) if there exists a quadrupole (M,i^0,g_{ab},Ω) consisting of a four-dimensional manifold M with a $C^{>1}$ differentiable structure[14] at $i^0\in M$, a Lorentzian metric g_{ab} of class $C^{>0}$ at i^0, and a real valued function Ω on M of class $C^{>2}$ at i^0, which are such that:

(i) There exists a diffeomorphism ψ of $\tilde{M}$ into M (by means of which we shall identify $\tilde{M}$ with $\psi(\tilde{M})\subset M$);

(ii) The boundary of $\tilde{M}$ as a subset of M consists of the null cone of $i^0\notin\tilde{M}$;

(iii) $g_{ab} = \Omega^2\tilde{g}_{ab}$ on $\tilde{M}$; and

(iv) $\Omega(i^0) = 0$, $d\Omega(i^0) = 0$, and $\nabla_a\nabla_b\Omega(i^0) = 2g_{ab}(i^0)$.

The quadrupole (M,i^0,g_{ab},Ω) will be referred to as *the completion* of the AFASI spacetime $(\tilde{M},\tilde{g}_{ab})$, and i^0 will be considered as representing spatial infinity for $(\tilde{M},\tilde{g}_{ab})$.

Suppose that (M,i^0,g_{ab},Ω) is a completion of the AFASI spacetime $(\tilde{M},\tilde{g}_{ab})$, and $\tilde{F}_{ab}$ is a generalized Maxwell field on $\tilde{M}$. We set $F_{ab} := \tilde{F}_{ab}$ on $\tilde{M}$ and say that $\tilde{F}_{ab}$ is asymptotically regular[15] (with respect to (M,g_{ab},i^0,Ω)) if ΩF_{ab} admits a regular direction-dependent limit at i^0. This in turn implies that $\lim_{\to i^0}\Omega F_{ab}$ gives rise to a direction-dependent tensor $\mathbf{F}_{ab}$ on the hyperboloid $\mathscr{H}$ of unit spacelike vectors at i^0 and that for $k = 1,2$

$$\partial_{a_1}\cdots\partial_{a_k}\mathbf{F}_{ab} = \lim_{\to i^0}\Omega^{1/2}\nabla_{a_1}[\cdots(\Omega^{1/2}\nabla_{a_k}\Omega F_{ab})\cdots]. \qquad (3.1)$$

We shall now demonstrate that if $\tilde{F}_{ab}$ is an asymptotically regular generalized Maxwell field then the constraints imposed upon $\mathbf{F}_{ab}$ by Eqs. (1.2) and (1.3) are the same as those which arise from Maxwell's equations.

To begin with let us consider Eq. (1.2). Since

$$\nabla_b F_{cd} = \Omega^{-3/2}[\Omega^{1/2}\nabla_b(\Omega F_{cd}) - 2\Omega F_{cd}\nabla_b\Omega^{1/2}], \qquad (3.2)$$

we see that Eqs. (1.2) and (2.8) imply that on $\tilde{M}$

$$\begin{aligned}&\Omega^{1/2}\nabla_b(\Omega F^{ab}) - 2\Omega F^{ab}\nabla_b\Omega^{1/2}\\ &\quad + \tfrac{1}{2}k\Omega[\Omega^{1/2}\nabla_b(\Omega F_{cd}) + 2\Omega F_{cd}\nabla_b\Omega^{1/2}\\ &\quad + 4\Omega F_{bd}\nabla_c\Omega^{1/2} - 4g_{bc}\Omega F_{ed}\nabla^e\Omega^{1/2}]\\ &\quad \times[\Omega*R*^{abcd} - g^{ae}g^{bf}\delta^{cdh}_{efl}\nabla^l\nabla_h\Omega\\ &\quad + 4(\nabla_m\Omega^{1/2})(\nabla^m\Omega^{1/2})g^{ae}g^{bf}\delta^{cd}_{ef}] = 0.\end{aligned}$$

Upon taking the limit of the above equation as we approach i^0, noting Eq. (3.1), along with the fact that $\Omega^{1/2}\mathbf{R}_{abcd}$ is a direction-dependent tensor at i^0, and $\lim_{\to i^0}\nabla^a\Omega^{1/2} = \eta^a$ (the unit tangent vector at i^0 to the spacelike curve used to evaluate the limit), we find that

$$\partial_b\mathbf{F}^{ab}(\eta) - 2\mathbf{F}^{ab}(\eta)\eta_b = 0. \qquad (3.3)$$

A similar argument applied to Eq. (1.3) (expressed in the form $\tilde{\nabla}_b*\tilde{F}^{ab} = 0$) shows that

$$\partial_b*\mathbf{F}^{ab}(\eta) - 2*\mathbf{F}^{ab}(\eta)\eta_b = 0. \qquad (3.4)$$

Equations (3.3) and (3.4) are identical to the equations which $\mathbf{F}_{ab}$ must satisfy on $\mathscr{H}$ when $\tilde{F}_{ab}$ is an asymptotically regular Maxwell field. Consequently, when $\tilde{F}_{ab}$ is asymptotically regular, the first order asymptotic behavior of the electromagnetic field at spatial infinity is governed by the same set of equations in the Einstein–Maxwell and generalized Einstein–Maxwell theories. In particular this implies that both theories can use the same expression for the total electric and magnetic charge of $(\tilde{M},\tilde{g}_{ab})$ (see p. 1555 of Ref. 5).

We shall now study the behavior of the energy–momentum tensor $\tilde{\mathscr{T}}_a{}^b$ at i^0 for an asymptotically regular generalized Maxwell field.

An examination of Eqs. (2.2)–(2.4) shows that due to Eq. (3.2)

$$\tilde{T}_a{}^b = \Omega^2\hat{T}_a{}^b, \quad \tilde{B}_a{}^b = \Omega^3\hat{B}_a{}^b, \quad \text{and} \quad \tilde{C}_a{}^b = \Omega^3\hat{C}_a{}^b, \qquad (3.5)$$

where $\hat{T}_a{}^b$, $\hat{B}_a{}^b$, and $\hat{C}_a{}^b$ admit direction-dependent limits at i^0. Hence, $\tilde{\mathscr{T}}_a{}^b = \tilde{T}_a{}^b + k(\tilde{B}_a{}^b + \tilde{C}_a{}^b)$, is dominated by $\tilde{T}_a{}^b$ as we approach spatial infinity. This observation appears to be a bit surprising in view of our previous work with $\tilde{\mathscr{T}}_a{}^b$ near null infinity, but it is in keeping with the behavior of $\tilde{\mathscr{T}}_a{}^b$ in Minkowski space discussed in the introduction. In order to explain this behavior, recall that the difference between the pullbacks of $\Omega^{-4}\tilde{\mathscr{T}}_a{}^c g_{cb}$ and $\Omega^{-4}\tilde{T}_a{}^c g_{cb}$ to I (in the case of null infinity) is due to the term presented in Eq. (2.10), which vanishes if there is no radiation incident upon null infinity. Thus the essential difference between the behavior of $\tilde{\mathscr{T}}_a{}^b$ at null and spatial infinity can be attributed to the fact that radiation fields can impinge upon null infinity, while they cannot reach spatial infinity.

The equations governing the gravitational field at spatial infinity in the generalized Einstein–Maxwell theory can be derived from the identity (see Eq. (13) of Ref. 4)

$$\nabla_{[a}(\Omega^{-1}C_{bc]}{}^{de}) = 2\Omega^{-2}\delta^{[d}_{[a}\nabla_b(\Omega^{-1}L_{c]}{}^{e]}) - 2\Omega^{-4}\delta^{[d}_{[a}\delta^{e]}_{b}L_{c]f}\nabla^f\Omega, \quad (3.6)$$

where $L_a{}^b$ is given by Eq. (2.5). Due to Eq. (3.5), we have $L_{ab} = \Omega^2\Psi_{ab}$, where Ψ_{ab} admits a regular direction-dependent limit at i^0, and therefore

$$\lim_{\to i^0}\left[2\delta^{[d}_{[a}\nabla_b(\Omega^{-1}L_{c]}{}^{e]}) - 2\Omega^{-2}\delta^{[d}_{[a}\delta^{e]}_{b}L_{c]f}\nabla^f\Omega\right] = 0. \quad (3.7)$$

Consequently we may employ Eqs. (3.6) and (3.7) to conclude that the field equations governing the asymptotic gravitational fields are the usual ones,[16] viz.,

$$\partial_{[a}\mathbf{C}_{bc]\,de}(\eta) = 3\eta_{[a}\mathbf{C}_{bc]\,de}(\eta),$$

on $\mathscr{H}$, where $\mathbf{C}_{bcde}(\eta) := \lim_{\to i^0}\Omega^{1/2}C_{bcde}$. Thus the presence of the generalized Maxwell field has no effect upon the first order behavior of the asymptotic gravitational field at spatial infinity.

In summary we see that as far as the properties of fields at spatial infinity is concerned, there is absolutely no difference between the Einstein-Maxwell and generalized Einstein–Maxwell field theories to first order.

ACKNOWLEDGMENTS

I would like to thank A. Ashtekhar, R. Geroch, and J. Wainwright for many interesting and informative discussions on the topics considered in this paper. I also wish to thank the Natural Sciences and Engineering Research Council of Canada for financial support. Lastly I would like to thank the relativity group of the University of Chicago for their hospitality during my visit from September 1979 to May 1980.

[1]For a discussion of the origin of the generalized Einstein–Maxwell field theory see G. W. Horndeski, J. Math. Phys. **17**, 1980 (1976).

[2]My notational conventions are the same as those employed in C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973) with the following exceptions: (i) tensorial indices in a 4-space will be denoted by lower case Latin letters; (ii) tensorial indices in a 3-space will be denoted by lower case Greek letters; and (iii) covariant differentiation with respect to $\tilde{g}_{ab}$ and g_{ab} will be denoted by $\tilde{\nabla}$ and ∇, respectively.

[3]G. W. Horndeski and J. Wainwright, Phys. Rev. D **16**, 1691 (1977).

[4]R. Geroch, "Asymptotic Structure of Space-Time," article in *Asymptotic Structure of Space-Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977), pp. 1–105.

[5]A. Ashtekar and R. O. Hansen, J. Math. Phys. **19**, 1542 (1978).

[6]The definition of an asymptote given here differs slightly from the one given on p. 8 of Ref. 4 in that we are not requiring $d\Omega$ to be a null covector field on I. However, we shall prove that this must be the case if various other conditions are satisfied.

[7]See, e.g., Theorem 8 on p. 39 of Ref. 4 and the ensuing discussion, or Sec. 14 of R. Penrose, Proc. R. Soc. Lond. Ser. A **284**, 159 (1965).

[8]G. W. Horndeski, J. Math. Phys. **20**, 1745 (1979).

[9]G. W. Horndeski, Phys. Rev. D **17**, 391 (1978).

[10]G. W. Horndeski, J. Math. Phys. **19**, 668 (1978).

[11]L. P. Eisenhart, *Riemannian Geometry* (Princeton, U.P., Princeton, N.J., 1965). Note that Eisenhart's Ricci tensor and Ricci scalar are the negative of ours.

[12]It is of interest to note that although $\Omega^{-4}\tilde{\mathscr{T}}_a{}^a \neq 0$ (in general) it does vanish in the limit as we approach I. Thus $\lim_{\to I}\Omega^{-4}\tilde{\mathscr{T}}_a{}^b$ is trace-free.

[13]See, e.g., pp. 51 and 63 of Ref. 4 for the definition of the News tensor and the Bondi energy–momentum vector. Strictly speaking, when discussing these quantities we must require the asymptote to be Minkowskian.

[14]For the definition of a $C^{>1}$ manifold structure at i^0, a $C^{>k}$ tensor field at i^0, and other terminology used in this section, see Appendix A of Ref. 5.

[15]Throughout this section when I refer to $\tilde{F}_{ab}$ as being asymptotically regular it will be in the sense of the present definition.

[16]This equation is the same as Eq. (11) of Ref. 5 due to the identity

$$\eta_{[m}\mathbf{C}_{ab\,]}{}^{cd} = 2\delta^{[d}_{[m}\mathbf{C}_{ab\,]}{}^{c]r}\eta_r,$$

which results from the fact that $0 = \delta^{cdrst}_{mabuv}\eta_r\mathbf{C}_{st}{}^{uv}$ in a 4-space. (Note that the coefficient of the first term on the right-hand side of Eq. (11) in Ref. 5 should be a 2.)

Gravitational and electromagnetic potentials of the stationary Einstein–Maxwell field equations

Thaddeus C. Jones [a)]
Physics Department, Montana State University, Bozeman, Montana 59717

(Received 2 March 1979; accepted for publication 20 August 1979)

We develop the necessary equations for the production of an infinite hierarchy of potentials for stationary Einstein–Maxwell spacetimes. Specific knowledge of these potentials are essential to the Kinnersley–Chitre method of solution generation. To elucidate their meaning we investigate the particular case of Reissner–Nordstrom.

1. INTRODUCTION

The discovery of various symmetries in nature has been instrumental in reinforcing the essential belief of the simplicity and unity of the universe. Thus, when it was determined that the inclusion of the Maxwell equations into the stationary Einstein equations not only maintained the original symmetry group but actually increased it, one had to sense that this could not be merely fortuitous.

In a remarkable series of papers,[1–4] Kinnersley and Chitre have explored some of the inherent ramifications of these symmetries. In particular, they have found a representation of the infinite parameter group of the Einstein–Maxwell equations through an infinite hierarchy of potentials. Recently, as a reward for Kinnersley's persistent faith that a perusal of these symmetries must lead to a method of solution generation, they have developed a method of generating new stationary solutions which are *guaranteed* to be *asymptotically flat*!

Instructive examples in the aforementioned papers have tended to relate to the details of the potentials for vacuum solutions. In this paper we will develop the pertinent functions which will enable us to deal with the complete set of Einstein–Maxwell equations and related potentials. In Sec. 2 we will derive the general form for the generating functions of the electromagnetic and gravitational potentials. Section 3 will be devoted to an examination of the Reissner–Nordstrom metric in order to consider the details of a specific example. In Sec. 4 the problems of gauge freedom and their relation to solution generation will be considered. Sections 5 and 6 will be devoted to specific results for the Reissner–Nordstrom metric, one of which will be consistent with the Harrison transformation.[5]

2. GENERATING FUNCTIONS FOR THE POTENTIALS

From paper II of the series we note the fundamental equations for the electromagnetic and gravitational potentials $\phi_A^{(n)}$ and $H_{AB}^{(n)}$:

$$\nabla\phi_A^{(n)} = i\rho^{-1} f_A{}^X \nabla\phi_X^{(n)}, \tag{2.1}$$

$$\nabla H_{AB}^{(n)} = -i\rho^{-1} f_A{}^X \tilde{\nabla} H_{XB}^{(n)}. \tag{2.2}$$

The recursive definitions were given as

$$\phi_A^{(n+1)} = i(M_A^{(1n)} + 2\phi_A K^{(1n)} + H_{AX}\phi^{(n)X}), \tag{2.3}$$

$$H_{AB}^{(n+1)} = i(N_{AB}^{(1n)} + 2\phi_A L_B^{(1n)} + H_{AX} H_B^{(n)X}), \tag{2.4}$$

where

$$\nabla^{(mn)}K = \phi_X^{(m)*}\nabla\phi^{(n)X}, \tag{2.5}$$

$$\nabla^{mn}L_B = \phi_X^{(m)*}\nabla H_B^{(n)X}, \tag{2.6}$$

$$\nabla^{mn}M_A = H_{XA}^{(m)*}\nabla\phi^{(n)X}, \tag{2.7}$$

$$\nabla^{mn}N_{AB} = H_{XA}^{(m)*}\nabla H_B^{(n)X}. \tag{2.8}$$

In paper III a generating function for the $H_{AB}^{(n)}$ was derived for a vacuum spacetime. We will now derive the equations for the general generating functions of $H_{AB}^{(n)}$ and $\phi_A^{(n)}$. Since $H_{AB}^{(n)}$ and $\phi_A^{(n)}$ are coupled, four generating functions are required:

$$Q(t) \equiv \sum_{n=0}^{\infty} t^n K^{(1n)}, \tag{2.9}$$

$$S_A(t) \equiv \sum_{n=0}^{\infty} t^n L_A^{(1n)}, \tag{2.10}$$

$$F_{AB}(t) \equiv \sum_{n=0}^{\infty} t^n H_{AB}^{(n)}, \tag{2.11}$$

$$D_A(t) \equiv \sum_{n=0}^{\infty} t^n \phi_A^{(n)}. \tag{2.12}$$

Differentiate Eqs. (2.3) and (2.4), multiply both sides of each equation by t^{n+1}, and sum. The result is

$$\nabla F_{AB} = it\{[H^*_{XA} + H_{AX} + 2\phi_A\phi^*_X]\nabla F^X{}_B + 2S_B\nabla\phi_A + F^X{}_B\nabla H_{AX}\}, \tag{2.13}$$

$$\nabla D_A = it\{[H^*_{XA} + H_{AX} + 2\phi_A\phi^*_X]\nabla D^X + 2Q\nabla\phi_A + D^X\nabla H_{AX}\}. \tag{2.14}$$

Regrouping the terms and defining

$$G^A{}_X \equiv \delta^A{}_X - it[H^{*A}_X + H^A{}_X + 2\phi^A\phi^*_X] = \delta^A{}_X[1 - 2tz] - 2it f^A{}_X, \tag{2.15}$$

we obtain

$$G^A{}_X\nabla F^X{}_B - itF^X{}_B\nabla H^A{}_X = 2it S_B\nabla\phi^A, \tag{2.16}$$

$$G^A{}_X\nabla D^X - it D^X\nabla H^A{}_X = 2it Q\nabla\phi^A. \tag{2.17}$$

It will be convenient to decouple these equations as far as possible. To initiate this result we abandon covariance and use the generalizations of Eqs. (2.1) and (2.2) which are

$$\nabla F_{AB} = -i\rho^{-1} f_A{}^X\tilde{\nabla}F_{XB}, \tag{2.18}$$

[a)]Former address. Now affiliated with the Shell Oil Company, Pacific Division Exploration, Entex Room 2244, Box 527, Houston, Texas 77079.

 0022-2488/80/071790-08$1.00

$$\nabla D_A = -i\rho^{-1} f_A{}^X \tilde{\nabla} D_X , \tag{2.19}$$

thereby obtaining (setting the index $A = 2$)

$$\frac{G^2{}_1 [\tilde{\nabla} F^2{}_B - i\rho^{-1} f_{12} \nabla F^2{}_B]}{i\rho^{-1} f_{11}} + G^2{}_2 \nabla F^2{}_B + itF^1{}_B \nabla H_{11} + itF^2{}_B \nabla H_{12} = -2itS_B \nabla \phi_1 , \tag{2.20}$$

$$\frac{G^2{}_1 [\tilde{\nabla} D^2 - i\rho^{-1} f_{12} \nabla D^2]}{i\rho^{-1} f_{11}} + G^2{}_2 \nabla D^2 + itD^1 \nabla H_{11} + itD^2 \nabla H_{12} = -2it\, Q \nabla \phi_1 . \tag{2.21}$$

Using Eq. (2.15) in Eqs. (2.20) and (2.21), we obtain

$$(1-2tz)\nabla F^2{}_B + 2t\rho \tilde{\nabla} F^2{}_B + it\, F^1{}_B \nabla H_{11} + it\, F^2{}_B \nabla H_{12} = -2it\, S_B \nabla \phi_1 , \tag{2.22}$$

$$(1-2tz)\nabla D^2 + 2t\rho \tilde{\nabla} D^2 + itD^1 \nabla H_{11} + itD^2 \nabla H_{12} = -2it\, Q \nabla \phi_1 . \tag{2.23}$$

Choose $B = 1$ in order to focus on F_{11} $[D_2]$:

$$-(1-2tz)\nabla F_{11} - 2t\rho \tilde{\nabla} F_{11} + itF_{21} \nabla H_{11} - itF_{11} \nabla H_{12} = -2itS_1 \nabla \phi_1 , \tag{2.24}$$

$$-(1-2tz)\nabla D_1 - 2t\rho \tilde{\nabla} D_1 + itD_2 \nabla H_{11} - itD_1 \nabla H_{12} = -2it\, Q \nabla \phi_1 . \tag{2.25}$$

We have two choices: Eliminate F_{21} $[D_2]$ algebraically and obtain

$$F_{11,x}[(1-2tz)H_{11,y} + 2t(x^2-1)H_{11,x}] + F_{11,y}[2t(1-y^2)H_{11,y} - (1-2tz)H_{11,x}] = itF_{11}[H_{11,x}H_{12,y} - H_{11,y}H_{12,x}] + 2itS_1[\phi_{1,x}H_{11,y} - \phi_{1,y}H_{11,x}], \tag{2.26}$$

$$D_{1,x}[(1-2tz)H_{11,y} + 2t(x^2-1)H_{11,x}] + D_{1,y}[2t(1-y^2)H_{11,y} - (1-2tz)H_{11,x}] = itD_1[H_{11,x}H_{12,y} - H_{11,y}H_{12,x}] + 2itQ[\phi_{1,x}H_{11,y} - \phi_{1,y}H_{11,x}] \tag{2.27}$$

or eliminates S_1 $[Q]$ and see that

$$F_{11,x}[2t(x^2-1)\phi_{1,x} + (1-2tz)\phi_{1,y}] + F_{11,y}[2t(1-y^2)\phi_{1,y} - (1-2tz)\phi_{1,x}] = itF_{11}[\phi_{1,x}H_{12,y} - \phi_{1,y}H_{12,x}] + itF_{21}[\phi_{1,y}H_{11,x} - \phi_{1,x}H_{11,y}], \tag{2.28}$$

$$D_{1,x}[2t(x^2-1)\phi_{1,x} + (1-2tz)\phi_{1,y}] + D_{1,y}[2t(1-y^2)\phi_{1,y} - (1-2tz)\phi_{1,x}] = itD_1[\phi_{1,x}H_{12,y} - \phi_{1,y}H_{12,x}] + itD_2[\phi_{1,y}H_{11,x} - \phi_{1,x}H_{11,y}] . \tag{2.29}$$

Similar equations follow for the other F_{AB} and D_A. Note that, given F_{1B} $[D_1]$, one may produce F_{2B} $[D_2]$ using Eqs. (2.18) and (2.19).

Consider the above equations for F_{1B} $[D_1]$; either they are coupled to S_B $[Q]$ or to F_{2B} $[D_2]$. This dilemma is troublesome in general, but fortunately the coupling term vanishes for cases where H_{11} and ϕ_1 (Ernst's ϵ and ϕ)[6] are functionally dependent, e.g., Reissner–Nordstrom, Kerr–Newman, and charged Tomimatsu–Sato solutions. In order to see this, consider the coupling term

$$\phi_{1,x}H_{11,y} - \phi_{1,y}H_{11,x} \propto \tilde{\nabla} H_{11} \cdot \nabla \phi . \tag{2.30}$$

Now, for the above cases,

$$H_{11} = \epsilon = f - |\Phi|^2 + i\psi = \frac{\xi - 1}{\xi + 1} , \tag{2.31}$$

$$\phi_1 = \Phi = \frac{q}{\xi + 1} , \tag{2.32}$$

so

$$\tilde{\nabla} H_{11} \cdot \nabla \phi_1 \propto \tilde{\nabla} \xi \cdot \nabla \xi = 0 \tag{2.33}$$

since

$$\tilde{\nabla} u \cdot \nabla u \equiv 0 \tag{2.34}$$

for any potential u. Ergo we are left with an equation involving only F_{1B} $[D_1]$ to solve in the instances of greatest physical interest.

3. REISSNER–NORDSTROM POTENTIALS

Referring back to Eqs. (2.30) and (2.31), we note that Ernst[7] has demonstrated that if ξ_0 is a vacuum solution to Einstein's equations, then

$$\xi = \xi_0 (1 - q^2)^{1/2} , \tag{3.1}$$

$$q \equiv e/m \tag{3.2}$$

is a solution to the Einstein–Maxwell equations. Specifically, using prolate spheroidal coordinates, $\xi_0 = x$ represents the Schwarzschild solution and

$$\xi = x(1 - q^2)^{1/2} \tag{3.3}$$

represents the Reissner–Nordstrom solution.

Using prolate spheroidal coordinates, where

$$x = \frac{r - m}{k},$$

$$y = \cos\theta, \tag{3.4}$$

$$k = (m^2 - e^2)^{1/2} ,$$

where k has been chosen so as to given the familiar result of e/r for the electromagnetic potential ϕ_1.

Defining

$$\beta \equiv (1 - q^2)^{-1/2} = mk^{-1} , \tag{3.5}$$

then for the Reissner–Nordstrom metric

$$f = (x^2 - 1)(x + \beta)^{-2} , \tag{3.6}$$

$$\psi = 0. \tag{3.7}$$

Using the equations in Sec. I of paper II, we discover

$$H_{AB} = [f_{AB} - A_A A_B - B_A B_B + \epsilon_{AB} K] + i[\psi_{AB} + A_A B_B + B_A A_B] , \tag{3.8}$$

where

$$\tilde{\nabla} B_A = \rho^{-1} f_A{}^X \nabla A_X . \tag{3.9}$$

Then, specifically, we derive the following:

$$A_A = \begin{pmatrix} \frac{q\beta}{x + \beta} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{e}{r} \\ 0 \end{pmatrix}, \tag{3.10}$$

$$B_A = \begin{pmatrix} 0 \\ q\beta y \end{pmatrix}, \tag{3.11}$$

$$\nabla K = i\nabla(A_1 B_2) \rightarrow K = \frac{iq^2\beta^2 y}{x+\beta}, \tag{3.12}$$

$$\psi_{AB} = \begin{pmatrix} 0 & -2\beta y \\ 2y(x-\beta) & 0 \end{pmatrix}, \tag{3.13}$$

$$H_{AB} = \begin{pmatrix} \dfrac{x-\beta}{x+\beta} & \dfrac{2iy(x^2-1)}{x+\beta} \\ -2i\beta y & (x+\beta)^2(y^2-1) - y^2q^2\beta^2 \end{pmatrix} \tag{3.14}$$

Next, we apply the method of characteristics to Eq. (2.26) and find

$$\frac{F_{11}^{-1}dF_{11}}{it\,[H_{12,y}H_{11,x} - H_{12,x}H_{11,y}]} = \frac{dx}{(1-2tz)H_{11,y} + 2t(x^2-1)H_{11,x}} = \frac{dy}{2t(1-y^2)H_{11,y} - (1-2tz)H_{11,x}} \tag{3.15}$$

Using Eqs. (3.13) and (3.14) for the particular example of the Reissner–Nordstrom metric, we obtain

$$\frac{dy}{dx} = \frac{(-1+2txy)}{2t(x^2-1)}, \tag{3.16}$$

$$\frac{dF_{11}}{F_{11}} = \frac{-dx}{x+\beta}. \tag{3.17}$$

The solutions are

$$a_1 = (2ty - x)(x^2-1)^{-1/2}, \tag{3.18}$$

$$F_{11} = \frac{a_2}{x+\beta}, \tag{3.19}$$

where a_1 and a_2 are constant along each characteristic curve. The general solution is given by

$$F_{11} = (x+\beta)^{-1}\gamma[(2ty-x)(x^2-1)^{-1/2}], \tag{3.20}$$

where γ is an arbitrary function of $(2ty-x)(x^2-1)^{-1/2}$. In order to determine the particular form of γ, use Eq. (2.18). Rewriting Eq. (2.18), we have

$$\tilde{\nabla} F_{AB} = i\rho^{-1} f_A{}^X \nabla F_{XB}. \tag{3.21}$$

We will use the vector identity $\nabla_2 \cdot \tilde{\nabla} u \equiv 0$, where u is any potential. Taking advantage of this gives

$$\nabla_2 \cdot (\rho^{-1} f_A{}^X \nabla F_{XB}) = 0, \tag{3.22}$$

$$\nabla_2 \cdot (\rho^{-1} f_A{}^X \nabla D_X) = 0. \tag{3.23}$$

Inserting Eq. (3.20) into (3.22), we obtain

$$F_{11} = t\,[\bar{K}(t)[x-2ty]S^{-1} + \bar{K}_1(t)](x+\beta)^{-1}, \tag{3.24}$$

where

$$S^2 \equiv 1 - 4txy + 4t^2(x^2+y^2-1). \tag{3.25}$$

Notice that $\bar{K}$ and $\bar{K}_1$ are functions of the expansion parameter t. Their particular form is to be determined by the specific choice of gauge imposed.

F_{12} $[D_1]$ satisfies the same characteristic equations and Eqs. (3.22) and (3.23) as did F_{11}, so

$$F_{12} = i[C(t)[x-2ty]S^{-1} + C_1(t)](x+\beta)^{-1}, \tag{3.26}$$

$$D_1 = t\,[E(t)[x-2ty]S^{-1} + E_1(t)](x+\beta)^{-1}. \tag{3.27}$$

Notice that F_{11} has an initial multiplier of t where F_{12} does not. This is to place them into conformity with the notation that $H^{(0)}_{AB} \equiv i\epsilon_{AB}$ so $F_{11} = 0 + tH^{(1)}_{11} + \ldots$, while $F_{12} = i + tH^{(1)}_{12} + \ldots$.

Using Eqs. (3.21), (3.24), (3.26), and (3.27) to determine F_{2B} $[D_2]$, we obtain

$$F_{21} = -\frac{i}{2}\{\bar{K}(t)[1-2ty(x-\beta) - 4t^2(x\beta+1-y^2)]S^{-1} - [2K_2(t) + 2K_1(t)ty]\}, \tag{3.28}$$

$$F_{22} = \frac{1}{2t}\{C(t)[1-2ty(x-\beta) - 4t^2(x\beta+1-y^2)]S^{-1} - [C_2(t) + 2C_1(t)t\beta y]\}, \tag{3.29}$$

$$D_2 = \frac{-i}{2}\{E(t)[1-2ty(x-\beta) - 4t^2(x\beta+1-y^2)]S^{-1} - [2E_2(t) + 2E_1(t)y]\}, \tag{3.30}$$

where $K_2(t)$, $C_2(t)$, and $E_2(t)$ are additional integration constants.

Excluding particular gauge conditions, we have now obtained the specific forms of the generating functions for the potentials $\phi^{(n)}_A$ and $H^{(n)}_{AB}$ of the Reissner–Nordstrom metric.

4. GAUGE CONDITIONS

Since the main application of this work lies in the area of solution generation, our foremost concern should be that the gauge conditions we choose are compatible with those used in equations which played essential roles in developing the method of solution generation.

In fact, Eqs. (3.14)–(3.20) of paper II were so used, and the associated selection of integration constants imply certain gauge conditions. We will now derive the specific relationships among the electromagnetic and gravitational potentials which are in agreement with those gauge conditions.

The equations in question are

$$K^{(mn)} - K^{(nm)*} = \phi^{(m)*}_X \phi^{(n)X}, \tag{4.1}$$

$$L^{(mn)}_B - M^{(nm)*}_B = \phi^{(m)*}_X H^X_B, \tag{4.2}$$

$$N^{(mn)}_{AB} - N^{(nm)*}_{BA} = H^{(m)*}_{XA} H^{(n)X}_B + \epsilon_{AB}\delta^m_0\delta^n_0, \tag{4.3}$$

$$K^{(m,n+1)} - K^{(m+1,n)} = 2iK^{(m1)}K^{(1n)} + iL_X^{(m1)}\phi^{(n)X} + \frac{i}{2}\delta_0^m\delta_0^n\,, \tag{4.4}$$

$$L_B^{(m,n+1)} - L_B^{(m+1,n)} = 2iK^{(m1)}L_B^{(1n)} + iL_X^{(m1)}H_B^{(n)X}\,, \tag{4.5}$$

$$M_A^{(m,n+1)} - M_A^{(m+1,n)} = 2iM_A^{(m1)}K^{(1n)} + iN_{AX}^{(m1)}\phi^{(n)X}\,, \tag{4.6}$$

$$N_{AB}^{(m,n+1)} - N_{AB}^{(m+1,n)} = 2iM_A^{(m1)}L_B^{(1n)} + iN_{AX}^{(m1)}H_B^{(n)X} - 3\epsilon_{AB}\delta^m{}_{-1}\delta_0^n\,. \tag{4.7}$$

Again, the most efficient procedure to ensure compliance with the above will be by means of a series of generating functions, i.e.,

$$\begin{aligned}
&K \equiv \sum K^{(mn)}r^m t^n, \quad {}'K \equiv \sum K^{(1n)}t^n = Q, \quad K' \equiv \sum K^{(m1)}r^m\,,\\
&L_A \equiv \sum L_A^{(mn)}r^m t^n, \quad {}'L_A \equiv \sum L_A^{(1n)}t^n = S_A, \quad L'_A \equiv \sum L_A^{(m1)}r^m\,,\\
&M_A \equiv \sum M_A^{(mn)}r^m t^n, \quad {}'M_A \equiv \sum M_A^{(1n)}t^n, \quad M'_A \equiv \sum M_A^{(m1)}r^m\,,\\
&N_{AB} \equiv \sum N_{AB}^{(mn)}r^m t^n, \quad {}'N_{AB} \equiv \sum N_{AB}^{(1n)}t^n, \quad N'_{AB} \equiv \sum N_{AB}^{(m1)}r^m\,.
\end{aligned} \tag{4.8}$$

Multiply Eqs. (4.1)–(4.3) by $r^m t^n$ and then sum. We acquire the following:

$$K - K^* = D_X^*(r)D^X(t)\,, \tag{4.9}$$

$$L_A - M_A^* = D_X^*(r)F^X{}_A(t)\,, \tag{4.10}$$

$$N_{AB} - N_{BA}^* = F_{XA}^*(r)F^X{}_B(t) + \epsilon_{AB} \tag{4.11}$$

Equations (4.4)–(4.7) yield

$$t^{-1}\left(K + \frac{ir}{2}\right) - r^{-1}\left(K - \frac{it}{2}\right) = 2iK'(r)Q(t) + iL'_X(r)D^X(t) + \tfrac{1}{2}i\,, \tag{4.12}$$

$$t^{-1}L_A - r^{-1}L_A = 2iK'(r)S_A(t) + iL'_X(r)F^X{}_A(t)\,, \tag{4.13}$$

$$t^{-1}M_A - r^{-1}[M_A + iD_A(t)] = 2iM'_A(r)Q(t) + iN'_{AX}(r)D^X(t)\,, \tag{4.14}$$

$$t^{-1}(N_{AB} - \epsilon_{AB}) - r^{-1}[N_{AB} + iF_{AB}(t)] = 2iM'_A(r)S_B(t) + iN'_{AX}(r)F^X{}_B(t)\,. \tag{4.15}$$

The recursive definitions given by Eqs. (2.3) and (2.4) give

$$D_A(t) = it\,[{}'M_A(t) + 2\phi_A Q(t) + H_{AX}D^X(t)]\,, \tag{4.16}$$

$$F_{AB}(t) - i\epsilon_{AB} = it\,[{}'N_{AB}(t) + 2\phi_A S_B(t) + H_{AX}F^X{}_B(t)]. \tag{4.17}$$

We now wish to eliminate unwanted primed quantities, e.g., ${}'N_{AB}$ and N'_{AB}. Therefore, in Eq. (4.1), set $m = 1$, multiply by λ^n, and sum.

$$\sum K^{(1n)}\lambda^n - \sum K^{(n1)*}\lambda^n = \sum \phi_X^*\phi^{(n)X}\lambda^n$$

or

$$Q - \sum K^{(m1)}r^m = \phi_X^* D^X\,. \tag{4.18}$$

Thus,

$$Q(t) - K'(t)^* = \phi_X^* D^X(t)\,.$$

In a like manner Eqs. (4.2) and (4.3) produce

$$S_A(t) - M'_A(t) = \phi_X^* F^X{}_A(t)\,, \tag{4.19}$$

$${}'N_{AB}(t) - N'^*_{BA}(t) = H_{XA}^* F^X{}_B(t)\,. \tag{4.20}$$

Setting $n = 1$, multiplying by λ^m, and summing over Eq. (4.2) yields

$$L'_A(r) - {}'M_A^*(r) = D_X^*(r)H^X{}_A\,. \tag{4.21}$$

This was the final bit of new information we were able to extract from Eqs. (4.1)–(4.7).

Using Eqs. (4.15), (4.17), (4.19), and (4.20), we obtain

$$\begin{aligned}
\left(\frac{r-t}{rt}\right)N_{AB} = {}& t^{-1}\epsilon_{AB} + 2iS_B(t)S_A^*(r) - 2iS_B(t)\phi_Z F^{*Z}{}_A(r) - r^{-1}F^X{}_B(t)F_{XA}^*(r)\\
&- 2i\phi_X^* S_A^*(r)F^X{}_B(t) - i(H_{ZX} + H_{XZ}^*)F^X{}_B(t)F^{*Z}{}_A(r)\,.
\end{aligned} \tag{4.22}$$

Equations (4.14), (4.17), (4.19), and (4.20) yield

$$\left(\frac{r-t}{rt}\right)M_A = 2iQ(t)S^*_A(r) - 2iQ(t)\phi_Z F^{*Z}{}_A(r) - 2iS^*_A(r)\phi^*_Z D^Z(t) - r^{-1}D^X(t)F^*_{XA}(r) - i(H_{ZX} + H^*_{XZ})F^{*Z}{}_A(r)D^X(t)\,. \tag{4.23}$$

Equations (4.13), (4.16), (4.18), and (4.21) give

$$\left(\frac{r-t}{rt}\right)L_A = 2iS_A(t)Q^*(r) - 2iS_A(t)\phi_Z D^{*Z}(r) - r^{-1}F^X{}_A(t)D^*_X(r) - 2i\phi^*_X F^X{}_A(t)Q^*(r) - i(H^*_{XZ} + H_{ZX})D^{*Z}(r)F^X{}_A(t)\,. \tag{4.24}$$

Equations (4.12), (4.16), (4.18), and (4.21) produce

$$\left(\frac{r-t}{rt}\right)K = \frac{i}{2}(1 - tr^{-1} - rt^{-1}) + 2iQ(t)Q^*(r) - 2iQ(t)\phi_Z D^{*Z}(r) - 2iQ^*(r)\phi^*_Z D^Z(t) - r^{-1}D^*_Z(r)D^Z(t) - i(H^*_{XZ} + H_{ZX})D^{*Z}(r)D^X(t)\,. \tag{4.25}$$

If we now set $r = t$ on both sides of Eqs. (4.22)–(4.25), there remains only algebraic combinations of the F_{AB}, D_A, Q, and S_A in addition to the known quantities ϕ_A and H_{AB}. Perforce, these results must have the same gauge conditions as the initial equations (4.1)–(4.7) contain.

It must be remarked that Eqs. (4.23) and (4.24) are complex conjugates while the choices of $A = 1$, $B = 2$ and $A = 2$, $B = 1$ result in another complex conjugate pair from Eqs. (4.22).

In conclusion, Eq. (4.22) yields three unique equations, Equation (4.23) [or Eq. (4.24)] yields two distinct equations, and Equation (4.25) gives one. We determine that, while we have a total of six distinct equations, a perusal of Eq. (3.24) and Eqs. (3.26)–(3.30) produces nine functions of the expansion parameter t, namely, $[\bar{K},\bar{K}_1,\bar{K}_2,C,C_1,C_2,E,E_1,E_2]$. In addition, S_A and Q contain three constants of integration; thus, while we have derived six constraint equations, we have a total of 12 functions of the expansion parameter.

Consider Eqs. (6.25) and (6.26) of paper I:

$$\phi_A \to \phi_A + a_A \quad \text{(two gauge functions)}\,, \tag{4.26}$$

$$H_{AC} \to H_{AC} - a^*_A\phi_C - \phi_A a^*_C - a^*_A a_C + i\alpha_{AC} \quad \text{(four gauge functions)}\,. \tag{4.27}$$

So we see that we have exactly the correct number of gauge freedoms in our equations. In order to determine in a complete fashion the specific form of the functions of the expansion parameter, we must choose six additional constraints which are to be fulfilled.

5. REISSNER–NORDSTROM POTENTIALS

An examination of the basic potentials for the Reissner–Nordstrom metric reveals that $S_2, D_1, \phi_1, H_{11}, H_{22}, F_{11}$, and F_{22} are real while $F_{12}, F_{21}, S_1, D_2, \phi_2, H_{12}, H_{21}$, and Q are imaginary. Therefore, our six basic equations reduce to

$$-F_{11}F_{21}t^{-1} - iS_1^2 + 2iS_1\phi_1F_{21} + 2iS_1\phi_2F_{11} = -iH_{11}F_{21}^2 + iH_{22}F_{11}^2 + i(H_{12} - H_{21})F_{11}F_{21}\,, \tag{5.1}$$

$$t^{-1} - F_{11}F_{22}t^{-1} - F_{21}F_{12}t^{-1} - 2iS_1S_2 + 2iS_2\phi_1F_{21} + 2iS_2\phi_2F_{11} + 2i\phi_1S_1F_{22} + 2i\phi_2S_1F_{12} = -i(H_{21} - H_{12})F_{22}F_{11} + i(H_{12} - H_{21})F_{12}F_{21} - 2iH_{11}F_{22}F_{21} + 2iH_{22}F_{12}F_{21}\,, \tag{5.2}$$

$$F_{12}F_{22}t^{-1} + iS_2^2 - 2iS_2\phi_1F_{22} - 2iS_2\phi_2F_{12} = iH_{11}F_{22}^2 - i(H_{12} - H_{21})F_{12}F_{22} - iH_{22}F_{12}^2\,, \tag{5.3}$$

$$F_{12}D_2t^{-1} + F_{22}D_1t^{-1} + 2iQS_2 - 2iQ\phi_1F_{22} - 2iQ\phi_2F_{12} - 2i\phi_1S_2D_2 - 2i\phi_2S_2D_1 = 2iH_{11}F_{22}D_2 - 2iH_{22}F_{12}D_1 - i(H_{12} - H_{21})F_{22}D_1 + i(H_{21} - H_{12})F_{12}D_2\,, \tag{5.4}$$

$$-F_{11}D_2t^{-1} - F_{21}D_1t^{-1} - 2iQS_1 + 2iQ\phi_1F_{21} + 2iQ\phi_2F_{11} + 2i\phi_1S_1D_2 + 2i\phi_2S_1D_1 = -2iH_{11}F_{21}D_2 + 2iH_{22}F_{11}D_1 + i(H_{12} - H_{21})F_{21}D_1 - i(H_{21} - H_{12})F_{11}D_2\,, \tag{5.5}$$

$$-1/4 + iD_1D_2t^{-1} - Q^2 + 2Q\phi_1D_2 + 2Q\phi_2D_1 = -H_{11}D_2^2 + H_{22}D_1^2 + (H_{12} - H_{21})D_1D_2\,. \tag{5.6}$$

Consider the latter equation. We choose to make the D_A as simple as possible and also to require that $\phi_A^{(n)} \to 0$ as $x \to \infty$; therefore, set $E(t) = E_2(t) = 0$. then we find

$$Q = iE_3(t) + itE_1(t)q\beta y(x + \beta)^{-1}, \tag{5.7}$$

$$D_1 = tE_1(t)(x + \beta)^{-1}, \tag{5.8}$$

$$D_2 = iE_1(t)ty\,. \tag{5.9}$$

Using Eq. (5.6), we determine $E_3(t)$ and $E_1(t)$ as

$$E_3(t) = 1/2[1 + 4t^2(\beta^2 - 1)]^{-1/2}, \tag{5.10}$$

$$E_1(t) = -q\beta[1 + 4t^2(\beta^2 - 1)]^{-1/2}\,. \tag{5.11}$$

The particular selection of signs was determined by the fact that $D_A = \phi_A + t\,\phi_A^{(2)}\cdots$. Five equations remain in which to determine the other eight unknowns, consistent of course with the above $E(t)$'s.

In order to exploit the rest of the gauge freedom, a combination of the F_{AB} found advantageous in paper IV for the generation of new solutions, i.e., $F_{A1} + itF_{A2}$, will be made as simple as possible. Equations (3.24), (3.26), (3.28), and (3.29) result in the following:

$$F_{11} + itF_{12} = (\bar{K} - C)\{t(x - 2ty)S^{-1}(x + \beta)^{-1}\} + (\bar{K}_1 - C_1)\{t(x + \beta)^{-1}\}, \tag{5.12}$$

$$F_{21} + itF_{22} = i\left(\bar{K}_2 - \frac{C_2}{2}\right) + ity(\bar{K}_1 - C_1) + \frac{i(C - \bar{K})}{2}S^{-1}\{1 - 2ty(x - \beta) - 4t^2(x\beta + 1 - y^2)\}. \tag{5.13}$$

Setting $\bar{K} = C$ will greatly reduce the complexity of the above. In addition, requiring $C_1 = \beta C$ and $\bar{K}_1 = -\beta\bar{K}$ will ensure a close similarity to the vacuum solutions given in paper IV.

The situation remains much too complicated algebraically until it is noticed that if the equations are correct for all x, y, they must hold in particular for $x = y = 0$. Luckily, this reduction still retains enough information for a complete determination of the remaining unknowns. Acting on this we obtain

$$\begin{aligned}
&\bar{K} = C = C_1\beta^{-1} = -\bar{K}_1\beta^{-1} = [1 - 4t^2]^{-1/2}[1 + 4t^2(\beta^2 - 1)]^{-1/2}\\
&\bar{K}_2 = 1/2[1 + 4t^2(\beta^2 - 1)]^{-1/2} - 1/2[1 + 4t^2(\beta^2 - 1)]^{-1/2}[1 - 4t^2]^{-1/2} - 1/2,\\
&C_2 = [1 + 4t^2(\beta^2 - 1)]^{-1/2} + [1 + 4t^2(\beta^2 - 1)]^{1/2}[1 - 4t^2]^{-1/2} - 1.
\end{aligned} \tag{5.14}$$

Various signs have been determined by noting the first term in the series $F_{AB} = i\epsilon_{AB} + \cdots$.

The above results do not represent the only possible specific representation of the potentials, or even the most interesting one. In fact, one could well wish to examine the results when agreement with the Harrison transformation is enforced.

6. THE HARRISON TRANSFORMATION

In finite form the Harrison transformation is given by Eq. (7.20) in paper I.

$$\Phi = \phi_1 \to \frac{\phi_1 + cH_{11}}{1 - 2c\phi_1 - c^2H_{11}}, \tag{6.1}$$

$$G_{11} = H_{11} \to \frac{H_{11}}{1 - 2c\phi_1 - c^2H_{11}}. \tag{6.2}$$

We wish to maintain the form of the Reissner–Nordstrom metric after the Harrison transformation is performed, e.g.,

$$f \to f' : (r^2 - 2mr + e^2)r^{-2} \to (r'^2 - 2m'r' + e'^2)(r')^2.$$

Let us define the following:

$$m' \equiv [r(1 - c^2) - 2ce][1 - c^2]^{-1}, \tag{6.3}$$

$$e' \equiv [e(1 + c^2) - 2mc][1 - c^2]^{-1}, \tag{6.4}$$

$$r' \equiv [r(1 - c^2) - 2ce - 2mc^2][1 - c^2]^{-1}. \tag{6.5}$$

Then

$$r^2 - 2mr + e^2 = r'^2 - 2m'r' + e'^2 \tag{6.6}$$

and

$$f \to f' = [r'^2 - 2m'r' + e'^2][(1 - c^2)r']^{-2}, \tag{6.7}$$

$$\phi_1 \to \phi'_1 = [c + e'/r'][1 - c^2]^{-1}, \tag{6.8}$$

$$H_{11} \to H'_{11} = 1 - c^2 - \frac{2m'}{r'}\frac{[1 - c^2]}{1 + c^2}. \tag{6.9}$$

Referring back to Sec. 3, we see that for the Reissner–Nordstrom metric the results should have the form

$$\begin{aligned}
&f' = [r'^2 - 2m'r'^2 + e'^2]r'^{-2},\\
&\phi'_1 = \frac{e'}{r'},\\
&H_{11} = \frac{(x' - \beta')}{(x' + \beta')} = 1 - \frac{2m'}{r'}.
\end{aligned} \tag{6.10}$$

We must, in general, be concerned with various changes and scale and gauge transformations if we are to achieve the appropriate final forms. We must also keep in mind that the Harrison transformation does not commute with the gauge transformations.

We will now elucidate the situation via an example:

$$H_{11} = 1 - \frac{2m}{r} = \frac{xk - m}{xk + m}, \tag{6.11}$$

$$\phi_1 = \frac{e}{r} = e(xk + m)^{-1}. \tag{6.12}$$

The above are potentials for the Reissner–Nordstrom metric. They reduce to the potentials for the Schwarzschild solution if we set $e = 0$.

Using Eqs. (6.3)–(6.5) and prolate spheroidal coordinates, we obtain

$$m^2 - e^2 = m'^2 - e'^2, \tag{6.13}$$

$$r - m = r' - m', \tag{6.14}$$

$$x = x', \tag{6.15}$$

$$k = k', \tag{6.16}$$

$$y = y'. \tag{6.17}$$

So, under the Harrison transformation

$$H_{11} \to \left[x'k' - \frac{(1 - c^2)}{1 + c^2}m'\right][(1 - c^2)(x'k' + m')]^{-1}, \tag{6.18}$$

$$\phi_1 \to \left[c + \frac{e'}{x'k' + m'}\right][1 - c^2]^{-1}, \tag{6.19}$$

applying the appropriate scale changes, we have

$$H_{11} \to [1 - c^2]\left[x'k' - \frac{(1 - c^2)}{1 + c^2}m'\right][x'k' + m']^{-1}, \tag{6.20}$$

$$\phi_1 \to c + \frac{e'}{x'k' + m'}. \tag{6.21}$$

Referring back to Eq. (4.26), we note that if

$$a_1 = -c, \tag{6.22}$$

then

$$\phi_1 \to \frac{e'}{x'k' + m'} = \frac{e'}{r'}. \tag{6.23}$$

This determines the necessary gauge transformation. From Eq. (4.27) we have

$$H_{11} \to H_{11} + 2c\phi_1 - c^2, \tag{6.24}$$

so

$$\begin{aligned} H_{11} \to{}& (1-c^2)\left[x'k' - \frac{(1-c^2)}{1+c^2}m'\right][x'k' + m']^{-1} \\ &+ 2c\left[c + \frac{e'}{x'k' + m'}\right] - c^2 \\ ={}& \frac{x'k' - m'}{x'k' + m'} = \frac{x' - \beta'}{x' + \beta'}. \end{aligned} \tag{6.25}$$

This is the result obtained in Eq. (3.14).

If we refer back to Sec. 3 of paper II, we find the infinitesimal transformations of the potentials for which the Harrison transformation is but a specific case. We may exponentiate these, as outlined in paper III, to acquire the finite form of the action of the Harrison transformation on the $H_{AB}^{(n)}$ and $\phi_A^{(n)}$. In addition, we may also procure finite forms of the gauge action. As a caveat we note that the $c_A^{(n)}$ infinitesimal transformation should be summed $s = 1,\ldots,k$ on the *first* sum and $s = 1,\ldots,k-1$ on the *second* sum. This avoids producing any extra linear terms.

We are now in a position to apply the Harrison transformation to the Schwarzschild potentials and acquire the results of Sec. 2.

Setting $e = 0$ (or equivalently $\beta = 1$) in Sec. 5, we have

$$F_{11} = \frac{t\,[(x - 2ty)S^{-1} - 1]}{(x+1)[1 - 4t^2]^{1/2}}, \tag{6.26}$$

$$F_{12} = \frac{i[(x - 2ty)S^{-1} + 1]}{(x+1)[1 - 4t^2]^{1/2}}. \tag{6.27}$$

Since $e = 0$, we may write the Harrison transformation for $H_{11}^{(n)}$ and $\phi_1^{(n)}$ as

$$H_{11}^{(n)} \to \frac{H_{11}^{(n)}}{1 - c^2 H_{11}}, \tag{6.28}$$

$$\phi_1^{(n)} \to \frac{cH_{11}^{(n)}}{1 - c^2 H_{11}}; \tag{6.29}$$

therefore,

$$F_{11} \to \frac{F_{11}}{1 - c^2 H_{11}}, \tag{6.30}$$

$$D_1 \to \frac{cF_{11}}{1 - c^2 H_{11}}. \tag{6.31}$$

Now

$$H_{12}^{(n)} \to \frac{H_{12}^{(n)} - ic^2 H_{11}^{(n+1)}}{1 - c^2 H_{11}} \tag{6.32}$$

so

$$\sum t^n H_{12}^{(n)} \to \frac{\sum t^n H_{12}^{(n)} - (ic^2/t)\sum H_{11}^{(n+1)} t^{n+1}}{1 - c^2 H_{11}} \tag{6.33}$$

or

$$F_{12} \to \frac{F_{12} - ic^2 t^{-1} F_{11}}{1 - c^2 H_{11}}. \tag{6.34}$$

The gauge transformation for $H_{11}^{(n+1)}$ is

$$H_{11}^{(n+1)} \to H_{11}^{(n+1)} + 2c\phi_1^{(n+1)} + ic^2 H_{12}^{(n)}; \tag{6.35}$$

thus,

$$F_{11} \to F_{11} + 2cD_1 + ic^2 t\, F_{12}. \tag{6.36}$$

Combining Eqs. (6.3) and (6.4) with $e = 0$ yields

$$e' = -2m'c/(1 + c^2) \tag{6.37}$$

or

$$\frac{1 + c^2}{1 - c^2} = \beta' \tag{6.38}$$

so

$$1 - c^2 H_{11} \to (1 - c^2)[x' + \beta'][x' + 1]^{-1}. \tag{6.39}$$

Applying Eqs. (6.30), (6.31), and (6.34) and using the appropriate scale changes, we find

$$F_{11} \to \frac{t(1 - c^2)}{[x' + \beta']}\left[\frac{(x' - 2ty')}{S'} - 1\right][1 - 4t^2]^{-1/2}, \tag{6.40}$$

$$D_1 \to \frac{tc}{[x' + \beta']}\left[\frac{(x' - 2ty')}{S'} - 1\right][1 - 4t^2]^{-1/2}, \tag{6.41}$$

$$\begin{aligned} F_{12} \to{}& \frac{i}{[x' + \beta']}\left[\frac{(x' - 2ty')}{S'}\right][1 - 4t^2]^{-1/2}, \\ &+ \frac{i(1 + c^2)}{(x' + \beta')(1 - c^2)}[1 - 4t^2]^{-1/2}. \end{aligned} \tag{6.42}$$

Next we use Eq. (6.36) and the result is

$$F_{11} \to \frac{t}{x' + \beta'}\left[\frac{(x' - 2ty')}{S'} - \beta'\right][1 - 4t^2]^{-1/2}. \tag{6.43}$$

Using a similar procedure for D_1, we find

$$D_1 \to \frac{t'q'\beta'}{x' + \beta'}[1 - 4t^2]^{-1/2}. \tag{6.44}$$

At this point we will, for convenience, drop the primes. Using Eqs. (2.9) and (2.19), we may derive

$$D_2 = it\,q\,\beta\,y(1 - 4t^2)^{-1/2} + \text{ constant } (t), \tag{6.45}$$

$$Q = \frac{it(\beta^2 - 1)y}{x + \beta}(1 - 4t^2)^{-1/2} + \text{ constant } (t). \tag{6.46}$$

Following the procedure of the previous section, we insert D_1, D_2, and Q into Eq. (5.6) and obtain

$$D_2 = iq\beta t\, y(1-4t^2)^{-1/2} + \frac{iq\beta^2}{2}[(1-4t^2)^{-1/2} - 1], \tag{6.47}$$

$$Q = \frac{it\, y(\beta^2-1)}{x+\beta}[1-4t^2]^{-1/2} + \frac{i}{2}[\beta^2[(1-4t^2)^{-1/2}-1] - (1-4t^2)^{-1/2}]. \tag{6.48}$$

Our methods have ensured that these are consistent with both the Harrison transformation and the required gauge conditions.

Proceedings as before in Sec. 5, we obtain the specific forms of the F_{AB} and S_A:

$$F_{11} = \frac{t}{x+\beta}\left[\frac{(x-2ty)}{S} - \beta\right][1-4t^2]^{-1/2}, \tag{6.49}$$

$$F_{12} = \frac{i}{x+\beta}\left[\frac{(x-2ty)}{S} + \beta\right][1-4t^2]^{-1/2}, \tag{6.50}$$

$$F_{21} = \frac{i}{2}[\beta^2 - 1 - \beta^2[1-4t^2]^{-1/2}] - i\beta t\, y[1-4t^2]^{-1/2} - \frac{i}{2S}[1-2ty(x-\beta) - 4t^2(x\beta+1-y^2)][1-4t^2]^{-1/2}, \tag{6.51}$$

$$F_{22} = \frac{1}{2t}[\beta^2 - 1 - \beta^2[1-4t^2]^{-1/2}] - \beta y[1-4t^2]^{-1/2} + \frac{1}{2tS}[1-2ty(x-\beta) - 4t^2(x\beta+1-y^2)][1-4t^2]^{-1/2}, \tag{6.52}$$

$$S_1 = \frac{iq\beta^2}{2}(1-[1-4t^2]^{-1/2}) - \frac{it\,q\beta^2 y}{x+\beta}[1-4t^2]^{-1/2} + \frac{2it^2 q\beta x}{S}[1-4t^2]^{-1/2} - \frac{it\,q\beta\, y(2ty+\beta)}{(x+\beta)S}[1-4t^2]^{-1/2}, \tag{6.53}$$

$$S_2 = \frac{q\beta^2}{2t}(1-[1-4t^2]^{-1/2}) - \frac{it\,q\beta^2 y}{x+\beta}[1-4t^2]^{-1/2} + \frac{2it^2 q\beta x}{S}[1-4t^2]^{-1/2} - \frac{it\,q\beta\, y}{(x+\beta)S}(2ty+\beta)[1-4t^2]^{-1/2}. \tag{6.54}$$

7. CONCLUSIONS

We now have in hand the particular equations necessary for the calculation of the generation potentials. We have gained experience in applying these equations to specific situations and have elucidated a few of the ramifications of the inherent gauge freedom that exists. Thus, we are now ready to apply the methods of paper IV to produce new asymptotically flat solutions. The results could be of great interest since the Harrison transformation and the procedures in paper IV do not commute. Thus, Schwarzschild → rotating→ charged should not be equivalent to Schwarzschild→ charge → rotating! Yet it is difficult to imagine what the result could be except the Kerr–Newman solution. In addition, since gauge transformations do not commute with Harrison transformations, how will the final solutions be affected by our choice of gauge?

[1]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977) (referred to as paper I).
[2]W. Kinnersley and D.M. Chitre, J. Math. Phys. **18**, 1538 (1977) (referred to as paper II).
[3]W. Kinnersley and D.M. Chitre, J. Math. Phys. **19**, 1926 (1978) (referred to as paper III).
[4]W. Kinnersley and D.M. Chitre, J. Math. Phys. **19**, 2037 (1978) (referred to as paper IV).
[5]B.K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[6]F.J. Ernst, Phys. Rev. **168**, 1415 (1967).
[7]Reference 6.

A solution of the Cauchy initial value problem in the nonsymmetric theory of gravitation

J. W. Moffat[a)]
Department of Physics, University of Toronto, Toronto, Ontario, Canada, M5S, 1A7

(Received 21 May 1979; accepted for publication 22 February 1980)

The Cauchy initial value problem in a theory of gravity based on a nonsymmetric Hermitian metric is investigated. The field equations are solved in terms of a series expansion of the metric $g_{\mu\nu}$. In the nth order the equations take the form of either second-order hyperbolic partial differential equations for g_{ij} $(i,j = 1,2,3)$ or a first-order equation for the vector gauge field W_i. Given initial data on a space-like surface S $(x^0 = 0)$, the integration forward in time of the equations can be performed, assuming that the field variables are reasonably smooth and analytic.

I. INTRODUCTION

A new theory of gravitation has been proposed[1,2] based on a nonsymmetric metric $g_{\mu\nu}$. It was shown that a rigorous spherically symmetric solution of the free-field equations exhibits world-line completeness if test particles move along paths in the non-Riemannian geometry. In addition to the energy-momentum tensor there occurs a new source in the form of a conserved fermion number current density, which plays a role in determining the structure of space–time. The cosmological solutions obtained in the theory[3,4] display different analytic properties from general relativity when $t \to 0$.

In Sec. II we shall give a derivation of the field equations with sources based on a Lagrangian density $\mathbf{L}$. A solution of the Cauchy initial-value problem is presented in Secs. II–V, using a general iterative solution of the vacuum equations.

II. DERIVATION OF THE FIELD EQUATIONS

We shall begin with a derivation of the field equations in the presence of phenomenological sources. The basic notation will be the same as in Ref. (1). We raise and lower suffixes by using the relation

$$g^{\mu\nu}g_{\sigma\nu} = g^{\nu\mu}g_{\nu\sigma} = \delta^{\mu}_{\sigma}\,. \tag{2.1}$$

A nonsymmetric affine connection $W^{\lambda}_{\mu\nu}$ is related to a (Hermitian) connection $\Gamma^{\lambda}_{\mu\nu}$ by the equation

$$W^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\mu\nu} - \tfrac{2}{3}\delta^{\lambda}_{\mu}W_{\nu}\,, \tag{2.2}$$

where $W_{\nu} = \frac{1}{2}(W^{\sigma}_{\nu\sigma} - W^{\sigma}_{\sigma\nu})$ is a purely imaginary vector gauge field. From (2.2) we have

$$\Gamma_{\mu} \equiv \Gamma^{\sigma}_{[\mu\sigma]} = 0. \tag{2.3}$$

A curvature tensor can be formed from the W connection:

$$R^{\sigma}_{\mu\nu\rho}(W) = (W^{\sigma}_{\mu\nu,\rho} - W^{\sigma}_{\alpha\nu}W^{\alpha}_{\mu\rho}) - (W^{\sigma}_{\mu\rho,\nu} - W^{\sigma}_{\alpha\rho}W^{\alpha}_{\mu\nu}) \tag{2.4}$$

and the contracted curvature tensor is

$$A_{\mu\nu}(W) = (W^{\beta}_{\mu\nu,\beta} - W^{\beta}_{\alpha\nu}W^{\alpha}_{\mu\beta}) - (W^{\beta}_{\mu\beta,\nu} - W^{\beta}_{\alpha\beta}W^{\alpha}_{\mu\nu}). \tag{2.5}$$

We can symmetrize $A_{\mu\nu}(W)$ with respect to μ and ν in the term $W^{\beta}_{\mu\beta,\nu}$ to give[5]

$$R_{\mu\nu}(W) = W^{\beta}_{\mu\nu,\beta} - \tfrac{1}{2}(W^{\beta}_{\mu\beta,\nu} + W^{\beta}_{\nu\beta,\mu}) - W^{\beta}_{\alpha\nu}W^{\alpha}_{\mu\beta} + W^{\beta}_{\alpha\beta}W^{\alpha}_{\mu\nu}\,. \tag{2.6}$$

This can be written as

$$R_{\mu\nu}(W) = A_{\mu\nu}(W) + \tfrac{1}{2}(W^{\beta}_{\mu\beta,\nu} - W^{\beta}_{\nu\beta,\mu}). \tag{2.7}$$

A study of the contractions of (2.4) with respect to σ and μ reveals that $R_{\mu\nu}(W)$ is a tensor.

By substituting (2.2) into (2.6), we get

$$R_{\mu\nu}(W) = R_{\mu\nu}(\Gamma) + \tfrac{2}{3}W_{[\mu,\nu]}\,, \tag{2.8}$$

where $W_{[\mu,\nu]} = \frac{1}{2}(W_{\mu,\nu} - W_{\nu,\mu})$ and

$$R_{\mu\nu}(\Gamma) = \Gamma^{\beta}_{\mu\nu,\beta} - \tfrac{1}{2}(\Gamma^{\beta}_{(\mu\beta),\nu} + \Gamma^{\beta}_{(\nu\beta),\mu}) - \Gamma^{\beta}_{\alpha\nu}\Gamma^{\alpha}_{\mu\beta} + \Gamma^{\beta}_{(\alpha\beta)}\Gamma^{\alpha}_{\mu\nu}\,. \tag{2.9}$$

Our Lagrangian density is

$$\mathbf{L} = \mathbf{g}^{\mu\nu}R_{\mu\nu}(W) + \mathbf{L}_m\,, \tag{2.10}$$

where $\mathbf{L}_m$ is the Lagrangian density for the phenomenological matter sources given by

$$\mathbf{L}_m = -(8\pi G/c^4)g^{\mu\nu}\mathbf{T}_{\mu\nu} + (8\pi a^2/3)W_{\mu}\mathbf{S}^{\mu}. \tag{2.11}$$

Here $\mathbf{T}_{\mu\nu}$ and $\mathbf{S}^{\mu}$ are the sources of the $g_{\mu\nu}$ field and we have used the notation $\mathbf{X}_{\mu\nu} = (-g)^{1/2}X_{\mu\nu}$. Moreover, $\mathbf{S}^{\mu}$ is a purely imaginary source vector density given by

$$\mathbf{S}^{\mu} = \left(\frac{3}{8\pi a^2}\right)\frac{\partial \mathbf{L}_m}{\partial W_{\mu}}\,, \tag{2.12}$$

where a is a new universal coupling constant (dimension of a length). The Hermitian tensor density $\mathbf{T}_{\mu\nu}$ is

$$\mathbf{T}_{\mu\nu} = -\left(\frac{c^4}{8\pi G}\right)\frac{\partial \mathbf{L}_m}{\partial g^{\mu\nu}}\,. \tag{2.13}$$

We can now use the Palatini method, varying g and W as independent field variables (such that δg and δW vanish at the boundaries of integration). The W variation gives

$$\mathbf{g}^{\mu\nu}{}_{,\sigma} + \mathbf{g}^{\rho\nu}W^{\mu}_{\rho\sigma} + \mathbf{g}^{\mu\rho}W^{\nu}_{\sigma\rho} - \mathbf{g}^{\mu\nu}W^{\rho}_{\sigma\rho} + \tfrac{2}{3}\delta^{\nu}_{\sigma}\mathbf{g}^{\mu\rho}W^{\beta}_{[\rho\beta]} + (4\pi a^2/3)(\mathbf{S}^{\nu}\delta^{\mu}_{\sigma} - \mathbf{S}^{\mu}\delta^{\nu}_{\sigma}) = 0. \tag{2.14}$$

Contracting over ν and σ and antisymmetrizing gives the field equation

$$\mathbf{g}^{[\mu\nu]}{}_{,\nu} = 4\pi a^2\mathbf{S}^{\mu}. \tag{2.15}$$

a)Supported by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/80/071798-04$1.00

By varying $g^{\mu\nu}$ in (2.10), we get

$$G_{\mu\nu}(W) = (8\pi G/c^4)T_{\mu\nu}\,, \tag{2.16}$$

where $G_{\mu\nu}(W)$ is the generalized Einstein tensor

$$G_{\mu\nu}(W) = R_{\mu\nu}(W) - \tfrac{1}{2}g_{\mu\nu}R(W), \tag{2.17}$$

with $R_{\mu\nu} = g_{\mu\alpha}g_{\beta\nu}R^{\alpha\beta}$ and $R = g^{\mu\nu}R_{\mu\nu}$.

We introduce another Hermitian connection $\Lambda^{\lambda}_{\mu\nu}$:

$$\Lambda^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\mu\nu} + D^{\lambda}_{\mu\nu}(S), \tag{2.18}$$

where $D^{\lambda}_{\mu\nu}$ is defined by the equation

$$g_{\rho\nu}D^{\rho}_{\mu\sigma} + g_{\mu\rho}D^{\rho}_{\sigma\nu} = -(4\pi a^2/3)S^{\rho}(g_{\mu\sigma}g_{\rho\nu} - g_{\mu\rho}g_{\sigma\nu} + g_{\mu\nu}g_{[\sigma\rho]}). \tag{2.19}$$

In terms of $\Lambda^{\lambda}_{\mu\nu}$ [and using (2.1)], the field equations (2.14) can be written

$$g_{\mu\nu,\sigma} - g_{\rho\nu}\Lambda^{\rho}_{\mu\sigma} - g_{\mu\rho}\Lambda^{\rho}_{\sigma\nu} = 0. \tag{2.20}$$

When the sources $T_{\mu\nu}$ and S^{μ} are zero, the field equations reduce to[6]

$$g_{\mu\nu,\sigma} - g_{\rho\nu}\Gamma^{\rho}_{\mu\sigma} - g_{\mu\rho}\Gamma^{\rho}_{\sigma\nu} = 0, \tag{2.21}$$

$$\mathbf{g}^{[\mu\nu]}{}_{,\nu} = 0, \tag{2.22}$$

$$R_{\mu\nu}(\Gamma) = \tfrac{2}{3}W_{[\nu,\mu]}\,. \tag{2.23}$$

The variational principle yields the four generalized Bianchi identities

$$[\mathbf{g}^{\alpha\nu}G_{\rho\nu}(\Gamma) + \mathbf{g}^{\nu\alpha}G_{\nu\rho}(\Gamma)]_{,\alpha} + g^{\mu\nu}{}_{,\rho}\mathbf{G}_{\mu\nu}(\Gamma) = 0. \tag{2.24}$$

In terms of (2.16), these are equivalent to

$$(g^{\alpha\nu}\mathbf{T}_{\rho\nu} + g^{\nu\alpha}\mathbf{T}_{\nu\rho})_{,\alpha} + g^{\mu\nu}{}_{,\rho}\mathbf{T}_{\mu\nu} + \tfrac{2}{3}W_{[\rho,\nu]}\mathbf{S}^{\nu} = 0. \tag{2.25}$$

The field equations are invariant under the gauge transformation

$$\widetilde{W}_{\mu} = W_{\mu} + \lambda_{,\mu}\,, \tag{2.26}$$

where λ is an arbitrary (pure imaginary) scalar field. Using this in the variational principle, we obtain a fifth identity

$$\mathbf{g}^{[\mu\nu]}{}_{,\nu,\mu} = 4\pi\mathbf{S}^{\mu}{}_{,\mu} = 0. \tag{2.27}$$

In the absence of sources, the identities (2.24) can be written in the covariant form

$$\mathrm{Re}[\mathbf{g}^{\alpha\nu}G_{\rho-\nu-;\alpha}(\Gamma)] = 0, \tag{2.28}$$

where we used the Einstein $+$ and $-$ notation for covariant differentiation with respect to Γ.[6]

For a macroscopic system the conserved four-vector density $\mathbf{S}^{\mu}$ can be written

$$\mathbf{S}^{\mu} = (i/c)nu^{\mu}\sqrt{-g}, \tag{2.29}$$

where n is the number density of fermions and u^{μ} is a velocity four-vector. In terms of Dirac spinors, $\mathbf{S}^{\mu}$ is

$$\mathbf{S}^{\mu} = -\bar{\psi}\gamma^{\mu}\psi\sqrt{-g}. \tag{2.30}$$

The new constant of integration l^2 that occurs in the static, spherically symmetric solution of the vacuum equations, discussed in Ref. (1), is given by

$$l^2 = -ia^2\int \mathbf{S}^4\, d^3x = a^2N, \tag{2.31}$$

where N is the number of fermions in a body.

The theory is based on two basic premises:

(i) The Hermitian length of a vector A^{μ} given by $(g_{\mu\nu}A^{\mu}\bar{A}^{\nu})$ is preserved under parallel transfer with respect to $\Lambda^{\lambda}_{\mu\nu}$ (or $\Gamma^{\lambda}_{\mu\nu}$ in free space).

(ii) A local *inertial* frame of reference exists such that at a point $x^{\mu} = x'^{\mu}$:

$$\Lambda^{\lambda}_{(\mu\nu)}\big|_{x=x'} = 0.$$

The general solution to the initial value problem would consist of finding a global solution of the partial differential equations (2.21)–(2.23) on a real manifold V_4. A solution can only be determined if the functions $g_{\mu\nu}$ and W_{μ} satisfy regularity conditions and obey boundary conditions. In the present work, we shall assume that V_4 is a differentiable manifold and that $g_{\mu\nu}$ is sufficiently analytic to allow a series expansion of $g_{\mu\nu}$ in a parameter ϵ to converge to a solution in V_4. The equation $ds^2 = 0$ defines at each point of V_4 a light cone of directions tangent to V_4 and the signature is of normal hyperbolic form $(-\ -\ -\ +)$. At certain singular points the metric ds^2 may change sign.[1] We will not consider this possibility in the present study of the Cauchy problem.

III. STRUCTURE OF THE FIELD EQUATIONS

The field equations of empty space (2.23) could be written in the form[6]

$$R_{(\mu\nu)}(\Gamma) = 0, \tag{3.1}$$

$$R_{[\mu\nu,\sigma]}(\Gamma) = 0. \tag{3.2}$$

In addition to the four identities (2.24), these equations satisfy the identity

$$\epsilon^{\mu\nu\sigma\rho}R_{[\mu\nu,\sigma]}(\Gamma)_{,\rho} = 0. \tag{3.3}$$

It is easy to show that the system of equations (2.22), (3.1), and (3.2) is compatible. There are 18 field equations but, owing to the existence of $6 = 4+2$ identities (2.24), (2.27), and (3.3), only 12 of the field equations are independent. On the other hand, only 12 of the 16 field variables $g_{\mu\nu}$ are independent, owing to the existence of four arbitrary coordinate transformations: $x'^{\mu} = (\partial x'^{\mu}/\partial x^{\alpha})x^{\alpha}$.

We shall not work with (3.2), for the antisymmetric part of Eq. (2.23) is more fundamental; it retains the important role played by the vector gauge field W_{μ}. Equations (2.22) and (2.23) constitute 20 field equations but, because these equations have $5 = 4+1$ identities (2.24) and (2.27) among them, only 15 of the equations are independent. Of the 20 field variables $g_{\mu\nu}$ and W_{μ} only 15 are independent, owing to the existence of four arbitrary coordinate transformations and the gauge transformation (2.26). Therefore the system of partial differential equations (2.22) and (2.23) constitutes a compatible system of equations.

IV. FORMULATION OF THE INITIAL-VALUE PROBLEM

We shall be given a three-dimensional hypersurface S oriented in space. A coordinate system is chosen such that the hypersurface S is described by the equation $x^0 = 0$. We are given the initial Cauchy data $g_{\mu\nu}$, $g_{\mu\nu,0}$, and W_i $(i = 1,2,3)$ on S and $g_{00} > 0$. This allows us to compute all the interior derivatives $g_{\mu\nu,i}$, $W_{i,j}$, etc. in S.

V. SEPARATION OF THE CAUCHY PROBLEM

As we are unable to find a simple closed-form solution for $\Gamma^{\lambda}_{\mu\nu}$ from the system of differential equations (2.21), we

shall solve the initial-value problem using a recursion solution of $\Gamma^{\lambda}_{\mu\nu}$, based on a weak-field expansion of $g_{\mu\nu}$ and $\Gamma^{\lambda}_{\mu\nu}$ (see Appendix). Equation (2.22) can be written in nth order as

$$\eta^{\alpha\beta}g^{(n)}_{[\beta\nu],\alpha} = C^{(n)}_{\nu}, \tag{5.1}$$

where $C^{(n)}_{\nu}$ contains products of g's of order $n-1$ and lower. In the nth order of approximation the field equations (2.23) take the form (see Appendix)

$$R^{(n)}_{(\mu\nu)}(\Gamma) = \tfrac{1}{2}\eta^{\rho\sigma}(g^{(n)}_{(\mu\sigma),\nu,\rho} + g^{(n)}_{(\nu\rho),\mu,\sigma} - g^{(n)}_{(\mu\nu),\rho,\sigma} - g^{(n)}_{(\rho\sigma),\mu,\nu}) + K^{(n)}_{(\mu\nu)} = 0, \tag{5.2}$$

$$R^{(n)}_{[\mu\nu]}(\Gamma) = -\tfrac{1}{2}\eta^{\rho\sigma}g^{(n)}_{[\mu\nu],\rho,\sigma} + K^{(n)}_{[\mu\nu]} + \tfrac{2}{3}W^{(n)}_{[\mu,\nu]} = 0, \tag{5.3}$$

where the $K^{(n)}_{(\mu\nu)}$ and the $K^{(n)}_{[\mu\nu]}$ contain products of g's and their derivatives of order $n-1$ and lower. Separating the spatial and time components of (5.2), we get

$$R^{(n)}_{(ij)} = -\tfrac{1}{2}g^{(n)}_{(ij),0,0} + N^{(n)}_{(ij)} + K^{(n)}_{(ij)} = 0, \tag{5.4}$$

$$R^{(n)}_{(i0)} = N^{(n)}_{(i0)} + K^{(n)}_{(i0)} = 0, \tag{5.5}$$

$$R^{(n)}_{00} = \tfrac{1}{2}g^{(n)}_{ii,0,0} + N^{(n)}_{00} + K^{(n)}_{00} = 0, \tag{5.6}$$

where the $N_{(\mu\nu)}$ can be expressed in terms of the initial data on S. Similarly, from (5.3) we have

$$R^{(n)}_{[ij]} = -\tfrac{1}{2}g^{(n)}_{[ij],0,0} + \tfrac{1}{2}\nabla^2 g^{(n)}_{[ij]} + K^{(n)}_{[ij]} + \tfrac{2}{3}W^{(n)}_{[i,j]} = 0, \tag{5.7}$$

$$R^{(n)}_{[i0]} = -\tfrac{1}{2}g^{(n)}_{[i0],0,0} + \tfrac{1}{2}\nabla^2 g^{(n)}_{[i0]} + K^{(n)}_{[i0]} + \tfrac{2}{3}W^{(n)}_{[i,0]} = 0, \tag{5.8}$$

while (5.1) becomes

$$g^{(n)}_{[i0],0} = g^{(n)}_{[ij],j} + C^{(n)}_{i} \tag{5.9}$$

The Cauchy separation of the set of equations (5.4)–(5.6) is familiar from general relativity.[7,8] A calculation of $N^{(n)}_{(i0)}$ gives

$$N^{(n)}_{(i0)} = -\tfrac{1}{2}g^{(n)}_{(j0),i,j} + \tfrac{1}{2}\nabla^2 g^{(n)}_{(0i)} + \tfrac{1}{2}g^{(n)}_{(jj),i,0} - \tfrac{1}{2}g^{(n)}_{(ji),0,j}. \tag{5.10}$$

Thus Eq. (5.5) represents only a constraint on the initial data and the interior derivatives of $g^{(n)}_{(\mu\nu)}$ in S. We also have for the combination

$$\eta^{00}R^{(n)}_{00} - \eta^{ij}R^{(n)}_{ij} = R^{(n)}_{00} + R^{(n)}_{ii} = N^{(n)}_{00} + N^{(n)}_{ii} + K^{(n)}_{00} + K^{(n)}_{ii} = 0. \tag{5.11}$$

If we consider the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$, then to nth order

$$G^{(n)0}_{0} = R^{(n)0}_{0} - \tfrac{1}{2}R^{(n)} = R^{(n)}_{00} - \tfrac{1}{2}(-R^{(n)}_{ii} + R^{(n)}_{00}) = \tfrac{1}{2}(N^{(n)}_{00} + N^{(n)}_{ii} + K^{(n)}_{00} + K^{(n)}_{ii}) = 0 \tag{5.12}$$

and

$$G^{(n)0}_{i} = R^{(n)0}_{i} = N^{(n)0}_{i} + K^{(n)0}_{i} = 0. \tag{5.13}$$

We see that the four equations

$$G^{(n)0}_{\lambda} = G^{(n)\lambda}_{0} = 0 \tag{5.14}$$

depend only on the initial data.

Since there is a one-parameter gauge freedom of the field equations under the gauge transformation (2.26), we can fix the gauge using the condition

$$W^{(n)}_{0} = 0. \tag{5.15}$$

We thereby arrive at the set of dynamical equations in nth order:

$$g^{(n)}_{(ij),0,0} = 2N^{(n)}_{(ij)} + 2K^{(n)}_{(ij)}, \tag{5.16}$$

$$g^{(n)}_{[ij],0,0} = \nabla^2 g^{(n)}_{[ij]} + K^{(n)}_{[ij]} + \tfrac{1}{3}W_{[ij]}, \tag{5.17}$$

$$W^{(n)}_{i,0} = \tfrac{3}{4}g^{(n)}_{[ij],0,j} - \tfrac{3}{4}\nabla^2 g^{(n)}_{[i0]} + L^{(n)}_{[i0]}. \tag{5.18}$$

where we have used Eq. (5.9) and $L^{(n)}_{[i0]}$ contains terms from $C^{(n)}_{i,0}$ and $K^{(n)}_{[i0]}$ that are known from the lower orders of approximation.

Let us again consider the compatibility of these equations. Equations (5.16)–(5.18) total 12 and together with (5.1) and (5.14) there are in all 20 equations but only 15 of these are independent, owing to the existence of the five identities (2.24) and (2.27). There are 20 unknown functions $g^{(n)}_{(\mu\nu)}$, $g^{(n)}_{[\mu\nu]}$, and $W^{(n)}_{\mu}$. However, only 15 of these functions are free to be determined in view of the four arbitrary choices of coordinates and the one gauge condition (5.15). Thus the system of equations (5.16)–(5.18) forms a compatible dynamical scheme.

VI. SOLUTION OF THE CAUCHY PROBLEM

We observe that the right-hand sides of (5.16)–(5.18) are known from the initial data in nth order: $g^{(n)}_{(ij)}$, $g^{(n)}_{(ij),0}$, $g^{(n)}_{[ij]}$, $g^{(n)}_{[ij],0}$, $g^{(n)}_{[i0]}$, and $W^{(n)}_{i}$ [$g^{(n)}_{[i0],0}$ can be found from Eq. (5.9) and the interior derivative of $g^{(n)}_{[ij]}$]. Equations (5.16) and (5.17) are both hyperbolic second-order partial differential equations in "normal form." They can be integrated forward in time, given reasonable smoothness and analyticity properties. Equation (5.18) is a first-order differential equation in time and it can be solved in every order to determine W_i. Thus, if the series expansions of the $g_{\mu\nu}$ and W_{μ} converge, then the time-evolution problem leads to a solution of the original system of equations (2.22) and (2.23), provided that the four compatibility conditions (5.14) converge and remain true once they are fulfilled for the time $x^0 = 0$.[9]

The proof of the existence and of the uniqueness of the solutions of (5.16)–(5.18) can be given if we use the Cauchy–Kowalewski theorem. Of course, we must still solve the Cauchy problem for the rigorous equations not using a series expansion about the flat-space metric $\eta_{\mu\nu}$, as is done in general relativity.[7] Work is in progress to solve this problem employing the theory of characteristic hypersurfaces[10] and the bicharacteristics of these hypersurfaces, determined from the equation

$$g_{\mu\nu}\frac{dx^{\mu}}{ds}\frac{dx^{\nu}}{ds} = 0. \tag{6.1}$$

The basic dynamical structure contained in the equations solved by the series expansion about flat space can still be seen to be preserved in the rigorous geometrical form of the field equations (2.21)–(2.23). We therefore expect that a Cauchy initial-value solution exists for the rigorous system of equations.

APPENDIX

We shall now give a description of the weak-field expansion of the field equations (2.21)–(2.23). This has been studied previously in the literature.[6,11] In the weak-field approximation the $g_{\mu\nu}$ are expanded in the series

$$g_{\mu\nu} = \eta_{\mu\nu} + \epsilon g^{(1)}_{\mu\nu} + \epsilon^2 g^{(2)}_{\mu\nu} + \cdots, \tag{A1}$$

where $\eta_{\mu\nu}$ is the Minkowski metric. We have also

$$\Gamma^{\lambda}_{\mu\nu} = \epsilon\Gamma^{\lambda\,(1)}_{\mu\nu} + \epsilon^2\Gamma^{\lambda\,(2)}_{\mu\nu} + \cdots,$$
$$W_{\mu} = \epsilon W^{(1)}_{\mu} + \epsilon^2 W^{(2)}_{\mu} + \cdots. \tag{A2}$$

If we expand $g^{\mu\nu}$ similarly,

$$g^{\mu\nu} = \eta^{\mu\nu} + \epsilon g^{(1)\mu\nu} + \epsilon^2 g^{(2)\mu\nu} + \cdots, \tag{A3}$$

then the $g^{(1)\mu\nu}$ and $g^{(2)\mu\nu}$, etc., can be found using the orthogonality condition (2.1):

$$\begin{aligned} g^{(1)\mu\nu} &= -\eta^{\mu\alpha}\eta^{\beta\nu}g^{(1)}_{\beta\alpha}, \\ g^{(2)\mu\nu} &= -\eta^{\mu\alpha}\eta^{\beta\nu}g^{(2)}_{\beta\alpha} + \eta^{\mu\alpha}\eta^{\beta\nu}\eta^{\rho\sigma}g^{(1)}_{\rho\alpha}g^{(1)}_{\beta\sigma}, \\ g^{(3)\mu\nu} &= -\eta^{\mu\alpha}\eta^{\beta\nu}g^{(3)}_{\beta\alpha} + \eta^{\mu\alpha}\eta^{\beta\nu}\eta^{\rho\sigma}g^{(2)}_{\rho\alpha}g^{(1)}_{\beta\sigma} \\ &\quad + \eta^{\mu\rho}\eta^{\alpha\gamma}\eta^{\beta\nu}g^{(2)}_{\gamma\alpha}g^{(1)}_{\beta\rho} - \eta^{\mu\rho}\eta^{\nu\beta}\eta^{\alpha\epsilon}\eta^{\delta\gamma}g^{(1)}_{\epsilon\delta}g^{(1)}_{\gamma\rho}g^{(1)}_{\beta\alpha} \end{aligned} \tag{A4}$$

It follows that

$$\begin{aligned} g^{(1)[\mu\nu]} &= -\eta^{\mu\alpha}\eta^{\beta\nu}g^{(1)}_{[\beta\alpha]}, \\ g^{(2)[\mu\nu]} &= -\eta^{\mu\alpha}\eta^{\beta\nu}g^{(2)}_{[\beta\alpha]} + \eta^{\mu\alpha}\eta^{\beta\nu}\eta^{\gamma\delta}(g^{(1)}_{(\gamma\alpha)}g^{(1)}_{[\beta\delta]} \\ &\quad + g^{(1)}_{[\gamma\alpha]}g^{(1)}_{(\beta\delta)}). \end{aligned} \tag{A5}$$

Up to the first approximation we find

$$\sqrt{-g} = 1 + \tfrac{1}{2}\eta^{\mu\nu}g^{(1)}_{(\mu\nu)}, \tag{A6}$$

so that to second approximation

$$\mathfrak{g}^{(2)[\mu\nu]}{}_{,\nu} = -(\eta^{\mu\alpha}\eta^{\beta\nu}g^{(2)}_{[\beta\alpha]})_{,\nu} - [\tfrac{1}{2}\eta^{\gamma\delta}\eta^{\mu\alpha}\eta^{\beta\nu}g^{(1)}_{(\gamma\delta)}g^{(1)}_{[\beta\alpha]} - \eta^{\mu\alpha}\eta^{\beta\nu}\eta^{\gamma\delta}(g^{(1)}_{(\gamma\alpha)}g^{(1)}_{[\beta\delta]} + g^{(1)}_{[\gamma\alpha]}g^{(1)}_{(\beta\delta)})]_{,\nu}. \tag{A7}$$

In the nth order of approximation Eq. (2.22) becomes (5.1).

Let us now consider Eq. (2.21). Cyclically permuting the indices $\mu\nu\sigma$ twice in (2.21) and solving, we get

$$g_{(\rho\sigma)}\Gamma^{\rho}_{\mu\nu} = [\sigma\mu\nu] - g_{[\mu\rho]}\Gamma^{\rho}_{\nu\sigma} - g_{[\rho\nu]}\Gamma^{\rho}_{\sigma\mu}, \tag{A8}$$

where

$$[\sigma\mu\nu] = \tfrac{1}{2}(g_{\sigma\nu,\mu} + g_{\mu\sigma,\nu} - g_{\nu\mu,\sigma}). \tag{A9}$$

Substituting the expansions (A1) and (A2) into the left-and right-hand sides of (A8) yields the recursion solution in the nth order

$$\Gamma^{(n)\lambda}_{\mu\nu} = \eta^{\lambda\sigma}[\sigma\mu\nu]^{(n)} - \eta^{\lambda\sigma}\sum_{m=1}^{n-1}(g^{(m)}_{(\rho\sigma)}\Gamma^{(n-m)\rho}_{\mu\nu} + g^{(m)}_{[\mu\rho]}\Gamma^{(n-m)\rho}_{\nu\sigma} + g^{(m)}_{[\rho\nu]}\Gamma^{(n-m)\rho}_{\sigma\mu}). \tag{A10}$$

In the first and second order this formula gives

$$\Gamma^{(1)\lambda}_{(\mu\nu)} = \tfrac{1}{2}\eta^{\lambda\sigma}(g^{(1)}_{(\sigma\nu),\mu} + g^{(1)}_{(\mu\sigma),\nu} - g^{(1)}_{(\nu\mu),\sigma}), \tag{A11}$$

$$\Gamma^{(2)\lambda}_{(\mu\nu)} = \tfrac{1}{2}\eta^{\lambda\sigma}(g^{(2)}_{(\sigma\nu),\mu} + g^{(2)}_{(\mu\sigma),\nu} - g^{(2)}_{(\nu\mu),\sigma}) - 2g^{(1)}_{[\beta\nu]}\Gamma^{(1)\beta}_{[\sigma\mu]} - 2g^{(1)}_{[\mu\beta]}\Gamma^{(1)\beta}_{[\nu\sigma]} - 2g^{(1)}_{(\alpha\beta)}\Gamma^{(1)\beta}_{(\mu\nu)}), \tag{A12}$$

$$\Gamma^{(1)\lambda}_{[\mu\nu]} = \tfrac{1}{2}\eta^{\lambda\sigma}(g^{(1)}_{[\sigma\nu],\mu} + g^{(1)}_{[\mu\sigma],\nu} - g^{(1)}_{[\nu\mu],\sigma}), \tag{A13}$$

$$\Gamma^{(2)\lambda}_{[\mu\nu]} = \tfrac{1}{2}\eta^{\lambda\sigma}(g^{(2)}_{[\sigma\nu],\mu} + g^{(2)}_{[\mu\sigma],\nu} - g^{(2)}_{[\nu\mu],\sigma}) - 2g^{(1)}_{[\beta\nu]}\Gamma^{(1)\beta}_{(\sigma\mu)} - 2g^{(1)}_{[\mu\beta]}\Gamma^{(1)\beta}_{(\nu\sigma)} - 2g^{(1)}_{(\sigma\beta)}\Gamma^{(1)\beta}_{[\mu\nu]}). \tag{A14}$$

In the calculation of the $R_{\mu\nu}(\Gamma)$ we can use the equation $\Gamma_{\mu} = \Gamma^{\alpha}_{[\mu\alpha]} = 0$, equivalent to Eq. (2.22), to obtain

$$R^{(1)}_{(\mu\nu)} = \Gamma^{(1)\alpha}_{(\mu\nu),\alpha} - \Gamma^{(1)\alpha}_{(\mu\alpha),\nu}, \tag{A15}$$

$$R^{(2)}_{(\mu\nu)} = \Gamma^{(2)\alpha}_{(\mu\nu),\alpha} - \Gamma^{(2)\alpha}_{(\mu\alpha),\nu} - \Gamma^{(1)\alpha}_{(\mu\beta)}\Gamma^{(1)\beta}_{(\alpha\nu)} - \Gamma^{(1)\alpha}_{[\mu\beta]}\Gamma^{(1)\beta}_{[\alpha\nu]} + \Gamma^{(1)\beta}_{(\alpha\beta)}\Gamma^{(1)\alpha}_{(\mu\nu)}, \tag{A16}$$

$$R^{(1)}_{[\mu\nu]} = \Gamma^{(1)\alpha}_{[\mu\nu],\alpha}, \tag{A17}$$

$$R^{(2)}_{[\mu\nu]} = \Gamma^{(2)\alpha}_{[\mu\nu],\alpha} - \Gamma^{(1)\alpha}_{(\mu\beta)}\Gamma^{(1)\beta}_{[\alpha\nu]} - \Gamma^{(1)\alpha}_{[\mu\beta]}\Gamma^{(1)\beta}_{(\alpha\nu)} + \Gamma^{(1)\beta}_{(\alpha\beta)}\Gamma^{(1)\alpha}_{[\mu\nu]}. \tag{A18}$$

Using (A10), we see from (A16) and (A18) that the field equations (2.23) can be written in the nth order of approximation as Eqs. (5.2) and (5.3).

[1] J. W. Moffat, Phys. Rev. D **19**, 3554 (1979).
[2] J. W. Moffat, Phys. Rev. D **19**, 3562 (1979).
[3] G. Kunstatter, J. W. Moffat, and P. Savaria, Phys. Rev. D **19**, 3559 (1979).
[4] G. Kunstatter, J. W. Moffat, and P. Savaria, Can. J. Phys. (to be published).
[5] K. Borchsenius, Nuovo Cimento A **46**, 403 (1979).
[6] A. Einstein and E. G. Straus, Ann. Math. **47**, 731 (1946).
[7] Y. Bruhat, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), p. 130.
[8] A. Lichnerowicz, *Théories Relativistes de La Gravitation et de L'électromagnétisme* (Masson, Paris, 1955).
[9] This is a subject of a separate investigation to be published elsewhere.
[10] A. H. Taub, Ill. J. Math. **1**, 370 (1957).
[11] A. Papapetrou, Proc. R. Irish Acad. Sect. A **52**, 69 (1948).

Diffusion in a periodic wind-tree model

J. Hardy and J. Weber
Laboratoire de Physique des Plasmas, Faculté des Sciences de Paris XI, Orsay 91405, France

(Received 28 November 1979; accepted 15 February 1980)

In this paper we study a two-dimensional wind-tree model where the diffusors are rectangles periodically disposed. One shows that the motion of the light particle depends strongly on the arithmetic properties of a number α. The light particle returns near its departure position at times which are in geometrical series; it is proved the particle escapes to infinity, but not faster than log$t$ loglog$t$, where t is the time.

1. INTRODUCTION

In the recent years, there has been considerable progress in understanding the equilibrium mechanisms of statistical mechanics. In a few cases one has proved properties such as ergodicity, K, or the Bernoulli property for realistic systems. The same success has not been obtained for the moment in transport theory: the onset of stochasticity is not sufficient in itself to fit the hydrodynamical laws with statistical mechanics. In fact, in this paper we present an example of abnormal diffusion in a model of Lorentz-gas, although it presents good equilibrium properties.

The lack of results in transport theory derived without approximations comes from several reasons; one is that there is no transport in finite systems. This can be explained easily by noticing that any transport coefficient occurs in a diffusion equation:

$$\frac{\partial F}{\partial t} = \theta \Delta F, \tag{1.1}$$

where F can be a concentration, a vorticity, or a temperature, and θ a diffusion coefficient, a viscosity , or a heat diffusivity, respectively. Equation (1.1) implies that if a perturbation is done in a dissipative media it (a particle or a vortex or an energy) diffuses to infinity according to the law

$$d \sim \sqrt{t}, \tag{1.2}$$

where d is the distance from the departure position at time zero, and t is the time. If the distance is bounded above, one can have any transport.

Do infinite systems mean infinite degrees of freedom? To deal with such systems involves a lot of technical difficulties. The functions and measures are defined by iterative processes just like the real numbers by iteration of an algorithm. One can produce measures which are not always absolutely continuous with each other or discontinuous functions. This fact could perhaps explain why very little is actually known about systems with infinite degrees of freedom, even for the equilibrium properties.

The periodic models provide a powerful means to get around the difficulties of infinite systems: they are infinite although their dynamical properties can be deduced from motion on finite tori. The choice which is made in this paper is to deal with a periodic Lorentz–Ehrenfest[1] model on a plane. The obstacles are fixed identical rectangles located periodically on a plane square lattice, and a light particle moves freely on the plane, except when it collides with a rectangle; in this case it is reflected according to the usual law of light reflection. The disposition of the rectangles is such that no trajectory can escape to infinity directly. The main results of this paper are

(i) The position reached by the wandering particle depends on the arithmetic properties of its position at time $t = 0$ and of the ratio of the sides of the rectangles.

(ii) If this ratio is rational, the particle either has a periodic bounded motion or escapes to infinity with a constant mean velocity.

(iii) If this ratio is irrational the particle escapes to infinity but comes back near the departure position at times (q_i, $i = 1,2,\cdots,\infty$), the time q_i being *grosso modo* in geometrical series.

(iv) In this case the wandering particle moves in a domain with a diameter increasing like log$t$ loglog$t$) (where t is the time). This means, of course, that there is abnormal diffusion.

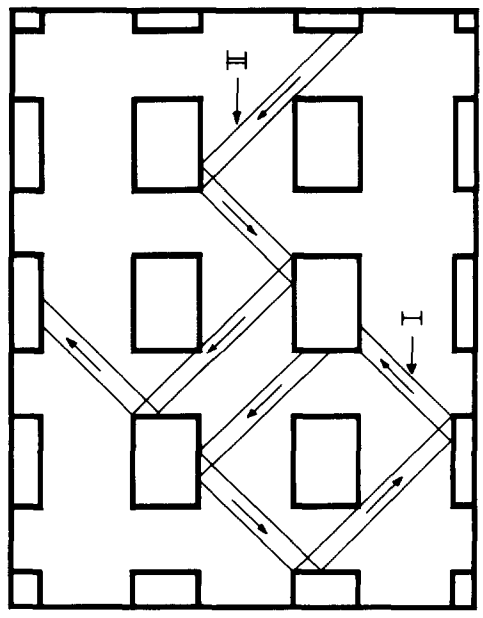


FIG. 1. Position of the diffusors: After four collisions one comes back to the same rectangle in case I and escapes in case II.

 0022-2488/80/071802-07$1.00

2. DEFINITION

The model used is this paper belongs to the general class of Lorentz-gas.[1] There are two kind of particles. The heavy ones, or trees, which consist of fixed parallel rectangles with sides of length $[(1-\alpha)/4, (1+\alpha)/4, |\alpha|<1]$ centered on a square lattice of the plane with a period two, as shown on Fig. 1. The light particles move freely between these obstacles with velocities of moduli $\sqrt{2}$, parallel to the bisectors of the lattice; the only interaction is the elastic scattering by the side walls of the rectangles. Actually, one may just consider a single light particle, the motion of the light particles being independent of each other. The light particle starts from a departure position with a definite velocity and will collide at times t_i (i an integer) with the rectangles R_i centered at even coordinates (x_i,y_i); the curvilinear coordinate of the impact point on R_i is denoted by s_i $(0\leqslant s_i<1)$. Clearly the wandering particle cannot escape directly to infinity, which means that the index i is not bounded above. Since the motion between two collisions is of no particular interest, we replaced, following Sinaï[2] and Gallavotti,[3] the continuous dynamical set $\{\mathbf{r}(t);\lambda(t)\}$ (the position and velocity, respectively, of the test particle at time t) by a discrete one $\{(x_i,y_i),s_i;\lambda(t_i)\}_{i=1,\ldots,\infty}$.

The direction of the velocity $\lambda(t)$, which can take only four different values (velocity is along the bisectors of the lattice) denoted 1, 2, 3, 4, changes suddenly at times t_i owing to the collisions. To avoid any ambiguity, we mean by $\lambda(t_i)$ the direction of λ just after a collision has occurred.

The time evolution law of the continuous model induces a time evolution for the discrete dynamical system. One can make two remarks which greatly simplify the dynamics.

First, the model being periodic, the curvilinear coordinate s_{i+1} and velocity λ_{i+1} at time t_{i+1} do not depend on the coordinates of the rectangle where the particle is located at time t_i, say (x_i,y_i). In the same way, the displacement $(x_{i+1}-x_i, y_{i+1}-y_i)$ depends on $[s_i,\lambda(t_i)]$ but not on (x_i,y_i). This means that the dynamics can be studied on the torus $[s,\lambda]$ and the positions (x_i,y_i) calculated as the sum of the displacements.

The second remark permits the elimination of velocity as a dynamical variable. In fact, one may suppose that there exist four laws giving s_{i+1} as a function of s_i corresponding to $\lambda_i = 1, 2, 3$, or 4.

As a matter of fact, one can see on Fig. 2 that on the perimeter of a rectangle, just after a collision there are only two possible directions for velocities, which are perpendicular to each other. This reduces the number of time evolution laws to two. But the velocity remains parallel to the initial one after even numbers of collisions and perpendicular to it after odd numbers of collisions. By choosing conventionally the initial velocity parallel to the second bisector and studying the position of the wandering particle after even numbers of collisions (which is sufficient to investigate diffusion), we get simply a dynamical system of dimension one. For technical reasons, we shall calculate the position after $4i$ collisions (i an integer). The reader is requested to convince himself, by drawing some trajectories on Fig. 1, that

$$s_{i+4} = s_i + \alpha \pmod 4. \tag{2.1}$$

This very simple result comes from the *a priori* strange definition for the rectangle sides, lengths $[(1-\alpha)/4, (1+\alpha)/4)]$; for non *ad hoc* models, for example squares and rectangles otherwise disposed on the elementary torus, one gets the interval exchange transformation[4] for the time evolution of the curvilinear coordinate s in place of a translation on the circle. The light particle returns to its departure rectangle after four collisions, unless the curvilinear coordinate s_i belongs to one of the four segments of lengths $\alpha/2$ (See Fig. 2):

$$I_1 = \tfrac{1}{4}[1-3\alpha, 1-\alpha] \pmod 1, \tag{2.2a}$$
$$I_2 = \tfrac{1}{4}[2-4\alpha, 2-2\alpha] \pmod 1, \tag{2.2b}$$
$$I_3 = \tfrac{1}{4}[3-3\alpha, 3-\alpha] \pmod 1, \tag{2.2c}$$
$$I_4 = \tfrac{1}{4}[4-4\alpha, 4-2\alpha] \pmod 1. \tag{2.2d}$$

In this case, it reaches a new rectangle and the displacement is given by:

$$x_{i+4}-x_i, y_{i+4}-y_i = \begin{cases} (4,-4), & \text{if } s_i\in I_1, \quad (2.3a)\\ (4,4), & \text{if } s_i\in I_2, \quad (2.3b)\\ (-4,4), & \text{if } s_i\in I_3, \quad (2.3c)\\ (-4,-4). & \text{if } s_i\in I_4. \quad (2.3d)\end{cases}$$

Then the coordinates of the rectangle reached after $4N$ collisions can be written as

$$x_{4N} = 4\sum_{i=1}^{N}(\chi_1+\chi_2-\chi_3-\chi_4)(s_0+i\alpha), \text{ mod } 1, \tag{2.4a}$$

$$y_{4N} = 4\sum_{i=1}^{N}(\chi_2+\chi_3-\chi_4-\chi_1)(s_0+i\alpha), \text{ mod } 1, \tag{2.4b}$$

where χ_μ, ($\mu = 1, 2, 3$, or 4) denote the characteristic functions of the segments I_μ on the circle with circumference of length one. In the next section we shall see how the random properties of a translation on the circle imply an anomalous diffusion for the model.

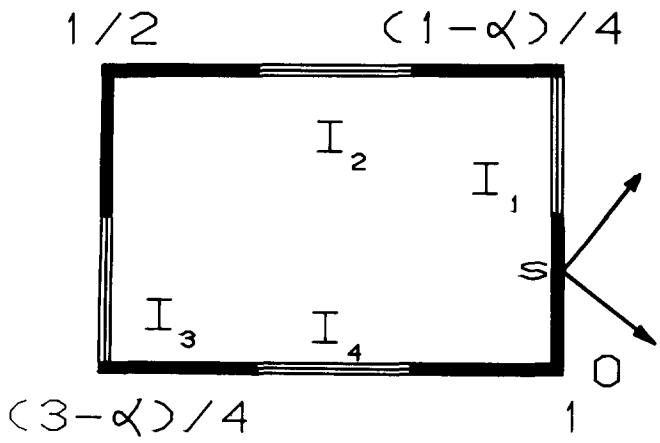


FIG. 2. Position of the four intervals $I_{1,2,3,4}$ on the rectangle.

3. STUDY OF THE DIFFUSION PROCESS

It appears that the study of dynamical systems for physical purposes often uses a probability measure. This can be done but is not very useful here; the calculus will appear clearer without it.

Before entering into the details of the translation on the circle, let us speak about the time. At time t_i the wandering particle has performed i collisions

$$t_i = \sum_{j=1}^{i} \Delta t_j\,(s_{j-1}). \tag{3.1}$$

the amount of time between two collisions is bounded above (there is no trajectory which escapes directly to infinity) and depends only on the curvilinear coordinate on a rectangle before the collision. With rational α the series $4k\alpha$ mod 4 ($k = 1,2,\cdots,\infty$) are periodic; times t_i ($t_i \to \infty$) are proportional to the number of collisions i. With α irrational, the rotation being ergodic, one can use the Birkhoff theorem

$$\lim_{i\to\infty} \frac{t_N}{N} = \lim_{i\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \Delta t_i(s_0 + 4i\alpha) = \langle \Delta t_i \rangle, \tag{3.2}$$

almost everywhere. Then in any case one can consider the number of collisions to be proportional to the time for large times.

A. Diffusion for α rational

Equations (2.4) giving an expression for the position of the rectangle reached after $4N$ collisions verify the following symmetry:

$$\mathbf{r}_{4N}(s_0) = [x_{4N}, y_{4N}](s_0) = -\mathbf{r}_{4N}(s_0 + \tfrac{1}{2} \bmod 1). \tag{3.3}$$

If one denotes α by p/q, one see that

(i) $s_0 + q\alpha = s_0$, mod 1,

(ii) $s_0 + \frac{q}{2}\alpha = s_0 + \frac{1}{2}$, mod 1, if q is even.

Then for q even

$$\mathbf{r}_{4q}(s_0) = \mathbf{r}_{2q}(s_0) + \mathbf{r}_{2q}(s_0 + \tfrac{1}{2}) = 0 \tag{3.4}$$

and the motion is periodic. For q odd one has

$$\mathbf{r}_{4kq}(s_0) = k\mathbf{r}_{4q}(s_0), \quad k \text{ an integer}, \tag{3.5}$$

and the motion is periodic on the torus but not on the plane; the escape to infinity is linear.

B. The case α irrational

It is already known that the points $s_0 + k\alpha$, k an integer, are dense on the circle and that the irrational rotation is ergodic with respect to the Lebesgue measure. In this paper we study this rotation in more detail. The case α irrational being more complicated than the rational case, we shall split the text in three parts: in the first one, this rotation is described precisely and give some useful notations are given, in the second part a maximization is given, and in the third a minimization. This is done, of course, because an exact formula for the distance cannot be given or it is very complicated. This is not surprising, since the model is supposed to reproduce something like a Brownian motion, which has to be a complicated thing.

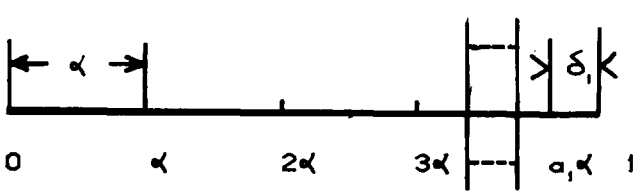


FIG. 3. The first jumps of length α on the unit circle.

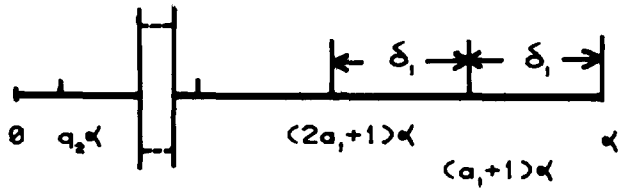


FIG. 4. Details of the points of the series $k\alpha$ entering in the segment $[0,\alpha]$.

1. Description of irrational rotation

In a first part, we shall suppose that $s_0 = 0$, and then translate the series of points $k\alpha$, k an integer, by s_0 mod 1. Consider the first $a_1 = I(1/\alpha)$ (i.e., the integer part of $1/\alpha$) of the points $k\alpha$, k integer which delimit $a_1 + 1$ segments. The first ones have length α and the last one, smaller, length $\delta_1 = [(1/\alpha) - I(1/\alpha)]\alpha$ (see Fig. 3). Successive points in the series $k\alpha$ will be inserted in these segments and, the last segment being smaller, the points $k\alpha$, k an integer, will move around the circle several times before falling in the small segment.

It is easy to see that in each of the a_1 long intervals one finds the same arrangement of points repeatedly; for this reason we shall study what happens in the first of them, that is $[0,\alpha]$. The point $a_1\alpha$ being to the left of the origin at a distance of δ_1, the successive points falling in the interval $[0,\alpha]$ start from α and perform a translation δ_1 to the left after each a_1 time interval (see Fig. 4). The last point at the right of the origin is called $q_2\alpha = (a_2a_1 + 1)\alpha$, where $a_2 = I[\alpha/\delta_1]$ and the distance to the origin is noted δ_2. In this case we see that we have the same problem of irrational rotation as in the original problem, except that intervals of time are enlarged (unity is replaced by a_1), that the length of the circle has a change of scale (unity is replaced by α) and the translation of α replaced by a translation $-\delta_1$. This remark will be the point of departure for a recurrence idea.

In the times following q_2, one reproduces the distribution of points of the interval $[0,\alpha]$ on the $a_1 - 1$ long intervals (length α); the first time a point enters the small interval of length δ_1 is $(q_2 + a_1)$. In general, the points of the series $k\alpha$ determine intervals of only three lengths on the circle,[5] but sometimes there are only two lengths. The first time this case occurs is a_1; one finds long intervals of lengths α and the small one of length δ_1. The second time corresponds to $(q_2 + a_1 - 1)$, just before a point in the center of the small interval of length δ_1; the a_1 small intervals of lenths δ_2 are to the right of the points $k\alpha$, k an integer, $0 < k < a_1$, and the others have lengths δ_1 (see Fig. 5).

FIG. 5. The points of the series $k\alpha$, k an integer less than $q_1 + q_2$, on the unit circle.

The small intervals (length δ_2) being separated from each other by long intervals (length δ_1), the points following $(q_2 + a_1)\alpha$ in the series $k\alpha$ will fall in the large intervals and produce the same arrangment of points in each of these intervals as in the interval $[a_1\alpha, 1]$. At each q_2 interval of time a new point falls in this segment, performing a translation of δ_2 to the right.

The recurrence appears more clearly by using the notation[6] of the continued fraction expansion of the number α. We define the continued fraction transformation by

$$T\alpha = \frac{1}{\alpha} - I\left(\frac{1}{\alpha}\right) \tag{3.6}$$

and the numbers a_i by

$$a_i = I\left(\frac{1}{T^{i-1}\alpha}\right). \tag{3.7}$$

Equations (3.6) and (3.7) permits one to express α by the fraction

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots \cfrac{}{a_i + T^i\alpha}}}. \tag{3.8}$$

If one reduces the fraction (3.8) to a unique denominator one gets

$$\alpha = \frac{1}{q_i}\left[p_i(-)^i\delta_i\right]. \tag{3.9}$$

The integers p_i and q_i are determined by the recurrence relations

$$q_0 = 1,\ q_1 = a_1,\ q_{n+1} = a_{n+1}q_n + q_{n-1}, \tag{3.10}$$

$$p_0 = 0,\ p_1 = 1,\ p_{n+1} = a_{n+1}p_n + p_{n-1}. \tag{3.11}$$

The number δ_i, very small, is given by the product

$$\delta_i = \prod_{j=0}^{i} T^j\alpha = \alpha T\alpha \cdots T^i\alpha. \tag{3.12}$$

The relation (3.8) is easy to obtain from definitions (3.6) and (3.7); in order to get (3.10–12) it is easier to use geometrical arguments. From (3.9) one can observe that $q_i\alpha$ approaches the origin from the right for i even and from the left for i odd. In any case, the distance from the origin decreases exponentially with the number i. In fact the continued fraction transformation being ergodic,[7] the Birkhoff theorem[8] asserts that

$$\lim_{N\to\infty}\frac{1}{N}\log\delta_N = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N}\log T^n\alpha = \int_0^1 \frac{\log x\,dx}{1+x} = -\frac{\pi^2}{12}, \tag{3.13}$$

almost everywhere with respect to the Gauss measure $d\mu = dx/(1+x)$. The Lebesgue measure being absolutely continuous with respect to the Gauss measure, (3.13) is also true almost everywhere with repect to the Lebesgue measure. The rationals p_n/q_n are the best approximations for α in the following sense: any rational p/q with a denominator q less than or equal to q_n is a less good approximation of α than p_n/q_n. The numbers q_n grow very fast with the index n. From its definition, Eq. (3.7), a_n is strictly positive and from Eq. (3.10) one gets the minoration

$$q_{n+1} = a_{n+1}q_n + q_{n-1} \geqslant 2q_{n-1}. \tag{3.14}$$

Since $q_0 = 1$, one has

$$q_n \geqslant 2^{n/2}. \tag{3.15}$$

For a majoration, from Khintchine[6] one has

$$\sup_n (q_n)^{1/n} < +\infty \text{ almost everywhere (in } \alpha). \tag{3.16}$$

We can come back to our recurrence. Let us suppose that, at time $(q_n + q_{n-1} - 1)$, just before entering the interval ($[q_{n-1}\alpha, 0]$ for n even and $[0, q_{n-1}\alpha]$ for n odd) of length δ_{n-1}, one finds in the circle only intervals with two lengths: δ_n or δ_{n-1}. The small ones (length δ_n) are to the right (resp. to the left) of the points $k\alpha$, k an integer, $0 \leqslant k < q_{n-1}$, for n even (resp. odd). These small intervals are separated from each other by large intervals of length δ_{n-1}. On Fig. 6 one shows the successive points of the series $k\alpha$ entering the interval $[q_{n-1}\alpha, 0]$ (resp. $[0, q_{n-1}\alpha]$).

The first point is $(q_n + q_{n-1})\alpha$. Each new point performs a jump of δ_n towards the origin; the lapse of time between two advents is q_n.

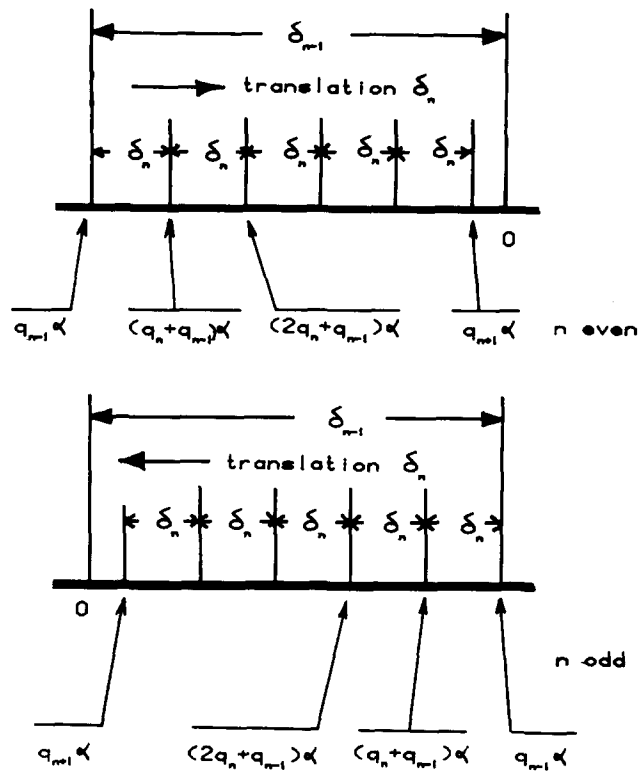


FIG. 6. Detail of the points of the series $k\alpha$ entering in the segment $[q_{n-1}\alpha, 0]$ in the case n even, and $[0, q_{n-1}\alpha]$ in the case n odd.

The point $k\alpha$ between these two times falls in the q_n large (length δ_{n-1}) intervals, reproducing exactly the same disposition of points as in the interval $[q_{n-1}\alpha,]$ (resp. $[0,q_{n-1}\alpha]$). At time $a_{n+1}q_n + q_{n-1} = q_{n+1}$, the interval $[q_{n-1}\alpha,0]$ (resp. $[0,q_{n-1}\alpha]$) contains a_{n+1} points, the other q_n large intervals $a_{n+1} - 1$. If you wait the time $(q_{n+1} + q_n - 1)$, you find two kinds of intervals, the small ones with length δ_{n+1} to the left (resp. to the right) of the points $k\alpha$, k an integer, $0 \leqslant k < q_n$, according to whether n is even or odd, and the big ones of length δ_n and the recurrence is done. A major difficulty in solving our diffusion problem is that if the successive points $k\alpha$, k an integer, $0 \leqslant k \leqslant N$, have gone around the circle several times it is not easy to say anything about the point $N\alpha$. The above description will permit us to replace this long rotation by successive rotations of lengths which are greater than one.

We shall use an expansion for an integer N already used[9] elsewhere. One takes the number q_M which is just less than N (i.e., $q_{M+1} > N$); the division of N by q_M gives

$$N = c_M q_M + N', \tag{3.17}$$

with $a_{M+1} \geqslant c_M \geqslant 0$, c_M an integer; $0 \leqslant N' < q_M$. By iteracting the same process on the remainder N', and so on, one gets an expansion for N of the form

$$N = \sum_{n=0}^{M} c_n\, q_n = \sum_{n=0}^{M(\alpha,N)} c_n\,(\alpha,N)\, q_n\,(\alpha). \tag{3.18}$$

This expansion looks like the decimal expansion of a number, but one has to note the following differences: the range of the digits (a_i) usually increases with the index i, and if $c_i = a_{i+1}$ then (3.10) implies that c_{i-1} disappears. To make this more clear, let us compare the α-expansion of the first integers with the decimal expansion in the case where all the a_i are equal to 9. By a–b we mean that in the α-expansion one finds the usual successive integers between a and b, as in their decimal expansion. The addition rules, Eq. (3.10), give the following α-expansion for the first integers

$$1,2\text{———}90,100,101\text{———}190,200,201, \tag{3.19}$$
$$\text{———}290,300,\cdots,890,900,901\text{———}909,1000,\cdots.$$

The rules of addition, Eq. (3.10), are not so simple as for the decimal expansion. The report of a carry, if one occurs, obeys different rules; in particular, this report affects two digits. In the case where all the a_i are equal to 9, we give the following example to compare with decimal addition

$$5672 + 3261 = 9032. \tag{3.20}$$

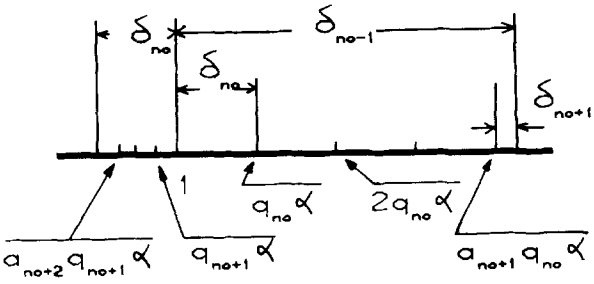


FIG. 7. Position of the points $k\alpha$, where in the α-expansion of k only digits of order higher than or equal to n_0 occur.

This expansion is very useful in determining where on the circle the point $N\alpha$ occurs; in fact, from Eq. (3.9), one has

$$N\alpha = \sum_{n=0}^{M} c_n\, q_n \left(\frac{p_n}{q_n} + (-)^n \frac{\delta_n}{q_n}\right) \pmod 1 \tag{3.21}$$
$$= \sum_{n=0}^{M} (-)^n c_n \delta_n \pmod 1. \tag{3.22}$$

Since the δ_n decrease fast with the index n, one sees that the position of the point $N\alpha$ depends principally on the low digits.

2. A majoration for the covered distance by the wandering particle at time N

We can now use the above notations to study the diffusion process. Equations. (2.4) will be summarized by

$$r_{4N} = \sum_{i=1}^{N} f(s_0 + i\alpha). \tag{3.23}$$

In Eq. (3.23) r_{4N} stands either for $\frac{1}{4}x_{4N}$ or $\frac{1}{4}y_{4N}$ and $f(s)$ is a periodic function of period unity, constant on intervals. One can use the expansion (3.18) of N to replace the sum (3.23) by:

$$r_{4N} = \sum_{M=1}^{c_M q_M} f(s_0 + n\alpha) + \sum_{n=1}^{c_1 q_1} f(s_0 + c_0\,\delta_0 + n\alpha)$$
$$+ \cdots + \sum_{n=1}^{c_M q_M} f(s_0 + \sum_{i=0}^{M-1} (-)^i c_i\,\delta_i + n\alpha). \tag{3.24}$$

Since the $\delta_n = \Pi_{i=0}^{n} T^i\alpha$ are less than one, Eq. (3.9) means that on each interval $[p/q_n, (p+1)/q_n]$, $q_n > p \geqslant 0$ q_n and p integers, one finds only a point of the series $k\alpha$ (mod 1), k an integer, $0 < k \leqslant q_n$. The function f being constant on intervals (see Eq. 2.4), one gets the following majoration

$$\left| \sum_{n=0}^{q_m} f(s + n\alpha) \right| \leqslant 2, \tag{3.25}$$

uniformly in s. By substituting (3.25) in (3.24) one obtains

$$|r_{4N}| \leqslant 2 \sum_{n=0}^{M} c_n. \tag{3.26}$$

The c_n enter in the α-expansion of the *time* N. Eq. (3.26) implies that when N belongs to the series of $\{q_i\}$ the wandering particle comes back near its departure position; in certain cases it comes back to the same rectangle. The sum (3.26) giving the diameter of the domain enclosing the particle has a serrated variation. By their definition, the c_n are less than a_{n+1}

$$|r_{4N}| \leqslant 2 \sum_{n=0}^{M} a_{n+1}, \tag{3.27}$$

and the sum on the right-hand side of Eq. (3.27) has been already evaluated by Kesten.[10] For every positive ϵ he gets the relation

$$\lim_{N\to\infty} P\left[\left| \frac{\Sigma_{n=0}^{M(N,\alpha)} a_{n+1}(\alpha)}{\log N \log\log N} - \frac{12}{\pi^2} \right| > \epsilon\right] = 0, \tag{3.28}$$

the measure P being the Lebesgue measure taken over α. The inequality *in measure* (3.28) implies only that there exists an infinite series N_i, i, an integer, such that

$$\sum_{n=0}^{M(N_i i,\alpha)} a_{n+1}(a) < \frac{12}{\pi^2}\log N_i \,\mathrm{loglog} N_i, \tag{3.29}$$

but one can find cases where this inequality is not verified, especially when an anomalous big a_i is encountered. This majoration implies an abnormal diffusion for the light particle. It is very hard to conjecture now if a normal diffusion could be found with other diffusors (for example with regular shapes like in the Sinaï's billards) or if a stochastic law for the diffusors' positions is necessary.

C. The wandering particle escapes

In what follows, we shall try to prove that although its diffusion is abnormal, the particle escapes to infinity.

The idea is to construct a series of N_i such that $r_{4\Sigma N_i}$ is unbounded. Let us symbolize the decomposition of a sum like in (3.24) in partial sums by

$$\sum_{n=1}^{N} = \sum_{n=1}^{c_0 q_0} + \sum_{n=1}^{c_1 q_1} + \cdots + \sum_{n=1}^{c_M q_M}. \tag{3.30}$$

Generally if we use the α-expansion of two numbers N_1 and $(N_1 + N_2)$ without case in the choice of N_1 and N_2, the decomposition in partial sums of $\Sigma_{n=1}^{N_1}$, $\Sigma_{n=1}^{N_2}$, and $\Sigma_{n=1}^{N_1+N_2}$ have no common part, because even the lower order digits are changed by an addition. The form given in (3.24) suggests finding a set of numbers $\{N_i\}$ such that $\Sigma_{n=1}^{N_i+N_j} = \Sigma_{n=1}^{N_i} + \Sigma_{n=1}^{N_j}$, in the sense of the decomposition (3.24). A set which has this property may be constructed with numbers N_i such that if $j > i$ the α-expansion of N_j contains only digits of strictly greater order than the α-expansion of N_i. Moreover, we shall select those sets of numbers for which the partial sums of (3.24) all have the same sign and are not zero.

If an infinite sequence of numbers N_i satisfying the properties described above exists, the light particle escapes to infinity, since

$$\left| r_{4\Sigma_1^{i} N_i} \right| \geqslant k. \tag{3.31}$$

As they play an important role in our demonstration, let us consider the integers k such that

$$k = \sum_{n > n_0} c_n q_n, \quad 0 \leqslant c_n \leqslant a_{n+1}. \tag{3.32}$$

The numbers k can be set in increasing order $(k_1, k_2, \cdots)$; two successive k's are generally separated by q_{n_0}, but sometimes by q_{n_0-1}. The corresponding points $k\alpha$ belong to the segment $[(-)^{n_0+1}\delta_{n_0}, (-)^{n_0}\delta_{n_0-1}]$ and are the iterates of a translation of $(-)^{n_0}\delta_{n_0}$, starting from the origin (see Fig. 7). The case where the lapse of time between two succsessive k's satisfying (3.32) is q_{n_0-1} corresponds to a jump over the point $(-)^{n_0+1}\delta_{n_0}$ [or $(-)^{n_0}\delta_{n_0-1}$] since we can identify these two points and consider the segment where the points k satisfying (3.22) fall as a circle).

As soon as one knows where the point $k_{i+1}\,\alpha$ occurs in this circle, the sum

$$\sum_{n=k_i+1}^{k_{i+1}} f(s + n\alpha)$$

is determined. From (3.25) this sum can have one of the values ± 2, ± 1, or 0 and is a function of the same kind as f; it is constant on intervals and of period $\delta_{n_0} + \delta_{n_0-1}$. By a change of scale one can use a function of period unity

$$\sum_{n=0}^{q_{n_0}} f(s_0 + n\alpha) = g(s'), \tag{3.33}$$

(where s' is a point between zero and one) to write

$$\sum_{n=0}^{k_i} f(s + n\alpha) = \sum_{n=o}^{i} g\left(s' + n\frac{T^{n_0}\alpha}{1 + T^{n_0}\alpha}\right). \tag{3.34}$$

The number α being irrational, it is easy to see that $-T^{n_0}\alpha/(1 + T^{n_0}\alpha)$ is also irrational and less than unity, and that the eight points which determine the four intervals I_1, I_2, I_3, I_4 [see Eq.(2.2)] do not belong to the same series of points $s'' + k\alpha$, k an integer, s'' belonging to the unit circle, and therefore, there exist nonnull intervals where the function g is not equal to zero.

We are now able to find series of N_i such that $r_{4\Sigma_i N_i}$ is unbounded.

The set of points $s_0 + n\alpha$, n an integer, being dense on the circle, the first N_0 such that

$$\sum_{n=0}^{N_0} f(s_0 + n\alpha) \neq 0 \tag{3.35}$$

exists and is finite. As in (3.18), one expands N_0

$$N_0 = \sum_{n=0}^{M_0} c_n q_n. \tag{3.36}$$

We shall denote the sign of (3.35) by ϵ_0. Let us consider the numbers k satisfying

$$k = \sum_{n > M_0} c_n q_n. \tag{3.37}$$

They can be arranged in increasing order k_i, i an integer. By Eq. (3.34), there is a point s', a function g of period unity constant on intervals, and a number $\alpha' = T^{M_0+1}\alpha/(1 + T^{M_0+1}\alpha)$ such that

$$\sum_{n=0}^{k_i} f(s_0 + \sum_{n=0}^{M_0} (-)^n c_m \delta_m + n\alpha) = \sum_{n=1}^{i} g(s' + n\alpha'). \tag{3.38}$$

The set of points $s' + n\alpha'$, n an integer, being dense on the unit circle, the first N_1 such that the sum in (3.38) is not zero exists and is finite. The number N_1 belongs to the set of k_i satisfying (3.32), with $n_0 = M_0 + 1$. Then the α-expansion of N_1 gives

$$N_1 = \sum_{n=M_0+1}^{M_1} c_n q_n. \tag{3.39}$$

Let us remark that the same coefficients c_n are used for N_1 and N_0 since the indices n are always different for the N_0 and N_1 α-expansions. One denotes the sign of the sum (3.38) by ϵ_1. Using the same process repeatedly, one gets a series of N_i i an integer, such that

$$\sum_{i \leqslant I} N_i = \sum_{n=0}^{M_I} c_n q_n. \tag{3.40}$$

If one separates the N_i in two sets corresponding to positive and negative ϵ_i one of the sets is infinite; we shall denote the N_i belonging to it by $N_{i_1}, N_{i_2}, \ldots, N_{i_n}$. The N_{i_j} being independent in the sense of (3.40) (their α-expansions do not contain

the same orders n), one gets

$$\left| \sum_{n=0}^{N_j} f(s_0 + n\alpha) \right| \geqslant j, \tag{3.41}$$

which proves that the wandering particle does not remain in a bounded domain.

The authors are gratefully indebted to M. Mendès France of the University of Bordeaux for very useful suggestions.

[1]E.H. Hauge, *Lecture notes in physics* N°. 31 (Springer, Berlin/Vienna, 1974).
[2]Ya.G. Sinaĭ, Russ. Math. Surv. **25**, 137 (1970).
[3]G. Gallovotti, *Lecture notes in physics* N°. **38**, 236 (Springer, Berlin/Vienna, 1974).
[4]M. Keane, Math. Z. **141**, (1975); Isr. J. Math. **26**, 188 (1977).
[5]J. Suranyi, Ann. Univ. Sci. Budap. Rolando Eotvas **1**, 107 (1958).
[6]J.F. Koksma, *Diophantische approximationen* (Chelsea, New York, 1936); and A. Khintchine, *Continued fractions* (Noordhoff, Groningen, The Netherlands, 1963).
[7]Y.V. Linnik, Lecture notes in math No. 45 (1968).
[8]V.I. Arnold and A. Avez, *Problèmes ergodiques de la mécanique classique* (Gauthier-Villars, Paris, 1967; Benjamin, New York, 1968).
[9]H. Kesten, Acta Arithmetica **XII, 2**, 193 (1966).
[10]H. Kesten, Acta Arithmetica **X**, 183 (1964). See also P. Levy, Comp. Math. **3**, 286 (1936).

The condensed phase of the imperfect Bose gas

M. Fannes and A. Verbeure
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

(Received 5 December 1979; accepted for publication 8 March 1980)

Correlation inequalities are used to prove rigorously the following for the imperfect Bose gas: (i) existence of condensation in and only in the ground state, (ii) unicity and existence of the limiting Gibbs states.

I. INTRODUCTION

Bose–Einstein condensation occurs in the exactly soluble model of the free Bose gas at low enough temperature. Furthermore, it is a general belief that interacting Bose systems will show a macroscopic occupation of the ground state. A way of expressing this belief is assuming that the so called Bogoliubov approximation is valid. Therefore, the condensation phenomenon is of interest. Moreover, the study of condensation is fundamental because it is a clue to our understanding of phase transitions.

As far as exact results are concerned, the Bose systems considered are those with superstable interactions.[1-3] This stability criterion ensures the existence of an equilibrium state for any value of the chemical potential. This condition excludes the free Bose system and also all mean field type interactions. In this work we treat the representative of the latter type, namely, the so called imperfect Bose gas,[4] representing one of the simplest models of an interacting Bose system showing condensation. Apart from a variety of physically intuitive arguments (see for example, Ref. 4), there exists a rigorous proof[5] of the existence of a singularity in the mean density as a function of the chemical potential.

This proof relies entirely on probabilistic methods. In this paper we use mainly the recently studied correlation inequalities[6] for equilibrium states. In the first place these inequalities provide a proof of condensation in and only in the ground state for sufficiently low temperature or sufficiently high density. This proof does not require first a complete solution of the problem. Therefore, it is hoped that a similar approach is successful for nonexactly soluble models. The second main result consists in showing the existence of the limiting Gibbs state and deriving the complete solution. Again this result is based on the proof that the correlation inequalities hold in the thermodynamic limit.

The appropriate setting of our system is the following. We consider only ν-dimensional systems with $\nu \geqslant 3$, as for $\nu < 2$ condensation is absent.[7] Denote by Λ the centered cubic box of side length L. A system of identical bosons in the box Λ is described by the algebra $\mathscr{A}_\Lambda$ of all bounded operators on the Fock space $\mathscr{H}_\Lambda$.

The algebra of local observables is then $\mathscr{A}_{\text{loc.}} = \cup_{\Lambda \in R^\nu} \mathscr{A}_\Lambda$ and the C^* algebra of the quasilocal observables $\mathscr{A}$ is obtained by taking the C^* completion of $\mathscr{A}_{\text{loc.}}$. An appropriate subalgebra to describe the commutation relations is the CCR algebra generated by the Weyl operators $\{ W(\phi) | \phi \in \mathscr{D} \}$, where $\mathscr{D}$ is the space of infinitely differentiable functions on R^ν with compact support and

$$W(\phi) = \exp i[a(\phi) + a^*(\phi)],$$

where a^* and a are the Fock creation and annihilation operators, respectively.

The model is specified by the local Hamiltonians H_L, with periodic boundary conditions

$$H_L = T_L - \mu_L N_L + \frac{\lambda}{2} \frac{N_L^2}{L}, \tag{1}$$

where

$$0 < \lambda \in R,$$

$$\tilde{\Lambda} = \left\{ k \in R^\nu \,\middle|\, \frac{kL}{2\pi} \in Z^\nu \right\},$$

$$N_L = \sum_{k \in \tilde{\Lambda}} a_{L,k}^+ a_{L,k},$$

$$a_{L,k} = a(\phi_k^L), \quad k \in \tilde{\Lambda},$$

$$\phi_k^L(x) = \frac{1}{L^{\nu/2}} e^{ikx}, \quad x \in \Lambda,$$

$$= 0, \quad x \notin \Lambda,$$

$$T_L = \sum_{k \in \tilde{\Lambda}} \epsilon(k) a_{L,k}^+ a_{L,k}, \quad \epsilon(k) = k^2/2.$$

Further on we use also the notations

$$N(\phi) = a^*(\phi) a(\phi) \quad \text{and} \quad N_{L,k} = N(\phi_k^L).$$

We study the equilibrium state of the system in the grand canonical ensemble. The key technique is the equivalence of the equilibrium conditions with the following correlation inequalities[6]:

$$\beta \omega_L(X^*[H_L, X]) \geqslant \omega_L(X^*X) \ln \frac{\omega_L(X^*X)}{\omega_L(XX^*)} \tag{2}$$

for all $X \in \mathscr{A}_\Lambda$ belonging to the domain of $[H_L, .]$, supplemented by the condition

$$\omega_L(N_L) = \rho L^\nu, \tag{3}$$

where ρ is any positive number standing for the mean density; ω_L is the canonical state of $\mathscr{A}_\Lambda$ determined by H_L. Note that $\exp(-\beta H_L)$ is trace class for all values of μ_L. Indeed, for all $-\mu_0 \in R^+$,

$$H_L = T_L - \mu_0 N_L + \frac{\lambda}{L^\nu} \left[N_L - \frac{(\mu_L - \mu_0) L^\nu}{2\lambda} \right]^2 - \frac{L^\nu}{4\lambda} (\mu_L - \mu_0)^2.$$

Furthermore, the state extends to any polynomial in the creation and annihilation operators. This can be used to extend the correlation inequalities to unbounded observables.

 0022-2488/80/071809-10$1.00

II. BOUNDS ON THE CORRELATION FUNCTION AND CONDENSATION

Lemma II.1: If $j \neq k_i$ $(i = 1,...,n)$, $k_i \neq k_{i'}$ for $i \neq i'$ and $n_i > 0$, $n_i \in N$, then

$$e^{|\epsilon(j)-\epsilon(k)|}\omega_L(N_{L,j}(N_{Lk_1}+1)^{n_1}(N_{Lk_2})^{n_2}\cdots(N_{Lk_m})^{n_m}) = \omega_L((N_{Lj}+1)(N_{L,k_1})^{n_1}\cdots(N_{L,k_m})^{n_m}).$$

Proof: It follows trivially from Eq. (2) by taking X successively equal to

$$a^*_{L,j}a_{L,k_1}(N_{L,k_1}{}^{n_1-1}N_{L,k_2}{}^{n_2}\cdots N_{L,k_n}{}^{n_n})^{1/2}$$

and

$$a^*_{L,k_1}a_{L,j}(N_{L,k_1}{}^{n_1-1}\cdots N_{L,k_n}{}^{n_n})^{1/2}. \quad \blacksquare$$

Lemma II.2: There exists a constant $C_1(\nu)$ such that

$$\omega_L(N_{L,j}) \leqslant \frac{\rho C_1(\nu)}{|j|^\nu}.$$

Proof: From Lemma II.1, it follows immediately that $|j| \to \omega_L(N_{L,j})$ is decreasing, and hence

$$\rho \geqslant \frac{1}{L^\nu}\sum_{|j|\geqslant|k|}\omega_L(N_{L,k}) \geqslant \frac{1}{L^\nu}\left\{\sum_{|k|\leqslant|j|}1\right\}\omega_L(N_{L,j}) \geqslant C_1(\nu)|j|^\nu\omega_L(N_{L,j}). \quad \blacksquare$$

Lemma II.3: For each $n \in N$,

(i)

$$\beta\omega_L\left(-\epsilon(k)N^{n+1}_{L,k}+(\mu_L-\lambda n_L)N^{n+1}_{L,k}+\frac{\lambda}{2}\frac{N^{n+1}_{L,k}}{L^\nu}\right) \geqslant \omega_L(N^{n+1}_{L,k})\ln\frac{\omega_L(N^{n+1}_{L,k})}{\omega_L[(N_{L,k}+1)^{n+1}]},$$

where $n_L = N_L/L^\nu$;

(ii)

$$\omega_L((\mu_L-\lambda n_L)N^n_{L,k}) \leqslant \omega_L\left(\frac{\lambda N_{L,j}N^n_{L,k}}{L^\nu}+\frac{\lambda N^n_{L,k}}{2L^\nu}+\epsilon(j\,N^n_{L,k})\right).$$

Proof: For (i) the result follows by putting

$$X = a_{L,k}N^{n/2}_{L,k}$$

in Eq. (2). One gets (ii) by taking

$$X = a_{L,j}\,N^{n/2}_{L,k}$$

in the inequality

$$\omega_L((X^*,[H_L,X\,]))\geqslant 0,$$

which itself is immediate from Eq. (2).

Lemma II.4: For $\nu \geqslant 3$ there exists a constant C_2 depending on λ and β, such that $\omega_L(N_{L,k})$ and $\omega_L(N^2_{L,k})^{1/2}$ are majorized by

$$\frac{\rho C_2}{|k|^{\max\{\nu/3,2\}+(2/3)}}$$

for $|k| \neq 0$.

Proof: We look for an upper bound of $\omega_L(N^3_{L,k})$ and then use the Hölder inequality in the form

$$\omega_L(N_{L,k}) \leqslant \omega_L(N^3_{L,k})^{1/3}, \quad \omega_L(N^2_{L,k})^{1/2} \leqslant \omega_L(N^3_{L,k})^{1/3}.$$

Using the inequality

$$a\ln\frac{a}{b} \geqslant a-b, \quad \text{if } 0 \leqslant a \leqslant b,$$

Lemma II.3(i) becomes

$$\beta\omega_L\left(-\epsilon(k)N^{n+1}_{L,k}+(\mu_L-\lambda n_L)N^{n+1}_{L,k}+\frac{\lambda}{2}\frac{N^{n+1}_{L,k}}{L^\nu}\right) \geqslant \omega_L(N^{n+1}_{L,k}-(N_{L,k}+1)^{n+1}). \tag{4}$$

Now we work on Lemma II.3(ii) to get

$$\omega_L[(\mu_L-\lambda n_L)N^{n+1}_{L,k}] \leqslant \omega_L\left(\frac{\lambda N_{L,0}N^{n+1}_{L,k}}{L^\nu}+\frac{\lambda N^{n+1}_{L,k}}{2L^\nu}\right), \tag{5}$$

by putting $j = 0$. Repeatedly using Lemma II.1 in the form

$$e^{-\beta\epsilon(k)}\omega_L[N_{L,0}(N_{L,k}+1)^{n+1}] \geqslant \omega_L(N_{L,0}N^{n+1}_{L,k}) \tag{6}$$

one gets

$$\frac{\omega_L(N_{L,0}N^{n+1}_{L,k})}{L^\nu} \leqslant \frac{\alpha_n\rho}{|k|^{2n+2}}, \quad \alpha_n \in R. \tag{7}$$

By inserting (7) in (5) and (5) in (4) one gets an inequality for $\omega_L(N^{n+1}_{L,k})$ in terms of $\omega_L(N^m_{L,k})$ $(m \leqslant n)$. By iteration of this procedure one arrives at an upper bound for $\omega_L(N^{n+1}_{L,k})$ in terms of $\omega_L(N_{L,k})$ and the inverse powers of $|k|$. Finally, using Lemma II.2 and specifying $n = 2$, one gets the Lemma. $\blacksquare$

Lemma II.5:For each $k \in \tilde{\Lambda}$ one has

$$\mu_L \leqslant \lambda\rho + \epsilon(k) + \frac{\lambda}{L^\nu}(\omega_L(N_{L,k})+\tfrac{1}{2})$$

and

$$\mu_\infty = \overline{\lim_{L\to\infty}}\,\mu_L \leqslant \lambda\rho.$$

Proof: Take $X = a^*_{L,k}$ in

$$\omega_L\{[X^*,[H_L,X\,]]\} \geqslant 0.$$

The last statement follows by using Lemma II.4 and by taking

$$\lim_{|k|\to 0}\overline{\lim_{L\to\infty}}. \quad \blacksquare$$

Lemma II.6: For $\lambda \neq 0$,

$$\omega_L\left[\left(\frac{N_{L,0}}{L^\nu}\right)^2\right] \leqslant \frac{\mu_L\rho}{\lambda}+\left(\frac{1}{\lambda\beta}+\frac{\rho}{2}\right)\frac{1}{L^\nu}.$$

Proof: Use Lemma II.3(i) with $k = 0$ and $n = 0$, and with the lower bound for the right hand side equal to -1. Using

$$\omega_L\left[\left(\frac{N_{L,0}}{L^\nu}\right)^2\right] \leqslant \omega_L\left(n_L\frac{N_{L,0}}{L^\nu}\right),$$

the lemma follows. $\blacksquare$

Note that the upper bound in Lemma II.6 is not valid for the free Bose gas. However, one can, using Lemma II.5, take the limit $\lambda \to 0$ after having performed the thermodynamic limit. Note that $\lim_{\lambda\to 0}$ and $\lim_{L\to\infty}$ are not interchangeable (see later on).

The upper bound of Lemma II.4 is not good enough at

large momenta for our purposes; we now derive a better bound.

Lemma II.7: For each $\delta > 0$ and L large enough, there exists a constant C_3 independent of δ and L, such that for $|k| \geqslant \delta$:

$$\omega_L(N_{L,k}) \leqslant \frac{1}{e^{c_k(L)} - 1} + b_k(L)\frac{1}{1 - e^{-c_k(L)}},$$

$$b_k(L) = \frac{C_3}{\delta^\nu L^{\nu/2}(e^{\beta\epsilon(k)} - 1)},$$

$$c_k(L) = \beta\left(\epsilon(k) - \frac{\lambda}{L^\nu} - \frac{\delta^2}{L}\right).$$

Proof: Choose j such that $|j| = \delta L^{-1/2}$ in Lemma II.1:

$$\omega_L(N_{L,j}\, N_{L,k}) \leqslant \frac{\omega_L(N_{L,j})}{e^{\beta[\epsilon(k) - \epsilon(j)]} - 1}.$$

By Lemma II.2, there exists a constant α such that

$$\frac{\omega_L(N_{L,j}\, N_{L,k})}{L^\nu} \leqslant \frac{\alpha}{\delta^\nu L^{\nu/2}(e^{\beta\epsilon(k)} - 1)}. \tag{8}$$

By Lemma II.3(ii) with $n = 1$:

$$\omega_L(\mu_L N_{L,k} - \lambda n_L N_{L,k}) \leqslant \frac{\lambda\alpha}{\delta^\nu L^{\nu/2}(e^{\beta\epsilon(k)} - 1)} + \omega_L(N_{L,k})\left(\frac{\lambda}{2L^\nu} + \frac{\delta^2}{L}\right). \tag{9}$$

Substitution of (8) and (9) in Lemma II.3(i) yields

$$-c_k(L)\omega_L(N_{L,k}) + b_k(L) \geqslant \omega_L(N_{L,k}) \ln\frac{\omega_L(N_{L,k})}{\omega_L(N_{L,k} + 1)},$$

where $b_k(L)$ and $c_k(L)$ are as above, and $C_3 = \lambda\alpha$. The Lemma follows from convexity arguments in the right-hand side. ∎

Lemma II.8: For each $\delta > 0$ and L large enough, one has for $|k| \geqslant \delta$:

$$\omega_L[(N_{L,k})^2] \leqslant \frac{1}{(e^{(1/2)c_k(L)} - 1)^2} + b'_k(L)\frac{1}{(1 - e^{-(1/2)c_k(L)})},$$

where

$$b'_k(L) = \frac{2C_3}{(e^{\beta\epsilon(k)} - 1)^2 L^{\nu/2}\delta^\nu} + \frac{C_3}{(e^{\beta\epsilon(k)} - 1)L^{\nu/2}\delta^\nu},$$

$$c_k(L) = \beta\left(\epsilon(k) - \frac{\lambda}{L^\nu} - \frac{\delta^2}{L}\right),$$

and where C_3 is the same constant as in Lemma II.7.

Proof: The proof goes along the same lines as that of Lemma II.7. One has, however, to use Schwartz's inequality in the right-hand side of Lemma II.3 with $n = 1$ to get

$$\omega_L[(N_{L,k})^2] \ln\frac{\omega_L[(N_{L,k})^2]}{\omega_L[(N_{L,k} + 1)^2]} \geqslant \omega_L[(N_{L,k})^2] \ln\frac{\omega_L[(N_{L,k})^2]}{\{\omega_L[(N_{L,k})^2]^{1/2} + 1\}}.$$ ∎

Now we prove the existence of condensation in the thermodynamic limit $L \to \infty$ taken with constant particle density ρ. The particular way the limit $L \to \infty$ is taken is irrelevant for condensation.

Theorem II.9: One has

(i) $$\lim_{L\to\infty} \frac{\omega_L(N_{L,0})}{L^\nu} \geqslant \rho - (2\pi)^{-\nu}\int_{R^\nu} dk\, \frac{1}{e^{\beta\epsilon(k)} - 1};$$

(ii) $$\lim_{\delta\to 0}\overline{\lim_{L\to\infty}} \sum_{0<|k|<\delta} \frac{\omega_L(N_{L,k})}{L^\nu} = 0;$$

(iii) for any $\rho > 0$, there exist a $\beta_c > 0$ such that for all $\beta > \beta_c$:

$$0 < \overline{\lim_{L\to\infty}} \frac{\omega_L(N_{L,0})}{L^\nu} \leqslant \overline{\lim_{L\to\infty}} \frac{\omega_L(N_{L,0})}{L^\nu} \leqslant \rho.$$

Proof: From (3)

$$\frac{\omega_L(N_{L,0})}{L^\nu} = \rho - \sum_{k\neq 0} \frac{\omega_L(N_{L,k})}{L^\nu}.$$

For any $\delta > 0$, split the k summation as follows:

$$\sum_{k\neq 0} = \sum_{0<|k|<\delta} + \sum_{|k|\geqslant\delta}.$$

Using the estimates of Lemmas II.4 and II.7 and taking the limit $L \to \infty$, one gets

$$\overline{\lim_{L\to\infty}} \frac{\omega_L(N_{L,0})}{L^\nu} \geqslant \rho - \rho C_3(2\pi)^{-\nu}\int_{|k|<\delta} \frac{dk}{|k|^{\max\{\nu/3,2\} + (2/3)}} - (2\pi)^{-\nu}\int_{|k|\geqslant\delta} \frac{dk}{e^{\beta\epsilon(k)} - 1},$$

which is a senseful bound, because $\nu \geqslant 3$. Now let δ tend to zero to get (i) and (ii). Furthermore, (iii) is immediate.

Note that Theorem II.9(iii) proves the condensation in the zeroth mode at sufficiently low temperature. Furthermore, Theorem II.9(ii) shows that the zeroth mode is the only one with macroscopic occupation.

Note also that the lower bound obtained for the critical temperature is the same as for the free Bose gas.[8]

III. THE LIMIT-CORRELATION INEQUALITY

The ultimate aim of this section is to obtain the correlation inequalities in the thermodynamic limit. To do so we first need more information about the structure of the representations induced by the limiting Gibbs states.

Lemma III.1: For any $\phi \in \mathscr{D}$ let $(1/2\pi)^{\nu/2}\int dx\, \phi(x)e^{ikx} = \tilde{\phi}(k)$; then

$$\overline{\lim_{L\to\infty}} |\omega_L(1 - W(\phi))|^2 \leqslant 2\left((2\pi)^\nu\bar{\rho}_0|\tilde{\phi}(0)|^2 + \int \frac{dk\,|\tilde{\phi}(k)|^2}{e^{\beta\epsilon(k)} - 1}\right) + \|\phi\|^2,$$

where

$$\bar{\rho}_0 = \overline{\lim_{L\to\infty}} \omega_L(N_{L,0})L^{-\nu}.$$

Proof: Using the Schwartz inequality and gauge invariance,

$$|\omega_L(1 - W(\phi))|^2 \leqslant \omega_L(2 - W(\phi) - W(-\phi)) = 2\omega_L(1 - \cos(a(\phi) + a^*(\phi))) = 4\omega_L\left(\sin^2\frac{a(\phi) + a^*(\phi)}{2}\right)$$

$$\leqslant \omega_L([a(\phi)+a^*(\phi)]^2)$$
$$= 2\omega_L(a^*(\phi)a(\phi)) + \|\phi\|^2.$$

Now using the estimates of Lemmas II.4 and II.7, one gets the result. ■

Lemma III.1 guarantees the existence of sufficiently many fields in the representation of any W^*-limit point of Gibbs states as $L\to\infty$ obtained by a net $(L_\alpha)_\alpha$. Let ω_β be such a limit point and $(\pi_\beta,\Omega_\beta,\mathscr{H}_\beta)$ its GNS triplet; denote also $\mathscr{M}_\beta = \pi_\beta(\mathscr{A})''$.

From Lemma III.1 we get the continuity of the map $t\in R\to\pi_\beta[W(t\phi)]$ for all $\phi\in\mathscr{D}$. Hence, by Stone's theorem there exists a self-adjoint field B_β such that

$$\pi_\beta(W(\phi)) = \exp iB_\beta(\phi)$$

and that each $B_\beta(\phi)$ is affiliated with the von Neumann algebra $\mathscr{M}_\beta$. The creation and annihilation operators, respectively, are then obtained in the standard way

$$a^*_\beta(\phi) = \tfrac{1}{2}[B_\beta(\phi) - iB_\beta(i\phi)],$$
$$a_\beta(\phi) = \tfrac{1}{2}[B_\beta(\phi) + iB_\beta(i\phi)].$$

Let $\bar{\mathscr{D}}$ be the completion of $\mathscr{D}$ with respect to the norm

$$|||\phi|||^2 = |\tilde\phi(0)|^2 + \int_{R^\nu} dk\,(1+(1/|k|^2))|\tilde\phi(k)|^2, \qquad \phi\in\mathscr{D}.$$

It is clear that $\bar{\mathscr{D}}$ is the direct sum of C with $L_2[\tilde R,(1+(1/|k|^2))\,dk]$. Its elements are denoted by $(\alpha,\tilde\phi)$ and $|||(\alpha,\tilde\phi)|||^2 = |\alpha|^2 + \int_{R^\nu} dk\,(1+(1/|k|^2)) \times|\hat\phi(k)|^2$. Also from Lemma III.1, for any sequence $(\phi_n)_{n\in N}$ of elements ϕ_n of $\mathscr{D}$ converging to $(\alpha,\phi)\in\bar{\mathscr{D}}$, the one-parameter groups of unitaries $\{\exp itB(\phi_n)|t\in R\}_{n\in N}$ converge strongly to a strongly continuous one-parameter group $\{\exp itB_\beta(\alpha,\phi)|t\in R\}$ of unitaries; where $B_\beta(\cdot,\cdot)$ stands for the extension of the self-adjoint fields to $\bar{\mathscr{D}}$. In the same way we extend the notation for the creation and annihilation operators.

Now we proceed with the study of the behavior of particle densities in the limit $L\to\infty$.

Lemma III.2:With the above notation, for all elements x, y in $\mathscr{A}_{\text{loc.}}$ one has the following:

(i) $$\lim_{L'\to\infty}\overline{\lim_{L\to\infty}}\,\omega_L\left[\left(\frac{N_L}{L^\nu}-\frac{N_{L'}}{L'^\nu}\right)^2\right]=0,$$

(ii) $$\lim_{L'\to\infty}\overline{\lim_{L\to\infty}}\,\left|\omega_L x[(e^{it\sum L'/L'^\nu},y)]\right|=0,$$

(iii) $$\lim_{t\to 0}\lim_{L'\to\infty}\overline{\lim_{L\to\infty}}\left|\omega_L\left[x(e^{it\sum L'/L'^\nu}-1)y\right]\right|=0,$$

(iv) for all $\epsilon>0$ and L', L'' large enough

$$\overline{\lim_{L\to\infty}}\left|\omega_L\left[x(e^{itN_{L''}/L''^\nu}-e^{itN_{L'}/L'^\nu})y\right]\right|<\epsilon,$$

(v) $$\lim_{L'\to\infty}\overline{\lim_{L\to\infty}}\left|\omega_L\left[x\left(\frac{N_L}{L^\nu}-\frac{N_{L'}}{L'^\nu}\right)y\right]\right|=0.$$

Proof:(i) One calculates

$$\frac{N_{L'}}{L'^\nu}$$

$$= \frac{1}{L'^\nu}\sum_{k'\in\Lambda'} a^*(\phi^{L'}_{k'})a(\phi^{L'}_{k'})$$
$$= \frac{1}{L'^\nu}\sum_{k'\in\Lambda'}\sum_{l_1,l_2\in\tilde\Lambda}(\phi^L_{l_1},\phi^{L'}_{k'})(\phi^{L'}_{k'},\phi^L_{l_2})a^*(\phi^L_{l_1})a(\phi^L_{l_2})$$
$$= \frac{1}{L'^\nu}\sum_{l_1,l_2\in\tilde\Lambda}(\phi^L_{l_1}\chi_{\Lambda'},\phi^L_{l_2})a^*(\phi^L_{l_1})a(\phi^L_{l_2}),$$

where χ_Λ is the characteristic function of the volume Λ. For $\Lambda'\subset\Lambda$,

$$\omega_L\left(\frac{N_L}{L^\nu}\frac{N_{L'}}{L'^\nu}\right)$$
$$= \frac{1}{L^\nu L'^\nu}\sum_{l_1,l_2\in\tilde\Lambda}(\phi^L_{l_1}\chi_{\Lambda'},\phi^L_{l_2})\omega_L[N_L a^*(\phi^L_{l_1})a(\phi^L_{l_2})]$$
$$= \frac{1}{L^{2\nu}}\sum_{l\in\Lambda}\omega_L[N_L a^*(\phi^L_l)a(\phi^L_l)] = \omega_L\left[\left(\frac{N_L}{L^\nu}\right)^2\right].$$

Furthermore,

$$\omega_L\left[\left(\frac{N_{L'}}{L'^\nu}\right)^2\right]$$
$$= \frac{1}{L'^{2\nu}}\sum_{l_1,l_2,l_3,l_4\in\tilde\Lambda}(\phi^L_{l_1}\chi_{\Lambda'},\phi^L_{l_2})$$
$$\times(\phi^L_{l_3}\chi_{\Lambda'},\phi^L_{l_4})\omega_L[a^*(\phi^L_{l_1})a(\phi^L_{l_2})a^*(\phi^L_{l_3})a(\phi^L_{l_4})].$$

Using gauge invariance,

$$\omega_L\left[\left(\frac{N_{L'}}{L'^\nu}\right)^2\right]$$
$$= \frac{1}{L'^{2\nu}}\sum_{l_1,l_2\in\tilde\Lambda}(\phi^L_{l_1}\chi_{\Lambda'}\phi^L_{l_1})(\phi^L_{l_2}\chi_{\Lambda'}\phi^L_{l_2})$$
$$\times\omega_L(N_{L,l_1}N_{L,l_2}) + \frac{1}{L'^{2\nu}}$$
$$\times\sum_{\substack{l_1,l_2\in\tilde\Lambda\\ l_1\neq l_2}}|(\phi^L_{l_1}\chi_{\Lambda'}\phi^L_{l_2})|^2\omega_L[N_{L,l_1}(N_{L,l_2}+1)]$$
$$= \omega_L\left[\left(\frac{N_L}{L^\nu}\right)^2\right] + \frac{1}{L'^{2\nu}}\sum_{\substack{l_1,l_2\in\tilde\Lambda\\ l_1\neq l_2}}|(\phi^L_l\chi_{\Lambda'}\phi^L_{l_2})|^2$$
$$\times\omega_L(N_{L,l_1}N_{L,l_2}) + \left(\frac{1}{L'^\nu}-\frac{L'^\nu}{L^\nu}\right)\rho.$$

Therefore, we get the following bound:

$$\overline{\lim_{L\to\infty}}\,\omega_L\left[\left(\frac{N_L}{L^\nu}-\frac{N_{L'}}{L'^\nu}\right)^2\right]$$
$$\leqslant\frac{\rho}{L'^\nu}+\overline{\lim_{L\to\infty}}\frac{1}{L'^{2\nu}}\sum_{\substack{l_1,l_2\in\tilde\Lambda\\ l_1\neq l_2}}|(\phi^L_{l_1}\chi_{\Lambda'}\phi^L_{l_2})|^2\omega_L(N_{L,l_1}N_{L,l_2})$$
$$\leqslant\frac{\rho}{L'^\nu}+\overline{\lim_{L\to\infty}}\frac{1}{L'^{2\nu}}\sum_{\substack{l_1,l_2\in\tilde\Lambda\\ l_1\neq l_2}}|(\phi^L_{l_1}\chi_{\Lambda'}\phi^L_{l_2})|^2\omega_L(N^2_{L,l_1})^{1/2}$$
$$\times\omega_L(N^2_{L,l_2})^{1/2}.$$

Using now the estimates of Lemmas II.4, 5, 6, and 8, one gets

$$\overline{\lim_{L\to\infty}}\frac{\omega_L(N^2_{L,0})^{1/2}}{L^\nu}\leqslant\rho$$

and there exists a function F element of $L^1(R^\nu,dk)$ such that

$$\overline{\lim_{L\to\infty}}\ \omega_L(N_{L,k}^2)^{1/2}\leqslant F(k),\quad \text{for } k\neq 0.$$

Hence,

$$\overline{\lim_{L\to\infty}}\ \omega_L\left[\left(\frac{N_L}{L^\nu}-\frac{N_{L'}}{L'^\nu}\right)^2\right]$$
$$\leqslant\frac{\rho}{L'^\nu}+2\rho\int_{R^\nu}dk\,F(k)\left|\frac{1}{L'^\nu}\int_{\Lambda'}dx e^{ikx}\right|^2$$
$$+\int_{R^\nu\times R^\nu}dk_1\,dk_2\,F(k_1)F(k_2)$$
$$\times\left|\frac{1}{L'^\nu}\int_{\Lambda'}dx\,e^{i(k_2-k_1)x}\right|^2.$$

Using Lebesgue dominated convergence, the right-hand side tends to zero as $L'\to\infty$ because the means tend almost everywhere to zero.

(ii) Take $y=W(\phi)$; $\phi\in\mathscr{D}$; for L' large enough, the support of ϕ is contained in Λ'; then

$$[e^{itN_{L'/L'^\nu}},W(\phi)]=e^{itN_{L'/L'^\nu}}[W(\phi)-W(e^{-i(t/L'^\nu)}\phi)]$$

and

$$\left|\omega_L(x[e^{itN_{L'/L'^\nu}},W(\phi)])\right|$$
$$\leqslant\|x\|\omega_L(|W(\phi)-W[e^{-it/L'^\nu}(\phi)]|^2)^{1/2}$$
$$\leqslant 2^{1/2}\|x\|\left|1-\omega_L[e^{-i\sin(t/L'^\nu)}W((e^{it/L'^\nu}-1)\phi)]\right|^{1/2}$$

and (ii) follows, using the estimate of Lemma III.1.

Proof of (iii): By (ii) it is sufficient to prove that

$$\lim_{t\to 0}\lim_{L'\to\infty}\overline{\lim_{L\to\infty}}\left|\omega_L[x(e^{itN_{L'/L'^\nu}}-1)]\right|=0.$$

Now

$$\left|\omega_L[x(e^{itN_{L'/L'^\nu}}-1)]\right|\leqslant\|x\|\,|t|\,\omega_L\left[\left(\frac{N_{L'}}{L'^\nu}\right)^2\right]^{1/2}.$$

But from Lemmas II.5, 6, and 8 and from (i),

$$\lim_{L'\to\infty}\overline{\lim_{L\to\infty}}\ \omega_L\left[\left(\frac{N_{L'}}{L'^\nu}\right)^2\right]=\overline{\lim_{L\to\infty}}\ \omega_L\left[\left(\frac{N_L}{L^\nu}\right)^2\right]<\infty.$$

Proof of (iv): By (ii) it is again sufficient to prove that for each $\epsilon>0$ and L', L'' large enough

$$\overline{\lim_{L\to\infty}}\left|\omega_L[x(e^{itN_{L'/L'^\nu}}-e^{itN_{L''/L''^\nu}})]\right|<\epsilon.$$

For this to hold it is clearly sufficient that

$$\lim_{L'\to\infty}\overline{\lim_{L\to\infty}}\left|\omega_L[x(e^{itN_{L'}/L'^\nu}-e^{itN_{L''}/L''^\nu})]\right|=0.$$

But this follows from (i) using an estimate of the type used in Lemma III.1.

Finally, (v) follows from (i) and (ii) using analogous techniques as before.

Proposition III.3: Let again ω_β be any W^*-limit point of the ω_L as L tends to infinity; with the notation above one has (i) for each L there exists a self-adjoint operator $N_{\beta,L}$ on $\mathscr{H}_\beta$ affiliated to $\mathscr{M}_\beta$ such that

$$\pi_\beta(e^{itN_L})=e^{itN_{\beta,L}},$$

(ii) $N_{\beta,L}/L^\nu$ converges in the generalized strong sense, as $L\to\infty$, to a self-adjoint operator n_β affiliated to the center of $\mathscr{M}_\beta$, and (iii) ω_β is locally normal.

Proof: (i) follows immediately from the assumption and from the proof of Lemma III.2(iii).

About (ii) the existence of the self-adjoint operator n_β follows from Lemmas III.2(iv) and (iii). That n_β is affiliated with the center of $\mathscr{M}_\beta$ follows from Lemma III.2(ii). Indeed, from (ii)

$$0=[\pi_\beta(x_1x_2),e^{itn_\beta}]\Omega_\beta=[\pi_\beta(x_1),e^{itn_\beta}]\pi_\beta(x_2)\Omega_\beta$$

for all $x_1,x_2\in\mathscr{A}_{\text{loc.}}$.

Finally, to prove (iii), from Lemma III.1 it follows that ω_β is a regular state; from Lemma III.2(v),

$$\overline{\lim_{L}}\ \omega_L\left(\frac{N_{L'}}{L'^\nu}\right)<\infty$$

and (iii) follows from Ref. 9.

*Proposition III.*4: Let ω_β be any W^*-limit point of the $\omega_L(L\to\infty)$; one has (i) for each finite L, there exists a self-adjoint operator $N^0_{\beta,L}$ on $\mathscr{H}_\beta$ affiliated with $\mathscr{M}_\beta$ such that

$$\pi_\beta(e^{itN_{L,0}})=e^{itN^0_{\beta,L}};$$

(ii) $N^0_{\beta,L}/L^\nu$ converges in the generalized strong sense, as $L\to\infty$, to a self-adjoint operator n^0_β affiliated with the center of $\mathscr{M}_\beta$.

Proof: The statements of Lemma III.2 hold also when N_L is replaced by $N_{L,0}$. However, the proof of these statements requires a minor change at two places, namely, in estimating

$$\omega_{L'}\left[\left(\frac{N_{L,0}}{L^\nu}-\frac{N_{L',0}}{L'^\nu}\right)^2\right]$$

and in computing

$$[e^{itN_{L',0/L'^\nu}},W(\phi)].$$

The proof of the proposition then goes along the same lines as that of Proposition III.3, but based on the adapted Lemma III.2. ■

Proposition III.5: With the above notation, $\Omega_\beta\in\mathrm{Dom}(n_\beta)$ and

$$\rho=(\Omega_\beta,n_\beta\Omega_\beta).$$

Proof: To settle the domain questions we use the following properties: (a) if A is a self-adjoint operator on a Hilbert space $\mathscr{H}$ and $\phi\in\mathscr{H}$ such that

$$\left\|\left(\frac{e^{itA}-1}{t}\right)\phi\right\|\leqslant C,$$

where C is a constant independent of t, then $\phi\in\mathrm{Dom}(A)$ and $\|A\phi\|\leqslant C$. This follows from: If $\psi\in\mathrm{Dom}(A)$,

$$|(A\psi,\phi)|=\lim_{t\to 0}\left|\left(\frac{e^{itA}-1}{-it}\psi,\phi\right)\right|\leqslant C\|\psi\|.$$

(b) If $A_n=A_n^*$ converges in the generalized strong sense to an operator A on $\mathscr{H}$ and if $\forall n:\phi\in\mathrm{Dom}(A_n)$ and $\|A_n\phi\|\leqslant C$, then $\phi\in\mathrm{Dom}(A)$ and $\|A\phi\|\leqslant C$. This follows from (a) and

$$\left\|\frac{e^{itA}-1}{it}\phi\right\|=\lim_n\left\|\frac{(e^{itA_n}-1)}{it}\phi\right\|$$
$$\leqslant\lim_n\|A_n\phi\|\leqslant C.$$

One also has

$$\lim_{t\to 0}\left(\phi,\frac{e^{iA}-1}{it}\phi\right)$$
$$=\lim_{t\to 0}\lim_{n\to\infty}\left(\phi,\frac{e^{itA}-1}{it}\phi\right)$$
$$=\lim_{t\to 0}\lim_{n\to\infty}\left\{(\phi\,|A_n\phi)+\left[\phi,\left(\frac{e^{itA}-1}{it}-A_n\right)\phi\right]\right\}.$$

But as $\phi\in\mathrm{Dom}(A_n)$ one has

$$\left|\left[\phi,\left(\frac{e^{itA}-1}{it}-A_n\right)\phi\right]\right|\leqslant\frac{t}{2}\|A_n\phi\|^2 \tag{10}$$

and, as $\|A_n\phi\|\leqslant C$, one gets

$$(\phi,A\phi)=\lim_{n\to\infty}(\phi,A_n\phi)=\lim_{n\to\infty}\lim_{t\to 0}\left(\phi,\frac{e^{itA_n}-1}{it}\phi\right). \tag{11}$$

Coming back to the imperfect Bose gas one has from the proof of Lemma III.2

$$\left|\left|\left(\frac{e^{it(N_{\beta,L}/L^\nu)}-1}{it}\right)\Omega_\beta\right|\right|^2\leqslant\overline{\lim_{L'\to\infty}}\,\omega_{L'}\left(\frac{N_L}{L^\nu}\right)^2\leqslant C, \tag{12}$$

where C is independent of L.

Therefore, $\Omega_\beta\in\mathrm{Dom}(N_{\beta,L})$ by (a), $\Omega_\beta\in\mathrm{Dom}(n_\beta)$ by (b) and Proposition III.3(ii), and by (11)

$$(\Omega_\beta,n_\beta\Omega_\beta)=\lim_{L\to\infty}\lim_{t\to 0}\left(\Omega_\beta,\frac{e^{it(N_{\beta,L}/L^\nu)}-1}{it}\Omega_\beta\right).$$

So

$$(\Omega_\beta,n_\beta\Omega_\beta)=\lim_{L\to\infty}\lim_{t\to 0}\lim_{\alpha}\omega_{L_\alpha}\left(\frac{e^{it(N_L/L^\nu)}-1}{it}\right)$$
$$=\lim_{L\to\infty}\lim_{t\to 0}\lim_{\alpha}\omega_{L_\alpha}\left[\frac{N_L}{L^\nu}+\left(\frac{e^{it(N_L/L^\nu)-1}}{it}-\frac{N_L}{L^\nu}\right)\right].$$

Using again (10) and the bound (12) one gets

$$(\Omega_\beta,n_\beta\Omega_\beta)\lim_{L\to\infty}\lim_{\alpha}\omega_{L_\alpha}\left(\frac{N_L}{L^\nu}\right)$$
$$=\lim_{L\to\infty}\lim_{\alpha}\omega_{L_\alpha}\left(\frac{N_{L_\alpha}}{L_\alpha^\nu}\right)$$
[by Lemma III.2(v)] ■
$$=\rho.$$

Proposition III.6: With the above notation, $\Omega_\beta\in\mathrm{Dom}(n_\beta^0)$and

$$\rho_0\equiv\lim_{\alpha}\omega_{L_\alpha}\left(\frac{N_{L_\alpha,0}}{L_\alpha^\nu}\right)=(\Omega_\beta,n_\beta^0\Omega_\beta).$$

Proof: The proof of this proposition goes exactly along the same lines as that of Proposition III.5 but is now based on Proposition III.4. ■

Lemma III.7: For all $\phi\in\mathscr{D}$ we have

$$\lim_{L\to\infty}|||(e^{i\tau_L}-e^{it\tau})(\tilde\phi(0),\tilde\phi)|||=0$$

uniformly in t on compacts, where

$$\tau(\tilde\phi(0),\tilde\phi)=(0,\epsilon\tilde\phi);\ k\in R^\nu,$$
$$\tau_L(\tilde\phi^L_k(0),\tilde\phi^L_k)=\epsilon(k)[\tilde\phi^L_k(0),\tilde\phi^L_k],\quad k\in\tilde\Lambda.$$

Proof: For L large enough, Λ contains the support of ϕ and by a straightforward computation

$$e^{it\tau_L}[\tilde\phi(0),\tilde\phi\,]=[\tilde\phi(0),\tilde\phi_{L,t}\,],$$

where

$$\tilde\phi_{L,t}=\sum_{l\in\Lambda}\langle\phi^L_l,\phi\rangle\,e^{it\epsilon(l)}\tilde\phi^L_l.$$

Also

$$|||(e^{it\tau_L}-e^{it\tau})[\phi(0),\tilde\phi\,]|||^2$$
$$=\int_{R^\nu}dk\left(1+\frac{1}{|k|^2}\right)|F_L(k,t)|^2,$$

where

$$F_L(k,t)=\tilde\phi_{L,t}(k)-e^{it\epsilon(k)}\tilde\phi(k).$$

We use now the Lebesgue dominated convergence theorem to prove that the second term vanishes as $L\to\infty$. First of all, it is a standard matter to prove that

$$\lim_L F_L(k,t)=0.$$

Next one finds an $L_2\{R^\nu,[1+(1/k^2)]dk\}$ bound for $F_L(k,t)$. This amounts to looking for a bound for $\tilde\phi_{L,t}(k)$. In fact, we obtain a bound uniform in t on compacts.

We shall need the following bounds for elements $\phi\in\mathscr{D}$: Consider the Fourier transform

$$\hat\phi(l)=\left(\frac{1}{2\pi}\right)^{\nu/2}\int dx\,\phi(x)e^{ilx},\quad l\in\tilde\Lambda.$$

For $|l|<1$,

$$|\tilde\phi(l)|\leqslant(1/2\pi)^{\nu/2}\|\phi\|_1.$$

For $|l|\geqslant 1$, let $\Delta=-\Sigma_{i=1}^\nu(\partial^2/\partial x_i^2)$, for any integer n

$$|\tilde\phi(l)|\leqslant\left(\frac{1}{2\pi}\right)^{\nu/2}\frac{\|\Delta^n\phi\|_1}{l^{2n}}.$$

Hence, choosing n large enough

$$\frac{1}{L^\nu}\sum_{l\in\Lambda}|\tilde\phi(l)|$$
$$\leqslant\left(\frac{1}{2\pi}\right)^{\nu/2}\left[\left(\int_{|l|\leqslant 1}dl\right)\|\phi\|_1+\left(\int_{|l|\geqslant 1}\frac{dl}{l^{2n}}\right)\|\Delta^n\phi\|_1\right]. \tag{13}$$

Consider also the discrete Laplacian: For $a\in l^p(\tilde\Lambda)$,

$$(\Delta_d a)_n=\sum_{i=1}^\nu(a_{n+e_i}+a_{n-e_i}-2a_n),$$

where $e_i=(0,\ldots,0,2\pi/L,0,\ldots,0)\in\tilde\Lambda$ for the ith site. One has for $a\in l^p(\tilde\Lambda)$ and $b\in l^q(\tilde\Lambda)[(1/p)+(1/q)=1]$ and $m\in N$,

$$\sum_{n\in\Lambda}a_n(\Delta_d^m b)_n=\sum_{n\in\Lambda}(\Delta_d^m a)_n b_n. \tag{14}$$

Note that

$$e^{-ilx}=\frac{(-\Delta_d)^m}{4^m(\Sigma_{i=1}^\nu\sin^2\pi x_i/L)^m}e^{-ilx},\quad\text{for } l\in\tilde\Lambda, x\in\Lambda;$$

therefore, using (14),

$$\frac{1}{L^\nu}\sum_{l\in\Lambda}e^{it\epsilon(l)}\tilde\phi(l)e^{-ilx}$$

$$= \frac{1}{L^\nu}\sum_{l\in\Lambda} e^{it\epsilon(l)}\tilde\phi(l)\frac{(-\Delta_d)^m}{4^m(\Sigma_{i=1}^\nu \sin^2\pi x_i/L)^m}e^{-ilx}$$

$$= \frac{1}{L^\nu}\sum_{l\in\Lambda}\frac{(-\Delta_d)^m[e^{it\epsilon(l)}\tilde\phi(l)]}{4^m(\Sigma_{i=1}\sin^2\pi x_i/L)^m}e^{-ilx}.$$

For $x\in\Lambda$,

$$\sum_{i=1}^{\nu}\sin^2\frac{\pi x_i}{L}\geqslant\frac{4}{L^2}x^2.$$

Hence,

$$\left|\frac{1}{L^\nu}\sum_{l\in\Lambda}e^{it\epsilon(l)}\tilde\phi(l)e^{-ilx}\right|$$

$$\leqslant\frac{1}{L^\nu}\sum_{l\in\Lambda}\frac{|\Delta_d^m[e^{it\epsilon(l)}\tilde\phi(l)]|}{2^{4m}x^{2m}/L^{2m}}.$$

Now we treat the l summation as in (13), namely, by considering the cases $|l|<1$ and $|l|\geqslant 1$. In both cases we bound the discrete Laplacian by a factor $(2\pi/L)^2$ times a polynomial in t of second degree. Therefore, for t in some compact, there exists a constant K such that

$$\left|\frac{1}{L^\nu}\sum_{l\in\Lambda}e^{it\epsilon(l)}\tilde\phi(l)e^{-ilx}\right|\leqslant\frac{K}{x^{2m}}. \tag{15}$$

Now we are in a position to find a bound for $F_L(k,t)$. Indeed, for Λ containing the support of Φ,

$$\tilde\phi_{L,t}(k)=\int_\Lambda dx\left[\frac{1}{L^\nu}\sum_{l\in\Lambda}e^{it\epsilon(l)}\tilde\phi(l)e^{-ilx}\right]e^{ikx}. \tag{16}$$

Again we consider two cases:

(a) $|k|<1$: As $\nu\geqslant 3$ it is sufficient to prove the boundedness of $\tilde\phi_{L,t}(k)$. But this follows from using the bound (13) in the region $|x|<1$ and from using the bound (15) in the complement.

(b) $|k|\geqslant 1$: Using

$$\frac{(-\Delta)^n e^{ikx}}{k^{2n}}=e^{ikx}$$

and performing partial integrations in (16), one reduces the problem to the situation treated in (a), replacing $\tilde\phi(l)$ by $l^{2n}\tilde\phi(l)$.

Combining (a) and (b) we have proved for t in some compact

$$|F_k(k,t)|\leqslant F(k),\quad \text{where } F\in L_2\left[R^\nu,\left(1+\frac{1}{k^2}\right)dk\right].$$

Finally, the uniformity in t is obtained using the formula

$$e^{it_1A}-e^{it_2A}=i\int_{t_2}^{t_1}ds\,A\,e^{isA}$$

and the bound of above for $l^2\tilde\phi(l)$ instead of $\tilde\phi(l)$.

Now we are in a position to consider the limit $L\to\infty$ of the correlation inequality, which we perform in different steps. First we introduce two commuting groups of automorphisms of the von Neumann algebra $\mathcal{M}_\beta$.

Lemma III.8: (i) the maps $\alpha_t^T: e^{iB_\beta[\tilde\phi(0),\tilde\phi]}\to e^{iB_\beta[e^{it\tau}[\tilde\phi(0),\tilde\phi]]}$, $\phi\in\mathcal{D}$, $t\in R$ extend to a weakly continuous group of automorphisms of $\mathcal{M}_\beta$. (ii) if ω_β is the W^*-limit point of the net $\{\omega_{L_\alpha}\}_\alpha$, then

$$\lim_\alpha\omega_{L_\alpha}\left\{y\left[T_{L_\alpha},\int ds f(s)\alpha_s^{T_{L_\alpha}}(x)\right]\right\}$$

$$=i\left(\Omega_\beta,\Pi_\beta(y)\int ds f'(s)\alpha_s^T[\Pi_\beta(x)]\Omega_\beta\right),$$

where x, y belong to $\mathcal{A}_{\text{loc.}}$ and f is any absolutely continuous function on R such that f and f' belong to $L_1(R,ds)$.

Proof: (i) If $\alpha_t^T e^{iB_\beta[\tilde\phi(0),\tilde\phi]}=e^{iB_\beta[e^{it\tau}[\tilde\phi(0),\tilde\phi]]}$, $\phi\in\mathcal{D}$, then clearly α_t^T extends to a group of * automorphisms of the C^* algebra generated by $[e^{iB_\beta[\tilde\phi(0),\tilde\phi]}|\phi\in\mathcal{D}]$. From the local normality of the state ω_β (Lemma III.1) this C^* algebra is weakly dense in $\mathcal{M}_\beta$. Therefore, α_t^T extends uniquely to a group of automorphisms of $\mathcal{M}_\beta$. Finally, the weak continuity in t follows again from Lemma III.1 and the invariance of the state under α^T which can be seen as follows: Let $y\in\mathcal{A}_{\text{loc.}}$ and $\phi\in\mathcal{D}$; then from Lemmas III.1 and III.7,

$$\lim_{L'}\overline{\lim_{L}}\ |\omega_L[y(\alpha_t^{T_L}-\alpha_t^{T_{L'}})W(\phi)]|=0$$

and then again from Lemmas III.1 and III.7,

$$\lim_\alpha\omega_{L_\alpha}[y\alpha_t^{T_{L_\alpha}}W(\phi)]$$

$$=\lim_{L'}[\Omega_\beta,\pi_\beta(y)e^{iB_\beta[e^{it\tau_L}[\tilde\phi(0),\tilde\phi]]}\Omega_\beta]$$

$$=[\Omega_\beta,\pi_\beta(y)e^{iB_\beta[e^{it\tau}[\tilde\phi(0),\tilde\phi]]}\Omega_\beta]. \tag{17}$$

The invariance of ω_β follows now from

$$\omega_L\circ\alpha_t^{T_L}=\omega_L.$$

(ii) follows from (17) and the Lebesgue dominated convergence argument. Indeed,

$$\lim_\alpha\omega_{L_\alpha}\left\{y\left[T_L,\int ds f(s)\alpha_s^{T_L}(x)\right]\right\}$$

$$=i\lim_\alpha\int ds f'(s)\omega_{L_\alpha}[y\alpha_t^{T_{L_\alpha}}(x)]$$

$$=i\int ds f'(s)[\Omega_\beta,\pi_\beta(y)\alpha_t^T\pi_\beta(x)\Omega_\beta]. \quad\blacksquare$$

Lemma III.9: (i) The maps $\alpha_s^N: a_\beta[\tilde\phi(0),\tilde\phi]\to e^{-is}a_\beta[\tilde\phi(0),\tilde\phi]$, $\phi\in\mathcal{D}$, $s\in R$, extend to a weakly continuous group of automorphism of $\mathcal{M}_\beta$. (ii) If ω_β is the W^*-limit point of the net $\{\omega_{L_\alpha}\}_\alpha$, then

$$\lim_\alpha\omega_{L_\alpha}\left\{y\left[-\mu_{L_\alpha}N_{L_\alpha}+\frac{\lambda}{2}\frac{N_{L_\alpha}^2}{L_\alpha^\nu},\int ds f(s)\alpha_s^{N_{L_\alpha}}(x)\right]\right\}$$

$$=i\left[\Omega_\beta,\pi_\beta(y)\int ds f'(s)\alpha_s^N(x)(-\mu+n_\beta)\Omega_\beta\right],$$

where x and y are local elements, and f is any absolutely continuous function on R such that f, f', and f'' belong to $L_1(R,ds)$.

Proof: In this case the proof is trivial because for any local element x, $\alpha_t^{N_L}(x)$ becomes L independent for L large enough and

$$\lim_L\pi_\beta[\alpha_t^{N_L}(x)]=\alpha_t^N\pi_\beta(x).$$

About (ii) note first that f, f', and f'' belonging to $L_1(R,ds)$, together with Proposition III.5, settle all domain questions in the formula. Furthermore,

$$\omega_L\left\{y\left[-\mu_L N_L+\frac{\lambda}{2}\frac{N_L^2}{L^\nu},\int ds\, f(s)\alpha_s^{N_L}(s)\right]\right\}$$
$$=i\omega_L\left\{y\left[\left(-\mu_L+\frac{\lambda}{2}n_L\right)\int ds\, f'(s)\alpha_s^{N_L}(x)+\frac{\lambda}{2}\,ds\, f'(s)\alpha_x^{N_L}(x)n_L\right]\right\}.$$

Note again that for x local and L large enough $\alpha_s^{N_L}(x)$ is L independent. The result follows using Lemma III.2(v) and Proposition III.3(ii). ■

Consider the set $\mathscr{E}$ of integrable functions f on R^2 absolutely continuous in both variables and with integrable partial derivatives. For $x\in\mathscr{M}_\beta$ and $f\in\mathscr{E}$ let

$$x_f=\int ds\, dt\, f(s,t)\alpha_s^T\alpha_t^N(x).$$

It is well known that the set $\mathscr{B}=\{x_f|x\in\mathscr{M}_\beta, f\in\mathscr{E}\}$ is dense in $\mathscr{M}_\beta$. As $\alpha_s^T\alpha_t^N=\alpha_t^N\alpha_s^T$ $(s,t\in R)$, it is clear that this set is in the intersection of the domains of the generators of the weakly continuous automorphism groups $\{\alpha_s^T\}_s$ and $\{\alpha_t^N\}_t$.

Theorem III.10: Let ω_β be a thermodynamic limit Gibbs states of the imperfect Bose gas; if $f_s=\partial f/\partial s$ and $f_t=\partial f/\partial t$ and $x_f\in\mathscr{B}$, then

$$i\beta\{x_f\Omega_\beta,[x_{f_s}-x_{f_t}(\mu-\lambda n_\beta)]\Omega_\beta\}\geqslant\|x_f\Omega_\beta\|^2\ln\frac{\|x_f\Omega_\beta\|^2}{\|x_f^*\Omega_\beta\|^2}. \tag{18}$$

Proof: The proof is now immediate from the local correlation inequality (2) and from Lemmas III.8 and III.9. ■

IV. THE CONDENSATE STATE

From above one has for $\phi\in\mathscr{D}$,

$$\pi_\beta[W(\phi)]=\exp iB_\beta[\tilde\phi(0),\tilde\phi];$$

also, more generally, using Lemma III.1,

$$\exp iB_\beta(\alpha,\tilde\phi)=\lim_L \pi_\beta\left\{W\left[\phi+[\alpha-\tilde\phi(0)]\left(\frac{2\pi}{L}\right)^{\nu/2}\phi_0^L\right]\right\}$$

because

$$\alpha=\tilde\phi(0)+[\alpha-\tilde\phi(0)]\left(\frac{2\pi}{L}\right)^{\nu/2}\phi_0^L(0)$$

and

$$\lim_{L\to\infty}\left|\left|\left|\left[0,\left(\frac{2\pi}{L}\right)^{\nu/2}\tilde\phi_0^L\right]\right|\right|\right|=0.$$

For $\phi\in\mathscr{D}$, we introduce $N_\beta(\phi)$ as the infinitesimal generator of the strongly continuous group of unitaries $\{\pi_\beta[\exp itN(\phi)]\}_{t\in R}$. The continuity follows from the local normality of the state ω_β. Another consequence of this is also

$$N_\beta(\phi)=a_\beta^*[\tilde\phi(0),\tilde\phi\,]a_\beta[\tilde\phi(0),\tilde\phi\,]. \tag{19}$$

This allows us to find an explicit expression for n_β^0, the generalized strong limit $L\to\infty$ of $N_\beta(\phi_0^L/L^{\nu/2})$:

$$n_\beta^0=s\text{-}\lim_L a_\beta^*\left(\frac{\tilde\phi_0^L(0)}{L^{\nu/2}},\frac{\tilde\phi_0^L}{L^{\nu/2}}\right)a\left(\frac{\tilde\phi_0^L(0)}{L^{\nu/2}},\frac{\phi_0^L}{L^{\nu/2}}\right)$$
$$=a^*\left[\left(\frac{1}{2\pi}\right)^{\nu/2},0\right]a\left[\left(\frac{1}{2\pi}\right)^{\nu/2},0\right]$$

by the $|||\cdot|||$-norm estimates of above. Define the operator

$$U_0=\frac{a_\beta[(1/2\pi)^{\nu/2},0]}{\rho_0^{1/2}}; \tag{20}$$

then

$$U_0U_0=n_\beta^0/\rho_0. \tag{21}$$

Note that U_0 is affiliated with the center $\mathscr{M}_\beta\cap\mathscr{M}_\beta'$ and hence a normal operator.

Lemma IV.1: Let f be any positive bounded measurable function f on R with compact support; then

$$\omega_\beta[n_\beta^0 f(n_\beta^0)(n_\beta-n_\beta^0)]=\left(\frac{1}{2\pi}\right)\int dk\,\frac{1}{e^{\beta\epsilon(k)}-1}\,\omega_\beta[n_\beta^0 f(n_\beta^0)].$$

Proof: For $\phi\in\mathscr{D}$ such that $\tilde\phi(0)=0$, take the observables

$$f(n_\beta^0)^{1/2}U_0a_\beta^*(0,\tilde\phi)\quad\text{and}\quad f(n_\beta^0)^{1/2}U_0^*a_\beta(0,\tilde\phi)$$

in the correlation inequality (18). The extension of this inequality to these unbounded operators is justified by the estimates given before. One gets

$$\beta\omega_\beta[n_\beta^0 f(n_\beta^0)a_\beta(0,\tilde\phi)a_\beta^*(0,\epsilon\tilde\phi)]\geqslant\omega_\beta\{n_\beta^0 f(n_\beta^0)[N_\beta(\phi)+\|\phi\|^2]\}\times\ln\frac{\omega_\beta\{n_\beta^0 f(n_\beta^0)[N_\beta(\phi)+\|\phi\|^2]\}}{\omega_\beta[n_\beta^0 f(n_\beta^0)N_\beta(\phi)]}, \tag{22}$$

$$-\beta\omega_\beta[n_\beta^0 f(n_\beta^0)a_\beta^*(0,\tilde\phi)a_\beta(0,\epsilon\tilde\phi)]\geqslant\omega_\beta[n_\beta^0 f(n_\beta^0)N_\beta(\phi)]\times\ln\frac{\omega_\beta[n_\beta^0 f(n_\beta^0)N_\beta(\phi)]}{\omega_\beta\{n_\beta^0 f(n_\beta^0)[N_\beta(\phi)+\|\phi\|^2]\}}. \tag{23}$$

Using Lemma III.1,

$$|\omega_\beta[n_\beta^0 f(n_\beta^0)a_\beta^*(0,\tilde\phi)a_\beta(0,\tilde\psi)]|\leqslant C\omega_\beta[N_\beta(\phi)]^{1/2}\omega_\beta[N_\beta(\psi)]^{1/2}\leqslant C'|||\phi|||'|||\psi|||',$$

where

$$|||\phi|||'=\left[\int dk\left(1+\frac{1}{|k|^2}\right)|\tilde\phi(k)|^2\right]^{1/2}.$$

Denote the closure of $\mathscr{D}$ with respect to the $|||\cdot|||'$-norm by $\bar{\mathscr{D}}'$; then by the Riesz representation theorem there exists a bounded operator $\tilde R$ on $\bar{\mathscr{D}}'$ such that

$$\omega_\beta[n_\beta^0 f(n_\beta^0)a_\beta^*(0,\tilde\phi)a_\beta(0,\tilde\psi)]=(\tilde\psi,\tilde R,\tilde\phi).$$

Furthermore,, Let $A\in\mathscr{A}_\Lambda$ and τ_x^Λ, $x\in R^\nu$ be the space translation automorphism over x with periodic boundary condition, and τ_x the usual ones; then

$$\tau_x^\Lambda A=\tau_x A$$

for all x such that $|x|<|L-L'|/2$. It follows that

$$\omega_\beta^0\tau_x=\omega_\beta,\quad\text{for all } x\in R^\nu.$$

Therefore, the operator $\tilde R$ is a multiplication operator by a continuous function:

$$(\tilde R\tilde\psi)(k)=\tilde R(k)\tilde\psi(k).$$

We may also assume $\tilde R(k)\neq0$ if $\omega_\beta[n_\beta^0 f(n_\beta^0)]\neq0$. Indeed, take S a set of measure different from zero and $\tilde R(k)=0$ for $k\in S$. Let ϕ be a function with support in S; then the left-hand side of inequality (4) is bounded; in order that

the right-hand side be bounded,

$$0=\omega_\beta\{n_\beta^0 f(n_\beta^0)[N_\beta(\phi)+\|\phi\|^2]\}=\omega_\beta[n_\beta^0 f(n_\beta^0)]\|\phi\|^2,$$

which is a contradiction. Now we write the inequalities obtained above by means of this operator $\tilde{R}$ and take a $\delta(k)$-convergent sequence $(\phi_n)_n$ of normalized functions ϕ_n in $\mathscr{D}'$ and derive

$$\left(1+\frac{1}{|k|^2}\right)\tilde{R}(k)=\frac{\omega_\beta[n_\beta^0 f(n_\beta^0)]}{e^{\beta\epsilon(k)}-1}.$$

Hence

$$\omega_\beta[n_\beta^0 f(n_\beta^0)(n_\beta-n_\beta^0)]$$

$$=\lim_L \frac{1}{L^\nu}\sum_{\substack{k\neq 0\\ k\in\Lambda}}\omega_\beta[n_\beta^0 f(n_\beta^0)a_\beta^*(0,\tilde{\phi}_k^L)a_\beta(0,\tilde{\phi}_k^L)]$$

$$=\lim_L \frac{1}{L^\nu}\sum_{\substack{k\neq 0\\ k\in\Lambda}}\omega_\beta[n_\beta^0 f(n_\beta^0)]\int dl\,|\tilde{\phi}_k^L(l)|^2\frac{1}{e^{\beta\epsilon(l)}-1}$$

$$=\left(\frac{1}{2\pi}\right)\omega_\beta[n_\beta^0 f(n_0)]\int dl\frac{1}{e^{\beta\epsilon(l)}-1}.$$

Lemma IV.2: With the above notation, $\mu=\lambda\rho$.

Proof: Take $(\phi_n)_n$ a sequence of elements of $\mathscr{D}$ such that $\tilde{\phi}_n(0)=0$, $\|\tilde{\phi}_n\|=1$, and $\lim_n(\tilde{\phi}_n,\epsilon\tilde{\phi}_n)=0$; then from (18)

$$\omega_\beta\{[a_\beta^*(0,\tilde{\phi}_n),-a_\beta(0,\epsilon\tilde{\phi}_n)+(\mu-\lambda n_\beta)a_\beta(0,\tilde{\phi}_n)]\}\geqslant 0$$

or

$$[\mu-\lambda\omega_\beta(n_\beta)]\|\tilde{\phi}_n\|^2-(\tilde{\phi}_n,\epsilon\tilde{\phi}_n)\leqslant 0.$$

Using Proposition III.5,

$$\mu\leqslant\lambda\rho.$$

Now take U_0 and U_0^* in (18):

$$\mu\omega_\beta(n_\beta^0)=\lambda\omega_\beta(n_\beta\, n_\beta^0).$$

From Lemma IV.1, letting $f_n\to 1$,

$$\omega_\beta[n_\beta^0(n_\beta-n_\beta^0)]=\rho_0\left(\frac{1}{2\pi}\right)^\nu\int dk\frac{1}{e^{\beta\epsilon(k)}-1}.$$

Hence,

$$\omega_\beta(n_\beta^0\, n_\beta)-\rho_0^2\geqslant\rho_0\left(\frac{1}{2\pi}\right)^\nu\int dk\frac{1}{e^{\beta\epsilon(k)}-1}$$

or

$$\frac{\mu}{\lambda}\geqslant\rho_0+\left(\frac{1}{2\pi}\right)^\nu dk\frac{1}{e^{\epsilon(k)}-1}.$$

But by Theorem II.9, $\mu/\lambda\geqslant\rho$. ■

Proposition IV.3: If $\rho_0>0$, one has (i) U_0 is unitary and $n_\beta^0=\rho_0$; (ii) $n_\beta=\rho$.

Proof: (i) Let $\{E_\lambda\,|\lambda\in R^+\}$ be the family of spectral projections of n_β^0 and for a measurable subset Δ of R^+, away from 0, denote

$$E_\Delta=\int_\Delta dE_\lambda.$$

From Lemma IV.1 one gets

$$\omega_\beta[E_\Delta(n_\beta-n_\beta^0)]=\left(\frac{1}{2\pi}\right)^\nu\left(\int dk\frac{1}{e^{\beta\epsilon(k)}-1}\right)\omega_\beta(E_\Delta).$$

Taking $x=f(n_\beta^0)U_0$ and $\bar{f}(n_\beta^0)U_0$ in (18), with $n_\beta^0\,|f(n_\beta^0)|^2=E_\Delta$, one finds, using Lemma IV.2,

$$\omega_\beta(E_\Delta n_\beta)=\frac{\mu}{\lambda}\omega_\beta(E_\Delta)=\rho\omega_\beta(E_\Delta).$$

Therefore,

$$\omega_\beta\left\{E_\Delta\left[n_\beta^0+\left(\frac{1}{2\pi}\right)^\nu\int dk\frac{1}{e^{\beta\epsilon(k)}-1}-\rho\right]\right\}=0.$$

It follows that the spetrum of n_β^0 is contained in $\{0,\rho_0'\}$, where

$$\rho_0'=\rho-\left(\frac{1}{2\pi}\right)^\nu\int dk\frac{1}{e^{\beta\epsilon(k)}-1}.$$

Let $n_\beta^0=\rho_0' E_{\rho_0'}$ be the spectral decomposition of n_β^0; then

$$\rho_0\omega_\beta(n_\beta^0)=\rho_0'\omega_\beta(E_{\rho_0'})\leqslant\rho_0\omega_\beta(E_{\rho_0'})$$

by Theorem II.9. But this implies $E_{\rho_0'}=1$ as $E_{\rho_0'}\in\mathscr{M}_\beta\cap\mathscr{M}_\beta'$ and (i) follows.

(ii) Taking $x=n_\beta^{1/2}U_0$ and $n_\beta^{1/2}U_0^*$ in (18), one gets, using (i) above,

$$\omega_\beta[n_\beta(\lambda n_\beta-\mu)]=0.$$

As $\omega_\beta(\lambda n_\beta-\mu)=0$ by Lemma IV.2,

$$\omega_\beta[(\lambda n_\beta-\mu)^2]=0$$

and (ii) follows. ■

Note that the proof of proposition IV.3 relies in an essential way on the condition $\lambda>0$. For the free Bose gas ($\lambda=0$) one can still conclude that $\mu=0$ (see Lemma IV.2) but it is known that the spectrum of n_β^0 is R^+. So proposition IV.3 is wrong in this case.

Note that as a result of proposition IV.3(i) we proved the existence of a phase operator for the condensed state.[10]

This result is also sufficient for the validity of the so-called Bogoliubov approximation. This approximation consists of replacing $a_\beta[(1/2\pi)^{\nu/2},0]$ by the c number $\rho_0^{1/2}e^{i\alpha}[\alpha\in(0,2\pi)]$. But consider now the decomposition

$$\omega_\beta(.)=\int_0^{2\pi}\omega_\beta[E(d\alpha).],$$

where

$$U_0=\int_0^{2\pi}E(d\alpha)e^{i\alpha}$$

is the spectral resolution of U_0.

It is clear that for the components $\omega_\beta[E(d\alpha).]$ the unitary operator U_0 is represented by $e^{i\alpha}$. Proving the validity of the Bogliubov approximation for the state consists in proving that the correlation inequality holds for these components, which is immediate from Theorem III.10. This result is an extension to the state of a result of Ref. 2 in the case of the imperfect Bose gas. Note that in this model the situation is particularly simple because the function $\alpha\to\omega_\beta[E(\alpha).]$ is constant due to the special type of the interaction.

Finally the solution of the condensed state is given as follows:

Theorem IV.4: For $\rho_0>0$ and $\phi\in\mathscr{D}$,

$$\omega_\beta[W(\phi)]=\exp\left[-\tfrac{1}{2}(\phi,K,\phi)\frac{1}{2\pi}\int_0^{2\pi}d\alpha\right]$$
$$\times\exp[2i\rho_0^{1/2}|\tilde{\phi}(0)|\cos\alpha],$$

where

$$(\tilde{K}\phi)(k) = \frac{1}{e^{\beta\epsilon(k)} - 1}\tilde{\phi}(k).$$

Proof: For $\phi \in \mathscr{D}$,

$$\begin{aligned}\omega_\beta[W(\phi)] &= (\Omega_\beta, e^{iB_\beta[\tilde{\phi}(0),0]} e^{iB_\beta(0,\tilde{\phi})}\Omega_\beta) \\ &= \left[\Omega_\beta, \int_0^{2\pi} E(d\alpha)\, e^{iB_\beta[\tilde{\phi}(0),0]} e^{iB_\beta(0,\tilde{\phi})}\Omega_\beta\right] \\ &= \int_0^{2\pi} \exp[e^{i\alpha}\rho_0^{1/2}\tilde{\phi}(0) + \text{c.c.}] \\ &\quad \times [\Omega_\beta, E(d\alpha)\, e^{iB_\beta(0,\tilde{\phi})}\Omega_\beta].\end{aligned}$$

As the components $\alpha \to [\Omega_\beta, E(d\alpha) X \Omega_\beta]$ with $X \in \{\pi_\beta[W(\phi)] \mid \tilde{\phi}(0) = 0, \phi \in \mathscr{D}\}''$ satisfy the correlation inequality for the free Bose system [by proposition IV.3(ii)], one has[6,11]

$$(\Omega_\beta, E(d\alpha) e^{iB_\beta(0,\tilde{\phi})}\Omega_\beta) = \frac{d\alpha}{2\pi}\exp[-\tfrac{1}{2}(\phi, K, \phi)].$$

ACKNOWLEDGMENTS

One of us (A.V.) thanks J. T. Lewis and J. V. Pulè for very illuminating discussions on the subject matter on many occasions,, especially during his visit to Dublin.

[1] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[2] J. Ginibre, Commun. Math. Phys. **8**, 26 (1968).
[3] G. Roepstorff, J. Stat. Phys. **18**, 191 (1978).
[4] K. Huang, *Statistical Mechanics* (Wiley, London, 1967).
[5] E. B. Davies, Commun. Math. Phys. **28**, 69 (1972).
[6] M. Fannes and A. Verbeure, Commun. Math. Phys. **55**, 125 (1977).
[7] P. C. Hohenberg, Phys. Rev. **158**, 383 (1967).
[8] J. T. Lewis and J. V. Pulè, Commun. Math. Phys. **36**, 1 (1974).
[9] G. F. Dell'Antonio and S. Doplicher, J. Math. Phys. **8**, 663 (1967).
[10] F. Rocca and M. Sirugue, Commun. Math. Phys. **34**, 111 (1973).
[11] F. Rocca, M. Sirugue, and D. Testard, Commun. Math. Phys. **19**, 119 (1970).

Stochastic electrodynamics of nonlinear systems. I. Partial in a central field of force

T. W. Marshall
Department of Mathematics, Manchester University, Manchester M13 9PL, England

P. Claverie
Institute de Biologie Physico-Chimique, Rue Pierre et Marie Curie, Paris 75005, France

(Received 31 May 1979; accepted for publication 23 August 1979)

We treat the motion of a charged particle in a central field plus a stochastic field corresponding to the zero-point field of quantum electrodynamics as a quasi-Markov process in the classical action variables. As special cases we consider the rigid rotator, the harmonic oscillator, and the Kepler system. We show that in the first two cases, but not in the third, the quasi-Markov process satisfies the detailed balance condition, which expresses the invariance of the process under time reversal. We consider the implications of this last result, and argue that it shows the possibility of a mechanical system being in equilibrium with respect to the radiation field, while the radiation field itself is not in equilibrium.

I. INTRODUCTION

Stochastic electrodynamics is the Brownian motion of a charged particle in a random electromagnetic field whose spectrum, at the absolute zero of temperature, is the same as that of the zero-point field in quantum electrodynamics. The frictional force is the normal Lorentz damping of classical electrodynamics. We are therefore dealing with a system which is entirely classical except for the presence of the zero-point field.

For the harmonic oscillator, such a model reproduces many features normally associated with quantum mechanics.[1] While this has been known for a long time, progress in extending the treatment to other force fields has been very limited,[2] mainly because a charged particle, moving on a classical orbit of given frequency, emits and absorbs energy from the electromagnetic field at all the harmonics of that frequency. Recently, however, Boyer[3] has noted that many results obtained for multiply periodic systems in the old quantum theory can be used also in stochastic electrodynamics.

The Fokker–Planck equation, which was introduced as part of old quantum theory[4] and subsequently became an important part of the general theory of Markov processes, already plays a central role in stochastic electrodynamics. It was used by one of us (see Ref. 1) as long ago as 1963, but the tool needed to calculate the drift and diffusion coefficients in it came to our attention only relatively recently. This is the multiply periodic expansion, which was well known in old quantum theory, and the use of which enables us to find the slow variation in the action variables when the electromagnetic forces are taken into account.

In the present article we seek to put the nonlinear problem in stochastic electrodynamics into the context of Markov processes and also to establish a connection with old quantum theory. A companion article[5] will show the Fokker–Planck equation found here may be related to a Fokker–Planck equation in complete (six-dimensional) phase space.

A central assumption is the validity of the *narrow line approximation*, according to which the system performs a large number of oscillations in the external field before the perturbing (that is damping and fluctuating) forces produce a significant change in its orbit. That this is a reasonable assumption with the orders of magnitude in stochastic electrodynamics has been demonstrated by Claverie and Diner.[6]

Our work has been greatly enriched by a study of the literature of the old quantum theory, and in this connection the recent semi-historical review article by Van Vleck and Huber[7] is of particular value. Indeed, the results obtained in the present paper follow directly from Van Vleck's treatment of radiation by multiply periodic systems.[8]

2. THE MULTIPLY PERIODIC EXPANSION

In this section we review the theory of action-angle variables, as it applies to the motion of a particle bound in a central field of force. A fuller treatment may be found in, for example, the textbook of Goldstein.[9]

The basic equations of motion are

$$m\ddot{\mathbf{x}} = \operatorname{grad} V(r), \tag{2.1}$$

where $r = |\mathbf{x}|$ is the distance from the center of force. The *action variables* (ξ_1, ξ_2, ξ_3) are defined as follows:

$$\xi_3 = L_3 = \text{component of angular momentum in a fixed direction}, \tag{2.2}$$

$$\xi_2 = L = \text{magnitude of angular momentum}, \tag{2.3}$$

$$\xi_1 = \frac{1}{2\pi} \oint \left(2m[H - V(r)] - \frac{L^2}{r^2}\right)^{1/2} dr, \tag{2.4}$$

where $H = T + V$ is the total energy (kinetic + potential) of the particle, and the integral is taken over a single complete oscillation of the variable r.

Equation (2.4) may be inverted to obtain H as a function of ξ_1 and ξ_2. Then the natural frequencies of the system are given by

$$\omega_1 = \frac{\partial H}{\partial \xi_1}, \quad \omega_2 = \frac{\partial H}{\partial \xi_2}. \tag{2.5}$$

For more general systems, H is a function of all the

 0022-2488/80/071819-07$1.00

action variables and the number of natural frequencies is equal to the number of degrees of freedom. In our case the derivative of H with respect to ξ_3 is zero, and hence ω_3 is zero. This leads to *degeneracy* in the orbits of the particle.

Associated with each action variable, ξ_i, there is an *angle variable*, η_i. To define these, we first form Hamilton's characteristic function

$$W(\mathbf{x},\boldsymbol{\xi}) = \int \{2m[H - V(r)] - (L^2/r^2)\}^{1/2}\, dr + L\theta + L_3\phi, \tag{2.6}$$

where θ is the polar angle in the plane of the orbit and ϕ is the azimuth of the line of nodes. Then

$$\eta_i = \frac{\partial W}{\partial \xi_i}. \tag{2.7}$$

The equations of motion (2.1), after the change of variables from $(\mathbf{x},\mathbf{p})$ to $(\boldsymbol{\xi},\boldsymbol{\eta})$, become

$$\dot{\xi}_i = 0, \tag{2.8}$$

$$\dot{\eta}_i = \omega_i, \tag{2.9}$$

that is, the action variables are constants of the motion, while the angle variables increase uniformly. It is these simple equations of motion which make the action-angle variables such useful coordinates, both in dynamical astronomy and in certain stochastic processes. Small extra perturbing terms on the right-hand side of (2.1) result in slow changes of the action variables. We have recently discovered[10] that there are other advantages. The drift and diffusion coefficients in the Fokker–Planck equation take a rather simple form when expressed in terms of these variables.

In the next section we shall see that a necessary part of the calculation of the coefficients is the obtaining of expressions for $\mathbf{x}$ and $\mathbf{p}$ in terms of $(\boldsymbol{\xi},\boldsymbol{\eta})$, and there is a generalization of Fourier's theorem, the *multiply periodic expansion theorem* (Ref. 9, p. 294) which gives precisely such an expression. This theorem states that any function F of $\mathbf{x}$ and $\mathbf{p}$ which is single valued on the phase-space $(\mathbf{x},\mathbf{p})$ may be expressed in the form

$$F(\mathbf{x},\mathbf{p}) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \sum_{n_3=-\infty}^{\infty} F_{\mathbf{n}}(\boldsymbol{\xi})\, e^{i\mathbf{n}\cdot\boldsymbol{\eta}}. \tag{2.10}$$

In the case where a particle moves in a central field of force, and F is $\mathbf{x}$ itself, an important simplification of this expansion occurs (Ref. 11, p. 138). Consider the case where the orbit lies in the (x,y) plane. Then

$$x + iy = \sum_{n=-\infty}^{\infty} c_n(\xi_1,\xi_2)\, e^{i(n\eta_1+\eta_2)}. \tag{2.11}$$

In this expression the summation over n_3 is, of course, absent, because of the restriction we have made on the plane of the orbit. But the summation over n_2 is also absent, because it takes the single value $n_2 = 1$. This expresses the uniform precession of the orbit, and is the reason for the selection rule for electric dipole radiation in old quantum theory. In the next section we shall see how this result may be extended to deal with the three-dimensional case.

3. THE FOKKER–PLANCK EQUATION

One of us has shown[10] how the Fokker–Planck equation in terms of the action variables may be obtained. Essentially the use of the narrow-line approximation permits us to treat the system as a *quasi-Markov process* in the action variables alone. The angle variables are needed only to find certain time averages which enter into the drift and diffusion coefficients. We shall be assuming implicitly that all the angle variables are uniformly distributed in the interval $(0, 2\pi)$, even for nonstationary states of the system.

The equation of motion is a generalized Langevin equation

$$m\ddot{\mathbf{x}} = -\operatorname{grad} V(r) + \mathbf{G} + \mathbf{F}(t), \tag{3.1}$$

where $\mathbf{G}$ is the Lorentz radiative reaction

$$\mathbf{G} = (2e^2/3c^3)\,\dddot{\mathbf{x}}, \tag{3.2}$$

and $\mathbf{F}(t)$ is the stationary stochastic force resulting from the particle's interaction, in the dipole approximation, with the vacuum electromagnetic field

$$\langle \mathbf{F}(t) \rangle = 0, \tag{3.3}$$

$$\langle F_i(t) F_j(t+\tau) \rangle = e^2 \delta_{ij} B_E(\tau). \tag{3.4}$$

In terms of the action variables, the Fokker–Planck equation is

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial \xi_i}\left(-A_i \rho + B_{ij} \frac{\partial \rho}{\partial \xi_j} \right), \tag{3.5}$$

where

$$A_i = \overline{G_j(t) \frac{\partial \xi_i}{\partial p_j}(t)} \tag{3.6}$$

and

$$B_{ij} = \tfrac{1}{2} e^2 \int_{-\infty}^{\infty} B_E(\tau)\, \overline{\frac{\partial \xi_i}{\partial p_k}(t) \frac{\partial \xi_j}{\partial p_k}(t+\tau)}\, d\tau. \tag{3.7}$$

In (3.6) and (3.7), the bar denotes time averaging over τ, and the time variation is that of the unperturbed motion, in which the effects of both $\mathbf{F}$ and $\mathbf{G}$ are to be neglected. We may use (2.10) to evaluate the right-hand sides of (3.6) and (3.7). This procedure is greatly simplified by the use of the result from classical mechanics, noted in Ref. 10, namely

$$\frac{\partial \xi_i}{\partial p_j} = \frac{\partial x_j}{\partial \eta_i}. \tag{3.8}$$

We obtain

$$A_i = -\frac{2}{3c^3} \sum_{\mathbf{n}} n_i (\mathbf{n}\cdot\boldsymbol{\omega})^3 d_{\mathbf{n}}^2, \tag{3.9}$$

$$B_{ij} = \tfrac{1}{2} \sum_{\mathbf{n}} n_i n_j S_E(\mathbf{n}\cdot\boldsymbol{\omega}) d_{\mathbf{n}}^2, \tag{3.10}$$

where $S_E(\omega)$ is the spectrum of the vacuum field, that is

$$S_E(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} B_E(\tau) d, \tag{3.11}$$

and $d_{\mathbf{n}}$ is the "oscillator strength" of the old quantum theory, that is

$$d_{\mathbf{n}}^2 = e^2(|x_{\mathbf{n}}|^2 + |y_{\mathbf{n}}|^2 + |z_{\mathbf{n}}|^2). \tag{3.12}$$

Returning to (3.6) and (3.7), we shall show in the Appendix that, because of the rotational symmetry, certain simple relations are satisfied by the coefficients A_i and B_{ij}.

These are

$$A_3 = (\xi_3/\xi_2)\,A_2, \tag{3.13}$$

$$B_{13} = (\xi_3/\xi_2)\,B_{12}, \tag{3.14}$$

$$B_{23} = (\xi_3/\xi_2)\,B_{22}, \tag{3.15}$$

$$B_{33} = \tfrac{1}{2}[1 + (\xi_3^2/\xi_2^2)]\,B_{22}. \tag{3.16}$$

We shall assume that the probability density also has rotational symmetry, so that its derivative with respect to ξ_3 is zero. In that case the Fokker–Planck equation reduces to

$$\begin{aligned}\frac{\partial\rho}{\partial t} = &\frac{\partial}{\partial\xi_1}\left[-A_1\rho + B_{11}\frac{\partial\rho}{\partial\xi_1} + B_{12}\frac{\partial\rho}{\partial\xi_2}\right]\\ &+\frac{\partial}{\partial\xi_2}\left[-A_2\rho + B_{12}\frac{\partial\rho}{\partial\xi_1} + B_{22}\frac{\partial\rho}{\partial\xi_2}\right]\\ &+\frac{1}{\xi_2}\left[-A_2\rho + B_{12}\frac{\partial\rho}{\partial\xi_1} + B_{22}\frac{\partial\rho}{\partial\xi_2}\right].\end{aligned} \tag{3.17}$$

This may be written in the form

$$\frac{\partial\rho}{\partial t} = \frac{1}{\xi_2}\frac{\partial}{\partial\xi_i}\xi_2\left[-A_i\rho + B_{ij}\frac{\partial\rho}{\partial\xi_j}\right], \tag{3.18}$$

where the summation convention now extends over $(i,j) = (1,2)$.

Finally, since ξ_1 and ξ_2, unlike ξ_3, are independent of the choice of axes, we may choose axes in the plane of the unperturbed orbit in calculating the drift and diffusion coefficients in (3.18). Then using the multiply period expansion (2.11), we find that the only nonvanishing oscillator strengths are given by $\mathbf{n} = (n,1,0)$, for which

$$d_{\mathbf{n}}^2 = e^2(|x_{\mathbf{n}}|^2 + |y_{\mathbf{n}}|^2) = e^2\,|c_n|^2. \tag{3.19}$$

The drift and diffusion coefficients of the reduced Fokker–Planck Eq. (3.18) then become

$$A_1 = -\frac{2e^2}{3c^3}\sum_{n=-\infty}^{\infty} n(n\omega_1 + \omega_2)^3\,|c_n|^2, \tag{3.20}$$

$$A_2 = -\frac{2e^2}{3c^3}\sum_{n=-\infty}^{\infty} (n\omega_1 + \omega_2)^3\,|c_n|^2, \tag{3.21}$$

$$B_{11} = \tfrac{1}{2}e^2\sum_{n=-\infty}^{\infty} n^2\,S_E(n\omega_1 + \omega_2)\,|c_n|^2, \tag{3.22}$$

$$B_{12} = \tfrac{1}{2}e^2\sum_{n=-\infty}^{\infty} nS_E(n\omega_1 + \omega_2)\,|c_n|^2, \tag{3.23}$$

$$B_{22} = \tfrac{1}{2}e^2\sum_{n=-\infty}^{\infty} S_E(n\omega_1 + \omega_2)\,|c_n|^2. \tag{3.24}$$

4. SOME SPECIAL CASES

To illustrate the method, and also to obtain some concrete models on which to base our later discussion of the detailed balance condition, we consider three potentials for which the Fokker–Planck coefficients may be explicitly calculated. They are

(i) the rigid rotator, $V(r) = \delta(r - r_0)$;
(ii) the linear oscillator, $V(r) = \frac{1}{2}m\omega_0^2 r^2$;
(iii) the Kepler system, $V(r) = -Ze^2/r$.

(i) In this case we put

$$c_0 = r_0 \quad\text{and}\quad c_n = 0 \quad (n \neq 0). \tag{4.1}$$

Also $H = \xi_2^2/2mr_0^2$, and therefore

$$\omega_1 = 0, \quad \omega_2 = \xi_2/mr_0^2. \tag{4.2}$$

The Fokker–Planck coefficients in (3.18) therefore become

$$A_1 = 0, \quad B_{11} = 0, \quad B_{12} = 0,$$

$$A_2 = -(2e^2/3c^3)\,\omega_2^3 r_0^2, \quad B_{22} = \tfrac{1}{2}e^2 S_E(\omega_2) r_0^2. \tag{4.3}$$

The stationary solution of this Fokker–Planck equation has been discussed by Boyer[12] and Claverie and Diner.[6] If we take the zero-point spectrum

$$S_E(\omega) = (2\hbar/3c^3)\,|\omega|^3, \tag{4.4}$$

then it is

$$\rho = (1/2\pi^2\hbar^2)\,e^{-2\xi_2/\hbar}. \tag{4.5}$$

Note that ρ is the density in the full phase-space ($\xi_2, \xi_3, \eta_2, \eta_3$). The reduced phase-space density, $W(\xi_2)$, is obtained by integrating over ξ_3, η_2 and η_3 and is

$$\begin{aligned}W(\xi_2) &= 8\pi^2\xi_2\rho\\ &= (4/\hbar^2)\,\xi_2\,e^{-2\xi_2/\hbar},\end{aligned} \tag{4.6}$$

which is the result given by Claverie and Diner.

(ii) For the linear oscillator (Ref. 11. p. 85)

$$H = \omega_0(2\xi_1 + \xi_2), \tag{4.7}$$

and therefore

$$\omega_1 = 2\omega_0 \quad\text{and}\quad \omega_2 = \omega_0. \tag{4.8}$$

The fact that $\omega_1 = 2\omega_2$ means that the orbit is closed, and leads to so-called "hidden degeneracy" in the quantum-mechanical energy levels. The only nonzero coefficients in the Fourier expansion (2.11) are c_0 and c_{-1}, which are related to ξ_1 and ξ_2 by

$$|c_0|^2 = 2(\xi_1 + \xi_2)/m\omega_0, \tag{4.9}$$

$$|c_{-1}|^2 = 2\xi_1/m\omega_0. \tag{4.10}$$

Hence the Fokker–Planck coefficients to be substituted in (3.18) for this case are

$$A_1 = -(4e^2\omega_0^2/3mc^3)\,\xi_1, \tag{4.11}$$

$$A_2 = -(4e^2\omega_0^2/3mc^3)\,\xi_2, \tag{4.12}$$

$$B_{11} = (2e^2\omega_0^2/3mc^3)\,\hbar\xi_1, \tag{4.13}$$

$$B_{12} = -(2e^2\omega_0^2/3mc^3)\,\hbar\xi_1, \tag{4.14}$$

$$B_{22} = (2e^2\omega_0^2/3mc^3)\,\hbar(2\xi_1 + \xi_2), \tag{4.15}$$

where we have again assumed the zero-point spectrum (4.4).

The stationary state in this case has been known at least since 1963.[1] It is

$$\rho = (1/\pi^3\hbar^3)\,e^{-(4\xi_1 + 2\xi_2)/\hbar}. \tag{4.16}$$

Again, we note that, as in the previous example, if we wish to express this as a reduced phase-space density in (ξ_1, ξ_2) we have to multiply by a factor $16\pi^3\xi_2$ to take account of integrations over ξ_3, η_1, η_2, and η_3.

(iii) For the Kepler system[9]

$$H = -mZ^2e^4/2(\xi_1 + \xi_2)^2, \tag{4.17}$$

and therefore

$$\omega_1 = \omega_2 = mZ^4e^4/(\xi_1 + \xi_2)^3. \tag{4.18}$$

As in the previous example, we have a closed orbit, which leads to degeneracy of the quantum-mechanical energy lev-

els. The Fourier expansion of the position vector on the Kepler orbit is[13]

$$x_n + iy_n = c_n = \frac{(\xi_1 + \xi_2)^2}{mZe^2} \times\left(\frac{J_n'(n\epsilon)}{n} + \frac{\eta}{\epsilon}\frac{J_n(n\epsilon)}{n}\right), \tag{4.19}$$

where

$$\eta = \xi_2/(\xi_1 + \xi_2) \text{ and } \epsilon = (1-\eta^2)^{1/2}. \tag{4.20}$$

Substituting in (3.20)–(3.24), we obtain for the Fokker–Planck coefficients

$$A_1 = -\frac{2mZ^4e^{10}}{3c^3}\frac{1}{(\xi_1+\xi_2)^5}\sum_{-\infty}^{\infty} n^3(n-1)f_n^2, \tag{4.21}$$

$$A_2 = -\frac{2mZ^4e^{10}}{3c^3}\frac{1}{(\xi_1+\xi_2)^5}\sum_{-\infty}^{\infty} n^3 f_n^2, \tag{4.22}$$

$$B_{11} = \frac{mZ^4e^{10}}{3c^3}\frac{\hbar}{(\xi_1+\xi_2)^5}\sum_{-\infty}^{\infty} n^2(n-1)^2|n|f_n^2, \tag{4.23}$$

$$B_{12} = \frac{mZ^4e^{10}}{3c^3}\frac{\hbar}{(\xi_1+\xi_2)^5}\sum_{-\infty}^{\infty} n^2(n-1)|n|f_n^2, \tag{4.24}$$

$$B_{22} = \frac{mZ^4e^{10}}{3c^3}\frac{\hbar}{(\xi_1+\xi_2)^5}\sum_{-\infty}^{\infty} n^2|n|f_n^2, \tag{4.25}$$

where

$$f_n = \frac{J_n'(n\epsilon)}{n} + \frac{\eta}{\epsilon}\frac{J_n(n\epsilon)}{n}. \tag{4.26}$$

It is a rather formidable tasks to obtain a closed form for the coefficients and then to find solutions of (3.18) with these coefficients. The very limited progress we have so far made will be reported in the next section.

5. THE DETAILED BALANCE CONDITION

The Fokker–Planck Eq. (3.5) may be expressed in the form

$$\frac{\partial\rho}{\partial t} + \text{div}\mathbf{J} = 0,$$

where

$$\mathbf{J} = \mathbf{A}\rho - \mathbf{B}\cdot\text{grad}\rho. \tag{5.1}$$

The quantity $\mathbf{J}$ may be called the *current density* associated with the probability density ρ.

We have shown in a separate article[10] that the condition that the process be invariant under time reversal, known as the condition of *detailed balance*, is that the current associated with the stationary distribution should be zero, that is

$$\mathbf{J}_0 = \mathbf{A}\rho_0 - \mathbf{B}\cdot\text{grad}\,\rho_0 = 0. \tag{5.2}$$

This implies that the drift vector $\mathbf{A}$ and the diffusion matrix $\mathbf{B}$ satisfy the *potential condition*[14,15]

$$\mathbf{B}^{-1}\mathbf{A} = -\text{grad}\phi \tag{5.3}$$

for some scalar ϕ.

Let us see whether the drift and diffusion coefficients obtained in the previous section satisfy this condition. We first examine the component of $\mathbf{J}_0$ associated with ξ_3.

$$J_{30} = A_3\rho_0 - B_{13}\frac{\partial\rho_0}{\partial\xi_1} - B_{23}\frac{\partial\rho_0}{\partial\xi_2} - B_{33}\frac{\partial\rho_0}{\partial\xi_3}. \tag{5.4}$$

Using the rotational symmetry, we have seen that the last term on the right-hand side is zero, and that the remaining terms satisfy the relations (3.13)–(3.15). Hence

$$J_{30} = \frac{\xi_3}{\xi_2}\left(A_2\rho_0 - B_{12}\frac{\partial\rho_0}{\partial\xi_1} - B_{22}\frac{\partial\rho_0}{\partial\xi_2}\right) = \frac{\xi_3}{\xi_2}J_{20}. \tag{5.5}$$

We have therefore found that, in order to establish detailed balance for three-dimensional systems with rotational symmetry, it is sufficient to show that the first two components of the current density vanish.

In the case of the three-dimensional rigid rotator, we are left with just one condition

$$-\frac{2e^2}{3c^3}\omega_2^3 r_0^2\rho - \frac{e^2}{3c^3}\omega_2^3 r_0^2 h\frac{\partial\rho_0}{\partial\xi_2} = 0. \tag{5.6}$$

This is satisfied by the stationary distribution (4.5). Indeed, it is generally the case that one-dimensional systems satisfy the detailed balance condition, and by (5.5) we have effectively reduced this system to one dimension.

Now consider the three-dimensional linear oscillator. We are left with two components of $\mathbf{J}_0$

$$J_{10} = -\frac{2e^2\omega_0^2}{3mc^3}\left(2\xi_1\rho_0 + \hbar\xi_1\frac{\partial\rho_0}{\partial\xi_1} - \hbar\xi_1\frac{\partial\rho_0}{\partial\xi_2}\right), \tag{5.7}$$

$$J_{20} = -\frac{2e^2\omega_0^2}{3mc^3}\left(2\xi_2\rho_0 - \hbar\xi_1\frac{\partial\rho_0}{\partial\xi_1} + \hbar(2\xi_1+\xi_2)\frac{\partial\rho_0}{\partial\xi_2}\right), \tag{5.8}$$

Again the stationary distribution (4.16) makes both of these components vanish. The vector $\mathbf{B}^{-1}\mathbf{A}$ has the components $(-4/\hbar, -2/\hbar, 0)$, so that (5.3) is satisfied with $\phi = (4\xi_1 + 2\xi_2)/\hbar$. The stationary density is a constant times $e^{-\phi}$, and this is generally the case where the potential condition is satisfied.

The rather satisfying trend apparently established by these two examples is, however, broken by our third case, that of the Kepler system. We begin by carrying out a change of variable

$$\xi' = \xi_1 + \xi_2, \quad \xi_2' = \xi_2. \tag{5.9}$$

Then the transformed Fokker–Planck equation is

$$\frac{\partial\rho}{\partial t} = \frac{\partial}{\partial\xi_i'}\left(-A_i'\rho + B_{ij}'\frac{\partial\rho}{\partial\xi_j'}\right), \tag{5.10}$$

where

$$A_1' = -\frac{2mZ^4e^{10}}{3c^3}\frac{1}{\xi_1'^5}\sum n^4 f_n^2, \tag{5.11}$$

$$A_2' = -\frac{2mZ^4e^{10}}{3c^3}\frac{1}{\xi_1'^5}\sum n^3 f_n^2, \tag{5.12}$$

$$B_{11}' = \frac{mZ^4e^{10}}{3c^3}\frac{\hbar}{\xi_1'^5}\sum n^4|n|f_n^2, \tag{5.13}$$

$$B_{12}' = \frac{mZ^4e^{10}}{3c^3}\frac{\hbar}{\xi_1'^5}\sum n^3|n|f_n^2, \tag{5.14}$$

$$B'_{22} = \frac{mZ^4e^{10}}{3c^3}\frac{\hbar}{\xi'^5_1}\sum n^2|n| f_n^2, \tag{5.15}$$

$$f_n = \frac{J'_n(n\epsilon)}{n} + \eta\frac{J_n(n\epsilon)}{n},$$
$$\eta = \xi'_2/\xi'_1 \text{ and } \epsilon = (1-\eta^2)^{1/2}. \tag{5.16}$$

The detailed balance condition in the form (4.3) requires that

$$\frac{\partial}{\partial\xi'_2}\left(\frac{B'_{22}A'_1 - B'_{12}A'_2}{B'_{11}B'_{22} - B_{12}}\right) = \frac{\partial}{\partial\xi'_1}\left(\frac{B'_{11}A'_2 - B'_{12}A'_1}{B'_{11}B'_{22} - B_{12}}\right), \tag{5.17}$$

and since both the bracketed expressions in (5.17) are functions of ξ'_1 and ξ'_2 through ϵ only, this simplifies to give

$$\frac{d}{d\epsilon}\left(\frac{B'_{22}A'_1 - B'_{12}A'_2}{B'_{11}B'_{22} - B'^2_{12}}\right) = -\eta\frac{d}{d\epsilon}\left(\frac{B'_{11}A'_2 - B'_{12}A'_1}{B'_{11}B'_{22} - B'^2_{12}}\right). \tag{5.18}$$

The series occurring on the right-hand sides of (5.11)–(5.15) may all be expressed as power series in ϵ, convergent for $|\epsilon| < 1$. The first few terms of each are as follows:

$$\sum n^4 f_n^2 = 1 + 3\epsilon^2 + \frac{45}{8}\epsilon^4 + \frac{35}{4}\epsilon^6 + \cdots, \tag{5.19}$$

$$\sum n^3 f_n^2 = 1 + \epsilon^2 + \epsilon^4 + \epsilon^6 + \cdots, \tag{5.20}$$

$$\sum n^4|n| f_n^2 = 1 + 7\epsilon^2 + \frac{717}{32}\epsilon^4 + \frac{7435}{144}\epsilon^6 + \cdots, \tag{5.21}$$

$$\sum n^3|n| f_n^2 = 1 + 3\epsilon^2 + \frac{179}{32}\epsilon^4 + \frac{1249}{144}\epsilon^6 + \cdots, \tag{5.22}$$

$$\sum n^2|n| f_n^2 = 1 + \epsilon^2 + \frac{33}{32}\epsilon^4 + \frac{151}{144}\epsilon^6 + \cdots. \tag{5.23}$$

The expressions occurring on the left- and right-hand sides of (5.18) have the expansions

$$-\frac{d}{d\epsilon}\left(\frac{B'_{22}A'_1 - B'_{12}A'_2}{B'_{11}B'_{22} - B'^2_{12}}\right) = \hbar^{-1}\left(\frac{1}{8}\epsilon - \frac{9}{32}\epsilon^3 + \cdots\right), \tag{5.24}$$

$$\eta\frac{d}{d\epsilon}\left(\frac{B'_{11}A'_2 - B'_{12}A'_1}{B'_{11}B'_{22} - B'^2_{12}}\right) = \hbar^{-1}\left(\frac{1}{8}\epsilon - \frac{17}{32}\epsilon^3 + \cdots\right), \tag{5.25}$$

and we conclude that the Kepler system does not satisfy the condition of detailed balance.

We have also investigated the detailed balance condition in the neighborhood of $\eta = 0$. The series in A'_1 and A'_2 can, in fact, be summed explictly and give

$$\sum n^4 f_n^2 = \tfrac{3}{2}\eta^{-5} - \tfrac{1}{2}\eta^{-3}, \tag{5.26}$$

$$\sum n^3 f_n^2 = \eta^{-2}. \tag{5.27}$$

The series in B'_{11}, B'_{12} and B'_{22} are more difficult to sum, and we have so far only been able to obtain the leading terms. We find that

$$\sum n^4|n| f_n^2 = \frac{4h\sqrt{3}}{\pi}\left[\eta^{-8} - \frac{13}{15}\eta^{-6} + O(\eta^{-5}\log\eta)\right], \tag{5.28}$$

$$\sum n^3|n| f_n^2 = \frac{4h\sqrt{3}}{3\pi}\left[\eta^{-5} - \frac{6}{5}\eta^{-3} + O(\eta^{-2}\log\eta)\right], \tag{5.29}$$

$$\sum n^2|n| f_n^2 = \frac{h\sqrt{3}}{\pi}\left[\eta^{-2} + \frac{2}{5}\log\eta + O(1)\right]. \tag{5.30}$$

We then obtain

$$\frac{d}{d\eta}\left(\frac{B'_{22}A'_1 - B'_{12}A'_2}{B'_{11}B'_{22} - B_{12}}\right) = -\frac{3\pi\sqrt{3}}{40h}\eta^2 + \cdots, \tag{5.31}$$

$$-\eta\frac{d}{d\eta}\left(\frac{B'_{11}A'_2 - B'_{12}A'_1}{B'_{11}B'_{22} - B'_{12}}\right) = \frac{54\pi\sqrt{3}}{125h}\eta\log\eta + \cdots, \tag{5.32}$$

so that the condition of detailed balance is not satisfied near $\eta = 0$ either.

The above results will make it very difficult to find a stationary solution of the Fokker–Planck equation for the Kepler system in stochastic electrodynamics. Nevertheless, the expression of the Fokker–Planck Eq. (5.1) in the form of a conservation law gives us some hint as to how to proceed. The physically accessible part of phase space is defined by the following inequalities:

$$0 \leqslant \xi'_1, \tag{5.33}$$

$$0 \leqslant \xi'_2 \leqslant \xi'_1, \tag{5.34}$$

$$-\xi'_2 \leqslant \xi'_3 \leqslant \xi'_2. \tag{5.35}$$

A norm-preserving solution of (5.1) will satisfy the boundary condition that the normal component of **J** vanishes on all the boundaries. Now, from (5.5) we see that, *for any probability density*, this condition is satisfied on the boundaries $\xi'_3 = \xi'_2$ and $\xi'_3 = -\xi'_2$, and from (5.19)–(5.23) it follows that, for $\xi'_2 = \xi'_1$, that is $\epsilon = 0$, the condition is again satisfied *for any probability density*.

Now Feller[16] has classified the types of boundary points which occur in the case of one-dimensional diffusion processes. If we limit our attention to common solutions of both the Fokker–Planck and Kolmogorov (or "forward" and "backward") equations, then they may be classified as "natural boundaries in the extended sense" (to include Feller's "natural", "entrance" and "exit" categories) and "regular" boundaries. In the latter case the norm-preserving condition that the current is zero at the boundary has to be imposed, while in the former it is automatically satisfied by all common solutions.

We have therefore found that all the finite boundaries in our phase-space are natural in the extended sense. This leads us to expect that it will be a fairly straightforward matter to find a minimizing functional for the stationary density, ρ_0,

and hence obtain a numerical solution for ρ_0, provided, of course, that such a stationary density exists.

6. SUMMARY AND DISCUSSION

We have shown that a mechanical system inateracting with the random field of stochastic electrodynamics may be treated as a quasi-Markov process in the action variables of the unperturbed motion. The resulting Fokker–Planck equation takes a fairly simple form for a particle moving in a central field, but except for certain rather special potentials it is still too complex for us to obtain even a stationary-state solution.

The special potentials lead to Fokker–Planck equations which satisfy the detailed balance condition, and therefore the process is time-reversal invariant in these cases. We have now been able to extend the analysis of a rather less simple system, the Kepler system, far enough to show that it does not satisfy this condition.

Boyer[3] has already pointed out the feature which is common to both the rigid rotator and the harmonic oscillator. For a given state of the system, that is for a given set of values of the action variables, it interacts only with one narrow band of frequencies in the zero-point spectrum. A general, nonlinear system will continually be transferring energy in the radiation field among the various harmonics of its natural frequencies.

Boyer found that, if we impose an extra condition called "radiation balance," namely that the mechanical system and the radiation field are both in a state of equilibrium, we come inevitably to the Rayleigh–Jeans distribution for the radiation and the Maxwell–Boltzmann distribution for the mechanical system. He expresses the view that this is a crisis for stochastic electrodynamics, which can be saved only by going over to a relativistic treatment.

We offer an alternative view. We have pointed out in another article[10] that the frictional force in stochastic electrodynamics is much less closely related to the stochastic force than in other forms of Brownian motion. Indeed in the customary, retarded-potential, form of electrodynamics, there is no relation between them at all. If we turn to the alternative, Wheeler–Feynman, form of electrodynamics, it becomes possible to relate them in principle, but even then the extreme long-range nature of the electromagnetic interaction means that the relation must be "global" rather than "local". And yet, in normal Brownian motion, the detailed balance condition does establish a relation between the frictional and stochastic forces.

Another feature of the super equilibrium studied by Boyer is that if we impose radiation balance we find not merely that the net amount of radiation absorbed at a given frequency is equal to the net amount emitted. We also find that such a balance is established separately for each state of the mechanical system, that is for each set of values of the action variables. This property, which has been called "super-detailed balance," was pointed out to us by Peña-Auerbach.[17]

We therefore propose that the failure of the general nonlinear system in stochastic electrodynamics to satisfy the conditions of either detailed balance or radiation balance be regarded as a reflection of the fact that stochastic electrodynamics, like quantum mechanics, is a local rather than a global theory. We speak of equilibrium between a mechanical system and the radiation field of the rest of the universe, even though the latter is, quite patently, *not* itself in a state of equilibrium. To the criticism, which may be made, that we should not therefore model it with a stationary Gaussian process, we reply, in the manner of Boltzmann, that you will have to wait an awfully long time to see any departure from stationarity, especially if you measure your time in atomic units.

We are grateful to Mr. E.J. Watson for showing us how to obtain the asymptotic expansions (5.28)–(5.30).

We give here the derivation of the relations (3.13)–(3.16).

APPENDIX

Putting the Lorentz radiative reaction (3.2) into (3.6), and putting also

$$\xi_3 = x_1 p_2 - x_2 p_1, \tag{A1}$$

we obtain

$$A_3 = (2e^2/3mc^3)\,\overline{x_1 \dddot{x}_2 - x_2 \dddot{x}_1} \tag{A2}$$

$$= (2e^2/3mc^3)\,\overline{(\mathbf{x} \wedge \ddot{\mathbf{x}})\cdot\hat{\mathbf{k}}}, \tag{A3}$$

where $\hat{\mathbf{k}}$ is the unit vector in the direction of the 3-axis. Similarly, putting

$$\xi_2^2 = (x_2 p_3 - x_3 p_2)^2 + (x_3 p_1 - x_1 p_3)^2 + (x_1 p_2 - x_2 p_1)^2, \tag{A4}$$

we obtain

$$A_2 = (2e^2/3mc^3)\,\overline{(\mathbf{x} \wedge \ddot{\mathbf{x}})\cdot\hat{\mathbf{L}}}, \tag{A5}$$

where $\hat{\mathbf{L}}$ is the unit vector in the direction of the angular momentum. Now $\ddot{\mathbf{x}}$ is in the plane of the orbit, and hence $\mathbf{x} \wedge \ddot{\mathbf{x}}$ is in the direction of $\mathbf{L}$, so

$$A_3 = \hat{\mathbf{L}}\cdot\hat{\mathbf{k}}\, A_2 = \frac{\xi_3}{\xi_2} A_2. \tag{A6}$$

The next two relations are obtained in a rather similar manner:

$$B_{i3} = \tfrac{1}{2}e^2 \int_{-\infty}^{\infty} B_E(\tau)\, \overline{\frac{\partial \xi_i}{\partial p_j}(t)\, \frac{\partial \xi_3}{\partial p_j}(t+\tau)}\, d\tau. \tag{A7}$$

Now both the vectors $\partial\xi_1/\partial\mathbf{p}$ and $\partial\xi_2/\partial\mathbf{p}$ are in the plane of the orbit, and therefore, by a repetition of the above argument

$$B_{13} = (\xi_3/\xi_2)\, B_{12} \quad \text{and} \quad B_{23} = (\xi_3/\xi_2)\, B_{22}. \tag{A8}$$

Finally, comparing B_{33} and B_{22},

$$B_{33} = \tfrac{1}{2}e^2 \int_{-\infty}^{\infty} B_E(\tau)\, \overline{x_1(t)x_1(t+\tau) + x_2(t)x_2(t+\tau)}\, d\tau, \tag{A9}$$

$$B_{22} = \tfrac{1}{2}e^2 \int_{-\infty}^{\infty} B_E(\tau)\, \overline{\mathbf{x}(t)\cdot\mathbf{x}(t+\tau)}\, d\tau. \tag{A10}$$

The 3-axis is, of course, fixed in space, but we are free to choose the direction of the others. We therefore take the 1-

axis as the line of nodes (that is the intersection of the plane of the orbit with the ecliptic). We also introduce a 2′-axis perpendicular to the 1-axis, but in the plane of the orbit. Then, clearly

$$x_1 = x_2' \cos i, \quad x_3 = x_2' \sin i, \quad \cos i = \xi_3/\xi_2. \quad \text{(A11)}$$

Also $\overline{x_1(t)x_1)(t+\tau)} = \overline{x_2'(t)x_2'(t+\tau)}$, and therefore

$$x_1(t)x_1(t+\tau) + x_2(t)x_2(t+\tau) = \tfrac{1}{2}\left[1 + (\xi_3^2/\xi_2^2)\right]\ \mathbf{x}(t)\cdot\mathbf{x}(t+\tau). \quad \text{(A12)}$$

Substituting (A12) in (A9) and comparing with (A10), we obtain the last of the required relations.

[1]T.W. Marshall, Proc. R. Soc. Ser. A **276**, 475 (1963).
[2]T.H. Boyer, Phys. Rev. D **13**, 2832 (1976).
[3]T.H. Boyer, Phys. Rev. A **18**, 1228 (1978).
[4]A. Fokker, Ann. Phys. **43**, 812 (1914); M. Planck, Ann. Phys. **52**, 491; and **53**, 241 (1917).
[5]P. Claverie, L. Pena-Auerbach, and S. Diner, article in preparation.
[6]P. Claverie, and S. Diner, Int. J. Quant. Chem. **12**, Supp. 1, 41 (1977).
[7]J.H. Van Vleck and D.L. Huber, Rev. Mod. Phys. **49**, 939 (1977).
[8]J.H. Van Vleck, Phys. Rev. **24**, 347 (1924).
[9]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, Mass., 1950), Chaps. 8 and 9.
[10]T.W. Marshall and P. Claverie, Proc. R. Soc. (to be published).
[11]M. Born, *Mechanics of the Atom* (Bell, London, 1927).
[12]T.H. Boyer, Phys. Rev. D **1**, 2257 (1970).
[13]L.D. Landau, and E.M. Lifshitz, *Classical Theory of Fields* (Pergamon, Oxford, 1962), Chap. IX.
[14]R.L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1963), Chap. 4.
[15]H. Haken, Z. Phys. **219**, 246 (1969).
[16]W. Feller, Ann. Math. **55**, 468 (1952).
[17]L. de la Peña-Auerbach, private communication.

Nonequilibrium fluctuations in driven systems

B. H. Lavenda
TEMA S.P.A., Via Marconi 29/1, 40122 Bologna and Istituto di Fisica Sperimentale, Via Tari 3, Napoli, Italy,

(Received 4 May 1979; accepted for publication 19 November 1979)

The validity of a fluctuation-dissipation theorem for nonequilibrium fluctuations in driven systems is discussed. It is found that the Gaussian assumption alone does not necessarily lead to a fluctuation-dissipation theorem but rather the Markov property must also be invoked. Consequences of the Gaussian–Markov assumption are: (i) the spectral density reduces to a sum of two independent contributions, one due to internal noise and the other to external excitations; and (ii) the mode of regression of nonequilibrium fluctuations is identical to the decay of fluctuations to equilibrium. Without the Markov assumption, time-varying external fields can induce additional correlations which: (i) influence the regression of fluctuations; and (ii) invalidate the fluctuation-dissipation theorem.

1. INTRODUCTION AND SUMMARY

One of the cornerstones of nonequilibrium statistical mechanics is the fluctuation-dissipation (FD) theorem[1] which represents a generalization of Nyquist's theorem[2] to nonequilibrium thermodynamic processes. Nyquist's theorem relates the electromotive force due to thermal agitation of electrons in a conductor to the resistive properties of the circuit. This is to say that the power spectrum or spectral density of the noise, $S(\epsilon;\omega)$, due to a fluctuating emf, ϵ, is related to the real part of the complex impedance $Z(\omega)$ by the formula

$$S(\epsilon,\omega) = 4k_B T\cdot \mathrm{Re}Z(\omega),$$

where k_B is Boltzmann's constant and T is the absolute temperature. The generalization of Nyquist's theorem, commonly referred to as the fluctuation-dissipation theorem, states that the change in the state of a system due to the action of an external (or fluctuating) force is accompanied by the absorption of energy whose rate is equal to that at which the energy is dissipated in the system.

There is, however an ambiguity to the FD theorem in that it apparently makes no difference as to whether the fluctuations occur spontaneously from equilibrium, i.e., internal noise, or whether they are the result of an imposed constraint. Let us recall that the Callen–Welton[1] derivation of the FD theorem requires the action of an external force, yet, the same results can be obtained by considering spontaneous fluctuations that are formally caused by the action of fictitious "random" forces.[3] That the system be insensitive to the way in which fluctuations occur is not intuitively obvious; Onsager and Machlup[4] have pointed out that this is, indeed, a very strong postulate: for a Gaussian process, it can be shown to be equivalent to the Markov assumption. Moreover, it is not self-evident that a FD theorem should be valid for systems in which fluctuations are caused by both random forces and external, time-varying constraints.[5]

We preface our remarks by clearly stating the limitations of our analysis. The first and foremost limitation is that the FD theorem is not valid beyond the linear approximation,[6] if the process happens to be nonlinear. This is to say that fluctuations in the emf need to be so small that nonlinear terms are unimportant. We shall guarantee this by treating a linear process so that there will always be present a Nyquist noise with an equilibrium temperature. In the Langevin description of a random process, in terms of noise sources (cf Sec. 2), the assumption that the random force is a Gaussian random variable together with the linearity of the process implies that the process is normal.

Within this limiting framework, we show that on account of the Gaussian–Markov postulate, we may carry over many of the equilibrium results to nonequilibrium fluctuations in driven systems. In particular, we obtain a FD theorem for the internal noise contribution to the spectral density even for slow, nonstationary processes for which it still makes sense to define a spectral density. We show that for Gaussian processes, the Markov assumption reduces the total spectral density to a sum of two, *independent* contributions: one due to internal noise for which there exists a FD theorem, relating the central second moments of the fluctuating variables to the dissipative properties of the system, and another contribution to the spectral density due to excitations that are caused by a time-varying, but otherwise arbitrary, external force. In other words, we show that the Gaussian assumption alone does not necessarily guarantee the validity of the FD theorem.

The FD theorem establishes a power balance between the rate at which energy is absorbed by the system and the rate at which energy is dissipated in the system. In physical terms, we may say that the system does not have the capacity to "store" energy (cf. Sec. 4), so that there cannot be noninstantaneous responses to external perturbations. This is guaranteed by the Markov assumption. We therefore conclude that for Gaussian processes, the Markov assumption is also required to ensure the existence of a FD theorem. Alternatively, for non-Gaussian–Markov processes, we are no longer able to express the spectral density as a superposition of internal noise and external excitations. Consequently, a FD theorem is no longer expected to hold.

Intuitively, we may say that regardless of whether fluctuations are occurring from an equilibrium or a nonequilibrium state, provided the process is Gaussian and the system does not remember how it got into the given state (i.e., the

 0022-2488/80/071826-07$1.00

Markov property), there exists a FD theorem and the regression of fluctuations is not influenced by time-varying, external constraints.

The latter statement can be considered as a generalization of the Onsager regression hypothesis to time-varying, external forces. We recall that the Onsager regression hypothesis states that the regression of spontaneous fluctuations obeys, *on the average*, the deterministic (phenomenological) law for the decay from the same nonequilibrium state back to equilibrium, when it has been produced by a constraint which is suddenly released.[7] Implicit in Onsager's hypothesis are: (i) the Markov assumption—the inability of the system to remember and (ii) the Gaussian assumption—the identification of the average with the most probable behavior. A regression theorem for Markov processes has been given by Lax[8] which does not make use of the Gaussian assumption. The utility of Lax's regression theorem is that it generalizes the Markov property from classical to quantum-mechanical systems by providing an appropriate factorization of the multitime density matrix. Limiting ourselves to the analysis of random processes that are both Markovian and Gaussian, which according to the generalized Doob theorem[9] must be linear Fokker–Planck processes (cf. Sec. 3), we find that the regression of fluctuations to a nonequilibrium-steady state in driven systems is identical to the regression of fluctuations to the state of equilibrium. This leads us to believe that Gaussian–Markov processes behave as if they were in equilibrium with respect to certain degrees of freedom while other degrees of freedom may be in a nonequilibrium state.

In Sec. 2 we briefly remark on the generalization of the Wiener–Khintchin theorem to slow, nonstationary processes. We then evaluate the autocorrelation function under nonequilibrium conditions by a method alternative to that of the characteristic function.[10] Under the Gaussian–Markov assumption, we find that fluctuations regress to a nonequilibrium steady state, or slowly evolving state, in the same way that they regress to the state of equilibrium. However, as a result of time-varying external fields, the mean value of the fluctuating variable is time-dependent and there will be an additional contribution to the spectral density at nonzero frequency due to external excitations. In order to gain an understanding of why external excitations have no effect on the regression of fluctuations, we perform a perturbation analysis in Sec. 3, using the path integral formulation of Gaussian nonequilibrium processes.[11,12] We find that the time-varying external field destroys the principle of microscopic reversibility since it can only create excitations and has no influence on the reverse transitions.

Finally, in Sec. 4 we show that, granting the Gaussian assumption, it is the Markov assumption that is responsible for our finding that the regression of fluctuations is not influenced by time-varying external fields. This is accomplished by comparing our results of Sec. 3 with those of a forced harmonic oscillator, to which we apply the operator formulation of nonequilibrium statistical thermodynamics.[12] We observe that the external field causes virtual transitions to and from neighboring excited states in the forced harmonic oscillator and this influences the way in which fluctuations regress. Moreover, the Markov assumption is violated since the system now remembers how it got into a given state. And on account of the fact that the system can store energy in the field, all responses need not be instantaneous; the system can manifest noninstantaneous responses to external perturbations.

2. THE WIENER–KHINTCHIN THEOREM FOR SLOW, NONSTATIONARY PROCESSES

The Wiener[13]–Khintchin[14] theorem relates the spectral density $S(\alpha;\omega)$ of a *stationary*, random process $\alpha(t)$ to the autocorrelation function $C(t)$ through a Fourier cosine transform:

$$S(\alpha;\omega) = 4\int_0^\infty C(t)\cos\omega t\, dt. \tag{2.1}$$

The derivation of (2.1) makes use of the property that the statistical characteristics of the random process are invariant under time translations. A generalization of the theorem to slow, nonstationary processes can be achieved under the conditions that the mean value $m(t) = \langle\alpha(t)\rangle$ and the autocovariance $A(t) = \langle[\alpha(t) - m(t)][\alpha(0) - m(0)]\rangle$ change slightly over times of the order of the correlation time τ_c, defined as

$$\tau_c \equiv \int_{-\infty}^{\infty} \tau A(\tau)\, d\tau \Big/ \int_{-\infty}^{\infty} A(\tau)\, d\tau. \tag{2.2}$$

This is to say that a spectral density may be defined as

$$S(\alpha;\omega,t) = 2\int_{-\infty}^{\infty} C(t,t+\tau)e^{-i\omega\tau}\, d\tau, \tag{2.3}$$

provided the following conditions are fulfilled[15]:

$$\tau_c \frac{dm}{dt} \ll m \quad \text{and} \quad \tau_c \frac{dA}{dt} \ll A. \tag{2.4}$$

The major difference between the stationary and slow, nonstationary definitions of the spectral density, (2.1) and (2.3) respectively, is that in the latter case, the autocorrelation function depends not only on the time difference but also upon the absolute time. In Sec. 3 we shall see that conditions (2.4) require the external force to vary slowly over times of the order of the inverse of the damping constant. Our analysis of fluctuations that occur in slow, nonstationary processes is now directed to the evaluation of the autocorrelation function appearing in (2.3).

Hashitsume[10] was probably the first to note the somewhat curious fact that time-varying exernal forces, acting on *linear* systems, have no influence on the autocovariance expression describing the regression of fluctuations. He analyzed the case of Brownian motion of a driven harmonic oscillator and found that the expression for the autocovariance was not modified by a time-varying electric field. Since this finding will be shown to be due to the Gaussian–Markov assumptions, it will suffice us to analyze the stochastic differential equation:

$$R\dot{\alpha} + s\alpha = X(t) + x(t). \tag{2.5}$$

By writing down Eq. (2.5), we have taken a Langevin description of the linear random process $\alpha(t)$, in which the noise source that generates the process is represented by the

random force $x(t)$. In the absence of the random force, Eq. (2.5) describes, for example, an RC circuit with an external emf, $X(t)$. We assume, for generality, that the external force is time-varying but is otherwise completely arbitrary.

In the Langevin noise source formulation, the fluctuating force is taken to represent thermal noise which is always stationary with a uniform spectrum. In order to be consistent with the fact that Eq. (2.5) describes, on the average, the deterministic evolution of the system, it is essential to assume that the conditional mean of the random force vanishes:

$$\langle x(t)\rangle_{\alpha_0} = 0, \quad t > t_0, \tag{2.6}$$

for any given initial value α_0. The Markovian property arises from the condition that the correlation function of the fluctuating force at two different times vanishes. In order to have a random process, it is necessary to assume that there is a sudden change in behavior at $t = t_0$, or, in other words, that the fluctuating force is δ-correlated

$$\langle x(t)x(t_0)\rangle = 2D\delta(t - t_0), \tag{2.7a}$$

where D is the diffusion coefficient, given by the Einstein relation:

$$D = k_B R^{-1}. \tag{2.7b}$$

Finally, the property that α is Gaussian follows from the observations that the average of the fluctuating forces over short time intervals is normal and that the process is linear.

The autocorrelation function is defined as

$$C(t,t_0) \equiv \langle \alpha(t)\alpha(t_0)\rangle, \tag{2.8}$$

where the angular brackets mean

$$\langle \alpha(t)\alpha(t_0)\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \alpha_0\alpha\omega_2(\alpha,\alpha_0)\,d\alpha\,d\alpha_0, \tag{2.9}$$

subject to $\alpha = \alpha(t)$ and $\alpha_0 = \alpha(t_0)$. It is important to bear in mind that we make no reference to an equilibrium ensemble in defining (2.8), as in the Kadanoff–Martin[16] definition of the equilibrium time correlation function.

In order to evaluate the autocorrelation function, we must first determine the two-joint probability density function (pdf), $\omega_2(\alpha,\alpha_0)$. Our task is simplified considerably by invoking the Markov assumption [cf. (2.7a)]. We can then write the two-joint pdf as a product of the conditional pdf $\omega_1(\alpha,t\,|\alpha_0,t_0)$ and the single pdf $\omega_0(\alpha_0)$, viz.,

$$\omega_2(\alpha,\alpha_0) = \omega_1(\alpha,t\,|\alpha_0,t_0)\cdot\omega_0(\alpha_0). \tag{2.10}$$

Substituting (2.10) into (2.9) reduces the autocorrelation function to a single integral

$$\langle \alpha(t)\alpha(t_0)\rangle = \int_{-\infty}^{\infty} \langle \alpha(t)\rangle_{\alpha_0}\alpha_0\omega_0(\alpha_0)\,d\alpha_0, \tag{2.11}$$

where $\langle \alpha(t)\rangle_{\alpha_0}$ is the conditional average:

$$\langle \alpha(t)\rangle_{\alpha_0} = \int_{-\infty}^{\infty} \alpha\omega_1(\alpha,t\,|\alpha_0,t_0)\,d\alpha. \tag{2.12}$$

We now have to evaluate the two-joint pdf (2.10).

Rather than employing the method of characteristic functions[10] to evaluate the pdf, we choose the more intuitive method of solving the Fokker–Planck equation

$$\frac{\partial}{\partial t}\omega = \frac{\partial}{\partial \alpha}\{(\gamma\alpha - X(t)/R)\omega\} + D\frac{\partial^2\omega}{\partial\alpha^2}, \tag{2.13}$$

where $\gamma = s/R$ and ω is any member of the hierarchy of pdf.

The explicit evaluation of any pdf from the Fokker–Planck equation depends on the imposed initial and boundary value conditions. Solving the Fokker–Planck equation (2.13) subject to the condition

$$\omega = \delta(\alpha - \alpha_0) \quad \text{for } t = 0, \tag{2.14}$$

say, by the method of variation of parameters,[17] we obtain the expression

$$\omega_1(\alpha,t\,|\alpha_0,0) = \{2\pi k_B[1 - \exp(-2\gamma t)]/s\}^{-1/2} \times \exp(-Q/2k_B), \tag{2.15}$$

for the conditional pdf, where

$$Q = s|\alpha - \alpha_0 e^{-\gamma t} - R^{-1} \times \int_0^t \exp[-\gamma(t-u)]X(u)\,du|^2/(1 - e^{-2\gamma t}). \tag{2.16}$$

If we allow a long enough time to elapse, it is natural to suppose that the system will forget its initial condition and the conditional pdf will go over into the single pdf. Assuming that the system has evolved from the state $\alpha_0 = 0$ at some very distance time in the past, $t_0 = -\infty$, due to the action of a time-varying, external force, we obtain the solution to the Fokker–Planck equation:

$$\omega_\infty(\alpha,t) = (s/2\pi k_B)^{1/2}\exp\left\{-s|\alpha - R^{-1} \times \int_{-\infty}^{t} \exp[-\gamma(t-u)]X(u)\,du|^2/2k_B\right\}, \tag{2.17}$$

which is the single pdf of the "aged"system. It should be noticed that (2.17) does not coincide with either a stationary $X = \text{const}$, or an equilibrium, $X = 0$, pdf. For slow, nonstationary processes, however, (2.17) will play the role of the single pdf in the autocorrelation function expression, (2.11). Now inserting expressions (2.15) and (2.17) into (2.10) and then into (2.9), we obtain the expression for the autocorrelation function:

$$C(t,t_0) = (k_B/s)\exp[-\gamma(t - t_0)] + \left\{R^{-1}\int_{-\infty}^{t} \exp[-\gamma(t-u)]X(u)\,du\right\} \times \left\{R^{-1}\int_{-\infty}^{t_0} \exp[-\gamma(t_0 - u)]X(u)\,du\right\}. \tag{2.18}$$

The autocorrelation function and autocovariance are related by

$$A(t) = C(t) - m^2(t). \tag{2.19}$$

Knowing that

$$m(t) = R^{-1}\int_{-\infty}^{t} \exp[-\gamma(t-u)]X(u)\,du, \tag{2.20}$$

which is a particular solution of the deterministic equation associated with the stochastic differential equation (2.5), we

get the expression for the autocovariance from (2.19):

$$A(t-t_0) = (k_B/s)\exp[-\gamma(t-t_0)]. \tag{2.21}$$

From this result, we conclude that for Gaussian–Markov processes, fluctuations regress in driven systems in exactly the same way as they do to the state of equilibrium. In other words, it shows that time-varying, external forces have no influence on the regression of fluctuations in Gaussian–Markov processes.

The spectral density, (2.3), can now be computed by taking the Fourier transform of the autocorrelation function, (2.18). We then obtain

$$S(\alpha;\omega,t_0) = S(\alpha - m;\omega) + S(X;\omega,t_0), \tag{2.22}$$

where the contribution of the internal noise to the spectral density is still given by Nyquist's theorem

$$\omega^2 S(\alpha - m;\omega) = 4k_B T\cdot \mathrm{Re}Y(\omega), \tag{2.23}$$

where $Y(\omega)$ is the admittance

$$Y(\omega) = (R + s/i\omega)^{-1}/T. \tag{2.24}$$

The important result which follows from (2.22) is that for Gaussian–Markov processes, the total spectral density is the sum of two, independent contributions: one coming from internal noise and the other from external excitations.

In the particular case of stationary, periodic excitations, $X(t) = X_0\exp(i\omega_0 t)$, the spectral density (2.22) reduces to

$$\omega^2 S(\alpha;\omega) = 4k_B T\cdot \mathrm{Re}Y(\omega) + 4\pi R^{-1}|X_0|^2 \times \mathrm{Re}Y(\omega_0)\delta(\omega+\omega_0). \tag{2.25}$$

In words, (2.25) states that periodic, external excitations will only contribute to the noise spectrum when they are in resonance with the natural frequencies of the system. In view of this result, it is interesting to recall Richardson's[5] conjecture that there need be no connection between the admittance function that describes the regression of fluctuations and the response of the system to external excitations. Expression (2.25) shows that the admittance functions must be identical for periodic, stationary excitations.

To conclude this section, we compare our results with those of Lax.[18] Lax took into account that fluctuations were occurring from a nonequilibrium steady state implicitly through the values of the coefficients in the Langevin equation. He then proposed that a Nyquist relation for nonequilibrium fluctuations be given by

$$\omega^2 S(\alpha;\omega) = 4k_B T\cdot \mathrm{Re}Y(\omega)\eta, \tag{2.26}$$

where η is the correction factor:

$$\eta = (s/k_B)\langle\alpha^2\rangle, \tag{2.27}$$

measuring the deviation from equilibrium where $\eta = 1$. Lax's analysis applies to fluctuations from a nonequilibrium steady state which is maintained by time-independent external constraints. The spectral density of such processes is given by

$$S(\alpha;\omega) = S(\alpha - m;\omega) + 4\pi m^2\delta(\omega). \tag{2.28}$$

The δ function in the second term is due to the fact that the condition of stationariness requires the mean value to be independent of time. From expression (2.28), we observe that the only difference in the spectral density of fluctuations occurring from equilibrium or a nonequilibrium steady state, maintained by time-independent external constraints, is a shift in the power spectrum. We therefore conclude that the case of time-independent external constraints is uninteresting since there are no qualitative differences in the spectral density for fluctuations occurring from equilibrium or a nonequilibrium steady state. Our approach differs from that of Lax's in that we have taken the constraints, which are now time-dependent, explicitly into account.

3. TIME-DEPENDENT TRANSITION PROBABILITIES

In the last section, we gave the conditions for which a spectral density can be defined for slow, nonstationary processes [cf. conditions (2.4)]. Considering the solution of the averaged Langevin equation, in which $\langle x(t)\rangle_{\alpha_0} = 0$, we observe that conditions (2.4) will be satisfied provided the external force varies slightly over times of the order of γ^{-1}. In this event, we can treat the external force as a small, time-dependent perturbation and apply time-dependent perturbation theory to the Fokker–Planck equation (2.13).

The Fokker–Planck equation may be formally written in the operator form:

$$\frac{\partial\omega}{\partial t} = (2k_B)^{-1}\hat{F}\omega, \tag{3.1}$$

where the symbol "$\hat{}$" is used to distinguish an operator from an ordinary function. Rather than studying the properties of the Fokker–Planck operator $\hat{F}$, it will prove advantageous to apply the "gradient transformation"[19]

$$\omega(\alpha,t) = \exp\left(-s\int \alpha\, d\alpha/2k_B\right)\cdot\phi(\alpha,t), \tag{3.2}$$

where[20]

$$\int \alpha\, d\alpha = \tfrac{1}{2}\alpha^2 - \tfrac{1}{2}\langle\alpha^2\rangle, \tag{3.3}$$

to Eq. (3.1) and study the properties of the Hamiltonian operator $\hat{H}$, defined by the Schrödinger-type equation:

$$\frac{\partial\phi}{\partial t} = -(\tfrac{1}{2}\gamma)\left(\frac{\gamma R\alpha^2}{2k_B} - \frac{X(t)\alpha}{k_B}\right)\phi - \frac{X(t)}{R}\cdot\frac{\partial\phi}{\partial\alpha} + \frac{k_B}{R}\cdot\frac{\partial^2\phi}{\partial\alpha^2} \equiv(\tfrac{1}{2}k_B^{-1})\hat{H}\phi. \tag{3.4}$$

For a vanishing external force, Eq. (3.4) becomes identical to the Schrödinger equation for a harmonic oscillator. This would lead us to suppose that for a small external force, the Hamiltonian $\hat{H}$ can be decomposed into an unperturbed part $\hat{H}_0$ and a perturbed part $\hat{H}_1$:

$$\hat{H} = \hat{H}_0 + \hat{H}_1. \tag{3.5}$$

$\hat{H}_0$ is a self-adjoint operator and possesses a complete set of eigenfunctions that correspond to those of a harmonic oscillator, viz.,

$$\phi_n = (2^n n!)^{-1/2}(s/2\pi k_B)^{1/4}H_n(\xi)\exp(-\tfrac{1}{2}\xi^2), \tag{3.6}$$

where

$$\xi = (s/2k_B)^{1/2}\alpha, \quad \text{and} \quad H_n(\xi) = (-1)^n\left(e^{\xi^2}\frac{d^n}{d\xi^n}e^{-\xi^2}\right), \tag{3.7}$$

i.e., H_n is the nth degree Hermite polynomial.

The transformed Fokker–Planck equation, (3.4), is comparable to the Schrödinger equation of a forced harmonic oscillator (cf. Sec. 4) in the presence of a vector potential. In analogy with the Feynman path integral formulation of quantum electrodynamics,[21] we define a kernel $G(\alpha,t\,|\alpha_0,t_0)$ which propagates the function ϕ from one point in α-space and time to another point at a later time according to the integral equation[12]

$$\phi(\alpha,t) = \int_{-\infty}^{\infty} G(\alpha,t\,|\alpha_0,t_0)\phi(\alpha_0,t_0)\,d\alpha_0. \tag{3.8}$$

This integral equation bears a striking resemblance to the Chapman–Kolmogorov equation and it is a consequence of the Markov assumption. In the absence of the external force, the perturbed Hamiltonian $\hat{H}_1$ vanishes and the kernel can be expanded as a bilinear sum of orthonormal functions, (3.6), multiplied by exponential decaying functions of time[12]

$$G_0(\alpha,t\,|\alpha_0,t_0) = \sum_{n=0}\phi_n(\alpha)\phi_n(\alpha_0)\exp[-\Gamma_n(t-t_0)], \tag{3.9}$$

with $\Gamma_n = n\gamma$. In the presence of time-varying external fields, the time dependence will no longer be given by a simple exponential factor and instead we write:

$$G(\alpha,t\,|\alpha_0,t_0) = \sum_{m=0}\sum_{n=0}\lambda_{nm}(t,t_0)\phi_n(\alpha)\phi_m(\alpha_0), \tag{3.10}$$

where the time-dependent transition probabilities $\lambda_{nm}(t,t_0)$ have now to be determined.

We express the kernel in the path integral form[12]

$$G(\alpha,t\,|\alpha_0,t_0) = \overline{\int\int}\exp\left[-(2k_B)^{-1}\int_{t_0}^{t}L(\alpha,\dot{\alpha},t)\,dt\right]D\alpha. \tag{3.11}$$

The connected integral sign denotes an average over all possible paths which begin in state α_0 at time t_0 and terminate in state α at the later time t. The Lagrangian L, in the present case, is

$$L(\alpha,\dot{\alpha},t) = \tfrac{1}{4}R\{(\dot{\alpha}-X(t)/R)^2 + \gamma^2\alpha^2 - 2\gamma\alpha X(t)/R\}, \tag{3.12}$$

which is seen to be a combination of the Lagrangian of a charged particle in an electromagnetic field and a forced harmonic oscillator. We could now apply time-dependent perturbation theory by expanding the exponential in (3.11) as a power series in the external force and calculate the time-dependent transition probabilities to any desired order in the external force.[22] However, if we note that our process is Gaussian and formally corresponds to a forced harmonic oscillator, which can be solved in closed form,[21] then we realize that the transition probabilities are also obtainable in closed form.

Using the Gaussian property, we equate average and most probable values so that the path average of the negative exponential of the action[11]

$$A = \int_{t_0}^{t}L(\alpha,\dot{\alpha},t)\,dt, \tag{3.13}$$

in expression (3.11) for the kernel can be replaced by

$$G(\alpha,t\,|\alpha_0,t_0) \propto \exp\left\{-(2k_B)^{-1}\left[\int_{t_0}^{t}L(\alpha,\dot{\alpha},t)\,dt\right]_{\min}\right\}. \tag{3.14}$$

Thus, the Lagrangian must satisfy the Euler–Lagrange equation:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\alpha}}\right) - \frac{\partial L}{\partial\alpha} = 0 \quad \text{or} \quad \ddot{\alpha} - \gamma^2\alpha = R^{-1}(\dot{X} - \gamma X), \tag{3.15}$$

which is similar to the equation of motion of a forced harmonic oscillator. The relation between the Euler–Lagrange equation, (3.15), which is a necessary condition for the extremum of the action (3.13), and the averaged Langevin equation can be deduced by writing the Euler–Lagrange equation (3.15) in canonical form.

Defining the conjugate momentum variable β in the usual fashion,

$$\beta \equiv \frac{\partial L}{\partial\dot{\alpha}}, \tag{3.16}$$

and taking the Legendre transform of the Lagrangian (3.12) with respect to the velocity, we obtain the Hamiltonian:

$$H(\alpha,\beta) = \frac{\partial L}{\partial\dot{\alpha}}\dot{\alpha} - L = \beta\dot{\alpha} - L, \tag{3.17}$$

whose corresponding operator has already been defined by Eq. (3.4) using the "quantization" condition[23]

$$\hat{\beta} = -2k_B\frac{\partial}{\partial\alpha}. \tag{3.18}$$

The canonical form of the Euler–Lagrange equation (3.13) is now given by the pair of equations:

$$\dot{\alpha} = \frac{\partial H}{\partial\beta} \quad \text{and} \quad \dot{\beta} = -\frac{\partial H}{\partial\alpha}. \tag{3.19}$$

Introducing the linear transformation

$$\xi = \tfrac{1}{2}(\alpha - \beta/s), \quad \eta = \tfrac{1}{2}(\alpha + \beta/s), \tag{3.20}$$

the canonical equations (3.19) are converted into the desired form:

$$\dot{\xi} + \gamma\xi = R^{-1}X(t), \tag{3.21}$$

$$\dot{\eta} - \gamma\eta = 0. \tag{3.22}$$

Equation (3.21) will be recognized readily as the averaged Langevin equation. In the case of a vanishing external force, the two equations are mirror images of one another; the general solution is a superposition of decaying and growing exponentials that manifest a symmetry in past and future.[24] The presence of the external force destroys this symmetry and also the reversibility of transitions between any two nonequilibrium states. We shall now investigate this symmetry breaking in greater detail.

Setting $\tau = t - t_0$, we write the general solution of the

Euler–Lagrange equation (3.15) as

$$\alpha(\tau) = \alpha_0 \cosh\gamma\tau + (\dot{\alpha}_0/\gamma)\sinh\gamma\tau + R^{-1} \times \int_0^\tau X(t)\exp[-\gamma(\tau - t)]\,dt. \tag{3.23}$$

The time integration in (3.13) can now be performed explicitly and we find

$$A = (s/2\sinh\gamma\tau)[(\alpha^2 + \alpha_0^2)\cosh\gamma\tau - 2\alpha\alpha_0 + 2(\alpha_0 - \alpha e^{\gamma\tau})\Phi + e^{\gamma\tau}\Phi^2], \tag{3.24}$$

where

$$\Phi = R^{-1}\int_0^\tau \exp[-\gamma(\tau - t)]X(t)\,dt. \tag{3.25}$$

The kernel can thus be written in the closed form:

$$G(\alpha,\tau|\alpha_0,0) = (se^{\gamma\tau}/4\pi k_B \sinh\gamma\tau)^{1/2}\exp(-A/2k_B). \tag{3.26}$$

The time-dependent transition probabilities in expression (3.10) are now obtainable by multiplying it by $\phi_n(\alpha)$ and $\phi_m(\alpha_0)$ and integrating over all α and α_0. The result can be written in the form of a matrix element of the kernel

$$\lambda_{nm}(\tau) = \langle\phi_n(\alpha)|G(\alpha,\tau|\alpha_0,0)|\phi_m(\alpha_0)\rangle. \tag{3.27}$$

Of particular interest is the probability of transition from the equilibrium or "ground" state to the first excited state which we shall refer to as the 0→1 transition. The transition probability is:

$$\lambda_{10}(\tau) = \langle\phi_1(\alpha)|G(\alpha,\tau|\alpha_0,0)|\phi_0(\alpha_0)\rangle = (s/k_B)^{1/2}\Phi. \tag{3.28}$$

Calculating the transition probability for the reverse transition, i.e., the 1→0 transition, we obtain the surprising result that $\lambda_{01}(\tau)$ vanishes. This implies that time-varying external fields destroy the principle of microscopic reversibility which equates the frequencies of forward and reverse transition at equilibrium.[25] In Sec. 4, we shall see that this symmetry breaking is due to the Markov assumption which, for Gaussian processes, means that the system does not remember how it got into the given state.

If the reverse transition probabilities did not vanish, then the virtual transitions to other states could take place leading to a modification of the autocovariance expression (2.21). That is, the transition probability for the decay of a fluctuation from the first excited state would be given by

$$\lambda_{11}(\tau) = \langle\phi_1(\alpha)|G(\alpha,\tau|\alpha_0,0)|\phi_1(\alpha_0)\rangle + \sum_n \langle\phi_1(\alpha)|G(\alpha,\tau|\alpha',t)|\phi_n(\alpha')\rangle \times \langle\phi_n(\alpha')|G(\alpha',\tau|\alpha_0,0)|\phi_1(\alpha_0)\rangle, \tag{3.29}$$

with $\tau > t \geqslant 0$. In expression (3.29), we have introduced a complete set of states to take into account virtual transitions. Due to the finding that the transition probabilities of all reverse transitions vanish, expression (3.29) gives the correct temporal dependence of the autocovariance (2.21), viz.,

$$\lambda_{11}(\tau) = \langle\phi_1(\alpha)|G(\alpha,\tau|\alpha_0,0)|\phi_1(\alpha_0)\rangle = \exp(-\gamma\tau). \tag{3.30}$$

Therefore, for Gaussian process that are also Markovian, we find that although time-varying, external fields create excitations, they have no influence on the decay of a fluctuation from a given state.

4. ELEMENTARY EXCITATIONS IN NONSTATIONARY PROCESSES

In this section, we show by comparison with the forced harmonic oscillator, that the inability of time-varying external fields to influence the decay of a fluctuation from a given nonequilibrium state is due to the Markov property. In Sec. 3 we treated the external field as a perturbation in a Schrödinger-type equation describing a harmonic oscillator. In analogy with quantum field theory, it would appear advantageous to introduce the creation and annihilation operators of second quantization[12]

$$\hat{a}^\dagger = \left(\frac{k_B}{s}\right)^{1/2}\left(\frac{s\alpha}{2k_B} - \frac{\partial}{\partial\alpha}\right), \tag{4.1}$$

$$\hat{a} = \left(\frac{k_B}{s}\right)^{1/2}\left(\frac{s\alpha}{2k_B} + \frac{\partial}{\partial\alpha}\right). \tag{4.2}$$

It is easily seen that these operators obey the commutation relations for Bose particles.

In terms of the creation and annihilation operators, the Hamiltonian of the transformed Fokker–Planck equation (3.4) is

$$(2k_B)^{-1}\hat{H} = \gamma\hat{a}^\dagger\hat{a} - (s/k_B)^{1/2}R^{-1}X(t)\hat{a}^\dagger, \tag{4.3}$$

which bears a striking resemblance to the Hamiltonian of a forced harmonic oscillator

$$(2k_B)^{-1}\hat{H}_f = \gamma\hat{a}^\dagger\hat{a} - (s/k_B)^{1/2}R^{-1}X(t)(\hat{a} + \hat{a}^\dagger). \tag{4.4}$$

The Hamiltonian (4.3) can be used to derive the Heisenberg equations of motion for the annihilation and creation operators, viz.,

$$\frac{d\hat{a}}{dt} = (2k_B)^{-1}[\hat{H},\hat{a}] = -\gamma\hat{a} + \left(\frac{s}{k_B}\right)^{1/2}R^{-1}X(t), \tag{4.5}$$

$$\frac{d\hat{a}^\dagger}{dt} = (2k_B)^{-1}[\hat{H},\hat{a}^\dagger] = \gamma\hat{a}^\dagger. \tag{4.6}$$

These equations of motion are analogous to the pair of canonical equations (3.21) and (3.22). Their solutions are

$$\hat{a}(t) = \exp(-\gamma t)\hat{a} + (s/k_B)^{1/2}R^{-1} \times \int_{-\infty}^{t} \exp[-\gamma(t-u)]X(u)\,du, \tag{4.7}$$

$$a^\dagger(t) = \exp(\gamma t)\hat{a}^\dagger, \tag{4.8}$$

which clearly display the asymmetry caused by the external field in the creation and destruction of elementary excitations.

The time dependencies of these operators can be written in a more compact form by defining the operators,

$$\hat{S}_0 = \exp(-\gamma\hat{a}^\dagger\hat{a}t), \tag{4.9}$$

$$\hat{S}_1 = \exp(\Xi^{(+)}\hat{a}), \tag{4.10}$$

with

$$\Xi^{(+)} = (s/k_B)^{1/2}R^{-1}\int_{-\infty}^{t}\exp(\gamma u)X(u)\,du. \tag{4.11}$$

With the aid of these operators, the solutions of the equations of motion, (4.7) and (4.8), can be written in the generic form

$$\hat{A}(t) = \hat{U}^{-1}(t)\hat{A}\hat{U}(t), \tag{4.12}$$

for any operator $\hat{A}$ and $\hat{U}(t)$ is the time development operator,

$$\hat{U}(t) \equiv \hat{S}_0 \hat{S}_1. \tag{4.13}$$

We are now in a position to express the moments as expectation values of particular combinations and products of the creation and annihilation operators over the ground state $|0\rangle = \phi_0$. Likewise, ϕ_1 corresponds to the first excited state, $|1\rangle$, and so on. The mean value $m(t)$ is given by the expectation value

$$\begin{aligned} m(t) &= (k_B/s)^{1/2} \langle 0 | \{ \hat{a}(t) + \hat{a}^\dagger(t) \} | 0 \rangle \\ &= (k_B/s)^{1/2} \langle 0 | \hat{U}^{-1}(t) \hat{a} \hat{U}(t) | 0 \rangle = (2.20). \end{aligned} \tag{4.14}$$

The autocorrelation function is given by the expression

$$\begin{aligned} A(t) &= (k_B/s) \langle 0 | \hat{a}(t) \hat{a}^\dagger(0) | 0 \rangle \\ &= (k_B/s) \langle 0 | \hat{U}^{-1}(t) \hat{a} \hat{U}(t) \hat{a}^\dagger | 0 \rangle \\ &= (k_B/s) \langle 0 | \{ \hat{a} + \Xi^{(+)} \} \hat{a}^\dagger | 0 \rangle \exp(-\gamma t) = (2.21). \end{aligned} \tag{4.15}$$

The steps leading to the final result in (4.15) have been written out in detail in order to demonstrate that the external field has no influence on the decay of fluctuations from a given state. In the language of second quantization, we say that the operator $\Xi^{(+)}\hat{a}^\dagger$ does not have a conjugate operator $\Xi^{(-)}\hat{a}$ which would induce de-excitations. It will now be appreciated that, in comparison with the forced harmonic oscillator, the presence of the conjugate operator $\Xi^{(-)}\hat{a}$ is incompatible with the Markov assumption.

In the case of the forced harmonic oscillator, whose Hamiltonian is given by (4.4), the operator $\hat{S}_1$ in (4.10) must be replaced by

$$\hat{S}_{1f} = \exp[\tfrac{1}{2}(\Xi^{(+)}\hat{a}^\dagger + \Xi^{(-)}\hat{a})], \tag{4.16}$$

where

$$\Xi^{(-)} = (s/k_B)^{1/2} R^{-1} \int_t^\infty \exp(-\gamma u) X(u)\, du. \tag{4.17}$$

Following the same procedure, we now find the mean value and autocovariance are given by

$$\begin{aligned} m_f(t) = (2R)^{-1} \Big\{ &\int_{-\infty}^{t} \exp[-\gamma(t-u)] X(u)\, du \\ &- \int_t^\infty \exp[\gamma(t-u)] X(u)\, du \Big\}, \end{aligned} \tag{4.18}$$

$$\begin{aligned} A_f(t) = (k_B/s) \exp(-\gamma t) \cdot \exp\Big[&-(\gamma/2k_B R) \int_{-\infty}^{t} du \\ &\times \int_t^\infty dv\, X(v) X(u) e^{-\gamma(v-u)} \Big], \end{aligned} \tag{4.19}$$

respectively. The autocovariance expression for the forced harmonic oscillator, (4.19), clearly shows that the time-varying external field has an "influence" on the decay of fluctuations from a given nonequilibrium state.

A result, similar to (4.19), was first obtained by Feynman[21] who used the forced harmonic oscillator as a model to study the interaction of charged particles with the electromagnetic field. Using his Lagrangian formulation of nonrelativistic quantum mechanics, Feynman found it possible to eliminate the coordinates of the field, considered as a set of uncoupled oscillators, and to recast the problem in terms of the coordinates of the particles alone. The effect of the field was to cause a delayed interaction between the particles and it was described by an "influence functional," similar to the second exponential factor in (4.19)[26]

The fact that the system now has the capability to store energy, as witnessed by noninstaneous responses to external perturbations, implies that the process has acquired a "memory" and is therefore no longer Markovian. The time-varying external field can now induce virtual transitions to and from neighboring excited states and these additional correlations manifest themselves in the mode of regression of fluctuations. Furthermore, since the system has the capacity to store energy, the power balance between the rate at which energy is absorbed by the system and the rate at which it is dissipated in the system, as required by the FD theorem, will no longer be fulfilled. Consequently, eventhough the process is Gaussian, a FD theorem will not be valid since it is no longer Markovian.

[1] H. B. Callen and J. H. Welton, Phys. Rev. **83**, 34 (1951). See also: R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
[2] H. Nyquist, Phys. Rev. **32**, 110 (1928).
[3] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, Oxford, 1959), p. 393.
[4] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).
[5] J. Richardson, IRE Trans. Inf. Theory **IT-1**, 62 (1955).
[6] N. G. Van Kampen, Physica **26**, 585 (1960).
[7] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).
[8] M. Lax, Phys. Rev. **157**, 213 (1967); **172**, 350 (1968).
[9] M. Lax, Rev. Mod. Phys. **38**, 359 (1966).
[10] N. Hashitsume, Prog. Theor. Phys. **15**, 369 (1956).
[11] B. H. Lavenda, Nuovo Cimento **7**, 229 (1977).
[12] B. H. Lavenda, F. Fittipaldi, and G. Peluso, Riv. Nuovo Cimento B **42**, 9 (1977).
[13] N. Wiener, Acta Acad. Sci. Hung. Math. **55**, 117 (1930); J. Math. Phys. (Cambridge, Mass.) **5**, 99 (1926).
[14] A. I. Khintchin, Ann. Math. **109**, 604 (1934).
[15] R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1963), Vol. 1, p. 131.
[16] L. P. Kadanoff and P.C. Martin, Ann. Phys. (N.Y.) **24**, 419 (1963).
[17] R. P. Feynman, *Statistical Mechanics: A Set of Lectures* (Benjamin, Reading, Mass., 1972), p. 50.
[18] M. Lax, Rev. Mod. Phys. **32**, 25 (1960).
[19] I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
[20] The presence of a correction term, $\frac{1}{2}\langle \alpha^2 \rangle$, is due to the fact that the random variable $\alpha(t)$ is not of bounded variation and the integral must be interpreted in the stochastic sense of *Itô* as explained in J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953), p. 443. The physical significance of the correction term has been discussed in B. H. Lavenda, Phys. Lett. A **71**, 304 (1979).
[21] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); Phys. Rev. **80**, 440 (1950).
[22] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965), pp. 143ff.
[23] N. Saitô and M. Namiki, Prog. Theor. Phys. **16**, 71 (1956).
[24] B. H. Lavenda and E. Santamato, Lett. Nuovo Cimento **26**, 27 (1979); **26**, 505 (1979).
[25] R. C. Tolman, *Principles of Statistical Mechanics* (Oxford Univ. Press., Oxford, 1938), p. 163.
[26] Noninstantaneous interactions have also been found in coupled, linear irreversible processes that, on account of a wide time scale separation between a primary process and secondary processes, simulate a heavy Brownian particle moving in a field of lighter Brownian particles; see Ref. 12.

Feynman amplitude circuit breakers [a)]

Richard R. Hampton
Villanova University, Villanova, Pennsylvania 19085

(Received 2 November 1979; accepted for publication 24 January 1980)

It is shown that point splitting may be used to regularize arbitrary Feynman amplitudes and Taylor subtraction terms of the type found in the Zimmermann formulation (BPHZ) of renormalized perturbation theory. Point splitting refers to the Fourier transformation of the Feynman integrand with respect to the internal (loop) momenta. It is shown that this has the effect in configuration space of breaking open the circuits of the diagram. The regularization may be removed, interchangeably, before or after the propagator epsilon is taken to zero in the sense of distribution theory. This result is proven elsewhere but is mentioned here for its importance. The result may make the method useful in establishing the unitarity of BPHZ. Two appendices contain a review of circuit-based graph theory and a demonstration that an arbitrary circuit based graph may sometimes have no realization as a graph.

1. INTRODUCTION

A point split Feynman amplitude is the Fourier transform with respect to the internal (i.e., loop) momenta of the integrand corresponding to the diagram for the amplitude. The method can be thought of as a generalization of the Wilson–Zimmermann work on point split operator products. Zimmermann introduced the notion of normal products[1] in which operator products are defined from Green's functions by the LSZ reduction formalism. The point at which the multiplied operators are to be evaluated appears as a special vertex in the Feynman diagrams for the Green's functions. These Green's functions are then calculated from the diagrams by using renormalized perturbation theory. The Wilson expansion[2] for the product of two operators at two different points [e.g., $A(x)A(x+\xi)$] led to the Wilson–Zimmermann proof[3] that this expansion is valid in perturbation theory. In this proof, the splitting of the special vertex in configuration space corresponds in momentum space to a Fourier transform with respect to the internal momenta of the lines incident on the special vertex. This approach was generalized to more complicated classes of operator products by Clark.[4]

In the present work, we study the effect of splitting every loop of an arbitrary Feynman diagram or subtraction term. Since we do not focus on any special vertex, and since the effect is to regularize the ultraviolet divergencies, it seems more appropriate to call the result circuit breaker regularization, but we will often use the earlier term.

The regularization of Feynman diagrams has been shown to be superfluous in that renormalized amplitudes may be calculated without it. In particular, this is true of the Zimmermann formulation (BPHZ)[2,5] of the Bogoliubov–Parasiuk–Hepp (BPH)[6,7] method of renormalization. Although a regularization may suggest a subtraction procedure, it may be possible to independently define the procedure. In the case of BPHZ, the need for regularization is obviated by the combining of subtraction terms before integrating with respect to the internal momenta.

Yet for other purposes, the need for regularization appears difficult to escape. For example, in Lagrangian field theory, it seems impossible to look for a relationship between divergent Lagrangian counter terms and the subtraction procedures of perturbation theory without regularization. Bogoliubov used Pauli–Villars for this purpose.[6]

Partly out of a desire to avoid nonphysical divergencies altogether, the axiomatic viewpoint has leaned toward using the renormalization procedure to define Green's functions which in turn may be used to construct a field theory, perhaps by way of the retarded functions.[8] This approach requires a demonstration of unitarity.

Zimmermann used Pauli–Villars regularization to sketch a heuristic argument for the unitarity of BPHZ.[1] His argument remained incomplete. As he pointed out, he needed to assume the validity of interchanging the propagator limit with the limit in which the regulator is removed.

Circuit breakers permit one to freely exchange these limits.[9,10] They may therefore be helpful in completing the proof of unitarity.

If regularization is, finally, to be thought an unnecessary complication, circuit breakers seem to minimize this complication. Instead of multiplying the integrand by strange factors, inserting additional lines in the diagram, we merely study the Fourier transforms of the integrands. The original value is then recovered as a particular value, that at the origin, of the transform. The study of a function and of its Fourier transform seem, in principle, hardly different. For this reason, point splitting appears to be a peculiarly natural way of regularizing.

2. THE FORMAL MEANING OF POINT SPLIT FEYNMAN AMPLITUDES

Consider a connected graph, Γ, consisting of vertices, $V_1\ V_2,\ldots,\ V_v$ and of lines, $L_{ab\sigma}$, where $L_{ab\sigma}$ is the σth line connecting V_a and and V_b. To each vertex V_a we assign an external momentum q_a (possibly fixed at zero) with $\Sigma_a q_a$

[a)]The research reported in this paper was supported in part by the Max-Planck-Institut für Physik, München, and by the binational German and American Fulbright Commision in Bonn.

 0022-2488/80/071833-09$1.00

$= 0$. To each line $L_{ab\sigma}$ we assign:

(1) an internal momentum $k_{ab\sigma}$ subject to the conditions

(a) $k_{ab\sigma} = -k_{ba\sigma}$

and

(b) $\sum_b k_{ab\sigma} = 0$,

(2) an external line momentum $q_{ab\sigma}$ subject to the conditions

(a) $q_{ab\sigma} = -q_{ba\sigma}$

and

(b) $\sum_{b\sigma} q_{ab\sigma} = q_a$,

(3) a total momentum

$p_{ab\sigma} = k_{ab\sigma} + q_{ab\sigma}$

and

(4) a propagator, $\tilde{\Delta}_{ab\sigma}(p_{ab\sigma})$.

We specify an arbitrary specific routing of the external momentum by means of a set of external momentum routing coefficients. These are numbers $c^{(j)}_{ab\sigma}$, which define a solution of the conditions in (2) through the equation

$$q_{ab\sigma} = \sum_{j=1}^{v} c^{(j)}_{ab\sigma} q_j. \tag{2.1}$$

Condition (2) subjects the external momentum routing coefficients to the equivalent conditions:

$$c^{(j)}_{ab\sigma} = -c^{(j)}_{ba\sigma} \quad \text{and} \quad \sum_{b\sigma} c^{(j)}_{ab\sigma} = \delta^j{}_a. \tag{2.2}$$

The set of vertices V_a for which q_a is fixed at zero will be called internal (int) and the complement of this set will be referred to as the external (ext).

Wherever no confusion results, we use the multi-index notation

$$(dk) = \prod_{\substack{ab\sigma \\ a>b}} dk_{ab\sigma}, \quad (q) = (q_1,\ldots,q_v), \quad (dq) = \prod_a dq_a,$$

$$(dq)_{\text{ext}} = \prod_{a\in\text{ext}} dq_a, \quad \text{and} \quad (dq)_{\substack{\text{ext} \\ a\neq 1}} = (dq)_{\text{ext}}/dq_1.$$

The unsubtracted amplitude associated with the graph is then "formally"

$$\tilde{\mathscr{F}}_\Gamma(q) = \int (dk) \left[\prod_{\substack{ab\sigma \\ a>b}} \tilde{\Delta}_{ab\sigma}(p_{ab\sigma})\right]\left[\prod_{\substack{a\in \\ \text{int}}} \delta(q_a)\right]\left[\delta\left(\sum_a q_a\right)\right] \times \left[\prod_{a\neq 1} \delta\left(\sum_{b\sigma} k_{ab\sigma}\right)\right]. \tag{2.3}$$

Note that the momentum conservation at the arbitrarily chosen first vertex is omitted since it is already implied by the remaining conditions.

In order to define a point split amplitude, one further assigns to each line:

(5) a breaking (i.e., splitting) parameter $\xi_{ab\sigma}$, subject, for convenience, to the condition

$\xi_{ab\sigma} = -\xi_{ba\sigma}$.

The point split amplitude is now formally obtained by replacing the integral over the internal momentum, in Eq. (2.3), with the Fourier transform with respect to the internal line momenta. The breaking parameters are then the coordinates conjugate to the internal line momenta.

We complete this section by formally examining the point split amplitude in coordinate space:

$$\begin{aligned}
&\mathscr{J}_\Gamma(z,\xi) \\
&= \int (dq) \exp(i\textstyle\sum_a q_a z_a)\, \tilde{\mathscr{F}}_\Gamma(q,\xi) \\
&= \int (dq) \exp(i\textstyle\sum_a q_a z_a) \int (dk) \left[\prod_{\substack{ab\sigma \\ a>b}} \tilde{\Delta}_{ab\sigma}(p_{ab\sigma}) \times \exp(ik_{ab\sigma}\xi_{ab\sigma})\right] \\
&\quad \times \left[\prod_{\substack{a\in \\ \text{int}}} \delta(q_a)\right]\left[\prod_{a\neq 1} \delta\left(\sum_{b\sigma} k_{ab\sigma}\right)\right]\delta\left(\sum_a q_a\right) \\
&= \int (dq)_{\text{ext}} (dk)(dx) \exp\left(i\sum_{\substack{a\in \\ \text{ext}}} q_a z_a + i\sum_a x_a \sum_{b\sigma} k_{ab\sigma}\right)\delta(x_1) \\
&\quad \times \delta\left(\sum_{\substack{a\in \\ \text{ext}}} q_a\right)\left[\prod_{\substack{ab\sigma \\ a>b}} \tilde{\Delta}_{ab\sigma}(p_{ab\sigma}) \exp(ik_{ab\sigma}\xi_{ab\sigma})\right] \\
&= \int (dq)_{\text{ext}} (dp)(dx) \exp\left[i\sum_{a\in\text{ext}} q_a z_a + \tfrac{1}{2}\sum_{ab\sigma} \xi_{ab\sigma}\left(p_{ab\sigma} - \sum_{j\in\text{ext}} c^{(j)}_{ab\sigma} q_j\right) + \sum_a x_a\left(-q_a + \sum_{b\sigma} p_{ab\sigma}\right)\right]\delta(x_1) \\
&\quad \times \left[\prod_{\substack{ab\sigma \\ a>b}} \tilde{\Delta}_{ab\sigma}(p_{ab\sigma})\right]\delta\left(\sum_{\substack{a\in \\ \text{ext}}} q_a\right) \\
&= \int (dq)_{\text{ext}} (dx) \exp\left[i\sum_{j\in\text{ext}} q_j\left(z_j - \tfrac{1}{2}\sum_{ab\sigma} \xi_{ab\sigma} c^{(j)}_{ab\sigma} - x_j\right)\right]\delta(x_1) \\
&\quad \times \left[\prod_{\substack{ab\sigma \\ a>b}} \Delta_{ab\sigma}(\xi_{ab\sigma} + x_a - x_b)\right]\delta\left(\sum_{\substack{a\in \\ \text{ext}}} q_a\right) \\
&= \int (dq)_{\substack{\text{ext} \\ a\neq 1}} (dx) \exp\left\{i\sum_{\substack{j\in\text{ext} \\ j\neq 1}} q_j\left[z_j - z_1 - \tfrac{1}{2}\sum_{ab\sigma} \xi_{ab\sigma}(c^{(j)}_{ab\sigma} - c^{(1)}_{ab\sigma}) - x_j + x_1\right]\right\}\delta(x_1) \prod_{\substack{ab\sigma \\ a>b}} \Delta_{ab\sigma}(\xi_{ab\sigma} + x_a - x_b) \\
&= \int (dx)\left[\prod_{\substack{j\in\text{ext} \\ j\neq 1}} \delta\left(z_j - z_1 - \tfrac{1}{2}\sum_{ab\sigma} \xi_{ab\sigma}(c^{(j)}_{ab\sigma} - c^{(1)}_{ab\sigma}) - x_j\right)\right] \\
&\quad \times \left[\prod_{\substack{ab\sigma \\ a>b}} \Delta_{ab\sigma}(\xi_{ab\sigma} + x_a - x_b)\right]\delta(x_1).
\end{aligned} \tag{2.4}$$

In the above, we have made use of the properties described in the line assignments (1) to (5). We now perform a translation in every x_j by an amount $(-z_1 + \frac{1}{2}\Sigma_{ab\sigma}\xi_{ab\sigma}c^{(1)}_{ab\sigma})$, and we assume, to keep the notation a little simpler, that $\Delta_{ab\sigma}(z) = \Delta_{ab\sigma}(-z)$. This results in

$\mathscr{J}_\Gamma(z,\xi)$

$$= \int (dx)_{\text{int}} \left[\prod_{\substack{ab\sigma \\ a>b \\ \wedge(a\in\text{ext})\wedge(b\in\text{int})}} \Delta_{ab\sigma}\left(z_a - \tfrac{1}{2}\sum_{gd\rho} \xi_{gd\rho}\, c^{(a)}_{gd\rho} - x_b \right) \right]$$
$$\times \left[\prod_{\substack{ab\sigma \\ (a>b)\wedge(a\in\text{int})\wedge(b\in\text{int})}} \Delta_{ab\sigma}(x_a - x_b + \xi_{ab\sigma}) \right]$$
$$\times \left[\prod_{\substack{ab\sigma \\ (a>b)\wedge(a\in\text{ext})\wedge(b\in\text{ext})}} \Delta\left(z_a - z_b - \tfrac{1}{2}\sum_{gd\rho} \xi_{gd\rho}(c^{(a)}_{gd\rho} - c^{(b)}_{gd\rho}) \right) \right]. \tag{2.5}$$

The geometrical effect of the splitting parameters is thus to change the $(x_a - x_b)$ of internal propagators by an amount $\xi_{ab\sigma}$, thus splitting apart, as it were, the end points of the internal line and, as it were, breaking open the loop, if there is one involved. The other effect is to introduce a kind of additive displacement of the external vertices, a displacement which is linear and homogeneous in (ξ) and which depends on the internal momentum routing coefficients.

This expression, Eq. (2.5), has the difficulty that it is not clearly defined for the case that the propagators are chosen to be Feynman propagators. Bogoliubov and Shirkov,[3] for example, derive

$$\Delta_F(x) = \frac{1}{4\Pi}\delta(x^2) - \frac{m}{8\pi\sqrt{x^2}}\theta(x^2) \times \left[J_1(m\sqrt{x^2}) - iN_1(m\sqrt{x^2}) \right] + \frac{im\theta(-x^2)}{4\pi^2\sqrt{-x^2}} K_1(m\sqrt{-x^2}),$$

where J_1, N_1, and K_1 are the Bessel, Neuman, and modified Bessel functions, respectively, of first order, and m is the mass. The term-by-term singularities of this expression make it inconvenient for studying the convergence properties of $\mathscr{F}(z, \xi)$. An examination in Feynman parameter space, however, proves quite tractable. In fact, the singularities effectively cancel for appropriately chosen breaking parameters in arbitrary neigborhoods of the breaking parameter origin.

3. RENORMALIZED AMPLITUDES: A REVIEW OF BPHZ AND FORM OF THE TYPICAL TERM

To be complete, point splitting will be studied in the context of a renormalization prescription. We choose BPHZ because it is clear, reasonably simple, and, in common with some, but not all other procedures, it is valid for all diagrams, including those with overlapping divergencies. We begin with a brief review of the Zimmermann forest formula.[1,5]

A Γ forest, U, is a set of nonoverlapping proper divergent subgraphs of the graph Γ. We designate the lines of any graph, γ, by $|\gamma|$. For any $I \subseteq |\Gamma|$, the reduced graph Γ/I is defined as the graph that results from reducing each line of I to a single vertex. Each line of Γ that was connected to the lines of I is then connected instead, at the same end, to this new vertex. The lines of I itself are deleted. We define

$$\bar{\Gamma}(U) = \Gamma / \bigcup_{i=1}^{c} |\gamma_i|,$$

where $\gamma_1, \ldots, \gamma_c$ are the maximal subgraphs of Γ in U. The set of all the $\tilde{\gamma}(U)$ with $\gamma \in U$ partitions Γ into disjoint parts so that every line and every vertex of Γ is a line or vertex of one and only one such reduced graph.

For each proper $\gamma \subseteq \Gamma$, subgraph momentum variables are defined in relationship to the subgraph in the same way as they previously were for the graph, i.e., as linear solutions to

$$k^{(\gamma)}_{ab\sigma} + q^{(\gamma)}_{ab\sigma} = k_{ab\sigma} + q_{ab\sigma}, \tag{3.1}$$

$$\sum_{b\sigma}{}^{\gamma} k^{(\gamma)}_{ab\sigma} = 0, \tag{3.2}$$

and

$$\sum_{b\sigma}{}^{\gamma} q^{(\gamma)}_{ab\sigma} = q^{(\gamma)}_a = \sum_{b\sigma}{}^{\gamma} (k_{ab\sigma} + q_{ab\sigma}), \tag{3.3}$$

where $\Sigma_{b\sigma}{}^{\gamma}$ means that the values of the indices $ab\sigma$ are restricted to those which indicate lines of γ. The solutions allowed are further restricted by the requirement that they be "admissible," i.e., of the form

$$k^{(\gamma)}_{ab\sigma} = k_{ab\sigma}(k),$$

or, in other words, independent of q. An example of such a solution is the set of canonical momenta.[1,5]

For convenience, we define forest variables $k^{(U)}_{ab\sigma}$ and $q^{(U)}_{ab\sigma}$ according to the rules

$$k^{(U)}_{ab\sigma} = k^{(\gamma)}_{ab\sigma}, \quad q^{(U)}_{ab\sigma} = q^{(\gamma)}_{ab\sigma} \tag{3.4}$$

where γ is the element of U for which $L_{ab\sigma} \in |\bar{\gamma}(U)|$.

Subgraph variables may be expressed through the use of momentum routing coefficients:

$$q^{(\gamma)}_{ab\sigma} = \sum_j c^{(j)\gamma}_{ab\sigma} q^{\gamma}_j, \quad \text{and} \quad k^{(\gamma)}_{ab\sigma} = \sum_j d^{(j)\gamma}_{ab\sigma} k_j, \tag{3.5}$$

with

$$\sum_{b\sigma} d^{(j)\gamma}_{ab\sigma} = 0, \quad \text{and} \quad \sum_{b\sigma} c^{(j)\gamma}_{ab\sigma} = \delta^j_a. \tag{3.6}$$

The forest formula then states that the renormalized Feynman amplitude in momentum space is the integral over any admissable set of independent (loop) momenta of

$$R_\Gamma = S_\Gamma \sum_{U\in F} \left[\prod_{\gamma\in U} (-t_\gamma S_\gamma) \right] I_\Gamma(U), \tag{3.7}$$

where

(a) $$I_\Gamma(U) = \prod_{\substack{ab\sigma: \\ L_{ab\sigma}\in|\Gamma|}} \tilde{\Delta}^{(\epsilon)}_{F_{ab\sigma}}(k^{(U)}_{ab\sigma} + q^{(U)}_{ab\sigma}),$$

(b) S_γ causes the variables of all subgraphs of γ to be expressed in terms of the variables of γ,

(c) t_γ is the Taylor operator of degree appropriate to γ (i.e., greater than or equal to the superficial degree of divergence of γ) in the variables q^γ_a,

(d) the product over graphs is taken according to the rule that whenever $\mu \subset \gamma$, then $(-t_\mu S_\mu)$ is placed to the right of $(-t_\gamma S_\gamma)$,

(e) F is the class of all Γ-forests, and

(f). the product over elements of the empty forest has unit value.

Lemma 3.1: The forest formula may be rewritten as

$$R_\Gamma = S_\Gamma \sum_{U\in F} I_{\bar\Gamma(U)} \prod_{\gamma\in U} [-t_\gamma S_\gamma I_{\bar\gamma(U)}(U)],$$

where the factors are written according to Eq. (3.7 d), and each Taylor operator operates on everything to the right of it.

Proof: Since $\bar\gamma_1 \cap \bar\gamma_2 = \emptyset$ unless $\gamma_1 = \gamma_2$,

$$I_\Gamma(U) = \prod_{\gamma\in U} [I_{\bar\gamma(U)}(U)]. \tag{3.8}$$

S_γ operators only on variables k^μ and q^μ such that $\mu \subset \gamma$, and t_γ operates only on expressions which are written in terms of q^γ. For this reason $t_\mu S_\mu$ commutes with any factors $I_{\bar\gamma(U)}$ such that either $\mu\subset\gamma$ or $\mu\cap\gamma = \emptyset$. Lemma 3.1 then follows directly from Eqs. (3.7) and (3.8).

Theorem 3.1. The variables $k_{ab\sigma}$ and $k^{(U)}_{ab\sigma}$ are related by an invertible linear transformation.

Proof: The fact that $k^{(U)}_{ab\sigma} = k^{(U)}_{ab\sigma}(U)$ is required by the definition of k^U, Eqs. (3.1)–(3.4), and the admissibility requirement. We will prove the existence of the inverse relation by an induction. To this purpose we define $U^0 = \emptyset$, and

$$U^j = U^{j-1} \cup \{\gamma |\ \gamma \text{ is minimal in } U - U^{j-1}\}. \tag{3.9}$$

We recall that an element of a forest is said to be minimal in the forest iff the forest contains no subdiagram of that element. It is clear that

$$(\exists j) \quad U^j = U, \tag{3.10}$$

since otherwise U could not have a finite number of elements.

Since $\bar\Gamma(\emptyset) = \Gamma$, Eq. (3.4) trivially gives for all lines

$$q^{(U^0)}_{ab\sigma} = q_{ab\sigma}, \text{ and } k^{(U^0)}_{ab\sigma} = k_{ab\sigma},$$

proving Theorem 3.1 for this case.

Now suppose that Theorem 3.1 is true for U replaced by U^j. If U is replaced by U^{j+1}, each line of Γ will fall into one of three categories:

(i) $L_{ab\sigma} \in |\bar\Gamma^{(U^{j+1})}| \subseteq |\bar\Gamma^{(U^j)}|$, that is, the line is not in any graph of either U^{j+1} or U^j with the possible exception of Γ itself. Then, by definition,

$$k_{ab\sigma} = k^{U^{j+1}}_{ab\sigma} = k^{U^j}_{ab\sigma}.$$

(ii) $\gamma \in U^{j+1}$, $\gamma \in U^j$, and $L_{ab\sigma} \in |\bar\gamma(U^j)|$. In this case we keep the previous solutions of Eqs. (3.1)–(3.3), that is

$$k^{U^{j+1}}_{ab\sigma} = k^{U^j}_{ab\sigma} = k^{\gamma}_{ab\sigma}.$$

(iii) $\gamma \in U^{j+1}$, and $\gamma \notin U^j$, and $L_{ab\sigma} \in |\bar\gamma(U^{j+1})|$.

For case (iii) we rewrite Eq. (3.1):

$$k_{ab\sigma} = k^{U^{j+1}}_{ab\sigma} + q^{U^{j+1}}_{ab\sigma} - q_{ab\sigma}, \tag{3.11}$$

where $q^{U^{j+1}}_{ab\sigma}$ is a linear combination of the q^γ_c. From Eq. (3.3) we find

$$q^\gamma_c = \sum_{b\sigma}{}^{\gamma} q^\gamma_{cb\sigma} = \sum_{b\sigma}{}^{\gamma} (k_{cb\sigma} + q_{cb\sigma}).$$

Now c is vertex of γ, and the members of a forest do not overlap. Furthermore, by the definition of U^j, γ is a maximal member (i.e., is not contained in any other member) of U^{j+1}. Therefore every line ending at c must be either in γ or in $\bar\Gamma(U^{j+1})$. Equation (3.2) with $\gamma = \Gamma$ requires

$$0 = \sum_{b\sigma} k_{cb\sigma} = \sum_{b\sigma}{}^{\bar\Gamma(U^{j+1})} k_{cb\sigma} + \sum_{b\sigma}{}^{\gamma} k_{cb\sigma}.$$

This results in

$$\begin{aligned} q^\gamma_c &= \sum_{b\sigma}{}^{\gamma} q_{cb\sigma} - \sum_{b\sigma}{}^{\bar\Gamma(U^{j+1})} k_{cb\sigma} \\ &= \sum_{b\sigma}{}^{\gamma} q_{cb\sigma} - \sum_{b\sigma}{}^{\bar\Gamma(U^{j+1})} k^{U^{j+1}}_{cb\sigma}, \end{aligned}$$

by case (i) of the present argument. Now we insert this result into $q^{U^{j+1}}_{ab\sigma}$ in the expression for $k_{ab\sigma}$, Eq. (3.11). The admissibility requirement is that the internal momenta be independent of external momenta. Applying this to Eq. (3.11) produces the conclusion that the combination

$$q^{U^{j+1}}_{ab\sigma} - q_{ab\sigma}$$

is independent of q_a. Therefore in case (iii) Eq. (3.11) exhibits $k_{ab\sigma}$ as a linear function of the $k^{U^{j+1}}_{cd\rho}$.

In summary, in cases (i) and (iii) we have explicitly exhibited $k_{ab\sigma}$ as a linear combination of the $k^{U^{j+1}}_{cd\rho}$.

If $L_{ab\sigma}$ is a line of case (ii), the induction hypothesis tells us that $k_{ab\sigma}$ is a linear combination of the $k^{U^j}_{cd\rho}$. Then either

$$k^{U^j}_{cd\rho} = k^{U^{j+1}}_{cd\rho},$$

for cases (i) and (ii), or $L_{cd\rho}$ belongs to case (iii), whereupon $L_{cd\rho} \in \bar\Gamma(U^j)$, and therefore

$$k^{U^j}_{cd\rho} = k_{cd\rho} = k_{cd\rho}\,(k^{U^{j+1}}),$$

by the argument of case (iii). We conclude that Theorem 3.1 is true by induction on j.

For convenience, we shall drop the operator S_Γ from the forest formula, Eq. (3.7). This has no effect on the value of R_Γ since S_Γ only determines the values used to express R_Γ. We examine the terms arising from any particular forest U in the expression of Lemma 3.1. Each $I_{\bar\gamma}(U)$ is expressed in terms of the subgraph variables k^U and q^U. Let us examine the case that γ is a minimal element of U. Then S_γ has no effect since $\bar\gamma(U) = \gamma$. The only factors to the right of $I_{\bar\gamma}(U)$ are associated with subgraphs which are disjoint from γ. It is clear that t_γ operates only the factor $I_\gamma(U)$, which is a product of propagators, one for each line of γ, of the form:

$$\begin{aligned} \tilde\Delta^{(\epsilon)}_{F_{ab\sigma}} = i/\{&(k^\gamma_{ab\sigma} + q^\gamma_{ab\sigma})^2 - M^2 \\ &+ i\epsilon[(k^\gamma_{ab\sigma} + q^\gamma_{ab\sigma})^2_E + M^2]\}. \end{aligned}$$

we here make the simplifying assumption that all propagators are associated with the same positive mass, M; it should be possible to generalize from this. Therefore, when γ is minimal, $t_\gamma I_\gamma(U)$ contains terms which, up to constant factors, are all of the form

$$\begin{aligned} &\prod_{\substack{ab\sigma: \\ L_{ab\sigma}\in|\gamma|}} \{(k^\gamma_{ab\sigma})^2 - M^2 + i\epsilon[(k^\gamma_{ab\sigma})^2_E + M^2]\}^{-n_{ab\sigma}} \\ &\times \prod_{a,\nu}(q_{a\nu})^{n_{a\nu}} \prod_{i,\mu}(k_{i\mu})^{n'_{i\mu}}. \end{aligned} \tag{3.12}$$

We proceed from this to develop the general form.

Theorem 3.2: The terms in R_Γ corresponding to a given Γ forest U are all, to within multiplicative constants, of the form:

$$I_{\bar\Gamma(U)}(U)\prod_{\substack{i=m\\ i=1\\ \nu=0}}^{\nu=3}(k_{i\nu}^{U})^{n_{i\nu}}\prod_{\substack{j=\upsilon\\ j=1\\ \mu=0}}^{\mu=3}(q_{j\mu})^{n'_{j\mu}} \times \prod_{\substack{ab\sigma\\ L_{ab\sigma}\in\\ |\Gamma|-|\bar\Gamma(U)|}} \{(k_{ab\sigma}^{U})^2 - M^2 + i\epsilon[(k_{ab\sigma}^{U})_E^2 + M^2]\}^{-n_{ab\sigma}}$$

where the k_i^{U} are any m independent internal forest momenta, and the various $n_{i\nu}, n'_{j\mu}$, and $n_{ab\sigma}$ are all non-negative integers.

Proof: We again use the U^j defined in the proof of Theorem 3.1. For $U = U^0 = \emptyset$, $\bar\Gamma = \Gamma$, and the theorem follows trivially with the various "n" exponents all zero.

Let $\gamma_1, \gamma_2, \ldots, \gamma_c$ be the maximal elements of $U = U^{j+1}$, and set $U' = U^j$. Suppose that Theorem 2.2 is true if U is replaced by U' and use this with Lemma 2.1 to write the terms associated with U in the form:

$$I_{\bar\Gamma}(U)\Big(\prod_{j=1}^{c}(-t_{\gamma_j}S_{\gamma_j}I_{\bar\gamma_j(U)})\Big) \times (I_{\bar\Gamma(U')})^{-1}I_{\bar\Gamma(U')}\Big(\prod_{i,\nu}(k_{i\nu}^{U'})^{n_{i\nu}}\Big) \times\prod_{j,\mu}(q_{j\mu}^{U'})^{n'_{j\mu}}\prod_{\substack{ab\sigma:\\ L_{ab\sigma}\in|\Gamma|-|\bar\Gamma(U')|}} \{(k_{ab\sigma}^{U'})^2 - M^2 + i\epsilon \times [(k_{ab\sigma}^{U'})_E^2 + M^2]\}^{-n_{ab\sigma}}.$$

Note that the lines of each $\bar\gamma_j(U)$ are excluded from the last product. Also for this last product, the lines all belong to case (ii) of Theorem 3.1, and therefore $k_{ab\sigma}^{U'}$ may be replaced by $k_{ab\sigma}^{U'}$. These factors are then all independent of q_j^{U} and therefore may be commuted to the left of the remaining Taylor operators. We next use Theorem 3.1 to rewrite the powers of $k_{i\nu}^{U}$ as polynomials in $k_{i\nu}^{U}$ which also commute with the Taylor operators. Using Equations (3.1) and (3.3), the factors of $q_a^{U'}$ may be rewritten as polynomials in q_a^{U} and k_j^{U}. The effect of the Taylor operators on these does not alter the form. The effect on the factors of $I_{\bar\gamma_j(U)}$ is just to produce more factors of the form of Eq. (3.12). Thus Theorem 3.2 is proved by induction on the U^j.

4. CIRCUIT BREAKER REGULARIZATION

We develop explicit forms for the point split, or open circuit, terms.

Definition 4.1: The point split Feynman amplitude for positive ϵ corresponding to a graph Γ (the Γ-psFa), $J_\Gamma^{(\epsilon)}(\xi,q)$ is

$$J_\Gamma^{(\epsilon)}(\xi,q) = \int (dk)\, R_\Gamma^{(\epsilon)}(k,q)\exp\Big(i\sum_{\substack{l:\\ l\in|\Gamma|}} k_l\xi_l\Big),$$

where $R_\Gamma^{(\epsilon)}(k,q)$ is given by the forest formula, Eq. (3.7), and

$$(dk) = \prod_{j=1}^{m} dk_j,$$

where m is the number of independent line momenta (or circuits) in Γ, and $k_1, k_2, \ldots, k_m$ constitute a basis.

As a result of Zimmermanns' convergence proof,[1,5] the Γ-psFa integral coverges absolutely and uniformly in ξ.

It would seem natural to define a point split forest term (psft) for a graph from the expression for the Γ-psFa by replacing $R_\Gamma^{(\epsilon)}$ by the typical forest term given in Theorem 3.2.

However we find it convenient to first express the typical forest term as an integral over a Feynman parameter space by recourse to the usual Gaussian form identity:

$$(\Delta)^n = \int_{(\beta)=0}^{(\infty)}(d\beta)\exp\Big(-\sum_{j=1}^{n}\beta_j\Delta^{-1}\Big). \tag{4.1}$$

We make the usual change of variables defined by

$$\lambda = \sum_{j=1}^{L}\beta_j,\quad \beta_j = \lambda\alpha_j,\quad \alpha_L = 1 - \sum_{j=1}^{L-1}\alpha_j. \tag{4.2}$$

Equation (4.2), together with the positivity of all β_j, result in the usual properties of the Feynman parameters:

$$\begin{aligned}&\text{(a) } \beta_L = \Big(1 - \sum_{j=1}^{L-1}\alpha_j\Big)\lambda,\\ &\text{(b) } 0 \leqslant \lambda \leqslant \infty,\\ &\text{(c) } 0 \leqslant \beta_j/\sum_{i=1}^{L}\beta_i = \beta_j/\lambda = \alpha_j \leqslant 1 \text{ and}\\ &\qquad 0 \leqslant \sum_{j=1}^{L-1}\alpha_j \leqslant 1,\\ &\text{(d) } 0 \leqslant \alpha_L \leqslant 1, \text{ and}\\ &\text{(e) } (d\beta) = d\lambda\,(d\alpha)\,\lambda^{L-1}.\end{aligned} \tag{4.3}$$

The region for the (α) integration, $\mathbf{R}_\alpha$, is defined by Eqs. (4.3c) and (4.3d).

We use Eq. (4.2) in Eq. (4.1), with $L = n$, and obtain

$$\begin{aligned}(\Delta)^n &= \int_{\mathbf{R}_\alpha} d\alpha \int_{\lambda=0}^{\infty} d\lambda\,\lambda^{n-1}\exp(-\lambda\Delta^{-1})\\ &= \frac{1}{(n-1)!}\int_{\beta=0}^{\infty} d\beta\,\beta^{n-1}\exp(-\beta\Delta^{-1}),\end{aligned} \tag{4.4}$$

where the last integral is obtained by another change of variables and the well-known value of the (α) integration. This results in

$$\prod_{\substack{l:\\ l\in|\Gamma|}}(\Delta_{F_l}^{(\epsilon)})^n = \prod_l[(n_l-1)!]^{-1} \times\int_{\beta_l=0}^{\infty} d\beta_l\,\beta_l^{n_l-1}\exp[-\beta_l(\Delta_{F_l}^{(\epsilon)})^{-1}] \tag{4.5}$$

$$= \prod_l[(n_l-1)!]^{-1}\int_{\mathbf{R}_\alpha}(d\alpha)\prod_l(\alpha_l^{n_l-1}) \times\int_{\lambda=0}^{\infty} d\lambda\,\lambda^{N-1}\exp\Big[-\lambda\sum_l\alpha_l(\Delta_{F_l}^{(\epsilon)})^{-1}\Big], \tag{4.6}$$

where $\sum_l n_l = N$. Equation (4.6) is obtained from Eq. (4.5) by using the change of variables, Eq. (4.2), once more, this time with L being the number of lines of Γ.

Definition 4.2: A point split forest term (psft) for a graph Γ is obtained from the expression for the Γ-psFa, Definition 4.1, by substituting the parameterized form of the typical forest term, Eq. (4.6), for $R_\Gamma^{(\epsilon)}$ and performing the

internal momenta (k) integrations before performing the parameter integrations.

Dropping multiplicative constants, the psft is then of the form:

$$T^{(\epsilon)}(\xi,q)=\prod_{a,\nu}(q_{a\nu})^{n'_{a\nu}}\int_{\mathbf{R}_\alpha}(d\alpha)\times\prod_l(\alpha_l^{n_l-1})\int_{\lambda=0}^{\infty}d\lambda\,\lambda^{N-1}\int(dk)\exp(G)\prod_{i,\mu}(k_{i\mu}^U)^{n_{i\mu}} \tag{4.7}$$

with

$$G=\sum_l\{-\lambda\alpha_l[\Delta_{F_l}^{(\epsilon)}(p_l^{(U)})]^{-1}+ik_l\xi_l\}. \tag{4.7a}$$

$$p_l^U=k_l^U+q_l^U=k_l+q_l,\quad \text{for } l\in|\bar\Gamma(U)| \tag{4.7b}$$
$$=k_l^U=k_l^\gamma \text{ for } l\in|\bar\Gamma(U)|, \text{ and}$$

$$(dk)=\prod_{i=1}^m dk_i. \tag{4.7c}$$

By Theorem 3.1

$$k_l=\sum_{i=1}^m d_l^i k_i^U \text{ and}$$
$$k_l^U=\sum_{i=1}^m d_l^{Ui}k_i^U, \tag{4.8}$$

where the lines have been numbered so that the first k_i^U are any m independent forest momenta. The numbers d_l^{Ui} and d_l^i are referred to as forest routing and momentum routing coefficients, respectively, and are related to circuit coefficients.[11]

For convenience we will suppress the line indices when they are summed over and use

$$\chi_i=d^i\xi. \tag{4.9}$$

Our principal result follows.

Theorem 4.1: Circuit breaking is a regularization. That is, for any psft there are real values of the breaking parameters in every neighborhood of the origin for which the integrals defining the psft converge; for these values of the breaking parameters, if the integrals over the internal momenta are performed first, then the integrals over the Feynman parameters, Eq. (4.7), are absolutely convergent (ac).

Remark: The necessary restrictions on the breaking parameters will be stated at the end.

Proof: We first study the case that all the $n_{i\mu}$ in Eq. (4.7) vanish. Ignoring the factors of external momenta, Eq. (4.7) may then be replaced by

$$T=\int_{\mathbf{R}_\alpha}(d\alpha)\prod_{l\in|\bar\Gamma|}(\alpha_l^{n_l-1})\int_{\lambda=0}^{\infty}d\lambda\,\lambda^{N-1}\int(dk^U)\times\exp\Big\{-\sum_{l\in|\bar\Gamma|}\lambda\alpha_l[\Delta_{F_l}^{(\epsilon)}(p_l^U)]^{-1}+i\sum_{j=1}^m(\chi_jk_j^U)\Big\}. \tag{4.10}$$

The internal momenta integrations are performed in the usual way by diagonalizing the positive semidefinite matrix A whose components are

$$A_{ij}=d^{Ui}\alpha d^{Uj}, \tag{4.11}$$

where the summation has been suppressed, as previously noted. The diagonalization is accomplished by means of an orthogonal transformation:

$$\mathcal{O}k\to k,\quad A\to A'=\mathcal{O}A\tilde{\mathcal{O}},\quad \det\mathcal{O}=1. \tag{4.12}$$

$\mathcal{O}$ defines rotated forest coefficients

$$D_l^i=\sum_{j=1}^m\mathcal{O}_{ij}d_l^{Uj}, \tag{4.13}$$

which become an orthogonal set on scaling by $\sqrt{\alpha_l}$:

$$\mathcal{D}_l^i=\sqrt{\alpha_l}\,D^i,\quad \text{and}\ \mathcal{D}^i\mathcal{D}^j=0\ \text{ for } i\neq j, \tag{4.14}$$
$$\mathcal{D}^i\mathcal{D}^j=(\mathcal{O}A\tilde{\mathcal{O}})_{ij}$$

According to Eq. (4.7b), the external forest momenta q_a^U enter into Eq. (4.10) only in the lines of $\bar\Gamma(U)$. Therefore we may, for convenience, redefine

$$q_l=0 \text{ for } l\notin|\bar\Gamma(U)|,$$
$$=q_{ab\sigma}=\sum_j c_{ab\sigma}^j q_j \text{ for } l=L_{ab\sigma}\in|\bar\Gamma(U)|, \tag{4.15}$$

and

$$Q_l=\sqrt{\alpha_l}\,q_l.$$

The application of Eq. 4.12 to 4.10 then results in

$$T=\int_{\mathbf{R}_\alpha}(d\alpha)\prod_{l\in|\bar\Gamma|}(\alpha_l^{n_l-1})\int_{\lambda=0}^{\infty}d\lambda\,\lambda^{N-1}\int(dk^U)\,e^G, \tag{4.16}$$

with

$$G=i\sum_j\Big(\Xi_jk_j^U+\lambda\{(\mathcal{D}^j)^2(k_j^u)^2+2(\mathcal{D}^jQ)k_j^U+i\epsilon[(\mathcal{D}^j)^2(k_j^U)_E^2+2(\mathcal{D}^jQ)k_{j_E}^U]\}\Big)+i\lambda[Q^2-M^2+i\epsilon(Q_E^2+M^2)],$$

where

$$\Xi_j=\sum_{i=1}^m\mathcal{O}_{ji}\chi_i. \tag{4.18}$$

Completing the squares in the internal momenta of Eq. (4.17) and applying the appropriate (Wick) rotations of contours in the complex k^U planes results in[9]

$$T=\left[\frac{\pi^2}{4(i+\epsilon)(1+\epsilon^2)^{1/2}}\right]^m\times\int_{\mathbf{R}_\alpha}(d\alpha)\frac{\prod_{l\in|\bar\Gamma|}\alpha_l^{n_l-1}}{\left[\prod_{j=1}^m(\mathcal{D}^j)^2\right]^2}\int_{\lambda=0}^{\infty}d\lambda\,\lambda^{N-2m-1}\,e^F, \tag{4.19}$$

where

$$F=i\sum_{j=1}^m\frac{(Q\mathcal{D}^j)\Xi_j}{(\mathcal{D}^j)^2}+i\lambda\left[Q^2-M^2+i\epsilon(Q_E^2+M^2)-\sum_{j=1}^m\frac{(Q\mathcal{D}^j)^2+i\epsilon(Q\mathcal{D}^j)_E^2}{(\mathcal{D}^j)^2}\right]+\frac{i}{4\lambda(1+\epsilon^2)}\sum_{j=1}^m\frac{-(\Xi_j)^2+i\epsilon(\Xi_j)_E^2}{(\mathcal{D}^j)^2}. \tag{4.20}$$

According to Eq. (4.14), the $\mathcal{D}^j$ constitute an orthogonal set. Therefore Bessel's inequality requires

$$0 \leqslant \sum_j \frac{(Q\mathscr{D}^j)_E^2}{(\mathscr{D}^j)^2} \leqslant Q_E^2 = \sum_l (q_l)_E^2\, \alpha_l \leqslant \sum_l (q_l)_E^2 \tag{4.21}$$

in $\mathbf{R}_\alpha$. Using Eq. (4.18) to invert the orthogonal transformation then results in

$$\operatorname{Re} F \leqslant H = -\epsilon \lambda M^2 - \frac{\epsilon}{4\lambda(1+\epsilon^2)} \sum_{i,j=1}^{m} (\chi_i A_{ij}^{-1} \chi_j)_E. \tag{4.22}$$

Therefore the integrand of Eq. (4.19) is uniformly bounded by

$$[\lambda^{N-2m-1}/(\det A)^2]\, e^H. \tag{4.23}$$

We break the λ integration into two parts:

$$\begin{aligned} \Lambda &= \int_{\lambda=0}^{\infty} d\lambda\, \lambda^{N-2m-1}\, e^H = \Lambda_1 + \Lambda_2, \\ \Lambda_1 &= \int_0^a d\lambda\, \lambda^{N-2m-1}\, e^H, \\ \Lambda_2 &= \int_a^{\infty} d\lambda\, \lambda^{N-2m-1}\, e^H. \end{aligned} \tag{4.24}$$

Choosing a sufficiently small, so that

$$0 < a \leqslant \epsilon/4(1+\epsilon^2) < \epsilon, \tag{4.25}$$

results in the integrand in Λ_1 being uniformly bounded by the integrand in

$$\begin{aligned} \Lambda_1 &\leqslant \int_{=0}^{a} d\lambda\, \lambda^{N-2m-1} \\ &\quad \times \exp\left[\left(\frac{\lambda - a}{\lambda} - 1\right) \sum_{i,j=1}^{m} (\chi_i A_{ij}^{-1} \chi_j)_E\right] \\ &< \epsilon^{N-2m} \exp\left[-\sum_{i,j=1}^{m} (\chi_i A_{ij}^{-1} \chi_j)_E\right]. \end{aligned} \tag{4.26}$$

Using

$$\begin{aligned} b &= \epsilon m^2, \\ c^{-1} &= \epsilon[4(1+\epsilon^2)]^{-1} \sum_{i,j=1}^{m} (\chi_i A_{ij}^{-1} \chi_j)_E, \quad \text{and} \\ n &= N - 2m - 1, \end{aligned} \tag{4.27}$$

as well as

$$d\lambda = \lambda^2 c\, d(e^{-(\lambda c)^{-1}}), \tag{4.28}$$

we integrate Λ^2 by parts twice. This results in

$$\begin{aligned} \Lambda_2 &= \exp[-ba - 1/(ca)] \\ &\quad \times [-ca^{n+2} + c^2(n+2)a^{n+3} - bc^2 a^{n+4}] \\ &\quad + c^2 \int_a^{\infty} d\lambda\, [(n+2)(n+3)\, \lambda^{n+2} \\ &\quad - (n+2)b\, \lambda^{n+3} - b(n+4)\, \lambda^{n+3} + b^2 \lambda^{n+4}] \\ &\quad \times \exp[-\lambda b - (c\lambda)^{-1}]. \end{aligned} \tag{4.29}$$

The analysis of this integral proceeds from two formulas obtained from circuit-based graph theory.[11] These formulas are generalizations of those obtained for the unsubtracted amplitude by Nakanishi.[13] A review of the necessary terminology and essential results may be found in Appendix I. These results are cast into a useful form by the next lemma. $\mathfrak{G}$ is the generalized, or circuit-based graph, for the psft.

Lemma 4.1: At any point of $\mathbf{R}_\alpha$ there exists a generalized circuit of $\mathbb{C}$ of $\mathfrak{G}$ and a line r of $\mathfrak{G}$ such that for some generalized chord set $\mathbf{T}^*$ in the class of all generalized chord sets $\mathscr{T}^*(\mathfrak{G})$,

(a) $$\prod_{l \in \mathbf{T}^*} \alpha_l = \sup_{\mathbf{T}_\dagger^* \in \mathscr{T}^*(\mathfrak{G})} \prod_{l \in \mathbf{T}_\dagger^*} \alpha_l$$

(b) $$\alpha_r = \inf_{\substack{l: \\ l \in \mathbf{T}^*}} \alpha_l,$$

(c) $$\mathbb{C} \cap \mathbf{T}^* = \{r\},$$

and

(d) $$\sum_{i,j=1}^{m} (\chi_i A_{ij}^{-1} \chi_j)_E \geqslant \left(\frac{1}{P}\right)\left(\frac{1}{\alpha_r}\right)\left(\sum_{i=1}^{m} (\chi_i d_i^{\mathbb{C}})_E\right)^2, \quad \text{with} \quad P > 0.$$

Proof: Properties (a) and (b) are obvious. Property (c) is true if one chooses for $\mathbb{C}$ that circuit which is generated by the line, r of $\mathbf{T}^*$ defined by (b). The last property then follows from the first three and the equations (Appendix I) for $\det A$ and A_{ij}^{-1}, using τ for the number of distinct chord sets in $\mathscr{T}^*(\mathfrak{G})$:

$$\begin{aligned} &\det A \\ &\leqslant \tau\left(\sup_{\mathbf{T}_\dagger^* \in \mathscr{T}^*(\mathfrak{G})} d_{\mathbf{T}_\dagger^*}^2\right) \prod_{l \in \mathbf{T}^*} \alpha_l = \tau D^2 \prod_{l \in \mathbf{T}^*} \alpha_l; \end{aligned} \tag{4.30}$$

therefore

$$\begin{aligned} &\sum_{i,j=1}^{m} (\chi_i A_{ij}^{-1} \chi_j)_E \\ &\quad \geqslant \frac{(\Sigma_{i=1}^m (\chi_i d_i^{\mathbb{C}})_E^2\, B_{\mathbb{C}}^2\, \Sigma_{\mathbf{T}_\dagger^* \in \mathscr{T}^*(\mathfrak{G}/\mathbb{C})} (\Pi_{l \in \mathbf{T}_\dagger^*} \alpha_l)}{\tau D^2 \prod_{l \in \mathbf{T}^*} \alpha_l} \\ &\quad \geqslant \frac{(1/P)(\Sigma_{i=1}^m \chi_i d_i^{\mathbb{C}})_E^2}{\prod_{l \in \mathbf{T}^*} \alpha_l} \left[\prod_{\substack{l: \\ l \in \mathbf{T}^* - \{r\}}} (\alpha_l)\right] \\ &\quad = \left(\frac{1}{P}\right)\left(\sum_{i=1}^{m} \chi_i d_i^{\mathbb{C}}\right)_E^2 \frac{1}{\alpha_r}, \end{aligned} \tag{4.31}$$

with

$$P = \tau D^2 / B_{\mathbb{C}}^2 > 0, \text{ and } D^2 = \sup_{\mathbf{T}_\dagger^* \in \mathscr{T}^*(\mathfrak{G})} d_{\mathbf{T}_\dagger^*}^2 > 0. \tag{4.32}$$

The positivity condition, Eq. (4.32), is a result from circuit-based graph theory. The next to the last line of Eq. (4.31) follows from the fact that[11]

$$\mathbf{T}^* - \{r\} \in \mathscr{T}^*(\mathfrak{G}/\mathbb{C}).$$

This completes the proof of Lemma 4.1.

We return to the proof of Theorem 4.1. Referring to Eq. (4.29), we repeat the integration by parts a total of $2m$ times. The resulting integrated terms contain factors of b and c which, by Eq. (4.27) and Lemma 4.1(d), are bounded in $\mathbf{R}_\alpha$ except on a set of measure zero. A remaining factor of $\exp[-1/(ca) - ba]$ is similarly bounded since it is certainly less than an arbitrary positive power of ca multiplied by an appropriate constant. Since the integration by parts introduces

$2m$ factors of c into the unintegrated terms, both types of terms are dominated in $\mathbf{R}_\alpha$ by

$$\frac{c^{2m}}{(\det A)^2} \leqslant \frac{P^{2m}}{[d^2 \Pi_{l\in\mathbb{T}^*}\alpha_l]^2} \frac{\alpha_r^{2m}}{[(\Sigma_{i=1}^m \chi_i d_i^{\mathbb{C}})_E^2]^{2m}} \leqslant P^{2m}/d^4 \chi_{\mathbb{C}}^{4m}, \tag{4.33}$$

with $0 < d = \inf_{\mathbb{T}^\dagger \in \mathcal{T}(\mathfrak{G})} d_{\mathbb{T}^\dagger}$.

Here we have made use of Lemma 4.1 and Appendix I; also see Eq. (4.34) (below).

Therefore the integrals converge absolutely and uniformly in $\mathbf{R}_\alpha$ providing only that the quantities

$$\chi_{\mathbb{C}} = \sum_{i=1}^m \chi_i d_i^{\mathbb{C}} \neq 0 \tag{4.34}$$

for every generalized circuit $\mathbb{C}$ of $\mathfrak{G}$. Recalling Eq. (4.9), this condition is seen to be a finite system of inequalities which are linear in the breaking parameters. That is, we need only choose breaking parameters which do not lie in any of a finite number of hyperplanes. Since this may always be done, we have finished the proof of Theorem 4.1 for the case that the $n_{i\mu} \neq 0$ (no factors of internal momenta in the numerator of the psft).

The case that the $n_{i\mu}$ of Eq. (4.7) are nonvanishing is treated similarly. Eq. (4.16) is modified by the introduction of factors of $(k_{i\mu}^U)^{n_{i\mu}}$. The (k^U) integrations are then dominated by the exponential, except on a set of values of λ and (α) of measure zero, whose exponent is given by Eq. (4.17). Therefore, except on the set of measure zero, the (k^U) integrations are ac uniformly with respect to ξ. For this reason, the factors of $(k_{i\mu}^U)$ may be replaced by derivatives with respect to $\chi_{i\mu}$, and these derivatives may be taken after performance of the (k^U) integrations. The result is Eq. (4.19) modified by factors of

$$\frac{\partial^n F}{\partial \chi_{i\mu}^n} = \frac{\partial^n}{\partial \chi_{i\mu}^n}\left[\frac{i}{4\lambda(1+\epsilon^2)}\right] \times \sum_{i,j}\left(-\chi_i A_{ij}^{-1}\chi_{jE} + i\epsilon\chi_i A_{ij}^{-1}\chi_{jE}\right)\Bigg], \tag{4.35}$$

with $n = 1,2$.

This introduces factors of $1/\lambda$ which do not affect the convergence of the λ integral since Eq. (4.26) can be replaced by

$$\Lambda_1 \leqslant \int_0^a d\lambda\, \lambda^{-n} \exp\left[\left(\frac{\lambda - a}{\lambda} - 1\right)\sum_{i,j}\chi_i A_{ij}^{-1}\chi_{jE}\right],$$

which is bounded by the dominance of the exponential near $\lambda = 0$.

The factors of A_{ij}^{-1} do not affect the convergence of the (α) integral since, according to Appendix I, these factors are (bounded) polynomials in (α) divided by $\det A$. Repetition of the integration by parts procedure of Eqs. (4.28) and (4.29) a total of m times for each factor of $1/\det A$ and repeated use of Eq. (4.33) produces the conclusion that the integrals remain convergent when the $n_{i\mu}$ are arbitrary. This concludes the proof of Theorem 4.1.

Corrollary to Theorem 4.1: The circuit breaker regularization remains valid for particles with spin.

Proof: The introduction of spin results only in altering the powers of internal and external momenta in the typical terms.

A subsequent paper will examine the distributional and regulator limits for circuit breaker regularization. It will be shown that the regulator and propagator ϵ limits may be freely interchanged after folding in a Schwartz test function.

ACKNOWLEDGMENTS

The author is indebted to John Lowenstein for suggesting that point splitting might be a regularization, to Wolfhart Zimmermann, Peter Breitenlohner, and Thomas Clark for encouraging comments, and to Eugene Speer for his valuable critical reading of the manuscript.

APPENDIX I: A BRIEF REVIEW OF CIRCUIT BASED GRAPH THEORY

*Definition I.*1: *A circuit based graph*, elsewhere called a *generalized graph*, or *g-graph*, $\mathfrak{G}$, is a set of lines to each of which is assigned an m-dimensional vector called the internal momentum of that line.

The internal momentum routing coefficients or forest routing coefficients serve as a representation of the set of internal momenta.

*Definition I.*2: A set of lines of a g-graph is said to be independent if the internal momenta of these lines constitute a linearly independent set.

*Definition I.*3. A *generalized chord set* or *g-chord*, $\mathbb{T}^*$, is a maximal independent set of lines of a g-graph.

*Definition I.*4: A *generalized tree*, or *g-tree*, $\mathbb{T}$, is the complement in the circuit based graph of the g-chord.

By using the internal momenta of any particular g-chord, $\mathbb{T}^* = \{l_1,\ldots,l_m\}$, as a basis, we may write the internal momentum of a line, of the circuit based graph as

$$p_l = \sum_{i=1}^m d_l^{\mathbb{C}_i} p_{l_i}. \tag{I.1}$$

*Definition I.*5: A *generalized circuit* or *g-circuit*, $\mathbb{C}_i$, is the set of lines for which the routing coefficients $d_l^{\mathbb{C}_i} \neq 0$.

The assignment of internal momenta thereby defines the circuits of the g-graph even though we have no idea of how to connect the lines into a graph (cf., Appendix II). The theory is in this way based upon circuit coefficients, in effect, rather than upon the more usual incidence matrix. This is the source of the terminology, "circuit based graphs."

*Definition I.*6: A class of g-circuits, $\{\mathbb{C}_1,\ldots,\mathbb{C}_m\}$, is called a *fundamental set* or *f-set* of g-circuits iff it is a maximal class of g-circuits for which

$$(\forall k{:}1 \leqslant k \leqslant m)\ \mathbb{C}_k \not\subseteq \bigcup_{\substack{j:\\ j\neq k}} \mathbb{C}_j.$$

We remark that in this definition no assumptions are made as to which chord sets give rise to the indicated circuits.

As a result of these definitions and particularly of the linear independence of the p_{l_i} with $\mathbb{T}^* = \{l_1,...,l_m\}$, it follows that

$$d_j^{\mathbb{C}_i} = \delta_j^i,$$

and (I.2)

$$\mathbb{T}^* \cap \mathbb{C}_i = \{l_i\}.$$

Therefore the g-circuits determined by Eq. (I.1) and Definition I.5 constitute an f-set.

It also follows that a g-tree is a maximal set of lines containing n circuits, and therefore that every g-circuit intersects every g-chord in at least one line.

An additional graph-like property of g-graphs is that every f-set of g-circuits, determined from an arbitrary assortment of g-chords, in turn determines at least one g-chord such that the second of Eqs. (I.2) is true, and the given f-set can be thought of as generated [in the sense of Eq. (I.1) and Definition I.5] by this g-chord.

Definition I.7: For any $\mathbb{B}$ such that $\mathbb{B} \subseteq \mathbb{G}$, the reduced g-graph $\mathbb{G}/\mathbb{B}$, is the g graph consisting of the lines of $\mathbb{G}$-$\mathbb{B}$ together with the internal momentum assignments which these lines have in $\mathbb{G}$.

Identifying $d_l^{\mathbb{C}_i} = d_l^{U_i}$, one may obtain two useful formulas relating to the matrix A of Eq. (4.11):

$$\det A = \sum_{\substack{\mathbb{T}^*: \\ \mathbb{T}^* \in \mathscr{T}^*(\mathbb{G})}} \left(\prod_{l \in \mathbb{T}^*} \alpha_l\right) d_{\mathbb{T}^*}^2,$$

with

$d_{\mathbb{T}^*} = \det d^{\mathbb{C}}$, where $d^{\mathbb{C}}$ is the matrix whose components are $d_j^{\mathbb{C}_i}$, while $\mathscr{T}^*(\mathbb{G})$ is the class of all g-chords in $\mathbb{G}$:

$$A_{ij}^{-1} = \frac{1}{\det A} \times \sum_{\substack{\mathbb{C}: \\ \mathbb{C} \mathbb{G}}} d_i^{\mathbb{C}} d_j^{\mathbb{C}} B_{\mathbb{C}}^2 \sum_{\substack{\mathbb{T}^*: \\ \mathbb{T}^* \in \mathscr{T}^*(\mathbb{G}/\mathbb{C})}} d_{\mathbb{T}^*}^2 \prod_{l \in \mathbb{T}^*} \alpha_l, \quad \text{(I.3)}$$

where the lines have been labeled so that the g chord whose internal momenta are used as a basis to define A is $\{1,2,...,m\}$, and $B_{\mathbb{C}}^2$ is positive definite; $C(\mathbb{G})$ is the class of all g-circuits in $\mathbb{G}$.

For a fuller development, the reader is referred to.[11]

APPENDIX II: ON THE QUESTION OF THE GRAPHICAL REPRESENTATION OF CIRCUIT BASED GRAPHS

We first note that every graph can be considered a circuit based graph: one need only assign independent internal momenta to any chord set of the graph and use the vertex conservation conditions to find the remaining momentum assignments. The g-chords, g-trees, and g-circuits of the graph are then identical with corresponding chord set, trees, and circuits of the graph.

It has been shown that the internal momenta assignments corresponding to a forest term [a term associated with any one forest in the forest formula, Eq. (3.7)] can also be thought of as arising from a single graph.[9]

In general, however, a circuit based graph can not always be thought of as having a realization as a graph. This is made clear by the following counter-example:

$$\mathbb{G} = \{1,2,3,4,5,6,7,8\},\ \mathbb{T}^* = \{1,2,3,4\},$$
$$\mathbb{C}_1 = \{1,5,6\},\ \mathbb{C}_2 = \{2,5,7\},\ \mathbb{C}_3 = \{3,6,7,8\},$$
$$\mathbb{C}_4 = \{4,5,8\}.$$

If there is a graphical representation, then $\mathbb{C}_i$-$\mathbb{T}^* = P_i$ must be a path. This permits only two different ways of connecting the lines of P_1 and P_2 (cf. Fig. 1.) The only ways of attaching the line 8, because P_3 and P_4 are both paths, all involve introducing a circuit. This is not permitted since $\mathbb{G}$-$\mathbb{T}^* = \mathbb{T}$ is a g-tree and must be realized as a tree. Therefore, for any assignment of moment consistent with the given g-chord and g-circuits, $\mathbb{G}$ cannot be realized as a graph.

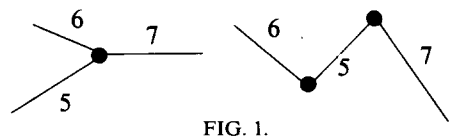


FIG. 1.

[1]W. Zimmermann, "Local operator products and renormalization in quantum field theory," in *Lectures on elementary particles and quantum field theory*, 1970 Brandeis lectures (MIT Press, Cambridge, MA, 1970), pp. 399–589.

[2]K. G. Wilson, "On products of quantum field operators at short distances," Cornell Report, Ithaca, New York (1964 (unpublished). K. G. Wilson, Phys. Rev. **179**, 1499 (1969).

[3]K. G. Wilson, and W. Zimmermann, Commun. Math. Phys. **24**, 87 (1972).

[4]T.E. Clark, Nucl. Phys. B **27**, 541 (1974).

[5]W. Zimmermann, Commun. Math. Phys. **15**, 208 (1969).

[6]N.N. Bogoliubov and D.V. Shirkov, *Introduction to the theory of quantized fileds* (Wiley, New York, 1959).

[7]K. Hepp, Commun Math. Phys. **2**, 301 (1966)

[8]V. Glaser, H. Lehmann, and W. Zimmermann, Nuovo Cimento **6**, 1122 (1957).

[9]R.R. Hampton, "The regularization of Feynman amplitudes by point splitting." Doctoral dissertation at New York University (1977) (unpublished).

[10]R.R. Hampton, (unpublished).

[11]R.R. Hampton, Commun. Math. Phys. **67**, 187–197 (1979).

[12]N. Nakanishi, *Graph theory and Feynman integrals* (Gordon and Breach, New York, 1971).

1/N expansion in two-dimensional lattice gauge theory

Yadin Y. Goldschmidt
Commissariat A L'Énergie atomique, Division de la Physique, Service de Physique Théorique, CEN-Saclay, Boîte Postale No 2, 91190 Gif-Sur-Yvette, France

(Received 23 January 1980; accepted for publication 22 February 1980)

We use the method of orthonormal polynomials to investigate the large N limit of the two-dimensional lattice gauge theory. The $1/N^2$ and $1/N^4$ corrections to the vacuum energy and the expectation value of the Wilson loop operator are calculated in the low temperature phase ($\lambda \equiv g^2 N < 2$). These corrections diverge at the transition point. In the high temperature phase we show that there are only corrections which fall exponentially as N becomes large, and hence there are no $1/N$ corrections in this phase.

1. INTRODUCTION

Recently much attention was devoted to the $1/N$ expansion as a tool for computation in gauge theories.[1-4] Here N is the order of the gauge group which is taken to be SU(N) or U(N). In a recent paper by D. Gross and E. Witten,[4] two-dimensional lattice gauge theory has been considered. Although the two-dimensional theory is somewhat trivial, because it can be reduced to a one plaquette problem, it has some interesting features from which one hopes to learn something about the behavior of the four-dimensional model. For example, it was found[4] that the two-dimensional lattice gauge theory has a third-order phase transition for $N = \infty$, as a function of $\lambda \equiv g^2 N$, where g^2 is the coupling constant of the model.

In order to carry out the computation of the vacuum energy and the expectation value of the Wilson loop the authors used the saddle point method, which was developed in a previous work by Brezin, Itzykson, Parisi, and Zuber[5] as a tool for calculating the $N = \infty$ limit of the ϕ^n theory. This method is not adequate for the computation of the $1/N$ corrections to the leading behavior. Recently another method wad developed by D. Bessis,[6] and simplified further by Parisi,[7] for the counting problem in ϕ^n theories. This method has the advantage that $1/N$ corrections can be calculated systematically. This method was extended and applied successfully to the two-matrix problem by M.L. Mehta.[8]

In this work we show how to extend the method of orthonormal polynomials due to Bessis to the case of integration over the set of unitary matrices. Using this method we reproduce the two phases previously obtained by the saddle point method. It is amusing that in contrast to the saddle point method the calculation of the low coupling regime turns out to be simpler than the strong coupling regime.

We proceed to calculate in the low temperature phase the next nonvanishing corrections which are of order $1/N^2$ and $1/N^4$. The coefficients of these corrections diverge as $\lambda \to 2$. In the high tempeature phase we are able to show that there are no $1/N$ corrections at all, but the corrections are exponentially small in N.

2. DEFINITION OF THE PROBLEM

We consider the lattice formulation of the two dimensional gauge theory.[9] The dynamical degrees of freedom are unitary matrices $U_{\mathbf{n},\mathbf{i}}$ associated with the links of the lattice, where $\mathbf{n}$ denotes the site and $\mathbf{i}$ the direction of the link. U is an $N \times N$ unitary matrix. The Wilson action is given by

$$S(U) = \sum_p \frac{1}{g^2} \mathrm{Tr}\left[\prod_p U + \mathrm{H.c.}\right], \tag{2.1}$$

where the sum is over all the plaquettes.

The vacuum-to-vacuum amplitude is

$$Z = \int (DU) \exp S(U) \tag{2.2}$$

and the vacuum energy density is

$$E_0 = -\frac{a^2}{V} \ln Z, \tag{2.3}$$

where V is the volume of the two-dimensional lattice, and a^2 is the volume of one plaquette. The expectation value of the Wilson loop operator is given by

$$W_L(g^2,N) = \frac{1}{N}\left\langle \mathrm{Tr} \prod_L U \right\rangle, \tag{2.4}$$

where L is a closed loop on the lattice.

It has been shown[3,4,10] that the functions defined above can be reduced to the one plaquette problem

$$Z = (z)^{V/a^2} \tag{2.5}$$

and

$$z(g^2,N) = \int dU \exp\left[\frac{1}{g^2} \mathrm{Tr}(U + U^+)\right]. \tag{2.6}$$

Similarly one can show that for a rectangular loop

$$W_L(g^2,N) = w^{R \cdot T}, \tag{2.7}$$

where

$$w(g^2,N) = \frac{1}{z} \int dU \frac{1}{N} \mathrm{Tr}U \exp\left[\frac{1}{g^2} \mathrm{Tr}(U + U^+)\right] \tag{2.8}$$

and $Ra \times Ta$ are the dimensions of the loop. It is easy to obtain w from z by noticing that

$$w = -\frac{\lambda^2}{2N^2} \frac{\partial \ln z}{\partial \lambda}. \tag{2.9}$$

Denoting the eigenvalues of U by $e^{i\theta_j}$, $j = 1,\ldots,N$ and denoting by T the unitary matrix that diagonalizes U, one

 0022-2488/80/071842-09$1.00

has

$$dU = \text{const} \times dT\left(\prod_{i=1}^{N} d\theta_i\right) |\Delta e^{i\theta_1},...,e^{i\theta_n})|^2, \tag{2.10}$$

where $\Delta(x_1,...,x_N)$ is the Vandermonde determinant

$$\Delta(x_1,...,x_N) = \prod_{i<j} (x_i - x_j). \tag{2.11}$$

Since the integrand in z does not depend on T, the T-integration can be performed trivially and one obtains

$$z(\lambda,N) = \text{const} \times \int_{-\pi}^{\pi} \prod_{i=1}^{N} \frac{d\theta_i}{2\pi} |\Delta(e^{i\theta_1},..,e^{i\theta_N})|^2 \times \exp\frac{2N}{\lambda} \sum_{j=1}^{N} \cos\theta_j. \tag{2.12}$$

3. THE METHOD OF ORTHONORMAL POLYNOMIALS

First let us note that the following identity holds:

$$\Delta(e^{i\theta_1},...,e^{i\theta_N}) = \det[e^{i(j-1)\theta_k}] = \det[P_{j-1}(e^{i\theta_k})]_{j,k=1,...,N}, \tag{3.1}$$

where

$$P_j(x) = x^j + \sum_{k=0}^{j-1} b_k x^k \tag{3.2}$$

are arbitrary polynomials with the coefficient of the highest power equal to one.

Let us now chose a system of polynomials which are orthonormal with respect to the measure

$$w(\theta) = e^{(2N/\lambda)\cos\theta}, \tag{3.3}$$

i.e.,

$$\int_{-\pi}^{\pi} \frac{d\theta}{2\pi} w(\theta) P_j(e^{i\theta}) P_k{}^*(e^{i\theta}) = h_j \delta_{jk}. \tag{3.4}$$

Note that Eq. (3.4) also defines the quantities h_j.

Letting $\alpha \equiv 2N/\lambda$ and using the shorthand notation $I_n \equiv I_n(\alpha)$ for the modified Bessel functions with argument α, it is easily shown, using the formula

$$I_n = I_{-n} = \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} w(\theta) e^{in\theta}, \tag{3.5}$$

that the polynomials P_n defined by (3.4) are given by

$$P_n(z) = c_n^{-1} \det \begin{bmatrix} I_0 & I_1 & I_2 & \cdots & I_n \\ I_1 & I_0 & I_1 & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{n-1} & \cdots & & I_0 & I_1 \\ 1 & z & z^2 & \cdots & z^n \end{bmatrix}, \tag{3.6}$$

where

$$c_n = \det[I_{k-l}]_{k,l=1,...,n}.$$

This is true since $P_n(e^{i\theta})$ is clearly orthonormal to $e^{-im\theta}$ for any $m < n$ with $w(\theta)$ as the measure. One notices that all the coefficients of $P_n(x)$ are real and therefore

$$P_j{}^*(e^{i\theta}) = P_j(e^{-i\theta}). \tag{3.7}$$

We further notice that

$$h_n = c_{n+1}/c_n. \tag{3.8}$$

Using (3.1) in (2.12) we expand the Vandermonde determinant and use the orthonormality properties of the polynomials to obtain

$$z(\lambda,N) = \text{const} \times N! \prod_{k=0}^{N-1} h_k. \tag{3.9}$$

The constant is to be taken such that $z(\infty,N) = 1$.

But for $\alpha = 0$, all I_n with $n \neq 0$ vanish; therefore all h_n are equal to unity and the constant in (3.9) has to be simply $1/N!$. Hence

$$z(\lambda,N) = \prod_{k=0}^{N-1} h_k = c_N = \det[I_{k-l}(\alpha)]_{k,l=1,...,N}. \tag{3.10}$$

This result agrees with a formula previously obtained by Bars and Green[3] using a completely different method. We need to know the asymptotic behavior of the product of h_j. To this end we shall now derive recursion relations for some of the coefficients of the polynomials P_j. Let us define the coefficients R_j, S_j by

$$z P_j(z) = P_{j+1}(z) + R_j P_j(z) + S_j P_{j-1}(z) + \cdots. \tag{3.11}$$

R_j, S_j are, of course, real. Other relations we shall need are

$$z^2 P_j(z) = P_{j+2}(z) + (R_j + R_{j+1}) P_{j+1}(z) + (S_j + S_{j+1} + R_j^2) P_j + \cdots, \tag{3.12}$$

$$z P_j'(z) = j P_j(z) + \left(\sum_{k=0}^{j-1} R_k\right) P_{j-1}(z) + \cdots, \tag{3.13}$$

$$z^2 P_j'(z) = j P_{j+1}(z) + \cdots. \tag{3.14}$$

Let us now calculate the following integral

$$\int \frac{d\theta}{2\pi} \frac{dw(\theta)}{d\theta} P_{j-1}(e^{i\theta}) P_j(e^{-i\theta}) = \frac{iN}{\lambda} \int \frac{d\theta}{2\pi} w(\theta)(e^{i\theta} - e^{-i\theta}) P_{j-1}(e^{i\theta}) P_j(e^{-i\theta}) = \frac{iN}{\lambda}(h_j - S_j h_{j-1}), \tag{3.15}$$

where we have used Eq. (3.11). On the other hand, the left-hand side of (3.15) is equal, using integration by parts, to

$$-i \int \frac{d\theta}{2\pi} w(\theta) e^{i\theta} P_{j-1}' P_j^* + i \int \frac{d\theta}{2\pi} w(\theta) P_{j-1} e^{-i\theta} P_j^{*\prime} = 0 + i\left(\sum_{0}^{j-1} R_k\right) h_{j-1}, \tag{3.16}$$

where we have used Eq. (3.13). From (3.15) and (3.16) one obtains

$$\frac{N}{\lambda}(h_j - S_j h_{j-1}) = h_{j-1} \sum_{0}^{j-1} R_k. \tag{3.17}$$

Defining

$$f_j = h_j/h_{j-1}, \tag{3.18}$$

(3.17) can be cast into the form

$$\frac{N}{\lambda}(f_j - S_j) = \sum_{0}^{j-1} R_k. \tag{3.19}$$

Similarly, computing in two different ways the integrals

$$\int d\theta\, e^{i\theta} \frac{dw(\theta)}{d\theta} P_j(e^{i\theta}) P_j(e^{-i\theta}), \tag{3.20}$$

$$\int d\theta\, e^{i\theta} \frac{dw(\theta)}{d\theta} P_{j-1}(e^{i\theta}) P_j(e^{-i\theta}), \tag{3.21}$$

one obtains, using relations (3.11)–(3.14), two other relations

$$\frac{N}{\lambda}(1 - S_{j+1} - S_j - R_j^2) = jR_j + \sum_0^j R_k \tag{3.22}$$

and

$$\frac{N}{\lambda}(R_j + R_{j-1}) f_j = j(1 - f_j). \tag{3.23}$$

Starting from

$$R_0 = I_1(\alpha)/I_0(\alpha), \tag{3.24}$$

one can use Eqs. (3.19), (3.22), and (3.23) as recursion relations for calculating all the coefficients R_j, S_j, f_j.

4. 1/N EXPANSION FOR $z(\lambda,N)$

We shall now evaluate the $1/N$ expansion of the quantity [see Eq. (2.43)]

$$-E_0(\lambda,N) = \frac{1}{N^2} \ln z(\lambda,N) = \frac{1}{N^2} \ln \prod_{i=0}^{N-1} h_i. \tag{4.1}$$

This last expression can be written in the form

$$-E_0(\lambda) = \frac{1}{N} \ln h_0 + \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{i}{N}\right) \ln f_i. \tag{4.2}$$

We start by evaluating the leading order of (4.2). Defining

$$x = \frac{j}{N}, \quad \epsilon = \frac{1}{N}, \tag{4.3}$$

we replace the discrete quantities f_j, R_j, S_j, by

$$f_j \sim f(x), \quad R_j \sim R(x), \quad S_j \sim S(x),$$
$$R_{j+1} \sim R(x+\epsilon) \sim R(x) \text{ etc.},$$
$$\frac{1}{N} \sum_0^j R_k \sim \int_0^x R(y)\, dy. \tag{4.4}$$

Then Eqs. (3.19), (3.22), and (3.23) take the form

$$\frac{1}{\lambda}[f(x) - s(x)] = \int_0^x R(y)\, dy, \tag{4.5a}$$

$$\frac{1}{\lambda}[1 - 2S(x) - R^2(x)] = xR(x) + \int_0^x R(y)\, dy, \tag{4.5b}$$

$$\frac{2}{\lambda} R(x) f(x) = x[1 - f(x)]. \tag{4.5c}$$

Before solving these equations let us note the constraint

$$0 \leqslant f(x) \leqslant 1 \text{ for } 0 \leqslant x \leqslant 1. \tag{4.6}$$

The argument goes as follows: clearly f_j is positive for any j because, by the definition (3.18), it is a quotient of two positive numbers. Moreover,

$$h_j = \int \frac{d\theta}{2\pi} w(\theta) P_j(e^{i\theta}) e^{i\theta} e^{-i\theta} P_j^*(e^{i\theta})$$
$$= h_{j+1} + R_j^2 h_j + S_j^2 h_{j-1} + \cdots, \tag{4.7}$$

where the dots stand for possible nonnegative terms only. Hence

$$f_{j+1} = 1 - R_j^2 - \frac{S_j^2}{f_j} - \cdots \leqslant 1 \tag{4.8}$$

because all f_j are positive.

From (4.5) one obtains the following integral equation for $R(x)$

$$\frac{x}{x + 2R(x)/\lambda} + \tfrac{1}{2}R^2(x) - \tfrac{1}{2} + \tfrac{1}{2}\lambda x R(x) - \tfrac{1}{2}\lambda \int_0^x R(y)\, dy = 0. \tag{4.9}$$

By diffentiating we obtain a first order nonlinear differential equation for $R(x)$, which is easily transformed into a differential equation for $V \equiv x + 2R/\lambda$

$$(\tfrac{1}{4}\lambda^2 V^3 - x)V' + V - \tfrac{1}{4}\lambda^2 V^3 = 0. \tag{4.10}$$

This equation has the trivial solution $V = 2/\lambda$, which corresponds to

$$R(x) = 1 - \tfrac{1}{2}\lambda x, \tag{4.10a}$$
$$f(x) = \tfrac{1}{2}\lambda x, \tag{4.10b}$$
$$S(x) = -\tfrac{1}{2}\lambda x(1 - \tfrac{1}{2}\lambda x). \tag{4.10c}$$

(4.10) has also the solution $v = x$, which corresponds to $R \equiv 0$ or $f \equiv 1$, but this is not a solution of (4.9).

The solution (4.10) is a good solution in the whole region $0 \leqslant x \leqslant 1$ only for $\lambda \leqslant 2$. For $\lambda > 2$, the condition (4.6) is not satisfied for $x > 2/\lambda$. For $\lambda > 2$ one should take instead the non analytic solution

$$R(x) = \begin{cases} 1 - \tfrac{1}{2}\lambda x & 0 \leqslant x \leqslant 2/\lambda, \\ 0 & 2/\lambda \leqslant x \leqslant 1, \end{cases} \tag{4.11a}$$

$$f(x) = \begin{cases} \tfrac{1}{2}\lambda x & 0 \leqslant x \leqslant 2/\lambda, \\ 1 & 2/\lambda \leqslant x \leqslant 1, \end{cases} \tag{4.11b}$$

$$S(x) = \begin{cases} -\tfrac{1}{2}\lambda x(1 - \tfrac{1}{2}\lambda x) & 0 \leqslant x \leqslant 2/\lambda, \\ 0 & 2/\lambda \leqslant x \leqslant 1. \end{cases} \tag{4.11c}$$

It remains to evaluate (4.2) by replacing, to leading order, the sum by an integral. We have to leading order

$$\frac{1}{N} \ln h_0 = \frac{1}{N} \ln I_0(2N/\lambda) = 2/\lambda + O(1/N). \tag{4.12}$$

Using (4.10b) one obtains

$$\int_0^1 (1 - x) \ln f(x) = \int_0^1 (1 - x) \ln \tfrac{1}{2}\lambda x = \tfrac{1}{2} \ln \tfrac{1}{2}\lambda - \tfrac{3}{4} \tag{4.13}$$

and using (4.11b) one obtains

$$\int_0^1 (1 - x) \ln f(x) = \int_0^{2/\lambda} (1 - x) \ln \tfrac{1}{2}\lambda x = 1/\lambda^2 - 2/\lambda. \tag{4.14}$$

Combining (4.12) with (4.13) and (4.14) we arrive at

$$-E_0(\lambda) = \begin{cases} 2/\lambda + \tfrac{1}{2} \ln \tfrac{1}{2}\lambda - \tfrac{3}{4} & \lambda \leqslant 2, \\ 1/\lambda^2 & \lambda \geqslant 2, \end{cases} \tag{4.15}$$

which agrees with the result obtained in Ref. 4.

We shall proceed to calculate the $1/N$ corrections to Eq. (4.15). We shall need in the sequel the following formulas:

$$\frac{1}{N}\ln h_0 = 2/\lambda + \frac{1}{2N}\ln\frac{\lambda}{4\pi N} + \frac{\lambda}{16N^2} + \frac{\lambda^2}{64N^3} + \frac{25}{3072}\frac{\lambda^3}{N^4} + \frac{13}{2048}\frac{\lambda^4}{N^5} + O\left(\frac{1}{N^6}\right), \tag{4.16}$$

$$\frac{1}{N}\sum_{p=1}^{N}\left(1-\frac{p}{N}\right)\ln\tfrac{1}{2}\lambda\,\frac{p}{N} \sim \tfrac{1}{2}\ln\tfrac{1}{2}\lambda - \tfrac{3}{4} - \frac{1}{2N} \times\ln\frac{\lambda}{4\pi N} + \frac{1}{N^2}\left(\tfrac{1}{12} - A - \tfrac{1}{12}\ln N\right) + \tfrac{1}{4}\sum_{k=2}^{\infty}\frac{B_{2k}}{(k-1)k}\frac{1}{N^{2k}}, \tag{4.17}$$

$$\frac{1}{N}\sum_{p=1}^{2N/\lambda}(1-p/N)\ln\tfrac{1}{2}\lambda\,\frac{p}{N} \sim -\frac{2}{\lambda} + \frac{1}{\lambda^2} - \frac{1}{2N}\ln\tfrac{1}{4}\frac{\lambda}{\pi N} + \frac{1}{N^2} \times\left(\tfrac{1}{24}\lambda - A - \tfrac{1}{12}\ln\frac{2N}{\lambda}\right) + \frac{1}{\lambda}\sum_{k=2}^{\infty}\frac{B_{2k}}{k(2k-1)}(\lambda/2N)^{2k} + \frac{1}{\lambda^2}\sum_{k=2}^{\infty}\frac{B_{2k}}{(k-1)k(2k-1)}(\lambda/2N)^{2k}. \tag{4.17a}$$

Here $A = 0.2487544770$ is the logarithm of Glaisher's constant and B_{2k} are Bernoulli numbers. We did not use Euler's formula to obtain the asymptotic expansions (4.17) and (4.18) because of the singularity of $(1-x)\ln\frac{1}{2}\lambda x$ at $x=0$, but rather used the asymptotic expansions[11] $\Sigma_{p=1}^{N}\ln px$ and $\Sigma_{p=1}^{N} p\ln px$. We shall also need Euler's formula, which reads

$$\frac{1}{N}\sum_{p=0}^{N} g(p/N) = \int_0^1 g(x)\,dx + \frac{1}{2N}[g(0)+g(1)] + \frac{B_2}{2!}\frac{1}{N^2}[g'(1)-g'(0)] - \frac{B^4}{4!}\frac{1}{N^4}[g'''(1)-g'''(0)] + \cdots + (-)^p\frac{B_{2p-2}}{(2p-2)!}\frac{1}{N^{2p+1}} \times[g^{(2p-3)}(1)-g^{(2p-3)}(0)] + (-)^{p+1}\frac{B_{2p}}{(2p)!} \times\frac{1}{N^{2p+1}}[g^{(2p)}(x_1)+\cdots+g^{(2p)}(x_p)]. \tag{4.18}$$

where $(i-1)/N < x_i < i/N$. This formula is valid for $g(x)$ of class C^{2p} on $[0,1]$. Let us now introduce the following expansions for the functions f_j, R_j, S_j:

$$f_j = f(x) + \frac{1}{N}f_1(x) + \frac{1}{N^2}f_2(x) + \frac{1}{N^3}f_3(x) + \frac{1}{N^4}f_4(x) + \cdots, \tag{4.19}$$

with similar expressions for R_j, S_j. One also has

$$R_{j-1} = R + \frac{1}{N}(R_1 - R') + \frac{1}{N^2}(R_2 - R_1' + \tfrac{1}{2}R'') + \frac{1}{N^3}(R_3 - R_2' + \tfrac{1}{2}R_1'' - \tfrac{1}{6}R''') + \frac{1}{N^4}(R_4 - R_3' + \tfrac{1}{2}R_2'' - \tfrac{1}{6}R_1''' + \tfrac{1}{24}R^{IV}) + \cdots, \tag{4.20}$$

$$R_j^2 = R^2 + \frac{1}{N}2RR_1 + \frac{1}{N^2}(R_1^2 + 2RR_2) + \frac{1}{N^3}(2RR_3 + 2R_1R_2) + \frac{1}{N^4}(2RR_4 + 2R_1R_3 + R_2^2) + \cdots, \tag{4.21}$$

and

$$S_{j+1} = S + \frac{1}{N}(S_1 + S') + \frac{1}{N^2}(S_2 + S_1' + \tfrac{1}{2}S'') + \frac{1}{N^3}(S_3 + S_2' + \tfrac{1}{2}S_1'' + \tfrac{1}{6}S''') + \frac{1}{N^4}(S_4 + S_3' + \tfrac{1}{2}S_2'' + \tfrac{1}{6}S_1''' + \tfrac{1}{24}S^{IV}). \tag{4.22}$$

Using (4.18) one easily obtains an expansion for $1/N \times \Sigma_0^j R_k$. Using these last expressions in the recursion formulas (3.19), (3.22), and (3.23), we obtain a set of three coupled equations, of integral and algebraic form, for any order in $1/N$. These equations are described in Appendix A. Having solved the equations up to a given order in $1/N^k$ one can proceed to solve the equations to the next order. Here we shall give only the results; the details are given in Appendix A. For $\lambda < 2$ we arrive at

$$R = 1 - \frac{\lambda}{2}x - \frac{1}{N}\frac{\lambda}{4} - \frac{1}{N^2}\frac{\lambda^2}{32}\frac{1}{(1-\frac{1}{2}\lambda x)^2} - \frac{1}{N^3}\frac{\lambda^3}{64}\frac{1}{(1-\frac{1}{2}\lambda x)^3} - \frac{1}{N^4}\frac{\lambda^4}{4096}\frac{50+11\lambda x}{(1-\frac{1}{2}\lambda x)^5} + \cdots, \tag{4.23a}$$

$$S = -\frac{\lambda}{2}x(1-\tfrac{1}{2}\lambda x) + \frac{1}{N^2}\frac{\lambda^3}{64}\frac{x(3-\lambda x)}{(1-\frac{1}{2}\lambda x)^2} + \frac{1}{N^4}\frac{9\lambda^5}{8192}\frac{x(10-\lambda x)}{(1-\frac{1}{2}\lambda x)^5}, \tag{4.23b}$$

$$f = \frac{\lambda}{2}x + \frac{1}{N^2}\frac{\lambda^3}{64}\frac{x}{(1-\frac{1}{2}\lambda x)^2} + \frac{1}{N^4}\frac{9\lambda^5}{8192}\frac{x(6+\lambda x)}{(1-\frac{1}{2}\lambda x)^5} + \cdots. \tag{4.23c}$$

For $\lambda > 2$, R, S, and f coincide with (4.23) in the region $x < 2/\lambda$, for $x > 2/\lambda$, $R \equiv 0$, $S \equiv 0$, $f \equiv 1$. The singularity at $x = 2/\lambda$ becomes more and more server in higher orders in $1/N$.

Since we know that for finite N, f_j has no divergence and remains between 0 and 1, the expansion given for f in (4.23b) has to represent some function which is finite for $x = 2/\lambda$.

Let us also comment on the result for $\lambda > 2, x > 2/\lambda$. It is true that Eqs. (3.19), (3.22), and (3.23) have a solution $R_j = 0, S_j = 0, f_j = 1$ for some $j > j_0$ provided $\Sigma_0^j R_k = N/\lambda$. But this solution is obtained only asymptotically. Let us define $\bar{f}_j = 1 - f_j$ and let us write the recursion relations in the form

$$\frac{N}{\lambda}(1 - \bar{f}_j - S_j) = \frac{N}{\lambda} - \sum_j^\infty R_k, \tag{4.24a}$$

$$\frac{N}{\lambda}(1 - S_j - S_{j+1} - R_j{}^2) = \frac{N}{\lambda} - \sum_{j+1}^\infty R_k + jR_j, \tag{4.24b}$$

$$\frac{N}{\lambda}(R_j + R_{j-1})(1 - \bar{f}_j) = j\bar{f}_j. \tag{4.24c}$$

Making the approximations

$$S_{j+1} \ll S_j, \quad R_j \ll R_{j-1}, \quad R_j{}^2 \ll S_j, \quad \bar{f}_j \ll 1.$$

we arrive at

$$\frac{N}{\lambda}(\bar{f}_j + S_j) = R_j, \tag{4.25a}$$

$$\frac{N}{\lambda}S_j = -jR_j, \tag{4.25b}$$

$$\frac{N}{\lambda}R_{j-1} = j\bar{f}_j. \tag{4.25c}$$

These equations have the solution

$$R_j = (N/\lambda)^2 \frac{1}{j(j+1)} R_{j-1}, \tag{4.26a}$$

$$S_j = -\frac{j\lambda}{N}R_j, \tag{4.26b}$$

$$\bar{f}_j = \frac{\lambda}{N}(j+1)R_j. \tag{4.26c}$$

Equation (4.26a) can be satisfied approximately by the solution

$$R_j = c(\lambda j/eN)^{-2j} = c(\lambda x/e)^{-2Nx}, \tag{4.27}$$

where $x = j/N$. Here c may depend on λ but not on N. So we see that this is an exponentially small solution as $N \to \infty$ and has no $1/N$ expansion.

Formula (4.27) is a good approximation for $\lambda x \gg 2$ but has to be modified in the vicinity of $\lambda x \gtrsim 2$, as we show in Appendix B.

In order to check the results (4.23) we compared them with the known series for R_0, R_1, S_1, f_1:

$$R_0 = \frac{I_1(2N/\lambda)}{I_0(2N/\lambda)} = 1 - \frac{\lambda}{4N} - \frac{\lambda^2}{32N^2} - \frac{\lambda^3}{64N^3} - \frac{25\lambda^4}{2048N^4} + \cdots, \tag{4.28a}$$

$$R_1 = \frac{I_1}{I_0}\frac{I_2 I_0 - I_1^2}{I_1^2 - I_0^2} = 1 - \frac{3\lambda}{4N} - \frac{L^2}{32N^2} - \frac{3\lambda^3}{64N^3} - \frac{121}{2048}\frac{\lambda^4}{N^4} + \cdots, \tag{4.28b}$$

$$S_1 = \frac{I_2 I_0 - I_1^2}{I_0^2} = -\frac{\lambda}{2N} + \frac{\lambda^2}{4N^2} + \frac{3\lambda^3}{64N^3} + \frac{\lambda^4}{32N^4} + \frac{125}{4096}\frac{\lambda^5}{N^5} + \cdots, \tag{4.28c}$$

$$f_1 = 1 - \frac{I_1^2}{I_0^2} = \frac{\lambda}{2N} + \frac{\lambda^3}{64N^3} + \frac{\lambda^4}{64N^4} + \frac{75}{4096}\frac{\lambda^5}{N^5} + \cdots. \tag{4.28d}$$

Substituting $x = 0$ in R of (4.23a) and $x = 1/N$ in (4.23) one can compare the two expansion and see the agreement. The terms of order $1/N^5$ predict that $f_5 \equiv 0$, $S_5 \equiv 0$.

Using the recursion relations (3.19), (3.22), and (3.23) we have carried out computer calculations of the various functions f_j, R_j, S_j, for various values of N and λ. The results agree very well with expression (4.23) with j/N substituted for x, except for x in the vicinity of $2/\lambda$. For $\lambda > 2$ and $j \sim N$ the results agree well with (4.26) or (4.27).

We shall write Eq. (4.23c) in the form

$$f(x) = \tfrac{1}{2}\lambda x\left(1 + \frac{\lambda^2}{32N^2}\frac{1}{(1 - \frac{1}{2}\lambda x)^2} + \frac{9\lambda^4}{4096}\frac{6 + \lambda x}{(1 - \frac{1}{2}\lambda x)^5} + \cdots\right). \tag{4.29}$$

Taking the logarithm we obtain

$$\ln f(x) = \ln\tfrac{1}{2}\lambda x + \frac{\lambda^2}{32N^2}\frac{1}{(1 - \frac{1}{2}\lambda x)^2} + \frac{\lambda^4}{2048N^4}\frac{26 + 5\lambda x}{(1 - \frac{1}{2}\lambda x)^5} + \cdots. \tag{4.30}$$

Hence

$$\frac{1}{N}\sum_{p=1}^N (1 - p/N)\ln f(p/N) = \frac{1}{N}\sum_{p=1}^N (1 - p/N)\ln\tfrac{1}{2}\lambda p/N + \int_0^1 (1 - x)\left[\frac{\lambda^2}{32N^2}\frac{1}{(1 - \frac{1}{2}\lambda x)^2} + \frac{\lambda^4}{2048N^4}\frac{26 + 5\lambda x}{(1 - \frac{1}{2}\lambda x)^5}\right]dx + \frac{1}{2N}\left[\frac{\lambda^2}{32N^2}\frac{1 - x}{(1 - \frac{1}{2}\lambda x)^2}\right]_0^1 + \frac{1}{12N^2}\left\{\frac{\lambda^2}{32N^2}\left[\frac{\lambda(1 - x)}{(1 - \frac{1}{2}\lambda x)^3} - \frac{1}{(1 - \frac{1}{2}\lambda x)^2}\right]\right\}_0^1 + \cdots. \tag{4.31}$$

After evaluating the various terms (4.31) reduces to

$$\frac{1}{N}\sum_{p=1}^N \left(1 - \frac{p}{N}\right)\ln f\left(\frac{P}{N}\right) = \frac{1}{N}\sum_{p=1}^N \left(1 - \frac{p}{N}\right)\ln\tfrac{1}{2}\lambda\frac{p}{N} + \frac{1}{N^2}\left(-\frac{\lambda}{16} - \frac{1}{8}\ln\left(1 - \frac{\lambda}{2}\right)\right) - \frac{\lambda^2}{64N^3} + \frac{1}{N^4}\left(\frac{3\lambda^3}{1024}\frac{1}{(1 - \frac{1}{2}\lambda)^3} - \frac{25\lambda^3}{3072}\right) + \cdots. \tag{4.32}$$

Combining Eqs. (4.16), (4.17), and (4.32) we finally obtain

$$-E_0(\lambda)=2/\lambda+\tfrac{1}{2}\ln\tfrac{1}{2}\lambda-\tfrac{3}{4}+\frac{1}{N^2}(\tfrac{1}{12}-A-\tfrac{1}{12}\ln N-\tfrac{1}{8}\ln(1-\tfrac{1}{2}\lambda))+\frac{1}{N^4}\left(\frac{3\lambda^3}{1024}\frac{1}{(1-\tfrac{1}{2}\lambda)^3}+\frac{1}{240}\right)+\cdots, \tag{4.33}$$

which is valid for $\lambda<2$. We note that the coefficients of the $1/N$ expansion become singular as $\lambda\to 2$.

We shall also write down the expression for the expectation value of the Wilson loop [see Eqs. (2.7)–(2.9)]. For $\lambda<2$ one has

$$w=1-\frac{\lambda}{4}-\frac{1}{N^2}\frac{\lambda^2}{32}\frac{1}{1-\frac{1}{2}\lambda}-\frac{1}{N^4}\frac{9\lambda^4}{2048}\frac{1}{(1-\frac{1}{2}\lambda)^4}+\cdots. \tag{4.34}$$

For $\lambda>2$ one cannot repeat blindly the manipulations done so far since it becomes evident immediately that one already encounters a divergent result at order $1/N^2$ due to the fact that the integral appearing on the right-hand side of (4.33) diverges at the point $x=2/\lambda$. As we will show immediately, there are no $1/N$ corrections in this phase, and therefore the $1/N^2$ corrections coming from (4.33) will have to cancel a term $(1/N^2)\ln N$ coming from (4.18) and therefore diverge. This means that one should seek a way to avoid passing through the singularity at $x=2/\lambda$. And in fact such a way does exist. Let us note the following identities (see Eq. (3.10)

$$\begin{aligned}c_\infty(2N/\lambda)&=\prod_{k=0}^{\infty}h_k(2N/\lambda)\\&=\prod_{k=0}^{N-1}h_k(2N/\lambda)\cdot\prod_{k=N}^{\infty}h_k(2N/\lambda)\\&=c_N(2N/\lambda)\cdot\prod_{k=N}^{\infty}h_k(2N/\lambda).\end{aligned} \tag{4.35a}$$

Hence

$$\ln c_N(2N/\lambda)=\ln c_\infty(2N/\lambda)-\ln\prod_{k=N}^{\infty}h_k(2N/\lambda). \tag{4.35b}$$

From this it follows that

$$\frac{1}{N^2}\ln Z(\lambda,N)=\frac{1}{N^2}\ln c_\infty(2N/\lambda)+\frac{1}{N^2}\sum_{k=1}^{\infty}k\ln f_{N+k}(2N/\lambda). \tag{4.36}$$

We shall now use a famous theorem due to G. Szegö[13] which gives the limit of a Töplitz determinant of infinite size. c_N of (3.10) is such a Töplitz determinant and Szegö's theorem implies (see Appendix C) that

$$c_\infty(2N/\lambda)=\exp(N^2/\lambda^2). \tag{4.37}$$

Therefore

$$\begin{aligned}\frac{1}{N^2}\ln Z(\lambda,N)&=\frac{1}{\lambda^2}+\frac{1}{N^2}\sum_{k=1}^{\infty}k\ln f_{N+k}(2N/\lambda)\\&\approx\frac{1}{\lambda^2}-\frac{1}{N^2}\sum_{k=1}^{\infty}k\bar f_{N+k}(2N/\lambda).\end{aligned} \tag{4.38}$$

Note that we need here only $\bar f_j$ with $j>N$ which, for $\lambda>2$, decreases exponentially with N!. So we avoid passing through the singularity by going from $x=\infty$ to $x=1$ instead of from $x=0$ to $x=1$. Of course this form is not suitable for $\lambda<2$ because then the singularily is at $x=2/\lambda>1$ and therefore one will again find divergent corrections to the result $1/\lambda^2$, which is known to be wrong in this phase.

An estimate to the correction in (4.38) for $\lambda\gg 2$ is obtained using Eqs. (4.26c) and (4.27):

$$\frac{1}{N^2}\ln Z(\lambda,N)-\frac{1}{\lambda^2}\simeq-\frac{1}{N^2}\bar f_{N+1}=-\frac{1}{N^2}C(\lambda)(\lambda/e)^{-2N}. \tag{4.39}$$

These corrections start to diverge for $\lambda<e$ (instead of $\lambda=2$) but, as we mentioned before, (4.27) has to be modified for λ close to 2.

We would like to add that our computer calculations verified the exponential decay of the corrections in the high temperature phase and agree with our theoretical estimates.

5. CONCLUSIONS

Using the method of orthonormal polynomials we were able to calculate the first few terms in the $1/N$ expansion in the low temperature phase. Up to the order we have calculated we found that there are only corrections with even powers of $1/N$. In fact, using the remark made after Eqs. (4.28), it is possible to show that the correction to $O(1/N^5)$ also vanishes. The fact that these terms vanish is due to nontrivial cancellations. For the continuous theory it was proved by 't Hooft[1], by topological agruments, that there are always corrections with even powers in $1/N$. This result still holds in perturbation theory (small coupling expansion) on the lattice.[12] Here this result seems to hold beyond perturbation theory, in two dimensions at least. We have found also that the $1/N^{2k}$ corrections diverge when $\lambda\to 2$ from below. In the high temperature phase we demonstrated that there are no $1/N$ corrections to the $N=\infty$ result, but the corrections fall exponentially with N. This is a new feature of the interesting phase transition which occurs in the two-dimensional lattice gauge theory in the large N limit.

ACKNOWLEDGMENTS

The author would like to thank Dr. E. Brezin and Dr. C. Itzykson for useful discussions, and especially Dr. M. Peskin for helpful advice and for his remarks on the manuscript. It is also a pleasure to thank Madame N. Tichit and Dr. J. M. Drouffe for help with the numerical work.

APPENDIX A

The equations to $O(1/N)$ are given by

$$\frac{1}{\lambda}(f_1-S_1)=\tfrac{1}{2}(1-R)+\int_0^x R_1\,dy, \tag{A1a}$$

$$-\frac{1}{\lambda}(2S_1+S'+2RR_1)$$

$$= xR_1 + \tfrac{1}{2}(1+R) + \int_0^x R_1\,dy, \tag{A1b}$$

$$\frac{1}{\lambda}(2R_1 f + 2Rf_1 - R'f) = -xf_1. \tag{A1c}$$

For $\lambda < 2, 0 < x < 1$ or $\lambda > 2, 0 < x < 2/\lambda$ these equations take the form

$$\frac{1}{\lambda}(f_1 - S_1) = \tfrac{1}{4}\lambda x + \int_0^x R_1\,dy, \tag{A2a}$$

$$-\frac{1}{\lambda}[2S_1 - \tfrac{1}{2}\lambda(1-\lambda x) + 2(1-\tfrac{1}{2}\lambda x)R_1] = xR_1 + 1 - \tfrac{1}{4}\lambda x + \int_0^x R_1\,dy, \tag{A2b}$$

$$\frac{1}{\lambda}(\lambda x R_1 + 2(1-\tfrac{1}{2}\lambda x)f_1 + \tfrac{1}{4}\lambda^2 x) = -xf_1, \tag{A2c}$$

which can be reduced to a single equation for R_1:

$$(1-\tfrac{1}{2}\lambda x)R_1' - \lambda R_1 - \tfrac{1}{4}\lambda^2 = 0. \tag{A3}$$

The solutions are

$$R_1 = -\lambda/4, \tag{A4a}$$

$$S_1 = 0, \tag{A4b}$$

$$f_1 = 0. \tag{A4c}$$

For $\lambda > 2$ and $x > 2/\lambda$ it is required, in order to avoid the singularity at $x = 2/\lambda$, to start from Eqs. (4.24) instead of (3.19), (3.22), and (3.23). In this region we obtain

$$\frac{N}{\lambda}(-f_1 + S_1) = \int_x^\infty R_1\,dy, \tag{A5a}$$

$$\frac{N}{\lambda}2S_1 = \int_x^\infty R_1\,dy - xR_1, \tag{A5b}$$

$$\frac{N}{\lambda}2R_1 = -xf_1. \tag{A5c}$$

The solution with the right boundary conditions is

$$R_1 = 0, \quad S_1 = 0, \quad f_1 = 0.$$

In the following we shall write explicitly only the equations valid for $\lambda < 2$, or $\lambda > 2$ and $x < 2/\lambda$. To order $1/N^2$ one obtains

$$\frac{1}{\lambda}(f_2 - S_2) = \int_0^x R_2\,dy + \tfrac{1}{2}(R_1(0) - R_1(x)) + \tfrac{1}{12}[R'(x) - R'(0)], \tag{A6a}$$

$$-\frac{1}{\lambda}(2S_2 + \tfrac{1}{2}S'' + R_1^2 + 2RR_2) = xR_2 + \int_0^x R_2\,dy + \tfrac{1}{2}[R_1(0) + R_1(x)] + \tfrac{1}{12}[R'(x) - R'(0)], \tag{A6b}$$

$$\frac{1}{\lambda}(2Rf_2 + 2R_2 f + \tfrac{1}{2}R''f - R_1'f) = -xf_2. \tag{A6c}$$

Using the previous known solutions these equations become

$$\frac{1}{\lambda}(f_2 - S_2) = \int_0^x R_2\,dy, \tag{A7a}$$

$$-\frac{1}{\lambda}\left[2S_2 + \frac{5\lambda^2}{16} + 2(1-\tfrac{1}{2}\lambda x)R_2\right] = xR_2 + \int_0^x R_2\,dy - \tfrac{1}{4}\lambda, \tag{(A7b)}$$

$$\frac{1}{\lambda}[2(1-\tfrac{1}{2}\lambda x)f_2 + \lambda x R_2)] = -xf_2. \tag{A7c}$$

These give

$$(2-\lambda x)R_2 - \lambda\int_0^x R_2\,dy + \lambda^2/16 = 0, \tag{A8}$$

from which we obtain

$$R_2 = -\frac{\lambda^2}{32}\frac{1}{(1-\frac{1}{2}\lambda x)^2}, \tag{A9a}$$

$$f_2 = \frac{\lambda^3 x}{64}\frac{1}{(1-\frac{1}{2}\lambda x)^2}, \tag{A9b}$$

$$S_2 = \frac{\lambda^3 x(3-\lambda x)}{(1-\frac{1}{2}\lambda x)^2}. \tag{A9c}$$

For the order $1/N^3$ one has.

$$\frac{1}{\lambda}(f_3 - S_3) = \int_0^x R_3\,dy + \tfrac{1}{2}(R_2(0) - R_2(x)) + \tfrac{1}{12}[R_1'(x) - R_1'(0)], \tag{A10a}$$

$$-\frac{1}{\lambda}(2S_3 + S_2' + \tfrac{1}{6}S''' + 2RR_3 + 2R_1R_2) = xR_3 + \int_0^x R_3\,dy + \tfrac{1}{2}(R_2(0) + R_2(x)) + \tfrac{1}{12}(R_1'(x) - R_1'(0)), \tag{A10b}$$

$$\frac{1}{\lambda}[f(2R_3 - R_2' + \tfrac{1}{2}R_1'' - \tfrac{1}{6}R''') + (2R_1 - R')f_2 + 2Rf_3)] = -xf_3, \tag{A10c}$$

which, using known solutions, reduce to

$$\frac{1}{\lambda}(f_3 - S_3) = \int_0^x R_3\,dy + \frac{\lambda^2}{64}\left[\frac{1}{(1-\frac{1}{2}\lambda x)^2} - 1\right], \tag{A11a}$$

$$-\frac{1}{\lambda}\left(2S_3 + \frac{\lambda^3}{32}\frac{1}{1-\frac{1}{2}\lambda x} + \frac{\lambda^3}{64}\frac{1+\lambda x}{(1-\frac{1}{2}\lambda x)^2} + \frac{\lambda^4}{64}\frac{x}{(1-\frac{1}{2}\lambda x)^3} + 2(1-\tfrac{1}{2}\lambda x)R_3 + \frac{\lambda^3}{64}\frac{1}{(1-\frac{1}{2}\lambda x)^2}\right) = xR_3 + \int_0^x R_3\,dy - \frac{\lambda^2}{64}\left(\frac{1}{(1-\frac{1}{2}\lambda x)^2} + 1\right), \tag{A11b}$$

$$\frac{1}{\lambda}\left(\tfrac{1}{2}\lambda x\left(2R_3 + \frac{\lambda^3}{32}\frac{1}{(1-\frac{1}{2}\lambda x)^3}\right) + 2(1-\tfrac{1}{2}\lambda x)f_3\right) = -xf_3. \tag{A11c}$$

These equations have the solution

$$R_3 = -\frac{\lambda^3}{64}\frac{1}{(1-\frac{1}{2}\lambda x)^3}, \tag{A12a}$$

$$f_3 = 0, \tag{A12b}$$

$$S_3 = 0. \tag{A12c}$$

To order $1/N^4$ the equations are:

$$\frac{1}{\lambda}(f_4 - S_4)$$
$$= \int_0^x R_4\,dy + \tfrac{1}{2}[R_3(0) - R_3(x)]$$
$$+ \tfrac{1}{12}[R_2'(x) - R_2'(0)] - \frac{B_4}{4!}[R'''(x) - R'''(0)], \tag{A13a}$$

$$-\frac{1}{\lambda}(2S_4 + \tfrac{1}{2}S_2'' + \tfrac{1}{24}S^{IV} + 2RR_4 + 2R_1R_3 + R_2^2)$$
$$= xR_4 + \int_0^x R_4\,dy + \tfrac{1}{2}[R_3(0) + R_3(x)]$$
$$+ \tfrac{1}{12}[R_2'(x) - R_2'(0)] - \frac{B_4}{4!}[R'''(x) - R'''(0)], \tag{A13b}$$

$$\frac{1}{\lambda}f(2R_4 - R_3' + \tfrac{1}{2}R_2'' - \tfrac{1}{6}R_1''' + \tfrac{1}{24}R^{IV})$$
$$+ f_2(2R_2 + \tfrac{1}{2}R'' - R_1') + 2Rf_4) = -xf_4. \tag{A13c}$$

These equations reduce to

$$\frac{1}{\lambda}(f_4 - S_4) = \int_0^x R_4\,dy + \frac{\lambda^3}{192}\frac{1}{(1-\frac{1}{2}\lambda x)^3} - \frac{\lambda^3}{192}, \tag{A14a}$$

$$-\frac{1}{\lambda}S_4 - \frac{1}{\lambda}(1-\tfrac{1}{2}\lambda x)R_4 - \frac{3\lambda^4}{512}\frac{x}{(1-\frac{1}{2}\lambda x)^4}$$
$$-\frac{\lambda^3}{2048}\frac{1}{(1-\frac{1}{2}\lambda x)^4} - \frac{\lambda^4}{256}\frac{x}{(1-\frac{1}{2}\lambda x)^3}$$
$$-\frac{3\lambda^3}{256}\frac{1}{(1-\frac{1}{2}\lambda x)^3} - \frac{\lambda^3}{128}\frac{1}{(1-\frac{1}{2}\lambda x)^2}$$
$$= \tfrac{1}{2}xR_4 + \tfrac{1}{2}\int_0^x R_4\,dy - \frac{\lambda^3}{192}\frac{1}{(1-\frac{1}{2}\lambda x)^3} - \frac{\lambda^3}{384}, \tag{A14b}$$

$$xR_4 + \frac{2}{\lambda}f_4 = \frac{\lambda^4 x}{1024}\frac{1}{(1-\frac{1}{2}\lambda x)^4}, \tag{A14c}$$

which have the solution

$$R_4 = -\frac{9\lambda^4}{512}\frac{1}{(1-\frac{1}{2}\lambda x)} + \frac{11\lambda^4}{2048}\frac{1}{(1-\frac{1}{2}\lambda x)^4}, \tag{A15a}$$

$$f_4 = \frac{9\lambda^5}{1024}\frac{x}{(1-\frac{1}{2}\lambda x)} - \frac{9\lambda^5}{4096}\frac{x}{(1-\frac{1}{2}\lambda x)^4}, \tag{A15b}$$

$$S_4 = \frac{9\lambda^4}{512}\frac{x}{(1-\frac{1}{2}\lambda x)^5} - \frac{27\lambda^4}{2048}\frac{1}{(1-\frac{1}{2}\lambda x)^4}$$
$$-\frac{9\lambda^4}{2048}\frac{1}{(1-\frac{1}{2}\lambda x)^3}. \tag{A15c}$$

APPENDIX B

One can substitute the following continuous approximation for the functions f_j, R_j, S_j, for $j > 2N/\lambda$.

$$f_j \to f(x) = \varphi(x)K(x)^{-Nx}, \tag{B1a}$$
$$S_j \to S(x) = s(x)K(x)^{-Nx}, \tag{B1b}$$
$$R_j \to R(x) = r(x)K(x)^{-Nx}, \tag{B1c}$$

For S_{j+1} and R_{j-1} one should take

$$S_{j+1} \to s(x)K(x)^{-Nx-1}\exp[-xK'(x)/K(x)], \tag{B2a}$$
$$R_{j-1} \to r(x)K(x)^{-Nx+1}\exp[xK'(x)/K(x)]. \tag{B2b}$$

Then the recursion relations (4.24) take the form

$$\varphi(x) + s(x) = 0, \tag{B3a}$$
$$-s(x)(1 + K^{-1}(x)\exp[xK'/K] = \lambda x r(x), \tag{B3b}$$
$$r(x)(1 + K\exp[xK'/K]) = \lambda x\varphi(x). \tag{B3c}$$

Denoting

$$L(x) = K(x)\exp[xK'(x)/K(x)], \tag{B4}$$

one finds

$$L^2 + (2-\beta^2)L + 1 = 0, \tag{B5}$$

where

$$\beta = \lambda x. \tag{B6}$$

This equation has the solution

$$L = \tfrac{1}{2}[\beta^2 - 2 + \beta\sqrt{\beta^2 - 4}]. \tag{B7}$$

Equation (B4) can be written as

$$\ln K(\beta) + \beta\frac{d\ln K(\beta)}{dx} = \ln L(\beta), \tag{B8}$$

which has the solution

$$\ln K(\beta) = \frac{1}{\beta}\Big[\int \ln L(\beta)\,d\beta + c\Big], \tag{B9}$$

where c is some constant.

For $\beta \gg 2$ one has $L(\beta) \sim \beta^2$ and therefore

$$K(\beta) \sim \frac{\beta^2}{e^2}. \tag{B10}$$

This coincides with (4.27), but of course the solution will be modified for $\beta \sim 2$. Assuming c is determined by the condition $K(2) = 1$, one can proceed to estimate K for $\beta \sim 2$.

APPENDIX C

Consider a Töplitz determinant

$$D_N = \det(d_{n-m}), \quad n,m = 1,\ldots N, \tag{C1}$$

where

$$d_n = \frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{-in\theta}\, d(e^{i\theta}) \tag{C2}$$

and

$$\ln d(e^{2ni}) - \ln d(e^{0i}) = 0. \tag{C3}$$

Szegö's Theorem states that

$$\lim_{N\to\infty}\frac{D_N}{\mu^N} = \exp\Big(\sum_{n=1}^{\infty} n g_{-n} g_n\Big), \tag{C4}$$

where

$$\mu = \exp\frac{1}{2\pi}\int_0^{2\pi} d\theta \ln d(e^{i\theta}) \tag{C5}$$

and

$$g_n = \frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{-in\theta}\ln d(e^{i\theta}), \tag{C6}$$

provided that as $|n| \to \infty$

$$C_n = O(K^{-|n|}), \tag{C7}$$

where $K > 1$.

In our case

$$d(e^{i\theta}) = \exp\frac{2N}{\lambda}\cos\theta \tag{C8}$$

and one finds immediately that

$$\mu = 1, \quad g_n = g_{-n} = \frac{N}{\lambda}\delta_{n,1}. \tag{C9}$$

Since for fixed $\alpha = 2N/\lambda$, and $n \to \infty$

$$I_n(\alpha) \sim \frac{1}{2\pi}\frac{1}{\sqrt{2n}}(2n/e\alpha)^{-n}, \tag{C10}$$

condition (C7) is satisfied and hence

$$D_\infty(2N/\lambda) = \exp(N^2/\lambda^2). \tag{C11}$$

[1]G. 't Hooft, Nuclear Physics B **72**, 461 (1974).
[2]E. Witten, in *Proceeedings of the* 1979 *Cargese Summer School*, "New Developments in Gauge Theories," and references therein.
[3]I. Bars and F. Green, I.A.S. Preprint, April 1979.
[4]D. J. Gross and E. Witten, Preprint Aug. 1979.
[5]E. Brezin, C. Itzykson, G. Parisi, and J. B. Zuber, Commun. Math. Phys. **59**, 35 (1978).
[6]D. Bessis, CEN-Saclay preprint DPHT 79/23, Feb. 1979.
[7]G. Parisi, unpublished. See also C. Itzykson and J. B. Zuber, CEN-Saclay Preprint DPHT 79/82, June 1979.
[8]M. L. Mehta, CEN-Saclay Preprint DPHT 124/79, Sept. 1979.
[9]K. Wilson, Phys. Rev. D **10**, 2445 (1974).
[10]R. Balian, J. M. Drouffe, and C. Itzykson, Phys. Rev. D **10**, 3376 (1974).
[11]E. R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, New Jersey, 1975).
[12]See B. E. Baaquie, Phys. Rev. D **16**, 2612 (1977). There will be infinitely many kinds of vertices but 't Hooft rules will still hold. I thank M. Peskin for this point.
[13]For a clear statement and proof of Szegö's Theorem see B. M. McCoy and T. T. Wu, in *The Two-Dimensional Ising Model* (Harvard U.P., Cambridge, Massachussetts, 1973) and references therein.

Narrow resonances as an eigenvalue problem and applications to high energy magnetic resonances: An exactly soluble model

A. O. Barut [a)]
International Centre for Theoretical Physics, Trieste, Italy

M. Berrondo [a)]
Institute for Advanced Studies, Hebrew University of Jerusalem, Jerusalem, Israel

G. García-Calderón
Instituto de Fisica, Universidad de Mexico, Mexico 20, D.F.

(Received 26 September 1979; accepted for publication 18 March 1980)

We formulate the problem of finding the narrow positive energy resonances in a deep potential well as an eigenvalue problem (thereby extending the scope of the discrete spectrum problem). We determine the number of resonances in an exactly soluble case. The method is then applied to a nonperturbative treatment of the magnetic resonances occurring in charge-dipole interactions, and the existence of the previously conjectured high mass narrow resonances in this model is proved.

I. INTRODUCTION

The study of quantum bound states as an eigenvalue problem is as old as the Schrödinger equation. Its mathematical theory is based on the Sturm–Liouville method[1] and Weyl's extension[2,3] thereof. In practice however, one finds that most states of composite systems are not truly stationary, but they can still be described as resonances. The energy spectrum is not discrete, but there is a spectral concentration around certain energies. In low-energy physics the photodecay of excited states is very well described in the framework of perturbation theory as "radiative effects" on the bound states. In problems involving radiationless decay, (e.g., autoionization in atoms, molecular predissociation, alpha decay in nuclei, to mention but a few) and especially in particle physics, the resonance behavior is an intrinsic effect and it must be accounted for from the beginning. In the case of sharp resonances, one can understand this decay starting from a potential function, though it might turn out to be energy dependent.

In the present paper we shall be particularly interested in a potential well like the one shown in Fig. 1, which can support positive-mass narrow resonances. Indeed, magnetic forces between spin-$\frac{1}{2}$ particles lead to effective radial potentials[4,5] of the type shown in Fig. 1, with one or more deep narrow wells, even in a relativistic treatment. In the single well model there appear at most a few resonances, but no negative energy bound states. A model for hadrons has recently been proposed[6] involving the magnetic interactions between stable particles (proton, electron, and neutrino). The narrow positive energy resonances in these models are missed if the magnetic dipole-dipole and charge-dipole potentials are treated perturbatively, as is done for the positronium, or even the charmonium. In our case, we find it necessary to solve the problem starting with the complete potential.

The purpose of the present work is to adopt analytic techniques to determine the spectrum of magnetic resonances for the relativistic and nonrelativistic cases. In fact, the theory amounts to a generalization of the usual eigenvalue problem to include also the case of resonances. The method is particularly useful if the resonances are narrow and isolated, i.e., nonoverlapping. The magnetic resonances are precisely of this type.

Physically one might question the use of the concept of a potential for relativistic problems at short distances. We must realize, however, that we are dealing with interactions of highly localized states. In such an event, as it is also true for bound states, the potential concept is proven to be superior to the use of perturbation theory.

In Sec. II we introduce the concept of a Gamow state, and its relevance to the problem of resonances, for the potential problem. Section III contains a general theory of the analytic solution for resonant states, and the particular case of a potential behaving like r^{-4} at the origin. In Sec. IV we use these results for the case of magnetic resonances with further corrections from the perturbation theory worked out in Sec. V for resonant states. We end with a brief discussion in Sec. VI.

II. GAMOW STATES

We start from the eigenvalue problem for λ^2

$$\left[\frac{d^2}{dr^2}+\lambda^2-V(\lambda,l,j;r)\right]\phi(r)=0, \quad r\geqslant 0, \tag{2.1}$$

where $V(r)$ is an effective potential which might depend on the linear momentum (or energy) λ, the orbital angular momentum l and the total angular momentum j, and presenting the general shape of Fig. 1. Let us assume that near the origin the potential is repulsive and behaves as

[a)]Permanent address: Physics Department, University of Colorado, Boulder, Colorado 80302.
[b)]On leave of absence from Instituto de Fisica, Universidad de Mexico, Mexico 20, D.F.

 0022-2488/80/071851-05$1.00

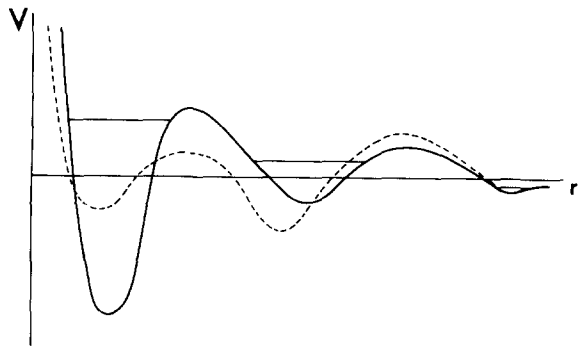


FIG. 1. Schematic radial potential for magnetic interactions Eq. (A9) at two different energies.

$$V(r) \to r^{-N}, \quad N > 0, \quad r \to 0^{+}. \tag{2.2}$$

Then $r = 0$ is a singular point of the differential equation (2.1). Asymptotically the potential vanishes as $r \to \infty$ (if it does not, we redefine λ^2 so that it does). Hence infinity is also a singular (irregular) point, due to the presence of the constant λ^2.

Physically, Eq. (2.1) will be taken to represent the radial equation for a charged particle in the field of a magnetic moment, or, alternatively, for a charged particle with a magnetic moment (both normal and anomalous) in the field of a fixed charge.[4,5] The square momentum λ^2 is taken as[7]

$$\lambda^2 = E^2 - m_r^2 \tag{2.3}$$

for relativistic problems. In the nonrelativistic limit

$$\lambda^2 = 2m_r E, \tag{2.4}$$

where E is the total energy and m_r is the reduced mass of the particle. The centrifugal potential $[l(l+1) - \alpha^2]/r^2$ and the Coulomb interaction α/r are both included in the potential $V(r)$. Its explicit expression is spelled out in the Appendix.

The Gamow states[8] are defined as the eigensolutions of Eq. (2.1), subject to boundary conditions

$$\phi(r = 0) = 0, \tag{2.5}$$

and purely outgoing wave behavior asymptotically

$$\phi_n'(r)/\phi_n(r) \to +i\lambda_n, \quad r \to \infty. \tag{2.6}$$

In general, these boundary conditions are not homogeneous[9] since the value of λ_n depends on the state. Furthermore, from flux conservation, it is easy to see that λ must be complex.[10] Hence, unless λ_n is purely imaginary ($\lambda_n^2 < 0$), the differential operator is not self-adjoint either. Bound states of the potential are included for the particular case $\lambda_n = +ik_n$ since, according to (2.6), the function $\phi(r)$ would vanish asymptotically in this case. As we should expect, for bound states we recover homogeneous, self-adjoint boundary conditions and the problem reduces to a (generalized) Sturm-Liouville problem.[1-3]

The interest in the more general Gamow states stems from the fact that the complex eigenvalues λ^2 correspond to the poles of the S matrix. Indeed, the physical radial wave functions behave asymptotically as

$$\phi(r) \to N[e^{-i\lambda r} - S(\lambda)e^{i\lambda r}], \quad r \to \infty.$$

From causality considerations, these poles can be shown.[10] to occur either on the lower half plane of λ or on the imaginary axis. The connection between isolated poles lying close to the real λ axis and narrow resonances is well established.[8,10-12]

III. ANALYTIC SOLUTIONS

We now turn our attention to the solutions of Eq. (2.1). Close enough to the origin, ϕ is determined from (2.2), so that

$$\phi(r) \to c_1 \exp(cr^{-s}) + c_2 \exp(-cr^{-s}), \quad r \to 0, \tag{3.1}$$

$$s = (N-2)/2, \quad \text{for } N > 2. \tag{3.2}$$

To have a regular solution at the origin, we must choose $c_1 = 0$, while asymptotically the Gamow states should behave like $\exp(i\lambda r)$, from Eqs. (2.5) and (2.6). To ensure the proper behavior, we write the solution as[13]

$$\phi(r) = F(r)\exp[g(r)], \tag{3.3}$$

and g is chosen appropriately. The new function F fulfills the equation

$$F'' + A(r)F' + B(r)F = 0, \tag{3.4}$$

with

$$A(r) = 2g'(r),$$
$$B(r) = g''(r) + g'^2(r) + \lambda^2 - V(r). \tag{3.5}$$

We now concentrate on the case in which $V(r)$ can be expanded as a power series in r^{-1}

$$V(r) = \sum_{k=1}^{N} \frac{v_k}{r^k}. \tag{3.6}$$

For a normal solution,[14] $g'(r)$ is a polynomial in r^{-1} (or in $r^{-1/2}$ for the case of odd N) so, recalling (3.2), we have

$$g(r) = \sum_{m=1}^{N-2} g_m r^{-m/2} + \nu \ln r + i\lambda r. \tag{3.7}$$

For even N, there are no fractional powers in the sum above. Furthermore, g_{N-2} is already fixed

$$g_{N-2} = -2\sqrt{v_N}/(N-2). \tag{3.8}$$

If we now expand the function F in a Taylor series

$$F(r) = \sum_{n=0} a_n r^n, \tag{3.9}$$

the coefficients $\{a_n\}$ fulfill a recursion relation determined by substituting (3.9) into (3.4), and a_0 can be taken as the normalization[9,14] constant.

For a more general class of potentials, the expression in (3.9) is a Laurent series,[15] i.e., it should also include negative powers. In any event, the recursion relations for the a's constitute a homogeneous set of linear algebraic equations. The condition for the existence of a nontrivial solution is that the associated (infinite) determinant[15] vanishes. For the case of Eq. (3.9), the determinant reduces to a semi-infinite one, in general. This is the quantization condition.

Let us now apply the general theory described above to the case of magnetic interactions. The dominant term at the origin for the effective potential,[4,5] Eq. (A9), yields $N = 4$, so that $s = 1$, from (3.2), and $c^2 = v_4$. We hence write $g(r)$, Eq. (3.7), as

$$g(r) = -\sqrt{v_4/r} + i\lambda r + \nu \ln r \tag{3.10}$$

and ν is chosen so that $B(r)$ in (3.5) has no term in r^{-3} i.e.,

$$2\nu c - 2c = v_3 . \tag{3.11}$$

The explicit expressions for (3.5) are

$$A(r) = 2c/r^2 + 2\nu/r + 2i\lambda ,$$

$$B(r) = \frac{1}{r^2}(\nu^2 - \nu + 2ic\lambda - v_2) + \frac{1}{r}(2i\nu\lambda - v_1) . \tag{3.12}$$

The recursion relation for the coefficients appearing in the Taylor expansion (3.9) is

$$2c(n+1)a_{n+1} + [D + n(n+2\nu-1)]a_n + [2i\lambda(n+\nu-1) - v_1]a_{n-1} = 0 , \tag{3.13}$$

where we have defined D as

$$D = \nu^2 - \nu + 2ic\lambda - v_2 , \tag{3.14}$$

and depends on v_2 and v_3, from the condition (3.11).

We shall next argue that the power series for $F(r)$ in (3.9) cannot involve very large terms. Hence a *sufficient* condition to obtain a convergent solution to Eq. (2.1) is that this series terminates. To this end, we look at the recursion relation for (3.13) for large values of n, in terms of the ratio between successive coefficients $\xi = a_{n+1}/a_n$,

$$2c\xi^2 + n\xi + 2i\lambda \simeq 0 , \tag{3.15}$$

with solutions

$$\xi_1 = -n/2c + o(1/n) ,$$

$$\xi_2 = -2i\lambda/n + o(1/n^2) . \tag{3.16}$$

For the first root, ξ_1, the series in (3.9) is an asymptotic series. The second root implies the wrong asymptotic behavior of the wavefunction $\phi(r) \to \exp(-i\lambda r)$. If the series in (3.9) terminates, i.e., $F(r)$ is a polynomial in r, we obtain a solution with the proper boundary conditions. This can be achieved by imposing two conditions in the three-term recursion relation (3.13).

Since we are looking for the magnetic resonances,[4,5] the Coulomb term is very weak in the region of interest and can be neglected, so we take $v_1 = 0$ in (3.13). In this case, the first of the two conditions is that $\nu = -M$, a negative integer, so that Eq. (3.13) will involve only two terms for the case $n = M + 1$

$$2c(M+2)a_{M+2} + [D + (M+1)(M+2)]a_{M+1} = 0 . \tag{3.17}$$

From Eq. (3.11) we see that this is a condition imposed on the form of the potential, namely that the ratio

$$v_3/\sqrt{v_4} = -2M - 2 \tag{3.18}$$

is a negative even integer.

In order that the series terminate, we now choose $a_{M+1} = 0$, so, from Eqs. (3.13) and (3.17), $a_{M+2} = a_{M+3} = \cdots = 0$, and $F(r)$ is a polynomial of degree M. The quantization condition for the complex momentum λ is the vanishing of the tridiagonal determinant of order $M+1$

$$\begin{vmatrix} D & 2c & 0 & \cdots & 0 \\ -2i\lambda M & D-2M & 4c & 0 & \\ 0 & -2i\lambda(M-1) & D+2(1-2M) & 6c & \\ \vdots & 0 & & & \vdots \\ 0 & 0 & \vdots & \vdots & -2i\lambda D + M(-M-1) \end{vmatrix} = 0 . \tag{3.19}$$

From the definition of D, Eq. (3.14), we see that it is linear in λ. Hence Eq. (3.19) is a secular equation for $2i\lambda$ which yields $(M+1)$ roots.

Let us summarize briefly. For a potential of the form

$$V(r) = v_4/r^4 - [2(M+1)\sqrt{v_4}/r^3] + v_2/r^2 , \tag{3.20}$$

the quantization condition for the eigenvalue λ is given by the secular equation (3.19), with

$$D = M^2 + M + 2i\sqrt{v_4}\lambda - v_2 , \tag{3.21}$$

where M is a quantum number, and v_2 and v_4 the parameters of the potential.

IV. THE MAGNETIC RESONANCES

In view of the application of the method developed in Sec. III to the magnetic resonances,[4,5] we shall first find the solutions explicitly for a few low-lying quantum numbers, namely, $M = 0$ and 1. For $M = 0$, the quantization condition is $D = 0$, so, using Eq. (3.21), we find

$$\lambda = -iv_2/2\sqrt{v_4} . \tag{4.1}$$

There is one bound state if $v_2 < 0$, or one antibound state if v_2 is positive. If $M = 1$, we have a determinant of order two, with solutions

$$\lambda = \frac{1}{2\sqrt{v_4}}\{(2 - v_2)i \pm [2(v_2 - 2)]^{1/2}\} . \tag{4.2}$$

In case $v_2 > 2$, there are two complex poles, λ and $-\lambda^*$, corresponding to an isolated resonance.[8-10] Otherwise λ is purely imaginary; at $v_2 = 2$ the two poles coincide and we have a "zero-energy" bound state or resonance.[16] Similarly, the complex poles for $M = 2,3,\cdots$ can be explicitly evaluated.

It is remarkable that magnetic spin–spin and spin-orbit interactions between two spin $\frac{1}{2}$-particles lead to a potential of the form (3.20). In the Appendix we have discussed a few models of charge-dipole interactions. In the case of the Dirac equation (including the anomalous magnetic moment) the effective potential is actually energy-dependent and quite complicated. But at very high energies, the potential has the characteristic $1/r^4$, $1/r^3$ and $1/r^2$ terms. Also, a scalar charged particle moving in a dipole field, or a massless particle with an anomalous magnetic moment moving in the Cou-

lomb field lead to the same effective potential of the form (3.20). To these we must add the spin–spin terms which will modify the coefficient of the $1/r^3$ terms.

The importance of the exactly soluble model reported here lies in the fact that we can now develop a perturbation theory around this new nonperturbative solution. Previously, magnetic interactions have been treated as perturbations around the Coulomb solution. Inspecting the potential wells shown in Fig. 1, we see that a perturbation theory around the Coulomb well can never reveal the existence of new states in the deep second (or third) well at shorter distances. This is truly a nonperturbative phenomenon, and a remarkable physical example of a quantum system with two distinct ground states.

The perturbation theory and numerical applications to magnetic resonances will be reported elsewhere.

V. CONCLUSION

We have extended the analytic methods used in the solution of bound state problems to the case of general Gamow states. The solutions are expanded in power series of the variable and the quantization condition appears as an (infinite in principle) secular equation which follows from the recursion relation for the expansion coefficients. The resulting energies are complex in general, the bound states being a particular case. For the particular case of potentials of the form (3.20), the corresponding determinant is finite and we obtain explicitly solutions for the poles of the S matrix.

APPENDIX

In this Appendix we reduce the Dirac equation for the case of a point charged dipole interacting with a fixed charge and including the anomalous magnetic moment of the dipole. This shall be taken as a model for the interaction of two electrons (of either charge), considering one of them as fixed. We shall follow closely Refs. 4. In natural units ($\hbar = c = 1$), and ϵ being the relative sign of the charges, the Dirac equation including the anomalous moment a is

$$\left\{\boldsymbol{\alpha}\cdot\mathbf{p} - \left(E - \frac{\epsilon e^2}{r}\right) + \beta m_r\right\}\Psi = -a\frac{\epsilon e^2}{2m}\frac{1}{r^2} i\beta\alpha_r\Psi, \quad \text{(A1)}$$

where α_r is the radial component of the Dirac $\boldsymbol{\alpha}$ matrix, m is the mass of the particle and m_r its reduced mass. Using the angular constants of the motion J^2, J_z, and

$$K = \beta(\boldsymbol{\sigma}\cdot\mathbf{L} + 1), \quad \text{(A2)}$$

we can separate the angular part and obtain two coupled equations[4] for the radial parts f and g

$$\begin{aligned} f'(r) &= \left(\frac{\kappa - 1}{r} + \frac{a\epsilon e^2}{2mr^2}\right) f(r) \\ &\quad + \left(m_r - E + \frac{\epsilon e^2}{r}\right) g(r), \\ g'(r) &= \left(-\frac{\kappa + 1}{r} - \frac{a\epsilon e^2}{2mr^2}\right) g(r) \\ &\quad + \left(m_r + E - \frac{\epsilon e^2}{r}\right) f(r), \end{aligned} \quad \text{(A3)}$$

where $\kappa = \pm(j + 1/2)$ is the eigenvalue of K. Introducing the radial component $u = rf(r)$ and eliminating g, we get a second-order equation

$$\begin{aligned} &\frac{d^2u}{dx^2} - \frac{A'}{A}\frac{du}{dx} - \left\{\left[\frac{\kappa(\kappa+1)}{x^2} + \frac{\epsilon(\kappa+1)}{x^3} + \frac{1}{4x^4}\right]\right. \\ &\quad \left. + \frac{A'}{A}\left(\frac{\kappa}{x} + \frac{\epsilon}{2x^2}\right) + AB\right\} u = 0, \end{aligned} \quad \text{(A4)}$$

in terms of the reduced variable

$$x = (m/\alpha a)r, \quad \text{(A5)}$$

where $a \cong \alpha/2\pi$ for the electron, and the functions A and B

$$\begin{aligned} A(x) &= a(\alpha/m)(m_r + E) - \epsilon\alpha/x, \\ B(x) &= a(\alpha/m)(m_r - E) + \epsilon\alpha/x, \end{aligned} \quad \text{(A6)}$$

depend on the energy.

To eliminate the first-order term in (A4) and normalize the x^{-4} term to one, we further introduce

$$\begin{aligned} &\Psi(y) = u/\sqrt{A}, \\ &y = 2x, \\ &\Lambda^2 = a^2(\alpha^2/4m^2)(E^2 - m_r^2). \end{aligned} \quad \text{(A7)}$$

The resulting equation has the form of a radial eigenvalue equation with an effective potential

$$\left[\frac{d^2}{dy^2} + \Lambda^2 - V(y)\right]\Psi(y) = 0, \quad \text{(A8)}$$

where

$$\begin{aligned} V(y) &= \frac{\kappa(\kappa+1)}{y^2} + \epsilon\frac{\alpha^3}{2\pi m}\frac{E}{y} + \frac{1}{y^2} \\ &\quad \times\left[\frac{\epsilon(\kappa+1)}{h(y)} + \frac{3}{4}\frac{1}{h^2(y)} - \alpha^2\right] \\ &\quad + \frac{1}{y^3}\left[2\epsilon(\kappa+1) + \frac{1}{h(y)}\right] + \frac{1}{y^4}, \end{aligned} \quad \text{(A9)}$$

where

$$h(y) = -\epsilon + \frac{\alpha}{4\pi}\frac{m_r + E}{m}y. \quad \text{(A10)}$$

The part of the potential which does not depend on the energy is given by

$$V_0(y) = \frac{\kappa(\kappa+1) - \alpha^2}{y^2} + \epsilon\frac{2(\kappa+1)}{y^3} + \frac{1}{y^4}. \quad \text{(A11)}$$

If the anomalous magnetic moment is neglected, the Dirac electron with its normal magnetic moment ($g = 2$) leads to Eq. (A8) with

$$\begin{aligned} &\Lambda^2 = (\alpha^2/m^2)(E^2 - m_r^2); \quad x \equiv (m/\alpha)r, \\ &V(x) = \frac{\kappa(\kappa+1) - \alpha^2}{x^2} + 2\epsilon\left(\frac{E}{m}\right)\frac{\alpha^2}{x} \\ &\quad + \epsilon\frac{\kappa+1}{x^2[(1 + E/m)x + 1]} \\ &\quad + \frac{3}{4}\frac{1}{x^2[(1 + E/m)x + 1]^2}. \end{aligned}$$

It is interesting that a scalar charged particle moving in the field of a fixed magnetic potential $\mathbf{A} = \mu(\boldsymbol{\sigma}\times\mathbf{r})/r$ also leads to the eigenvalue equation (A8) with

$$\Lambda^2 = (E^2 - m_r^2)e^2\mu^2, \quad y = r/e\mu\,,$$

$$V(y) = \frac{l(l+1) - \alpha^2}{y^2} + \epsilon\,\frac{2C(l,j)}{y^3} + \frac{1}{y^4}\,.$$

Here

$$C(l,j) = -l + 1, \quad \text{for } l = j + \tfrac{1}{2}$$
$$= l, \quad \text{for } l = j - \tfrac{1}{2}\,,$$

in the state with orbital angular momentum l and total angular momentum j.[17] Furthermore, the same equation is obtained for a particle with total magnetic moment μ in the field of a fixed charge in a different formalism.[18]

Finally, the pure anomalous magnetic moment limit of (A1) (i.e., dropping from the beginning the term $\epsilon e^2/r$ on the left-hand side) also leads immediately to this same potential (A11).

[1]E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second-Order Differential Equations*, (Oxford U. P., Oxford, England 1946).
[2]H. Weyl, Math. Ann. **68**, 220 (1910).
[3]H. V. McIntosh, in *Group Theory and its Applications to Quantum Mechanics*, edited by K. Loebl (Academic, New York, 1975), Vol. III, p. 333.
[4]A. O. Barut and J. Kraus, Phys. Letters B **59**, 175 (1975); J. Math. Phys. **17**, 506 (1976).
[5]A. O. Barut and J. Kraus, Phys. Rev. D **16**, 161 (1977).
[6]A. O. Barut, in *Lecture Notes in Physics* (Springer, Berlin, 1979), Vol. 94, pp. 490–8; Surv. High Energy Phys. **1** (2), 113 (1980).
[7]In natural units such that $\hbar = c = 1$, and $e^2 = \alpha$, the fine structure constant.
[8]G. Gamow, Z. Phys. **51**, 204 (1928); A. J. F. Siegert, Phys. Rev. **56**, 750 (1939); J. Humblet and L. Rosenfeld, Nucl. Phys. **26**, 529 (1961).
[9]M. Berrondo and G. García-Calderón, in *Quantum Science*, edited by J. L. Calais, O. Goscinski, J. Linderberg, and Y. Ohrn (Plenum, New York, 1976), p. 295.
[10]See, e.g., H. M. Nussenzveig, *Causality and Dispersion Relations*, (Academic, New York, 1974).
[11]See, e.g., J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[12]R. E. Peierls, Proc. R. Soc. London Ser. A **253**, 16 (1959).
[13]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chapter XVII.
[14]G. García-Calderón and R. E. Peierls, Nucl. Phys. A **265**, 443 (1976).
[15]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Vol. IV.
[16]A. O. Barut, J. Math. Phys. **21**, 568 (1980).
[17]A. O. Barut and G. Strobel (to be published).
[18]C. Piron and F. Reuse, Helv. Phy. Acta **51**, 146 (1978); F. Reuse, Helv. Phys. Acta **51**, 157 (1978).

Orbital angular momentum of "lumps" [a)]

S. Ahmed and H. Schiff
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta, Canada, T6G 2J1

(Received 18 September 1979; accepted for publication 22 February 1980)

We formulate the external orbital angular momentum of a "lump" so that $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ on any space-like hyperplane. The systems considered are scalar, spinor and vector fields with $e^{i\omega t}$ time dependence, each interacting with its own electromagnetic field. Particular attention is given to cases involving nonzero dipole moments.

1. INTRODUCTION

Recent interest in localized classical fields ("lumps") as possible models for extended particles has emphasized the question of their stability. In this paper we direct our attention to the description of their angular momentum.

If we imagine a classical particle at rest, then its external orbital angular momentum is zero and we may consider its angular momentum as being entirely intrinsic. If the particle is a composite, then the intrinsic angular momentum may consist of internal "orbital" and "spin" angular momenta. However, for such a bound state the division into "orbital" and "spin" is not in general meaningful.

If we consider now a field theory with localized solutions, to what extent can we define the external orbital angular momentum of a "lump" so that it conforms to what we intuitively expect for a classical particle? More specifically, given a classical relativistic field theory, can we ensure that the expression for the external orbital angular momentum is independent of the space-like surface on which it is evaluated?

Here we will examine this question within the framework of three complex fields—a scalar field, a spinor field, and a vector field, each interacting with its own electromagnetic field gauge invariantly. That is, each field with its electromagnetic field constitutes a self-contained free system (without singularities).

Assuming a harmonic ($e^{i\omega t}$) time dependence for the complex fields and restricting the space-like surface to hyperplanes, we show that it is possible to have a meaningful description of the external orbital angular momentum.

Given a Lagrangian $\mathscr{L}(\phi)$ (we choose $c = 1$, $x_4 = it$), the invariance of $\int \mathscr{L}\, d^4x$ with respect to variations in x and ϕ yields the conservation laws[1]

$$\partial_\mu G_\mu = 0, \tag{1.1}$$

where the Noether current G_μ can be written

$$G_\mu = \mathscr{L}\, \delta x_\mu + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_\kappa)} \delta\phi_\kappa + \partial_\tau (W_{\tau\mu\nu}\, \delta x_\nu). \tag{1.2}$$

Included in (1.2) is a divergence involving the quantity $W_{\tau\mu\nu}$, which is antisymmetric in $(\tau\mu)$. This will not contribute to (1.1) but one must ensure that no surface contributions appear at spatial infinity for the integrals of energy, momentum, and angular momentum. In addition to this divergence we will add appropriate divergence terms to the Lagrangian as needed.

Under a translation,

$$\delta x_\nu = \epsilon_\nu, \quad \delta\phi_\kappa = -\epsilon_\nu \partial_\nu \phi_\kappa,$$

the Noether current becomes

$$G_\mu = \epsilon_\nu T_{\mu\nu},$$

with the energy-momentum tensor

$$T_{\mu\nu} = T^c_{\mu\nu} + \partial_\tau W_{\tau\mu\nu}, \tag{1.3}$$

where

$$T^c_{\mu\nu} = \mathscr{L}\, \delta_{\mu\nu} - \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_\kappa)} \partial_\nu \phi_\kappa \tag{1.4}$$

is the usual canonical energy-momentum tensor. In general, the $T_{\mu\nu}$ defined by (1.3) is not symmetric.

Under a rotation

$$\delta x_\nu = \omega_{\nu\sigma} x_\sigma,$$

$$\delta\phi_\kappa = -\omega_{\nu\sigma} x_\sigma \partial_\nu \phi_\kappa + \tfrac{1}{2} i \omega_{\nu\sigma} (S_{\nu\sigma})_{\kappa\beta} \phi_\beta,$$

the Noether current is

$$G_\mu = \omega_{\nu\sigma} \left(x_\sigma T_{\mu\nu} + \frac{i}{2} \frac{\partial \mathscr{L}}{\partial(\partial_\mu \phi_\kappa)} (S_{\nu\sigma})_{\kappa\beta} \phi_\beta + W_{\sigma\mu\nu} \right). \tag{1.5}$$

Making use of the antisymmetry of $\omega_{\nu\sigma}$ and $S_{\nu\sigma}$, one obtains from (1.5) the total angular momentum density

$$J_{\mu\nu\lambda} = x_\mu T_{\lambda\nu} - x_\nu T_{\lambda\mu} - i \frac{\partial \mathscr{L}}{\partial(\partial_\lambda \phi_\kappa)} (S_{\mu\nu})_{\kappa\beta} \phi_\beta + (W_{\mu\lambda\nu} - W_{\nu\lambda\mu}). \tag{1.6}$$

The total three-angular-momentum is

$$J_k = \tfrac{1}{2} \epsilon_{klm} \int J_{lm\lambda}\, d\sigma_\lambda. \tag{1.7}$$

We divide J_k into "orbital" and "intrinsic" parts,

$$J_k = L_k + S_k, \tag{1.8}$$

where

$$L_k = \epsilon_{klm} \int x_l T_{\lambda m}\, d\sigma_\lambda \tag{1.9}$$

and

$$S_k = \epsilon_{klm} \int \left(-\frac{i}{2} \frac{\partial \mathscr{L}}{\partial(\partial_\lambda \phi_\kappa)} (S_{lm})_{\kappa\beta} \phi_\beta + W_{l\lambda m} \right) d\sigma_\lambda. \tag{1.10}$$

In Secs. 2, 3, and 4, which deal with the scalar, spinor, and vector fields, respectively, we will show that it is possible

[a)]Work supported in part by the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/80/071856-07$1.00

to make $L_k = 0$ in the rest frame on an arbitrary space-like hyperplane and also consider the case for a moving system.

2. SCALAR FIELD

For the Lagrangian we take

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}^2 + (D_\mu^* \Phi^*)(D_\mu \Phi) + V(\Phi^* \Phi) + \mathscr{L}_d, \tag{2.1}$$

where $\mathscr{L}_d$ is the divergence

$$\mathscr{L}_d = K\partial_\mu(A_\mu \partial_\nu A_\nu - A_\nu \partial_\nu A_\mu) \tag{2.2}$$

and K is an arbitrary constant. Also, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $D_\mu = \partial_\mu - ieA_\mu$, and $V(\Phi^*\Phi)$ is a suitable self-coupling function of $\Phi^*\Phi$, with $\Phi(\mathbf{r},t) = \phi(r)e^{i\omega t}$. The $e^{i\omega t}$ time dependence leads to a time independent energy-momentum tensor that thus defines the rest frame of the system. Also, the electric current is time independent so we can choose a time-independent electromagnetic four-potential A_μ.

We consider first the integrations for L_k and S_k in (1.9) and (1.10) on the hyperplane $x_4 = \text{const.}$ ($d\sigma_4 = -i\,d^3x$). One obtains for their integrands

$$T_{4m} = -(1-2K)A_{4,n}A_{n,m} - 2KA_{\alpha,\alpha}A_{4,m} - (\omega - ieA_4)(\phi^*_{,m}\phi - \phi^*\phi_{,m}) + W_{n4m,n} \tag{2.3}$$

and

$$\frac{i\partial\mathscr{L}}{\partial(\partial_4 A_\kappa)}(S_{lm})_{\kappa\beta}A_\beta = -(1-2K)(A_l A_{4,m} - A_m A_{4,l}), \tag{2.4}$$

where we have used in (2.4) the spin matrix appropriate for vectors

$$(S_{\mu\lambda})_{\kappa\beta} = -i(\delta_{\mu\kappa}\delta_{\lambda\beta} - \delta_{\mu\beta}\delta_{\lambda\kappa}). \tag{2.5}$$

To simplify the expression (1.9) involving (2.3) we make use of the equation for the space and time parts of A_μ,

$$A_{n,mn} - A_{m,nn} + ie(\phi^*_{,m}\phi - \phi^*\phi_{,m}) - 2e^2A_m\phi^*\phi = 0, \tag{2.6}$$

$$A_{4,nn} + ie(\omega - ieA_4)2\phi^*\phi = 0. \tag{2.7}$$

Multiplying (2.6) by A_4 and (2.7) by A_m, adding and integrating with respect to $i\epsilon_{klm}x_l\,d^3x$ yields the identity

$$i\epsilon_{klm}\int x_l[A_{n,mn}A_4 + (A_m A_{4,nn} - A_{m,nn}A_4) + ieA_4(\phi^*_{,m}\phi - \phi^*\phi_{,m}) + 2ie\omega A_m\phi^*\phi]\,d^3x = 0. \tag{2.8}$$

Now, adding (2.8) to (1.9) using (2.3), we can write

$$L_k = i\epsilon_{klm}\int x_l[(1-2K)(A_{n,m}A_4)_{,n} + (A_m A_{4,nn} - A_{m,nn}A_4) + (2KA_{n,n}A_4)_{,m} + \omega(\phi^*_{,m}\phi - \phi^*\phi_{,m}) + 2ie\omega A_m\phi^*\phi - W_{n4m,n}]\,d^3x. \tag{2.9}$$

Integrating by parts, we have

$$\epsilon_{klm}\int x_l(A_{n,m}A_4)_{,n}\,d^3x = \epsilon_{klm}\int A_l A_{4,m}\,d^3x, \tag{2.10}$$

$$\epsilon_{klm}\int x_l(A_m A_{4,nn} - A_{m,nn}A_4)\,d^3x = 2\epsilon_{klm}\int A_l A_{4,m}\,d^3x, \tag{2.11}$$

and

$$\epsilon_{klm}\int x_l(2KA_{n,n}A_4)_{,m}\,d^3x = 0. \tag{2.12}$$

Thus

$$L_k = -i\epsilon_{klm}\int(2K-3)A_l A_{4,m}\,d^3x - i\epsilon_{klm} \times \int x_l\omega[(\phi^*\phi_{,m} - \phi^*_{,m}\phi) - 2ieA_m\phi^*\phi]\,d^3x - i\epsilon_{klm}\int W_{n4m,n}\,d^3x. \tag{2.13}$$

For $K = \frac{3}{2}$ the first integral in (2.13) is zero. In the second integral we make use of (2.6) to write the integrand

$$-\omega[(\phi^*_{,m}\phi - \phi^*\phi_{,m}) + 2ieA_m\phi^*\phi] = (i\omega/e)(A_{m,nn} - A_{n,nm}). \tag{2.14}$$

In the right side of (2.14) $\epsilon_{klm}\int x_l A_{n,nm}\,d^3x = 0$, so (2.13) becomes

$$L_k = -i\epsilon_{klm}\int x_l[(i\omega/e)A_{m,nn} + W_{n4m,n}]\,d^3x. \tag{2.15}$$

The first term in (2.15) can be written

$$\int x_l A_{m,nn}\,d^3x = \int \partial_n(x_l A_{m,n})\,d^3x - \int \partial_l A_m\,d^3x. \tag{2.16}$$

Both integrals in (2.16) can be expressed as surface integrals that are nonzero only if the system has a magnetic-dipole moment. In that case (2.15) suggests that, in order that $L_k = 0$, the simplest expression for $W_{\mu\lambda\nu}$ should be linear in the derivatives of A_σ and include a 4-vector whose fourth component would match ω. The most general form for such a combination antisymmetric in $(\mu\lambda)$ can be expressed as follows[2]:

$$W_{\mu\lambda\nu} = (A_{\nu,\mu}a_\lambda - A_{\nu,\lambda}a_\mu) + (A_{\mu,\nu}b_\lambda - A_{\lambda,\nu}b_\mu) + (A_{\mu,\lambda} - A_{\lambda,\mu})c_\nu. \tag{2.17}$$

The expression (2.17) involves three arbitrary 4-vectors a_μ, b_μ, and c_μ, which, in this paper, we will take to be constant. We thus find that

$$W_{n4m,n} = A_{m,nn}a_4 + A_{n,nm}b_4 - (A_{4,mn}b_n + A_{4,nn}c_m). \tag{2.18}$$

Hence, choosing $a_4 = -i\omega/e$ and substituting in (2.15), we get

$$L_k = i\epsilon_{klm}\int x_l(A_{4,mn}b_n + A_{4,nn}c_m)\,d^3x. \tag{2.19}$$

The first term in (2.19) can be expressed as a surface integral by partial integration:

$$\epsilon_{klm}\int x_l A_{4,mn}\,d^3x = \epsilon_{klm}\int\partial_m(x_l A_{4,n})\,d^3x = \epsilon_{klm}\int_s \frac{1}{r}x_m x_l A_{4,n}r^2\,d\Omega = 0.$$

The second term on the other hand in (2.19) gives a

nonzero contribution if the system has an electric dipole moment, consequently either $c_m = 0$ or we would have to exclude solutions with electric dipole moments. However, the above considerations apply only to the expression for L_k on the hyperplane $x_4 = \text{const}$. More than three 4-vectors are needed to take into account the requirement that L_k should be zero on any space-like surface; also there must be no surface contributions at spatial infinity to the energy, momentum, and total angular momentum due to dipole terms in $W_{\mu\lambda\nu}$, as well as surface terms due to nonzero electric charge.

To augment the three 4-vectors introduced in Eq. (2.17) we can add to (2.17) terms linear in the 4-vectors, of the form

$$(x_{\sigma_1}\partial_{\sigma_1}x_{\sigma_2}\partial_{\sigma_2}\cdots x_{\sigma_s}\partial_{\sigma_s})W^s_{\mu\lambda\nu} \equiv (x_\sigma\partial_\sigma)^s W^s_{\mu\lambda\nu}, \tag{2.20}$$

where $W^s_{\mu\lambda\nu}$ is similar to (2.17) with a_μ, b_μ, and c_μ replaced by a new triplet a^s_μ, b^s_μ, and c^s_μ, i.e.,

$$W^s_{\mu\lambda\nu} = (A_{\nu,\mu}a^s_\lambda - A_{\nu,\lambda}a^s_\mu) + (A_{\mu,\nu}b^s_\lambda - A_{\lambda,\nu}b^s_\mu) + (A_{\mu,\lambda} - A_{\lambda,\mu})c^s_\nu. \tag{2.21}$$

Any number of triplets of 4-vectors (say $N+1$) can now be used to construct

$$W_{\mu\lambda\nu} = \sum_{s=0}^{N} (x_\sigma\partial_\sigma)^s W^s_{\mu\lambda\nu}. \tag{2.22}$$

We observe that (2.22) is not translationally invariant, resulting in origin-dependent contributions to the angular momentum from terms linear in the translation parameters if the system has electric charge. However, these terms can be made to cancel by choosing suitable relations between the 4-vectors.

If we confine ourselves to spacelike hyperplanes, then $d\sigma_\lambda$ for any hyperplane can be expressed in terms of $d\sigma^0_\tau$, corresponding to the hyperplane $x_4 = \text{const}$ by a boost

$$d\sigma_\lambda = a_{\lambda\tau}\, d\sigma^0_\tau = a_{\lambda 4}\, d\sigma^0_4 = -ia_{\lambda 4}\, d^3x. \tag{2.23}$$

Let $\tau_{\mu\nu} = \partial_n W_{n\mu\nu}$, then for the 4-momentum to be unaffected by the addition of $W_{\mu\lambda\nu}$ on any hyperplane requires

$$\int \tau_{\lambda\mu}\, d\sigma_\lambda = -ia_{\lambda 4}\int \tau_{\lambda\mu}\, d^3x = 0. \tag{2.24}$$

Since $a_{\lambda 4}$ is arbitrary, we require

$$\int \tau_{\lambda\mu}\, d^3x = 0. \tag{2.25}$$

Now $\int \tau_{ij}\, d^3x = 0$ because these integrals convert to surface integrals involving derivatives of the magnetic-dipole potential. Hence we need consider only

$$\int \tau_{\mu 4}\, d^3x = 0 \quad \text{and} \quad \int \tau_{4\mu}\, d^3x = 0. \tag{2.26}$$

Contributions to the integrals in (2.26) come only from the electric charge and one needs at least two 4-vectors to nullify them.

For the total angular momentum to be unaffected by the addition of $W_{\mu\lambda\nu}$ on any hyperplane requires, again using (2.20),

$$\int [(x_\mu\tau_{\lambda\nu} - x_\nu\tau_{\lambda\mu}) + (W_{\mu\lambda\nu} - W_{\nu\lambda\mu})]\, d^3x = 0. \tag{2.27}$$

For each value of λ there are six homogeneous equations for the 4-vectors, consequently (2.27) requires in general at least five 4-vectors. In addition, the orbital angular momentum must be zero on every hyperplane,

$$L_k = \epsilon_{klm}\int x_l T_{\lambda m}\, d\sigma_\lambda = 0, \tag{2.28}$$

which represents 12 inhomogeneous equations for the 4-vectors

$$\epsilon_{klm}\int x_l(T^c_{\lambda m} + \tau_{\lambda m})\, d^3x = 0, \tag{2.29}$$

where $T^c_{\lambda m}$ is the canonical energy-momentum tensor (1.4).

The nine equations of (2.29) for $\lambda = 1,2,3$ make an interesting statement relating the symmetry of the solutions of the field equations to the magnetic dipole moment. Namely, if not all of the integrals $\epsilon_{klm}\int x_l T^c_{jm}\, d^3x$ equal zero, then the system must have a magnetic dipole moment in order that (2.29) be satisfied, because the coefficients of the constant 4-vectors in the integrals $\epsilon_{klm}\int x_l\tau_{jm}\, d^3x$ are proportional to the magnetic dipole moment.

Thus, with the use of (2.22), we are able to ensure that $L_k = 0$ in the rest frame on every space-like hyperplane and that the addition of $W_{\mu\lambda\nu}$ does not change the integrals for the conserved quantities.

With $L_k = 0$ the intrinsic angular momentum is equal to the total angular momentum. Making use of (2.15), (2.27), and (2.4), with $K = \frac{3}{2}$, we have

$$S_k = \epsilon_{klm}\int \left(2iA_l\, A_{4,m} + \frac{\omega}{e}x_l\, A_{m,nn}\right) d^3x. \tag{2.30}$$

Writing the asymptotic dipole term as

$$A_m = \epsilon_{mpq}\,\mu_p x_q/r^3, \tag{2.31}$$

one gets for the second term in (2.30)

$$\epsilon_{klm}\int x_l\, A_{m,nn}\, d^3x = -8\pi\mu_k. \tag{2.32}$$

In vector notation, then

$$\mathbf{S} = \int 2(\mathbf{A}\times\mathbf{E})\, d^3x - \frac{8\pi\omega}{e}\boldsymbol{\mu}. \tag{2.33}$$

The results (2.15) and (2.33) for $\omega = 0$ correspond to those previously obtained.[3]

We now examine the orbital angular momentum when the system is moving.

First, it is easy to show that with $L_k = 0$ in the rest frame L_k, for the moving system is orthogonal to the velocity on every spacelike hyperplane.

Consider the 4-vector

$$V'_\tau = \epsilon_{\tau\sigma\mu\nu}v'_\sigma\int x'_\mu T'_{\lambda\nu}\, d\sigma'_\lambda \tag{2.34}$$

for the moving system, on an arbitrary hyperplane, where v'_σ is the system's 4-velocity. The fourth component of (2.34) is

$$V'_4 = \epsilon_{4ijk}v'_i\int x'_j T'_{\lambda k}\, d\sigma'_\lambda = \mathbf{v}'\cdot\mathbf{L}'.$$

In terms of rest-frame quantities,

$$V'_4 = a_{4\beta}V_\beta = a_{4\beta}\epsilon_{\beta\sigma\mu\nu}v_\sigma \int x_\mu T_{\lambda\nu}\, d\sigma_\lambda$$

$$= -a_{4i}\epsilon_{4ijk}v_4 \int x_j T_{\lambda k}\, d\sigma_\lambda,$$

since $v_i = 0$ in the rest frame. So

$$V'_4 = -v_4 a_{4i} L_i = 0,$$

because $L_i = 0$ in the rest frame.

Since the total angular momentum is independent of the hyperplane, it follows that the component of the intrinsic angular momentum in the direction of motion is independent of the hyperplane. In other words, with $L_k = 0$ in the rest frame one obtains a meaningful helicity for the system. However, we can go further. Making use of the constant 4-vectors available in (2.22), we can obtain a hyperplane-independent orbital angular momentum for the moving system as follows.

In the frame in which the system is moving with velocity $\mathbf{v}'$ we write

$$L'_k = \tfrac{1}{2}\epsilon_{klm}L'_{lm}, \tag{2.35}$$

where

$$L'_{lm} = \int (x'_l T'_{\lambda m} - x'_m T'_{\lambda l})\, d\sigma'_\lambda \tag{2.36}$$

on an arbitrary space-like hyperplane. In terms of rest-frame quantities,

$$L'_{lm} = a_{l\mu}a_{m\nu}L_{\mu\nu} = a_{l4}a_{mj}L_{4j} + a_{lj}a_{m4}L_{j4},$$

since $L_{ij} = 0$ in the rest frame. Therefore

$$L'_k = \epsilon_{klm}a_{lj}a_{m4}L_{j4}, \tag{2.37}$$

with

$$L_{j4} = \int (x_j T_{\lambda 4} - x_4 T_{\lambda j})\, d\sigma_\lambda. \tag{2.38}$$

Now $\int T_{\lambda j}\, d\sigma_\lambda = 0$, since the integral is just the 3-momentum that is zero on every hyperplane. Thus

$$L_{j4} = \int x_j T_{\lambda 4}\, d\sigma_\lambda. \tag{2.39}$$

The arbitrary hyperplane in (2.39) is related to the hyperplane $x_4 = \text{const}$ by $d\sigma_\lambda = \bar{a}_{\lambda 4}\, d\sigma_4^0 = -i\bar{a}_{\lambda 4}\, d^3x$, where the $\bar{a}_{\lambda 4}$ are boost coefficients. So

$$L_{j4} = -i\bar{a}_{\lambda 4} \int x_j T_{\lambda 4}\, d^3x. \tag{2.40}$$

We can use our available constant 4-vectors to make each of the integrals in (2.40) zero, ensuring also that the translationally dependent terms from $W_{\mu\lambda\nu}$ give no contribution. Now (2.39) is the defining integral for the center-of-mass coordinates on the hyperplane in the rest frame and we have arranged to place the center of mass at the origin of the coordinate system independent of the hyperplane. With a displacement of the origin to an arbitrary position R_i, L_{j4} now becomes

$$L_{j4} = R_j \int T_{\lambda 4}\, d\sigma_\lambda = iR_j M_0, \tag{2.41}$$

where M_0 is the rest mass of the system. Thus (2.35) becomes

$$L'_k = i\epsilon_{klm}a_{lj}a_{m4}R_j M_0 = \epsilon_{klm}R_l v'_m(\gamma M_0) \tag{2.42}$$

or

$$\mathbf{L}' = \mathbf{R}\times\mathbf{p}, \tag{2.43}$$

independent of the hyperplane in the frame where the system has momentum $\mathbf{p} = \gamma M_0\mathbf{v}'$. This final result is actually the justification for referring to $\epsilon_{klm}\int x_l T_{\lambda m}\, d\sigma_\lambda$ as the external orbital angular momentum.

In the following sections the above arguments regarding the external orbital angular momentum apply as well, *mutatis mutandis*, to both the spinor and vector fields.

3. SPINOR FIELD

The Lagrangian is taken as

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}^2 - \bar{\Psi}\gamma_\mu D_\mu \Psi + V(\bar{\Psi}\Psi) + \mathcal{L}_d, \tag{3.1}$$

where $\Psi(\mathbf{r},t) = \psi(\mathbf{r})e^{i\omega t}$ is a four-component spinor, $V(\bar{\Psi}\Psi)$ is a suitable self-coupling function, the γ_μ are the Dirac matrices, and

$$\mathcal{L}_d = K\partial_\mu(A_\mu\partial_\nu A_\nu - A_\nu\partial_\nu A_\mu)$$

is the same as (2.2). For the integrand in (1.9) we have

$$T_{4m} = -(1-2K)A_{4,n}A_{n,m} - 2KA_{\alpha,\alpha}A_{4,m} + \bar{\psi}\gamma_4\psi_{,m} + W_{n4m,n} + W'_{n4m,n}, \tag{3.2}$$

where $W_{\mu\lambda\nu}$ is the same as (2.22) and we have added the divergence of

$$W'_{\mu\lambda\nu} = i\bar{\psi}\gamma_\nu S'_{\mu\lambda}\psi, \tag{3.3}$$

where the spinor spin matrix

$$S'_{\mu\lambda} = \tfrac{1}{2}\Sigma_{\mu\lambda} = (1/4i)(\gamma_\mu\gamma_\lambda - \gamma_\lambda\gamma_\mu), \tag{3.4}$$

so that

$$W'_{n4m} = \tfrac{1}{2}i\bar{\psi}\gamma_m\Sigma_{n4}\psi = -\tfrac{1}{2}i\bar{\psi}\gamma_4\Sigma_{nm}\psi + \tfrac{1}{2}\bar{\psi}\gamma_4\delta_{mn}\psi. \tag{3.5}$$

For the integrand of (1.10) we have

$$-\frac{i}{2}\frac{\partial\mathcal{L}}{\partial(\partial_4 A_k)}(S_{lm})_{\kappa\beta}A_\beta - \frac{i}{2}\frac{\partial\mathcal{L}}{\partial(\partial_4\psi)}S'_{lm}\psi + W_{l4m} + W'_{l4m},$$

where the first term is obtained from (2.4) and the second term

$$-\frac{i}{2}\frac{\partial\mathcal{L}}{\partial(\partial_4\psi)}S'_{lm}\psi = \frac{i}{4}\bar{\psi}\gamma_4\Sigma_{lm}\psi. \tag{3.6}$$

In order to simplify the expression for the orbital angular momentum, we will need the field equations for A_k, A_4, ψ, and $\bar{\psi}$; these are

$$-A_{m,nn} + A_{n,nm} - ie\bar{\psi}\gamma_m\psi = 0, \tag{3.7}$$

$$A_{4,nn} + ie\bar{\psi}\gamma_4\psi = 0, \tag{3.8}$$

$$\gamma_k\psi_{,k} + \omega\gamma_4\psi - ie\gamma_\nu A_\nu\psi - V'\psi = 0, \tag{3.9}$$

$$\bar{\psi}_{,k}\gamma_k - \omega\bar{\psi}\gamma_4 + ie\bar{\psi}\gamma_\nu A_\nu + \bar{\psi}V' = 0, \tag{3.10}$$

where

$$V' = \frac{\partial V}{\partial(\bar{\psi}\psi)}.$$

Multiplying (3.7) by A_4 and (3.8) by A_m and adding yields the identity

$$A_{n,nm}A_4 + A_m A_{4,nn} - A_{m,nn}A_4 - ie\bar{\psi}(\gamma_m A_4 - \gamma_4 A_m)\psi = 0. \quad (3.11)$$

Also, multiplying (3.9) by $\bar{\psi}\gamma_4\gamma_m$ and (3.10) by $\gamma_m\gamma_4\psi$ and subtracting yields the identity

$$(\bar{\psi}\gamma_4\psi_{,m} - \bar{\psi}_{,m}\gamma_4\psi) - i\partial_n(\bar{\psi}\gamma_4\Sigma_{nm}\psi) - 2\omega\bar{\psi}\gamma_m\psi + 2ie\bar{\psi}(\gamma_m A_4 - \gamma_4 A_m)\psi = 0. \quad (3.12)$$

Now, multiplying (3.12) by $\frac{1}{2}$, adding it to (3.11), and integrating with respect to $i\epsilon_{klm}x_l\, d^3x$, we get the identity

$$i\epsilon_{klm}\int x_l\Big(A_{n,nm}A_4 + (A_m A_{4,nn} - A_{m,nn}A_4) + \tfrac{1}{2}(\bar{\psi}\gamma_4\psi_{,m} - \bar{\psi}_{,m}\gamma_4\psi) - \frac{i}{2}\partial_n(\bar{\psi}\gamma_4\Sigma_{nm}\psi) - \frac{i\omega}{e}A_{m,nn}\Big)d^3x = 0, \quad (3.13)$$

where we have used (3.7) to obtain the last term in (3.13) by the relation

$$\epsilon_{klm}\int x_l\frac{i\omega}{e}A_{m,nn}\,d^3x = +\epsilon_{klm}\int x_l\omega\bar{\psi}\gamma_m\psi\,d^3x.$$

Using (3.2), the orbital angular momentum can be written

$$L_k = -i\epsilon_{klm}\int x_l[-(1-2K)A_{4,n}A_{n,m} - 2KA_{\alpha,\alpha}A_{4,m} + \tfrac{1}{2}(\bar{\psi}\gamma_4\psi_{,m} - \bar{\psi}_{,m}\gamma_4\psi) + W_{n4m,n} + W'_{n4m,n}]\,d^3x, \quad (3.14)$$

where we have used, for $l \neq m$,

$$\int x_l\bar{\psi}\gamma_4\psi_{,m}\,d^3x = -\int x_l\bar{\psi}_{,m}\gamma_4\psi\,d^3x.$$

To (3.14) we now add (3.13). For $K = \frac{3}{2}$ the terms that depend only on A_μ, including $W_{n4m,n}$, cancel, as was the case for the scalar field, leaving

$$L_k = -i\epsilon_{klm}\int x_l\Big(\frac{i}{2}\partial_n(\bar{\psi}\gamma_4\Sigma_{nm}\psi) + W'_{n4m,n}\Big)d^3x = 0,$$

using (3.5).

The A_μ contribution to the intrinsic angular momentum is formally identical to the scalar-field case. So, making use of (2.33), (3.5), and (3.6), we get

$$\mathbf{S} = \int[2(\mathbf{A}\times\mathbf{E}) - \tfrac{1}{4}\bar{\psi}\gamma_4\boldsymbol{\Sigma}\psi]\,d^3x - \frac{8\pi\omega}{e}\boldsymbol{\mu}. \quad (3.15)$$

4. VECTOR FIELD

For the vector-field Lagrangian we use the one introduced by Corben and Schwinger[4] with the mass term $m^2\Phi^*_\mu\Phi_\mu$ replaced by a suitable self-coupling term $V(\Phi^*_\mu\Phi_\mu)$

$$\mathscr{L} = -\tfrac{1}{4}F^2_{\mu\nu} - \tfrac{1}{2}(D^*_\mu\Phi^*_\nu - D^*_\nu\Phi^*_\mu)(D_\mu\Phi_\nu - D_\nu\Phi_\mu) + \tfrac{1}{2}ie\gamma F_{\mu\nu}(\Phi^*_\mu\Phi_\nu - \Phi^*_\nu\Phi_\mu) + V(\Phi^*_\mu\Phi_\mu) + \mathscr{L}_d + \mathscr{L}'_d, \quad (4.1)$$

where γ is an arbitrary dipole parameter and again we use the divergence

$$\mathscr{L}_d = K\partial_\mu(A_\mu\partial_\nu A_\nu - A_\nu\partial_\nu A_\mu),$$

plus the divergence

$$\mathscr{L}'_d = C\partial_\mu(\Phi^*_\mu\partial_\nu\Phi_\nu - \Phi^*_\nu\partial_\nu\Phi_\mu), \quad (4.2)$$

where C is a constant. The time dependence is again of the form $\Phi_\mu(\mathbf{r},t) = \phi_\mu e^{i\omega t}$.

For the energy-momentum tensor one gets

$$T_{4m} = -[(1-2K)A_{4,n}A_{n,m} - ie\gamma(\phi^*_n\phi_4 - \phi^*_4\phi_n)A_{n,m} + 2KA_{\alpha,\alpha}A_{4,m}] - (1-C)(\phi^*_{n,m}\phi_{4,n} + \phi^*_{4,n}\phi_{n,m}) + ieA_n(\phi^*_{n,m}\phi_4 - \phi^*_4\phi_{n,m}) + (\omega - ieA_4)\times(\phi^*_{n,m}\phi_n - \phi^*_n\phi_{n,m}) - C(\phi^*_{n,n}\phi_{4,m} + \phi^*_{4,m}\phi_{n,n}) + W_{n4m,n} + W''_{n4m,n}, \quad (4.3)$$

where $W_{\mu\lambda\nu}$ is the same as (2.22) and we choose

$$W''_{\mu\lambda\nu} = -ie(\gamma-1)[A_\mu(\phi^*_\lambda\phi_\nu - \phi^*_\nu\phi_\lambda) - A_\lambda(\phi^*_\mu\phi_\nu - \phi^*_\nu\phi_\mu)] + 2ie(\phi^*_\mu\phi_\lambda - \phi^*_\lambda\phi_\mu)A_\nu + [h_\mu(\phi^*_\lambda\phi_\nu - \phi^*_\nu\phi_\lambda) - h_\lambda(\phi^*_\mu\phi_\nu - \phi^*_\nu\phi_\mu)], \quad (4.4)$$

where h_μ is a constant 4-vector.

For the intrinsic angular momentum (1.10) we find

$$S_k = -i\epsilon_{klm}\int\{(1-2K)A_l A_{4,m} + (1-C)\times(\phi^*_l\phi_{4,m} + \phi^*_{4,m}\phi_l) - ie\times[(\gamma-1)A_l(\phi^*_m\phi_4 - \phi^*_4\phi_m) - 2\phi^*_l\phi_m A_4] - 2\omega\phi^*_l\phi_m + W_{l4m} + W''_{l4m}\}\,d^3x. \quad (4.5)$$

For the orbital angular momentum we will need the field equations. The ones for A_m, A_4 are, in terms of ϕ_m and ϕ_4,

$$-A_{m,nn} + A_{n,nm} - ie\gamma(\phi^*_m\phi_n - \phi^*_n\phi_m)_{,n} - ie(\phi^*_{n,m}\phi_n - \phi^*_{m,n}\phi_n) + ie(\phi^*_n\phi_{n,m} - \phi^*_n\phi_{m,n}) + 2e^2A_m(\phi^*_n\phi_n + \phi^*_4\phi_4) - e^2A_n(\phi^*_m\phi_n + \phi^*_n\phi_m) - ie(\phi^*_{4,m}\phi_4 - \phi^*_4\phi_{4,m}) - ie(\omega - ieA_4)(\phi^*_m\phi_4 + \phi^*_4\phi_m) = 0, \quad (4.6)$$

$$-A_{4,nn} - ie\gamma(\phi^*_4\phi_n - \phi^*_n\phi_4)_{,n} + ie(\phi^*_{4,n}\phi_n - \phi^*_n\phi_{4,n}) - e^2A_n(\phi^*_n\phi_4 + \phi^*_4\phi_n) + 2ie\omega\phi^*_n\phi_n + 2e^2A_4\phi^*_n\phi_n = 0. \quad (4.7)$$

The equations for Φ_m, Φ_4 are

$$D_\nu(D_m\Phi_\nu - D_\nu\Phi_m) - \Phi_m V' - ie\gamma\Phi_\nu F_{m\nu} = 0, \quad (4.8)$$

$$D_\nu(D_4\Phi_\nu - D_\nu\Phi_4) - \Phi_4 V' - ie\gamma\Phi_\nu F_{4\nu} = 0. \quad (4.9)$$

We now construct the following identity,

$$[A_4(4.6) - A_m(4.7)] + [\Phi^*_4(4.8) - \Phi^*_m(4.9)] + [\Phi_4(4.8)^* - \Phi_m(4.9)^*] = 0, \quad (4.10)$$

where, for example, $\Phi_4(4.8)^*$ means the complex conjugate of the left side of (4.8) multiplied by Φ_4, etc.

After carrying out the time derivatives in (4.10), we obtain an expression of the form

$$M_m - \omega N_m = 0, \quad (4.11)$$

where the term independent of ω

$$M_m = A_{n,nm}\,A_4(A_m\,A_{4,nn} - A_4\,A_{m,nn}) + (\phi^*_{n,nm}\phi_4 + \phi^*_4\phi_{n,nm}) + (\phi^*_m\phi_{4,nn} - \phi^*_{m,nn}\phi_4) + (\phi^*_{4,nn}\phi_m - \phi^*_4\phi_{m,nn})$$
$$+ ie(\gamma-1)\,[(\phi^*_4\phi_n - \phi^*_n\phi_4)A_m - (\phi^*_m\phi_n - \phi^*_n\phi_m)A_4]_{,n} + ie\gamma(\phi^*_n\phi_4 - \phi^*_4\phi_n)A_{n,m}$$
$$+ 2ie\,[(\phi^*_4\phi_m - \phi^*_m\phi_4)A_n]_{,n} - ie(\phi^*_4\phi_{n,m} - \phi^*_{n,m}\phi_4)A_n - ie(\phi^*_{n,m}\phi_n - \phi^*_n\phi_{n,m})A_4,$$

and

$$N_m = 2ieA_m(\phi^*_n\phi_n + \phi^*_4\phi_4) + (\omega - ieA_4)(\phi^*_4\phi_m + \phi^*_m\phi_4)$$
$$+ (\phi^*_m\phi_{n,n} - \phi^*_{n,n}\phi_m) - ieA_n(\phi^*_m\phi_n + \phi^*_n\phi_m) + (\phi^*_{4,m}\phi_4 - \phi^*_4\phi_{4,m}).$$

From (4.11) we construct the integral identity

$$i\epsilon_{klm}\int x_l(M_m - \omega N_m)\,d^3x = 0 \tag{4.12}$$

and add it to the orbital angular momentum involving (4.3) to get

$$L_k = i\epsilon_{klm}\int x_l\{[(1-2K)A_4\,A_{n,m}]_{,n} + (1-C)(\phi^*_{n,m}\phi_4 + \phi^*_4\phi_{n,m})_{,n} + (A_m\,A_{4,nn} - A_4\,A_{m,nn}) + (\phi^*_m\phi_{4,nn} - \phi^*_{m,nn}\phi_4)$$
$$+ (\phi^*_{4,nn}\phi_m - \phi^*_4\phi_{m,nn}) + ie(\gamma-1)\,[(\phi^*_4\phi_n - \phi^*_n\phi_4)A_m - (\phi^*_m\phi_n - \phi^*_n\phi_m)A_4]_{,n} + 2ie\,[(\phi^*_4\phi_m - \phi^*_m\phi_4)A_n]_{,n}$$
$$+ [2KA_{n,n}\,A_4 + C(\phi^*_{n,n}\phi_4 + \phi^*_4\phi_{n,n})]_{,m} + \omega[-N_m - (\phi^*_{n,m}\phi_n - \phi^*_n\phi_{n,m})] - W_{n4m,n} - W''_{n4m,n}\}\,d^3x. \tag{4.13}$$

Integrating (4.13) by parts, we obtain

$$L_k = -i\epsilon_{klm}\int [(2K-3)A_l\,A_{4,m} + (C-3)(\phi^*_l\phi_{4,m} + \phi^*_{4,m}\phi_l) + ie(\gamma-3)$$
$$\times(\phi^*_m\phi_4 - \phi^*_4\phi_m)A_l + 2ie(\gamma-1)\,\phi^*_l\phi_m\,A_4]\,d^3x - i\epsilon_{klm}$$
$$\times\int \omega x_l[+N_m + (\phi^*_{n,m}\phi_n - \phi^*_n\phi_{n,m})]\,d^3x - i\epsilon_{klm}\int x_l(W_{n4m,n} + W''_{n4m,n})\,d^3x. \tag{4.14}$$

The second integral in (4.14) can be simplified by making use of (4.6). Multiplying that equation by $i\epsilon_{klm}x_l$ and integrating yields an identity that can be written, after one partial integration, as

$$i\epsilon_{klm}\int x_l\{A_{n,nm} - A_{m,nn} - ie\gamma(\phi^*_m\phi_n - \phi^*_n\phi_m)_{,n} - ie[2ieA_m(\phi^*_n\phi_n + \phi^*_4\phi_4) + (\omega - ieA_4)(\phi^*_m\phi_4 + \phi^*_4\phi_m)$$
$$- ie(\phi^*_m\phi_n + \phi^*_n\phi_m)A_n + (\phi^*_{4,m}\phi_4 - \phi^*_4\phi_{4,m}) + (\phi^*_{n,m}\phi_n - \phi_{n,m}\phi^*_n)$$
$$- (\phi^*_{n,n}\phi_m - \phi^*_m\phi_{n,n})]\}\,d^3x + i\epsilon_{klm}\int 2ie\phi^*_l\phi_m\,d^3x = 0. \tag{4.15}$$

With the help of (4.15), the second integral in (4.14) can be expressed as

$$-i\epsilon_{klm}\int\omega x_l[+N_m + (\phi^*_{n,m}\phi_n - \phi^*_n\phi_{n,m})]\,d^3x$$
$$= \left(\frac{i\omega}{e}\right)i\epsilon_{klm}\int x_l[A_{n,nm} - A_{m,nn} - ie\gamma(\phi^*_m\phi_n - \phi^*_n\phi_m)_{,n}]\,d^3x - i\epsilon_{klm}\int 2\omega\phi^*_l\phi_m\,d^3x. \tag{4.16}$$

Since the integral of $A_{n,nm}$ is zero and integrating the γ bracket by parts, the orbital angular momentum (4.14) becomes

$$L_k = -i\epsilon_{klm}\int [(2K-3)A_l\,A_{4,m} + (C-3)(\phi^*_l\phi_{4,m} + \phi^*_{4,m}\phi_l)]\,d^3x - i\epsilon_{klm}\int x_l\left(\frac{i\omega}{e}A_{m,nn} + W_{n4m,n}\right)d^3x$$
$$- i\epsilon_{klm}\int\{ie[(\gamma-3)(\phi^*_m\phi_4 - \phi^*_4\phi_m)A_l + 2(\gamma-1)\,\phi^*_l\phi_m\,A_4] - 2\omega(\gamma-1)\,\phi^*_l\phi_m + x_lW''_{n4m,n}\}\,d^3x. \tag{4.17}$$

Putting $K = \frac{3}{2}$, $C = 3$ makes the first integral zero. The second integral is zero, as for the scalar case. In the third integral, using (4.4) with $h_l = 0$ and $h_4 = \omega(\gamma-1)$ and integrating the last term in (4.17) by parts makes that integral zero.

Finally, proceeding to the intrinsic angular momentum (4.5), we obtain

$$S_k = -i\epsilon_{klm}\int [-2A_l\,A_{4,m} - 2(\phi^*_l\phi_{4,m} + \phi^*_{4,m}\phi_l)$$
$$- 2ieA_l(\phi^*_m\phi_4 - \phi^*_4\phi_m) - 2\gamma(\omega - ieA_4)\phi^*_l\phi_m]\,d^3x - (8\pi\omega/e)\mu_k. \tag{4.18}$$

5. DISCUSSION

For Abelian systems in which the complex fields have $e^{i\omega t}$ time dependence, we have shown that, with the use of a number of auxiliary 4-vectors, it is possible to define an energy-momentum tensor (not necessarily symmetric) that yields an orbital angular momentum $\mathbf{L} = \mathbf{r}\times\mathbf{p}$ independent of the space-like hyperplane on which it is evaluated. We expect that this should also be possible for systems with more

complicated time dependence, but the analysis would then be appreciably more complicated because of the time dependence of the energy-momentum tensor. Also, we would expect the above to apply as well to non-Abelian fields.

The selection of hyperplanes for the space-like surfaces was dictated by the choice of making the 4-vectors in $W_{\mu\lambda\nu}$ constant. For more general spacelike surfaces the 4-vectors could be considered as functions of the coordinates. This would lead to coupled first-order partial differential equations for the 4-vectors, but whether a consistent set of solutions exists is still to be determined.

[1]See, for example, D. Lurié, *Particles and Fields* (Interscience, New York, 1968).

[2](2.17) can also be written in terms of the vector spin tensor,

$$W_{\mu\lambda\nu} = i(S_{\mu\lambda})_{\kappa\beta}[A_{\nu,\kappa}a_\beta + A_{\kappa,\nu}(b_\beta + c_\beta)] + i(S_{\lambda\nu})_{\kappa\beta}A_{\mu,\kappa}c_\beta + i(S_{\nu\mu})_{\kappa\beta}A_{\lambda,\kappa}c_\beta.$$

[3]D.S. Phillips and H. Schiff, J. Phys. A **12**, 999 (1979).

[4]H.C. Corben and J. Schwinger, Phys. Rev. **58**, 953 (1940).

Classical supersymmetric particles [a)]

Carlos A. P. Galvao [b)]
Joseph Henry Laboratories, Princeton, University, Princeton, New Jersey 08544

Claudio Teitelboim [c)]
The Institute for Advanced Study, Princeton, New Jersey 08540

(Received 12 December 1979; accepted for publication 1 February 1980)

The dynamics of classical spinning particles is studied from the point of view of gauge supersymmetry. The central idea is that the natural way of introducing intrinsic spin degrees of freedom into a physical system is to take the square root of the Hamiltonian generators of the system without spin, which is equivalent to rendering the system gauge supersymmetric. This is accomplished by describing the spin degrees of freedom by means of "anticommuting c-numbers" (odd Grassman algebra elements) and relying on Dirac's theory of constrained Hamiltonian systems. The requirement of gauge supersymmetry fixes completely the action principle and leaves neither room nor need for *ad hoc* subsidiary conditions on the relative direction of the spin and the velocity as in the more traditional treatments. Both massive and massless particles free and in interaction with electromagnetic and gravitational fields are discussed. It is found that there exists a supergauge in which the spin tensor of a massive particle in a gravitational field is transported in parallel but the particle does not follow a geodesic. Massless particles on the other hand have the property of possessing a supergauge where their helicity is conserved and in which at the same time the worldline is a geodesic. Special attention is paid to the meaning and properties of the supergauge transformations. The main aspects of that discussion are applicable to more complicated systems such as supergravity. In particular phenomena such as necessity of invoking the equations of motion to close the gauge are analyzed.

I. INTRODUCTION

This article approaches the old problem of defining and studying the dynamics of classical spinning particles from what one could perhaps call a "modern" point of view, namely that of gauge supersymmetry. The central idea is that the natural way of introducing intrinsic angular momentum (spin) degrees of freedom on a physical system is to take the square root of the Hamiltonian generators of the system without spin (see for example Ref. 1 and references therein). This is accomplished by relying on Dirac's theory of constrained Hamiltonian systems.

The square root procedure, actually discovered by Dirac himself in connection with the electron wave equation, is equivalent in today's parlance to requiring that the spin degrees of freedom should enter in such a way so as to make the system invariant under a gauge supersymmetry transformation. The demand that the system by supersymmetric fixes completely the action principle and there is neither room nor necessity for physically significant *ad hoc* restrictions on the relative directions of the spin and the velocity of the system as in more traditional approaches.[2-4]

The price for demanding that the system be supersymmetric is that the spin degrees of freedom must be classically described by odd elements of a Grassmann algebra ("anticommuting c-numbers") and have no direct physical significance. This a reflection of the fact that what is being described is a microscopic spinning particle (of spin one-half in our case) and not a macroscopic rotating system.

However the fact that the classical spin degrees of freedom do not have a direct physical meaning should not be taken as indicating that the approach itself is devoid of utility. First of all, this way of treating the problem puts it in direct relationship with more complicated systems such as supergravity.[5] Indeed it appears that a classical spinning particle may be regarded as supergravity in zero dimensions. This analogy permits one to gain insight both into the dynamics of the spinning particle and also on some puzzling aspects supergravity itself—such as the closure of the gauge algebra—which also appear here in a technically simpler context.

Second, a well-defined canonical quantization procedure exists for these systems so that the classical theory can be used as a starting point for canonical quantization. Although this possibility is quite important for more complicated systems such as supergravity it is really not so useful in our case where the quantum system as a primary goal is well understood (Dirac equation in an external field). It may be nevertheless valuable as testing ground for quantization procedures of more complicated gauge supersymmetric theories.

Furthermore there is a reason to believe[6] that a semiclassical approximation for systems described by anticommuting c-numbers may be developed. If this is so the results obtained here—such as the fact that there exists a supergauge in which a massless spinning particle in a gravitational field follows a geodesic, might be translated in a precise manner into a clear physical statement.

The plan of the article is as follows. Section II reviews

[a)]Research supported by National Science Foundation Grant No. PHY79-19887 to the Institute for Advanced Study.
[b)]On leave of absence from Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, and Dept. de Fisica, UFRN, Natal, Brasil. Supported by CNPq, Brasil.
[c)]Alfred P. Sloan Research Fellow.

 0022-2488/80/071863-18$1.00

the dynamics of nonrelativistic spinning particles with particular attention paid to boundary terms in the action principle and to conservation laws. Section III develops, starting from the Dirac equation, the relativistic, supersymmetric generalization of the action given in Sec. II. It contains a detailed general discussion of the meaning of the gauge invariances of the action both in phase and configuration spaces. The phenomena of a gauge algebra that "closes only when the equations of motion hold" which has played an important role in supergravity theory is also found to arise here. Section IV discusses Poincaré invariance of the action and the notion of spin for both massive and massless particles. In the latter case it is shown that the notion of spin vector is not a meaningful one but that one has to deal only with the helicity. Section V treats the dynamics of a supersymmetric particle under the action of an external electromagnetic field as a problem of interest in itself but also as preparation for Sec. VI which deals with the gravitational interaction. In that last section the case of massive and massless particles is considered separately. In the massive case there exists a supergauge where the spin is "covariantly conserved" (transported in parallel) but the particle does not follow a geodesic, whereas in the massless case, in a special supergauge, the helicity is also conserved and in that same supergauge the orbital motion is geodesic. In this sense massless spinning particles couple minimally to the gravitational field whereas massive ones do not.

Finally Appendix A presents some standard material on Hamiltonian dynamics on a Grassmann algebra whereas Appendix B contains a discussion of the Dirac equation in curved space–time.

Some well-known results of other investigators are rederived for the sake of clarity and in order to set the new work in proper perspective. Thus, Secs. II, IIa–c, V, and Appendix A are to a large extent contained in the work of Berezin and Marinov[7] and Casalbuoni *et al.*[8] The presentation of Secs. IIId–f, IV, and VI seems to be largely new. Appendix B contains standard material and is included mostly to fix the notation. We would like, however, to stress that we have no pretense whatever of systematically reviewing the interesting and numerous previous works in this field.

II. NONRELATIVISTIC SPIN ONE-HALF PARTICLE

A. Constraints

As a simple example we outline the discussion of the nonrelativistic spinning particle. The analysis of this system will provide us with concepts we will need for the development of the more complex systems discussed in the following sections.

We consider a nonrelativistic free particle (interactions will be considered for the relativistic case in Secs. V and VI below) with position coordinates $x^i(t)$, $i = 1,2,3$. For the purpose of describing the spin degrees of freedom we associate with the particle three real anticommuting variables

$$\theta^i = \theta^i(t)$$

in addition to the position coordinates. The θ^i will be assumed to behave as the components of a vector under spatial rotations.

In order to construct the kinetic part of the free Lagrangian L we recall that it must be an even function (see Appendix A) and since the θ's are odd it is not possible to use $\dot\theta^2$ as a kinetic term. However, we now have a new possibility which was not available in the ordinary case, namely the term $\dot\theta\cdot\theta$ which is not a total derivative. Thus, we write the Lagrangian for the free nonrelativistic spin one-half particle as

$$L = \tfrac{1}{2}m\dot{\mathbf{x}}^2 + (i/2)\,\dot{\boldsymbol\theta}\cdot\boldsymbol\theta\,. \tag{II.1}$$

It is clear that the equations of motion derived from the Lagrangian (II.1) are invariant under rotations and also under Galilean transformations if the θ's remain unchanged under the latter. It is worthwhile to observe that the Lagrangian (II.1) exhibits a general feature which is a distinctive property of Fermi systems, namely the linearity in the time derivative. As it will be seen in the following sections the quantum mechanics based on (II.1) is that of a nonrelativistic spin $\frac{1}{2}$ particle.

To pass to the Hamiltonian we define the conjugate momenta

$$p_i = \frac{\partial L}{\partial \dot x^i} = m\dot x_i\,, \tag{II.2}$$

$$\pi_k = \frac{\partial L}{\partial \dot\theta_k} = \frac{i}{2}\theta_k\,, \tag{II.3}$$

whose generalized Poisson brackets are (see Appendix A):

$$\{x^i, p_j\} = \delta^i_{\ j}\,, \tag{II.4}$$

$$\{\theta^k, \pi_l\} = \delta^k_{\ l}\,, \tag{II.5}$$

all others being zero.

From Eq. (II.3) we obtain the primary constraints

$$\chi_k = \pi_k - \frac{i}{2}\theta_k \approx 0\,, \tag{II.6}$$

which are consequences of the linearity of L in $\dot\theta$. To check whether there are secondary constraints we write down the total Hamiltonian

$$H_T = H_c + \lambda^i\chi_i \tag{II.7}$$

$$= \frac{\mathbf{p}^2}{2m} + \lambda^i\chi_i\,, \tag{II.8}$$

where the λ^i are anticommuting Lagrange multipliers. The consistence conditions

$$\dot\chi^i = \{\chi^i, H_T\} \approx 0$$

lead to $\lambda^i = 0$. Therefore we have no secondary constraints and the χ_i are second class:

$$\{\chi_i, \chi_k\} = i\delta_{ik}\,. \tag{II.9}$$

It is now possible to introduce Dirac brackets based on the second class constraints χ_k; after this is done one may consider $\chi_k = 0$ as strong equations thus eliminating the π_k from the theory, which leaves the θ_i as the only Fermi variables, with modified brackets

$$\{\theta^i, \theta^j\}^* = i\delta^{ij}\,. \tag{II.10}$$

B. The action principle

The equations of motion are obtained by extremizing the action under small deformations of the history of the system. The allowed deformations must obey boundary con-

ditions the number of which is equal to the number of integration constants in the general solution of the equations of motion. Now, for the Bose variables x_i, we demand as usual that x_i be fixed at the initial and final times. It is not possible however to impose a similar requirement on the θ's, as this would imply two boundary conditions for a first-order differential equation. As a consequence the action for our system cannot be taken to be just the time integral of the Lagrangian but must be supplemented by a boundary term. Instead of following an inductive procedure we will simply exhibit the correct action principle and show that it satisfies all the necessary requirements.

We write the action as

$$S = \int_{t_1}^{t_2} L\, dt + (i/2)\, \boldsymbol{\theta}(t_1)\cdot \boldsymbol{\theta}(t_2) \tag{II.11}$$

and state that the solution of the equations of motion are those histories which yield no variation of S under the conditions

$$\delta \mathbf{x}(t_1) = 0, \quad \delta \mathbf{x}(t_2) = 0, \tag{II.12}$$

$$\delta \boldsymbol{\theta}(t_1) + \delta \boldsymbol{\theta}(t_2) = 0. \tag{II.13}$$

In order to verify that this action principle is a suitable one we must verify two properties: (i) Extremization of S under conditions (II.12,13) should yield just the equations of motion without additional restrictions; (ii) Those equations of motion should have a unique solution (with possible exception of some "unfortunate" choices of the boundary conditions) consistent with arbitrary given values of $\mathbf{x}(t_1)$, $\mathbf{x}(t_2)$, $\boldsymbol{\theta}(t_1) + \boldsymbol{\theta}(t_2)$.

The x-dependence of the action is the usual one so it need not concern us any longer. The novelty is in the θ part. If we vary $\theta(t)$ we find

$$\begin{aligned}\delta S &= \int_{t_2}^{t_1} dt\, i\dot{\boldsymbol{\theta}}\cdot\delta\boldsymbol{\theta} + \frac{i}{2}[\delta\boldsymbol{\theta}(2)\cdot\boldsymbol{\theta}(2) - \delta\,\boldsymbol{\theta}(1)\cdot\boldsymbol{\theta}(1)] \\ &\quad + \frac{i}{2}\,[\delta\boldsymbol{\theta}(1)\cdot\boldsymbol{\theta}(2) + \boldsymbol{\theta}(1)\cdot\delta\,\boldsymbol{\theta}(2)] \\ &= \int_{t_2}^{t_1} dt\, i\dot{\boldsymbol{\theta}}\cdot\delta\boldsymbol{\theta} - \frac{i}{2}\,[\delta\boldsymbol{\theta}(\mathbf{1}) + \delta\,\boldsymbol{\theta}(2)]\cdot[\boldsymbol{\theta}(1) - \boldsymbol{\theta}(2)],\end{aligned}$$

where we abbreviated $\delta\,\boldsymbol{\theta}(t_1) = \delta\,\boldsymbol{\theta}(1)$, etc. The boundary term vanishes on account of condition (II.13) and extremization of S yields just

$$\dot{\boldsymbol{\theta}}_k = 0 \tag{II.14}$$

as needed.

Now, let us suppose that $\boldsymbol{\theta}(1) + \boldsymbol{\theta}(2)$ is given as $2\boldsymbol{\xi}$, say. In that case there is a unique solution to Eq. (II.14) with that boundary condition, namely $\boldsymbol{\theta}(t) = \boldsymbol{\xi}$ for all t so that requirement (ii) is fulfilled.

It should be mentioned that had we attempted to fix $\boldsymbol{\theta}(1) - \boldsymbol{\theta}(2)$ instead of $\boldsymbol{\theta}(1) + \boldsymbol{\theta}(2)$ we could have also satisfied requirement (i) by properly adjoining the boundary term to the action, but we would not have been able to satisfy (ii) since in that case the boundary condition would not have fixed the solution of the equation of motion.

The action can be rewritten in Hamiltonian form as

$$S = \int_{t_1}^{t_2} dt \left(\dot{\mathbf{x}}\cdot\mathbf{p} + \frac{i}{2}\,\dot{\boldsymbol{\theta}}\cdot\boldsymbol{\theta} - \frac{\mathbf{p}^2}{2m}\right) + \frac{i}{2}\,\boldsymbol{\theta}(1)\cdot\boldsymbol{\theta}(2). \tag{II.15}$$

C. Conserved quantities

Having an action principle we can discuss conservation laws. The action (II.11) is invariant under translations, rotations, and Galilean transformations. The θ's are only affected by spatial rotations so that in the other two cases we just obtain the standard results for a spinless particle. Let us therefore analyze the case of rotations. Under that transformation we write:

$$\delta x^i = \omega^i_{\ j}\, x^j, \tag{II.16a}$$

$$\delta p^i = \omega^i_{\ j}\, p^j, \tag{II.16b}$$

$$\delta \theta^i = \omega^i_{\ j}\, \theta^j, \tag{II.16c}$$

with $\omega_{ij} = -\omega_{ji}$.

The action (II.11) is clearly invariant under this transformation. On the other hand, following Noether's procedure, we can rewrite the variation of the action as

$$\begin{aligned}\delta S = \delta\mathbf{x}\cdot\mathbf{p}\,\big|_{t_1}^{t_2} &- \frac{i}{2}[\delta\boldsymbol{\theta}(1) + \delta\boldsymbol{\theta}(2)]\cdot[\boldsymbol{\theta}(1) - \boldsymbol{\theta}(2)] \\ &+ (\text{terms vanishing when the equations of motion hold}).\end{aligned} \tag{II.17}$$

Now, the θ term in the above equation can be rewritten in this case as

$$\begin{aligned}&\frac{i}{2}\,\omega_{jk}\,[\theta^k(1) + \theta^k(2)][\theta^j(1) - \theta^j(2)] \\ &\quad = \frac{i}{2}\,\omega_{jk}\,[\theta^j(2)\theta^k(2) - \theta^j(1)\theta^k(1)]\,.\end{aligned} \tag{II.18}$$

If we now insert (II.18) into (II.17) and recall (II.16) we find that

$$J_{ik} = L_{ik} + S_{ik} \tag{II.19}$$

is a constant of motion, where

$$L_{ik} = x_i\, p_k - x_k\, p_i\,, \tag{II.20}$$

$$S_{ik} = i\theta_i\theta_k\,. \tag{II.21}$$

The dynamical variable J_{ik} is the generator of rotations and should then be identified as the total angular momentum. It splits into an orbital part L_{ik} which is not invariant under translations or Galilean transformations and an intrinsic or spin part S_{ik} which is invariant under those two kinds of transformations. In terms of their Dirac brackets L_{ik} and S_{ik} obey the customary algebra. For example, if we define the spin vector

$$S_i = -\tfrac{1}{2}\epsilon_{ijk}S_{jk}\,, \tag{II.22}$$

we have

$$\{S_i\,, S_j\}^* = \epsilon_{ijk}S_k\,. \tag{II.23}$$

It is interesting to mention here that had we neglected the surface term in (II.11) and applied Noether's procedure naively to the action $S = \int L\, dt$ we would have arrived at a definition for the spin of opposite sign to (II.21). From the point of view of conservation laws this would be of no importance for a free particle given that L_{ik} and S_{ik} are conserved independently in that case. But if interactions are brought in the wrong sign would give a wrong answer since only the sum of L_{ik} and S_{ik} is conserved in that case (and not their difference). However if one wants to identify J_{ik} as the gener-

ator of rotations the wrong sign of S_{ik} would give the wrong answer even for the free particle as J_{ij} would not generate the transformations (II.16a,b,c) in such case.

1. Quantization

In order to show that the system described by the action (II.11) corresponds to the classical limit of a nonrelativistic spin one-half particle we briefly outline the canonical quantization procedure for that system.

The key step is to convert the dynamical variables of the classical theory into operators by means of the prescription

$$\{A,B\}^* \to (i\hbar)^{-1}[\hat{A},\hat{B}]_{\pm}\,, \tag{II.24}$$

where the plus sign denotes an anticommutator (to be used when A and B are odd) and the minus sign corresponds to a commutator (to be used when at least A or B is even).

Applying this prescription to (II.10) we obtain

$$[\hat{\theta}_i\,,\hat{\theta}_j]_+ = \hbar\delta_{ij}\,, \tag{II.25}$$

from which we conclude that we have a Clifford algebra. In terms of operators acting on a vector space we know that there is just one irreducible representation of this algebra. In this representation we have

$$\hat{\theta}_i = \left(\frac{\hbar}{2}\right)^{1/2}\sigma_i\,, \tag{II.26}$$

where the σ_i are the Pauli matrices.

The quantum spin vector is then given by

$$\hat{S}_i = -\frac{i}{2}\,\epsilon_{ijk}\hat{\theta}_j\hat{\theta}_k = -\frac{\hbar}{4}\,i\epsilon_{ijk}\sigma_j\sigma_k = i\frac{\hbar}{2}\,\sigma_i\,. \tag{II.27}$$

From Eq. (II.27) we see that our theory corresponds to that of a (nonrelativistic) spin one-half particle.

III. RELATIVISTIC SPIN ONE-HALF PARTICLE

A. Constraints

We shall now proceed to discuss the relativistic generalization of the model studied in the last section. A naive generalization would simply replace the three θ_k by a four vector θ^μ and would keep the form of the action (II.11) otherwise unchanged. However that procedure would increase the number of Fermi degrees of freedom and it would not therefore be possible to interpret the new system as corresponding to a relativistic spin one-half particle.

In order to find the correct classical description we will work "backwards" from the known quantum mechanical equations of a relativistic spinning particle, namely the Dirac equation. This way to proceed seems the most direct and convincing one, as there appears to be no purely classical argument of similar simplicity. Conversely, one obtains in this way valuable insights which can be used in other cases when the quantum description is not fully understood and a classical analog is desirable as a starting point. The most valuable of these insights appears to be the idea of supersymmetry which emerges directly from the classical limit of the Dirac equation as we shall see below.

Let us therefore start from the Dirac equation

$$(\hbar\gamma^\mu\partial_\mu + m)\,\psi = 0, \tag{III.1a}$$

which implies the Klein–Gordon equation

$$(-\hbar^2\Box^2 + m^2)\,\psi = 0. \tag{III.1b}$$

We first introduce the operators

$$\hat{\theta}^\mu = i(\hbar/2)^{1/2}\,\gamma_5\gamma^\mu\,, \tag{III.2a}$$

$$\hat{\theta}_5 = (\hbar/2)^{1/2}\,\gamma_5\,, \tag{III.2b}$$

which obey the anticommutation rules

$$[\hat{\theta}^\mu\,,\hat{\theta}^\nu] = -\hbar\eta^{\mu\nu}\,, \tag{III.3a}$$

$$[\hat{\theta}_5\,,\hat{\theta}_5] = -\hbar\,, \tag{III.3b}$$

and recall that

$$[\hat{x}^\mu\,,\hat{p}_\nu] = i\hbar\delta^\mu{}_\nu\,, \tag{III.4}$$

with

$$\hat{p}_\mu = (\hbar/i)\,\partial_\mu \tag{III.5}$$

in the coordinate representation.

Now, the key step in passing to the classical limit is to interpret Eqs. (III.1a,b) as two first class constraints acting on allowed states. To this end we first multiply (III.1a) from the left by γ_5 in order to rewrite it as

$$(\hat{\theta}^\mu\hat{p}_\mu + m\hat{\theta}_5)\,\psi = 0 \tag{III.6a}$$

and, of course (III.1b) reads

$$(\hat{p}^\mu\hat{p}_\mu + m^2)\,\psi = 0. \tag{III.6b}$$

The advantage of form (III.6a) over (III.1a) is that it is homogeneous in the θ's which will become odd Grassmann variables in the classical theory because they obey anticommutation rules [Eqs. (III.3a,b)] in the quantum case.

The classical theory is now formulated in terms of four pairs of Bose (commuting) real variables $x^\mu\,,p_\nu$ and *five* (anticommuting) real variables θ^μ,θ_5 the only nonvanishing Dirac brackets among which are

$$\{\theta^\mu\,,\theta^\nu\}^* = i\eta^{\mu\nu}\,, \tag{III.7a}$$

$$\{\theta_5\,,\theta_5\}^* = i\,, \tag{III.7b}$$

$$\{x^\mu\,,p^\nu\}^* = \eta^{\mu\nu}\,. \tag{III.7c}$$

The dynamics of the system is fully contained in the classical analog of (III.6a,b), the first class constraints

$$\mathcal{S} = \theta^\mu p_\mu + m\theta_5 \approx 0, \tag{III.8a}$$

$$\mathcal{H} = p^2 + m^2 \approx 0\,, \tag{III.8b}$$

which obey the algebra

$$\{\mathcal{S}\,,\mathcal{S}\}^* = i\mathcal{H}, \tag{III.9a}$$

$$\{\mathcal{S}\,,\mathcal{H}\}^* = 0, \tag{III.9b}$$

$$\{\mathcal{H}\,,\mathcal{H}\}^* = 0. \tag{III.9c}$$

We have here denoted our brackets with a star in order to make contact with the nonrelativistic results of Sec. II and with the action functional treated below.

B. Relativistic action functional

1. Hamiltonian form

Since the whole dynamics of the theory is contained in the constraints (III.8a,b) we know that the effective Hamiltonian of the theory will simply be a linear combination of $\mathcal{S}$ and $\mathcal{H}$ with arbitrary Lagrange multipliers. Moreover the

structure of the kinetic term in the action is fixed by the brackets (III.7a,b,c), which according to our experience with the nonrelativistic particle, correspond to a term of the form $\dot{x}p + (i/2)\dot{\theta}\theta$. Therefore our action functional including the appropriate surface terms is

$$S = \int_{\tau_1}^{\tau_2} d\tau \Big[\dot{x}^\mu p_\mu + \frac{i}{2}(\dot{\theta}^\mu \theta_\mu + \dot{\theta}_5\theta_5) - N(\tau)(p^2 + m^2) - iM(\tau)(\theta^\mu p_\mu + m\theta_5)\Big] + \frac{i}{2}[\theta^\alpha(1)\theta_\alpha(2) + \theta_5(1)\theta_5(2)]\,, \tag{III.10}$$

where N and M are, respectively, even and odd real Lagrange multipliers.

The equations of motion are obtained by demanding that (III.10) should not change under small variations of $x^\mu, p_\mu, \theta^\alpha, \theta_5, N, M$ subject to the conditions

$$\delta x^\mu(1) = 0 = \delta x^\mu(2)\,, \tag{III.11a}$$

$$\delta\theta^\mu(1) + \delta\theta^\mu(2) = 0\,, \tag{III.11b}$$

$$\delta\theta_5(1) + \delta\theta_5(2) = 0\,. \tag{III.11c}$$

2. Lagrangian form

Although (III.10) is satisfactory as an action principle it is interesting to eliminate the momenta p_μ in favor of the velocities $\dot{x}^\mu$ in order to obtain an equivalent action principle in Lagrangian form. This is of interest to analyze as it brings to light some features which also appear in more complicated problems, and which may be potential sources of confusion, especially when the Lagrangian form is taken as the starting point in the construction of a theory.

The elimination of the momenta is achieved by observing that in general if we have an action of the form $S[f(\tau), g(\tau)]$ and the equations $\delta S/\delta g = 0$ may be solved to express $g = g[f]$ then we may insert this solution for g back into the action and obtain a new action $\bar{S}[f] = S\{f, g[f]\}$ which upon extremization with respect to f gives the correct equation of motion for that variable. If on the other hand it is not possible to solve for g as function of f, the insertion of $\delta S/\delta g = 0$ into the action will lead in general to equations of motion for the remaining variables which are different from the ones obtained from the original action principle, thus rendering the procedure illegitimate in that case.

We now apply this technique taking for g the set (p_μ, N). From $\delta S/\delta p_\mu = 0$ and $\delta S/\delta N = 0$ we obtain, respectively,

$$\dot{x}^\mu - 2Np^\mu - iM\theta^\mu = 0\,, \tag{III.12a}$$

$$p^2 + m^2 = 0\,, \tag{III.12b}$$

from which it follows that

$$N(\tau) = (1/2m)\sqrt{-z^2}\,, \tag{III.13}$$

with

$$z^\mu = \dot{x}^\mu - iM\theta^\mu \tag{III.14}$$

and

$$p^\mu = m z^\mu/\sqrt{-z^2}\,. \tag{III.15}$$

Substituting (III.13) and (III.15) in (III.10) yields

$$\bar{S} = \int_1^2 d\tau \Big[-m\sqrt{-z^2} + \frac{i}{2}(\dot{\theta}^\mu\theta_\mu + \dot{\theta}_5\theta_5) - iMm\theta_5\Big] + \frac{i}{2}[\theta^\alpha(1)\theta_\alpha(2) + \theta_5(1)\theta_5(2)]\,. \tag{III.16}$$

Extremization of $\bar{S}$ with respect to $x^\mu, \theta^\mu, \theta_5, M$ under conditions (III.11a,b,c) gives the correct equations of motion for x^μ, θ^μ, and θ_5.

Now it is tempting to try to eliminate $M(\tau)$ and θ_5. From $\delta\bar{S}/\delta\theta_5 = 0$ and $\delta\bar{S}/\delta M = 0$ we obtain

$$\dot{\theta}_5 = mM(\tau) \tag{III.17}$$

and

$$\theta^\mu p_\mu + m\theta_5 = 0, \tag{III.18}$$

with p_μ given by (III.15). However, these two equations do not permit to eliminate M and θ_5 as a function of the remaining variables due to the boundary conditions (III.11c) on θ_5. In fact, if we use the identities

$$\sqrt{-z^2} = (-\dot{x}^2 + 2iM\theta\cdot\dot{x})^{1/2} = \sqrt{-\dot{x}^2}\Big(1 - \frac{iM\theta\cdot\dot{x}}{\dot{x}^2}\Big),$$

$$\frac{1}{\sqrt{-z^2}} = \frac{1}{\sqrt{-\dot{x}^2}}\Big(1 + \frac{iM\theta\cdot x}{\dot{x}^2}\Big),$$

we obtain

$$\theta^\mu p_\mu = \frac{m}{\sqrt{-\dot{x}^2}}\theta\cdot\dot{x}\,.$$

Therefore we can write

$$M = \frac{1}{m}\frac{d}{d\tau}\Big(\frac{\theta\cdot\dot{x}}{\sqrt{-\dot{x}^2}}\Big), \tag{III.17a}$$

$$\theta_5 = -\frac{\theta\cdot x}{\sqrt{-\dot{x}^2}}\,. \tag{III.18a}$$

However it is not permissible to use (III.18a) to express θ_j and M as a function of the other variables because since $\dot{x}$ is free at the end points the right-hand side of (III.18a) does not fulfill the boundary condition (III.11c).

It is not legitimate either to use (III.17) to eliminate M in favor of θ_5 as is sometimes done.[7] Indeed, that procedure would be equivalent to using $\bar{S}' = \bar{S} - \int d\tau\,\lambda(\tau)[\dot{\theta}_5 - mM(\tau)]$ as an action instead of $\bar{S}$, where $\lambda(\tau)$ is an anticommuting Lagrange multiplier. This action gives rise to

$$\frac{\delta\bar{S}'}{\delta\lambda} = \dot{\theta}_5 - mM(\tau), \quad \frac{\delta\bar{S}}{\delta x^\mu} = \Big(\frac{\delta\bar{S}'}{\delta x^\mu}\Big)_{\dot{\theta}_5 = mM},$$

$$\frac{\delta\bar{S}}{\delta\theta^\mu} = \Big(\frac{\delta\bar{S}'}{\delta\theta^\mu}\Big)_{\dot{\theta}_5 = mM},$$

$$\frac{\delta\bar{S}'}{\delta M} = \frac{\delta\bar{S}}{\delta M}_{\dot{\theta}_5 = mM} + m\lambda(\tau), \quad \frac{\delta\bar{S}'}{\delta\theta_5} = \dot{\lambda}(\tau)\,.$$

Thus, by extremizing $\bar{S}'$ with respect to θ_5 we obtain $\dot{\lambda} = 0$ but not $\lambda = 0$, and instead of $(\delta\bar{S}/\delta M) = 0$ we find the weaker statement $(d/d\tau)(\delta\bar{S}/\delta M) = 0$. We see that it is not correct to use Eq. (III.17) to eliminate $M(\tau)$ from the action and it is therefore not possible to have an action principle in terms of θ_μ, θ_5, and x^μ only.

3. Gauge invariance

In this section we will denote collectively $\mathscr{H}$ and $\mathscr{S}$ by

ϕ_a. This will make the calculations more compact and will also bring out clearly that most of the discussion below is not restricted to the system at hand but is of general applicability.

The Hamiltonian for the spinning particle is a linear combination of the constraints $\mathscr{H}$ and $\mathscr{S}$:

$$H = N\mathscr{H} + iM\mathscr{S} \equiv N^a\phi_a \,. \tag{III.19}$$

Since the constraints are first class they are automatically preserved in time under the evolution generated by the Hamiltonian (III.19). Therefore the functions N (an even real element of the Grassmann algebra) and M (an odd real element) are not restricted by the equations of motion, but remain rather arbitrary functions of time. As a consequence, if we give the values of the canonical variables at an initial time τ_0 their value at a later time τ will not be uniquely determined but will depend on the choice of N and M for $\tau > \tau_0$.

Now, by definition one regards as physically meaningful only that information which can be predicted unambiguously from the initial conditions. Hence one should consider two histories which evolve from a given initial condition with two different choices for the functions N^a as physically equivalent or, as one says, as being related by a "gauge transformation."

The infinitesimal mapping which relates two histories which branch off at τ_0 due to two different choices N^a, $\bar{N}^a$ for $\tau < \tau_0$ of the multipliers in the Hamiltonian (III.19) is generated by the constraints themselves, i.e., it takes the form

$$\delta F(\tau) = \{F, \epsilon^a(\tau)\,\phi_a\}\,, \tag{III.20}$$

with

$$\epsilon^a(\tau_0) = 0,$$

for any function F of the canonical variables of the theory. If τ is infinitesimally close to τ_0 Eq. (III.20) follows simply by subtracting

$$\bar{F}(\tau_0 + d\tau) = F(\tau_0) + \{F, N^a\phi_a\}\,d\tau \tag{III.21a}$$

from

$$\bar{F}(\tau_0 + d\tau) = F(\tau_0) + \{F, \bar{N}^a\phi_a\}\,d\tau\,, \tag{III.21b}$$

so that

$$\bar{F}(\tau_0 + d\tau) - F(\tau_0 + d\tau) = \{F, \epsilon^a\phi_a\}\,,$$

with

$$\epsilon^a(\tau_0 + d\tau) = \bar{N}^a(\tau_0) - N^a(\tau_0)\,. \tag{III.22}$$

For later times the simple form (III.22) is not valid but Eq. (III.20) still holds after iterations due to the first class property of the ϕ_a [Equation (III.23) is understood to hold when ϵ^a and η^a do not depend on the canonical variables. The only reasons why we do not write dropping ϵ and η simply $\{\phi_a, \phi_b\} = \kappa_{ab}{}^c\phi_c$ is that some minus signs arise for odd ϕ's and it is better bookkeeping to always keep the parameter attached to the corresponding generator.]

$$\{\epsilon^a\phi_a, \eta^b\phi_b\} = \epsilon^a\kappa_{ab}{}^c\eta^b\phi_c\,. \tag{III.23}$$

Equation (III.20) may be implemented as a symmetry transformation of the action provided one adds a corresponding prescription for the multipliers N^c:

$$\delta N^c = \dot{\epsilon}^c + \epsilon^a\kappa_{ab}{}^c N^b\,. \tag{III.24}$$

(Note that for $\tau = \tau_0$ (III.24) reduces to (III.22) as should be the case.)

In fact if one studies the variation of an action of the form

$$\delta S = \int_{\tau_1}^{\tau_2} (\dot{q}\cdot p + \frac{i}{2}\theta\cdot\theta - N^a\phi_a)\,d\tau + \frac{i}{2}\,\theta(1)\,\theta(2)\,, \tag{III.25}$$

one finds that it reduces to a boundary term

$$\delta S = \epsilon^a\left(p\frac{\partial\phi_a}{\partial p} - \phi_a\right)\Bigg|_{\tau_1}^{\tau_2} - \frac{i}{2}[\theta(2) - \theta(1)] \times [\{\theta, \epsilon^a\phi_a\}(1) + \{\theta, \epsilon^a\phi_a\}(2)]\,. \tag{III.26}$$

(A nonvanishing first class Hamiltonian H' may be added to the $N^a\phi_a$ term without changing the conclusions.)

4. Transformations of histories versus transformations of states

The boundary term in (III.26) is crucial for the interpretation of the transformations (III.20), (III.24). In fact only if that term is zero is the action invariant. Now it is easy to convince oneself that in the case at hand expression (III.26) vanishes for arbitary values of x, p, and θ if and only if

$$\epsilon^a(\tau_1) = \epsilon^a(\tau_2) = 0. \tag{III.27}$$

Now, the fact the requirement of invariance of the action does not restrict the time dependence of the ϵ^a at points other than the endpoints is the origin of the arbitrariness of the N^a in the solutions of the equations of motion and hence is what indicates that the transformations (III.20), (III.24) relates physically indistinguishable histories.

The necessity of fixing the ϵ^a at the endpoints τ_1, τ_2 according to (III.27) points to another crucial fact. Namely it indicates that the correspondence between histories defined by (III.20), (III.24) cannot be refined to a correspondence between states at a given time. In fact, in order to deduce a mapping between states from a mapping between histories one must collapse two equivalent histories to individual points by letting τ_2 approach τ_2. If the mapping is not trivial in that limit [i.e., if $\epsilon(\tau_1) \neq 0$] the two histories will collapse into different points and one will obtain from the mapping between histories a mapping between individual points in phase space, i.e., a mapping between states.

In our case since (III.27) must hold it is not possible to deduce a mapping between states from (III.20), (III.24). Therefore it would be mistaken to conclude that (III.20) maps a state at a given time τ onto another, physically indistinguishable state at that same time. That this conclusion is correct could have been anticipated by recalling that the motion of the system is itself of the form (III.20) and interpreting it as a transformation which does not change the physical state would have led to the conclusion that there is no dynamics at all in the system.

5. Configuration space

Additional light is shed on the preceding remarks and on some other points of interest by passing to configuration space as described in Sec. III B.

One would expect that the gauge freedom associated with the appearance of arbitrary functions of time in the solution of the equations of motion would be also present in configuration space. This is indeed the case but the correspondence is not as direct as one could think and some subtleties arise.

The subtleties in question are not just a peculiarity of the problem at hand here but they also occur in more complicated systems such as supergravity theory (see, for example, Ref. 9).

Generally speaking the complications in going to configuration space come from the fact that the algebraic structure of the gauge transformation is not preserved by this "projection" operation. In fact if we write (III.20)

$$\delta F = \{F, u\mathscr{H} + iv\mathscr{S}\}, \tag{III.28}$$

and we apply this and (III.24) to our variables we find

$$\delta x^\mu = 2up^\mu + iv\theta^\mu, \tag{III.29a}$$

$$\delta\theta^\mu = vp^\mu, \tag{III.29b}$$

$$\delta\theta_5 = vm, \tag{III.29c}$$

$$\delta M = \dot{v}, \tag{III.29d}$$

$$\delta p^\mu = 0, \tag{III.30a}$$

$$\delta N = \dot{u} + ivM. \tag{III.30b}$$

Now we can use Eqs. (III.29a–d) to define a transformation between our configuration space variables x^μ, θ^μ, θ_5, M by replacing p_μ by its expression

$$p_\mu = m(-z^2)^{-1/2} z_\mu, \tag{III.31a}$$

with

$$z^\mu = \dot{x}^\mu - iu\theta^\mu. \tag{III.31b}$$

One can easily convince oneself, either by a general reasoning or by explicit calculation that Eqs. (III.29a–d) leave the configuration space action

$$\bar{S} = \int_{\tau_1}^{\tau_2} d\tau \left[-m\sqrt{-z^2} + \frac{i}{2}(\dot{\theta}^\mu\theta_\mu + \dot{\theta}_5\theta_5)\right] + \left[\frac{i}{2}\theta^\alpha(1)\theta_\alpha(2) + \theta_5(1)\theta_5(2)\right] \tag{III.32}$$

invariant. Therefore (III.29a–d) define a gauge transformation in configuration space.

So far everything appears to be quite straightforward. However if we study the composition of two transformations the simplicity seems to fade away. Consider, for example, the commutation of two transformations generated by $\mathscr{S}$. That is, look at the successive action of two transformations of the type (III.29a–d) with parameters v_1, v_2 and $u_1 = u_2 = 0$. The commutator is obtained by letting v_2 act first and v_1 second and then subtracting from that result what is obtained by the same operation with the roles of v_1 and v_2 reversed. Now, in phase space the result of that commutator acting on all the dynamical variables is equivalent to another transformation of type (III.29) with $v = 0$ and

$$u = iv_1v_2. \tag{III.33}$$

This is a consequence of the closure relation

$$\{\mathscr{S}, \mathscr{S}\} = i\mathscr{H}. \tag{III.34}$$

However in configuration space the same operation yields.

$$\Delta x^\mu = 2iv_1v_2p^\mu, \tag{III.35a}$$

$$\Delta\theta^\mu = 2iv_1v_2m(-z^2)^{-1/2}[\dot{\theta}^\mu + (1/m^2)(p\cdot\dot{\theta})p^\mu], \tag{III.35b}$$

$$\Delta\theta_5 = 0, \tag{III.35c}$$

$$\Delta M = 0. \tag{III.35d}$$

Now, Eqs. (III.35a,c,d) are in agreement with the phase space result, but (III.35b) is not. In fact according to the phase space prescription one should have $\Delta\theta^\mu = 0$ instead of expression (III.35b). The difficulty can be traced to the fact that in phase space the momentum p_μ is invariant under both $\mathscr{S}$ and $\mathscr{H}$ [Eq. (III.30a)], but if we calculate its variations from Eqs. (III.31), (III.35) we obtain the nonvanishing result

$$\delta p_\mu = 2up_\mu + ivm(-z^2)^{-1/2}[\dot{\theta}_\mu + (1/m^2)(p\cdot\dot{\theta})p_\mu]. \tag{III.36}$$

Therefore we see that in configuration space the transformations (III.29) do not form a closed algebra in spite of the fact that their phase space counterpart does. It is interesting to realize, however, that if the equations of motion hold the right-hand side of (III.36) vanishes and so does (III.29b). Thus we learn that the price paid for reducing the number of dynamical variables (i.e., passing from phase space to configuration space) is that the equations of motion get mixed up with the algebraic properties of the gauge transformations. This statement is sometimes phrased in field theoretic systems by saying that "the gauge algebra closes only on shell."

Suppose now that one were given the configuration space action to start with. He would then have regarded the lack of closure of the gauge algebra as an undesirable fact and could have cleverly realized that by introducing the additional variables p_μ, N in the description of the system the algebra can be made to close without utilizing the equations of motion. In that context one would regard p_μ N as "auxiliary variables" used merely to close the algebra.

The necessity of introducing auxiliary variables to close the gauge algebra appears to be a general feature of gauge supersymmetric systems. It should however be observed here that in more complicated cases it is not sufficient to go to phase space to close the algebra. This can be seen from taking the commutator of two transformations of the form (III.20) with parameters $\epsilon_1{}^a$, $\epsilon_2{}^a$. Using the Jacobi identity then yields

$$F = \{F, \{\epsilon_1{}^a\phi_a, \epsilon^b{}_2\phi_b\}\} = \{F, \epsilon_1{}^a\kappa_{ab}{}^c\epsilon_2{}^b\phi_c\} + \{F, \epsilon^a{}_1\kappa_{ab}{}^c\epsilon^b{}_2\}\phi_c. \tag{III.37}$$

Now, the first term on the right side of (III.37) is again a transformation of the form (III.20) with

$$\epsilon^c = \epsilon_1{}^a\kappa_{ab}{}^c\epsilon_2{}^b, \tag{III.38}$$

but the second term is not of that type unless the $\kappa_{ab}{}^c$ are independent of the dynamical variables ("c-numbers"). In our case the $\kappa_{ab}{}^c$ are indeed independent of the dynamical variables (and deserve to be called structure *constants*) but in more complicated systems (such as general relativity and

supergravity) this is not so. When the $\kappa_{ab}{}^{c}$ do not have the same value all over phase space the transformation (III.37) is of the form (III.20) only when $\phi_c \approx 0$, i.e., again the algebra closes only "on shell" (or, more precisely, "weakly"). In that case it may be necessary to introduce additional auxiliary fields beyond phase space to close the algebra. [So far there appears to be no systematic way to deal with this problem. For example, in pure general relativity the $\kappa_{ab}{}^{c}$ are field dependent but the gauge algebra of the configuration space action ("general covariance", "reparametrization invariance"—see below for its analog here) closes "off shell". Therefore in that case eliminating the momenta actually appears to simplify the problem! See in this context the idea of "rank" of a theory introduced in Ref. 10.]

6. Reparametrization invariance

Equations (III.35a–d) are the result of a composition of transformations of the form (III.29a–d) each of which leaves the action invariant. It follows therefore that the action should also be invariant under a transformation under which only the θ^{μ} change by an amount

$$\delta\theta^{\mu} = \alpha(\tau)[\dot{\theta}^{\mu} + (1/m^2)(p\cdot\dot{\theta})\,p^{\mu}]\,, \tag{III.39}$$

where α is an arbitrary function of time. The action (III.32) is indeed invariant under (III.39) as one can easily verify.

Now as we already saw the transformation (III.39) is not of the form (III.29b) and so it is a new gauge invariance of the action. It does not however introduce a third arbitrary function besides N and M in the general solution of the equations of motion. This follows simply from the fact that there are only two first class constraints $\mathscr{H}$ and $\mathscr{S}$ in the problem, and can also be seen by observing that (III.39) does not map an extremal history onto a different one, since it vanishes when the equations of motion hold. We learn therefore that superficial counting of the independent gauge symmetries of the configuration space action may lead to erroneous conclusions.

Equation (III.29) is not the only "spurious" gauge symmetry of the theory. Others appear from relating Eqs. (III.29 to reparametrization invariance as we proceed to do now.

If we insert expression (III.31a) for p_{μ} in the transformations (III.29) and define

$$\xi = 2um/\sqrt{-z^2} \tag{III.40a}$$

and

$$\epsilon = v - (2umM/\sqrt{-z^2}), \tag{III.40b}$$

so that

$$v = \epsilon + \xi M. \tag{III.40c}$$

We obtain after some rearrangements

$$\delta x^{\mu} = i\epsilon\theta^{\mu} + \xi\dot{x}^{\mu}\,, \tag{III.41a}$$

$$\delta\theta^{\mu} = \epsilon p^{\mu} + \xi\dot{\theta}^{\mu} + \tilde{\delta}\theta^{\mu}\,, \tag{III.41b}$$

$$\delta\theta_5 = \epsilon m + \xi\dot{\theta}_5 + \tilde{\delta}\theta_5\,, \tag{III.41c}$$

$$\delta M = \dot{\epsilon} + \xi M, \tag{III.41d}$$

with

$$\tilde{\delta}\theta^{\mu} = \xi\,(Mp^{\mu} - \dot{\theta}^{\mu})\,, \tag{III.42a}$$

$$\tilde{\delta}\theta_5 = \xi\,(Mm - \dot{\theta}_5)\,. \tag{III.42b}$$

Now, the ϵ-dependence of those equations is just the v-dependence of Eqs. (III.29) and is what we call a supersymmetry transformation. We set therefore $\epsilon = 0$ and study the remainder:

$$\delta x^{\mu} = \xi\dot{x}^{\mu}\,, \tag{III.43a}$$

$$\delta\theta^{\mu} = \xi\dot{\theta}^{\mu} + \tilde{\delta}\theta^{\mu}\,, \tag{III.43b}$$

$$\delta\theta_5 = \xi\dot{\theta}_5 + \tilde{\delta}\theta_5\,, \tag{III.43c}$$

$$\delta M = (d/d\tau)(\xi M)\,. \tag{III.43d}$$

The reason for isolating the $\tilde{\delta}$-part in Eqs. (III.42a,b) is that if we set $\tilde{\delta} = 0$ in Eqs. (III.42a,b), what is left, namely

$$\delta x^{\mu} = \xi\dot{x}^{\mu}\,, \tag{III.44a}$$

$$\delta\theta^{\mu} = \xi\dot{\theta}^{\mu}\,, \tag{III.44b}$$

$$\delta\theta_5 = \xi\dot{\theta}_5\,, \tag{III.44c}$$

$$\delta M = (d/d\tau)(\xi M)\,, \tag{III.44d}$$

is just the effect of a change of parameter $\tau \to \tau + \xi(\tau)$ under which x^{μ}, θ^{μ}, and θ_5 transform as scalars while M appears transforming as "density". One can see by inspection that the action (III.32) is invariant under such a reparametrization. But this can be so only if the action is separately invariant under $\tilde{\delta}$ alone and again one can verify that this is indeed the case.

Now $\tilde{\delta}$ defined by (III.42a,b) is yet another "spurious" gauge transformation different from (III.39) but sharing with it the property of reducing to the identity when the equations of motion hold. Thus we see that the "spurious" transformations proliferate quickly when the algebra closes up to the equations of motion. Nevertheless from the Hamiltonian point of view the situation is much clearer as everything is contained in the two first class constraints $\mathscr{H}$, $\mathscr{S}$.

Lastly we should emphasize that the reparametrization (III.35) results as combination of a transformation generated by $\mathscr{H}$ with a particular supersymmetry transformation [recall Eq. (III.39)] and a "spurious" transformation (III.42a,b). Furthermore it can be obtained in this way only in configuration space. Therefore calling $\mathscr{H}$ the "generator of reparametrizations" as is sometimes done appears to be an ill-advised terminology.

On the other hand the fact that $\mathscr{H}$ and $\mathscr{S}$ are related to reparametrizations in the way just discussed illustrates clearly that those constraints cannot be thought of as generating transformations which relate two physically indistinguishable states. In fact a reparametrization is an operation which continuously maps each point of the worldline onto another different point on the same line and hence clearly relates states at different times. Further one clearly cannot speak of reparametrization of a simple point but has to deal with a whole segment of a line in agreement with our previous discussion on transformations of histories versus transformations of states. One can also apprehend here the necessity for having the transformation to be the identity at the end points [Eqs. (III.27)] as if it were not so one would be mapping a segment of the worldline not onto itself but on a different one.

IV. RELATIVISTIC CONSERVATION LAWS

As discussed in the previous sections the action (III.10) is invariant under reparametrizations and supersymmetry transformations. The corresponding parameter may vary arbitrarily as a function of the time τ which is the characteristic feature of a "gauge transformation." The conserved quantities associated with gauge transformations are just the constraints which generate the transformations and which have numerical value zero for all possible solutions of the equations of motion. In this sense these conservation laws are trivial.

However besides its gauge invariance the action (III.10) is not changed by other transformations, the parameters of which must be identical for all times. The corresponding conservation laws are then nontrivial. The transformations in question are, of course, those of the Poincaré group which we pass on to study now.

We define the Poincaré behavior of our dynamical variables by

$$\delta x^{\mu} = \omega^{\mu}{}_{\nu} x^{\nu} + \epsilon^{\mu} , \tag{IV.1a}$$

$$\delta p_{\mu} = \omega_{\mu}{}^{\nu} p_{\nu} , \tag{IV.1b}$$

$$\delta\theta^{\mu} = \omega^{\mu}{}_{\nu}\theta^{\nu} , \tag{IV.1c}$$

$$\delta\theta_5 = 0 , \tag{IV.1d}$$

with $\omega_{\mu\nu} = -\omega_{\nu\mu}$. Applying Noether's procedure for the transformations (IV.1a–d) to the action (III.10) yields now

$$0 = \delta S = \epsilon^{\mu}[p_{\mu}(2) - p_{\mu}(1)] + \tfrac{1}{2}\omega^{\mu\nu}[J_{\mu\nu}(2) - J_{\mu\nu}(1)] , \tag{IV.2}$$

where

$$J^{\mu\nu} = x^{\mu}p^{\nu} - x^{\nu}p^{\mu} + i\theta^{\mu}\theta^{\nu} \tag{IV.3}$$

are the canonical generators of Lorentz rotations which we identify as the total angular momentum of the particle. Together with p_{μ} these generators obey the Poincaré group algebra,

$$\{x^{\mu}, p_{\nu}\}^{*} = \delta^{\mu}{}_{\nu} , \tag{IV.4a}$$

$$\{J^{\mu\nu}, x^{\alpha}\}^{*} = -x^{\mu}\eta^{\nu\alpha} + x^{\nu}\eta^{\mu\alpha} , \tag{IV.4b}$$

$$\{J^{\mu\nu}, p_{\alpha}\}^{*} = -p^{\mu}\delta^{\nu}{}_{\alpha} + p^{\nu}\delta^{\mu}{}_{\alpha} , \tag{IV.4c}$$

$$\{J^{\mu\nu}, J^{\alpha\beta}\}^{*} = J^{\mu\alpha}\eta^{\nu\beta} - J^{\mu\beta}\eta^{\nu\alpha} - J^{\nu\alpha}\eta^{\mu\beta} + J^{\nu\beta}\eta^{\mu\alpha} . \tag{IV.4d}$$

Since the Hamiltonian is a linear combination of the constraints with arbitrary coefficients, the fact that $J_{\mu\nu}$ and p_{ν} are conserved quantities may be restated as saying that both of these quantities have zero Dirac brackets with the constraints $\mathscr{H}$ and $\mathscr{S}$. This assertion may of course be checked directly. Thus, in particular the conserved quantities are "supersymmetric invariant."

A. The definition of spin

We will now deal with the problem of separating the total angular momentum $J_{\mu\nu}$ into an orbital part $L_{\mu\nu}$ and a spin part $S_{\mu\nu}$ in analogy with the nonrelativistic splitting (II.19–21).

The spin will be defined as the projection of the total angular momentum on the subspace orthogonal to p_{μ}. This definition is satisfactory when the particle has a mass different from zero since the spin so defined is unambiguous and has the following properties:

(1) It is conserved for a free particle. In particular it is invariant under supersymmetry transformations.

(2) It is translation invariant ("intrinsic" property).

(3) It obeys the algebra of the Lorentz group.

If the particle has zero mass the above definition becomes ambiguous and is therefore not satisfactory. That case will be treated separately.

A. Massive particles

As stated above we define the spin tensor by

$$S^{\mu\nu} \approx \mathscr{M}^{\mu}{}_{\rho}\mathscr{M}^{\nu}{}_{\lambda} J^{\rho\lambda} , \tag{IV.5}$$

where $\mathscr{M}^{\mu}{}_{\rho}$ is the projection operator on the subspace orthogonal to p:

$$\mathscr{M}^{\mu}{}_{\rho} \approx \delta^{\mu}{}_{\rho} + (1/m^2)\, p^{\mu}p_{\rho} . \tag{IV.6}$$

Since $\mathscr{M}$ is defined in terms of p only and both p_{μ} and $J^{\alpha\beta}$ are conserved, it follows immediately that $S^{\mu\nu}$ is conserved. This means that the brackets of $S^{\mu\nu}$ with both $\mathscr{S}$ and $\mathscr{H}$ vanish, so in particular the spin tensor is supersymmetrically invariant.

Furthermore, since

$$\mathscr{M}^{\mu}{}_{\rho}\, p^{\rho} \approx 0, \tag{IV.7}$$

it follows that $S^{\mu\nu}$ is translation invariant. In fact under a translation a_{μ}

$$J_{\mu\nu} \to J_{\mu\nu} + a_{\mu} p_{\nu} - a_{\nu} p_{\mu} \tag{IV.8}$$

and the terms containing a drop out when projected with $\mathscr{M}$ on account of (IV. 8). In other words

$$\{S^{\mu\nu}, p_{\alpha}\}^{*} \approx 0. \tag{IV.9}$$

Lastly, it is straightforward to verify that

$$\{S^{\mu\nu}, S^{\alpha\beta}\}^{*} \approx C^{\mu\nu\alpha\beta}_{\lambda\rho} S^{\lambda\rho} , \tag{IV.10}$$

where

$$C^{\mu\nu\alpha\beta}_{\lambda\rho} = -\eta^{\nu\beta}\delta^{\alpha}{}_{\lambda}\delta^{\mu}{}_{\rho} + \eta^{\mu\beta}\delta^{\alpha}{}_{\lambda}\delta^{\nu}{}_{\rho} - \eta^{\nu\alpha}\delta^{\beta}{}_{\rho} + \eta^{\mu\alpha}\delta^{\nu}{}_{\lambda}\delta^{\beta}{}_{\rho} \tag{IV.11}$$

are the structure constants of the Lorentz group. [The tensor (IV.6) is a projection operator only when $p^{\mu}p_{\mu} + m^2 = 0$, i.e., when the constraint $\mathscr{H} = 0$ holds. Therefore the definition (IV.5) and Eqs. (IV.7–10) hold only weakly. This is suffcient for our purposes.]

If we insert the explicit expression (IV.3) for $J^{\rho\lambda}$ into (IV.5) we find after using the constraint $\mathscr{S} \approx 0$ and $\mathscr{H} \approx 0$:

$$S^{\mu\nu} \approx i\theta^{\mu}\theta^{\nu} + (i/m)\,\theta_5(\theta^{\mu}p^{\nu} - \theta^{\nu}p^{\mu}) . \tag{IV.12}$$

Equation (IV.12) gives the spin for a relativistic massive particle.

Now, it is straightforward to verify that

$$S^{\nu\mu}p_{\nu} \approx 0 \tag{IV.13}$$

and

$$S^{\mu\nu}\theta_{\nu} \approx 0 . \tag{IV.14}$$

As a consequence of Eqs. (IV.13,14) and (IV.12) it follows that we also have

$$S^{\mu\nu}\dot{x}^{\nu}\approx 0\,. \tag{IV.15}$$

It is worthwhile to realize that as a consequence of (III.12a) $\dot{x}^{\nu}$ is not proportional to p^{ν} in a general supergauge. However in spite of that fact $S^{\mu\nu}$ is orthogonal to both $\dot{x}^{\nu}$ and p^{ν}. In the approaches where only commuting variables are used (see Ref. 11 and references therein) conditions (IV.14), (IV.16) cannot be simultaneously enforced and a great deal of controversy has arisen in the past over which one is to be preferred. In our approach both conditions follow simultaneously from the action principle.

B. Massless particles

We want to extend the discussion of the preceding sections to include classical massless spin one-half particles. Since a rest system for such particles does not exist we cannot proceed exactly along the same lines as in the $m\neq 0$ case. So, certain modifications of the procedure are necessary; however, the starting point once again consists of the constraints which [cf. Eqs. (III.8a,b)] now assume the form

$$\mathscr{S}=\theta^{\mu}p_{\mu}\approx 0\,, \tag{IV.16}$$

$$\mathscr{H}=p^2=-p^{0^2}+\mathbf{p}^2\approx 0\,. \tag{IV.17}$$

The spin tensor should be defined through a projection of the total angular momentum which possesses no component along p which is now null. This cannot be done by means of the projection operator (III.9) which is not defined when $m\to 0$. In fact as we shall see below it appears that no nonambiguous projection operators can be constructed.

A natural definition for the subspace complementary to that of p when p is null is achieved by introducing two spacelike vectors $e_{(r)}$ and a additional null vector k_{μ} obeying the relations

$$p\cdot k=1, \tag{IV.18a}$$

$$p\cdot e_{(r)}=0\,, \tag{IV.18b}$$

$$k\cdot e_{(r)}=0\,, \tag{IV.18c}$$

$$e_{(r)}\cdot e_{(s)}=\delta_{rs}\,. \tag{IV.18d}$$

In this basis any arbitrary vector v^{μ} can be decomposed as

$$\begin{aligned} v^{\mu}&=v_p\,p^{\mu}+v_k k^{\mu}+v_r e^{\mu}_{(r)}\\ &=(v\cdot k)\,p^{\mu}+(v\cdot p)\,k^{\mu}+(v\cdot e_{(r)})\,e^{\mu}_{(r)}\,. \end{aligned} \tag{IV.19}$$

Once p is given the vectors k, $e_{(r)}$ are determined only up to a Lorentz transformation which leaves p invariant. Using (IV.20) it is straightforward to find the infinitesimal form of that ambiguity. In fact the most general variation compatible (IV.18) turns out to be (allowing also for a change δp in p):

$$\delta k^{\mu}=-(-k\cdot\delta p)\,k^{\mu}+u^{r}e^{\mu}_{(r)}\,, \tag{IV.20}$$

$$\delta e^{\mu}_{(r)}=-(e_{(r)}\cdot\delta p)\,k^{\mu}-u_r\,p^{\mu}+u_r{}^{s}e^{\mu}_{(s)}\,, \tag{IV.21}$$

with $u^{r}{}_{s}\equiv u_{rs}=-u_{sr}$.

The coefficients u^{r}, u_{rs} in (IV.21) are arbitrary. There are therefore three arbitrary parameters in (IV.20).

Since we are interested in constructing a translational invariant $S^{\mu\nu}$, the projection we are interested in is that which has no component along p, namely that for which p itself has zero projection [see Eq. (IV.7)]. This is given by

$$\tilde{v}^{\mu}=v^{\mu}-v_p\,p^{\mu}=(v\cdot p)\,k^{\mu}+(v\cdot e_{(r)})\,e^{\mu}_{(r)}\,, \tag{IV.22}$$

which may be rewritten as

$$\tilde{v}^{\mu}=\tilde{\mathscr{M}}^{\mu}{}_{\nu}v^{\nu}\,, \tag{IV.23}$$

with

$$\tilde{\mathscr{M}}^{\mu}{}_{\nu}=\delta^{\mu}{}_{\nu}-p_{\nu}\,k^{\mu}\,. \tag{IV.24}$$

The operator (IV.23) is an attractive choice, but unfortunately it suffers from the major drawback of being ambiguous due to the fact that as mentioned above the null vector k is determined by relations (IV.19) only up to the subgroup of the Lorentz group which leaves p invariant, and under these transformations $\tilde{\mathscr{M}}^{\mu}{}_{\nu}$ is *not* invariant, in contrast with what happens in the massive case. The unavailability of a good projection operator means that one cannot define a spin *tensor* $S^{\mu\nu}$ for a massless particle.

There is however a possibility in the massless case which is not present in the massive case. Namely one can define a *scalar* projection of the total angular momentum which is unambiguous, translational invariant and conserved. The scalar in question is the component of $J_{\alpha\beta}$ on the $\mathbf{e}_{(1)},\mathbf{e}_{(2)}$ plane, which we will denote by Σ:

$$\Sigma=J^{\mu\nu}e^{(1)}_{\mu}e^{(2)}_{\nu}\,. \tag{IV.25}$$

The key point is that this projection is invariant under the Lorentz transformations which leave p invariant. This can be seen as follows. On account of (IV.21) we have

$$\begin{aligned} \delta(e^{(1)}e^{(2)})&=e^{(1)}_{\mu}(-u^2p_{\nu}+u^2_1e^{(2)}_{\nu})+e^{(2)}_{\nu}(-u^1p_{\mu}+u^1_2e^{(2)}_{\mu})\\ &=-(e^{(1)}_{\mu}p_{\nu}u^2+e^{(2)}_{\nu}p_{\mu}u^1)+u_{21}(e^{(1)}_{\mu}e^{(1)}_{\nu}-e^{(2)}_{\nu}e^{(2)}_{\mu})\,. \end{aligned} \tag{IV.26}$$

Now, if we contract (IV.26) with $J_{\mu\nu}$ given by (IV.3) we obtain zero on account of the constraints (IV.16,17) and of the properties (IV.18) of the basis. Therfore Σ is unambiguously defined by (IV.25). Furthermore it is evident that Σ is translation invariant since the $e^{(r)}_{\mu}$ are orthogonal to p_{μ}. This means that Σ represents an "intrinsic" property of the particle, much in the same way as the spin tensor $S^{\mu\nu}$ does for the massive case.

The physical meaning of Σ is put in evidence simply by realizing that it represents the projection of the total angular momentum along the direction of motion in any Lorentz frame. Thus, Σ is the helicity of the particle. To see this we start from the relation

$$\epsilon_{\mu\nu\rho\sigma}\,k^{\rho}p^{\sigma}=-\operatorname{sign}(p^0)(e_{(1)\mu}e_{(2)\nu}-e_{(1)\nu}e_{(2)\mu})\,, \tag{IV.27}$$

which follows readily from (IV.18) by considering the frame where $e_{(1)}=(0,0,1,0)$, $e_{(2)}=(0,0,0,1)$. Thus we have the alternative expression for the helicity

$$\Sigma=-\tfrac{1}{2}\operatorname{sign}(p^0)\,\epsilon_{\mu\nu\rho\sigma}\,J^{\mu\nu}k^{\rho}p^{\sigma}\,. \tag{IV.28}$$

Now, since (IV.27) has the same value for any choice of k compatible with (IV.19) we may choose, for example,

$$k^{\mu}=(2|\mathbf{p}|^2)^{-1}(-p^0,\mathbf{p}) \tag{IV.29}$$

without loss of generality. Inserting now (IV.29) into (IV.28) yields

$$\Sigma = -(2|\mathbf{p}|^2)^{-1}\,\mathrm{sign}(p^0)\,\epsilon_{mnro}J^{mn}p^r p^0$$
$$= +\tfrac{1}{2}\epsilon_{mnr}J^{mn}\frac{p^r}{|\mathbf{p}|} = \mathbf{J}\cdot\frac{\mathbf{p}}{|\mathbf{p}|}. \quad \text{(IV.30)}$$

Thus, Σ is indeed the helicity of the particle. An explicity expression for it may be obtained by inserting the form (IV.3) of J into (IV.30). This yields

$$\Sigma = i\epsilon_{ijk}\theta^j\theta^k\frac{p^i}{|\mathbf{p}|}. \quad \text{(IV.31)}$$

A useful expression for Σ may be obtained writing

$$\theta^{(r)} = e^{(r)}_\mu\,\theta^\mu$$

and returning to expression (IV.25). This gives

$$\Sigma = i\theta^{(1)}\theta^{(2)}. \quad \text{(IV.32)}$$

Finally let us determine the vector S^μ for this case. We have according to (IV.3)

$$S_\mu = -\tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}J^{\nu\rho}p^\sigma \quad \text{(IV.33)}$$
$$= -\tfrac{1}{2}i\epsilon_{\mu\nu\rho\sigma}\theta^\nu\theta^\rho p^\sigma. \quad \text{(IV.34)}$$

However we may write

$$\theta^\rho = (\theta\cdot k)\,p^\rho + (\theta\cdot p)k^\rho + \theta^{(r)}e^\rho_{(r)}$$
$$\approx (\theta\cdot k)\,p^\rho + \theta^{(r)}e^\rho_{(r)}. \quad \text{(IV.35)}$$

If we now insert (IV.35) into (IV.34) we find

$$S_\mu = -\tfrac{1}{2}i\epsilon_{\mu\nu\rho\sigma}e^\nu_{(r)}\,e^\rho_{(s)}\,p^\sigma\theta^{(r)}\theta^{(s)},$$

which according to the inverse relation of (IV.27) yields

$$S_\mu = i\theta^{(1)}\theta^{(2)}(k_\mu p_\sigma - k_\sigma p_\mu)\,p^\mu$$
$$\approx (i\theta^{(1)}\theta^{(2)})\,p_\mu.$$

So we have, on account of (IV.32)

$$S_\mu \approx \Sigma p_\mu. \quad \text{(IV.36)}$$

It should be emphasized here that there is no more information on spin contained in the vector S_μ than that carried by the scalar Σ. In particular it is at least misleading to say that (IV.36) shows that for massless particles the spin points point along the momentum. As we have seen before the spatial components S_i *do not* generate rotations around an arbitrary axis and hence cannot be interpreted as the spin of the particle (except in the rest frame which here does not exist!). The spin is given by the spatial components S^{ij} of the antisymmetric spin tensor which cannot be defined for a massless particle. Thus, the only meaningful concept is the component of spin along the direction of motion, Σ. But, again, one cannot infer from it expressions for the component of the spin along any other axis. The latter concept does not exist for a massless particle.

We should emphasized here that the conclusions of this section are, of course, familiar and well-known in the context of relativistic quantum mechanics. The only reason for having gone into such detail in the analysis is to show that the same results can already be reached at the classical level.

V. ELECTROMAGNETIC INTERACTION OF CLASSICAL SPIN ONE HALF PARTICLES

A. Constraints

In the presence of an electromagnetic field the Dirac equation is

$$\gamma^\mu(\partial_\mu - ieA_\mu(x))\,\psi(x) + m\psi(x) = 0, \quad \text{(V.1)}$$

where $A_\mu(x)$ is the vector potential.

Introducing the anticommuting classical variables defined in (III.2a,b) we obtain the following first class constraint

$$\mathscr{S} = \theta^\mu P_\mu + m\theta_5 \approx 0, \quad \text{(V.2)}$$

where

$$P_\mu = p_\mu - eA_\mu. \quad \text{(V.3)}$$

Here p_μ is canonically conjugate to x^μ:

$$\{x^\mu, p_\nu\}^* = \delta^\mu{}_\nu, \quad \text{(V.4a)}$$
$$\{p_\mu, p_\nu\}^* = 0, \quad \text{(V.4b)}$$

while P_μ is not. It obeys, for example

$$\{P_\mu, P_\nu\}^* = eF_{\mu\nu}, \quad \text{(V.5)}$$

as one can easily verify from (V.3). (We keep using the star in the bracket to make contact with our earlier notation. It reminds us that the momenta conjugate to the θ's have been eliminated. No gauge has been fixed, though.) Using (V.5) one can check that the constraint $\mathscr{S}$ satisfies relations (I.27,a–c) with

$$\mathscr{H} = P^\mu P_\mu + m^2 - ieF_{\mu\nu}\theta^\mu\theta^\nu \approx 0. \quad \text{(V.6)}$$

The action principle can now be written as

$$S = \int_{\tau_1}^{\tau_2}\left\{\dot{x}^\mu P_\mu + \frac{i}{2}(\dot{\theta}^\mu\theta_\mu + \dot{\theta}_5\theta_5) + eA_\mu\dot{x}^\mu - [N\mathscr{H} - iM\mathscr{S}]\right. . \quad \text{(V.7)}$$

From a physical point of view the variable P_μ is more relevant than p_μ because the former is invariant under changes of the electromagnetic gauge

$$A_\mu \to A_\mu + \frac{1}{e}\partial_\mu\Lambda, \quad \text{(V.8)}$$

whereas the latter is not.

This point is an interesting one and deserves some elaboration. In the quantum problem one has, under an electromagnetic gauge transformation

$$\psi(x) \to \psi'(x) = e^{i\Lambda(x)}\psi(x), \quad \text{(V.9a)}$$

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\Lambda(x), \quad \text{(V.9b)}$$

$$\partial_\mu \to \partial'_\mu = \partial_\mu. \quad \text{(V.9c)}$$

The primed variables satisfy then an equation identical in form to (V.1).

Now, in order to interpret correctly (V.9a–c) in the classical theory one has to distinguish between the transformation of the states and that of the dynamical variables as it is the latter which has a direct analog in the Hamiltonian mechanics.

To this end we start from (V.9) which say that states are changed according to a unitary operator

$$|\psi'\rangle = U|\psi\rangle,$$

where U is diagonal in the basis where x is diagonal and with matrix elements

$$U(x,x') = e^{i\Lambda(x)}\delta(x,x'). \quad \text{(V.10)}$$

The dynamical variables transform then as

$$A' = UAU^{-1}. \tag{V.11}$$

If we apply (V.11) to the dynamical variables of our theory we find

$$\hat{x}' = \hat{x}, \tag{V.12a}$$

$$\hat{p}' = \frac{1}{i}\frac{\partial}{\partial x} + \frac{\partial \Lambda}{\partial x} = \hat{p} + \frac{\partial \Lambda}{\partial \hat{x}}, \tag{V.12b}$$

$$\hat{\theta}'^{\mu} = \hat{\theta}^{\mu}, \tag{V.12c}$$

$$\hat{\theta}' = \hat{\theta}_5. \tag{V.12d}$$

Equation (V.12a) follows immediately from (V.10) whereas (V.12b) is obtained most easily by using the fact that p' and x' obey canonical commutations relations. Finally (V.12c) and (V.12d) follow simply from the fact that the operator U reduces to the identity in the "internal index" which labels the different components of the spinor ψ.

Equations (V.12a–d) can be applied as they stand to the classical dynamical variables. They have to be supplemented by

$$A_\mu(x) \to A_\mu(x) + \frac{1}{e}\partial_\mu \Lambda, \tag{V.13}$$

in order to leave the Hamiltonian invariant.

Now, it should be stressed that even though we may allow Λ to be a function of x and τ in the classical theory the invariance of the Hamiltonian under (V.12a–d) and (V.13) is not associated with any first class constraint. The reason for this is that on account of (V.13) the transformation does not only alter the dynamical variables but it also explicitly alters the external field $A_\mu(x)$ and therefore it is not a canonical transformation and has no canonical generator. [If the dynamics of the electromagnetic field were included in the action principle (i.e., if we were dealing with a closed system) the transformations (V.11,12) would be a gauge symmetry with an associated first class constraint generator.]

The variable P_μ is the classical analog of the gauge covariant derivative operator

$$\frac{1}{i}\nabla_\mu = \frac{1}{i}\partial_\mu - eA_\mu,$$

which is not changed at all under the transformations (V.9), (V.12a–d) and which obeys the characteristic commutation rules with itself

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)\psi = F_{\mu\nu}\psi,$$

in agreement with (V.5).

The canonical momentum p, on the other hand, is the classical analog of $(1/i)(\partial/\partial x)$ and obeys

$$(\partial_\mu \partial_\nu - \partial_\mu \partial_\nu)\psi = 0,$$

in agreement with (V.4b).

B. Supergauge constraints and the equations of motion

The equation of motion for a dynamical variable $F(\tau)$ is

$$\dot{F}(\tau) = \{F, H\}^*, \tag{V.14}$$

with the Hamiltonian given by

$$H = N(\tau)\mathscr{H} + iM(\tau)\mathscr{S} \approx 0, \tag{V.15}$$

where $N(\tau)$ and $M(\tau)$ are arbitrary functions of the parameter τ. Taking $F = \theta_5$ in (V.14) gives

$$\dot{\theta}_5 = mM(\tau). \tag{V.16}$$

This result tells us that the variable θ_5 can be made to take on any prescribed value during the evolution of the system by appropriately choosing $M(\tau)$. We can use this arbitrariness to choose the supergauge so that

$$\theta_5(\tau) \approx 0. \tag{V.17}$$

[This supergauge choice may of course be also made for the free case and will also be made for a gravitating particle in Sec. VI. One may use Dirac's technique to find the modification of the brackets induced by the gauge condition (V.17). Those brackets may then be used to study the motion of a position operator in quantum mechanics. We refer the reader to Ref. 12 for the treatment of that problem for the free particle. Since we confine our attention here just to the equation of motion we do not need to use the modified brackets.] Now, from (V.16) we see that also here (V.17) implies

$$M(\tau) \approx 0. \tag{V.18}$$

Thus, when working in this gauge we can drop the term M from the Hamiltonian.

Having fixed the supergauge freedom we are still free to fix the parametrization of the world line. The obvious choice is to take the parameter τ as the proper time (in the supergauge $\theta_5 \approx 0$), namely one would like to have

$$\dot{x}^\mu \dot{x}_\mu = -1. \tag{V.19}$$

Condition (V.19) can be attained by an appropriate choice of the function $N(\tau)$, which henceforth remains no longer arbitrary. In fact it follows from the equations of motion for x^μ that

$$\dot{x}^\mu = 2NP^\mu, \tag{V.20}$$

which combined with (V.19) and (V.6) yields

$$N(\tau) = \tfrac{1}{2}(m^2 - eF_{\mu\nu}S^{\mu\nu})^{-1/2}, \tag{V.21}$$

where

$$S^{\mu\nu} = i\theta^\mu \theta^\nu$$

is the spin tensor (III.15).

From the canonical point of view one would like to obtain (V.21) from a gauge condition on the dynamical variables of the theory, much in the same way as (V.17) follows from (V.16). While this is indeed possible in the free case [the gauge condition is $x^0 = (p^0/m)\tau$] there appears to be no simple way to achieve a similar step in the interacting case. We therefore simply choose $N(\tau)$ without fixing a gauge. While this is permissible to examine the equations of motion it has the drawback of not permitting the elimination of any of the canonical variables from the theory.

We observe that (V.21) is well defined since there is no possibility for the quantity under the square root sign in (V.21) to become zero, since $F_{\mu\nu}S^{\mu\nu}$ is in a different part of the algebra from the real number m^2. Thus, one has

$$N(\tau) = \frac{1}{2m}\left(1 - \frac{e}{m}F_{\mu\nu}S^{\mu\nu}\right)^{-1/2} = \frac{1}{2m}\left(1 + \frac{e}{2m^2}F_{\mu\nu}S^{\mu\nu} - \frac{e^2}{8m^4}(F_{\mu\nu}S^{\mu\nu})^2\right)$$

$$\approx \frac{1}{2m}\left(1+\frac{e}{2m^2}F_{\mu\nu}S^{\mu\nu}\right). \tag{V.22}$$

The second equality in (V.22) follows from the fact that a product of more than four θ's vanishes identically whereas the last (weak) equality is a consequence of the constraint $\theta\cdot p\approx 0$ which leaves only three independent θ's. Since N never changes sign we may conclude that the trajectories never bend backwards in time.[13,14]

Using (V.22), Eq. (V.20) assumes the form

$$\dot{x}^{\mu}=\frac{1}{m}\left(1+\frac{e}{2m^2}F_{\mu\nu}S^{\mu\nu}\right)P^{\mu}. \tag{V.23}$$

Finally, assuming that Maxwell equations $\partial^{\mu}F_{\mu\nu}=0$ hold we obtain the following equations of motion for the dynamical variables:

$$\dot{P}^{\alpha}\approx\frac{e}{m}\left(1+\frac{e}{2m^2}F_{\lambda\rho}S^{\lambda\rho}\right)F^{\alpha}{}_{\mu}P^{\mu}+\frac{e}{2m}(\partial^{\alpha}F_{\mu\nu})S^{\mu\nu}, \tag{V.24}$$

$$\dot{\theta}^{\alpha}\approx\frac{e}{m}\left(1+\frac{e}{2m^2}F_{\lambda\rho}S^{\lambda\rho}\right)F^{\alpha}{}_{\mu}\theta^{\mu}, \tag{V.25}$$

$$\dot{S}^{\alpha\lambda}\approx\frac{e}{m}(F_{\mu}{}^{\alpha}S^{\mu\lambda}+F_{\mu}{}^{\lambda}S^{\alpha\mu}), \tag{V.26}$$

$$\dot{S}^{\alpha}\approx\frac{e}{m}F^{\alpha}{}_{\mu}S^{\mu}, \tag{V.27}$$

$$\ddot{x}^{\alpha}\approx\frac{e}{m}\left(1+\frac{e}{2m^2}F_{\lambda\rho}S^{\lambda\rho}\right)F^{\alpha}{}_{\mu}\dot{x}^{\mu}+\frac{e}{2m^2}(\partial^{\alpha}F_{\mu\nu})S^{\mu\nu}+\frac{e}{2m^3}(F_{\mu\nu}S^{\mu\nu})P^{\alpha}. \tag{V.28}$$

We see from (V.24–28) that the presence of the spin introduces alterations in the Lorentz force arising from the coupling of the spin with the electromagnetic field and its first derivatives. The appearance of electromagnetic field gradients is an indication that in a physical sense the particle is "extended." On the other hand the equation of motion for spin, Eq. (V.26), is just the equation of motion for a particle of gyromagnetic ratio

$$g=2.$$

In fact, in the rest frame $\dot{x}^{i}=0$ we have

$$\frac{ds^{i}}{dt}=\frac{e}{m}F^{i}{}_{j}S^{j},$$

which means that the magnetic moment is

$$\boldsymbol{\mu}=2\left(\frac{e}{2m}\mathbf{S}\right),$$

so the gyromagnetic ratio is equal to two (see Ref. 8).

VI. GRAVITATIONAL INTERACTION OF CLASSICAL SPIN ONE HALF PARTICLES

A. Constraints

We deal in this section with the motion of a spinning test particle in a gravitational field. This problem has been analyzed by several authors in terms of commuting variables for the spin. Our analysis will treat the basic spin degrees of freedom as anticommuting. The system in consideration will therefore be a microscopic (spin one-half) particle and the interaction will be totally fixed by the requirement of gauge supersymmetry invariance. There will be no arbitrary elements (such as nonequivalent choices of subsidiary conditions for the spin) in the dynamics, in contrast with earlier treatments.

The treatment is quite analogous to the one given in the previous section for the electromagnetic coupling. We start from the Dirac equation in a gravitational field (See Appendix 2 for conventions.):

$$\gamma^{\mu}(\partial_{\mu}-\tfrac{1}{4}\omega_{\mu AB}\gamma^{A}\gamma^{B})\psi(x)+m\psi(x)=0. \tag{VI.1}$$

The classical first class constraint associated with (VI.1) is

$$\mathcal{S}=\theta^{\mu}\mathcal{P}_{\mu}+m\theta_{5}\approx 0, \tag{VI.2}$$

where

$$\mathcal{P}_{\mu}=p_{\mu}-\frac{i}{2}\omega_{\mu AB}\theta^{A}\theta^{B}. \tag{VI.3}$$

The analog in this case of the electromagnetic gauge freedom is the possibility of rotating arbitrarily the tetrad field:

$$L^{\mu}{}_{A}(x)\rightarrow\epsilon_{A}{}^{B}(x)\,L^{\mu}{}_{B}(x), \tag{VI.4}$$

with $\epsilon_{AB}=-\epsilon_{BA}$.

Under (VI.4) the connection changes as

$$\omega_{\mu AB}\rightarrow\omega_{\mu AB}+\nabla_{\mu}\epsilon_{AB} \tag{VI.5}$$

with

$$\nabla_{\mu}\epsilon_{AB}=\partial_{\mu}\epsilon_{AB}-\omega_{\mu A}{}^{C}\epsilon_{CB}-\omega_{\mu B}{}^{C}\epsilon_{AC}.$$

The vector $\mathcal{P}_{\mu}$ is invariant under a local rotation of the frame, whereas the canonical momentum p_{μ} changes as

$$p_{\mu}\rightarrow p_{\mu}+\frac{i}{2}\nabla_{\mu}\epsilon_{AB}. \tag{VI.6}$$

The odd dynamical variables θ^{A} also rotate as

$$\theta^{A}\rightarrow\theta^{A}+\epsilon^{A}{}_{B}\theta^{B}.$$

On the other hand x^{μ},θ_{5} and

$$\theta^{\mu}\equiv L_{A}{}^{\mu}(x)\theta^{A}$$

are invariant under rotations of the local tetrad.

Equations (VI.5) and (VI.6) are the analogs of (V.13) and (V.12b). Since the gravitational field is a given background, there is no conservation law or constraint associated with the invariance under local tetrad rotations, exactly as it occurred for the electromagnetic case.

The only nonzero brackets among the basic canonical variables are

$$\{x^{\mu},p_{\nu}\}^{*}=\delta^{\mu}{}_{\nu}, \tag{VI.7a}$$

$$\{\theta^{A},\theta^{B}\}^{*}=i\eta^{AB}, \tag{VI.7b}$$

$$\{\theta_{5},\theta_{5}\}^{*}=i. \tag{VI.7c}$$

Other useful brackets are

$$\{\theta^{\mu},\mathcal{P}_{\alpha}\}^{*}=-\Gamma^{\mu}_{\alpha\beta}\theta^{\beta}, \tag{VI.8}$$

$$\{\theta^{C},\mathcal{P}_{\alpha}\}^{*}=\omega_{\alpha}{}^{C}{}_{A}\theta^{A}, \tag{VI.9}$$

$$\{\mathcal{P}_{\alpha},\mathcal{P}_{\beta}\}^{*}=\frac{i}{2}R_{\alpha\beta\mu\nu}\theta^{\mu}\theta^{\nu}. \tag{VI.10}$$

The dynamical variable $\mathcal{P}_{\alpha}$ is the classical analog of the gravitational covariant derivative

$$\frac{1}{i}\nabla_\mu\psi = (\partial_\mu - \tfrac{1}{4}\omega_{\mu AB}[\gamma^A, \gamma^B])\psi,$$

of a Dirac Spinor ψ, and Eq. (VI.10) corresponds to the basic property (definition) of the curvature tensor

$$(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)\psi = \tfrac{1}{8}R_{\mu\nu AB}[\gamma^A, \gamma^B]\psi.$$

A new first class constraint is obtained by taking the bracket of (VI.2) with itself, which yields

$$\{\mathscr{S}, \mathscr{S}\}^* = i\mathscr{H},$$

with

$$\mathscr{H} = g^{\mu\nu}\mathscr{P}_\mu\mathscr{P}_\nu + m^2 \approx 0. \tag{VI.11}$$

The action principle is now

$$S = \int_{\tau_1}^{\tau_2} d\tau\left\{\dot{x}^\mu\mathscr{P}_\mu + \frac{i}{2}(\dot{\theta}^A\theta_A + \dot{\theta}_5\theta_5) + \frac{i}{2}\omega_{\mu AB}\theta^A\theta^B\dot{x}^\mu - N\mathscr{H} - iM\mathscr{S}\right\}. \tag{VI.12}$$

It is interesting to realize that the analog of the term

$$F_{\mu\nu}\theta^\mu\theta^\nu,$$

which appeared in the electromagnetic case is absent from (VI.11). This is due to the fact that the contraction

$$R_{\mu\nu AB}\theta^A\theta^B\theta^\mu\theta^\nu = (\epsilon^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu})\theta^0\theta^1\theta^2\theta^3 \tag{VI.13}$$

is identically zero.

Now, if one applies twice the Dirac operator appearing in (VI.1) one finds that (VI.1) implies

$$(\nabla^\mu\nabla_\mu + m^2 + \tfrac{1}{4}R)\psi = 0. \tag{VI.14}$$

This equation differs by the $\frac{1}{4}R$ term from what one would obtain by simply replacing $\mathscr{P}_\mu$ by $-i\nabla_\mu$ in $\mathscr{H}$ given by (VI.12) and demanding $\mathscr{H}\psi = 0$. The origin of the scalar curvature term in (IV.14) can be understood by recalling that classically one has

$$\theta^\mu\theta^\nu + \theta^\nu\theta^\mu = 0$$

so that (VI.13) holds, whereas quantum mechanically this equation is replaced by

$$\hat{\theta}^\mu\hat{\theta}^\nu + \hat{\theta}^\nu\hat{\theta}^\mu = \hbar g^{\mu\nu}$$

so that (V.13) is no longer valid in the latter case. Thus, the $\frac{1}{4}R$ term in (VI.14) can be traced to a "factor ordering" problem which one faces only upon quantization and which is not present classically.

The correct quantum equation is of course (VI.14) with the curvature term included as it follows directly from the Dirac equation. However, in the classical limit one should demand that the Compton wave length $\hbar/mc$ be much smaller than the radius of curvature $1/\sqrt{|R|}$ of the background. This amounts to neglecting R in front of m^2 in (VI.14) in that limit, in agreement with our earlier reasoning.

B. Supergauge fixation and equations of motion

1. Nonzero mass particle

The reasoning given for the electromagnetic coupling may be literally taken over to the present case. Equations (V.14)–(V.20) remain valid, and in place of (V.21) we obtain simply

$$N = \frac{1}{2m}. \tag{VI.15}$$

When (V.15) is used the equations of motion read

$$\frac{D\theta^\alpha}{D\tau} = 0, \tag{VI.16}$$

$$\frac{DS^{\alpha\beta}}{D\tau} = 0, \tag{VI.17}$$

$$\frac{D\mathscr{P}^\alpha}{D\tau} = \frac{1}{2m}R^\alpha{}_{B\mu\nu}\mathscr{P}^\beta S^{\mu\nu}, \tag{VI.18}$$

$$\frac{D^2x^\mu}{D\tau^2} = \frac{1}{m}R^\mu{}_{\lambda\alpha\beta}\dot{x}^\lambda S^{\alpha\beta}. \tag{VI.19}$$

The result that a classical massive spinning particle does not follow a geodesic in spacetime due to the coupling of the spin tensor to the curvature, Eq. (VI.19), is not a new one. In fact this problem has been studied by Papapetrou[2] and Eqs. (VI.17–19) are formally similar to the equations obtained by him. However a key element in Papapetrou's case is the choice of a subsidiary condition on the spin tensor to obtain a system of equations capable of predicting the future from a specification of initial conditions. In his approach different choices of subsidiary conditions give rise to physically different equations of motion (which have indeed been proposed by other authors[3,4]). In the present analysis however the theory is fully determined by the requirement of gauge supersymmetry invariance and the only ambiguity is that inherent in the choice of the supergauge, which is by definition physically harmless (much as the choice of the parametrization of the worldline).

It is also interesting to anticipate here a result of the next section where we shall see that for a massless particle there exists a supergauge where the worldline is a geodesic. One may therefore wonder whether one could also devise, for $m \neq 0$, a gauge choice (other than $\theta_5 \approx 0$) which would lead to the same conclusion. Whereas we have at the moment of this writing no proof that this cannot be done, there is no resonable condidate for that gauge condition and we believe that there is no supergauge where the particle's worldline will be a geodesic.

2. Zero rest-mass particle

The preceding treatment applies only when the rest mass of the test particle is not zero. While we did not discuss electromagnetic couplings of zero mass particles as this seems to be a totally academic problem, given that charged particles with zero rest mass do not appear to exist, it is of interest to consider gravitational couplings of massless spin $\frac{1}{2}$ particles. The analysis should apply, for example, to neutrinos.

For $m = 0$ the constraints (VI.2) and (VI.11) reduce to

$$\mathscr{S} = \theta^\mu\mathscr{P}_\mu \approx 0, \tag{VI.20}$$

$$\mathscr{H} = g^{\alpha\beta}\mathscr{P}_\alpha\mathscr{P}_\beta \approx 0, \tag{VI.21}$$

and the Hamiltonian is

$$H = N(\tau)g^{\alpha\beta}\mathscr{P}_\alpha\mathscr{P}_\beta + iM(\tau)\theta^\mu\mathscr{P}_\mu \approx 0. \tag{VI.22}$$

We see that the variable θ_5 drops out from the formalism and the supergauge cannot be fixed by setting $\theta_5 \approx 0$. There is however a natural supergauge choice in this case which is implemented in terms of the basis defined in

(IV.19a–d) with the p_μ of flat space replaced by $\mathscr{P}^\mu$. We fix the gauge by setting

$$\theta^p = \theta \cdot k \approx 0 . \tag{VI.23}$$

The gauge choice (VI.23) is not suitable for $m \neq 0$ since the vector k is not available in that case.

It should be noticed that since k is fixed only up to a Lorentz transformation which leaves $\mathscr{P}^\mu$ invariant the gauge condition (VI.23) is ambiguous. We will show that this ambiguity may be exploited to select a particular evolution for k so that the corresponding worldline is remarkably simple: it becomes a geodesic. Therefore the main result of this section will be that there exists a supergauge where a massless spin one-half particle follows a geodesic. This conclusion is in agreement with the results of Mashhoon[15] who studied the motion of massless spinning particles in a gravitational field along the lines of Refs. 2–4. It does not seem possible to achieve a similar result when $m \neq 0$ as we discussed in the preceding section. [Incidentally, another consequence of the ambiguity (IV.20) in the definition of k is to make the derivatives $\partial k / \partial p$ not well defined. Therefore condition (VI.23) cannot be used straightforwardly to pass to a new bracket where it holds strongly. This is relevant to the problem of defining quantum mechanical position operators for a massless particle, an issue that we plan to discuss elsewhere.]

The analysis proceeds as follows. Before fixing a gauge the equations of motion read

$$\frac{dx^\mu}{d\tau} = 2N\mathscr{P}^\mu + iM\theta^\mu, \tag{VI.24}$$

$$\frac{D\theta^\mu}{D\tau} = M\mathscr{P}^\mu , \tag{VI.25}$$

$$\frac{D\mathscr{P}^\mu}{D\tau} = iR^\mu{}_{\beta\rho\sigma}\,\theta^\rho\theta^\sigma(2N\mathscr{P}^\beta + iM\theta^\beta) , \tag{VI.26}$$

where $D/D\tau$ is the covariant derivative along the worldline. After imposing (VI.23) there are only two nonzero components of θ^μ left:

$$\theta_{(r)} = \theta_\mu e^\mu{}_{(r)} , \quad r = 1,2. \tag{VI.27}$$

In order to obtain their equations of motion we recall Eqs. (IV.20,21) which give the general variation of k^μ and $e^\mu{}_{(r)}$. Applied to the change along the worldline of the particle they give

$$\frac{Dk^\mu}{D\tau} = -\left(k \cdot \frac{D\mathscr{P}}{D\tau}\right) k^\mu + u^r e^\mu{}_{(r)} , \tag{VI.28}$$

$$\frac{De^\mu{}_{(r)}}{D\tau} = -\left(e_{(r)} \cdot \frac{D\mathscr{P}}{D\tau}\right) k^\mu - u_r\,\mathscr{P}^\mu + u_r{}^s e^\mu{}_{(s)} , \tag{VI.29}$$

with u^r, $u_{rs} = -u_{sr}$ arbitrary. It follows from (IV.27) and (VI.29) that

$$\frac{d\theta_{(r)}}{d\tau} = u_r{}^s\theta_{(s)} , \tag{IV.30}$$

which is particular implies that the helicity

$$\Sigma = i\theta_{(1)}\theta_{(2)}$$

is a constant of motion,

$$\frac{d\Sigma}{d\tau} = 0 . \tag{VI.31}$$

In order to examine the equation of motion for x^μ we need to know the value of $M(\tau)$, which is obtained by demanding that the gauge condition (VI.23) be preserved in time:

$$0 = \frac{D}{D\tau}(\theta_\mu k^\mu) = M + u^r\theta_{(r)} .$$

So, we find

$$M = -u^r\theta_{(r)} . \tag{VI.32}$$

We will also need

$$\dot{M} = -(\dot{u}^s + u^r u_r{}^s)\,\theta_{(s)} \equiv v^s\theta_{(s)} . \tag{VI.33}$$

Now, covariantly differentiating (VI.24) with respect to time and using (VI.25) and (VI.26) we find

$$\frac{D^2x^\mu}{D\tau^2} = 2\dot{N}\mathscr{P}^\mu + iM\theta^\mu + 2NiR^\mu{}_{\beta\rho\sigma}\theta^\rho\theta^\sigma(2N\mathscr{P}^\beta + iM\theta^\beta) ,$$

which taking into account (VI.32) and (VI.33) simplifies to

$$\frac{D^2x^\mu}{D\tau^2} = 2\dot{N}^\mu + i(v^s e^\mu{}_{(r)} + 4N^2R^{\mu krs})\,\theta_{(r)}\theta_{(s)} . \tag{VI.34}$$

Now, we have

$$R^{\mu krs} = R^{pkrs}\mathscr{P}^\mu + R^{mkrs}e^\mu{}_{(m)} ,$$

so that we obtain

$$\frac{D^2x^\mu}{D\tau^2} = (2\dot{N} + 4N^2iR^{pkrs}\theta_{(r)}\theta_{(s)})\,\mathscr{P}^\mu + i(v^s\delta^{mr} + 4N^2R^{mkrs})\,\theta_{(r)}\theta_{(s)}\,e^\mu{}_{(m)} . \tag{VI.35}$$

The second term in the right side of (VI.35) vanishes if the arbitrary functions v^s are chosen as

$$v_s = -8N^2\epsilon_{sr}R^{rk12} , \tag{VI.36}$$

where ϵ_{sr} is the two-dimensional alternating tensor defined by $\epsilon_{rs} = -\epsilon_{sr}$, $\epsilon_{12} = 1$.

After v_s are fixed by (VI.36) one fixes N by demanding

$$\dot{N} + 2N^2R^{pkrs}i\theta_{(r)}\theta_{(s)} = 0$$

or

$$\dot{N} + 2N^2R^{pk12}\Sigma = 0 , \tag{VI.37}$$

which yields

$$N(\tau) = N_0\left(1 - 4N_0\Sigma\int_0^\tau R^{pk12}(\bar{\tau})\,d\bar{\tau}\right), \tag{VI.38}$$

where N_0 is an arbitrary constant. [In order for the parameterization to be nondegenerate N must have a nonzero component along unity in the Grassmann algebra. According to (VI.38) it is sufficient that N_0 satisfies that demand in order for $N(\tau)$ to be acceptable. One may take for example $N_0 = 1$.]

It is interesting to notice that on account of (VI.33) and (VI.35) the function $M(\tau)$ is not zero in this gauge. In fact from (VI.33) we have

$$M(\tau) = \int_0^\tau v^s\theta_{(s)} = \int_0^\tau -N^2\,\epsilon_{sr}\theta^{(s)}R^{rk12}\,d\bar{\tau} . \tag{IV.39}$$

This last relation shows in particular that $M(\tau)$ is unaltered by a time-dependent rotation of the spatial vectors $e_{(r)}$, the orientation of which remains therefore open. [The freedom in u^r has however been fixed by (VI.35).]

With N,M given by (VI.38) and (VI.39) we have

$$\frac{D^2x^\mu}{D\tau^2}=0, \tag{VI.40}$$

the geodesic equation. [Observing Eqs. (VI.31) and (VI.40) one cannot help feeling as if he were working on general relativity before the advent of tensor calculus. In fact, one should be able to develop a proper geometrical formulation by means of a "super-covariant derivative" in the (x,θ) "superspace" in which those equations could be merged into a single geometrical statement which would make no reference to a particular supergauge.]

Lastly it is interesting to write down the equations for the propagation of the tetrad $\mathscr{P}^\mu, k^\mu, e^\mu{}_{(r)}$. They are readily obtained from (IV.24,28,29) and read as follows:

$$\frac{D\mathscr{P}^\mu}{D\tau}=\frac{\dot N}{N}\mathscr{P}^\mu-i\frac{\dot M}{2N}\theta^{(r)}e^\mu{}_{(r)}\,. \tag{VI.41}$$

$$\frac{Dk^\mu}{D\tau}=-\frac{\dot N}{N}k^\mu+u^r e^\mu{}_{(r)}\,, \tag{VI.42}$$

$$\frac{De^\mu{}_{(r)}}{D\tau}=i\frac{\dot M}{2N}\theta_{(r)}k^\mu-u_r\,\mathscr{P}^\mu+u_r{}^s e^\mu{}_{(s)}\,, \tag{VI.43}$$

with N, M, u^r given by (VI.38, 39, 32) and $u_r{}^s$ arbitrary.

We see that the local tetrad is not parallelly transported. There is in effect no way to fix M and N so that $\mathscr{P}_\mu$ will be parallelly transported, as one can easily see from (VI.26).

ACKNOWLEDGMENTS

This article is based in part (especially Secs. II, III, and Appendix 1) on notes of a course on constrained Hamiltonian systems given by one of us (C.T.) at Princeton University in the fall of 1976. We are grateful to C. Aragao de Carvalho and L. Yaffe for making available to us their lecture notes. C.A.G. wishes to express his acknowledgments for the kind hospitality extended to him by the Department of Physics of Princeton University where his part in this work has been done. He also wishes to thank Drs. J.J. Giambiagi, L.C. Gomes, and A.C. Olinto from CBPF/CNPq for their constant encouragement.

APPENDIX 1: ANTICOMMUTING CLASSICAL VARIABLES

In what follows we will merely describe the formal properties of classical anticommuting variables (odd variables) and the canonical formulation of its dynamics. More details can be found in the existing literature.[7,8,17]

The following properties of even variables (commuting c-numbers) and odd variables (anticommuting c-numbers) are dictated by the idea that these variables are the classical limit of Bosonic and Fermionic operators, respectively,

$$\theta^\alpha\theta^\beta+\theta^\beta\theta^\alpha=0 \quad \text{for } \theta^\alpha,\ \theta^\beta \text{ odd}, \tag{1.1a}$$

$$\theta^\alpha q^i-q^i\theta^\alpha=0 \quad \text{for } \theta^\alpha \text{ odd},\ q^i \text{ even}, \tag{1.1b}$$

$$q^iq^j-q^jq^i=0 \quad \text{for } q^i,\ q^j \text{ even}. \tag{1.1c}$$

A function of a set of odd variables can be defined by formal expansion,

$$f(\theta^\alpha)=f_0+f_\alpha\theta^\alpha+f_{\alpha\beta}\theta^\alpha\theta^\beta+\cdots. \tag{1.2}$$

The series necessarily terminates if the θ^α are finite in number. Derivatives are defined by

$$\delta f=\delta\theta^\alpha\frac{\partial f}{\partial\theta^\alpha}, \tag{1.3}$$

with $\delta\theta^\alpha$ on the left (left derivative).

Another useful concept is that of complex conjugation (involution) which will be the classical analogue of Hermitian conjugation for operators. It has the properties

$$(AB)^*=B^*A^*, \tag{1.4a}$$

$$(A^*)^*=A. \tag{1.4b}$$

A variable is called real if $A^*=A$ and purely imaginary if $A^*=-A$. Note that for odd variables $\theta^\alpha\theta^\beta$ will be purely imaginary if θ^α and θ^β are real.

Canonical formalism in the presence of odd variables

The equations of motion for a given system are assumed to follow by formally extremizing an action functional of the form

$$S=\int_{\tau_1}^{\tau_2}L(q^i,q^i,\theta^\alpha,\dot\theta^\alpha)\,d\tau\,. \tag{1.5}$$

The Lagrangian L must be taken to be even so that the evenness or oddness of a variable is preserved during the dynamical evolution.

The canonical momenta are defined as usual by

$$p_i=\frac{\partial L}{\partial\dot q^i} \tag{1.6}$$

$$=\frac{\partial L}{\partial\dot\theta^\alpha} \tag{1.7}$$

and the canonical Hamiltonian is

$$H=\dot q^ip_i+\dot\theta^\alpha\pi_\alpha-L\,. \tag{1.8}$$

The θ^α have been put in the left of π_α in (1.8) because, on account of our convention (1.3) for the derivative, only in this way we get that H is a function of the coordinate and momenta:

$$\delta H=q^i\delta p_i+\dot\theta^\alpha\delta\pi_\alpha-\delta qi\frac{\partial L}{\partial q_i}-\delta\theta^\alpha\frac{\partial L}{\partial\theta^\alpha}. \tag{1.9}$$

From this point on one can proceed formally in the same way as in the usual canonical theory for even variables.

There are however two new features which appear when odd variables are introduced in the theory, namely the generalization of the Poisson bracket and the fact that, as a consequence its properties, the second class constraints, may be now odd in number.

If we carry out the extremization of the action functional

$$S=\int(\dot q^ip_i+\dot\theta^\alpha\pi_\alpha-H)\,d\tau\,, \tag{1.10}$$

we find the following equations of motion:

$$qi=\frac{\partial H}{\partial p_i},\quad p_i=-\frac{\partial H}{\partial\dot q}, \tag{1.11}$$

$$\dot\theta^\alpha=-\frac{\partial H}{\partial\pi_\alpha},\quad \dot\pi_\alpha=-\frac{\partial H}{\partial\dot\theta^\alpha}. \tag{1.12}$$

Demanding now that

$$\dot{F} = \{F,H\}, \tag{1.13}$$

one can show that the Poisson bracket in Eq. (1.13) has the general form

$$\{F,G\} = \left(\frac{\partial F}{\partial q^i}\frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial G}{\partial q^i}\right) + \epsilon_F\left(\frac{\partial F}{\partial \theta^\alpha}\frac{\partial G}{\partial \pi_\alpha} + \frac{\partial F}{\partial \pi_\alpha}\frac{\partial G}{\partial \theta^\alpha}\right), \tag{1.14}$$

where ϵ_F is $+1$ if F is even and -1 otherwise. The bracket (1.15) obeys generalized versions of Jacobi identity which take different forms depending on the even or odd character of the functions involved.

The Dirac bracket is defined by the same formula as in the ordinary case, namely

$$\{F,G\}^* = \{F,G\} - \{F,\chi_m\}\, C^{mn}\{\chi_m,G\} \tag{1.15}$$

and it may be shown that it has the same algebraic properties as the (generalized) Poisson bracket.

If the constraints χ_m are second class the inverse matrix C^{mn} exists in the generalized case. However it is important to realize that if some of the χ_m are odd functions the existence of C^{mn} does not imply that the χ_m are even in number. As an immediate consequence of this fact the dimensionality of the physical phase space may be odd.

APPENDIX 2: SUMMARY OF NOTATIONS AND CONVENTIONS FOR RIEMANN SPACETIME AND THE GENERAL RELATIVISTIC DIRAC EQUATION

We consider a Riemann manifold with metric tensor g of hyperbolic signature and local Minkowski structure. The tetrad field components $L^{\mu}{}_{(A)}(x), \mu,A = 0,1,2,3$, are defined by

$$L^{\mu}{}_{(A)}L^{\nu}{}_{(B)}\, g_{\mu\nu} = \eta_{AB} = \mathrm{diag}(-1,+1,+1,+1), \tag{2.1a}$$

$$L_{\mu}{}^{(A)}L_{\nu}{}^{(B)}\,\eta_{AB} = g_{\mu\nu}\,. \tag{2.1b}$$

We denote by $\Gamma^{\alpha}_{\mu\nu}$ the Christoffel symbols associated with the metric tensor $g_{\mu\nu}(x)$. The covariant derivative of a vector field $V_\lambda(x)$ is defined by

$$V_\lambda(x)_{\|\mu} = \partial_\mu V_\lambda - \Gamma^{\alpha}_{\mu\nu}V_\alpha \tag{2.2}$$

and the Riemann tensor $R^{\alpha}{}_{\beta\mu\nu}$ is defined by

$$V_{\beta\|\mu\|\nu} - V_{\beta\|\nu\|\mu} = R^{\alpha}{}_{\beta\mu\nu}\,V_\alpha\,. \tag{2.3}$$

The constant Dirac matrices γ^A constitute a representation of Clifford algebra associated with the local Minkowski metric satisfying the anticommutation relation

$$[\gamma^A,\gamma^B]_+ = \gamma^A\gamma^B + \gamma^B\gamma^A = 2\eta^{AB}\,. \tag{2.4}$$

A representation for the Clifford algebra associated with the spacetime metric $g_{\mu\nu}(x)$ can be obtained from (2.4) with the help of the tetrad vectors,

$$[\gamma^\mu(x),\gamma^\nu(x)]_+ = 2g^{\mu\nu}(x)\,, \tag{2.5}$$

with

$$\gamma^\mu(x) = L^{\mu}{}_{(A)}(x)\,\gamma^A\,. \tag{2.6}$$

The extension of the special relativistic Dirac equation

$$\gamma^\mu\partial_\mu\,\psi(x) + m\psi(x) = 0 \tag{2.7}$$

to general relativity is obtained (see, for example, Refs. 18 and 19) by means of the minimal coupling to the gravitational field in the sense that we make the substitution $\partial_\mu \to \nabla_\mu$, where ∇_μ is the operator of covariant derivative for the spinor field $\psi(x)$, defined by

$$\nabla_\mu = \partial_\mu - \Gamma_\mu\,,$$

where $\Gamma_\mu(x)$ are the Ricci rotation coefficients (or spin connections) given by

$$\Gamma_\mu = -\tfrac{1}{8}L^{\nu}{}_{(A)}L_{(B)\|\nu}(\gamma^A\gamma^B - \gamma^B\gamma^A) = -\tfrac{1}{4}\omega_{\mu AB}\gamma^A\gamma^B\,. \tag{2.8}$$

Thus, the general relativistic Dirac equation is

$$\gamma^\mu(\partial_\mu + \tfrac{1}{4}\omega_{\mu AB}\gamma^A\gamma^B)\,\psi(x) + m\psi(x) = 0\,. \tag{2.9}$$

Note added in proof: While this paper was in press several colleagues kindly sent us copies of their interesting papers on the classical theory of spinning particles (Refs. 20–25). As we stated in the introduction, we make no pretense whatsoever of reviewing here the extensive literature in this field and we apologize for the many references that we must undoubtedly have left out of our bibliography. Also, since this paper was written, an approach to the quantization of gravity and supergravity based on their analogy with the theory of the point particle has been developed by one of us (Refs. 26, 27). The boundary term (II.26) plays a crucial role in that approach.

[1] C. Teitelboim, in "Current Trends in the Theory of Fields," American Institute of Physics, New York, 1978.
[2] A. Papapetrou, Proc. R. Soc. London Ser. A **209**, 48 (1951).
[3] W. Tulczyjew, Acta Phys. Pol. **15**, 389 (1956).
[4] L.I. Schiff, Proc. Nat. Acad. Sci. **46**, 87 (1960).
[5] For a recent review, see J. Scherk and P. van Nieuwenhuizen, Phys. Rep., in press.
[6] C. Teitelboim, "The Semiclassical Approximation for Fermions: A Progress Report," in Proceedings of the First Chilean Symposium on Theoretical Physics held in Santiago, Chile, 1978 (Editorial de la Universidad Tecnica del Estado, Santiago, Chile, 1980).
[7] F.A. Berezin and M.S. Marinov, Ann. Phys. (N.Y.) **104**, 336 (1977).
[8] R. Casalbuoni, Nuovo Cimento A **33**, 115 (1976); **33**, 289 (1976); A. Barducci, R. Casalbuoni, and L. Lusana, "Supersymmetry and the Pseudoclassical Relativistic Electron," preprint, Firenze (1976).
[9] See, for example, E.S. Fradkin and M.S. Vasiliev, "*S* matrix for theories that admit closure of the algebra with the aid of auxiliary fields: The auxiliary fields in supergravity," Princeton preprint, 1978, and references therein.
[10] E.S. Fradkin and T.E. Fradkina, Phys. Lett. B **72**, 343 (1978).
[11] A.J. Hanson and T. Regge, Ann. Phys. (N.Y.) **87**, 498 (1971).
[12] C. Bachas, Phys. Lett. B **29**, 62 (1978).
[13] E.C.G. Stueckelberg, Hel. Phys. Acta **14**, 588 (1941).
[14] S. Hojman and T. Regge, in *Studies in Mathematical Physics: Essays in Honour of Valentine Bargmann*, edited by E.H. Lieb and B. Simon (Princeton U.P., Princeton, N.J., 1976).
[15] B. Mashhoon, Ann. Phys. (N.Y.) **89**, 254 (1975).
[16] V. Bargmann, L. Michel, and V.L. Telegdi, Phys. Rev. Lett. **2**, 435 (1959).
[17] F.A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[18] D.R. Brill and J.A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).
[19] A. Lichnerowicz, in *Relativity Groups and Topology*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, London, 1963).

[20]A. Barducci, R. Casalbuoni, and L. Lusanne, Nucl. Phys. B **124**, 93, 521 (1972); Phys. Lett. B **64**, 319 (1976).
[21]A. P. Balachandran, G. Marmo, B. S. Shagerstam, and A. Stern, Phys. Lett. B **89**, 199 (1980); Nucl. Phys. B **162**, 385 (1980).
[22]A. P. Balachandran, Per Salomonson, B. S. Shagerstam, and J. O. Winnberg, Phys. Rev. D **15**, 2308 (1977).
[23]L. Brink, R. Marnelius, and P. Salomonson, "A superspace formulation of the spinning particle or supergravity in one space–time dimension: A revisit," Göteborg preprint 77, 31 (1977).
[24]R. Marnelius and B. Nilsson, "Manifestly conformally covariant field equations and a possible origin of the Higgs mechanism," Göteborg preprint 75, 72 (1979).
[25]F. Ravndal, "Supersymmetric Dirac particles in external fields," University of Oslo Report 79, 19 (1979).
[26]C. Teitelboim, "Proper time approach to quantum supergravity," Institute for Advanced Study preprint, May 1980.
[27]C. Teitelboim, "Proper time approach to the quantization of the gravitational field," Institute for Advanced Study preprint, March 1980.

Perturbation expansion of multichannel Regge poles

S. Bosanac[a)]
R. Bošković Institute, 41001 Zagreb, Croatia, Yugoslavia

(Received 19 June 1979; accepted for publication 21 August 1979)

A perturbation method for the complex angular momentum poles of the multichannel S-matrix is developed. The inelastic residues of the S-matrix are also treated in a similar fashion. It is shown that the poles and residues are the second order in coupling between the channels. The closed channel Regge poles are also discussed and it is shown that their contribution in the perturbation series is finite, regardless of the asymptotic behavior of potential.

1. INTRODUCTION

In the Regge theory for the elastic scattering amplitude, the essential idea is to replace the partial wave sum by an integral in the complex angular momentum plane.[1,2] After suitable transformations, the final form of the scattering amplitude is given in two parts: (a) the sum over the complex angular momentum poles of the S matrix and (b) an integral along the imaginary angular momentum axis. The original derivation of Regge has been modified to take into account when the hard core is present, e.g., atom–atom potential.[3] It has been found that relatively few Regge poles contribute to the scattering amplitude, compared to the large number of partial waves involved in the original sum. However, the difficult part is to calculate the Regge poles and the appropriate residues of the S matrix, but a suitable method can be found to overcome this.[4]

Regge theory also gives a simple description of the scattering events in terms of orbiting and surface waves.[5] Each Regge pole represents either an orbiting state or a surface wave and gives a characteristic contribution to the scattering amplitude.

It is therefore tempting to use the Regge theory in the inelastic collisions. However, as in the elastic collision case, we must have a method for calculating the Regge poles and the appropriate residues. In this work we propose a perturbation method for solving this task. Instead of calculating the exact poles, we decouple the set of equations and define the Regge poles and residues for the uncoupled channels. When the coupling is "switched on" these poles move to another place in the angular momentum plane and the shift is determined by the magnitude of coupling. Similarly, the residues change their values.

The article is divided into two parts. In the first part we describe the perturbation method for the Regge poles and residues in the elastic collisions.

In the second part the theory is generalized for the inelastic collisions. It is shown that the shift of the elastic Regge poles and residues is second order in the coupling, which indicates that the inelastic scattering amplitude is well described by the unperturbed poles. Hence, not all the Regge poles in the uncoupled channels contribute equally to the scattering amplitude. Channels which have poles close to the real axis dominate the scattering amplitude, and therefore all the poles with the large imaginary part can be neglected in the first approximation. This conclusion is reached from the discussion of the elastic scattering amplitude.[3,5]

The perturbation method, described here, has been developed for the poles of the Green's function by Titchmarsh.[6] However, we give two new points: (a) perturbation expansion for the residues and (b) extension of the theory to the multichannel equations.

2. FORMAL DERIVATION OF PERTURBATION SERIES FOR ELASTIC SCATTERING

The radial Schrödinger equation for a perturbed system is

$$\varphi'' = \left[V_0 + \epsilon V' + (\lambda^2 - \tfrac{1}{4})/r^2 - k^2 \right] \varphi , \tag{2.1}$$

where $\epsilon V'$ is a perturbation. The regular solution has the form of a linear combination

$$\varphi = j^+ f^-(r) + j^- f^+(r) \sim j^+ e^{ikr} + j^- e^{-ikr} , \tag{2.2}$$

where $j^\pm$ are the Jost functions and $f^\pm(r)$ are the two linearly independent irregular solutions of (2.1). The Regge poles are defined as the solutions of equation $j^- = 0$. Since j^- is ϵ dependent, it follows that λ_n, being the nth root of the last equation, is also ϵ dependent. Hence, for a small perturbation, we can write

$$\lambda_n = \lambda_n^0 + \epsilon \lambda_n^1 + \tfrac{1}{2}\epsilon^2 \lambda_n^2 . \tag{2.3}$$

Postponing the discussion about the restrictions of the expansion series (2.3) to Sec. 3, let us find the coefficients λ_n^m, $m = 0,1,2,\cdots$.

The roots of the equation $j^- = 0$ are assumed to be simple; therefore they can be calculated from

$$\lambda_n = \frac{1}{2\pi i} \oint d\lambda \, \lambda \frac{1}{j^-} \frac{\partial j^-}{\partial \lambda} , \tag{2.4}$$

where the contour integration encircles only the nth root. It can be shown that for real k the solutions of (2.4) are complex[2] with no value in the fourth quadrant, while for imaginary k they are either real or imaginary.

From (2.4) the coefficients in the series (2.3) are deduced as follows:

$$\lambda_n^0 = \lambda_n \ (\epsilon = 0), \quad \lambda_n^1 = \frac{\partial \lambda_n}{\partial \epsilon} \ (\epsilon = 0),$$

$$\lambda_n^2 = \frac{\partial^2 \lambda_n}{\partial \epsilon^2} \ (\epsilon = 0) . \tag{2.5}$$

[a)]The work was supported by grant NSF F6F006-Y.

 0022-2488/80/071881-07$1.00

Because j^- is an analytic function, the contour integral is independent of ϵ and the derivatives with respect to ϵ can be taken into the integral. Thus, for λ_n^1 we get

$$\lambda_n^1 = \frac{1}{2\pi i}\oint d\lambda\,\lambda\left[\frac{1}{j^-}\frac{\partial^2 j^-}{\partial\lambda\partial\epsilon} - \frac{\partial j^-}{\partial\lambda}\frac{\partial j^-}{\partial\epsilon}\frac{1}{(j^-)^2}\right]. \quad (2.6)$$

The first integral in (2.6) is simply

$$a = \lambda_n^0 \frac{\partial^2 j^-}{\partial\lambda\partial\epsilon}\left(1\Big/\frac{\partial j^-}{\partial\lambda}\right)\ (\epsilon = 0), \quad (2.7)$$

while the second is

$$b = \left(\lambda_n^0\Big/\frac{\partial j^-}{\partial\lambda}\right)\frac{\partial^2 j^-}{\partial\lambda\partial\epsilon} + \left(1\Big/\frac{\partial j^-}{\partial\lambda}\right)\frac{\partial j^-}{\partial\epsilon}. \quad (2.8)$$

The coefficient λ_n^1 is now

$$\lambda_n^1 = a - b = -\left(1\Big/\frac{\partial j^-}{\partial\lambda}\right)\frac{\partial j^-}{\partial\epsilon}, \quad (2.9)$$

where the limit $\epsilon\to 0$ is assumed. By noticing that the S matrix is defined as

$$S = (j^+/j^-)e^{i\pi(\lambda+1/2)} \quad (2.10)$$

and the residues by

$$\beta_n = \lim_{\lambda\to\lambda_n}(\lambda - \lambda_n)S, \quad (2.11)$$

the expression (2.9) can be put into another form:

$$\lambda_n^1 = -\beta_n^0\frac{\partial}{\partial\epsilon}S^{-1}\ (\epsilon = 0), \quad (2.12)$$

where β_n^0 is the residue of the unperturbed S matrix. This form is often more convenient than (2.9).

Similarly, we get for λ_n^2

$$\lambda_n^2 = -\left(1\Big/\frac{\partial j^-}{\partial\lambda}\right)\left(\frac{\partial^2 j^-}{\partial\epsilon^2} + 2\lambda_n^1\frac{\partial^2 j^-}{\partial\lambda\partial\epsilon} + (\lambda_n^1)^2\frac{\partial^2 j^-}{\partial\lambda^2}\right), \quad (2.13)$$

where we have used (2.9). Higher-order coefficients are calculated in an analogous manner.

Let us turn our attention to the residues of the S matrix, defined by (2.11). Since λ_n and the S matrix are ϵ dependent, the residues are also ϵ dependent. For a small perturbation, they are also expanded in a power series

$$\beta_n = \beta_n^0 + \epsilon\beta_n^1 + \tfrac{1}{2}\epsilon^2\beta_n^2 + \cdots, \quad (2.14)$$

where

$$\beta_n = \frac{1}{2\pi i}\oint d\lambda\,\frac{j^+}{j^-}e^{i\pi(\lambda+1/2)} \quad (2.15)$$

and the contour encircles the nth pole. The definition (2.11) is the result of the contour integration of (2.15).

To find the coefficients in (2.14) we proceed in an analogous manner as for λ_n. We get for β_n^1

$$\beta_n^1 = \frac{1}{2\pi i}\oint d\lambda\left(\frac{1}{j^-}\frac{\partial j^+}{\partial\epsilon} - \frac{j^+}{(j^-)^2}\frac{\partial j^-}{\partial\epsilon}\right)e^{i\pi(\lambda+1/2)}. \quad (2.16)$$

The result of the first integration is

$$a = \beta_n^0\frac{1}{j^+}\frac{\partial j^+}{\partial\epsilon}, \quad (2.17)$$

while the second integral is

$$b = \lim_{\lambda\to\lambda_n}\frac{\partial}{\partial\lambda}\left(\frac{j^+}{(j^-)^2}(\lambda - \lambda_n)^2\frac{\partial j^-}{\partial\epsilon}e^{i\pi(\lambda+1/2)}\right). \quad (2.18)$$

By taking derivatives of the expression in the bracket, and subtracting b from a, we get for β_n^1

$$\beta_n^1 = \frac{\beta_n^0}{j^+}\frac{\partial j^+}{\partial\epsilon} + i\pi\lambda_n^1\beta_n^0 - \left(\beta_n^0\Big/\frac{\partial j^-}{\partial\lambda}\right)\frac{\partial^2 j^-}{\partial\epsilon\partial\lambda} + \left(\lambda_n^1\Big/\frac{\partial j^-}{\partial\lambda}\right)\frac{\partial j^+}{\partial\lambda}e^{i\pi(\lambda+1/2)} - \left(\lambda_n^1\beta_n^0\Big/\frac{\partial j^-}{\partial\lambda}\right)\times\frac{\partial^2 j^-}{\partial\lambda^2}, \quad (2.19)$$

or in a more compact form

$$\beta_n^1 = \lambda_n^1\beta_n^0\frac{\partial}{\partial\lambda}\ln\left(\frac{\partial s^{-1}}{\partial\epsilon}\Big/\frac{\partial s^{-1}}{\partial\lambda}\right), \quad (2.20)$$

where S^{-1} is the inverse of (2.10) and λ_n^1 is given by (2.9) or (2.12).

3. CONNECTION OF EXPANSION COEFFICIENTS WITH PERTURBATION POTENTIAL

It has been shown how to calculate the expansion coefficients of λ_n and β_n and how they are related to the derivatives of the Jost functions. Derivatives of j^- with repect to λ can be taken without difficulty since the result is independent of the perturbation potential. However, derivatives in the variable ϵ can only be taken after solving (2.1) first and subsequently letting $\epsilon\to 0$.

Let us for the the moment assume that the limit $\epsilon\to 0$ of φ exists. Formally we can write the integral equation for $j^\pm$ in terms of V'^2

$$j^\pm = j_0^\pm \pm \frac{\epsilon}{2ik}\int_0^\infty f_0^\pm(r)V'(r)\varphi(r)\,dr, \quad (3.1)$$

where $j_0^\pm$ are the Jost functions of the unperturbed system. Since $\varphi(r)$ is a solution of

$$\varphi(r) = \varphi_0(r) - \frac{\epsilon}{2ik}\int_0^r K(r,r')V'(r')\varphi(r')\,dr', \quad (3.2)$$

where $\varphi_0(r)$ is a regular solution of the unperturbed system, $K(r,r')$ is defined as

$$K(r,r') = f_0^+(r)f_0^-(r') - f_0^+(r')f_0^-(r). \quad (3.3)$$

The Jost functions $j^\pm$ are given as the power series in ϵ. It is essential to realize that solving (3.2) by iteration produces an absolutely convergent series for all k and ϵ, provided the first moment of V' is finite,

$$\int_0^\infty dr\,r|V'(r)| < \infty, \quad (3.4)$$

and $\mathrm{Re}(\lambda) > 0$. This is an impotant result because it shows that φ can be calculated by specifying ϵ, λ, and k only, i.e., φ has no branching points or other types of singularity. The restriction $\mathrm{Re}(\lambda) > 0$ imposes a limit on the variation of λ but for a certain class of potentials the analytic continuation into $\mathrm{Re}(\lambda) < 0$ is possible. However, in our applications, we will assume the potentials with a hard core, for which the Regge poles are confined to the first and third quadrant. Since the

poles of the third quadrant do not enter the Regge representation of the scattering amplitude,[3] the power series obtained by solving (3.2) can be used in the calculation of the coefficients (2.5).

Calculating derivatives of $j^{\pm}$ is now straightforward; thus for example

$$\frac{\partial j^-}{\partial \epsilon} = -\frac{1}{2ik}\int_0^\infty f_0^-(r)V'(r)\varphi_0(r)\,dr\,; \tag{3.5}$$

hence the coefficient λ_n^1 is

$$\lambda_n^1 = \frac{1}{2ik}\left(1\Big/ j_0^+\frac{\partial j_0^-}{\partial \lambda}\right)\int_0^\infty V'(r)\varphi_0^2(r)\,dr\,. \tag{3.6}$$

Similarly one calculates other derivatives, whence β_n^1 and λ_n^2 are obtained:

$$\beta_n^1 = \lambda_n^1\beta_n^0\left[i\pi - \left(1\Big/\frac{\partial j_0^-}{\partial\lambda}\right)\frac{\partial^2 j_0^-}{\partial\lambda^2}\right] + \left(\beta_n^0\Big/ ikj_0^+\frac{\partial j_0^-}{\partial\lambda}\right)\int_0^\infty V'\varphi_0\frac{\partial\varphi_0}{\partial\lambda}\,dr \tag{3.7}$$

and

$$\lambda_n^2 = -\left(1\Big/\frac{\partial j_0^-}{\partial\lambda}\right)\left[(\lambda_n^1)^2\frac{\partial^2 j^-}{\partial\lambda^2} - \frac{\lambda_n^1}{ik}\frac{\partial}{\partial\lambda}\times\int_0^\infty f_0^-\varphi_0 V'\,dr - \frac{1}{2k^2}\int_0^\infty f_0^- V'\,dr\times\int_0^r K(r,r')V'(r')\varphi_0(r')\,dr'\right]. \tag{3.8}$$

4. PERTURBATION EXPANSION OF INELASTIC REGGE POLES FOR SPHERICALLY SYMMETRIC POTENTIALS

It was shown in the previous sections how to develop a perturbation method for the Regge poles and the appropriate residues in the elastic collisions. The same technique can be generalized to the inelastic Regge poles, taking the coupling matrix as a perturbation. In this article we will treat the spherically symmetric potentials since generalization to the nonspherical potentials is not straightforward, involving analytic continuation of the S matrix in two angular momentum variables.[8]

We write the radial Schrödinger equation for inelastic collision as

$$\varphi'' = [V + (\lambda^2 - \tfrac{1}{4})/r^2 - k^2]\varphi\,, \tag{4.1}$$

where V is a potential matrix. It is given by

$$V = V_0 + \epsilon V'\,, \tag{4.2}$$

where V_0 is a diagonal matrix and V' has zeros on the diagonal. The channel energy matrix k^2 is diagonal,

$$k^2_{m,n} = k_n^2\delta_{m,n}\,. \tag{4.3}$$

For $\epsilon = 0$ the set of Eq. (4.1) is uncoupled and the solutions are n separate Schrödinger equations

$$\varphi_j^{0''} = [V_{j,j} + (\lambda^2 - \tfrac{1}{4})/r^2 - k_j^2]\varphi_j^0,\quad j = 1,\ldots,n\,. \tag{4.4}$$

In an analogy to the elastic case, we define the Jost functions as

$$\varphi = f^-(r)j^+ + f^+(r)j^- \underset{r\to\infty}{\sim} f_0^-(r)j^+ + f_0^-(r)j^-\,, \tag{4.5}$$

where $f_0^{\mp}$ are the irregular solutions of the uncoupled Eqs. (4.4). The Jost functions $j^{\pm}$ are now the $(n\times n)$ matrices. From the Jost functions we find the S matrix

$$S = \exp[i(\pi/2)(\lambda+\tfrac{1}{2})]k^{\frac{1}{2}}j^+(j^-)^{-1}k^{-\frac{1}{2}} \times\exp[i(\pi/2)(\lambda+\tfrac{1}{2})]\,. \tag{4.6}$$

The Regge poles are the singular points of S in the λ variable and these are determined from $\mathrm{Det}(j^-) = 0$. This condition is equivalent to asking which linear transformation on (4.5) produces a solution with only the outgoing waves, i.e., to find A with the property $j^-A = 0$. The set of equations has a solution if and only if $\mathrm{Det}(j^-) = 0$.

We will from now use a more convenient representation of the S matrix then that given by (4.6). It was shown by Newton[9] that the diagonal elements of S are

$$S_{m,m} = \exp[i\pi(\lambda+\tfrac{1}{2})]\,D(-k_m)/D \tag{4.7a}$$

and the off-diagonal

$$S^2_{m,n} = S_{m,m}\cdot S_{n,n} - [D(-k_m,-k_n)/D] \times\exp[2i\pi(\lambda+\tfrac{1}{2})]\,, \tag{4.7b}$$

where $D = \mathrm{Det}(j^-)$. The function $D(-k_m)$ is equal to D with k_m replaced by $-k_m$. This is also the case with $D(-k_m,-k_n)$, where k_m and k_n in D are replaced by its negative values. The indices m and n refer to the open channels.

The poles in the angular momentum variable λ of the S matrix are now the roots of the equation

$$D = 0\,. \tag{4.8}$$

In the case of no coupling, i.e., $\epsilon = 0$, the Jost function reduces to a diagonal form; hence

$$D = j_1^0 j_2^0\cdots j_n^0 = 0\,, \tag{4.9}$$

defining n independent equations of the type $j_m^0 = 0$, $m = 1,\ldots,n$. Each equation defines a set of Regge poles, designated by $\lambda_{p,m}^0$ where the index p refers to the pth pole of the mth uncoupled Jost function. When the coupling is "switched on," each pole $\lambda_{p,m}$ moves in the λ plane and if ϵ is small, we can write

$$\lambda_{p,m} = \lambda_{p,m}^0 + \epsilon\lambda_{p,m}^1 + \tfrac{1}{2}\epsilon^2\lambda_{p,m}^2 + \cdots\,. \tag{4.10}$$

The coefficients of (4.10) can be calculated by a procedure developed for the single channel case. Thus, for example, $\lambda_{p,m}^1$ is

$$\lambda_{p,m}^1 = -\frac{\partial D}{\partial\epsilon}\Big/\frac{\partial D}{\partial\lambda},\quad \epsilon = 0,\quad \lambda = \lambda_{p,m}^0\,. \tag{4.11}$$

To find derivatives of D we use the representation[10]

$$D = \mathrm{Det}(j^-) = \exp\{\mathrm{Tr}[\ln(j^-)]\}\,, \tag{4.12}$$

where Tr designates the trace of the matrix.

Therefore we have

$$\frac{\partial D}{\partial\epsilon} = D\,\mathrm{Tr}\left((j^-)^{-1}\frac{\partial j^-}{\partial\epsilon}\right). \tag{4.13}$$

In the limit $\epsilon\to 0$, the Jost function j^- is diagonal. As it will be shown in Sec. 6, in the same limit the derivative $\partial j^-/\partial\epsilon$ has zeros on the diagonal, whence

$$\lambda_{p,m}^1 \sim \frac{\partial D}{\partial\epsilon} = 0,\quad \epsilon = 0\,. \tag{4.14}$$

In other words, the leading expansion coefficients in the series (4.10) is exactly zero, and the series starts with the order ϵ^2. The coefficient $\lambda^2_{p,m}$ is simply obtained from (2.13) with D replacing j^-. Taking into account (4.14) we get

$$\lambda^2_{p,m} = -\left(1\Big/\frac{\partial D}{\partial\lambda}\right)\frac{\partial^2 D}{\partial\epsilon^2}. \tag{4.15}$$

The second derivative of D is determined from (4.13) being equal to

$$\frac{1}{D}\frac{\partial^2 D}{\partial\epsilon^2} = \mathrm{Tr}\left((j^-)^{-1}\frac{\partial^2 j^-}{\partial\epsilon^2}\right) - \mathrm{Tr}\left((j^-)^{-1}\frac{\partial j^-}{\partial\epsilon}(j^-)^{-1}\frac{\partial j^-}{\partial\epsilon}\right), \tag{4.16}$$

where we have used (4.14). We can simplify (4.16) by expanding the Jost matrix j^- in a power series of ϵ; hence for each matrix element we have

$$j^-_{r,s} = j^{\bar 0}_s\delta_{r,s} + \epsilon j^{\bar 1}_{r,s} + \tfrac{1}{2}\epsilon^2 j^{\bar 2}_{r,s} + \cdots, \tag{4.17}$$

giving for $\lambda^2_{p,m}$

$$\lambda^2_{p,m} = \left(1\Big/\sum_s\frac{1}{j^-_s}\frac{\partial j^-_s}{\partial\lambda}\right)\left(2\sum_{s>r}\frac{j^{\bar 1}_{s,r}j^{\bar 1}_{r,s}}{j^{\bar 0}_s j^{\bar 0}_r} - \sum_s\frac{j^{\bar 2}_{s,s}}{j^{\bar 0}_s}\right). \tag{4.18}$$

By taking into account that $j^{\bar 0}_m \to 0$, the last expression is

$$\lambda^2_{p,m} = \left(1\Big/\frac{\partial j^-_m}{\partial\lambda}\right)\left(2\sum_{s\neq m}\frac{j^{\bar 1}_{m,s}j^{\bar 1}_{s,m}}{j^{\bar 0}_s} - j^{\bar 2}_{m,m}\right). \tag{4.19}$$

It will be shown in Sec. 6 how the matrix elements of j^- are related to the perturbation matrix.

5. PERTURBATION EXPANSION FOR THE RESIDUES

The residues are defined analogously to the elastic ones

$$\beta = \mathrm{Res}(s) = \lim_{\lambda\to\lambda_{p,m}}(\lambda-\lambda_{p,m})s. \tag{5.1}$$

Using the matrix elements notation (4.7) we can write for the diagonal residues

$$\beta^{p,r}_{m,m} = \lim_{\lambda\to\lambda_{p,r}}(\lambda-\lambda_{p,r})s_{m,m}, \tag{5.2}$$

and for the off-diagonal residues

$$\beta^2_{m,n} = \beta_{m,m}\beta_{n,n}, \tag{5.3}$$

showing that the off-diagonal elements of β are not independent quantities. In other words, diagonal residues determine (5.1) completely, up to a sign. This implies that only the diagonal elements of β have to be expanded in the perturbation series since the off-diagonal residues are given by (5.3). We therefore write only one index, i.e.,

$$\beta^{p,r}_{m,m} \equiv \beta^{p,r}_m, \tag{5.4}$$

where the superscript designates the Regge pole $\lambda_{p,r}$ essential for the residue.

We now write for $\beta^{p,r}_m$

$$\beta^{p,r}_m = \beta^{p,r}_m(0) + \epsilon\beta^{p,r}_m(1) + \tfrac{1}{2}\epsilon^2\beta^{p,r}_m(2) + \cdots, \tag{5.5}$$

where

$$\beta^{p,r}_m = \frac{1}{2\pi i}\oint d\lambda\,\frac{D(-k_m)}{D}\exp[i\pi(\lambda+\tfrac{1}{2})]. \tag{5.6}$$

$\beta^{p,r}_m(0)$ is equal to zero for $r\neq m$ since both $D(-k_m)$ and D are zero in the limit $\lambda\to\lambda_{p,r}$. In other words, the channel m is not, in the zeroth order, perturbed by the presence of a Regge pole in the channel r.

Other coefficients in (5.5) are found by the procedure already described in Sec. 2. Thus for example $\beta^{p,r}_m(1)$ is given by (2.19) with replacement $j^-\to D$ and $j^+\to D(-k_m)$. By the arguments given in Sec. 4 it can be shown that

$$\beta^{p,r}_m(1) = 0; \tag{5.7}$$

therefore the leading term in (5.5) is of the order ϵ^2. Noticing that (4.14) applies to $D(-k_m)$ as well, the coefficient $\beta^{p,r}_m(2)$ is given by

$$\frac{\partial^2\beta^{p,r}_m}{\partial\epsilon^2} = \frac{1}{2\pi i}\oint d\lambda\left(\frac{1}{D}\frac{\partial^2 D(-k_m)}{\partial\epsilon^2} - \frac{D(-k_m)}{D^2}\frac{\partial^2 D}{\partial\epsilon^2}\right)\times\exp[i\pi(\lambda+\tfrac{1}{2})], \tag{5.8}$$

where the first integral is

$$a = \beta^{p,r}_m(0)\frac{1}{D(-k_m)}\frac{\partial^2 D(-k_m)}{\partial\epsilon^2},\quad \epsilon = 0, \tag{5.9}$$

while the second is

$$b = i\pi\left(\beta^{p,r}_m(0)\Big/\frac{\partial D}{\partial\lambda}\right)\frac{\partial^2 D}{\partial\epsilon^2} + \frac{\partial D(-k_m)}{\partial\lambda}\left(\beta^{p,r}_m(0)\Big/D(-k_m)\frac{\partial D}{\partial\lambda}\right)\frac{\partial^2 D}{\partial\epsilon^2} + \left(\beta^{p,r}_m(0)\Big/\frac{\partial D}{\partial\lambda}\right)\frac{\partial^3 D}{\partial\epsilon^2\partial\lambda} - \left[\beta^{p,r}_m(0)\Big/\left(\frac{\partial D}{\partial\lambda}\right)^2\right]\frac{\partial^2 D}{\partial\epsilon^2}\frac{\partial^2 D}{\partial\lambda^2}. \tag{5.10}$$

The coefficient $\beta^{p,r}_m(2)$ is now given by

$$\beta^{p,r}_m(2) = a - b. \tag{5.11}$$

Let us assume that $r\neq m$, i.e., the diagonal element of β is from the mth channel while the unperturbed Regge pole is in the rth channel. The unperturbed residue $\beta^{p,r}_m(0)$ is zero in that case. Therefore, the terms from a and b not having $D(-k_m)$ in the denominator are exactly zero; hence the coefficient $\beta^{p,r}_m(2)$ gets the contribution only from a and the second term in b, giving

$$\beta^{p,r}_m(2) = \left(\exp[i\pi(\lambda^0_{p,r}+\tfrac{1}{2})]\ \frac{\partial j\ r}{\partial\lambda}\right)\Big[j^2_{r,r}(-k_m) - \frac{j^+_m}{j^-_m}j^2_{r,r} + 2\sum_{s\neq r}\left(\frac{j^+_m}{j^-_m}\frac{j^1_{s,r}j^1_{r,s}}{j_s} - \frac{j_{s,r}(-k_m)j^-_{r,s}(-k_m)}{j^-_s(-k_m)}\right), \tag{5.12}$$

where the index $-k_m$ indicates that in the matrix element of j^-, k_m is replaced by $-k_m$.

The case when $r = m$ is more complicated and will be treated in the following section.

6. CONNECTION WITH THE PERTURBATION POTENTIAL

In the expression for the expansion coefficients of the Regge poles and residues of the S matrix, the essential role is played by the matrix elements of the Jost function. It is

therefore necessary to relate them to the perturbation potential, defined in (4.2).

The regular solution of (4.1), if (4.2) is assumed, is

$$\varphi = \varphi_0 + \frac{\epsilon}{2i} k^{-1} \int_0^r K(r,r') V'(r') \varphi(r')\, dr', \tag{6.1}$$

where $\varphi_0(r)$ is the regular solution of the uncoupled equations (4.1), and

$$K(r,r') = f_0^+(r') f_0^-(r) - f_0^+(r) f_0^-(r'), \tag{6.2}$$

where

$$(f_0^{\pm})_{p,s} \sim \delta_{p,s} e^{\mp i k_p r}. \tag{6.3}$$

By taking the limit $r \to \infty$ in (6.1) and comparing with (4.5), we obtain the integral equation for the Jost functions

$$j^{\pm} = j_0^{\pm} \pm \frac{\epsilon}{2i} k^{-1} \int_0^{\infty} f_0^{\pm}(r') V'(r') \varphi(r')\, dr'. \tag{6.4}$$

In the first approximation we take $\varphi = \varphi_0$ and $j^{\pm}$ is to the first order in ϵ given by

$$j^{\pm} = j_0^{\pm} \pm \frac{\epsilon}{2i} k^{-1} \int_0^{\infty} f_0^{\pm}(r') V'(r') \varphi_0(r')\, dr'. \tag{6.5}$$

Since $V'(r)$ has zero diagonal elements, it follows that

$$j^1_{s,s} = 0, \tag{6.6}$$

where we have used the notation of (4.17). In general, the coefficients $j^1_{p,s}$ are

$$j^1_{p,s} = -\frac{1}{2ik_p} \int_0^{\infty} f_p^-(r) V'_{p,s}(r) \varphi_s(r)\, dr. \tag{6.7}$$

The second-order correction to the Jost function is obtained by iterating (6.1) once and subsequently replacing φ in (6.4), in which case we get

$$j^2_{p,s} = \frac{1}{2k_p} \sum_q \frac{1}{k_q} \int_0^{\infty} dr f_p^-(r) V_{p,q}(r) \times \int_0^r dr'\, K_{q,q}(r,r') V_{q,s}(r') \varphi_s(r'). \tag{6.8}$$

We are now able to write the explicit relationship between (4.19) and the coupling potential, hence obtaining

$$\lambda^2_{p,m} = -\left(2k_m j_m^+ \frac{\partial j_m^-}{\partial \lambda}\right)^{-1} \sum_{s \neq m} \frac{1}{k_s j_s^-} \left(\int_0^{\infty} \varphi_m V_{m,s} f_s^-\, dr \times \int_0^r \varphi_s V_{s,m} \varphi_m dr' + \int_0^{\infty} \varphi_m V_{m,s} \varphi_s\, dr \times \int_r^{\infty} f_s^- V_{s,m} \varphi_m\, dr' \right). \tag{6.9}$$

Similarly, we get the expression for $\beta^{p,q}_m$, defined by (5.13),

$$\beta^{p,q}_m(2) = \exp[i\pi(\lambda^0_{p,q} + \tfrac{1}{2})] \left(\frac{j_m^+}{j_m^-} - 1 \right) \lambda^2_{p,q} - \frac{\exp[i\pi(\lambda^0_{p,q} + \tfrac{1}{2})]}{2k_q k_m j_q^+ j_m^+ j_m^-} \left(\frac{\partial j_q^-}{\partial \lambda} \right)^{-1} \times \left(\int_0^{\infty} \varphi_q V_{q,m} \varphi_m\, dr \right)^2. \tag{6.10}$$

The most complicated case is when $r = m$. The reason is that $\lambda_{p,m}$ while being a zero of D is not a zero of $D(-k_m)$. Hence, all the terms in (5.9) and (5.10) must be included in the expansion coefficient $\beta^{p,m}_m(2)$ and the result is

$$\begin{aligned} \beta^{p,m}_m(2) = \beta^{p,m}_m(0) \Bigg\{ & \lambda^2_{p,m} \left[i\pi - \left(1 \Big/ \frac{\partial j_m^-}{\partial \lambda} \right) \frac{\partial^2 j_m^-}{\partial \lambda^2} \right] \\ & - \frac{1}{2k_m j_m^+} \sum_{q \neq m} \frac{1}{k_q j_q^-} \left(\int_0^{\infty} dr f_m^+ V_{m,q} f_q^- \right. \\ & \times \int_0^r dr' \varphi_q V_{q,m} \varphi_m + \int_0^{\infty} dr f_m^+ V_{m,q} \varphi_q \\ & \times \left. \int_r^{\infty} dr' f_q^- V_{q,m} \varphi_m \right) - \left(1 \quad 2k_m j_m^+ \frac{\partial j_m^-}{\partial \lambda} \right) \frac{\partial}{\partial \lambda} \\ & \times \sum_{q \neq m} \frac{1}{k_q j_q^-} \\ & \times \left(\int_0^{\infty} \varphi_m V_{m,q} f_q^-\, dr \int_0^r \varphi_q V_{q,m} \varphi_m\, dr' \right. \\ & + \left. \left. \int_o^{\infty} \varphi_m V_{m,q} \varphi_q\, dr \int_r^{\infty} dr' f_q^- V_{q,m} \varphi_m \right) \right\}. \end{aligned} \tag{6.11}$$

In this way we found the relationship between the expansion coefficients of the Regge poles and residues and the perturbation potential. However, the results need further discussion revealing some of the properties of the Regge poles and residues.

7. DISCUSSION

In the analysis of the scattering amplitude it is very often needed to calculate the contribution of the closed channels. In the conventional methods, e.g., the distorted wave approximation, the closed channels contribute in the second order of perturbation, which often turns out to be infinite.[11] Therefore, such methods are only suitable for certain class of potentials e.g., exponentially decaying potentials, for which this contribution is finite.

In our procedure the contribution of the closed channels to the scattering amplitude is given via the Regge poles and residues, for which the perturbation series was developed. It is therefore worth noticing that the coefficients $\lambda^2_{p,m}$ and $\beta^{p,r}_m(2)$ are finite without imposing any restrictions on the form of the potential, except that it should go to zero faster than r^{-2}. This is simply obtained from (6.9).

Let us also notice that the leading term in the perturbation series of both the Regge poles and residues is of the order ϵ^2. The first-order contribution in ϵ is exactly zero.

Let us, on the example of a two channel problem, discuss the various poles. We will assume that one channel is closed, i.e., $k^2_{22} < 0$. The other channel is open, meaning $k^2_{11} > 0$. Furthermore we will assume that V_{22} is typical of the atomic potentials, i.e., it has a hard core and a minimum at $r = r_0$. For $r \to \infty$, it goes to zero as some inverse power of r.

For $k^2_{22} < 0$ the function

$$p^2 = V_{22} + \frac{\lambda^2 - \frac{1}{4}}{r^2} - k^2_{22} \tag{7.1}$$

has two turning points and the Regge poles corresponding to V_{22} (the unperturbed poles) are given from the WKB solution of the Schrödinger equation. The appropriate Jost functions can be calculated and are given by

$$j^+ = -\cos\left(\int_a^b p\,dr + \pi/2\right) \times \exp\left[k_{22}b - \int_b^\infty (|p| - k_{22})\,dr\right] \tag{7.2}$$

and

$$j^- = 2\sin\left(\int_a^b p\,dr + \pi/2\right) \times \exp\left[-k_{22}b + \int_b^\infty (|p| - k_{22})\,dr\right], \tag{7.3}$$

where

$$p^2(a) = p^2(b) = 0\,. \tag{7.4}$$

The Regge poles, being the roots of j^-, are then the solutions of the equation

$$\int_a^b p\,dr = (n + \tfrac{1}{2})\pi, \quad n = 0,1,2,\cdots \tag{7.5}$$

in the variable λ. The residues are also obtained and are given by

$$\beta_n = \left(1/2\lambda_n \int_a^b \frac{dr}{r^2 p}\right) \exp\left[i\pi\lambda_n + 2k_{22}b - 2\int_b^\infty (p - k_{22})\,dr\right]. \tag{7.6}$$

The poles, the solution of (7.5), are either real or imaginary. The real poles correspond to the bound states, when they are half integer. The imaginary poles give negative centrifugal terms hence they lower p^2. In other words, the states with imaginary λ are the solutions of the Schrödinger equation when k^2 is below the minimum of the potential. Such states are meaningless in a one-channel problem, but in the inelastic collisions they must also be included.

The poles corresponding to the open channel are complex and are in the first and third quadrant of the λ plane. For simplicity let us assume that k_{11}^2 is much larger than the minimum of V_{11}. The imaginary part of such poles is large.[3,5] The result is that they only contribute to the forward space of the differential cross section, giving the rainbow and the diffraction oscillations.[5] In our treatment these poles will not be considered.

Let us now discuss the closed channel poles when the perturbation is included.

A. Imaginary poles

The poles from the closed channels do not contribute to $\beta_1^p(0)$; hence by introducing coupling we have approximately

$$\beta_1^p \sim \frac{i}{2}\left[\left(\frac{j_1^+}{j_1^-} - 1\right)\lambda_p^2 + \left(i/\frac{\partial j_2^-}{\partial\lambda}\right)\frac{(\int_0^\infty \varphi_2 V_{1,2}\varphi_1\,dr)^2}{2k_1|k_2|j_2^+ j_1^+ j_1^-}\right] \times \exp(-\pi|\lambda_p^0|) \tag{7.7}$$

and

$$\lambda_p^2 = \left(1/j_2^+ \frac{\partial j_2^-}{\partial\lambda}\right)\frac{i}{2|k_2|}\frac{1}{k_1 j_1^-} \times\left(\int_0^\infty \varphi_2 V_{2,1} f_1^-\,dr \int_0^r \varphi_1 V_{1,2}\varphi_2\,dr' + \int_0^\infty \varphi_2 V_{2,1}\varphi_1\,dr \int_r^\infty dr' f_1^- V_{1,2}\varphi_2\right). \tag{7.8}$$

Let us prove that any perturbation $V_{2,1}$ gives a contribution to λ_p^0 with a positive real part. This is essential since the exact poles λ_p lie in the first quadrant. The leading term in the perturbation expansion, therefore, has to give a positive real contribution. The real part is

$$\mathrm{Re}(\lambda_p^2) = \tfrac{1}{2}(\lambda_p^2 + \lambda_p^{2*})\,. \tag{7.9}$$

From the relationship[2]

$$ikj^+\frac{\partial j^-}{\partial\lambda} = -\lambda\int_0^\infty \frac{\varphi^2}{r^2}\,dr \tag{7.10}$$

it is easily deduced that the product $j_2^+(\partial j_2^-/\partial\lambda)$ is positive imaginary. Hence, (7.9) becomes

$$\mathrm{Re}(\lambda_p^2) = \tfrac{1}{4}\left|\frac{1}{k_2 k_1 j_2^+}\left(\frac{\partial j_2^-}{\partial\lambda}\right)^{-1}\right|\frac{1}{j_1^- j_1^+} \times\left(\int_0^\infty \varphi_2 V_{2,1}\varphi_1\,dr\right)^2, \tag{7.11}$$

being indeed positive. In (7.11) we have used the fact that φ_1 and φ_2 are real.

For large $|\lambda_p^0|$, the residue β_1^p is small and therefore the contribution of the imaginary poles is negligible. However, there are cases when the imaginary poles are close to the real axis. Provided the integrals in (7.7) and (7.8) are not negligible, the residue β_1^p have a noticeable value. In that case the contribution of such poles to the scattering amplitude is approximately

$$f_{1,1} \sim 1 + [(1 - 4\lambda_p^2)/4]\cos^2(\theta/2)\,, \tag{7.12}$$

being a nonoscillatory function of angle. Therefore, the imaginary poles, at best, give a contribution which is a smooth function of the angle θ.

B. Real poles

The residues and the poles are given by (7.7) and (7.8), however, the exponential factor $\exp(-\pi|\lambda_p^0|)$ is replaced by $\exp(-\pi\lambda_p^0)$. By similar arguments as for the imaginary poles, the imaginary part of λ_p^2 can be proved to be positive. From

$$\mathrm{Im}(\lambda_p^2) = \tfrac{1}{2}(\lambda_p^2 - \lambda_p^{2*}) \tag{7.13}$$

and using (7.10) we find that (7.13) is equal to (7.11), however, Im replacing Re. The properties of (7.6) are now essentially determined by the integrals appearing in (7.7) and (7.8). If the integrals are small, the residues are small, hence, the contribution of such poles is negligible. However, if λ_p is half integer, the scattering amplitude is approximately

$$f_{1,1} \sim P_{\lambda_p - \frac{1}{2}}(-\cos\theta)\,, \tag{7.14}$$

having a noticable contribution to the differential cross section. A similar thing happens in the elastic collisions, howev-

er, in the inelastic collisions this effect is due to the closed channels. We talk about the Feschbach resonances. This short discussion of the scattering amplitude reveals the consistency of the perturbation method approach for calculating the inelastic Regge poles. We have shown the consistency for a two-channel case, but generalization to more channels is straightforward.

8. CONCLUSION

We have shown a method for generating the inelastic Regge poles and residues by a perturbation approach. In each unperturbed channel we have defined a set of poles and developed a method for finding their positions when a perturbation is introduced. The unperturbed channels represent the elastic collisions; therefore this procedure provides one way of describing the inelastic differential cross section in terms of the elastic cross sections of each channel. By knowing the properties of the Regge poles and the residues of the unperturbed system we deduce characteristic features of the inelastic cross section using such a perturbation method.

There is an interesting feature of this perturbation method not directly related to the collisions. If the same technique is applied for the poles of the S matrix in the energy variable, formula (3.8) shows that the definition of a complete set of functions, in the usual perturbation theory for the bound states, is redundant. If in this formula λ is replaced by k, the second-order perturbation coefficient for the energy does not involve a sum/integral over the complete set of functions. The coefficients depend only on the solution of that particular eigenstate.

[1]A. Bottino, A.M. Longoni, and T. Regge, Nuovo Cimento **8**, 107 (1962).
[2]V. de Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965).
[3]S. Bosanac, Molec. Phys. **35**, 1057 (1978).
[4]S. Bosanac, J. Math. Phys. **19**, 789 (1978).
[5]S. Bosanac, Phys. Rev. A **19**, 125 (1979).
[6]E.C. Titchmarsh, *Eigenfunction Expansions* (Clarendon, Oxford, 1958), Vol. II (19.20).
[7]R.G. Newton, *The Complex j-Plane* (Benjamin, New York, 1964).
[8]J.M. Charap and E.J. Squires, Ann. Phys. (N.Y.) **20**, 145 (1962); **21**, 8 (1963).
[9]R.G. Newton, J. Math. Phys. **2**, 29 (1958).
[10]K.M. Watson and J. Nuttall, *Topics in Several-Particle Dynamics* (Holden-Day, San Francisco, 1967).
[11]R.G. Newton, Ann. Phys. (N.Y.) **4**, 29 (1958).

Partition permuting array approach to few-body Hamiltonian models of nuclear reactions

R. Goldflam, K. L. Kowalski and W. Tobocman
Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106

(Received 12 June 1979; accepted for publication 15 February 1980)

The approximate Hamiltonian formulation of many-body scattering recently proposed by Polyzou and Redish is used to derive approximate Baer–Kouri–Levin–Tobocman (BKLT) integral equations. It is shown that these equations have connected kernels after a finite number of iterations and the approximation defined by them is unitary. Integral equations imbedding the approximation in the exact theory are also given. Some lemmas which relate to the connectivity of operator products are established and are employed to prove the connected-kernel properties of both the BKLT as well as the imbedding integral equations

I. INTRODUCTION

The standard model for nuclear reactions at energies below the threshold for pion production consists in the use of the nonrelativistic Schrödinger equation description of the dynamics of the N nucleons which comprise the entire system.[1] One of the set of assumptions which make up this picture is the supposition that the internucleon interactions are mediated by static potentials of the pairwise, three-particle, etc. varieties. Given some particular initial state, such as a nucleon incident upon a nucleus, the correct formulation of the problem requires the specification of the boundary conditions appropriate to *all* of the possible outgoing configurations. Thus, a correct starting point for a complete discussion of nuclear reactions within this standard picture is a set of well-defined N-particle scattering integral equations.[2] The condition that the kernels of these equations become connected operators after a finite number of iterations is thought to be an important criterion in determining such a set and this has led to the construction of many alternative equations.[2]

These equations are forbiddingly complex and it is unrealistic to propose calculational procedures which depend upon anything but the solution of some approximate version of them. Thus, except when $N \leqslant 4$, the practical calculation of nuclear reactions inevitably involves the introduction of effective few-body models whose specific realization is designed to reflect the dominant dynamical characteristics using relatively few degrees of freedom. In most of the traditional treatments[1] of nuclear reactions, "few" is, in fact, equal to 2 as, e.g., in the optical model, resonating group, and coupled channels methods. It is only relatively recently that three-body models have begun to be developed seriously although calculations using these methods are still relatively rare.[3] The number of systematic developments of such models has been, until recently, even rarer.

Polyzou and Redish[4] have found an elegant formulation of the N-particle scattering problem which facilitates the introduction of certain types of few-body models in a consistent and controlled fashion. Central to their work is the introduction of an approximate N-particle Hamiltonian $H(A)$ appropriate to a reaction mechanism A. Here A refers to the dominant channels which are used to describe the reaction mechanism. $H(A)$ defines a unitary model for transitions among the cluster states which enter into its spectral resolution. It is possible to develop connected-kernel integral equations for the scattering operators which correspond to these transitions.[4] Also, Ref. 4 contains a systematic procedure for calculating corrections to the model represented by $H(A)$. Other advantages of this general type of approach, e.g., in comparison with those simple few-body models (for reactions *or* bound states) which assume the effective indivisibility of the constituent "bodies," are discussed at greater length in Ref. 4.

One of the advantages of the model Hamiltonian approach of Ref. 4 is that the model problem is stated in terms of definitions of the partition Hamiltonians, external interactions, transition operators, and the Green functions and the resolvent identities satisfied by them that are formally the same as in the untruncated problem. Thus any of the diverse methods[2] for the development of connected-kernel partition-labeled scattering integral equations can be formally applied without substantial modification. In the present article we are concerned with the application of the partition permuting array (PPA)[5] approach to the model problem defined by $H(A)$.

As one might expect, the introduction of the PPA for the model scattering equations is trivial to carry out. What is not obvious is that this technique still works to yield the connected-kernel model counterparts of the Baer–Kouri–Levin–Tobocman (BKLT) N-particle equations. Similar questions were addressed in a previous article[6] which contains an approach alternative to that of Ref. 4 for introducing few-body models within the context of the BKLT formalism. Connectivity questions are handled with the aid of a lemma proved in Appendix A concerning the structure of the class of operator products which appears in PPA-generated theories. This generalizes some results proven in Ref. 6. Finally, we establish a new procedure for imbedding the model problem within the full N-particle dynamics. This provides a method for calculating controlled corrections to the model problem which is alternative to the method of Ref. 4 but is similar to the modification of the Feshbach projection operator technique employed in Ref. 6.

 0022-2488/80/071888-09$1.00

The work of Ref. 4 contains the application and development of some of the most sophisticated techniques available in N-particle scattering theory. Since these are ideas that are as of yet not widely known, we review the relevant aspects of Ref. 4 for the sake of intelligibility.

II. REACTION MECHANISMS

In order to be able to treat even two-cluster initiated nuclear reactions in a comprehensive fashion, it is necessary to allow for the possibility of arbitrary intermediate and final configurations of the N particles which comprise the projectile and the target. There are many of these configurations for even moderately large N and thus an efficient notation is needed to handle this complexity. Central to such a notation is the idea of *partitions* of the N particles into sets $a,b,...$, of $n_a,n_b,...$, clusters of particles, where the order of the groupings of the particles within the clusters and the order of the clusters themselves are regarded as immaterial. For example, for $N = 6$ a three-cluster partition is $(123)(45)(6) = (312)(45)(6) = (45)(6)(312)$. The two partitions which correspond to $n_a = 1$ and $n_a = N$ are often denoted as $a = 1$ and $a = 0$, respectively.

Evidently partitions codify the possible asymptotic states of the N-particle system where interactions take place only among the particles within a cluster. A particular partition actually can be identified with a physical asymptotic state only if it is dynamically possible for all of the particles in each of the clusters to be bound. Such partitions are said to be *stable*. We remark that $a = 0$ is always regarded as a stable partition.

The partition indexing of the Hamiltonians for the asymptotic states as well as the description of interactions among the various clusters requires a precise definition of the interaction internal or external to a partition. This can be done with the aid of the idea of a partition b, say, being *contained in* another partition a which we denote by $a \supseteq b$, where we include the possibility of equality. Here $a \supset b$ means that b can be obtained from a by breaking up one or more of its clusters into clusters containing fewer particles. The alternative possibility where b cannot be obtained in such a manner is denoted by $a \not\supset b$ and then b is said to be *not contained* in a.

The N-particle system is presumed to be described by a Hamiltonian H which we suppose decomposes into the kinetic energy H_0 of the system and a potential V which represents the interaction among the N particles:

$$H = H_0 + V. \tag{2.1}$$

The potential V in general decomposes into the sum

$$V = \sum_a [V]_a, \tag{2.2}$$

where $[V]_a$ is the a-connected part of V. We recall that an N-particle operator A is said to be a-connected if its momentum-space matrix elements have the form

$$\langle \mathbf{p}'_1,...,\mathbf{p}'_N | A | \mathbf{p}_1,...,\mathbf{p}_N \rangle = A_a(\mathbf{p}'_1,...,\mathbf{p}'_N | \mathbf{p}_1,...,\mathbf{p}_N) \times \prod_{\alpha=1}^{n_a} \delta[\mathbf{P}'(\alpha) - \mathbf{P}(\alpha)], \tag{2.3}$$

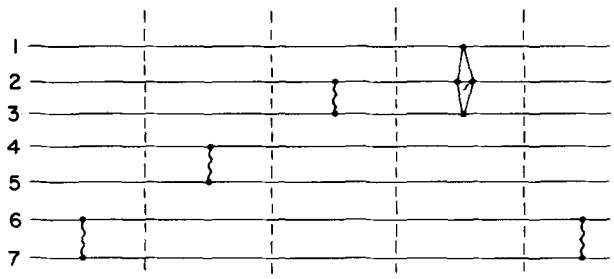


FIG. 1. Disconnected structure of the operator described in the text. The wavy lines connecting the solid horizontal nucleon lines represent pair interactions. The diamond denotes the three-particle force and the vertical dotted lines correspond to the free Green functions G_0.

where

$$\mathbf{P}(\alpha) \equiv \mathbf{p}_{j_\alpha} + \cdots + \mathbf{p}_{k_\alpha}$$

is the momentum of the c.m. of the αth cluster, $(j_\alpha,...,k_\alpha)$ correspond to the particles in the αth cluster, and the functions $A_a(\mathbf{p}'_1,...,\mathbf{p}'_N|\mathbf{p}_1,...,\mathbf{p}_N)$ possess no δ-function singularities. Evidently (2.3) is related to the invariance of A with respect to independent translations of the clusters of particles corresponding to the partition a. This can occur only because there are no interactions linking these clusters. The relationship between the connectivity structure of operators and their invariance properties with respect to the various translation groups associated with the different clusterings of the N particles has been investigated recently by Polyzou.[7] We remark that an operator which is 1-connected is usually referred to as being *fully connected* or, simply, *connected*. For such operators (2.3) contains only a δ function in the total momentum.

Figure 1 provides a graphical illustration of the preceding ideas regarding connectivity. Let $V_{(ij)}$ represent the pair interaction between particles i and j and $V_{(123)}$ a three-particle potential. The free N-particle Green function $G_0 = (z - H_0)^{-1}$, where z is a complex parametric energy, is 0-connected; however, the product $V_{(67)}G_0V_{(45)}G_0V_{(23)}G_0V_{(123)}G_0V_{(67)}$ depicted in Fig. 1 is (123)(45)(67)-connected. If $N = 8$ then this operator product is (123)(45)(67)(8)-connected. We note that $V_{(23)}G_0V_{(123)}$, e.g., commutes with the total momenta of the clusters (56) and (67).

The partitions which enter into the sum on the right side of (2.2) are of a restricted type which conforms to the usual description of V as a sum of pairwise, three-body, etc., forces. For example, a pair interaction $V_{(ij)}$ connects together only the particles i and j and therefore is $a(ij)$-connected. Here $a(ij)$ is the $(N-1)$ cluster partition in which particles i and j are contained in a single cluster (Fig. 1). The three-particle force $V_{(123)}$ in Fig. 1 evidently is $a(123)$-connected, where $a(123)$ is the $(N-2)$-cluster partition in which particles 1,2, and 3 are contained in a single cluster. It is clear from these examples that corresponding to our usual ideas about interparticle potentials, $[V]_a = 0$ unless a is a partition which contains only a single cluster with more than one particle within it.[8] We note in particular that $[V]_0 = 0$, while $[V]_1$ refers to a possible N-particle potential. For the sake of

simplicity in parts of the subsequent development we suppose that $[V]_1 = 0$, although this fully connected interaction can be incorporated without changing our essential results.[4]

The interaction V_a *internal* to the partition a is *defined* by

$$V_a = \sum_{a \supsetneq b} [V]_b, \tag{2.4}$$

and this leads to the introduction of the partition Hamiltonian

$$H_a = H_0 + V_a. \tag{2.5}$$

If a is a stable partition the maximally connected eigenstates of H_a are the states $|\phi_a\rangle$ for which all n_a clusters are in bound configurations, e.g., a two-deuteron state for $N = 4$. The states $|\phi_a\rangle$ are possible asymptotic channel states. The scattering eigenstates $|\psi_a^{(+)}\rangle$ of H_a evolve from the maximally connected eigenstates of some other partition Hamiltonian H_b with $a \supset b$.

Several different definitions of the notion of a channel are used in work on nuclear reactions. In this article we adapt the definition introduced in Ref. 4, namely, a *channel* (ν_a) corresponding to a stable partition a refers to the quantum numbers of the internal (bound) states of the various clusters represented by a maximally connected eigenstate $|\phi_a(\nu_a)\rangle$ of H_a, where we suppress the dependence upon the momenta of the cluster centers of mass. These internal quantum numbers are designated collectively as ν_a.

Thus, both the scattering and bound eigenstates H_a can be denoted as $|\hat{\phi}_a(\nu_b)\rangle$, where b is a stable partition for which $a \supsetneq b$. For $a \supset b$, we have $|\hat{\phi}_a(\nu_b)\rangle = |\psi_a^{(+)}(\nu_b)\rangle$, while $|\hat{\phi}_a(\nu_a)\rangle = |\phi_a(\nu_a)\rangle$. In both instances, the dependencies upon the momenta of the c.m. of each of the n_b clusters has again been suppressed.

Now, let $\mathscr{A}_0$ be the set of all partitions of the N particles with two or more clusters and let $\bar{\mathscr{A}}_0 \subseteq \mathscr{A}_0$ be the subset consisting only of stable partitions. Then the set A_0 of all physical channels, in the previous sense of the term,[4] is

$$A_0 = \{\nu_b \mid \text{ all } b \in \bar{\mathscr{A}}_0\}. \tag{2.6}$$

A *reaction mechanism* (RM) can then be thought of in an abstract way as a subset A of the set A_0. Evidently the notion of a RM takes on content when A is identified as the dominant set of channels for a particular system in a specific energy range and, moreover, that this can be used to simplify the scattering integral equations. The formalism of Ref. 4 provides a vehicle for carrying out this simplification. It should be noted that there are alternative approaches to the question of systematic approximations to N-particle scattering where the preceding definition of A RM is inappropriate.[6]

An important aspect of the development in Ref. 6 is the association of a particular set of partitions $\mathscr{A}$ with a RM A. This concept is also highly relevant to the work of Ref. 4 although it is not used there explicitly. Such an association is not unique. Implicitly in Ref. 4 and explicitly in our subsequent work $\mathscr{A}$ is taken to be[6]

$$\mathscr{A} = \{a \mid a \neq 1,\ a \supseteq b,\ \nu_b \in A\}. \tag{2.7}$$

For example, for $N = 4$ and $b_2 = (12)(34)$, $b_3 = (12)(3)(4)$ if A consists of a single two-cluster channel ν_{b_2} (e.g., a two-deuteron state in the four-nucleon case) and a single three-cluster channel ν_{b_3} (e.g., a deuteron plus two free nucleons), then

$$\mathscr{A} = \{(12)(34),(123)(4),(124)(3),(12)(3)(4)\}. \tag{2.8}$$

We note that it is possible to have $\mathscr{A} = \mathscr{A}_0$ even if $A \subset A_0$. A *few-body RM* A is one for which $n_a \leqslant 4$ for all $a \in \mathscr{A}$.

The set $\mathscr{A}$ yields a criterion for the preservation of the dynamical integrity of the N-particle dynamics. The statement of this criterion as well as the exposition of much of our subsequent analysis concerning connectivity is greatly facilitated by the use of a simple binary rule of combination among the partitions.[4,7-9] The *union* of two partitions a and b of N objects is the partition denoted as $a \cup b = b \cup a$ which not only satisfies $a, b \subseteq a \cup b$ but also satisfies $a \cup b \subseteq c$, where c is any partition of N objects for which $a, b \subseteq c$. For example, $b_2 \cup b_3 = b_2$ in the case of the partitions of the preceding paragraph. We see that $a \cup b$ is the partition with the largest number of clusters which contains both a and b. Evidently we can define in an obvious manner the union of several partitions, e.g., $a \cup b \cup c$, etc.

A partition $\hat{a}(A)$ *characteristic* of the RM A can then be defined as the union of all the partitions in $\mathscr{A}$:

$$\hat{a}(A) \equiv \bigcup_{a \in \mathscr{A}} a. \tag{2.9}$$

The reaction mechanism A is said to be *trivial* if $\hat{a}(A) \neq 1$. For the RM implicit in (2.8) we see that $\hat{a}(A) = (1234)$ and so we have a nontrivial case. Because the union of two distinct two-cluster partitions is 1, A is trivial if and only if $\mathscr{A}$ contains only a single two-cluster partition. We note that $\mathscr{A}$ always includes at least one two-cluster partition.[6] We explore the dynamical consequences of a trivial RM in the next section.

III. RM HAMILTONIANS AND INTERACTIONS

For any partition a we have the following resolution of the identity in terms of the eigenstates of H_a on the N-particle space $\mathscr{H}_N$:

$$I = \sum_b \sum_{\nu_b \in A_0} \Delta_{a,b} \left\{ \prod_{i=1}^{n_b} \int (d\mathbf{P}_i) \right\} |\hat{\phi}_a(\nu_b)\rangle \langle \hat{\phi}_a(\nu_b)|, \tag{3.1}$$

where the integrations are over the c.m. momenta of the n_b clusters of the stable partition b and where

$$\Delta_{a,b} = 1, \quad \text{if } a \supseteq b,$$

and

$$\Delta_{a,b} = 1, \quad \text{if } a \not\supseteq b, \tag{3.2}$$

Equation (3.1) amounts to the decomposition of $\mathscr{H}_N$ into the direct sum of the subspaces $\mathscr{H}_a[\nu_b]$ spanned by the states $\{|\hat{\phi}_a(\nu_b)\rangle\}$ for a given ν_b and a. A projector onto $\mathscr{H}_a[\nu_b]$ is then

$$P_a(\nu_b) = \left\{ \prod_{i=1}^{n_b} \int (d\mathbf{P}_i) \right\} |\hat{\phi}_a(\nu_b)\rangle \langle \hat{\phi}_a(\nu_b)|. \tag{3.3}$$

This allows us to rewrite (3.1) in the more concise form

$$I = \sum_b \sum_{\nu_b \in A_0} \Delta_{a,b} P_a(\nu_b). \tag{3.4}$$

In the absence of N-body forces Polyzou and Redish[4] have shown that the full Hamiltonian H possesses the decomposition

$$H = \sum_a {}' C_a H_a, \tag{3.5}$$

into the partition Hamiltonians. Here $C_a = (-1)^{n_a}(n_a - 1)!$ and the prime on the summation over partitions indicates the omission of $a = 1$. The combination of (3.4) and (3.5) yields a channel decomposition of H:

$$H = \sum_b \sum_{\nu_b \in A_0} H(\nu_b), \tag{3.6}$$

where

$$H(\nu_b) \equiv \sum_a {}' C_a H_a \Delta_{a,b} P_a(\nu_b). \tag{3.7}$$

Equation (3.6) suggests the introduction of an RM Hamiltonian $H(A)$ by suitably restricting the summation over channels to those in the set A:

$$H(A) \equiv \sum_b \sum_{\nu_b \in A} H(\nu_b). \tag{3.8}$$

In a similar manner a projection operator corresponding to A and a particular partition a is [cf. Eq. (3.4)]

$$P_a(A) = \sum_b \sum_{\nu_b \in A} \Delta_{a,b} P_a(\nu_b). \tag{3.9}$$

Then (3.8) becomes

$$H(A) = \sum_a C_a H_a P_a(A). \tag{3.10}$$

A comparison of Eqs. (3.5) and (3.10) suggests the definition

$$H_a(A) = P_a(A) H_a = H_a P_a(A), \tag{3.11}$$

as the counterpart of the partition Hamiltonian H_a when the dynamics is truncated by the introduction of a reaction mechanism. We note that $P_a(A_0) = I$, $H(A_0) = H$, and $H_a(A_0) = H_a$.

It is evident that $H(A)$ possesses a cluster decomposition

$$H(A) = \sum_a {}' [H(A)]_a, \tag{3.12}$$

where $[H(A)]_a$ is the a-connected part of $H(A)$. However, it is not obvious whether the counterparts of Eqs. (2.4) and (2.5), namely,

$$H_a(A) = \sum_c {}' \Delta_{a,c} [H(A)]_c, \tag{3.13}$$

are valid. In this regard we note that $[H]_0 = H_0$ and that (2.5) can be rewritten as

$$H_a = \sum_c {}' \Delta_{a,c} [H]_c. \tag{3.14}$$

In Ref. 4 it is shown that (3.13) is valid and that, moreover, $[H(A)]_c$ has the structure

$$[H(A)]_c = \sum_b {}' \sum_{\nu_b \in A} \Delta_{c,b} h_c(b, \nu_b), \tag{3.15}$$

where $h_c(b,\nu_b)$ is a c-connected operator which vanishes if $c \not\supset b$.

We can now see that (3.12) can be expressed in the form

$$H(A) = H_a(A) + V^a(A), \tag{3.16}$$

where the operator

$$V^a(A) = \sum_c \bar{\Delta}_{a,c} [H(A)]_c, \tag{3.17}$$

involves only the parts of $H(A)$ external to the partition a. This is because

$$\bar{\Delta}_{a,b} \equiv 1 - \Delta_{a,b}, \tag{3.18}$$

vanishes unless $a \not\supset b$. For $A = A_0$, (3.16) becomes the usual partition-degenerate decomposition of the full Hamiltonian

$$H = H_a + V^a, \quad (\text{any } a), \tag{3.19}$$

where

$$V^a = \sum_b \bar{\Delta}_{a,b} [V]_b. \tag{3.20}$$

$V^a(A)$ represents the interaction among the clusters of a and thus it generates the full N-particle dynamics appropriate to A. If

$$V^a(A) = 0, \quad \text{for some } a \neq 1, \tag{3.21}$$

then the RM does not contain interactions linking all N particles under all circumstances. The significance of the characteristic partition $\hat{a}(A)$ is that we always have

$$V^{\hat{a}(A)} = 0. \tag{3.22}$$

But (3.22) takes on content [cf. (3.21)] only if $\hat{a}(A) \neq 1$ since $V^1(A) \equiv 0$. Thus, if the RM is trivial the full N-particle dynamics has been lost in any formalism for which $V^a(A)$ represents the interactions for the truncated problem.

On the other hand, suppose that for some $a \neq 1$, Eq. (3.21) is satisfied. Then, it follows from Eqs. (3.15) and (3.17) that for any $c \not\subset a$ there are no partitions $b \subset c$ for which $\nu_b \in A$. Thus, if $d \in \mathscr{A}$, then necessarily $d \subset a$ and therefore $\hat{a}(A) \subset a$. So, if (3.21) is satisfied, the RM is necessarily trivial. Henceforth, we assume that A is nontrivial.

It is important to emphasize again that the preceding characterization of a RM as well as the designations trivial and nontrivial reaction mechanisms are nonunique and formalism-dependent. In alternative approaches either of the incomplete standard variety or of the full N-particle type[6] the notions of Ref. 4 concerning reaction mechanisms are not necessarily relevant.

IV. PARTITION PERMUTING ARRAY EQUATIONS

We next consider the truncated scattering problem which is stated in terms of the RM Green functions[4]

$$G(A) \equiv [z - H(A)]^{-1}, \tag{4.1}$$

$$G_a(A) \equiv [z - H_a(A)]^{-1}, \tag{4.2}$$

where z is a complex parametric energy. We consistently suppress the dependence upon z except in Appendix C. It is important to note that $G(A)$ and $G_a(A)$ are defined on the entire N-particle Hilbert space. This is made explicit by the spectral resolution[4]

$$G_a(A) = P_a(A) G_a + Q_a(A)/z, \tag{4.3}$$

where we denote the complement of the projection $P_a(A)$ as

$$Q_a(A) = I - P_a(A). \tag{4.4}$$

RM transition operators $T^{(\pm)}_{a,b}(A)$ can be defined in a manner consistent with the interpretation of $G(A)$ and $G_a(A)$ by

$$G(A) = G_a(A) + G_a(A)T^{(+)}_{a,b}(A)G_b(A) \tag{4.5a}$$

$$= G_b(A) + G_a(A)T^{(-)}_{a,b}(A)G_b(A). \tag{4.5b}$$

When a and b are stable partitions we have on shell with $\nu_\alpha, \nu_b \in A$

$$\langle \phi_a(\nu_a)|G_a(A)^{-1}G(A)G_b(A)^{-1}|\phi_b(\nu_b)\rangle = \langle \phi_a(\nu_a)|G_a^{-1}G(A)G_b^{-1}|\phi_b(\nu_b)\rangle, \tag{4.6}$$

and also

$$\langle \phi_a(\nu_a)|G_a^{-1}G(A)G_b^{-1}|\phi_b(\nu_b)\rangle = \langle \phi_a(\nu_a)|T^{(\pm)}_{a,b}(A)|\phi_b(\nu_b)\rangle. \tag{4.7}$$

Equation (4.7) clearly relates $T^{(\pm)}_{a,b}(A)$ to the residues of the same pole singularities which appear in the untruncated Green function G.

We find with the aid of the resolvent identities

$$G(A) - G_a(A) = G_a(A)V^a(A)G(A) = G(A)V^a(A)G_a(A), \tag{4.8}$$

that, e.g.,

$$T^{(+)}(A) = \overline{V}(A)\mathcal{S} + \overline{V}(A)\mathcal{S}G(A)\overline{V}(A), \tag{4.9}$$

where here we have introduced, for the sake of brevity, a matrix notation[2,10] in the partition indices with $T^{(\pm)}(A) = (T^{(\pm)}_{a,b}(A))$, $\overline{V}(A) = (V^a\delta_{a,b})$ and $\mathcal{S}_{a,b} = 1$ for all a,b. We note that $G(A)$ has no matrix structure so that $[G(A),\mathcal{S}] = 0$. The expression (4.9) suggests the introduction of partition coupling schemes $\mathcal{V}(A)$ which satisfy

$$\overline{V}(A)\mathcal{S} = \mathcal{V}(A)\mathcal{S}, \tag{4.10}$$

or, in component form,

$$V^b(A) = \sum_a \mathcal{V}^{b,a}(A). \tag{4.10'}$$

The array $\mathcal{V}^{b,a}(A)$ is, in general, not diagonal. The significant aspect of (4.9) and (4.10) is that $T^{(+)}(A)$ as given by (4.9) is invariant under the replacement (4.10). One sees that the resolvent identities (4.8) can also be written in a form which is invariant with respect to the substitution (4.10), e.g.,

$$G(A)\mathcal{S} = \widehat{G}(A)\mathcal{S} + \widehat{G}(A)\overline{V}(A)\mathcal{S}G(A), \tag{4.11}$$

where $\widehat{G}(A) = (G_a(A)\delta_{a,b})$. One can then deduce[2,10] the coupled integral equations satisfied by the transition operators $T^{(+)}_{b,a}(A)$

$$T^{(+)}(A) = \mathcal{V}(A)\widehat{G}(A)\mathcal{S}\widehat{G}(A)^{-1} + \mathcal{V}(A)\widehat{G}(A)T^{(+)}(A). \tag{4.12}$$

We confine our attention, henceforth, to the case where

$$\mathcal{V}(A) = \overline{V}(A)W(A). \tag{4.13}$$

Here $W(A)$ is a *partition-coupling array*[5] which is a numerical matrix in the partition indices with defining property

$$\mathcal{P}W(A)\mathcal{S} = \mathcal{P}\mathcal{S}, \tag{4.14}$$

where $\mathcal{P}$ is a projector on an arbitrary set of partitions. Equations (4.12) and (4.13) require that (4.14) be valid for $\mathcal{P} = I$, and hence for all projectors. When $\mathcal{P} \neq I$, (4.14) can be used, together with some additional constraints on $W(A)$, to simplify Eq. (4.12).

Equation (4.14) is a relatively weak condition and it is known[5,6] that $W(A)$ must be severely restricted in order that (4.12) possess connected kernels after a finite number of iterations. We let $\mathcal{B}(A)$ denote the set of all two-cluster partitions contained in $\mathcal{A}$ and $\mathcal{P}_2(A)$ be the projector on this set.[6] We require that $W(A)$ satisfy

$$W(A)\mathcal{P}_2(A) = W(A). \tag{4.15}$$

If (4.12) is multiplied on the left by $\mathcal{P}_2(A)$ and we use (4.14) in the special case $\mathcal{P} = \mathcal{P}_2(A)$ we obtain a set of equations which couple together only those operators $T^{(+)}_{\beta,\alpha}(A)$ with $\alpha,\beta \in \mathcal{B}(A)$, viz.,

$$T^{(+)}_{\beta,\alpha}(A) = V^\beta(A)\sum_{\lambda\in\mathcal{B}(A)} W_{\beta,\lambda}(A)G_\lambda(A)[G_\alpha(A)^{-1} + T^{(+)}_{\lambda,\alpha}(A)]. \tag{4.16}$$

All of the indices in (4.16) are restricted to $\mathcal{B}(A)$ and in this sense they are minimally coupled.[2,10] In the special case that A is a trivial RM we have $T^{(+)}_{\beta,\alpha}(A) \equiv 0$. We remark that $\mathcal{P}_2(A)W(A) \neq W(A)$ so that the operators $T^{(+)}_{b,\alpha}(A)$, e.g., with $b \notin \mathcal{B}(A)$ but $\alpha \in \mathcal{B}(A)$ can be determined in terms of the solution of (4.16) since we can use (4.16) as a quadrature rule in such cases.

When A is not trivial we want (4.16) to be a set of connected-kernel equations in the sense that after a finite number of iterations of (4.16) the kernels of the resultant equations are connected operators. Previous experience[5,6] suggests that we choose $W_{\beta,\alpha}(A)$, with $\beta,\alpha \in \mathcal{B}(A)$, to be a *partition permuting array* (PPA) in order to achieve this.

It is useful to review briefly the concept of a PPA.[5] Let us imagine some pertinent set [say $\mathcal{B}(A)$ or $\mathcal{B}(A_0)$] of two-cluster partitions to be ordered in some definite but arbitrary fashion α_i, $i = 1,\ldots,N_2$, where N_2 is the number of partitions in this set. A PPA over this set is then defined as

$$W_{\alpha_i,\beta} = \delta_{\widehat{\alpha}_{i+1},\beta}, \tag{4.17}$$

where

$$\widehat{\alpha}_{i+1} = \alpha_{i+1}\overline{\delta}_{i,N_2} + \alpha_1\delta_{i,N_2}, \tag{4.18}$$

and $\overline{\delta}_{i,j} = 1 - \delta_{i,j}$. The structure (4.17) of the PPA ensures that in a product of k operators, each of which has the form $U^\alpha W_{\alpha,\beta}$, no index appears more than once in any intermediate summations and successive U^α's in this product have different indices which cycle through the set:

$$[UW\cdots UW]_{\alpha_i,\beta} = U^{\alpha_i}U^{\alpha_{i+1}}\cdots U^{\alpha_{i+k-1}}\delta_{\alpha_{i+k},\beta}, \tag{4.19}$$

if $k < N_2 + 1 - i$. If $k > N_2 + 1 - i$, we have a repetition of indices on the U's as one moves to the right of the product but only after all of the indices in the set have been enumerated. This property is crucial for connectivity considerations.

Now from (4.8) it follows that $G_a(A)$ possesses the cluster decomposition

$$G_a(A) = \sum_a \Delta_{a,b}[G(A)]_b. \tag{4.20}$$

This is because the right-hand sides of (4.8) contain only sums of terms of connectivities $b \not\subseteq a$. Here we have used the

cluster decompositions of Sec. III as well as the theorem[4,7] that the product of an a-connected and a b-connected operator is $(a\cup b)$-connected. This theorem also allows us to deduce that when $W(A)$ is a PPA, the kernel of Eqs. (4.16) can be placed into the form

$$U^{\beta}(A) \equiv V^{\beta}(A) W_{\beta,\lambda(\beta)}(A) G_{\lambda(\beta)}(A)$$

$$= \sum_{b} \bar{\Delta}_{\beta,b} \{u(\beta)\}_{b}, \tag{4.21}$$

where

$$\{u(\beta)\}_{b} \equiv \sum_{c} \Delta_{\lambda(\beta),c} [H(A)]_{b} [G(A)]_{c}, \tag{4.22}$$

and $\lambda(\beta) \in \mathscr{B}(A)$ is *the* partition for which $W_{\beta,\lambda(\beta)} = 1$. From (4.22) we see that $\{u(\beta)\}_{b}$ is a sum of operators each with connectivities $b \cup c \supseteq b$, and $b \in \mathscr{A}$.

Now let $\mathscr{B}(A,b)$ be the subset of $\mathscr{B}(A)$ which consists of those two-cluster partitions for which $\beta \supseteq b$. Then, if $\mathscr{B}(A)$ contains $N_2(A)$ elements, $N_2(A)$ iterations of (4.16) yields equations with the typical kernel

$$\{U^{\beta}(A) \cdots U^{\gamma}(A)\} U^{\beta}(A)$$

$$= \sum_{b} \bar{\Delta}_{\beta,b} \{U^{\beta}(A) \cdots U^{\gamma}(A)\} \{u(\beta)\}_{b}. \tag{4.23}$$

The indices of the operators within the curly brackets in (4.23) contain a nonrepeating enumeration of all the elements of $\mathscr{B}(A)$ and thus of $\mathscr{B}(A,b)$ for any $b \in \mathscr{A}$. It follows by Lemma 2 of Appendix A that each term in the sum on the right-hand side of (4.23) is a connected operator. Thus, we see that the integral equations (4.16) for the truncated scattering problem acquire connected kernels after a finite number of iterations.

We reach our principal conclusion that the general reaction theory of Ref. 4 can be realized in the form of the type of scattering equations developed in Refs. 5. Equations (4.16) constitute an alternative to the RM integral equations proposed in Ref. 4. Both sets of equations possess distinctive features; however, Eqs. (4.16) have two properties which are worthy of special note. First, they are minimally coupled among only the two-cluster partitions relevant to the RM and second they possess a more or less conventional inhomogeneous term structure.

The results of Ref. 4 concerning the RM equations which do not depend the choice of a particular channel coupling scheme [cf. (4.10)] carry over without substantial modification to the present analysis. We have especially in mind the proof in Ref. 4 of the discontinuity relations satisfied by the RM transition operators $T^{(+)}_{a,b}(A)$ and the resultant restricted unitarity relations. This can also be established for the theory defined by Eqs. (4.12)–(4.16). In this proof, which is outlined in Appendix C, Eqs. (4.14) are needed to define $T^{(+)}_{a,b}(A)$ when the partitions a and b are not both contained in $\mathscr{B}(A)$. In order to do this, we identify $\mathscr{P}$ in (4.14) with, e.g., the projector on the entire set $\mathscr{A}$, while still retaining properties (4.15) and (4.17) on $\mathscr{B}(A)$. This sort of procedure is a characteristic of unitary proofs for PPA equations,[11] and it is a consequence of the minimally coupled nature of (4.16). This again emphasizes one of the principal attributes of minimally coupled equations, namely that the remaining operators $T^{(+)}_{a,b}(A)$ can be calculated by what amount to quadrature rules in terms of the operators $T^{(+)}_{\alpha,\beta}(A)$, with $\alpha,\beta \in \mathscr{B}(A)$.

We conclude that Eqs. (4.14)–(4.16) define a unitary theory with respect to the set of transition amplitudes which are appropriate to the RM A. The precise nature of these RM unitarity relations and the consequent RM optical theorem are elaborated upon more fully in Ref. 4 and in Appendix C. We remark that these unitarity relations further support the identification of the operators $T^{(+)}_{a,b}(A)$ with transition operators of the truncated theory.

Finally, we note that for practical calculations involving reactions initiated from the two-cluster state $|\phi_{\alpha}(\nu_{\alpha})\rangle$ the inhomogeneous term in (4.16) can be simplified by the following modification of the Lippmann identity:

$$G_{\lambda}(A) G_{\alpha}(A)^{-1} |\phi_{\alpha}(\nu_{\alpha})\rangle = \delta_{\lambda,\alpha} |\phi_{\alpha}(\nu_{\alpha})\rangle. \tag{4.24}$$

Equation (4.24) is interpreted as a half-on-shell equation which is valid for $\lambda,\alpha \in \mathscr{A}$ only. In this connection we note that $G_{\beta}(A) = z^{-1}$ if $\beta \notin \mathscr{A}$. However, as pointed out in Ref. 12, Eq. (4.24) must be used with care in the inhomogeneous term of scattering integral equations of the generic form

$$T = B + KT. \tag{4.25}$$

The solution of (4.25) can be written as

$$T = \sum_{p=0}^{N-1} (K)^{p} B + (1 - K)^{-1} (K)^{N} B, \tag{4.26}$$

where $N \geqslant 0$ is an integer. In the case of (4.16) we have, using a matrix notation restricted to the set $\mathscr{B}(A)$,

$$(K)^{p} B$$
$$= \bar{V}(A) [W(A) \hat{G}(A) \bar{V}(A)]^{p} W(A) \{\hat{G}(A) \mathscr{S} \hat{G}(A)^{-1}\}. \tag{4.27}$$

The central question[12] regarding the applicability of (4.24) when (4.27) operates half-on-shell on $|\phi_{\alpha}(\nu_{\alpha})\rangle$ is whether there are terms to the left of the curly brackets in (4.27) which in combination with $G_{\beta}(A)$ can generate an α-connected pole singularity which cancels the vanishing of $G_{\beta}(A) G_{\alpha}(A)^{-1} |\phi_{\alpha}(\nu_{\alpha})\rangle$ when $\beta \neq \alpha$. For finite p such terms could only arise from the Green functions which appear only in the combinations $G_{\lambda}(A) V^{\lambda}(A)$. The α-connected parts of operators of the form $(\beta \neq \alpha)$

$$G_{\lambda}(A) V^{\lambda}(A) \cdots G_{\nu}(A) V^{\nu}(A) G_{\beta}(A),$$

consist of a sum of products of the b-connected parts of $G(A)$ with $b \subset \alpha$ but $b \neq \alpha$ and a finite number of potential terms $[V(A)]_{c}$, $c \subseteq \alpha$. Such products lack the infinite number of interaction terms contained in α but not in any of the b's needed to generate the required singularity.[12] For N large enough so that $(K)^{N}$ is connected there is no further chance for cancellation when (4.26) operates on the states $|\phi_{\alpha}(\nu_{\alpha})\rangle$ and we conclude by the results of Ref. 12 that the use of the Lippmann identity in connection with (4.16), and the usual BKLT equations[5] in particular, is legitimate.

V. EMBEDDING EQUATIONS

The statement of an approximation procedure is complete only if it is known how the approximate quantities $(T^{(+)}_{a,b}(A))$ are related to their exact counterparts $(T^{(+)}_{a,b})$.

This relationship is useful, however, only if it suggests a means of systematically improving the approximation. Both of these attributes are realized in the form of a connected-kernel integral equation for $T^{(+)}_{a,b}$ in terms of the RM transition operators and the part $H(A')$ of the total Hamiltonian not included in $H(A)$:

$$H(A') \equiv H - H(A). \tag{5.1}$$

Such an integral equation is termed an *imbedding equation*.

In Ref. 4 it is shown that

$$\begin{aligned}\langle \phi_a(\nu_a)|T^{(+)}_{a,b}|\phi_b(\nu_b)\rangle &= \langle \phi_a(\nu_a)|T^{(+)}_{a,b}(A)|\phi_b(\nu_b)\rangle \\ &\quad + \langle \phi_a(\nu_a)|G_a(A)^{-1}G(A)\widehat{T}G(A)G_b(A)^{-1}|\phi_b(\nu_b)\rangle,\end{aligned} \tag{5.2}$$

where $\widehat{T}$ is an operator which satisfies the connected-kernel integral equation

$$\widehat{T} = zH(A')G(A') + H(A')G(A')H(A)G(A)\widehat{T}, \tag{5.3}$$

and $G(A') \equiv [z - H(A')]^{-1}$. Equation (5.2) is essentially a two-potential-type connection between $T^{(+)}_{a,b}$ and $T^{(+)}_{a,b}(A)$. It represents an extreme form of an imbedding relationship in the sense that the integral equation (5.3) does not involve $T^{(+)}_{a,b}(A)$ except implicitly through $G(A)$.

We consider next an alternative imbedding equation similar to that developed in Ref. 6. We observe that [cf. Eq. (4.12)]

$$\mathscr{P}_2(A)T^{(+)}(A) = \Lambda(A)\overline{\mathscr{V}}(A)\widehat{G}(A)\mathscr{S}\widehat{G}(A)^{-1}, \tag{5.4}$$

where

$$\Lambda(A) = \mathscr{P}_2(A_0) + \overline{\mathscr{V}}(A)\widehat{G}(A)\Lambda(A), \tag{5.5}$$

$$\overline{\mathscr{V}}(A) = \mathscr{P}_2(A)\overline{V}(A)\overline{W}(A). \tag{5.6}$$

Here, $\overline{W}(A) = \mathscr{P}_2(A)W(A)\mathscr{P}_2(A)$ is a matrix with elements $\overline{W}_{\alpha,\beta}(A) = W_{\alpha,\beta}(A)$ for $\alpha,\beta \in \mathscr{B}(A)$ and is zero otherwise with $W_{\alpha,\beta}(A)$ defined by (4.17)–(4.18) over the $N_2(A)$ elements of $\mathscr{B}(A)$. It is more convenient to work with $\Lambda(A)$ rather than $T^{(+)}(A)$ because of the simpler inhomogeneous term. In the limiting case $A = A_0$ we have

$$\Lambda = \mathscr{P}_2(A_0) + \overline{\mathscr{V}}(A_0)\overline{G}\Lambda, \tag{5.7}$$

where we have set $\Lambda = \Lambda(A_0)$.

$\Lambda(A)$ differs from $\mathscr{P}_2(A_0)$ only on the set $\mathscr{B}(A)$. Thus Eq. (5.5) is minimally coupled on $\mathscr{B}(A)$. It is important to note that given $\Lambda(A)$ the transition operators $T^{(+)}_{\alpha,\beta}(A)$ for $\alpha,\beta \in \mathscr{B}(A)$ are determined via (5.4). The remaining transition operators $T^{(+)}_{a,b}(A)$, with $a,b \in \mathscr{A}$, are then obtained using (4.16) as quadrature rules. This allows us to restrict the imbedding problem to the expression of Λ in terms of $\Lambda(A)$. This relationship is easily found from Eqs. (5.5) and (5.7) to be the integral equation

$$\Lambda = \Lambda(A) + \Lambda(A)[\overline{\mathscr{V}}(A_0)\widehat{G} - \overline{\mathscr{V}}(A)\widehat{G}(A)]\Lambda. \tag{5.8}$$

Equation (5.8) is our alternative imbedding equation and bears an even more striking resemblence to the two-potential formula. It is proved in Appendix B that (5.8) possesses a connected-kernel after a finite number of iterations.

VI. SUMMARY

The recently proposed approximate reaction theory of Polyzou and Redish[4] has been applied to the problem of deriving approximate connected-kernel scattering integral equations of the partition-permuting-array[5] type. We show that the condition for obtaining such equations is that all the two-cluster partitions which are contained in the set of all partitions relevant to a reaction mechanism A are coupled by the partition permuting array $W(A)$. This is proven with the aid of a lemma which we establish in Appendix A which refers to the connectivity of typical operator products occurring in PPA theories.

The approximate PPA equations which we derive have two advantages over the approximate but differently coupled equations found in Ref. 4. First, they couple together only two-cluster-partition-labeled channels and, second, the inhomogeneous terms of these equations have the usual (Born) structure.

We also obtain two forms of imbedding equation. These are equations which imbed the approximation in the exact theory and which can also be used to generate more refined approximations as well. One of these forms is the connected-kernel imbedding equation of Ref. 4 while the other is similar to the equation derived in Ref. 6. In the latter case we demonstrate that the imbedding equation has a kernel which becomes connected after a finite number of iterations.

In conclusion we make two remarks concerning the applicability of the present work. First, we note that although our derivation of the PPA equations relies heavily upon the reaction theory of Ref. 4, our discussion is easily modified to include and/or develop other approximate reaction theories such as that contained in Ref. 6. Second, it is not yet clear whether any of the approximate connected-kernel equations developed here and in Ref. 4 and 6 are optimal from the point of view of practical calculation of nuclear reactions. The determination of the practical calculational characteristics of such approximate reaction theories is one of the most pressing problems in many-particle scattering theory at the present time.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant No. PHY 78-26595.

APPENDIX A

We generalize some previous arguments[5,6,13] regarding the connectedness of the products of the types of operators which appear in PPA theories. These results are used in the text.

For a fixed partition b, let $\mathscr{B}(b)$ represent the set of all two-cluster partitions $\alpha \supseteq b$. Then

Lemma 1: If $d \not\subseteq \alpha$ for all $\alpha \in \mathscr{B}(b)$, then $d \cup b = 1$.

Proof: Assume that $d \cup b \neq 1$. Then there is at least one two-cluster partition β such that $d \cup b \subseteq \beta$. Thus $d \subseteq \beta$ and $\beta \in \mathscr{B}(b)$ which is a contradiction and establishes the lemma.

Now let α be a two-cluster partition and call

$$\mathscr{O}^{\alpha} \equiv \sum_{\bar{a}} \overline{\Delta}_{\alpha,a}[\mathscr{O}^{\alpha}]_{\bar{a}(a)}, \tag{A1}$$

where $[\mathscr{O}^{\alpha}]_{\bar{a}(a)}$ is an $\bar{a}$-connected operator with $\bar{a} \supseteq a$. Evidently a product of operators of this type,

$$\mathcal{O}^{\alpha}\cdots\mathcal{O}^{\gamma}=\sum_{a,\ldots,g}\bar{\Delta}_{\alpha,a}\cdots\bar{\Delta}_{\gamma,g}[\mathcal{O}^{\alpha}]_{\bar{a}(a)}\cdots[\mathcal{O}^{\gamma}]_{\bar{g}(g)}, \tag{A2}$$

is a sum of $(\bar{a}\cup\cdots\cup\bar{g})$-connected operators. Clearly $(\bar{a}\cup\cdots\cup\bar{g})\subseteq\alpha,\ldots,\gamma$. Thus, it follows that:

Lemma 2: If $[B]_b$ is a b-connected operator and the operators $\mathcal{O}^{\alpha}$ are defined as in (A1), then the product $\mathcal{O}^{\alpha}\cdots\mathcal{O}^{\gamma}[B]_b$ of operators (in any order) is connected if $\alpha,\ldots,\gamma$ includes an enumeration of all the two-cluster partitions in $\mathscr{B}(b)$.

APPENDIX B

We prove here that the imbedding equation (5.8) has a connected kernel after a finite number of iterations. In Eq. (5.8) all operators are defined on the set of all two-cluster partitions $\mathscr{B}(A_0)$. For this reason, all matrix equations in this Appendix are to be regarded as defined on the set $\mathscr{B}(A_0)$.

First, we consider the appropriate form (5.5) which after $N_2(A)\equiv n$ iterations becomes by the arguments of Sec. IV,

$$A(A)=\mathscr{P}_2(A_0)+\sum_{k=1}^{n-1}[\bar{V}(A)\bar{W}(A)\hat{G}(A)]^k+A_c(A), \tag{B1}$$

where $A_c(A)$ is a fully-connected operator and we recall that $\bar{W}(A)$ is the coupling array with elements $\bar{W}_{\alpha,\beta}(A)=W_{\alpha,\beta}(A)$ for $\alpha,\beta\in\mathscr{B}(A)$. Then we see that the kernel, K, of Eq. (5.8) can be written in the form

$$K=K^{\text{disc}}+K_c, \tag{B2}$$

where the fully connected part K_c is proportional to $A_c(A)$. The remainder, K^{disc}, is not necessarily connected and has the structure

$$K^{\text{disc}}=A^{\text{disc}}(A)\bar{\mathscr{U}}(A), \tag{B3}$$

where $A^{\text{disc}}(A)$ is the disconnected part of $A(A)$ and $[\bar{W}=\bar{W}(A_0)]$

$$\bar{\mathscr{U}}(A)\equiv U(A_0)\bar{W}-\bar{U}(A)\bar{W}(A). \tag{B4}$$

We have used a matrix version of the notation introduced in (4.21). Specifically, we have $[U(A_0)]_{\beta,\alpha}=U^{\beta}(A_0)\delta_{\beta,\alpha}$ and $[\bar{U}(A)]_{\beta,\alpha}=\bar{U}^{\beta}(A)\delta_{\beta,\alpha}$ with $\bar{U}^{\beta}(A)$ given by (4.21) but with $\bar{W}_{\beta,\lambda(\beta)}(A)$ replaced by $\bar{W}_{\beta,\lambda(\beta)}(A)$. We also use the notation $\mathscr{U}^{\beta}(A)=U^{\beta}(A_0)-\bar{U}^{\beta}(A)$. We observe that $\bar{U}(A)\bar{W}(A)$ is nonzero only on $\mathscr{B}(A)$ while $U(A_0)\bar{W}$ is nonzero on the set of all two-cluster partitions.

Let us consider an enumeration of all the two-cluster partitions in $\mathscr{B}(A_0)$, namely, $\alpha_1,\ldots,\alpha_n,\alpha_{n+1},\ldots,\alpha_{N_2}$, such that the first n partitions are contained in $\mathscr{B}(A)$. We choose [cf. (4.17), (4.18)]

$$W(A)_{\alpha_i,\beta}=\delta_{\hat{\alpha}_{i-1}(A),\beta}, \quad i\leqslant n, \tag{B5a}$$

where

$$\hat{\alpha}_{i+1}(A)=\alpha_{i+1}\bar{\delta}_{i,n}+\alpha_1\delta_{i,n}, \tag{B5b}$$

and

$$W(A)_{\alpha_i,\beta}=0, \quad i>n. \tag{B5c}$$

Also, we take

$$W(A_0)_{\alpha_i,\beta}=W(A)_{\alpha_i,\beta}, \quad i<n, \qquad =\delta_{\hat{\alpha}_{i+1},\beta}, \quad i\geqslant n, \tag{B6a}$$

where

$$\hat{\alpha}_{i+1}=\alpha_{i+1}\bar{\delta}_{i,N_2}+\alpha_1\delta_{i,N_2}. \tag{B6b}$$

Then

Lemma 1: $\{\bar{\mathscr{U}}(A)^{N_2}\}_{\alpha_i,\beta}$ is either fully connected or zero.

Proof: If $i<n$ the product develops a string of operators from $\mathscr{U}^{\alpha_i}$ to $\mathscr{U}^{\alpha_{n-1}}$. But $\mathscr{U}^{\alpha_n}(A)=\mathscr{U}^{\alpha_n}(A_0)-\bar{U}^{\alpha_n}(A)$. So the product is really the sum of two strings, one which continues out to $U^{\alpha_{N_2}}$ and then back to $\mathscr{U}^{\alpha_{i-1}}$ and enumerates the N_2 partitions of $\mathscr{B}(A_0)$, and one which goes back to U^{α_1} and up through the cycle again but enumerates the n partitions of $\mathscr{B}(A)$ plus N_2-n more factors one of which can be taken to be $\bar{U}^{\alpha_n}(A)$. The proof then follows using Lemma 2 of Appendix A. For the first string of terms the relevant $[B]_b$ operator in Lemma 2 has $b=0$ while for the second the argument used in connection with Eqs. (4.23) is appropriate. The product $[\bar{\mathscr{U}}(A)^{N_2}]_{\alpha_i,\beta}$ is zero if β is not consistent with the evolution of either string. The proof for $i\geqslant n$ involves only a cyclic permutation of the preceding argument.

Now

$$A^{\text{disc}}(A)=I+\sum_{k=1}^{n-1}[\bar{U}(A)\bar{W}(A)]^k. \tag{B7}$$

Thus

$$K^{\text{disc}}=\bar{\mathscr{U}}(A)+\hat{U}(A), \tag{B8}$$

where

$$\hat{U}(A)=\sum_{k=1}^{n-1}\{[\bar{U}(A)\bar{W}(A)]^k\}\bar{\mathscr{U}}(A). \tag{B9}$$

We note that

$$[\hat{U}(A)]_{\alpha_i,\beta}=0, \quad \text{for } i>n, \tag{B10}$$

and so

$$\{(K^{\text{disc}})^{N_2}\}_{\alpha_i,\beta}=\{[\bar{\mathscr{U}}(A)]^{N_2}\}_{\alpha_i,\beta}, \quad \text{for } i>n. \tag{B11}$$

If $i\leqslant n$ then

$$\{(K^{\text{disc}})^{N_2}\}_{\alpha_i,\beta}=\{[\bar{\mathscr{U}}(A)]^{N_2}\}_{\alpha_i,\beta}+R_{\alpha_i,\beta}, \tag{B12}$$

where $R_{\alpha_i,\beta}$ is a sum of products of p $\hat{U}(A)$'s and q $\bar{\mathscr{U}}(A)$'s, where $N_2\geqslant p\geqslant 1$ and $N_2-1\geqslant q\geqslant 0$. Each of these products contains an enumeration of all the partitions in $\mathscr{B}(A)$ and each contains at least one term $\hat{U}(A)$ of connectivity $b\neq 0$. Thus by Lemma 2 of Appendix A and Lemma 1 we have:

Lemma 2 $[(K^{\text{disc}})^{N_2}]_{\alpha_i,\beta}$ is either fully connected or zero. From (B2) and the preceding lemma we see that the kernel of the imbedding equation (5.7) becomes connected after N_2 iterations. These results are similar to those of Ref. 6.

APPENDIX C

We consider some of the pertinent aspects of the truncated unitarity relations satisfied by the model transition operators $T^{(+)}_{a,b}(A,z)$. These operators are solutions of Eqs. (4.16) when a and b are restricted to the set of two-cluster partitions $\mathscr{B}(A)$; the extension to arbitrary partitions in $\mathscr{A}_0$

is defined by Eqs. (4.12)–(4.15). It will be necessary for us to indicate the parametric energy z explicitly at times; we use the notation $T^{(+)}_{a,b}(A, \pm)$ to represent the limiting cases where $z = E \pm i0$.

In Ref. 4 it is shown in detail and with careful attention to the singularity structure of $T^{(+)}_{a,b}(A, \pm)$ that the model transition operators which are defined by (4.9) satisfy a truncated unitarity relation which is appropriate to the reaction mechanism A. In particular, the unitarity sum is of the form

$$-2\pi i \sum_{d\in\mathscr{A}} T^{(+)}_{a,d}(A,+)\,\delta[E - H_d(A)] \times P_d(\nu_d)[T^{(+)}_{b,d}(A,+)]^\dagger, \tag{C1}$$

and involves only those channels within the RM A. We remark that $a,b\in\mathscr{A}_0$ in general although only the cases when $a,b\in\mathscr{A}$ are of actual interest.

A special choice for $\mathscr{V}(A)$ is employed in Ref. 4 within the integral equations (4.12) for $T^{(+)}(A, \pm)$. These particular integral equations are used for some aspects of the proof of the unitarity relations. This should not obscure the fact that the results of Ref. 4 apply to any set of operators which are given by (4.9). Thus, we can exploit all of the results of Ref. 4 in regard to unitarity in connection with *any* set of operators defined by Eqs. (4.12), provided the solutions of these equations are unique, which we suppose is ensured by the connected-kernel property, because Eqs. (4.12) follow from (4.9). The formal recovery of (4.9) from (4.12) is not obvious but follows with the help of a Green function matrix $\mathscr{G}(A)$ which is defined by

$$\mathscr{G}(A) = \widehat{G}(A) + \widehat{G}(A)\mathscr{V}(A)\mathscr{G}(A) \tag{C2a}$$

$$= \widehat{G}(A) + \mathscr{G}(A)\mathscr{V}(A)\widehat{G}(a). \tag{C2b}$$

Then

$$T^{(+)}(A) = \mathscr{V}(A)\mathscr{G}(A)\mathscr{S}\widehat{G}(A)^{-1}. \tag{C3}$$

The matrices involved in Eqs. (C2), (C3) are defined with respect to partition indices over the entire set $\mathscr{A}_0$. If we multiply (C2b) on the left by $\widehat{G}(A)^{-1}\mathscr{S}G(A)$ we obtain[2]

$$\mathscr{G}(A)\mathscr{S} = G(A)\mathscr{S} = \mathscr{S}G(A), \tag{C4}$$

where we recall that $G(A)$ has no matrix structure. Combining (C3) and (C4) we obtain

$$T^{(+)}(A) = \overline{V}(A)G(A)\mathscr{S}\widehat{G}(A)^{-1}, \tag{C5}$$

from which (4.9) follows if we use (4.8).

Equations (4.12)–(4.18) define a set of PPA integral equations for the operators $T^{(+)}_{a,b}(A)$ for all $a,b\in\mathscr{A}_0$. These integral equations have kernels which become connected after a finite number of iterations and so we presume that this implies uniqueness. The solutions of these equations are given by (4.9) and therefore they satisfy the unitarity relations derived in Ref. 4.

[1]P.E. Hodgson, *Nuclear Reactions and Nuclear Structure* (Oxford U.P., London, 1971). In accordance with the scope of this book we include elastic, inelastic breakup, and rearrangement, collisions under the general designation of nuclear reactions.

[2]K.L. Kowalski, in *Few-Body Systems and Nuclear Forces*, edited by H. Zingl, M. Haftel, and H. Zankel (Springer-Verlag, Berlin, 1978), Vol. II, p. 393 and references cited therein.

[3]E.F. Redish, in *Modern Three-Hadron Physics*, edited by A.W. Thomas (Springer-Verlag, Berlin, 1977), p. 181.

[4]W.N. Polyzou and E.F. Redish, Ann. Phys. (N.Y.), **119**, 1 (1979).

[5]M. Baer and D.J. Kouri, J. Math. Phys. **14**, 1637 (1973); W. Tobocman, Phys. Rev. C **9**, 2466 (1974); D.J. Kouri and F.S. Levin, Phys. Lett. B **50**, 42 (1974). Some of these works used the terminology *channel* permuting array. In the N-particle case this is a misnomer.

[6]R. Goldflam and W. Tobocman, Phys. Rev. C **20**, 904 (1979).

[7]W. Polyzou, J. Math. Phys. **21**, 306 (1980).

[8]K.L. Kowalski, W. Polyzou, and E.F. Redish (to be published).

[9]T. Donnellan, *Lattice Theory* (Pergamon, New York, 1968).

[10]K.L. Kowalski, Phys. Rev. C **16**, 7 (1977).

[11]P. Benoist-Gueutal, Phys. Lett. B **56**, 413 (1975); D.J. Kouri, F.S. Levin and W. Sandhas, Phys. Rev. C **13**, 1825 (1976); K.L. Kowalski, Phys. Rev. C **15**, 42 (1977).

[12]P. Benoist-Gueutal, M. L'Huillier, E.F. Redish, and P.C. Tandy, Phys. Rev. C **17**, 1924 (1978); M. L'Huillier, E.F. Redish, and P.C. Tandy, J. Math. Phys. **19**, 1276 (1978).

[13]Y. Avishai, Nucl. Phys. A **161**, 621 (1971).

A multigroup criticality condition for a space-independent multiplying assembly via the Chapman–Kolmogorov equation

I. Aviram and S. Goshen
Nuclear Research Center-Negev, P. O. Box 9001, Beer-Sheva, Israel

(Received 19 June 1979; accepted for publication 23 August 1979)

We consider the multigroup time-dependent probability distribution of neutrons in a space-independent, source-free multiplying assembly. The first-order partial differential equation for the probability generating function can be derived from the forward Chapman–Kolmogorov equations of the stochastic process, and solved—in the principle—by the method of characteristics. It is well known that the asymptotic extinction probability is related to the singular point of the system of characteristic equations. An analysis of the location of this point in the hyperspace of dummy variables associated with neutrons of various velocity groups leads to a multigroup criticality condition for the assembly.

1. INTRODUCTION

Space-independent stochastic models of nuclear reactors have been used in the past by many authors[1-8] to describe the probability distribution of neutrons, and delayed-neutrons precursors, and/or of various moments thereof. These models mainly served the purpose of investigating the velocity-independent reactor kinetics in the presence of a weak source. The usual procedure was to set up the Chapman–Kolmogorov equations of the corresponding stochastic process, to construct the partial differential equation for the probability generating function which can be solved—at least in principle—by the method of characteristic trajectories. Knowledge of the probability generating function enables one to obtain through successive differentiation the complete probability distribution and the moments. This, however, can be done explicitly only in a rather limited number of simple cases[9-11] (e.g., one-velocity, binary fission). Bell[6] has shown how one can derive the asymptotic properties of the distribution by investigating the singular points of the system of first-order differential equations of the characteristic trajectories; the complete solution is not needed for this purpose.

In this note we derive the stochastic equations for the probability distribution as a function of time in a space-independent (point) multiplying assembly with time-independent, velocity-dependent cross sections. This set of equations to our knowledge, has not been previously presented in the literature. The main result of this work is a criticality condition for such a multiplying system, based on an analysis of the singular points of the system of characteristic equations. The analysis was partly inspired by Bell's work.

2. BASIC EQUATIONS

Our basic equations are standard Chapman–Kolmogorov forward equations for the set of probabilities $p(\mathbf{n};t) = p(n_1,...,n_g,...,n_G;t)$ of finding at time t in the system n_1 neutrons in the velocity group 1,...,n_G neutrons in group G, given that a set of initial conditions have been specified at $t = 0$. As we shall assume a source-free assembly, at least one neutron of some kind must be present at $t = 0$. The neutrons can undergo down-scattering (slowing down) from group g to group g' described by a triangular matrix of normalized scattering cross sections $\Lambda_{gg'}$, $g = 1,2,...,G$, $g' = g,...,G$, such that $\Lambda_g^s = \Sigma_{g'=g}^G \Lambda_{gg'}$; they can be captured or leak away from the medium, the normalized cross section for group g being Λ_g^c; finally, they may induce fission with a normalized cross section of Λ_g^f. The cross sections satisfy: $\Lambda_g^s + \Lambda_g^c + \Lambda_g^f = 1$, for every g. Let τ_g be the mean lifetime of a neutron in group g. Let c_j be the probability that j neutrons of any velocity are produced in a fission event. We shall assume for simplicity that the distribution c_j is independent of the velocity of the neutron which has induced the fission event. We have the normalization $\Sigma_{j=0}^J c_j = 1$, where usually $J = 5$. The fission spectrum is specified by the numbers η_g, giving the probability of a fission neutron being born in group g. They satisfy $\Sigma_{g=1}^G \eta_g = 1$. It will be sometimes convenient to use the vector notation $\boldsymbol{\eta} = (\eta_1,...,\eta_G)$. Let $\mathbf{j} = (j_1, j_2,..., j_G)$ be a G-dimensional vector whose components are natural numbers or zero, and let $\sigma(\mathbf{j}) = j_1 + \cdots + j_G$ denote the sum of the components. Denote by $w_j(\mathbf{j})$ the probability that from among j neutrons born in a fission event, j_1 are born into group 1,..., j_G are born into group G. It is easy to see that,

$$w_j(\mathbf{j}) = \eta_1^{j_1}\binom{j}{j_1}\eta_2^{j_2}\binom{j_2 - j_1}{j_2}\cdots\eta_{G-1}^{j_{G-1}} \times\binom{j - j_1 - \cdots - j_{G-2}}{j_{G-1}}\eta_G^{j_G}, \tag{1}$$

and it satisfies the normalization condition

$$\sum_{\sigma(\mathbf{j})=j} w_j(\mathbf{j}) = (\eta_1 + \cdots + \eta_G)^j = 1, \tag{2}$$

where the summation is taken over all possible partitions of the integer j into G integer terms j_g such that $0 \leqslant j_g \leqslant j$, repetitions being allowed.

The differential equation for $p(\mathbf{n};t)$ is obtained by expressing $p(\mathbf{n};t+dt)$ in terms of $p(\mathbf{n};t)$ through enumeration of all the possible events that can occur during the interval dt. Neglecting terms of order $(dt)^2$ and higher, dividing by dt, and passing to the limit $dt \to 0$, we find:

 0022-2488/80/071897-05$1.00

$$\frac{dp(\mathbf{n};t)}{dt} = -\sum_{g=1}^{G}\frac{n_g}{\tau_g}p(\mathbf{n};t) + \sum_{g=1}^{G}\frac{n_g+1}{\tau_g}\Lambda_g^c\, p(\mathbf{n}+\boldsymbol{\delta}_g;t) + \sum_{g=1}^{G}\sum_{g'=g}^{G}\frac{n_g+1-\delta_{gg'}}{\tau_g} \times \Lambda_{gg'}\, p(\mathbf{n}+\boldsymbol{\delta}_g-\boldsymbol{\delta}_{g'};t) + \sum_{g=1}^{G}\sum_{j=0}^{J}\sum_{\sigma(\mathbf{j})=j}\frac{n_g+1-j_g}{\tau_g} \times \Lambda_g^f c_j\, w_j(\mathbf{j})\, p(\mathbf{n}+\boldsymbol{\delta}_g-\mathbf{j};t), \tag{3}$$

where $\delta_{gg'}$ is the Kronecker's delta and $\boldsymbol{\delta}_g = (0,0,...,1,...,0)$ is a vector with all zero components, except for a one in the gth position. There is one Eq. (3) for every conceivable vector $\mathbf{n}$.

The first term on the RHS in Eq. (3) is related to the event: "Nothing happens during interval dt;" the second term accounts for capture and leakage; the third term accounts for transfer among velocity groups; finally, the fourth term accounts for fission events.

We now introduce the probability generating function $F(\mathbf{x};t)$ where $\mathbf{x} = (x_1,...,x_G)$:

$$F(\mathbf{x};t) = \sum_{n_1=0}^{\infty}\cdots\sum_{n_G=0}^{\infty} x_1^{n_1}\cdots x_{x_G}^{n_G}\, p(\mathbf{n};t). \tag{4}$$

The normalization condition of the distribution $p(\mathbf{n};t)$ is equivalent to

$$F(\mathbf{1};t) = 1. \tag{5}$$

It follows that the series (4) converge uniformly in the hypercube $-1 \leqslant x_1, x_2, ..., x_G \leqslant 1$. The probabilities $p(\mathbf{n};t)$ are given in the usual way:

$$p(\mathbf{n};t) = \frac{1}{n_1!\cdots n_G!}\left(\frac{\partial^{n_1+\cdots+n_G}F}{\partial x_{x_1}^{n_1}\cdots\partial x_G^{n_G}}\right)_{\mathbf{x}=\mathbf{0}}. \tag{6}$$

In particular, the extinction probability is

$$p(\mathbf{0};t) = F(\mathbf{0};t). \tag{7}$$

Multiplying Eq. (3) by $x_1^{n_1}\cdots x_G^{n_G}$ and summing over $n_1,...,n_G$ from 0 to ∞, we find that $F(\mathbf{x};t)$ satisfies the equation

$$\frac{\partial F}{\partial t} = \sum_{g=1}^{G}\frac{H_g(\mathbf{x})}{\tau_g}\frac{\partial F}{\partial x_g}, \tag{8}$$

where

$$H_g(\mathbf{x}) = -x_g + \Lambda_g^c + \sum_{g'=g}^{G}\Lambda_{gg'}\, x_{g'} + \Lambda_g^f\,\psi(\mathbf{x}) \tag{9}$$

and

$$\psi(\mathbf{x}) = \sum_{j=0}^{J} c_j(\boldsymbol{\eta}\cdot\mathbf{x})^j. \tag{10}$$

Evidently, $\psi(\mathbf{1}) = 1$, and for any t,

$$H_g(\mathbf{1}) = 0, \quad g = 1,2,...,G. \tag{11}$$

$F(\mathbf{x};t)$ is a nondecreasing function of x_g $(g = 1,...,G)$, and $0 \leqslant F \leqslant 1$ in the hypercube $0 \leqslant x_1,...,x_G \leqslant 1$. In order to solve Eq. (8) we shall assume that F is known at time $t = 0$:

$$F(\mathbf{x};0) = f(\mathbf{x}) \leqslant 1. \tag{12}$$

If, for example, one neutron of group g is present at time $t = 0$, then $f(\mathbf{x}) = x_g$; or, if one fission neutron is present at $t = 0$, then $f(\mathbf{x}) = \boldsymbol{\eta}\cdot\mathbf{x}$, etc. The function f must obviously satisfy $f(\mathbf{1}) = 1$. Equation (8) is a first-order partial differential equation, and may be solved by the method of characteristic trajectories. If we define the functions of time $x_g(t)$, $g = 1,...,G$, such that

$$\dot{x}_g \equiv \frac{dx_g}{dt} = -\frac{H_g(\mathbf{x})}{\tau_g}, \quad g = 1,2,...,G, \tag{13}$$

then the system (13) specifies a family of trajectories in the $(G+1)$-dimensional space of $(\mathbf{x};t)$. The trajectories never intersect each other, each one of them being specified, say, by the point $\mathbf{x}(0) = \boldsymbol{\xi}$, whose components $(\xi_1,...,\xi_G)$ are given as initial conditions. Along any trajectory the function F satisfies the ordinary differential equation $dF/dt = 0$, and hence $F(\mathbf{x},t) = \text{const.}$, where the constant is obtained by evaluating $f(\mathbf{x})$ at the point $\boldsymbol{\xi}(\xi_1,...,\xi_G)$ where the particular trajectory intersects the plane $t = 0$. The quantities H_g are evidently functions of time, through the intermediate of $\mathbf{x}$.

3. ANALYSIS OF SINGULAR POINTS

By Eq. (11), the point $\mathbf{x} = \mathbf{1}$ in the G-dimensional space is a singular point of the system (13), and the trajectory emerging from this point is a "straight line" parallel to the t-axis. It can be shown that this is a saddle point (see, e.g., Ref. 12, Chap. II). We show subsequently that there is only one other singular point $\mathbf{x} = \mathbf{x}_0$ for which

$$H_g(\mathbf{x}_0) = 0, \quad g = 1,...,G, \tag{14}$$

whose components are real, positive numbers, either all less than one, or all bigger than one, when the physical system is supercritical or subcritical, respectively. Indeed, let us assume that $0 < x_{01}, x_{02}, ..., x_{0G} < 1$. The trajectory emerging from $\mathbf{x}_0$ is a "straight line" parallel to the t axis. Figure 1 illustrates the situation for $G = 2$. The curves $H_1(\mathbf{x}) = 0$ and $H_2(\mathbf{x}) = 0$ divide the unit square into four regions in which dx_1/dt and dx_2/dt conserve their signs. The characteristic trajectories thus form a divergent bundle around the trajectory corresponding to $\mathbf{x} = \mathbf{x}_0$. Such a point is called an unstable two-tangent node. After a sufficiently long time, t_∞, only those trajectories having emerged in the immediate neighborhood of $\mathbf{x}_0$ will still be found within the boundaries of the unit square. This means that, with the exception of the point $\mathbf{x} = \mathbf{1}$,

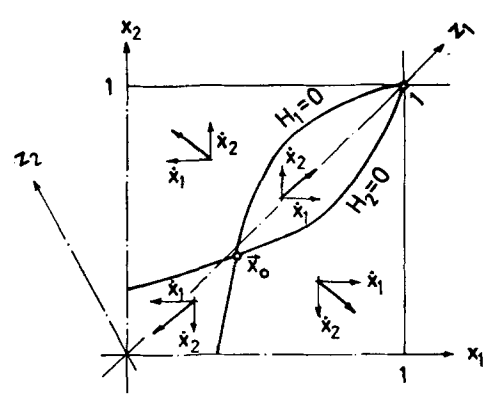


FIG. 1. The curves $H_1 = 0$, and $H_2 = 0$. The singular point $\mathbf{x}_0$ is located in the unit square when the system is supercritical.

$$F(\mathbf{x},t>t_\infty)\simeq F(\mathbf{x}_0,0)=f(\mathbf{x}_0)=\text{const} \tag{15}$$

throughout the unit square. This argument can be directly extended to any G: Just replace the words "unit square" by "unit hypercube." Equation (15) shows that only the free term $p(\mathbf{0},t)$ of the series (4) survives after t_∞. Thus, the asymptotic extinction probability is

$$p(\mathbf{0},t>t_\infty)=f(\mathbf{x}_0), \tag{16}$$

and is less than 1 if and only if $\mathbf{x}_0$ is within the unit hypercube. The asymptotic probability of having a persisting chain reaction $1-p(\mathbf{0},t>t_\infty)$ will then be finite, and this is only possible in a supercritical assembly. Due to the monotonicity of $f(\mathbf{x})$, the closer $\mathbf{x}_0$ is to $\mathbf{1}$, the larger is the extinction probability. If both singular points coincide $\mathbf{x}_0=\mathbf{1}$, the system is just critical. If $\mathbf{x}_0$ is outside the unit hypercube, the system is subcritical. (A similar argument has been presented by Bell[6] for prompt neutrons and delayed neutrons' precursors.) Let us mention that although negative or complex roots of the system (4) may exist, no physical meaning can be attached to them.

4. CRITICALITY CONDITION

In order to analyze the conditions under which $\mathbf{x}_0$ lies within the unit hypercube it is convenient first to transform to a new set of variables $\mathbf{z}(z_1,z_2,\ldots,z_G)$ defined by:

$$\begin{aligned} z_1 &= \boldsymbol{\eta}\cdot\mathbf{x}=\eta_1 x_1+\cdots\eta_G x_G\,,\\ z_g &= -x_1+x_g\,,\quad g=2,3,\ldots,G. \end{aligned} \tag{17}$$

It is also convenient to write the function ψ [Eq. (10)] in the form

$$\psi(\mathbf{x})=\sum_{j=0}^{J} b_j(\boldsymbol{\eta}\cdot\mathbf{x}-1)^j, \tag{18}$$

where the b_j are related to the moments of the distribution c_k:

$$b_j=\frac{1}{j!}\sum_{k=0}^{J} k(k-1)\cdots(k-j+1)\,c_k\,. \tag{19}$$

In particular $b_0=1$, and $b_1\equiv\nu$ is the familiar average number of neutrons born in a fission event.

Let us write

$$\psi(\mathbf{x})=1+\nu(\boldsymbol{\eta}\cdot\mathbf{x}-1)+\phi(\mathbf{x}), \tag{20}$$

where $\phi(\mathbf{x})$ includes the terms of (18) from $j=2$ onwards. Inverting the system (17), and substituting everything in the system $H_g(\mathbf{x})=0$, we find

$$(\gamma_g-1)(z_1-1)+\Lambda_g^f\phi(z_1)+\sum_{g'=2}^{G} a_{gg'}z_{g'}=0,\qquad g=1,2,\ldots,G, \tag{21}$$

where the numbers $a_{gg'}$ form a rectangular $G\times(G-1)$ matrix, defined by

$$a_{gg'}=\eta_{g'}(1-\Lambda_g^s)+\Lambda_{gg'}-\delta_{gg'},\qquad g=1,2,\ldots,G\quad g'=2,3,\ldots,G, \tag{22}$$

and the quantity

$$\gamma_g=\Lambda_g^s+\nu\Lambda_g^f \tag{23}$$

may be called "average number of secondaries per collision in group g." Each equation of the system (21) represents a $(G-1)$-dimensional manifold (a hypersurface). A solution of the system is a point common to all hypersurfaces. The main advantage of the transformation (17) is that only one variable, namely z_1, enters nonlinearly in the equations of system (21).

Let us now fix the value of $z_1=Z$. This equation represents a hyperplane orthogonal to the z_1 axis. Substitute $z_1=Z$ into (21). Each of the resulting equations represents a $(G-2)$-dimensional manifold (a linear hypercurve) lying on the hyperplane $z_1=Z$:

$$(\gamma_g-1)(Z-1)+\Lambda_g^f\phi(Z)+\sum_{g'=2}^{G} a_{gg'}z_{g'}=0,\qquad g=1,\ldots,G. \tag{24}$$

Let us introduce the following notation:

$$D_G[\mathbf{h}]=\begin{vmatrix} h_1 & a_{12} & \cdots & a_{1G}\\ h_2 & a_{22} & \cdots & a_{2G}\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ \cdot & \cdot & & \cdot\\ h_G & a_{G2} & \cdots & a_{GG}\end{vmatrix} \tag{25}$$

is the determinant of a square $G\times G$ matrix formed by bordering the rectangular matrix $a_{gg'}$ with a vector column $\mathbf{h}$. The necessary condition that the linear equations have a common solution is then

$$D_G[(\boldsymbol{\gamma}-\mathbf{1})(Z-1)+\boldsymbol{\Lambda}^f\phi(Z)]=0,$$

or

$$P(Z)=(Z-1)\,D_G[\boldsymbol{\gamma}-\mathbf{1}]+\phi(Z)\,D_G[\boldsymbol{\Lambda}^f]=0, \tag{26}$$

where $P(Z)$ is a polynomial of degree J in Z. We have

$$P(1)=0, \tag{27}$$

$$P(0)=D_G[\boldsymbol{\Lambda}^c+c_0\,\boldsymbol{\Lambda}^f], \tag{28}$$

$$P'(1)=D_G[\boldsymbol{\gamma}-\mathbf{1}], \tag{29}$$

$$P'(0)=-D_G[\boldsymbol{\Lambda}^c+(1-c_1)\,\boldsymbol{\Lambda}^f], \tag{30}$$

$$P''(Z)=\phi''(Z)\,D_G[\boldsymbol{\Lambda}^f]. \tag{31}$$

Clearly, $\phi''(Z)>0$, for $Z\geqslant 0$.

In order to examine the properties of the determinant D_G, it will be convenient to extend the array $a_{gg'}$ so as to include a_{g1}, $g=1,2,\ldots,G$. The extended array now forms a square matrix of order G. The following relations are easily verified:

$$\sum_{g'=1}^{G} a_{gg'}=0\quad(\text{all } g), \tag{32}$$

$$a_{gg'}>0\quad(\text{all } g'\neq g), \tag{33}$$

$$a_{gg}<0\quad[\text{follows from (32), (33)}]. \tag{34}$$

These properties allow us to establish an important result:

$$(-1)^{G-1}D_G[\mathbf{h}]>0 \tag{35}$$

as long as $\mathbf{h}$ is a column vector of nonnegative numbers (a sufficient condition). The proof of (35) is given in the Appendix.

Denoting $\bar{P}(Z)=(-1)^{G-1}P(Z)$, and using (35), it is seen that $\bar{P}(0)>0$, $\bar{P}'(0)<0$, and $\bar{P}''(Z)>0$ for $Z\geqslant 0$. It fol-

lows that the polynomial $\bar{P}(Z)$ has, in addition to $Z=1$, a single positive root $Z=z_0$ which is located as follows:

$$0<z_0\leqslant 1, \quad \text{iff } \bar{P}'(1)=(-1)^{G-1}D_G\,[\gamma-\mathbf{1}]\geqslant 0, \tag{36}$$

$$z_0>1, \quad \text{iff } \bar{P}'(1)=(-1)^{G-1}D_G\,[\gamma-\mathbf{1}]<0. \tag{37}$$

Let us now assume that $0<z_0\leqslant 1$. The function ψ in Eq. (9) is then

$$\psi=\sum_{j=0}^{J} c_j z_0^j \leqslant 1.$$

If we examine the equations of the system (14) $H_g(\mathbf{x}_0)=0$ in succession starting with $g=G$, we find: For $g=G$:

$$x_{0G}(1-\Lambda_{GG})=\Lambda^{\,c}_{G}+\Lambda^{\,f}_{G}\psi,$$

or since $\Lambda_{GG}=\Lambda^{\,s}_{G}=1-\Lambda^{\,c}_{G}-\Lambda^{\,f}_{G}$, we have

$$0<x_{0G}=(\Lambda^{\,c}_{G}+\Lambda^{\,f}_{G}\psi)/(\Lambda^{\,c}_{G}+\Lambda^{\,f}_{G})\leqslant 1. \tag{38}$$

Now, suppose that $0<x_{0G},x_{0G-1},\ldots,x_{0g+2},x_{0g+1}\leqslant 1$. Then

$$x_{0g}(1-\Lambda_{gg})=\sum_{g'=g+1}^{G} x_{g'}\Lambda_{gg'}+\Lambda^{\,c}_{g}+\Lambda^{\,f}_{g}\psi.$$

But

$$\Lambda_{gg}=\Lambda^{\,s}_{g}-\sum_{g'=g+1}^{G}\Lambda_{gg'}=1-\sum_{g'=g+1}^{G}\Lambda_{gg'}-\Lambda^{\,c}_{g}-\Lambda^{\,f}_{g},$$

whence

$$0<x_{0g}=\frac{\Sigma_{g'=g+1}^{G}x_{g'}\Lambda_{gg'}+\Lambda^{\,c}_{g}+\Lambda^{\,f}_{g}\psi}{\Sigma_{g'=g+1}^{G}\Lambda_{gg'}+\Lambda^{\,c}_{g}+\Lambda^{\,f}_{g}}\leqslant 1. \tag{39}$$

It is thus proved by complete backward induction that all x_{0g} satisfy $0<x_{0g}\leqslant 1$, i.e., $\mathbf{x}_0$ lies inside the unit hypercube, a necessary condition for this being $0<z_0\leqslant 1$. It can be shown in exactly the same way that all $x_{0g}>1$, provided that $z_0>1$, and no singular points exist such that some of the x_{0g} are less than one, and the others are larger than one. The inequality (36) can thus be regarded as a criticality condition for the physical system under consideration.

ACKNOWLEDGMENT

The authors express their thanks to J. Avron for very helpful discussions.

APPENDIX

Let A be the $G\times G$ matrix with elements $a_{gg'}$ defined by (22), including $g'=1$:

$$A=\begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1G} \\ a_{21} & a_{22} & \cdots & a_{2G} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{G1} & a_{G2} & \cdots & a_{GG} \end{Vmatrix}. \tag{A1}$$

Let A_{rr} be the principal minor of order $G-1$ corresponding to the diagonal element a_{rr}. All principal[13] minors of order $G-1$ are strictly diagonally dominant:

$$|a_{gg}|>\sum_{\substack{g'=1\\ g'\neq g,\,g'\neq r}}^{G}|a_{gg'}|, \quad g=1,\ldots,r-1,r+1,\ldots,G, \tag{A2}$$

a result which follows directly from Eqs. (32) and (33).

Lemma 1: The sign of all principal cofactors of A is $(-1)^{G-1}$.

Proof: From relation (A2) it follows that all principal minors of order $G-1$ are nonsingular by Theorem 1.8 of Ref. 14. The corresponding determinants (principal cofactors) are clearly nonzero. Moreover, let us continuously alter all off-diagonal elements of a principal minor of order $G-1$ until they vanish. The minor remains strictly diagonally dominant, and its determinant will not change sign, for it never vanishes in the process. The sign of any principal cofactor therefore will be that of the product of its diagonal elements, i.e., $(-1)^{G-1}$.

Lemma 2: All cofactors of A corresponding to the elements of a row, are equal.

Proof: The determinant of A vanishes because of Eq. (32). Let us expand the determinant of A in terms of the elements of row g. Then

$$a_{g1}A_{g1}+\cdots+a_{gG}A_{gG}=0. \tag{A3}$$

Here A_{gr} denotes the cofactor of the element a_{gr}. Add ϵ to the element a_{gr}, and subtract ϵ from the element a_{gs}. Since relation (32) still holds for all rows after this alteration, we have

$$a_{g1}A_{g1}+\cdots+(a_{gr}+\epsilon)A_{gr}+\cdots +(a_{gs}-\epsilon)A_{gs}+\cdots+a_{gG}A_{gG}=0. \tag{A4}$$

Subtracting Eq. (A4) from Eq. (A3), we find that $A_{gr}=A_{gs}$. Hence all cofactors of a row g are equal.

Theorem: Let A be the $G\times G$ matrix of Eq. (A1), whose elements $a_{gg'}$ satisfy the relations (32) and (33). Then all cofactors (principal or not) have the same sign $(-1)^{G-1}$.

Proof: By Lemma 2, all cofactors of row g are equal to A_{gg}. In particular, they all have the sign $(-1)^{G-1}$. Combining this result with Lemma 1 concludes the proof of the theorem.

In order to establish the result of Eq. (35) we replace in A the first column of elements a_{g1}, $g=1,\ldots,G$, by a column vector $\mathbf{h}(h_1,\ldots,h_G)$ whose elements are all nonnegative, and expand the corresponding determinant in terms of the elements of the first column. We then have

$$\operatorname{sign}(h_1A_{11}+h_2A_{21}+\cdots+h_GA_{G1})=(-1)^{G-1} \tag{A5}$$

or

$$(-1)^{G-1}D_G\,[\mathbf{h}]>0.$$

[1]R.P. Feynman, F. De Hoffman, and R. Serber, J. Nucl. Energy **3**, 64 (1956).
[2]L.I. Pal, Proc. 2nd Intern. Conf. Peaceful Uses Atomic Energy, Geneva **16**, 687 (1958).
[3]G.E. Hansen, Nucl. Sci. Eng. **8**, 709 (1960).
[4]H. Hurwitz, Jr., G.E. Res. Lab. Rep. No. 60-RL (2405E) (1960).
[5]H. Hurwitz, Jr., D.B. MacMillan, J.H. Smith, and M.L. Storm, Nucl. Sci. Eng. **15**, 166 (1963); and **15**, 187 (1963).

[6]G.I. Bell, Ann. Phys. **21**, 243 (1963).
[7]G.I. Bell, W.A. Anderson, and D. Galbraith, Nucl. Sci. Eng. **16**, 118 (1963).
[8]D.B. MacMillan, Nucl. Sci. Eng. **39**, 329 (1970).
[9]F. Norelli, V.M. Jorio, and N. Pacilio, J. Nucl. Energy **2**, 67 (1975).
[10]D.R. Harris, "Stochastic Fluctuations in a Power Reactor," USAEC Report WAPD-TM-190, Westinghouse Electric Corp. (1958).
[11]J. Lewins, Atomkernenergie **23**, 172 (1974).
[12]G. Sansone and R. Conti, *Nonlinear Differential Equations* (Pergamon, New York, 1964).
[13]As a matter of fact, principal minors of lower order too are strictly diagonally dominant, but only the result for order $G-1$ will be needed in the proof.
[14]R.S. Varga, *Matrix Interative Analysis* (Prentice-Hall, Englewood Cliffs, N.J., 1962).

Shift-operator techniques for the classification of multipole-phonon states. I. Properties of shift operators in the R(5) group

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

(Received 10 October 1979; accepted for publication 28 December 1979)

With a view to obtaining an orthogonal solution to the state labeling problem of the nuclear quadrupole-phonon states, independent operators which shift the eigenvalues of the R(3) Casimir operator L^2 are constructed. Expressions which connect quadratic products of these shift operators with R(5) invariants are given.

1. INTRODUCTION

In a set of papers[1,2] shift operators were constructed out of the enveloping algebra of the generators l_i $(i = 0, \pm)$ of R(3) and the components $T(j,\mu)$, $\mu = -j,...,j$ of an arbitrary $(2j+1)$ dimensional tensor representation of R(3). The commutators of the $T(j,\mu)$ with the generators $l_0, l_\pm$ of R(3) are supposed to be of the standard form:

$$[l_\pm, T(j,\mu)] = [(j \mp \mu)(j \pm \mu + 1)]^{1/2} T(j, \mu \pm 1), \quad (1.1)$$

$$[l_0, T(j,\mu)] = \mu T(j,\mu), \quad (1.2)$$

where the positive real value of the square root has been taken into account. The constructed shift operators, denoted by O_l^k, $k = -j,...,j$, change the l value, where $l(l+1)$ is the eigenvalue of the R(3) Casimir operator L^2, by k, without altering the eigenvalue of l_0. A general analysis of the operators O_l^k for arbitrary j values has been given by Hughes and Yadegar.[2] They are generally defined as follows[2]:

$$O_l^k = \gamma_0^k(l,m) Q_0 + \sum_{\mu=1}^{j} [\gamma_\mu^k(l,m) Q_{+\mu} + (-1)^{j+k} \gamma_\mu^k(l,-m) Q_{-\mu}], \quad (1.3)$$

where for $\mu = 0,...,j$ and $k \geqslant 0$

$$\gamma_\mu^k(l,m) = (-1)^{j+3l-m} \{(j+k)!(j-k)!(2l+j+k+1)! \times (l-m-\mu)!(l+m+k)!(l-m+k)!/ (2j)!(2l-j+k)!(l+m+\mu)!((l-m)!)^2\}^{-1/2} \times \begin{pmatrix} j & l & l+k \\ \mu & -\mu-m & m \end{pmatrix} \quad (1.4)$$

and

$$Q_{\pm\mu} = T(j, \mp\mu) l_\pm^\mu, \quad \mu \geqslant 0. \quad (1.5)$$

On the other hand the expression for O_l^{-k} $(k \geqslant 0)$ follows also from Eq. (1.3) due to the relation

$$O_l^{-k} = O_{-(l+1)}^k. \quad (1.6)$$

The symbol in the large parentheses in Eq. (1.4) represents a Wigner 3-j symbol.

In the present paper we apply Hughes' general theory to the R(5) group, whose symmetric representations play an important role in the classification of the nuclear quadrupole-phonon states. The group generators are defined in terms of the creation $(b_{2\mu}^+)$ and annihilation $[(-1)^\mu b_{2-\mu}]$ operators of these phonon states by Weber *et al.*[3] Here we shall be concerned with a little different form for these generators, such that they fulfill the commutation relations (1.1) and (1.2); i.e.,

$$l_0 = \sqrt{10}(b_2^+ b_2)_0^1, \quad l_\pm = \mp 2\sqrt{5}(b_2^+ b_2)_{\pm 1}^1,$$
$$T(j=3,\mu) \equiv q_\mu = (b_2^+ b_2)_\mu^3, \quad \text{with} \quad \mu = -3,...,3. \quad (1.7)$$

The q_μ form a seven-dimensional irreducible tensor representation with respect to the R(3) subgroup of R(5), generated by l_0 and $l_\pm$ which are defined by (1.7).

The R(5) group possesses two Casimir operators, respectively of second- and fourth-order in the generators. The second-order one provides us with the seniority quantum number. For the symmetric irreducible representations of interest the fourth-order invariant is not linearly independent of the second-order one and by this of no use for the labeling of the degenerated l states.[4] In a previous paper[5] the authors have deduced the general form of the fourth-order R(3) scalar operator which can play a fundamental role in the orthogonal specification of the quadrupole-phonon states. Moreover they have proved that the O_l^0 operator (1.3) belongs to that class of operators. The apparatus for obtaining its eigenvalues in an analytic way consists of the l-shift operators which we now like to study.

2. EXPLICIT FORMS FOR THE SHIFT OPERATORS

Since in deriving the general expression (1.3) no use was made of the commutation relations of the $T(j,\mu)$ among themselves, it is in fact not assumed that the l_i and $T(j,\mu)$ generate a group. Therefore the explicit forms of the shift operators will be valid not for the R(5) group only, but for each group having a set of generators forming a seven-dimensional tensor representation of R(3).

[a)] Aangesteld Navorser bij het Nationaal Fonds voor Wetenschappelijk Onderzoek, Belgium.

 0022-2488/80/071902-04$1.00

The explicit forms of the shift operators O_l^k follow immediately from Eqs. (1.3)–(1.6) by introducing $j = 3$. The analytic expressions tabulated by Yutsis *et al.*[6] have been used. The following results were obtained:

$$O_l^0 = -(2/\sqrt{5})[3l(l+1) - 5l_0 - 1\ ]q_0 l_0 - (3/5)^{1/2}[l(l+1) - 5l_0(l_0+1) - 2\ ]q_{-1}l_+ + \sqrt{6}(l_0+1)q_{-2}l_+^2 + q_{-3}l_+^3 + (3/5)^{1/2}[l(l+1) - 5l_0(l_0-1) - 2]\ q_{+1}l_- + \sqrt{6}(l_0-1)q_{+2}l_-^2 - q_{+3}l_-^3, \tag{2.1}$$

$$O_l^{+1} = -(2/\sqrt{5})(l^2+2l-5l_0^2)(l-l_0+1)(l+l_0+1)q_0 - (1/\sqrt{15})(l^2-10ll_0-15l_0^2-3l - 25l_0-10)(l-l_0+1)q_{-1}l_+ + (2/3)^{1/2}(l+3l_0+4)(l-l_0+1)q_{-2}\ l_+^2 + (l-l_0+1)q_{-3}l_+^3 - (1/\sqrt{15})(l^2+10ll_0-15l_0^2-3l+25l_0-10)(l+l_0+1)q_{+1}l_- + (2/3)^{1/2}(l-3l_0+4)(l+l_0+1)q_{+2}\ l_-^2 + (l+l_0+1)q_{+3}l_-^3, \tag{2.2}$$

$$O_l^{+2} = 2\sqrt{5}\ l_0(l+l_0+2)(l+l_0+1)(l-l_0+2)(l-l_0+1)q_0 + (5/3)^{1/2}(l+3l_0+3)(l+l_0+2)(l-l_0+2)(l-l_0+1)q_{-1}1_+ + (2/\sqrt{6})(2l+3l_0+6)(l-l_0+2)(l-l_0+1)q_{-2}\ l_+^2 + (l-l_0+2)(l-l_0+1)q_{-3}\ l_+^3 - (5/3)^{1/2}(l-3l_0+3)(l-l_0+2)(l+l_0+2)(l+l_0+1)q_{+1}l_- - (2/\sqrt{6})(2l-3l_0+6)(l+l_0+2)(l+l_0+1)q_{+2}\ l_-^2 - (l+l_0+2)(l+l_0+1)q_{+3}\ l_-^3, \tag{2.3}$$

$$O_l^{+3} = 2\sqrt{5}(l+l_0+3)(l+l_0+2)(l+l_0+1)(l-l_0+3)(l-l_0+2)(l-l_0+1)q_0 + \sqrt{15}(l+l_0+3)(l+l_0+2)(l-l_0+3)(l-l_0+2)(l-l_0+1)q_{-1}l_+ + \sqrt{6}(l+l_0+3)(l-l_0+3)(l-l_0+2)(l-l_0+1)q_{-2}\ l_+^2 + (l-l_0+3)(l-l_0+2)(l-l_0+1)q_{-3}\ l_+^3 + \sqrt{15}(l+l_0+3)(l+l_0+2)(l+l_0+1)(l-l_0+3)(l-l_0+2)q_{+1}l_- + \sqrt{6}(l+l_0+3)(l+l_0+2)(l+l_0+1)(l-l_0+3)q_{+1}\ l_-^2 + (l+l_0+3)(l+l_0+2)(l+l_0+1)q_{+3}\ l_-^3, \tag{2.4}$$

$$O_l^{-k} = O_{-(l+1)}^{+k} \quad (k = 1,2,3). \tag{2.5}$$

The O_l^{+k} operators are "one-side" operators; they act as shift operators upon states to the right and not upon states to the left. They are assumed to act upon eigenvectors of L^2 and l_0, which will be denoted by the kets $|l,a_l,m\rangle$ where m is the eigenvalue of l_0 and a_l represents the other labels for their unique specification. Note that O_l^0 is invariant under the replacement of l by $-(l+1)$. Therefore O_l^0 is an R(3) scalar operator. The internal structure of l_i and q_μ in terms of the quadrupole-phonon creation and annihilation operators [see Eq. (1.7)] does not play a role in the construction of these shift operators. This means that the expressions (2.1)–(2.5) for the shift operators remain valid for every group having amongst its generators a subset of generators of the type (1.7). In a later paper, where we plan to study the labeling problem of the octupole-phonon states, this specific property will be fully considered.

3. THE PRODUCT OPERATORS AND THEIR MUTUAL RELATIONS

With the aid of the introduced O_l^k various scalar R(3) operators which of course commute with L^2 and l_0 can be constructed. The ones with which we shall be concerned will be of the type $O_{l+k}^{-k}O_l^{+k}$. In his study of the SU(3) group Hughes[1] also introduces product operators of the form $O_{l\pm1}^{\mp1}O_{l\pm2}^{\mp1}O_l^{\pm2}$, etc., consisting of three shift operators. In the present study it is very complicated and in practice not possible to derive expressions for such triple product operators. However, in the following paper it will be clear that such forms are not needed for the derivation of the eigenvalues of O_l^0. Because of their scalar character the quadratic product operators must be expressible in terms of the other available scalar operators, i.e., L^2, l_0, O_l^0 and the R(5) second-order Casimir operator defined by[3]

$$V^* = -\tfrac{1}{10}L^2 - \sum_{\mu=-3}^{3}(-1)^\mu q_\mu q_{-\mu}, \tag{3.1}$$

having an eigenvalue of $-\frac{1}{2}v(v+3)$.

The explicit expressions of these product operators in terms of the q_μ, $l_\pm$, and l_0 generators involve the l and m values of the states upon which they act. In order, however to considerably simplify their calculation we have restricted our considerations to the case when they act on states of zero m value. Let us further remark that when the rightest operator in the product works on a $|l,a_l,m=0\rangle$ state, it creates a state proportional to $|l+k,b_{k+l},m=0\rangle$, since the O_l^{+k} does not affect the m quantum number. This has a consequence that the leftest operator also acts upon a $m=0$ state. Since the study of these product forms is closely related to the derivation of the O_l^0 eigenvalues, which are independent of m,[2] this seemingly drastic condition will not seriously detract from the generality of subsequent calculations. The expressions for the product operators are even with the above simplification rather tedious to derive. We shall only briefly sketch how to construct them, without giving the results, but

more attention will be paid to the specific relations which mutually yield between them.

The quadratic product operators $O_{l+k}^{-k} O_l^{+k}$ ($k = -3, -2, \ldots, +2, +3$) consist of terms composed of two q_μ and six or less l_i operators. In order to reach a one to one relation between the before-mentioned scalar operators it is clear that all operators should be brought, into a so-called standard form. Therefore in the expressions of the product operators the two q_μ generators have been shifted in each term to the extreme left-hand side by using the commutation relations (1.1). Secondly the q_μ's are placed in such a way that the absolute value of the μ index of the one on the far left is not smaller than those of the one on the far right. To perform that operation explicit use has been made of the commutation relations between the several components of the q operator. The relations are summarized in the Appendix. After these manipulations three distinct kinds of terms occur in the expressions of the $O_{l+k}^{-k} O_l^{+k}$ scalar operators:

(i) terms quadratic in the q operators, i.e.,

$q_\mu q_\nu l_+^\alpha l_-^\beta$, with $|\mu| \geqslant |\nu|$ and $\mu + \nu = \beta - \alpha$;

(ii) terms linear in the q operators, i.e.,

$q_\mu l_+^\alpha l_-^\beta$, with $\mu = \beta - \alpha$;

(iii) terms independent of the q operators, i.e.,

$l_+^\alpha l_-^\beta$ with $\alpha = \beta$.

Taking into account that under the adopted condition ($m = 0$), $L^2 = l_+ l_- = l_- l_+$, the l_i factors in these terms can be simplifed as follows:

(a) If $\alpha = \beta$:

$$l_+^\alpha l_-^\alpha = l_-^\alpha l_+^\alpha = l(l+1)[l(l+1)-2]\cdots \times[l(l+1)-(\alpha-1)(\alpha-2)] \times[l(l+1)-\alpha(\alpha-1)];$$

(b) If $\alpha \neq \beta$: (1) $\alpha > \beta$

$$l_+^\alpha l_-^\beta = l_+^{\alpha-\beta}(l_+^\beta l_-^\beta),$$
$$l_-^\alpha l_+^\beta = l_-^{\alpha-\beta}(l_-^\beta l_+^\beta),$$

while the terms between rounded brackets are of the type (a).

(2) $\alpha < \beta$

$$l_+^\alpha l_-^\beta = (l+\beta)(l+\beta-1)\cdots(l+\beta-\alpha+1) \times(l-\beta+1)(l-\beta+2)\cdots(l-\beta+\alpha)\, l_-^{\beta-\alpha},$$
$$l_-^\alpha l_+^\beta = (l+\beta)(l+\beta-1)\cdots(l+\beta-\alpha+1) \times(l-\beta+1)(l-\beta+2)\cdots(l-\beta+\alpha)\, l_+^{\beta-\alpha}.$$

Finally one finds the following types of terms in the expressions of the product operators

(i) $q_\mu q_\nu l_+^{\mu+\nu}$ with $\mu + \nu \geqslant 0$ and $q_\mu q_\nu l_+^{\mu-\nu}$ with $\mu + \nu \leqslant 0$;

(ii) $q_\mu l_-^\mu$ and $q_{-\mu} l_+^\mu$ with $\mu > 0$;

(iii) terms without operators.

Once the $O_{l+k}^{-k} O_l^{+k}$ brought into the standard forms discussed it is quite easy to look for relations between them and the other scalar operators constructed with a lower number of generators. It is rather straightforward to observe that in order to achieve a proper identification between the several occurring terms, one always needs four of the seven product operators. By this it also follows that among the various relations which can be constructed only four independent ones exist. Since our final aim consists of finding the eigenvalues of O_l^0 it is obvious to retain in each relation the $(O_l^0)^2$ term. It is now a matter of practical reasons to choose besides the $(O_l^0)^2$ term the three other occurring product operators. For the construction of the first relation we have chosen $k = 1$, 2, and 3, while for the second one the k values $+1$, -1, and -2 have been considered. The following relations could finally be constructed:

$$\begin{aligned} &5(l+3)^2(2l+5)^2 (O_l^0)^2 \\ &\quad + \sqrt{2}(l+1)(l+2)(l+3)^2(2l+3)(2l+5)^2 O_l^0 \\ &= -\frac{15(l+3)^2(l+4)(2l+3)}{(l+1)^2} O_{l+1}^{-1} O_l^{+1} \\ &\quad -3\frac{(l+4)(2l+5)(2l+9)}{(l+1)(l+2)^2} O_{l+2}^{-2} O_l^{+2} \\ &\quad -\frac{(2l+3)(l^2+7l+15)}{(l+1)(l+2)^2(l+3)^2} O_{l+3}^{-3} O_l^{+3} \\ &\quad -4(l+1)^2(l+2)^2(l+3)^2(2l+3)^2(2l+5)^2 V^* \\ &\quad -(2/5)l(l+1)^2(l+2)^2(l+3)^2 \\ &\quad \times(2l+3)^2(2l+5)^2(l+7), \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} &(l-1)(2l+1)(8l^2-41l+41)(O_l^0)^2 \\ &\quad -(\sqrt{2}/15)l(l+1)(l-1)(2l-1)(2l+1) \\ &\quad \times(4l^3-22l^2+31l-33)O_l^0 \\ &= \frac{(l-1)^2(l^2-5l+9)(2l-1)}{(l+1)^2} O_{l+1}^{-1} O_l^{+1} \\ &\quad -\frac{(l+1)(l^3-2l^2-25l+41)(2l-1)}{l^2} O_{l-1}^{+1} O_l^{-1} \\ &\quad +\frac{(l+1)(2l+1)(2l^2+2l-9)}{l^2(l-1)^2} O_{l-2}^{+2} O_l^{-2} \\ &\quad +4l^2(l+1)^2(l-1)^2(2l-1)^2(2l+1)V^* \\ &\quad +(1/75)l^2(l+1)^2(l-1)^2(2l-1)^2 \\ &\quad \times(2l+1)(8l^3-14l^2-58l-216). \end{aligned} \tag{3.3}$$

Using the fact that O_l^{+k} and O_l^{-k} go over into each other on replacing l by $-(l+1)$, two other equations can be easily derived from (3.2) and (3.3):

$$\begin{aligned} &5(l-2)^2(2l-3)^2(O_l^0)^2 \\ &\quad -\sqrt{2}l(l-1)(l-2)^2(2l-1)(2l-3)^2 O_l^0 \\ &= -15(l-2)^2(l-3)(2l-1)/l^2 O_{l-1}^{+1} O_l^{-1} \\ &\quad -\frac{3(l-3)(2l-3)(2l-7)}{l(l-1)^2} O_{l-2}^{+2} O_l^{-2} \\ &\quad -\frac{(2l-1)(l^2-5l+9)}{l(l-1)^2(l-2)^2} O_{l-3}^{+3} O_l^{-3} \\ &\quad -4l^2(l-1)^2(l-2)^2(2l-1)^2(2l-3)^2 V^* \\ &\quad -(2/5)(l+1)l^2(l-1)^2(l-2)^2 \\ &\quad \times(2l-1)^2(2l-3)^2(l-6), \end{aligned} \tag{3.4}$$

and

$$(l+2)(2l+1)(8l^2+57l+90)\ (O_l^0)^2$$

$$
\begin{aligned}
&-(\sqrt{2}/15)l(l+1)(l+2)(2l+3)\\
&\times(2l+1)(4l^3+34l^2+87l+90)\,O_l^0\\
&=-\frac{(l+2)^2(2l+3)(l^2+7l+15)}{l^2}\,O_{l-1}^{+1}\,O_l^{-1}\\
&+\frac{l(2l+3)(l^3+5l^2-18l-63)}{(l+1)^2}\,O_{l+1}^{-1}\,O_l^{+1}\\
&+\frac{l(2l+1)(2l^2+2l-9)}{(l+1)^2(l+2)^2}\,O_{l+1}^{-2}\,O_l^{+2}\\
&-4l^2(l+1)^2(l+2)^2(2l+3)^2(2l+1)V^*\\
&+(1/75)l^2(l+1)^2(l+2)^2(2l+3)^2\\
&\times(2l+1)(8l^3+38l^2-6l+180). \qquad (3.5)
\end{aligned}
$$

Once more we like to insist on the fact that these four relations are only valid when they are acting to the left upon $m=0$ states. It is now also possible to derive from (3.2)–(3.5) all other existing relations between the considered scalar operators. However, the four mentioned formulas will prove to be extremely useful in a following paper where the eigenvalues of O_l^0 will be determined.

4. REVIEW OF SOME PROPERTIES OF THE SHIFT AND PRODUCT OPERATORS

Several formulas, concerning the matrix elements of the shift operators will prove essential in the following papers. Most of them have been derived by Hughes[1] and will be merely stated here. The $O_l^{\pm k}$ are related by Hermiticity

$$
\begin{aligned}
&\langle l,a_l,m|(O_l^{+k})^\dagger|l+k,b_{l+k},m\rangle\\
&\quad=\beta_{kl}\langle l,a_l,m|O_{l+k}^{-k}|l+k,b_{l+k},m\rangle, \qquad (4.1)
\end{aligned}
$$

with[2]

$$
\beta_{kl}=(2l+1)/(2l+2k+1). \qquad (4.2)
$$

The matrix elements of $O_{l+k}^{-k}O_l^{+k}$ and $O_l^{+k}O_{l+k}^{-k}$ are inter-related:

$$
\begin{aligned}
&\sum_{a_l}\langle l,a_l,m|O_{l+k}^{-k}O_l^{+k}|l,a_l,m\rangle\\
&\quad=\sum_{b_{l+k}}\langle l+k,b_{l+k},m|O_l^{+k}O_{l+k}^{-k}|l+k,b_{l+k},m\rangle. \qquad (4.3)
\end{aligned}
$$

Two important relationships exist between the matrix elements of $O_l^{\pm k}$ and those of $O_{l\pm k}^{\mp k}O_l^{\pm k}$, i.e.,

$$
\begin{aligned}
&\langle l,a_l,m|O_{l+k}^{-k}O_l^{+k}|l,a_l,m\rangle\\
&\quad=\frac{1}{\beta_{kl}}\sum_{b_{l+k}}|\langle l+k,b_{l+k},m|O_l^{+k}|l,a_l,m\rangle|^2\\
&\quad=\beta_{kl}\sum_{b_{l+k}}|\langle l,a_l,m|O_{l+k}^{-k}|l+k,b_{l+k},m\rangle|^2, \qquad (4.4)
\end{aligned}
$$

and

$$
\begin{aligned}
&\langle l,a_l,m|O_{l-k}^{+k}O_l^{-k}|l,a_l,m\rangle\\
&\quad=\beta_{k,l-k}\sum_{b_{l-k}}|\langle l-k,b_{l-k},m|O_l^{-k}|l,a_l,m\rangle|^2\\
&\quad=\frac{1}{\beta_{k,l-k}}\sum_{b_{l-k}}|\langle l,a_l,m|O_{l-k}^{+k}|l-k,b_{l-k},m\rangle|^2. \qquad (4.5)
\end{aligned}
$$

APPENDIX: THE COMMUTATION RELATIONS $[q_\mu,q_\nu]$

$$
\begin{aligned}
&[q_0,q_{\pm1}]=\pm\frac{1}{\sqrt{10}}q_{\pm1}-\frac{\sqrt{3}}{10}l_\pm,\\
&[q_{\mp1},q_{\pm2}]=\frac{1}{2\sqrt{10}}l_\pm,\\
&[q_0,q_{\pm2}]=\pm\frac{1}{\sqrt{10}}q_{\pm2},\\
&[q_{\mp2},q_{\pm3}]=\mp\frac{1}{\sqrt{5}}q_{\pm1}-\frac{\sqrt{3}}{10\sqrt{2}}l_\pm,\\
&[q_0,q_{\pm3}]=\mp\frac{1}{\sqrt{10}}q_{\pm3},\\
&[q_{\mp1},q_{\pm3}]=\mp\frac{1}{\sqrt{5}}q_{\pm2},\\
&[q_{-1},q_{+1}]=\frac{1}{\sqrt{10}}q_0+\frac{1}{10}l_0,\\
&[q_{\mp1},q_{\mp2}]=\mp\frac{1}{\sqrt{5}}q_{\mp3},\\
&[q_{-2},q_{+2}]=-\frac{1}{\sqrt{10}}q_0-\frac{1}{5}l_0,\\
&[q_{-3},q_{-2}]=[q_{-3},q_{-1}]=0,\\
&[q_{-3},q_{+3}]=-\frac{1}{\sqrt{10}}q_0+\frac{3}{10}l_0,\\
&[q_{+1},q_{+3}]=[q_{+2},q_{+3}]=0.
\end{aligned}
$$

[1] J.W.B. Hughes, J. Phys. A **6**, 48 (1973).
[2] J.W.B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978); and refs. quoted therein.
[3] H.J. Weber, M.G. Huber, and W. Greiner, Z. Phys. **190**, 25 (1966).
[4] G. Vanden Berghe and H. De Meyer, J. Phys. A (to be published).
[5] H. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 486 (1980).
[6] A.P. Yutsis, I.B. Levinson, and V.V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translations, Jerusalem, 1962).

Shift-operator techniques for the classification of multipole-phonon states. II. Eigenvalues of the quadrupole shift operator O_l^0

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

(Received 10 October 1979; accepted for publication 28 December 1979)

By the aid of previously derived relations involving shift operators and their products, a set of rules is set up for the determination of eigenvalues of the scalar Hermitian shift operator O_l^0 in the space of R(5) nuclear quadrupole-phonon states. The eigenvalues are listed for all seniority states with $v < 8$.

I. INTRODUCTION

In the present paper we shall make use of the shift operators O_l^k with $k = 0, \pm 1, \pm 2, \pm 3$ introduced by Hughes and Yadegar[1] and in the preceding paper,[2] (to be referred to as I), and more precisely of the quadratic relations (I.3.2–I.3.5) existing between these shift operators and the Casimir operator of the R(5) group V^*, in order to solve the quadrupole state labeling problem completely.

Our major concern is to present a procedure for deriving in a pure analytical way (i.e., without having to take recourse in numerical routines) the eigenvalue spectrum of the Hermitian scalar operator O_l^0.

It has been known for a long time[3] that the quadrupole-phonon states (spin-2 boson states) correspond to the totally symmetric irreducible representations of the U(5) group or equally well of the SU(5) group if one discards irrelevant phase factors. The SU(5) Casimir operator N which counts the number of phonons in a state delivers a first label for classifying the state. Other nontrivial labels can be obtained by considering the chain of groups $SU(5) \supset R(5) \supset R(3)$. Indeed, the R(5) Lie algebra contains the quadratic Casimir operator V^* which is the seniority operator explicitly given in I, and which produces the eigenvalue $-v(v+3)/2$ whereby the seniority v denotes the number of phonons which do not occur coupled in zero angular momentum pairs. The biquadratic R(5) invariant, however, is not independent of V^* for the class of symmetric wavefunctions and is therefore useless for classifying purposes. Furthermore, by reducing SU(5) states into R(5) representation space no v degeneracies occur. Therefore we are left with the reduction of R(5) representations in the representation space of R(3), which is labeled itself by the total angular momentum l and its projection m. By the aid of the theory of group characters and by the use of symmetric Young tables[3] it can be demonstrated that in the latter reduction l degeneracies do occur. The actual l multiplicities in the reduction of R(5) into R(3) are listed for $1 \leqslant v \leqslant 9$ in Table I. It is precisely these l degeneracies which we intend to raise by means of the scalar operator O_l^0, which in addition to the operators N, V^*, L^2, and l_0 then serves as the so-called fifth label generating operator.

Since the operator O_l^0 has been constructed such that its eigenvalues are independent of the angular momentum projection m, and also since in the reduction of R(5) into R(3) m degeneracies are absent, it is harmless to set $m = 0$ throughout. As a consequence, it is convenient to restrict our attention at the $N = v$ and zero-m-projection quadrupole-phonon states

$$|v,l,\alpha_{v,l}\rangle \equiv |N = v,v,l,m = 0,\alpha_{v,l}\rangle \,, \tag{1.1}$$

whereby $\alpha_{v,l}$ denotes the eigenvalue of the O_l^0 operator, or

$$O_l^0|v,l,\alpha_{v,l}\rangle = \alpha_{v,l}|v,l,\alpha_{v,l}\rangle \,. \tag{1.2}$$

If a v state has l multiplicity equal to n for some l value, the corresponding eigenvalue equation (1.2) must have n distinct roots. In particular, if the state considered is not l degenerated there is a unique eigenvalue $\alpha_{v,l}$, which we call therefore a singular one. Furthermore, we shall consistently omit to label the state explicitly by a singular eigenvalue, thus writing

$$|v,l\rangle \equiv |v,l,\alpha_{v,l}\rangle \,, \tag{1.3}$$

if $|v,l\rangle$ is not l degenerated.

It has already been noticed that the particular construction of O_l^0 in terms of R(5) generators ensures us that O_l^0 necessarily commutes with the operators L^2, l_0, N, and V^*. Although the complete solution of the quadrupole-phonon state classification problem is therefore guaranteed in advance, a fact which will be confirmed subsequently, it cannot be claimed that it provides us with a unique solution of the problem. Indeed, it was already suggested by Weber *et al.*[3] that there should be at least one operator of the form

$$Q(k_1k_2k_3k_4k_5) = (2k_5+1)^{1/2}\{[(b^+b)^{k_1}(b^+b)^{k_2}]^{k_5}[(b^+b)^{k_3}(b^+b)^{k_4}]^{k_5}\}^{(0)} \quad (k_1,k_2,k_3,k_4 = 1 \text{ or } 3)\,, \tag{1.4}$$

whereby b^+ and b are the elementary spin-2 phonon creation and annihilation operators respectively, which could serve as a fifth label generating operator. That this is indeed the case for no less than ten operators of the kind (1.4) (apart from trivial symmetries) has been proven very recently by the present authors.[4] The proof followed as a by-product in the analysis of another label-generating operator, denoted by S, of which the particular form in terms of so-called canonical operators[5] allowed the numerical calculation of eigenvalues and the explicit determination of quadrupole-phonon

[a] Aangesteld navorser bij het Nationaal Fonds voor Wetenschappelijk Onderzoek, Belgium.

 0022-2488/80/071906-07$1.00

TABLE I. l-multiplicity table for maximum seniority states with $v \leqslant 9$. For each v are listed the l^n, l denoting the angular momentum and n the multiplicity. Blanks refer to nonoccurring states.

$v=1$		2															
$v=2$		2		4													
$v=3$	0		3	4		6											
$v=4$		2		4	5	6		8									
$v=5$		2		4	5	6	7	8		10							
$v=6$	0		3	4		6^2	7	8	9	10		12					
$v=7$		2		4	5	6	7	8^2	9	10	11	12		14			
$v=8$		2		4	5	6	7	8^2	9	10^2	11	12	13	14		16	
$v=9$	0		3	4		6^2	7	8	9^2	10^2	11	12^2	13	14	15	16	18

eigenstates.[6] Although very handsome in practical calculations, the mentioned operators display the common disadvantage that eigenvalues can only be obtained within a certain degree of accuracy and with reference to a large set of interrelated phonon states.[4,6] It is clear that by taking into consideration the operator O_l^0 such troubles must be avoided.

Another solution to the quadrupole classification problem as presented by Chacon *et al*,[7] is based on the property that maximum angular momentum projection states can be uniquely factorized in a particular set of elementary permissible diagrams (epd). Since within this scheme the fifth label is associated to the number of zero angular momentum coupled triplets of phonons occurring in the state, the analytical determination of its value for a particular state is obvious. Nevertheless, the lack of an operator generating the label, may be regarded as a weak point of the theory.

The outline of the present paper is the following. In Sec. 2 we obtain unambiguously the eigenvalues of O_l^0 on all maximum angular momentum states which from Table I are known to be of the form $|v,2v,\alpha_{v,2v}\rangle$. In Sec. 3 a chain procedure is presented for calculating the eigenvalues on the states which are not l degenerated, while in Sec. 4 the results of the procedure are resumed in a set of generally valid formulas. In Sec. 5 we tackle the problem of l degeneracies and at the end we list the eigenvalues belonging to the states of the form (1.1) for which $v \leqslant 7$.

Finally, let us mention that the following analysis may be considered also as a preliminary exercise for handling on the same lines of investigation the octupole-phonon classification problem too, for which at present no really satisfactory solution is known, since to our knowledge the only attempt undertaken so far in this direction,[8] relies on a utmost artificial ordering prescription.

2. THE MAXIMUM ANGULAR MOMENTUM STATES

We calculate now in a general way for an arbitrary maximum angular momentum state $|v,2v\rangle$ the singular eigenvalue $\alpha_{v,2v}$ of the scalar shift operator O_{2v}^0. The singularity of all $\alpha_{v,2v}$ is easily verified in Table I.

To this purpose we replace in Eq. (I.3.2) l formally by $2v$. Since $2v$ is the maximum value which l can attain for a given seniority ($v = N$), the operator products $O_{2v+1}^{-1} O_{2v}^{+1}$, $O_{2v+2}^{-2} O_{2v}^{+2}$, and $O_{2v+3}^{-3} O_{2v}^{+3}$, clearly yield zero contributions when acting on a $|v,2v\rangle$ state. In this way we thus immediately obtain from (I.3.2) a first relation containing $(O_{2v}^0)^2$ and O_{2v}^0 acting on an arbitrary but fixed $|v,2v\rangle$ state. We find after some cancellations that:

$$5(O_{2v}^0)^2|v,2v\rangle + 2\sqrt{2}(v+1)(2v+1)(4v+3)\, O_{2v}^0|v,v\rangle$$
$$= -\tfrac{16}{5}v(v+1)^2(2v+1)^2(4v+3)^2(2v+7)|v,2v\rangle$$
$$+8v(v+1)^2(2v+1)^2(4v+3)^2(v+3)|v,2v\rangle\ . \quad (2.1)$$

Solving this quadratic equation two different expressions are obtained for $(O_{2v}^0)|v,2v\rangle$, of which for a given seniority only one can be valid for producing the eigenvalue. In order to resolve this apparent ambiguity we can exploit the property that no states of the form $|v,2v-1\rangle$ ever exist, as may be seen from Table I. Consequently $O_{2v-1}^{+1} O_{2v}^{-1}|v,2v\rangle$ is identically zero and therefore equation (I.3.5) with 1 replaced by $2v$ immediately produces a second relation involving $(O_{2v}^0)^2$ $|v,2v\rangle$, $O_{2v}^0|v,2v\rangle$, and $|v,2v\rangle$ terms with v-dependent coefficients. We find that:

$$(16v^2+57v+45)(O_{2v}^0)^2|v,2v\rangle - (2\sqrt{2}/15)v(2v+1)(4v+3)$$
$$\times(16v^3+68v^2+87v+45)\, O_{2v}^0|v,2v\rangle$$
$$= -8v^3(v+1)(v+3)(2v+1)^2(4v+3)^2|v,2v\rangle$$
$$-(16/75)v^2(v+1)(2v+1)^2(4v+3)^2$$
$$\times(16v^3+38v^2-3v+45)|v,2v\rangle\ . \quad (2.2)$$

Eliminating $(O_{2v}^0)^2|v,2v\rangle$ from (2.1) and (2.2) it is straightforward to deduce that:

$$O_{2v}^0|v,2v\rangle = (2\sqrt{2}/5)v(v+1)(2v+1)(4v+3)|v,2v\rangle\ , \quad (2.3)$$

which on account of the definition (1.2) and the nondegeneracy of the maximum angular momentum states is equivalent to:

$$\alpha_{v,2v} = (2\sqrt{2}/5)v(v+1)(2v+1)(4v+3)\ . \quad (2.4)$$

It is seen from this expression that all eigenvalues $\alpha_{v,2v}$ are nonnegative and raise in value with increasing v value. In particular $\alpha_{0,0} = 0$ which is in agreement with the observation that since $N = v$, the state $|0,0\rangle$ necessarily denotes the vacuum.

3. A CHAIN PROCEDURE FOR CALCULATING SINGULAR EIGENVALUES

We continue in the present section our analysis for the states $|v,l\rangle$ which are not l degenerated. By careful inspection of the multiplicity Table I, it is readily noticed that none of the Eqs. (I.3.2–I.3.5) with $l > 2$ can give rise to a relation containing $(O_l^0)^2|v,l\rangle$ and $O_l^0|v,l\rangle$, and whereby each of the

three occurring shift-operator products of the form $O_{l-k}^{+k} O_{l}^{-k}$ ($k = \pm 1, \pm 2, \pm 3$), acting on the $|v,l\rangle$ state gives zero contribution.

As a demonstration of this statement let us study in general the $|v,2v-2\rangle$ states with $v>2$, which are all nondegenerated. Since states of the form $|v,2v-1\rangle$ do not exist (see Table I) we can use Eq. (I.3.2) with l replaced by $2v-2$, to obtain a relation involving $(O_{2v-2}^{0})^2|v,2v-2\rangle$, $O_{2v-2}^{0}|v,2v-2\rangle$, $O_{2v}^{-2} O_{2v-2}^{+2}|v,2v-2\rangle$, and v-dependent constant terms. As is the case with any of the other equations (I.3.3–I.3.5) there remains a product of different shift operators acting on the $|v,2v-2\rangle$ state. However, as at this stage we know already by (2.3) the result of O_{2v}^{0} acting on a maximum angular momentum state $|v,2v\rangle$ with the same seniority v, we can use Eq. (I.3.3) with l replaced by $2v$ in order to obtain an expression for the operator product $O_{2v-2}^{+2} O_{2v}^{-2}$ acting on the $|v,2v\rangle$ state, which does not contain other unknowns. Moreover, we have to remind that the property (I.4.3) reduces to

$$\langle v,l|O_{l+k}^{-k} O_{l}^{+k}|v,l\rangle = \langle v,l+k|O_{l}^{+k} O_{l+k}^{-k}|v,l+k\rangle \quad (k=\pm 1,\pm 2,\pm 3), \tag{3.1}$$

if both $|v,l\rangle$ and $|v,l+k\rangle$ are nondegenerated states. Therefore it suffices to multiply the equation for $O_{2v-2}^{0}|v,2v-2\rangle$ at the left with $\langle v,2v-2|$ and to multiply the expression for $O_{2v-2}^{+2} O_{2v}^{-2}|v,2v\rangle$ at the left with $\langle v,2v|$ to obtain with the help of (3.1) a quadratic equation in the expectation value $\langle v,2v-2|O_{2v-2}^{0}|v,2v-2\rangle = \alpha_{v,2v-2}$. Carrying out these operations we finally arrive at:

$$\begin{aligned} 5^2\alpha_{v,2v-2}^2 &+ 2\sqrt{2}5v(2v-1)(4v-1)\,\alpha_{v,2v-2} \\ &= -2^6 5v^2(v+1)(v-1)(2v-1)^2(4v+5) \\ &\quad + 2^3 v^2(2v-1)^2(4v-1)^2(v^2+9v+10). \end{aligned} \tag{3.2}$$

This equation does not enable us to obtain the eigenvalue uniquely; instead we find the following two expressions:

$$\alpha_{v,2v-2} = \begin{cases} (2\sqrt{2}/5)v(2v-1)(4v^2-5v-14), & \text{(3.3a)} \\ \text{or} \\ -(2\sqrt{2}/5)v(2v-1)(4v^2-v-15), & \text{(3.3b)} \end{cases}$$

of which for a determinate v value only one produces the correct eigenvalue. Furthermore we cannot claim at the present stage that by varying seniority it is always the same expression that gives the eigenvalue.

In contrast to the case of maximum angular momentum states we do not have the possibility here to construct a second equation of the form (3.2) which should allow us to calculate a unique eigenvalue expression. Indeed, it is easily verified that starting with another equation of the type (I.3.2–I.3.5) acting on a $|v,2v-2\rangle$ state than (I.3.3), one necessarily arrives at an equation which besides the eigenvalue $\alpha_{v,2v-2}$ also contains at least one of the eigenvalues $\alpha_{v,2v-2-k}$ with $k=1,2$, or 3, which are unknown so far.

For further use it is convenient to introduce a binary tree (each vertex having two underlying branches) of which the top vertex or root corresponds to the expression (2.4) for the $\alpha_{v,2v}$ eigenvalue and whereby the two vertices at one lower level correspond to one of the expression (3.3) for the $\alpha_{v,2v-2}$ eigenvalue each.

Now, the foregoing procedure can be repeated without fundamental alterations for obtaining expressions for the $\alpha_{v,2v-3}$ eigenvalues ($v>2$) since all the corresponding $|v,2v-3\rangle$ states ($v>2$) exist and are nondegenerate. However, the calculations must be repeated independently for both possible expressions (3.3). Consequently, we shall end up with four expressions for $\alpha_{v,2v-3}$ eigenvalues. These expressions are then associated to the four vertices of the binary tree at the second lower lying level with respect to the root level. Of course, our binary tree ends at the first lower lying level if the state $|v,2v-3\rangle$ does not exist which only happens for $v=2$.

To obtain explicit expressions for $\alpha_{v,2v-3}$ similar to (3.3), one proceeds as follows. Replace in (I.3.2) l formally by $2v-3$, replace in Eq. (I.3.5) l formally by $2v-2$ and replace finally in Eq. (I.3.3) l by $2v$. In virtue of the property (3.1) the terms $O_{2v-2}^{-1} O_{2v-3}^{+1}|v,2v-3\rangle$ and $O_{2v}^{-3} O_{2v-3}^{+3}|v,2v-3\rangle$ can be eliminated from the first equation by use of the other equations. The eigenvalue $\alpha_{v,2v}$ which enters the resulting equation being uniquely determined by (2.4), one obtains for each of the expressions (3.3) for $\alpha_{v,2v-2}$ entering that same equation, a quadratic equation in $\alpha_{v,2v-3}$ alone. This confirms our prior statement concerning the number of distinct expressions existing for $\alpha_{v,2v-3}$.

The same reasoning now also applies for determining expressions associated to $|v,2v-4\rangle$ states which exist and which are nondegenerated. From Table I it follows that this is the case if $v>3$. For $v=3$ the binary tree ends at the level of the four distinct $\alpha_{v,2v-3}$ values. If $v>3$ the binary tree can be extended with a level of eight vertices corresponding to eight expressions for $\alpha_{v,2v-4}$. The final step consists in deriving 16 possible expressions for the $\alpha_{v,2v-5}$ eigenvalues of the $|v,2v-5\rangle$ states which for $v>4$ all exist and are nondegenerated. For $v=4$ the tree ends at the foregoing level. Furthermore, since for $v>5$, the $|v,2v-6\rangle$ states are all degenerated we have to stop here our tree generating procedure momentarily. The important feature of the foregoing mechanism is that once a particular eigenvalue, say $\alpha_{v,l}$ with $2v-5 \leqslant l \leqslant 2v-2$ is rigorously determined for a fixed seniority v, we immediately can read off all the eigenvalues $\alpha_{v,l'}$, for which $l'>l$ by following in the binary tree the unique chain which starts at the vertex corresponding to the exact $\alpha_{v,l}$ value and ends at the root which corresponds to the rigorous $\alpha_{v,2v}$ value. The eigenvalues are generated by replacing in the expressions associated to the intermediate vertices of the chain, v by its actual value. The only problem remaining is to find a method by which the eigenvalue associated to the deepest lying level of the binary tree can be rigorously fixed.

To this aim, let us remark that a variety of expressions for calculating $\alpha_{v,l}$ values has been obtained so far by starting at the maximum angular momentum states. There is, however, no particular reason, why one should not introduce a tree-generating procedure beginning at the lowest angular momentum state too. By inspection of Table I it appears that in contrast to the maximum angular momentum states, the minimum value of the angular momentum is no longer linearly dependent on the seniority v. Instead, we have to distinguish states by the divisibility of v with respect to the numeral 3.

In particular if v is a triple or $v = 3k$ ($k = 1,2,\cdots$), the state with lowest angular momentum is the nondegenerated $|3k,0\rangle$ state, whereas the $|3k,1\rangle$ and $|3k,2\rangle$ states do not exist. The eigenvalue $\alpha_{3k,0}$ of the zero momentum state is easily determined by taking into consideration the explicit form (I.2.1) of the operator O_l^0. Since therein l_0 may be set equal to zero,[2] all the remaining terms have at the right-hand side at least one operator $l_\pm$, which shifts the angular momentum projection m by one unit. Thus, as for $|3k,0\rangle$ states this projection is necessarily confined to zero, we immediately deduce that

$$\alpha_{3k,0} = 0 . \tag{3.4}$$

If we now let Eq. (I.3.2) act on $|3k,0\rangle$, and Eq. (I.3.4) on $|3k,3\rangle$, the first relation so obtained can be used to eliminate on account of the property (3.1), the term $O_0^{+3} O_3^{-3} |3k,3\rangle$ from the second equation. Furthermore, $\alpha_{3k,0}$ being rigorously known, the latter equation is immediately transformed into a quadratic equation in $\alpha_{3k,3}$. The solutions are found to be independent of k (or v); i.e.,

$$\alpha_{3k,3} = \begin{cases} 24\sqrt{2}, & (3.5a) \\ \text{or} \\ -18\sqrt{2}. & (3.5b) \end{cases}$$

It is very convenient to introduce again a binary tree for representing eigenvalue expressions. The root is according to (3.4) assigned the zero value of $\alpha_{3k,0}$, while the two underlying vertices contain respectively the expressions (3.5a) and (3.5b). Moreover, since all $|3k,4\rangle$ states are nondegenerated a similar procedure as before can be used, such that the tree is extended to include four expressions for $\alpha_{3k,4}$.

Finally, if $v = 3k \pm 1$ ($k = 1,2,\cdots$) the state with lowest angular momentum is the $|3k \pm 1,2\rangle$ state, and no simple argument as the one leading to (3.4) can be invoked. Instead, it is possible to obtain with the help of Eq. (I.3.3) whereby l is replaced by the numeral 2, and on account of the fact that no $|3k \pm 1,3\rangle$ states exist, a quadratic equation in the unknown $\alpha_{3k\pm1,2}$. A second quadratic equation cannot be obtained from Eq. (I.3.4) since with $l = 2$, infinities would arise. We thus find the following two solutions for $\alpha_{3k\pm1,2}$:

$$\alpha_{3k\pm1,2} = \begin{cases} \frac{6}{5}\sqrt{2}(5v+9) & (3.6a) \\ \text{or} \\ -\frac{6}{5}\sqrt{2}(5v+6) & (3.6b) \end{cases} \qquad (v = 3k \pm 1).$$

These expressions can be associated respectively to the two vertices of a binary tree, which are the immediate successors of, for our purpose, a meaningless root. The derivation of four expressions for $\alpha_{3k\pm1,4}$, eight expressions for $\alpha_{3k\pm1,5}$, sixteen expressions for $\alpha_{3k\pm1,6}$, and, finally, thirty-two expressions for $\alpha_{3k\pm1,7}$, all of which are associated to nondegenerated states (see Table I), then follows the same pattern as before. Obviously, for low k values the eigenvalue trees must end such that no levels associated to nonexisting states are included.

At this point we have for any fixed v value two eigenvalue trees at our disposition: one generated from the highest angular momentum state and one generated from the lowest angular momentum state. There remains now the problem of linking the two trees in such a way that in both we can isolate a unique chain up to the root of which the vertices are associated to the rigorous eigenvalues.

Actually, the linking is most easily performed if both trees end at a level associated to the same angular momentum value, say l. Since the same rigorous eigenvalue $\alpha_{v,l}$ necessarily appears at the end level of both trees, the eigenvalue chains can be isolated immediately. As an example, let us mention the case whereby $v = 2$. The upper tree ends at the expressions (3.3) with v replaced by 2, giving $-96\sqrt{2}/5$ or $12\sqrt{2}/5$. The lower tree, however, ends at the expressions (3.6) with $v = 2$, giving $114\sqrt{2}/5$ or $-96\sqrt{2}/5$. The linking is uniquely performed with $\alpha_{2,2} = -96\sqrt{2}/5$, being therefore the true eigenvalue. It is readily verified that with this simple procedure the part of Table I, for which $l \leqslant 5$ is completely covered.

However, the possibility of linking the upper and lower tree may also be guaranteed if both trees end at levels for which the respective associated angular momentum values do not differ by more than 3 units one from the other. More precisely, let us assume that the upper tree ends at $l = \text{lup}$ and the lower at $l = \text{low}$, such that the condition

$$p = \text{lup} - \text{low} \leqslant 3 ,$$

is fulfilled. Choosing at random an eigenvalue $\alpha_{v,\text{low}}$ out of the lower tree we can by suitable combination of Eqs. (I.3.2)–(I.3.5) obtain an expression for the unknown expectation value $\langle v,\text{low}|O_{\text{lup}}^{-p} O_{\text{low}}^{+p}|v,\text{low}\rangle$ in terms of $\alpha_{v,\text{low}}$-dependent quantities. Similarly, picking up out of the upper tree one of the possible $\alpha_{v,\text{lup}}$ eigenvalues, we obtain by suitable combinations of Eqs. (I.3.2)–(I.3.5) an expression for $\langle v,\text{lup}|O_{\text{low}}^{+p} O_{\text{lup}}^{-p}|v,\text{lup}\rangle$ involving $\alpha_{v,\text{lup}}$ dependent quantities only. On account of the property (3.1) both expressions thus obtained should have the same numerical value. If this is indeed the case we have picked up the correct eigenvalues $\alpha_{v,\text{low}}$ and $\alpha_{v,\text{lup}}$, if not the calculation must be repeated for other combinations of likely $\alpha_{v,\text{low}}$ and $\alpha_{v,\text{lup}}$ expressions in the trees, until (3.1) is rigorously satisfied. Then the chains of correct singular eigenvalues are immediately isolated in both trees. Obviously, the present method for linking trees is very convenient for jumping one or two nonsingular eigenvalues. By inspection of Table I it is thus readily verified that all singular eigenvalues $\alpha_{v,l}$ with $l \leqslant 7$ may be calculated. There remains of course the problem of treating degenerated states.

4. CLOSED FORMULAS FOR CALCULATING SINGULAR EIGENVALUES

By carrying out the tree-generating and chain-producing algorithms as discussed in Sec. 3, in full detail, we have found explicitly all singular eigenvalues $\alpha_{v,l}$, for $v < 8$. These eigenvalues can be resumed in a set of handsome formulas.

Starting from the highest l value one obtains:

$$\alpha_{v,2v} = \tfrac{2}{5}\sqrt{2}v(v+1)(2v+1)(4v+3) \qquad (v \geqslant 1), \tag{4.1}$$

$$\alpha_{v,2v-2} = \tfrac{2}{5}\sqrt{2}v(2v-1)(4v^2-5v-14) \qquad (v \geqslant 2), \tag{4.2}$$

$$\alpha_{v,2v-3} = \tfrac{2}{5}\sqrt{2}(4v-3)(2v^3-v^2-17v+1) \qquad (v \geqslant 3), \tag{4.3}$$

$$\alpha_{v,2v-4} = \tfrac{2}{5}\sqrt{2}(v-1)(8v^3-38v^2-v-60) \qquad (v \geqslant 4), \tag{4.4}$$

$$\alpha_{v,2v-5} = \tfrac{2}{5}\sqrt{2}(8v^4 - 42v^3 - 77v^2 + 258v - 150) \quad (v \geqslant 5). \tag{4.5}$$

We observe that independent of the v value under consideration, the eigenvalues are always associated with the vertices of a unique chain in the upper tree. This remarkable property verified so far for $v \leqslant 7$, gives strong evidence for predicting that (4.1)–(4.5) remain valid formulas for $v > 7$ too. In fact this has been shown already for the particular formula (4.1) in Sec. 2.

Starting now from the lowest l value we obtain:

If $v = 3k$ $(k = 1,2,\cdots)$:

$$\alpha_{v,0} = 0, \tag{4.6}$$

$$\alpha_{v,3} = -18\sqrt{2}, \tag{4.7}$$

$$\alpha_{v,4} = 42\sqrt{2}. \tag{4.8}$$

If $v = 3k + 1$ $(k = 0,1,2,\cdots)$:

$$\alpha_{v,2} = \tfrac{6}{5}\sqrt{2}(5v + 9) \quad (v \geqslant 1), \tag{4.9}$$

$$\alpha_{v,4} = -12\sqrt{2}(3v + 4) \quad (v \geqslant 4), \tag{4.10}$$

$$\alpha_{v,5} = 18\sqrt{2}(3v + 1) \quad (v \geqslant 4), \tag{4.11}$$

$$\alpha_{v,6} = 6\sqrt{2}(5v + 36) \quad (v \geqslant 4), \tag{4.12}$$

$$\alpha_{v,7} = -\tfrac{6}{5}\sqrt{2}(150v + 401) \quad (v \geqslant 7). \tag{4.13}$$

If $v = 3k - 1$ $(k = 1,2,\cdots)$:

$$\alpha_{v,2} = -\tfrac{6}{5}\sqrt{2}(5v + 6) \quad (v \geqslant 2), \tag{4.14}$$

$$\alpha_{v,4} = 12\sqrt{2}(3v + 5) \quad (v \geqslant 2), \tag{4.15}$$

$$\alpha_{v,5} = -18\sqrt{2}(3v + 8) \quad (v \geqslant 5), \tag{4.16}$$

$$\alpha_{v,6} = -6\sqrt{2}(5v - 21) \quad (v \geqslant 5), \tag{4.17}$$

$$\alpha_{v,7} = \tfrac{6}{5}\sqrt{2}(150v + 49) \quad (v \geqslant 5). \tag{4.18}$$

We observe that in the lower tree the eigenvalue chain to be selected is dependent on the divisibility of v with respect to 3. Thus, keeping the angular momentum fixed and increasing the seniority with 3 units (which is equivalent since $N = v$, to enlarging the state by a triplet of zero-coupled phonons, the same formula can be used for calculating the eigenvalue. Moreover, by increasing v consecutively by 3,6,··· units, the eigenvalue is anytime increased by a fixed l-dependent positive or negative amount. From these observations and from the fact that the $|3k,0\rangle$ states, which can also be written as $[\otimes |3,0\rangle]^k$, are the ones which are completely annihilated by the scalar shift operator, it is apparent that we must attribute within the present formalism, a particular status to the state $|3,0\rangle$. Although O_l^0 does not count the number of zero-coupled triplets occurring in a state, the striking resemblance to Chacons' state labeling solution partially based on (3,0) epd factorization, cannot be denied.

A remarkable feature of the various eigenvalue expressions listed, is that the set (4.14)–(4.18) can be obtained from the set (4.9)–(4.13) in the same order, by simply changing in the right-hand sides of the latter set v by $-v-3$. Moreover, the right-hand sides of (4.6)–(4.8) being constants, are invariant to this transformation. This feature clearly finds its origin in the fact that Eqs. (I.3.2)–(I.3.5) are only dependent on the seniority v by the product $-v(v+3)/2$—the V^* eigenvalue— and that this product is invariant for the mentioned transformation.

It is also worthwhile to notice that all singular eigenvalues calculated so far can be written as $(6\sqrt{2}/5)$ times an integer value. As it is, moreover, very tempting to claim that Eqs. (4.6)–(4.18) remain valid for $v > 7$, this property is expected to be demonstrated by all singular eigenvalues.

Above we have listed eigenvalue expressions only. Nevertheless, for the sake of completeness and for a clearer insight in the consistency of the shift-operator technique, it is useful to give the various used expressions for the expectation values $\langle v,l|O_{l+k}^{-k} O_l^{+k}|v,l\rangle$ too. We have found while generating the upper tree

$$\langle v,2v|O_{2v-2}^{+2} O_{2v}^{-2}|v,2v\rangle = (2^7/3)v^4(v-1)(2v+1)^2(2v-1)^3(4v+1),$$

$$\langle v,2v|O_{2v-3}^{+3} O_{2v}^{-3}|v,2v\rangle = 2^9v^4(v-1)^3(v-2)(2v-1)^2(4v-1)(4v+1)(4v-3),$$

$$\langle v,2v-2|O_{2v-3}^{+1} O_{2v-2}^{-1}|v,2v-2\rangle = (2^5/3)(v-1)^2(v-2)(2v-1)(4v-3)(4v-1)(2v+1)^2,$$

$$\langle v,2v-2|O_{2v-4}^{+2} O_{2v-2}^{-2}|v,2v-2\rangle = (2^8/3)(v-1)^3(v-2)(v-3)(2v-1)^2(2v-3)^2(2v+1)(4v-1), \tag{4.19}$$

$$\langle v,2v-2|O_{2v-5}^{+3} O_{2v-2}^{-3}|v,2v-2\rangle = 2^9(v-1)^3(v-2)^3(v-3)(v-4)(2v-3)^2(4v-5)(4v-3)(4v-1),$$

$$\langle v,2v-3|O_{2v-4}^{+1} O_{2v-3}^{-1}|v,2v-3\rangle = 2^6(v-1)(v-3)(2v-1)(2v+1)(2v-3)^2(4v-3),$$

$$\langle v,2v-3|O_{2v-5}^{+2} O_{2v-3}^{-2}|v,2v-3\rangle = (2^7/3)(v-1)(v-2)^2(v-3)(v-4)(2v-1)(2v+1)^2(2v-3)^2(4v-5),$$

$$\langle v,2v-4|O_{2v-5}^{+1} O_{2v-4}^{-1}|v,2v-4\rangle = (2^6/3)(v-2)^2(v-4)(2v-1)^2(2v+1)(4v-5)(4v-3).$$

Similarly, starting at the lowest angular momentum values we obtained for $v = 3k$:

$$\langle v,0|O_3^{-3}O_0^{+3}|v,0\rangle = 2^5 3^4 5 v(v+3)\,, \tag{4.20}$$
$$\langle v,3|O_4^{-1}O_3^{+1}|v,3\rangle = 2^7\ 3\ 5\ 7(2v+1)(2v+5)\,.$$

For $v = 3k+1$:

$$\langle v,2|O_4^{-2}O_2^{+2}|v,2\rangle = 2^{10}3^3 5(v-1)(2v+5)\,,$$
$$\langle v,2|O_5^{-3}O_2^{+3}|v,2\rangle = 2^7 3^6 5^3 7(v-1)(v+5)\,,$$
$$\langle v,4|O_5^{-1}O_4^{+1}|v,4\rangle = 2^6 3\ 5^2 7(v+5)(2v+5)\,,$$
$$\langle v,4|O_6^{-2}O_4^{+2}|v,4\rangle = 2^9 3^3 5^3 7(v+5)(2v-1)\,, \tag{4.21}$$
$$\langle v,4|O_7^{-3}O_7^{+3}|v,4\rangle = 2^5 3^6 5^3 7^3 11(v-4)(v+5)\,,$$
$$\langle v,5|O_6^{-1}O_5^{+1}|v,5\rangle = 2^6 3^6 5(2v-1)(2v+5)\,,$$
$$\langle v,5|O_7^{-2}O_5^{+2}|v,5\rangle = 2^8 3^3 5\ 7^2 11(v-4)(2v+5)\,,$$
$$\langle v,6|O_7^{-1}O_6^{+1}|v,6\rangle = 2^7 3\ 7^2 11(v-4)(2v-1)\,.$$

and, finally, for $v = 3k-1$:

$$\langle v,2|O_4^{-2}O_2^{+2}|v,2\rangle = 2^{10}3^3 5(v+4)(2v+1)\,,$$
$$\langle v,2|O_5^{-3}O_2^{+3}|v,2\rangle = 2^7 3^6 5^3 7(v-2)(v+4)\,,$$
$$\langle v,4|O_5^{-1}O_4^{+1}|v,4\rangle = 2^6 3\ 5^2 7(v-2)(2v+1)\,,$$
$$\langle v,4|O_6^{-2}O_4^{+2}|v,4\rangle = 2^9 3^3 5^3 7(v-2)(2v+7)\,, \tag{4.22}$$
$$\langle v,4|O_7^{-3}O_4^{+3}|v,4\rangle = 2^5 3^6 5^3 7^3 11(v-2)(v+7)\,,$$
$$\langle v,5|O_6^{-1}O_5^{+1}|v,5\rangle = 2^6 3^6 5(2v+1)(2v+7)\,,$$
$$\langle v,5|O_7^{-2}O_5^{+2}|v,5\rangle = 2^8 3^3 5\ 7^2 11(v+7)(2v+1)\,,$$
$$\langle v,6|O_7^{-1}O_6^{+1}|v,6\rangle = 2^7 3\ 7^2 11(v+7)(2v+7)\,.$$

Remark that the set of expectation values for $v = 3k-1$, can again be obtained from the corresponding set for $v = 3k+1$, by formally changing in the latter one at the right-hand side v by $-v-3$, and keeping v at the left-hand side unchanged. On the other hand the right-hand sides of (4.20) are invariant to this transformation. We also note that all expressions (4.19)–(4.22) are nonnegative for any positive integer value of v. This is in complete agreement with the properties (I.4.4)–(I.4.5), showing that expectation values of the form $\langle v,l|O_{l+k}^{-k}O_l^{+k}|v,l\rangle$ can always be written as sums of positive semidefinite quantities. It can thus happen that an expectation value becomes zero for particular v and l values, indicating an account of the same properties that the state $|v,l+k\rangle$ does not occur. Let us check this from the formulas listed above by a few examples. From (4.20) it follows that $\langle v,2v-2|O_{2v-4}^{+2}O_{2v-2}^{-2}|v,2v-2\rangle$ is zero for $v = 2$ or $v = 3$, since v is necessarily an integer and greater than 1. This indicates that the states $|2,0\rangle$ and $|3,2\rangle$ do not occur, whereas the other states of the form $|v,2v-4\rangle$ with $v \geqslant 4$ always exist, a fact which is readily verified from Table I. Similarly it follows from (4.20) that $\langle v,4|O_7^{-3}O_4^{+3}|v,4\rangle$ is only zero for $v = 4$, showing that all the $|3k+1,7\rangle$ states occur with the exception of the $|4,7\rangle$ state. This statement is again confirmed by Table I.

So far singular eigenvalues have been obtained by a tree-generating and chain-isolating mechanism which is essentially based on the property (3.1). In Sec. 5 it will be demonstrated that by using this property in the extended form valid for degenerated states too, nonsingular eigenvalue calculations can be handled by the same mechanism.

5. EIGENVALUES FOR DEGENERATED STATES

In this section we derive a procedure for calculating the O_l^0 eigenvalues for a set of $|v,l\rangle$ states having the same quantum numbers $N = v,v,l$ and $m = 0$. According to the general theory these eigenvalues $\alpha_{v,l}$ should lift the degeneracy. For simplicity and in agreement with the multiplicity values in Table I, we only consider doubly degenerated states which we assume orthonormalized and which we denote by $|v,l,(1)\rangle$ and $|v,l,(2)\rangle$. Herein (1) and (2) are merely numbers to distinguish the independent states, and they differ from the O_l^0 eigenvalues which we denote by $\alpha_{v,l}(1)$ and $\alpha_{v,l}(2)$, respectively.

Let us further suppose that for the other occurring states with the same seniority v, the O_l^0 eigenvalues are singular and have been rigorously found yet. With this assumption we can tackle the calculation of the doubly degenerated $|6,6\rangle$ and $|7,8\rangle$ states since by our former tree-generating mechanism all other eigenvalues $\alpha_{v,l}$ with $v = 6$ or $v = 7$ have been obtained in (4.1)–(4.5), (4.6)–(4.8), and (4.9)–(4.13). If we choose one of the Eqs. (I.3.2)–(I.3.5), say for instance (I.3.2), acting on both the orthogonal states $|v,l,(1)\rangle$ and $|v,l,(2)\rangle$, the two resulting relations when multiplied at the left-hand sides by $\langle v,l,(1)|$ and $\langle v,l,(2)|$, respectively, can be added to yield a single relation involving apart from constant terms the combinations

$$\alpha_{v,l}^2(1) + \alpha_{v,l}^2(2), \quad \alpha_{v,l}(1) + \alpha_{v,l}(2),$$

and

$$\sum_{j=1}^{2} \langle v,l,(j)|O_{l+k}^{-k}O_l^{+k}|v,l,(j)\rangle\,,$$

with $k = 1,2,3$. Since, furthermore, the properties (I.4.4) and (I.4.5) can be written for the present case in the form:

$$\sum_{j=1}^{2} \langle v,l,(j)|O_{l+k}^{-k}O_l^{+k}|v,l,(j)\rangle$$
$$= \langle v,l+k\,|O_l^{+k}O_{l+k}^{-k}|v,l+k\rangle\,. \tag{5.1}$$

It is only a matter of letting the appropriate equations of the form (I.3.2)–(I.3.5) act on the nondegenerated states $|v,l+k\rangle$ with $k = 1,2,3$ of which the eigenvalues are known already, in order to obtain after elimination of the expectation values of the shift operator products, a relation involving the combinations $\alpha_{v,l}^2(1) + \alpha_{v,l}^2(2)$ and $\alpha_{v,l}(1) + \alpha_{v,l}(2)$ alone. If we then repeat this complete procedure starting with any other of the Eqs. (I.3.2)–(I.3.5) we obtain a system of two quadratic equations from which it is straightforward to calculate the solutions $\alpha_{v,l}(1)$ and $\alpha_{v,l}(2)$. Obviously, since we have originally four equations at our disposition we could have deduced a system of four quadratic equations in $\alpha_{v,l}(1)$ and $\alpha_{v,l}(2)$ too, of which only two are needed for finding the solution. This apparent overcompleteness should allow one to incorporate the eigenvalue calculation of doubly degenerated states into our previous tree-generating mechanism, such that in the resulting equations other eigenvalues appear whose value must be chosen in a set of likely values. Since we do not need such procedure for calculating the $|6,6\rangle$ and

TABLE II. The values of $5\alpha_{v,l}/6\sqrt{2}$ for $v \leq 7$, $\alpha_{v,l}$ denoting the O_l^0 eigenvalues.

l \ v	1	2	3	4	5	6	7
0			0			0	
2	14	−16		29	−31		44
3			−15			−15	
4		110	35	−160	200	35	−250
5				195	−345		330
6			420	280	−20	$\frac{15}{2}[-5 \pm (9881)^{1/2}]$	355
7					799	−26	−1451
8				1140	915	490	$\frac{5}{2}[269 \pm (139 \times 1531)^{1/2}]$
9						2065	895
10					2530	2200	1630
11							4325
12						4914	4459
13							
14							8680

$|7,8\rangle$ eigenvalues, there is no reason for going into much more detail here.

The results of the foregoing analysis when applied on the $|6,6\rangle$ and $|7,8\rangle$ states, can be found in Table II, which contains all numerical eigenvalues up to seniority $v = 7$. We note that in contrast to the singular eigenvalues, the ones associated to doubly degenerated states cannot be written as $(6\sqrt{2})/5$ times an integer value, anymore. However, their sum still remains except for this factor an integer value.

6. DISCUSSION

In the present paper we have succeeded in calculating rigorously a large set of quadrupole-phonon state eigenvalues with the use of only four independent quadratic relations between shift operators which were derived in I. It is important to notice that there was thus no need for having at our disposal equations involving products of more than two shift operators.

The obtained eigenvalues can be used for labeling the states if these cannot be distinguished by the other quantum numbers. In fact it is known from the general theory that the O_l^0 spectrum provides us with a complete solution of the state-labeling problem. The classification of the quadrupole-phonon states therefore being completely accomplished, there only remains the problem of constructing these states explicitly, and to analyze their relationship to the states as previously constructed by the present authors.[6] For details concerning these problems the reader is referred to a subsequent paper.

[1] J.W.B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978), and references quoted therein.
[2] G. Vanden Berghe and H.E. De Meyer, J. Math. Phys. **21**, 1902 (1980).
[3] H.J. Weber, M.G. Huber, and W. Greiner, Z. Phys. **190**, 25 (1966).
[4] G. Vanden Berghe and H.E. De Meyer, J. Phys. A (to be published).
[5] H.E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1913 (1980).
[6] G. Vanden Berghe and H.E. De Meyer, Nucl. Phys. A **323**, 302 (1979).
[7] E. Chacon, M. Moshinsky, and R.T. Sharp, J. Math. Phys. **17**, 668 (1976).
[8] S.G. Rohozinski, J. Phys. G **4**, 1075 (1978).

Shift operator techniques for the classification of multipole-phonon states. III. The quadrupole-phonon eigenstates of O_l^0

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

(Received 10 October 1979; accepted for publication 28 December 1979)

It is shown that the non-l-degenerated normalized eigenstates of O_l^0 can be identified with previously constructed quadrupole-phonon states which have been labeled by the intermediate angular momentum values. The l-degenerated O_l^0 eigenstates are proven to be identical to the corresponding eigenstates of the fifth label generating operator S. From the analysis of the relationship between O_l^0 and S, a rigorous expression for S in terms of canonical operators is withdrawn, while it also becomes possible to obtain the S eigenvalues in exact irrational form.

1. INTRODUCTION

In two preceding papers[1,2] (to be referred to as I and II) we have outlined the construction of the R(3) scalar shift operator O_l^0 and the calculation of its eigenvalues in the space of quadrupole-phonon states $|N=v,v,l,m=0\rangle$. Hereby v denotes the seniority, whereas the number of phonons N is chosen equal to v and the angular momentum projection m is set equal to zero momentarily. Moreover, it was shown that the O_l^0 eigenvalues $\alpha_{v,l}$ are m-independent. Furthermore, the various results in II have been derived with reference only to the existence of mutually orthogonal and normalized quadrupole-phonon states $|v,l\rangle$, and not to their explicit forms. For this we could rely on strict group-theoretical results[3] which were summarized in Table I of Paper II. As an obvious continuation of the preceding investigation we intend in the present paper to construct these O_l^0 eigenstates explicitly.

It goes without saying that the construction of such eigenstates of which the labeling is completed by means of exact calculable irrational O_l^0 eigenvalues is very important from the theoretical point of view. However, for practical purposes it may be preferable to express quadrupole-phonon states in terms of coupled elementary phonon-creation operators $b^+_{2\mu}$ acting on a vacuum, thereby labeling them by an overcomplete set of intermediate coupling angular momentum values. This has been realized recently by the present authors[4] by use of numerical computer routines. A basis of independent orthonormalized phonon states has been selected in accordance with a simple counting mechanism, not associated to any operator. In this paper we therefore also want to investigate the relationship between the latter set of states and our actual O_l^0 eigenstates. Finally, in order to raise l degeneracies we introduced also a label-generating operator S,[5] which could be numerically expressed in terms of so-called canonical operators $O^{I_2 I_3 \cdots J_N}_{J_1 J_2 \cdots J_N}$ as follows:

$$S = 0.8658 O^{235}_{235} + O^{246}_{246} + 0.5804 O^{468}_{468}\,, \tag{1.1}$$

with

$$O^{J_1 J_2 J_3}_{J_1 J_2 J_3} = \sum_{m_1 m_2 m_3 m_4} \sum_{m_1' m_2' m_3' m_4'} \sum_{M_1 M_2 M_1' M_2' M_3} \langle 2m_1 2m_2 | J_1 M_1\rangle \times \langle J_1 M_1 2m_3 | J_2 M_2\rangle \langle J_2 M_2 2m_4 | J_3 M_3\rangle \times \langle 2m_1' 2m_2' | J_1 M_1'\rangle \times \langle J_1 M_1' 2m_3' | J_2 M_2'\rangle \langle J_2 M_2' 2m_4' | J_3 M_3\rangle \times b^+_{m_1} b^+_{m_2} b^+_{m_3} b^+_{m_4} b_{m_1'} b_{m_2'} b_{m_3'} b_{m_4'}\,. \tag{1.2}$$

Herein $(-1)^\mu b_{-\mu}$ are the phonon-annihilation operators and the bracket factors denote Clebsh–Gordon coefficients. As an example we have shown how the S eigenvalues can be used to distinguish the doubly degenerated $|6,6\rangle$ states[6] and how these states can be constructed as linear combinations of basis states.

As a consequence it appears interesting to analyze the relationship between O_l^0 and S too. We shall demonstrate how this leads us to a simple procedure for calculating S eigenvalues rigorously and for obtaining also in an exact way the coefficients in the right-hand side (rhs) of (1.1).

2. THE CONSTRUCTION OF O_l^0 EIGENSTATES

For reasons of simplicity, let us start with the construction of the maximum angular momentum states, denoted in II as $|v,2v\rangle$. It is obvious that apart from a normalization factor $\mathcal{N}_v$ to be determined later, such states must be uniquely expressible in terms of phonon-creation operators acting on the vacuum state, i.e.,

$$\begin{aligned}|v,2v\rangle &= \mathcal{N}_v \{[(b^+b^+)^4 b^+]^6 \cdots b^+\}^{2v}_{m=0} |0,0\rangle \\ &= \mathcal{N}_v |0246\cdots 2v\rangle_{m=0}\,,\end{aligned} \tag{2.1}$$

whereby at the last line we introduce the notation of Ref. 4.

Since in fact $\mathcal{N}_v$ is independent of the particular choice of m, it is convenient to set momentarily at the rhs of (2.1) m equal to its maximum value $m=2v$. Then the intermediate projections $m_1,\ldots,m_v$ become all restricted to $+2$, and $\mathcal{N}_v$ is readily found to be given by $1/(v!)^{1/2}$. The O^0_{2v} eigenstates $|v,2v\rangle$ are thus identical to the normalized states $\mathcal{N}_v |024\cdots 2v\rangle$ listed elsewhere.[4]

As a next step we consider the O^0_{2v-2} eigenstates $|v,2v-2\rangle$ which are all nondegenerated, and are assumed to be normalized to unity already. Bearing in mind that the shift operator O^{-2}_{2v} when acting on a $|v,2v\rangle$ state leaves the seniority v and the angular momentum projection $m=0$ un-

[a] Bevoegdverklaard navorser bij het Nationaal Fonds voor Wetenschappelijk Onderzoek (Belgium).

 0022-2488/80/071913-05$1.00

changed while decreasing the angular momentum by two units, it follows that

$$O_{2v}^{-2}\,|v,2v\rangle = a|v,2v-2\rangle\,. \tag{2.2}$$

The states $|v,2v\rangle$ and $|v,2v-2\rangle$ being normalized to unity, we have to determine the value of the proportion parameter a. To this aim we multiply (2.2) on both sides at the left with $\langle v,2v-2|$ and use the properties (I.4.1) and (I.4.2) in order to obtain consecutively

$$\begin{aligned} a &= \langle v,2v-2|\,O_{2v}^{-2}|v,2v\rangle \\ &= \frac{1}{\beta_{2,2v-2}}\,\langle v,2v-2|(O_{2v-2}^{+2})^{\dagger}|v,2v\rangle \\ &= \frac{4v+1}{4v-3}\,\langle v,2v|\,O_{2v-2}^{+2}\,|v,2v-2\rangle^{*}\,, \end{aligned} \tag{2.3}$$

Therefore

$$\langle v,2v|\,O_{2v-2}^{+2}\,|v,2v-2\rangle = \frac{4v-3}{4v+1}\,a^{*}\,, \tag{2.4}$$

and consequently

$$\begin{aligned} &\langle v,2v-2|\,O_{2v}^{-2}|v,2v\rangle\,\langle v,2v|\,O_{2v-2}^{+2}\,|v,2v-2\rangle \\ &\quad = \frac{4v-3}{4v+1}\,|a|^{2}\,, \end{aligned}$$

or on account of the unicity and normalization properties of the $|v,2v\rangle$ states

$$\begin{aligned} |a|^{2} &= \frac{4v+1}{4v-3}\,\langle v,2v-2|\,O_{2v}^{-2}\,O_{2v-2}^{+2}\,|v,2v-2\rangle \\ &= \frac{4v+1}{4v-3}\,\langle v,2v|\,O_{2v-2}^{+2}\,O_{2v}^{-2}|v,2v\rangle\,. \end{aligned} \tag{2.5}$$

Since the expectation values at the rhs of (2.5) have been obtained in (II.4.19) already, we immediately find, except for a trivial phase factor, a to be given by

$$a = 2^{3}v^{2}(2v+1)(2v-1)(4v+1)\left(\frac{2}{3}\,\frac{(v-1)(2v-1)}{4v-3}\right)^{1/2}. \tag{2.6}$$

On the other hand we found by means of the construction formalism of Ref. 4, that there could be selected for each v value a unique quadrupole-phonon $|v,2v-2\rangle$ state (see Tables 1–4 of Ref. 4) which is written as

$$|v,2v-2\rangle = \mathcal{N}_{2v-2}\,|02I_{2}I_{3}\cdots I_{v}\rangle\,, \tag{2.7}$$

with $I_{v} = 2v-2$. Herein $\mathcal{N}_{2v-2}$ is the normalization factor for which the numerical values with $2\leqslant v\leqslant 6$ are found in Ref. 4. It is evident that the state $|v,2v-2\rangle$ as given by (2.7), being unique and normalized to unity is identical to the state $|v,2v-2\rangle$ as obtained by (2.2) and (2.6), namely

$$|v,2v-2\rangle = a^{-1}\,O_{2v}^{-2}|v,2v\rangle\,, \tag{2.8}$$

with a as given in (2.6).

As long as we restrict ourselves to nondegenerated states $|v,l\rangle$, the foregoing reasoning can be easily repeated, yielding that in the shift operator formalism $|v,l\rangle$ may always be constructed as

$$|v,l\rangle = a'_{k,l}\,O_{l+k}^{-k}\,|v,l+k\rangle\,, \tag{2.9}$$

where k is chosen from among the values ± 1, ± 2, ± 3 such that $|v,l+k\rangle$ is nondegenerated, and whereby the proportion factor $a'_{k,l}$ is generally given by

$$a'_{k,l} = \frac{2l+1}{2l+2k+1}\,(\langle v,l|\,O_{l+k}^{-k}\,O_{l}^{+k}|v,l\rangle)^{-1/2}\,. \tag{2.10}$$

Moreover $|v,l\rangle$ as determined by (2.9) and (2.10), is completely identical to the corresponding normalized $|v,l\rangle$ state expanded in terms of states belonging to the basis of quadrupole-phonon states selected in Ref. 4. We note that in some cases the state $|v,l\rangle$ is expanded as

$$|v,l\rangle = \mathcal{N}_{v,l}^{1}|\,0202J_{4}\cdots J_{v}\,\rangle + \mathcal{N}_{v,l}^{2}|02I_{2}I_{3}\cdots I_{v}\rangle\,, \tag{2.11}$$

with $J_{v} = I_{v} = l$, whereby $\mathcal{N}_{v,l}^{1}$ and $\mathcal{N}_{v,l}^{2}$ are coefficients whose numerical value is listed in Tables 1–4 of Ref. 4. We also remark from (2.9) that even in the present scheme different ways of generating the $|v,l\rangle$ state are at hand, each corresponding to a different k value, i.e.,

$$\begin{aligned} |v,l\rangle &= a'_{k_1,l}\,O_{l+k_1}^{-k_1}\,|v,l+k_1\rangle \\ &= a'_{k_2,l}\,O_{l+k_2}^{-k_2}\,|v,l+k_2\rangle\,, \quad (k_1\neq k_2)\,. \end{aligned} \tag{2.12}$$

For l-degenerated states $|v,l\rangle$ the explicit determination of the eigenstates by use of shift operators is somewhat more complicated. In order to keep the demonstration of the general procedure as simple as possible we confine it to the construction of the doubly degenerated orthonormal $|6,6\rangle$ states which we denote as $|6,6,(1)\rangle$ and $|6,6,(2)\rangle$, respectively, whereas the corresponding eigenvalues of O_{6}^{0} are written as $\alpha_{6,6}(1)$ and $\alpha_{6,6}(2)$. Since the $|6,7\rangle$ and $|6,8\rangle$ states are nondegenerated we are allowed to define

$$O_{7}^{-7}|6,7\rangle = \mu|6,6,(1)\rangle + \nu|6,6,(2)\rangle\,, \tag{2.13}$$

$$O_{8}^{-2}|6,8\rangle = \mu'|6,6,(1)\rangle + \nu'\,|6,6,(2)\rangle\,. \tag{2.14}$$

The problem is now to determine the four parameters μ,μ',ν, and ν' in terms of known quantities. From (2.13) it follows with the help of the properties (I.4.1) and (I.4.2) that

$$\begin{aligned} \langle 6,6,(1)|\,O_{7}^{-1}|6,7\rangle = \mu &= \frac{1}{\beta_{1,6}}\,\langle 6,6,(1)|\,(O_{6}^{+1})^{\dagger}|6,7\rangle \\ &= (15/13)\langle 6,7|\,O_{6}^{+1}|6,6,(1)\rangle^{*}\,, \end{aligned} \tag{2.15}$$

or

$$(13/15)\,\mu^{*} = \langle 6,7|\,O_{6}^{+1}|6,6,(1)\rangle\,. \tag{2.16}$$

Similarly, we find that

$$\begin{cases} (13/15)\nu^{*} = \langle 6,7|\,O_{6}^{+1}|6,6,(2)\rangle\,, \\ (13/17)\,\mu'^{*} = \langle 6,8|\,O_{6}^{+2}|6,6,(1)\rangle\,, \\ (13/17)\,\nu'^{*} = \langle 6,8|\,O_{6}^{+2}|6,6,(2)\rangle\,. \end{cases}$$

By considering the product $\langle 6,6,(1)|\,O_{7}^{-1}|6,7\rangle\,\langle 6,7|\,O_{6}^{+1}|6,6,(2)\rangle$ we deduce on account of the uniqueness of the normalized $|6,7\rangle$ state, and by use of (2.15) and (2.16) that

$$(13/15)|\mu|^{2} = \langle 6,6,(1)|\,O_{7}^{-1}O_{6}^{+1}|6,6,(1)\rangle\,, \tag{2.17}$$

It is straightforward to obtain in addition that

$$(13/15)|\nu|^{2} = \langle 6,6,(2)|\,O_{7}^{-1}O_{6}^{+1}|6,6,(2)\rangle\,, \tag{2.18}$$

$$(13/17)|\mu'|^{2} = \langle 6,6,(1)|\,O_{8}^{-2}O_{6}^{+2}|6,6,(1)\rangle\,, \tag{2.19}$$

$$(13/17)|\nu'|^{2} = \langle 6,6,(2)|\,O_{8}^{-2}O_{6}^{+2}|6,6,(2)\rangle\,, \tag{2.20}$$

and also that

$$(13/15)\mu\nu^* = \langle 6,6,(1)|\, O_7^{-1}O_6^{+1}|6,6,(2)\rangle\,, \quad (2.21)$$

$$(13/15)\mu^*\nu = \langle 6,6,(2)|\, O_7^{-1}O_6^{+1}|6,6,(1)\rangle\,, \quad (2.22)$$

$$(13/17)\mu'\nu'^* = \langle 6,6,(1)|\, O_8^{-2}O_6^{+2}|6,6,(2)\rangle\,, \quad (2.23)$$

$$(13/17)\mu'^*\nu' = \langle 6,6,(2)|\, O_8^{-2}O_6^{+2}|6,6,(1)\rangle\,, \quad (2.24)$$

In contrast to the case of no l degeneracies, none of the expressions at the rhs of (2.17)–(2.24) is directly obtained by the tree-generating mechanism of II. Indeed, by letting act on the states $|6,7\rangle$, $|6,8\rangle$ the suitable equations of the form (I.3.2)–(I.3.5) it is quite easy to derive the values of the expectation values $\langle 6,7|\, O_6^{+1}O_7^{-1}|6,7\rangle$ and $\langle 6,8|\, O_6^{+2}O_8^{-2}|6,8\rangle$, which on account of the property (I.4.3) are related to the expectation values in (2.15)–(2.18) in the following way:

$$\begin{aligned}\langle 6,7|\, O_6^{+1}O_7^{-1}|6,7\rangle &= \langle 6,6,(1)|\, O_7^{-1}O_6^{+1}|6,6,(1)\rangle \\ &\quad + \langle 6,6,(2)|\, O_7^{-1}O_6^{+1}|6,6,(2)\rangle\,, \end{aligned} \quad (2.25)$$

$$\begin{aligned}\langle 6,8|\, O_6^{+2}O_8^{-2}|6,8\rangle &= \langle 6,6,(1)|\, O_8^{-2}O_6^{+2}|6,6,(1)\rangle \\ &\quad + \langle 6,6,(2)|\, O_8^{-2}O_6^{+2}|6,6,(2)\rangle\,, \end{aligned} \quad (2.26)$$

By this it is proven that we have already at our disposition the values of $|\mu|^2 + |\nu|^2$ and $|\mu'|^2 + |\nu'|^2$. On the other hand, if we let Eq. (I.3.5), which for the present case ($v = 6$, $l = 6$) reads

$$\begin{aligned}2^7 3^2 5\ \ 13(O_6^0)^2 - \sqrt{2}\, 2^6 3^4 5^2 7\ \ 13(O_6^0) &= \frac{2\ 3^4 5^3}{7^2} O_7^{-1}O_6^{+1} + \frac{3^2 5^2 13}{2^5 7^2} O_8^{-2}O_6^{+2} \\ &\quad + 2^{10}3^7 5^2 7^2 13 + 2^{11}3^7 5\ \ 7^2 13\,, \end{aligned} \quad (2.27)$$

act at the right on $|6,6,(1)\rangle$, and multiply thereafter both sides at the left with $\langle 6,6,(1)|$ we have, since $\alpha_{6,6}(1)$ is known, a numerical expression for the sum $2^6 3^2 5\langle 6,6,(1)|\, O_7^{-1}O_6^{+1}|6,6,(1)\rangle + 13\langle 6,6,(1)|\, O_8^{-2}O_6^{+2}|6,6,(1)\rangle$, or on account of (2.15) and (2.17) also a numerical expression for the sum $2^6 3\ \ 17|\mu|^2 + 13|\mu'|^2$. Replacing $|6,6,(1)\rangle$ by $|6,6,(2)\rangle$ a numerical expression is also obtained for $2^6 3\ \ 17|\nu|^2 + 13|\nu'|^2$ by use of $\alpha_{6,6}(2)$. Having found in this way four independent equations in the unknowns $|\mu|^2$, $|\mu'|^2$, $|\nu|^2$, and $|\nu'|^2$, a solution for μ,μ', ν, and ν' is readily deduced when leaving aside phase factors. With this solution the state $|6,6,(1)\rangle$ and $|6,6,(2)\rangle$ can be constructed from (2.13) and (2.14) in terms of the expressions $O_7^{-1}|6,7\rangle$ and $O_8^{-2}|6,8\rangle$. However, the mentioned phase factors are in the present case not completely irrelevant, since they must be chosen such that $|6,6,(1)\rangle$ and $|6,6,(2)\rangle$ are mutually orthogonal. In fact, this supplementary condition can be mathematically implied by, for instance, letting (2.27) act on $|6,6,(1)\rangle$, while multiplying thereafter both sides at the left with $\langle 6,6,(2)|$. By this we obtain an expression which on account of the orthogonality of $|6,6,(1)\rangle$ and $|6,6,(2)\rangle$ reads:

$$2^6 3\ \ 17\mu^*\nu + 13\mu'^*\nu' = 0\,. \quad (2.28)$$

Interchanging the roles of $|6,6,(1)\rangle$ and $|6,6,(2)\rangle$, the complex conjugate of (2.28) is found. Now (2.28) indicates for instance that by restricting μ,μ', ν, and ν' to real values, the sign of three of these four quantities can be chosen in an arbitrary way whereafter the sign of the remaining fourth quantity is unambiguously fixed.

Let us draw once more attention to the fact that for constructing states in terms of shift operators acting on one or more other selected states, we have usually more than one possibility for choosing the latter. In the present case, for example, instead of starting from (2.13) and /or (2.14), we could have started from similar equations for $O_9^{-3}|6,9\rangle$, $O_4^{+2}|6,4\rangle$, $O_3^{+3}|6,3\rangle$, as well. Of course, although various expressions for $|6,6,(1)\rangle$ and $|6,6,(2)\rangle$ exist, these orthogonal states are, once normalized, unique.

Previously we have constructed in a numerical way two independent orthonormalized $N = 6$, $v = 6$, $l = 6$ states[4] as linear combinations of the basis states $|0202246\rangle$, $|0220246\rangle$, and $|0223456\rangle$. This construction was, however, not unique as long as there was missing a label-generating operator for distinguishing the states. To overcome this problem we have introduced[5] the operator S given in (1.1), for which it was shown[6] that it generates two independent orthonormalized $N = 6$, $v = 6$, $l = 6$ states which are in a unique way expressible as linear combinations of the foregoing basis states.

Having thus at our disposition the $|6,6\rangle$ eigenstates of O_l^0 and those of S, we can ask whether these states coincide or not. For answering this question let us be reminded that it was shown[5] that any two operators which contain at most four elementary phonon-creation and four annihilation operators, and which commute with L^2, l_0, V^*, and N, are necessarily dependent on each other, in the sense that the first, say T_1, may be expressed in terms of the second, say T_2, as follows:

$$\begin{aligned}T_1 = \mathcal{F}(T_2, &N, N^2, L^2, V^*, N^3, L^2N, V^*N, N^4, \\ &L^2N^2, V^*N^2, L^4, V^{*2}, L^2V^*)\,, \end{aligned} \quad (2.29)$$

whereby $\mathcal{F}$ is a linear functional. Moreover, although the eigenvalue spectrum of T_1 generally differs from the T_2 spectrum, both operators have the same eigenstates. In our present case we can replace T_1 by O_l^0 and T_2 by S, showing that O_l^0 must be expressible in terms of S by a relation of the kind (2.29), and that the O_l^0 eigenstates as constructed for instance by means of (2.13) and (2.14) are identical to the numerically constructed eigenstates of S.[6] In Sec. 3 we shall derive the relationship between O_l^0 and S in full detail.

3. THE RELATION BETWEEN O_l^0 AND S

At this point there is still a fundamental difference in working with either O_l^0 or S. Indeed, the operator S acts in the space of the symmetric states of the group SU(5), while so far O_l^0 only acts in the subspace of the symmetric R(5) states, with the consequence that for the latter operator only seniority states with $N = v$ had to be considered. However, since R(5)⊂SU(5) we can quite easily determine the action of O_l^0 on SU(5) states too. Therefore we notice that for the latter states one necessarily has that $N = v + 2n$, whereby n is zero or a positive integer, and that one can write

$$|\, N = v + 2n, v, l\,\rangle = [\otimes |2,0\rangle]^n \otimes |\, N = v, v, l\,\rangle\,. \quad (3.1)$$

TABLE I. Eigenvalues of the operators $5O_l^0/6\sqrt{2}, N, N^2, \ldots, L^2V^*, S$ on selected states $|N,v,l\rangle$.

$\|N,v,l\rangle$	$5O_l^0/6\sqrt{2}$	N	N^2	L^2	V^*	N^3	L^2N	V^*N	N^4	L^2N^2	V^*N^2	L^4	V^{*2}	L^2V^*	S
$\|1,1,2\rangle$	14	1	1	6	−2	1	6	−2	1	6	−2	36	4	−12	0
$\|2,0,0\rangle = \|2,0\rangle \otimes \|0,0,0\rangle$	0	2	4	0	0	8	0	0	16	0	0	0	0	0	0
$\|2,2,2\rangle$	−16	2	4	6	−5	8	12	−10	16	24	−20	36	25	−30	0
$\|2,2,4\rangle$	110	2	4	20	−5	8	40	−10	16	80	−20	400	25	−100	0
$\|3,3,0\rangle$	0	3	9	0	−9	27		−27	81	0	−81	0	81	0	0
$\|3,1,2\rangle = \|2,0\rangle \otimes \|1,1,2\rangle$	14	3	9	6	−2	27	18	−6	81	54	−18	36	4	−12	0
$\|3,3,3\rangle$	−15	3	9	12	−9	27	36	−27	81	108	−81	144	81	−108	0
$\|3,3,4\rangle$	35	3	9	20	−9	27	60	−27	81	180	−81	400	81	−180	0
$\|3,3,6\rangle$	420	3	9	42	−9	27	126	−27	81	378	−81	1764	81	−378	0
$\|4,0,0\rangle = [\otimes \|2,0\rangle]^2 \otimes \|0,0,0\rangle$	0	4	16	0	0	64	0	0	256	0	0	0	0	0	0
$\|4,4,2\rangle$	29	4	16	6	−14	64	24	−56	256	96	−224	36	196	−84	0
$\|4,2,2\rangle = \|2,0\rangle \otimes \|2,2,2\rangle$	−16	4	16	6	−5	64	24	−20	256	96	−80	36	25	−30	0
$\|4,4,4\rangle$	−160	4	16	20	−14	64	80	−56	256	320	−224	400	196	−280	0
$\|4,2,4\rangle = \|2,0\rangle \otimes \|2,2,4\rangle$	110	4	16	20	−5	64	80	−20	256	320	−80	400	25	−100	0
$\|4,4,6\rangle$	280	4	16	42	−14	64	168	−56	256	672	−224	1764	196	−588	60/7

Furthermore, since the action of O_l^0 is confined to the space of $|N=v,v,l\rangle$ states we immediately find that

$$O_l^0|N=v+2n,v,l\rangle = [\otimes|2,0\rangle]^n \otimes \{O_l^0|N=v,v,l\rangle\} = \alpha_{v,l}|N=v+2n,v,l\rangle \quad (n=0,1,2,\cdots), \tag{3.2}$$

if $|N=v,vl\rangle$ is an eigenstate of O_l^0 with eigenvalue $\alpha_{v,l}$, showing that all states $|N=v+2n,v,l\rangle$ are also eigenstates of O_l^0 with the same eigenvalue as $|N=v,v,l\rangle$.

We now have all elements to express O_l^0 in terms of S and other operators such as in (2.29) by comparing the action of both operators on a particular set of common eigenstates. More specifically we draw attention to the fact that

$$S|N,v,l\rangle = 0 \quad \text{if} \quad N<4. \tag{3.3}$$

On the other hand it follows from Racah algebra that

$$O^{235}_{235}|4,4,5\rangle = 9|4,4,5\rangle, \tag{3.4}$$

$$O^{246}_{246}|4,4,6\rangle = (60/7)|4,4,6\rangle, \tag{3.5}$$

$$O^{468}_{468}|4,4,8\rangle = 24|4,4,8\rangle, \tag{3.6}$$

and with the second of these results substituted in (1.1) we rigorously obtain that

$$S|4,4,6\rangle = (60/7)|4,4,6\rangle. \tag{3.7}$$

In Table I we list the eigenvalues of $5/6\sqrt{2}\ O_l^0, N, N^2, L^2, V^*, N^3, L^2N, V^*N, N^4, L^2N^2, V^*N^2, L^4, V^{*2}, L^2V^*$, and S for all their common eigenstates for which, in particular, the S eigenvalue is rigorously known. These eigenvalues are used to express $5/6\sqrt{2}\,O_l^0$ as a linear combination of the fourteen operators $N, N^2, \ldots, L^2V^*$, and S. As expected, the system of fifteen linear equations in fourteen unknown coefficients is consistent, and leads to the unique solution

$$\begin{aligned}\frac{5}{6\sqrt{2}}O_l^0 = {} & \frac{2160}{7}N - \frac{1980}{7}N^2 - \frac{113}{4}L^2 \\ & + \frac{675}{14}V^* + \frac{540}{7}N^3 + 30L^2N \\ & - \frac{180}{7}V^*N - \frac{45}{7}N^4 \\ & - 5L^2N^2 + \frac{30}{7}V^*N^2 + \frac{3}{8}L^4 \\ & + \frac{25}{14}V^{*2} + \frac{5}{2}L^2V^* + \frac{154}{3}S.\end{aligned} \tag{3.8}$$

By this formula it thus also becomes possible to calculate an S eigenvalue if the corresponding O_l^0 eigenvalue is known. In particular, if we consider the state $|4,4,5\rangle$, we know from Table II of Paper II that $(5/6\sqrt{2})\,O_5^0|4,4,5\rangle = 195|4,4,5\rangle$. Making use of (3.8) it is readily calculated that

$$S|4,4,5\rangle = (600/77)|4,4,5\rangle. \tag{3.9}$$

On the other hand the expression (1.1) for S in terms of canonical operators showed that the only contribution to the S eigenvalue originates from the action of the O^{235}_{235} term only, which was found in (3.4). We therefore deduce that the O^{235}_{235} coefficient in the S expansion (1.1) is exactly $(600/77)/9 = 200/231$. In the same way we calculate from (3.8) that

$$S|4,4,8\rangle = (2145/154)|4,4,8\rangle, \tag{3.10}$$

while on the other hand this eigenvalue is produced uniquely by the O^{468}_{468} term in (1.1). It follows from (3.6) that the accompanying coefficient is exactly 65/112. Therefore, we can replace now the numerically determined expansion (1.1) for S by the following expression:

$$S = \frac{(2^3)(5^2)}{(3)(7)(11)}O^{235}_{235} + O^{246}_{246} + \frac{(5)(13)}{(2^4)(7)}O^{468}_{468}. \tag{3.11}$$

We note that the previously obtained numerical values of (1.1) approximate the exact corresponding values of (3.11) with an error smaller than 5×10^{-5}.

Finally, we can exploit Eq. (3.8) to calculate rigorously the S eigenvalues associated with the states $|6,4,6\rangle$, $|6,6,6,(1)\rangle$, and $|(6,6,6,(2)\rangle$ with $N=6$ and $l=6$, which we determined numerically as 50.6502, 79.2477, and 50.2009 in a previous paper.[6] From Table II of Paper II we find that

$$(5/6\sqrt{2})\,O_6^0|6,4,6\rangle = 280|6,4,6\rangle,$$

$$(5/6\sqrt{2})\,O_6^0|6,6,6,(1)\rangle = (15/2)[-5+(9881)^{1/2}]|6,6,6,(1)\rangle,$$

and

$$(5/6\sqrt{2})\,O_6^0|6,6,6,(2)\rangle = (15/2)[-5-(9881)^{1/2}]|6,6,6,(2)\rangle.$$

Substituting the latter results consecutively in (3.8) we deduce that

$$S|6,4,6\rangle = \frac{(2^2)(3)(5^2)(13)}{(7)(11)}|6,4,6\rangle, \tag{3.12}$$

$$S|6,6,6,(1)\rangle = \frac{(3^2)(5)}{(2^2)(7)(11)}[443+(9881)^{1/2}]\,|6,6,6,(1)\rangle, \tag{3.13}$$

$$S|6,6,6,(2)\rangle = \frac{(3^2)(5)}{(2^2)(7)(11)}[443-(9881)^{1/2}]|6,6,6,(2)\rangle, \tag{3.14}$$

showing that the mentioned values numerically obtained by Racah algebra[6] are very close to the exact values of (3.12)–(3.14), namely within 10^{-5} of the relative difference.

[1] G. Vanden Berghe and H.E. De Meyer, J. Math. Phys. **21**, 1902 (1980).
[2] H.E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906 (1980).
[3] H.J. Weber, M.G. Huber, and W. Greiner, Z. Phys. **190**, 25 (1966).
[4] G. Vanden Berghe and H.E. De Meyer, Nucl. Phys. A **323**, 302 (1979).
[5] H.E. De Meyer and G. Vanden Berghe, J. Math. Phys. (to be published).
[6] G. Vanden Berghe and H.E. De Meyer, J. Phys. A (to be published).

On some solutions of the polaron model

G. A. Ringwood
Department of Mathematics, University of Durham, and Department of Physics, University of Antwerpen
J. T. Devreese
Institute of Applied Mathematics and Department of Physics, University of Antwerpen (R.U.C.A. and U.I.A.), c/o Universiteitsplein 1, 2610 Wilrijk, Belgium

(Received 7 August 1979; accepted for publication 10 December 1979)

A paper by one of the present authors (JTD) and others obtained solutions of the classical polaron model in which the electron rotates about the stationary polaron center. The radius of gyration of the electron was, however, shown to diverge with the radius of the Brillouin zone. In this paper the analysis is carried out more extensively and concisely and it is found that there are solutions for which the radius of gyration remains finite.

1. INTRODUCTION

Exact solutions to classical field theories are interesting not only in their own right but also for semiclassical approximations to the quantum mechanical problem. This paper extends results on the behavior of some classical solutions of the polaron model found in Ref. 1.

Reference 1 gives the first term for the series and asymptotic expansions of the solutions for a restricted range of the electron rotation frequency. The present paper gives the complete series and asymptotic expansions for the whole range of the electron frequency. In the frequency range that was neglected in Ref. 1 some interesting structure appears.

In Ref. 1 a Gaussian cutoff was used to facilitate the analytic evaluation of integrals. In the present paper the more natural Heaviside cutoff is employed. Where results in the two papers are duplicated the different cutoffs lead to unimportant differences in numerical factors. (A choice of cutoff represents a choice of the short range behavior of the interaction. It is the long range behavior which is essentially being examined in this model.)

2. SERIES EXPANSION

The Frölich Hamiltonian for the classical polaron model is

$$p^2/2m + \sum_{\mathbf{k}} \omega a_{\mathbf{k}}^{\dagger} a_{\mathbf{k}} + \sum_{\mathbf{k}} (V_k a_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{r}) + V_k^{\dagger} a_{\mathbf{k}}^{*} \exp(-i\mathbf{k}\cdot\mathbf{r})),$$

where $V_k = -2i\pi^{1/2}\alpha^{1/2}(2m\omega)^{-1/4}\omega k^{-1}\hbar V^{-1/2}$. The equations of motion are

$$\dot{\mathbf{p}} = -i \sum_{\mathbf{k}} \mathbf{k}(V_k a_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{r}) - V_k^{*} a_{\mathbf{k}}^{\dagger} \exp(-i\mathbf{k}\cdot\mathbf{r})), \tag{2.1}$$

$$a_{\mathbf{k}}(t) = -i\hbar^{-1} V_k^{*} \exp(-i\omega^{+}t) \times \int_{-\infty}^{t} dt' \exp(-i\mathbf{k}\cdot\mathbf{r}(t') + i\omega^{+}t'), \tag{2.2}$$

$a_{\mathbf{k}}(-\infty)$ assumed zero.

The paper by Evrard *et al.*[1] suggests a trial solution

$$x(t) + iy(t) = u \exp(i\Omega t), \quad z(t) = 0.$$

Expanding $\exp(i\mathbf{k}\cdot\mathbf{r})$ in terms of Bessel functions, with this trial solution, the integral in Eq. (2.2) can be done. Substituting the result in Eq. (2.1) and appealing to cylindrical symmetry gives, after some rearrangement, the consistency condition

$$mu\Omega^2 = (\partial/\partial u) \sum_{\mathbf{k}} |V_k|^2 \sum_{n=-\infty}^{\infty} J_n^2(ku\sin\theta)/(\Omega n - \omega^{+}).$$

The trial solution has energy

$$E = \tfrac{1}{2} mu\Omega^2 + \sum_{\mathbf{k}} |V_k|^2 (2 + \omega(\partial/\partial\omega)) \times \sum_{n=-\infty}^{\infty} J_n^2(ku\sin\theta)/(\Omega n - \omega^{+}).$$

It is clear that viable solutions depend on the successful interpretation of

$$8\pi^3 V^{-1} \sum_{\mathbf{k}} k^{-2} \sum_{n=-\infty}^{\infty} J_n^2(ku\sin\theta)/(\Omega n - \omega^{+}). \tag{2.3}$$

Using a Sommerfeld–Watson transformation the sum over n can be performed:

$$\sum_{n=-\infty}^{\infty} J_n^2(ku\sin\theta)/(\Omega n - \omega^{+}) = -\pi\Omega^{-1} \csc(\pi\omega/\Omega) J_{-\omega/\Omega}(ku\sin\theta) J_{\omega/\Omega}(ku\sin\theta).$$

If the product of two Bessel functions is replaced by the integral representation

$$2\pi^{-1} \int_0^{\pi/2} dt\, J_0(2ku\sin\theta\cos t)\cos(2\omega t/\Omega),$$

the summation over $\mathbf{k}$ (which, as is customary, is replaced by an integral over a spherical Brillouin zone of radius K) can be done (the angular integrations first) so that line (2.3) becomes

$$-4\pi(u\Omega)^{-1} \csc(\pi\omega/\Omega) \times \int_0^{\pi/2} dt\, \mathrm{Si}(2Ku\cos t)\cos(2\omega t/\Omega)\sec t.$$

The consistency condition and the energy are then given by, respectively,

$$\Omega^3/\omega^3 = -4\gamma K \csc(\pi\omega/\Omega) u^{-1}(\partial/\partial u) f((Ku)^2, \omega/\Omega),$$

 0022-2488/80/071918-03$1.00

$$E = -4\gamma m\omega^2 K(\omega/\Omega)\csc(\pi\omega/\Omega)(\tfrac{1}{2}u(\partial/\partial u) + 2 - \pi(\omega/\Omega)\cot(\pi\omega/\Omega) + \omega(\partial/\partial\omega)) \times f((Ku)^2, \omega/\Omega),$$

where

$$f(\xi,\rho) = \pi^{-1}\xi^{-1/2}\int_0^{\pi/2} dt\, \mathrm{Si}(2\xi^{1/2}\cos t)\cos(2\rho t)\sec t$$

and

$$\gamma = \alpha(\hbar/2m\omega)^{3/2}.$$

Substituting the power series expansion for the sine integral in the definition of f and performing the integration term by term leads to the following power series expansion for f:

$$f(\xi,\rho) = \sum_{n=0}^{\infty}(-1)^n\xi^n/(2n+1)^2\Gamma(n+1+\rho) \times\Gamma(n+1-\rho).$$

The small Ku approximation to the consistency condition and the energy is given by taking the first two terms in the series:

$$\Omega^2/\omega^2 \simeq 1 + 8(9\pi)^{-1}\gamma K^3,$$

$$E \simeq -4\pi^{-1}\gamma m\omega^2 K[1 - 2(9\pi)^{-1}(Ku)^2(1-\omega^2/\Omega^2)^{-2}].$$

The first term in the energy is the lattice polarization energy

$$-\sum_{\mathbf{k}}(\hbar\omega)^{-1}|V_k|^2.$$

3. ASYMPTOTIC EXPANSION

As in general the derivative of an asymptotic series is not the asymptotic series of the derivative, any differentiations have to be performed first. The method of obtaining an asymptotic series from a power series[2] will be demonstrated on f. The procedure for its derivatives follows the same pattern. The power series for f is

$$f(\xi,\rho) = \sum_{n=0}^{\infty}(-1)^n\mu(n,\rho)\xi^n,$$

where

$$\mu(z,\rho) = ((2z+1)^2\Gamma(z+1+\rho)\Gamma(z+1-\rho))^{-1}.$$

Consider the integral

$$(2i)^{-1}\int_C dz\,\mu(z,\rho)\xi^z\csc(\pi z),$$

around the square $-N+1/2+iM/2$, $-N+1/2-iM/2$, $M-N+1/2-iM/2$, $M-N+1/2+iM/2$ with M and N integers, $M>N>2$. It can be argued that the contribution from the line integrals along the bottom, right-hand side and top of the square tend to zero as M tends to infinity for $|\arg\xi| < \pi$. Thus the integral along the line $-N+1/2+i\infty$ to $-N+1/2-i\infty$ is equal to π times the sum of all the residues to the right of it. The residue at the pole $z=n$ of $\csc(\pi z)$ is $\pi^{-1}(-1)^n\mu(n,\rho)\xi^n$ so that the sum includes $f(\xi,\rho)$. As $|\xi|$ tends to infinity with $|\arg\xi| < \pi$ the line integral along $-N+1/2+i\infty$ to $-N+1/2-i\infty$ can be shown to be of order $|\xi|^{-N}$ so that the conditions for an asymptotic series are satisfied. The only pole of $\mu(z,\rho)$ at $z=-1/2$ gives the leading contribution to the asymptotic series:

$$f(\xi,\rho) \sim \tfrac{1}{4}\xi^{-1/2}\cos(\pi\rho)[\ln\xi - \psi(\tfrac{1}{2}+\rho) - \psi(\tfrac{1}{2}-\rho)] - \sum_{n=1}^{\infty}(-1)^n\xi^{-n}/(2n-1)^2\Gamma(1+\rho-n)\Gamma(1-\rho-n).$$

The asymptotic series for the derivatives of f are found similarily:

$$(\partial/\partial\xi)f(\xi,\rho) \sim -\tfrac{1}{8}\xi^{-1/2}\cos(\pi\rho)[\ln\xi - 2 - \psi(\tfrac{1}{2}+\rho) - \psi(\tfrac{1}{2}-\rho)] - \sum_{n=2}^{\infty}(-1)^n\xi^{-n}(n-1)/(2n-3)^2\Gamma(2+\rho-n)\Gamma(2-\rho-n);$$

$$(\partial/\partial\rho)f(\xi,\rho) \sim -\tfrac{1}{4}\xi^{-1/2}\cos(\pi\rho)[\pi\tan(\pi\rho)\ln\xi + \psi'(\tfrac{1}{2}+\rho) - \psi'(\tfrac{1}{2}-\rho) - \pi\tan(\pi\rho)(\psi(\tfrac{1}{2}+\rho) + \psi(\tfrac{1}{2}-\rho))] - \sum_{n=1}^{\infty}(-1)^n\xi^{-n}(\psi(1-\rho-n) - \psi(1+\rho-n))/(2n-1)^2\Gamma(1+\rho-n)\Gamma(1-\rho-n).$$

The two leading terms in the asymptotic series give the asymptotic form of the consistency condition and the energy for large Ku:

$$u^3\Omega^3/\omega^3 \sim 2\gamma\cot(\pi|\omega/\Omega|)[\ln(Ku) - 1 - \psi(\tfrac{1}{2}+|\omega/\Omega| = +|\omega/\Omega|) - \cot(\pi|\omega/\Omega|) - 3\pi/2].$$

$$E \sim \pi\gamma m\omega^2 u^{-1}|\omega/\Omega|[(\ln(Ku) - \psi(\tfrac{1}{2}+|\omega/\Omega|))(2\pi|\omega/\Omega|\csc^2(\pi\omega/\Omega) - 3\cot(\pi|\omega/\Omega|)) + 2|\omega/\Omega|\cot(\pi|\omega/\Omega|)\psi'(\tfrac{1}{2}+|\omega/\Omega|) - \cot(\pi|\omega/\Omega|) - 3\pi/2].$$

4. CONCLUSIONS

The asymptotic form of the solution obtained in Ref. 1 is applicable only for $|\omega/\Omega| < 1$ whereas above no such restriction holds. When $|\omega/\Omega| < 1$ the solution takes the form

$$u^3 \sim 2\gamma\pi^{-1}|\omega/\Omega|^2\ln(Ku). \qquad (4.1)$$

For K sufficiently greater than 1 and $|\omega/\Omega|$ fixed (4.1) will have two solutions for u. One solution is of the order of $1/K$ and is not admissable as it does not lie in the asymptotic regime. As $|\omega/\Omega|$ increases the term on the right-hand side of 4.1 is modified by a resonance factor $\pi|\omega/\Omega| \times\cot(\pi|\omega/\Omega|)$. As $|\omega/\Omega|$ approaches an integer the radius of gyration of the electron diverges. (The behavior is not apparent in Ref. 1.)

For $|\omega/\Omega| < 1$ the energy is negative

$$E \sim -\gamma m\omega^2 u^{-1}\ln(Ku),$$

but as $|\omega/\Omega|$ increases it is modified by a factor

$$\pi|\omega/\Omega| \csc^2(\pi|\omega/\Omega|)(\tfrac{3}{2}\sin(2\pi|\omega/\Omega|) - 2\pi|\omega/\Omega|).$$

This changes sign when $|\omega/\Omega|$ is approximately 1/4. The energy diverges positively at the resonance positions.

In general the radius of gyration diverges with the cutoff K but the modification provided by the resonance factor $\pi|\omega/\Omega|\cot(\pi|\omega/\Omega|)$ has the effect that when the gyration frequency is completely out of phase, $|\omega/\Omega|$ half-odd-integer, the angular momentum takes a constant value

$$u^3\Omega^3 = \gamma\pi\omega^3$$

independent of the cutoff. Thus there is a countable set of bound states which persist in the limit in which K tends to infinity. However, the energy of these solutions still diverges.

[1]R. Evrard, J.T.Devreese, E. Kartheuser, and C.C. Grosjean, Nuovo Cimento B **12**, 118 (1972).
[2]M.A. Evgrafov, *Asymptotic Estimates of Entire Functions* (Gordon and Breach, New York, 1961).

Quantum Heisenberg ferromagnets and stochastic exclusion processes

Lawrence E. Thomas [a)]
Department of Mathematics, University of Virginia, Charlottesville, Virginia 22903

(Received 21 December 1979; accepted for publication 4 March 1980)

Let H be the isotropic spin-s quantum ferromagnetic Heisenberg Hamiltonian associated with a finite volume in $\mathbb{Z}^d$. In an appropriate representation, the restriction of $-H$ to the N-spin wave sector $-H_N$ can be regarded as the generator of a Markov semigroup $\exp(-tH_N)$. For $s=\frac{1}{2}$ the corresponding process $X_N(t)$ is identified as a simple N-particle exclusion process; for higher spin it is a more complicated exclusion process involving different species of particles. A Feynman–Kac, Cameron–Martin perturbation formula for $\exp(-tH_N)$ is developed in the $s=\frac{1}{2}$ case, and the formula is used to obtain some thermodynamic correlation inequalities.

I. INTRODUCTION

Let $H=-\Sigma'_{m,m'\in\Lambda}[\mathbf{J}(m)\cdot\mathbf{J}(m')-s^2]$ be the isotropic spins-s Heisenberg Hamiltonian associated with a finite rectangular volume $\Lambda\subset\mathbb{Z}^d$. Here, Σ' denotes summation over nearest neighbors, with periodic boundary conditions. Then H commutes with $\hat{N}=\Sigma_{m\in\Lambda}[(J_3(m)+s]$ and acts invariantly on the set of states $\mathscr{H}_N$ whose eigenvalue under $\hat{N}$ is N ($N=0,1,2,\cdots,2s|\Lambda|$). Denote by H_N the restriction of H to $\mathscr{H}_N$.

The subspace $\mathscr{H}_N$ can be realized as a (finite dimensional) l^2 space in such a manner that the off-diagonal part of $-H_N$ is positive, so that in this representation $-H_N$ generates a positivity preserving semigroup. If the underlying measure for the l^2 space is chosen appropriately so that the ground state of $-H_N$ is realized as a constant function, the semigroup is Markovian. [For Schrödinger operators, the representation is that associated with the Jacobi method of multiplicative variation (Ref. 1, p. 458).]

Of course, the Hamiltonian need not be isotropic to permit this representation. But if it is isotropic, the process $X_N(t)$ associated with the Markov semigroup can be described explicitly and is conceptually simple. For example, in the spin-$\frac{1}{2}$ case $X_N(t)$ is a simple N-particle exclusion process, one which has been studied as a model for time evolution of *classical* lattice systems.[2] In the spin-1 case, one can represent $X_N(t)$ as a more complicated exclusion-like proces involving two sorts of particles with creation and annihilation.

We note that the idea of associating a stochastic process with the isotropic spin-$\frac{1}{2}$ case is not new. Powers,[3] and earlier Hurst and Sherman,[4] based on an idea of Dirac related it to a random walk on the permutation group of Λ to obtain correlation inequalities. For related ideas, see Powers[5] on resistance inequalities.

In Sec. III, we give a path space integral expression for the spin-$\frac{1}{2}$ semigroup $\{e^{-tH_N}\}$ and perturbations of it of the sort $H_N\to H_N+\delta V+\delta W$ with δV a multiplication operator, δW a "drift" term (Feynman–Kac, Cameron–Martin formula). In Sec. IV, we illustrate the use of the formula by obtaining some thermodynamic correlation inequalities [see Lemmas 4.1, 4.2, and Remarks]. The first inequality, originally obtained by Hurst, Sherman, and Powers[3,4] in the spin-$\frac{1}{2}$ case, is a special case of the first Griffiths inequality. The second inequality, Lemma (4.2), is closely related to a diamagnetic inequalities for Schrödinger operators, and can be interpreted as a bound on short range order.

II. MARKOV SEMIGROUP PROPERTY AND EXAMPLE

We recall a fact concerning finite dimensional Markov semigroups and we give an abstract finite dimensional version of the method of multiplicative variation.

Proposition 2.1 (cf. Ref. 6, p. 114): *In a finite dimensional state space $\{e^{tG}\}$ is a Markov semigroup iff G is a matrix with entries satisfying $g_{ij}\geqslant 0$ for $i\neq j$ and $\Sigma_j g_{ij}=0$, for each i,* i.e., *G has positive off-diagonal elements and annihilates constants.*

Lemma 2.2: *Let $\mathscr{H}$ be a finite dimensional Hilbert space with orthonormal basis $\{e_1,e_2,\ldots,e_n\}$, and let $\alpha_1,\alpha_2,\ldots,\alpha_n$ be a set of complex numbers, $\alpha_i\neq 0$, $i=1,\ldots n$.* Define *U: $\mathscr{H}\to l^2(\{1,2,\ldots,n\},\mu)$ with $\mu(\{k\})=|\alpha_k|^2$ and $Ue_j=f_j$ with $f_j(k)=(1/\alpha_j)\delta_{jk}$. Then U is unitary. Suppose further $\phi_0=\Sigma\alpha_j e_j\in\mathscr{H}$. Then $U\phi_0$ is identically one on $\{1,2,\ldots,n\}$.*

Proof: The operator $U$is clearly unitary. We have $U\phi_0(k)=\Sigma_j\alpha_j f_j(k)=\Sigma_j\alpha_j(1/\alpha_j)\delta_{jk}=1$.

Proposition 2.3: (Method of Multiplicative Variation): Let *$\mathscr{H}$, U,* and *ϕ_0 be as in Lemma* 2.2, *with $\alpha_i>0$, $i=1,\cdots,n$. Suppose G acts in $\mathscr{H}$ so that $Ge_j=\Sigma_i g_{ij}e_i$ with $g_{ij}\geqslant 0$ for $i\neq j$ and $G\phi_0=0$. Then $\tilde{G}\equiv UGU^{-1}$ is the generator of a Markov semigroup.*

Proof: The operator $\tilde{G}$ annihilates constants since $\tilde{G}\mathbb{1}=\tilde{G}U\phi_0=UG\phi_0=0$. On the other hand, the matrix for $\tilde{G}$is easily computed to be $\tilde{g}_{ij}=\alpha_i^{-1}g_{ij}\alpha_j$ which has positive off-diagonal elements. ∎

We now apply Proposition 2.3 to H_N. The state space is identified with the orthonormal basis for $\mathscr{H}_N$, $\{\text{const}\times J_+^{l_1}(m_1)\cdots J_+^{l_r}(m_r)\Omega,\ m_i\in\Lambda,\ \Sigma_{i=1}^r l_i=N\}$. (The vector $\Omega=\otimes|-s\rangle$ is the normalized ground state in which the J_3 spin at each site is $-s$.) The off-diagonal part of $-H_N$ is just $\frac{1}{2}\Sigma'[J_+(m)J_-(m')+J_-(m)J_+(m')]$; with the canonical choice of positive Clebsch–Gordan coefficients for the raising and lowering operators (Ref. 7, p. 25), one sees that the off-diagonal part of $-H_N$ has positive entries. Note further that $-H$ annihilates Ω, and commutes with $\Sigma_{m\in\Lambda}J_i(m)$, $i=1,2,3$ and in particular with $\Sigma_{m\in\Lambda}J_+(m)$. Thus, H_N annihilates $\phi_0\equiv(1/N!)[\Sigma_{m\in\Lambda}J_+(m)]^N\Omega$ which

a)Partially supported by NSF-MCS 74-07313-A03.

 0022-2488/80/071921-04$1.00

has all its Fourier coefficients strictly positive. By Proposition 2.3, we have the following result:

Theorem 2.4: *Let H_N be the restriction of the spin-s Heisenberg Hamiltonian H to the N-spin wave sector. Then $-H_N$ is unitarily equivalent to the generator of a Markov semigroup.*

Example 1: (spin-$\frac{1}{2}$ $\sim$ simple exclusion process): The normalized N-spin wave states are given by $O_N = \{J_+(m_1)J_+(m_2)\cdots J_+(m_N)\Omega \,|m_1,\cdots,m_N \in \Lambda$ and distinct$\}$ and the constant function is identified with $(1/N!)\,[\Sigma\, J_+(m_i)]^N\,\Omega = \Sigma\, J_+(m_1)\cdots J_+(m_N)\,\Omega$, where the latter sum extends over all $(m_1,\ldots,m_N)$ with m_i distinct. The state $J_+(m_1)\cdots J_+(m_N)\Omega$ corresponds to a configuration of N particles with positions $m_1,\ldots,m_N$ in Λ. Consider first the off-diagonal part of $-H_N$, i.e., $\Sigma'\frac{1}{2}\,[J_+(m)J_-(m') + J_-(m)J_+(m')]$. This operator has the effect of causing a particle to jump with weight $\frac{1}{2}$ to a nearest neighbor site if that site is unoccuppied, i.e., if the jump is not forbidden. If none of the particles are neighboring, there are $2dN$ such jumps possible (d is the lattice dimension).

We turn next to the diagonal part of $-H$. This part $\Sigma'[J_3(m)J_3(m') - \frac{1}{4}]$, applied to a vector in O_N, is equal to (has eigenvalue) $-\frac{1}{2}\times$ number of occupied–unoccupied "bonds." If the particles are the not neighboring, there are $2dN$ occupied–occupied bonds. If some are neighboring, there are $2dN - j$ such bonds, where j is twice the number of occupied–occupied bonds, or in other words where j is the number of forbidden jumps. In summary, this part of $-H_N$ can be written

$$\sum[J_3(m)\,J_3(m') - \tfrac{1}{4}] = -dN + \tfrac{1}{2}j. \tag{2.1}$$

This gives us that $\{P^n\}$ with $P \equiv 1 + (1/2dN)(-2H_N)$ is (unitarily equivalent to) a discrete time Markov semigroup. At each step a particle is selected at random and moved in a random direction one step unless the jump is forbidden in which case the particle configuration remains the same. The total probability of the configuration remaining the same is proportional to the number of forbidden jumps. Thus, P is the transition probability for a discrete time N particle simple exclusion process (cf. Ref. 2).

We now make the time continuous by randomizing the jumps according to a Poisson distribution.

$$K_t = e^{-dNt}\sum\frac{(dNt)^n}{n!}P^n = e^{-tH_N}, \tag{2.2}$$

giving us a continuous time N-particle exclusion process.

Example 2: (convex linear combinations of generators): Suppose H and H' are spin-s isotropic Heisenberg Hamiltonians with nearest and next nearest neighbor ferromagnetic interactions, respectively. Then $-[\alpha H + (1-\alpha)H']$, $0 \leqslant \alpha \leqslant 1$, has positive off-diagonal matrix entries and annihilates $[\Sigma_m\, J_+(m)]^N\Omega$, and so is equivalent to a Markov semigroup generator.

III. PATH SPACE REPRESENTATION OF SPIN-$\frac{1}{2}$ SEMIGROUP AND PERTURBATIONS

Let $X_N(t) = [x_1(t), x_2(t),\ldots,x_N(t)]$ be the configurations of the N particles in Λ at time t (see example 1 of the previous section). As a function of time, we assume $X_N(t)$ is right continuous and piecewise constant. Let $\delta V(X_N)$be a real function of the particle configuration, and $\delta p(X_N, X'_N)$ a complex function of (configurations) X (configurations), which is zero unless X_N can be obtained from X'_N by moving a single coordinate of X'_N to a nearest neighbor site. Define the set of paths

$$ds_N(\,Y_0,Y_1,\ldots,Y_N;\,t_0,t_1,\ldots,t_n,t\,) = \{X_N(\cdot)|X_N \text{ jumps within } t_1 \pm (dt_1/2),\ldots,t_n \pm (dt_n/2) \text{ and between jumps assumes the values } Y_0,\ldots,Y_n\}. \tag{3.1}$$

We define a path space measure μ_N such that the measure of ds_N is

$$\mu_N(ds_N) = \exp\Big\{-\sum_{j=1}^{n+1}[1+\delta V(\,Y_{j-1}\,)](t_j - t_{j-1})\Big\} \times \prod_{j=1}^{n} p(\,Y_j,\,Y_{j-1}\,)dt_1,\ldots,dt_n, \tag{3.2}$$

where $t_{n+1} \equiv t$ and

$$p(\,Y',\,Y) = p_0(\,Y',\,Y) + \delta p(\,Y',\,Y), \tag{3.3}$$

with

$$p_0(\,Y',\,Y) = \frac{1}{2dN}, \quad \text{if } Y' \text{ is obtained from } Y \text{ by moving a single coordinate of } Y \text{ to an unoccupied nearest neighbor site,}$$
$$= \frac{j}{2dN}, \text{ if } Y' = Y \text{ and } j \text{ is number of forbidden jumps in configuration } Y. \tag{3.4}$$

Next define P_t to be the integral operator with kernel

$$P_t(Y,Y_0) = \mu_N(\{X_B(\cdot)|X_N(0) = Y_0, X_N(t) = Y\}). \tag{3.5}$$

Proposition 3.1: *The measure μ_N is bounded on the space of paths $\{X_N(s)|X_N(0) = Y_0\ 0 \leqslant s \leqslant t\}$. The operators $\{P_t\}$ form a semigroup with generator G_N given by*

$$G_N f(\,Y) = \sum_{Y'} p(\,Y',\,Y)f(\,Y') - [1+\delta V(\,Y)]\,f(\,Y). \tag{3.6}$$

Proof: Let a,b be constants such that $\inf[1+\delta V(\,Y)] \geqslant -a$, $\sup|\,p(\,Y,\,Y')| \leqslant b$. Then $|\,\mu_N[ds_N(\,Y_0,\ldots,Y_n;0,t_1,\ldots,t_n,t\,)]| \leqslant e^{at}b^n dt_1\ldots dt_n$, while an estimate on the number of paths with n jumps is $(2dN)^n$. Integrating over the jumping times and summing over n we obtain an estimate on the path measure of $\exp(at + 2dNbt)$.

That $\{P_t\}$ is a semigroup follows readily from that fact that

$$\int_{t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n \leqslant t} \mu_N\,[ds_N(Y_0,Y_1,\cdots,Y_n;t_0,t_1,\ldots,t_n,t\,)]$$
$$= \sum_{i=0}^{n}\int_{t_0 \leqslant t_1 \leqslant \cdots \leqslant t_i \leqslant s \leqslant t_{i+1} \leqslant \cdots \leqslant t_n \leqslant t} \mu_N\,[ds_N(\,Y_0,\,Y_1,\ldots, \times Y_n;\,t_0,\,t_1,\ldots,t_n,\,t\,)]$$
$$= \sum_{i=0}^{n}\int_{t_0 \leqslant t_1 \leqslant \cdots \leqslant t_i \leqslant s} \mu_N\,[ds_N(Y_0,Y_1,\cdots,Y_i;t_0,t_1,\ldots,t_i,s)] \times \int_{s \leqslant t_{i+1} \leqslant \cdots \leqslant t_n \leqslant t} \mu_N\,[(ds_N(Y_i,\cdots Y_n;s,t_{i+1},\cdots,t_n,t\,)].$$

Its generator can be determined by differentiating P_t and setting $t = 0$. Since paths with n jumps have measure $O(t^n)$, Eq. (3.6) follows easily from Eqs. (3.2) and (3.5).

Let μ_N^0 and G_N^0 be the measure and semigroup generator of proposition (3.1) and corresponding to $\delta V = \delta p = 0$. Comparing Eqs. (3.4) and (3.6) with example 1 of the previous section, one see that

$$G_N^0 = -\frac{1}{dN} H_N. \tag{3.7}$$

Finally, we develop a perturbation formula for μ_N in terms of μ_N^0, assuming $\delta p(Y', Y) = 0$ unless $Y' \neq Y$ and Y' is obtained from Y by moving a single coordinate of Y to a neighboring site. We write

$$\exp\Big[-\sum_{j=0}^{n} \delta V(Y_j)(t_{j+1} - t_j)\Big] = \exp\Big\{-\int_{t_0}^{t} \delta V[X_N(s)]\, ds\Big\},$$

and since $\delta p(Y, Y) = 0$,

$$\prod_{j=1}^{n} p(Y_j, Y_{j-1})$$

$$= \prod_{j=1}^{n} p_0(Y_j, Y_{j-1}) \prod_{j=1}^{n} [1 + 2dN\delta p(Y_j, Y_{j-1})]$$

$$= \prod_{j=1}^{n} p_0(Y_j, Y_{j-1}) \exp\Big[2dN \sum_{j=1}^{n} \delta p(Y_j, Y_{j-1}) - \frac{(2dN)^2}{2} \sum_{j=1}^{n} \delta p^2(Y_j, Y_{j-1}) + O(\delta p^3)\Big].$$

Combining these two expression, we obtain the following:

Proposition 3.2: *Assume $\delta p(Y', Y) = 0$ unless Y' is obtained from Y by moving a single coordinate of Y to a neighboring site. Then μ_N is given by the perturbation formula*

$$\mu_N[ds_N(Y_0, Y_1, \cdots, Y_n; t_0, t_1, \ldots, t_n, t)]$$

$$= \exp\Big\{-\int_{t_0}^{t} ds\delta V[X_N(s)] + 2dN \sum_{j=1}^{n} \delta p(Y_j, Y_{j-1}) - \tfrac{1}{2}(2dN)^2 \sum_{j=1}^{n} \delta p^2(Y_j, Y_{j-1}) + O(\delta p)^3\Big\},$$

$$\times \mu_N^0[ds_N(Y_0, Y_1, \cdots, Y_n; t_0, t_1, \ldots, t_n, t)]. \tag{3.8}$$

The generators G_N and G_N^0 are related by

$$G_N f(Y) = G_N^0 f(Y) + \sum_{Y'} \delta p(Y', Y) f(Y') - \delta V(Y) f(Y). \tag{3.9}$$

Remark: The perturbations δV and δp can be represented in terms of spin operators. We do not do so for δV here, but we note that the operator $J_+(m)J_-(m')$ corresponds to

$\delta p(Y, Y') = 1$, if y is obtained from Y'
by moving a particle at m' to m,
$= 0$, otherwise. (3.10)

IV. THERMODYNAMIC CORRELATION INEQUALITIES FOR SPIN-½ HEISENBERG MODEL

Let H be the isotropic spin-½ Heisenberg Hamiltonian.

Lemma 4.1[3,4]:

$$\sum_{i}^{3} \mathrm{tr} J_i(m) e^{-(\beta - t)H} J_i(m') e^{-tH} > 0, \quad m, m' \in \Lambda, \ 0 < t \leqslant \beta.$$

Proof: Consider first

$$\mathrm{tr}\, J_-(m) e^{-(\beta - t)H} J_+(m') e^{-tH}$$

$$= \sum_{N} \sum_{m_1, \ldots, m_N \in \Lambda} \langle m_1, m_2, \ldots, m_N | J_-(m) e^{-(\beta - t)H} \times J_+(m') e^{-tH} | m_1, \ldots, m_N \rangle, \tag{4.1}$$

with $|m_1, m_2, \ldots, m_N\rangle = J_+(m_1) \cdots J_+(M_N)\Omega$. Each term in the sum has a path space interpretation; start the particles at $m_1, \ldots m_N$ and propagate them for a time dNt [recall $G_N^0 = -(1/dN) H_N$)]. Inject another particle at m', if possible, symmetrize, and propagate the $N + 1$ particles for a time $d(N + 1)(\beta - t)$ and then "kill" them all at $m_1, \ldots, m_n, m$. Each of the terms is nonnegative and some are strictly positive. Thus, Eq. (4.1) is strictly positive.

Now let $U = \Pi_{m \in \Lambda} e^{i\pi J_3(m)}$. Then U is unitary, commutes with H, and satisfies $UJ_{\pm}(m)U^{-1} = J_{\mp}(m)$. By the invariance of the trace under unitary automorphisms, we obtain $\mathrm{tr}\, J_+(m) e^{-(\beta - t)H} J_-(m') e^{-tH} > 0$, which added to Eq. (4.1) gives

$$\sum_{i=1}^{2} \mathrm{tr} J_i(m) e^{-(\beta - t)H} J_i(m') e^{-tH} > 0.$$

[Terms involving $J_+(m)\, J_+(m')$ or $J_-(m)\, J_-(m')$ are zero since particle number is not conserved.] Again using rotation invariance, we obtain the lemma. ■

Remarks: Lemma 4.1 is a slight generalization of Refs. 3 and 4 since they consider the J_i's at equal times only. It should be clear that a similar argument will work for an isotropic Heisenberg model of arbitrary spin. The argument can also be used to show that products of higher degree in the $J_{\pm}$'s have positive thermal expectation.

Let $(A, B)_\beta$ denote the thermal duHamel expectation

$$(A, B)_\beta = \frac{1}{\beta Z} \int_0^\beta \mathrm{tr}(A^* e^{-(\beta - t)H} B e^{-tH})\, dt,$$
$$Z = \mathrm{tr}\, e^{-\beta H}. \tag{4.2}$$

Lemma 4.2: *Let $\mathbf{A} = \Sigma' \mathbf{h}(m, m') \cdot \mathbf{J}(m) \times \mathbf{J}(m')$, where $\mathbf{h}$ is real, Σ' is a sum over nearest neighbors, and $\times$ denotes the vector product. Then*

$$\sum_i (A_i, A_i)_\beta \leqslant \frac{3d|\Lambda|}{\beta} \sup_{m, m' \in \Lambda} h^2(m, m'). \tag{4.3}$$

Remarks: This inequality is closely related to the diamagnetic and paramagnetic inequalities of Schrödinger operators with magnetic field. The global rotation symmetry of H can be localized, giving rise to a "gauge" transformation $H \to H' = U(g) H U^{-1}(g)$, with $U(g) = \exp i\Sigma_{m \in \Lambda} \Sigma_{k=1}^{3} g_k(m) J_k(m)$. In particular, if $U = \exp i\epsilon \Sigma_{m \in \Lambda} g(m) J_3(m)$, then $H' = H + (\epsilon/2)\Sigma'[g(m) - g(m')][\mathbf{J}(m) \times \mathbf{J}(m')]_3 + O(\epsilon^2)$ with gauge field $h(m, m') = \epsilon[g(m) - g(m')]$. For this field, inequality (4.3) is essentially that obtained from the Bogoliubov inequality (cf. Ref. 8, Eq. (13)] $(C, H], [C, H])_\beta \leqslant (1/\beta)\langle [[C, H], C]\rangle_\beta$, with $C = \Sigma' h(m) J_3(m)$. [The right-hand side is $O(|\Lambda|/\beta) \sup h^2$)]. Inequality (4.3) is a generalization of this inequality to gauge fields which cannot be gauged away, i.e., which are not curl-free.

The quantity $\mathbf{A}$ is a measure of spin alignment and classically would be zero if all the spins pointed in the same direction. There is a *lower* bound associated with the individ-

ual terms of **A**, which for spin $\frac{1}{2}$ is

$$[\mathbf{J}(m)\times\mathbf{J}(m')]^2 = J^2(m)J^2(m') - [\mathbf{J}(m)\cdot\mathbf{J}(m')]^2 - \mathbf{J}(m)\cdot\mathbf{J}(m') \geqslant 1/4.$$

Proof of Lemma (4.2): Let $B = (1/dN)\,A_3 = (1/2dNi)\,\Sigma'\,h(m,m')[J_+(m)J_-(m') - J_-(m)J_+(m')]$. By Eq. (3.10), B corresponds to

$$\delta p(Y,Y') = \pm\frac{1}{2dNi}h(m,m'), \quad \text{if } Y \text{ obtained from } Y' \text{ by moving a particle at } m' \text{ to } m \text{ (or } m \text{ to } m'),$$
$$= 0, \quad \text{otherwise,}$$

and by Eq. (3.8),

$$|\mu_N\,[ds_N(Y_0,\cdots,Y_n;0,t_1,\ldots,t_n,t)]\,| \leqslant \exp\left[-\frac{1}{2}(2dN)^2\sum_{j=1}^{n}\delta p^2(Y_j,Y_{j-1}) + O(\delta p)^3\right]_{\mu_N^0(ds_N)}$$
$$\leqslant \exp\{1/2\,nh^2 + O(h^3)\}\,\mu_N^0[ds_N(Y_0,\ldots,Y_n;\ O,t_1,\ldots,t_n,t)], \tag{4.4}$$

with $h = \sup|h(m,m')|$. Note that n is the number of jumps in the path and is a random variable.

We have that if $H'_N = H_N + A_3$, $\mathrm{tr}_N e^{-\beta H'_N} = \int \mu_N(ds_N)$ where integration is over all N paths $X_N(s)$, $0\leqslant s\leqslant dN\beta$ starting at any configuration $(m_1,\ldots,m_N)$ and returning to that (permuted) configuration. Inequality (4.4) and Eqs. (3.2) and (3.7) give

$$|\mathrm{tr}_N e^{-\beta H'_N}| \leqslant \int e^{1/2nh^2 + O(h^3)}\mu_N^0(ds_N)$$
$$= e^{-dN\beta(1-e^{1/2h^2})}\sum_{\{Y_i\}} e^{-dN\beta e^{1/2h^2}} \times \frac{(dN\beta e^{1/2h^2})^n}{n!}\prod_j^n p_0(Y_j,Y_{j-1})$$
$$= e^{-dN\beta(1-e^{1/2h^2})}\mathrm{tr}_N e^{-\beta e^{h^2}H_N}$$
$$\leqslant e^{1/2dN\beta h^2}\,\mathrm{tr}_N\,e^{-\beta H_N}, \tag{4.5}$$

by Eq. (3.2), and where we have retained $O(h^2)$ terms only. If we multiply $h(\cdot,\cdot)$ by a scale factor ϵ and expand to order ϵ^2, inequality (4.5) gives

$$\mathrm{tr}_N e^{-\beta(H_N+\epsilon A_3)} = (\mathrm{tr}_N e^{-\beta H_N})[1 + 1/2\,\beta^2\epsilon^2(A_3,A_3)]_N + \cdots$$
$$\leqslant (\mathrm{tr}_N e^{-(\beta H_N)})(1 + 1/2dN\beta\epsilon^2h^2 + \cdots) \tag{4.6}$$

since $\mathrm{tr}_N(e^{-\beta H_N}A_3) = 0$. Here,

$$(A,B)_{N\beta} \equiv \mathrm{tr}_N(e^{-\beta H_N})^{-1}\frac{1}{\beta}\int_0^\beta \mathrm{tr}_N\,e^{-(\beta-t)H_N}A\,\dagger e^{-tH_N}B.$$

Inequality (4.6) gives

$$(A_3,A_3)_{N\beta} \leqslant \frac{dN}{\beta}h^2. \tag{4.7}$$

Inequality (4.7) holds N sector by N sector. It implies that $(A_3,A_3)_\beta \leqslant (d\,|\Lambda\,|/\beta)h^2$. Rotation invariance gives the lemma. ■

ACKNOWLEDGMENT

The basic ideas presented were arrived at in conversations with D.G. Babbitt to whom I express my thanks.

[1] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I.
[2] F. Spitzer, Adv. Math. **5**, 246 (1970).
[3] R. T. Powers, Lett. Math. Phys. **1**, 125 (1976).
[4] C. A. Hurst and S. Sherman, Phys. Rev. Lett. **22**, 1357 (1969).
[5] R.T. Powers, J. Math. Phys. **17**, 1910 (1976); and Commun. Math. Phys. **51**, 151 (1976).
[6] J. Lamperti, *Stochastic Processes* (Springer, New York, 1977).
[7] I. M. Gelfand, R. A. Minlos, and Z. Y. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications*) (Pergamon, New York, 1963).
[8] G. Roepstorff, Commun. Math. Phys. **46**, 253 (1976).

Partial wave expansion of the RKKY indirect exchange interaction in metals

Ulf Larsen
H.C. Ørsted Institute, Physics Laboratory I, University of Copenhagen, Universitetsparken 5, DK 2100 Copenhagen Ø, Denmark

(Received 25 May 1979; accepted for publication 21 August 1979)

Using the full partial wave series for the exchange coupling constant $J(\mathbf{k},\mathbf{k}') = \Sigma_{L=0}^{\infty}(2L+1) \times P_L(\hat{k}\cdot\hat{k}')J_L(k,k')$, in standard notation, for an *s-d* type exhange interaction between a local spin (electronic or nuclear) and free conduction electrons in metals, the corresponding full partial wave series for the long range, indirect, exchange interaction (Ruderman–Kittel–Kasuya–Yosida, RKKY) between two such local spins is derived.

In particular, the asymptotic long range form of the exchange coupling constant is $\mathcal{F}(R) \sim -(9\pi n^2/E_F)J_{\mathrm{RKKY}}^2 \cos(2k_F R)/(2k_F R)^3$, where $J_{\mathrm{RKKY}} = \Sigma_{L=0}^{\infty}(2L+1)(-1)^L \times J_L(k_F,k_F)$. For constant $J_L(k,k') = J_L$ the exchange coupling $\mathcal{F}(R)$ is exactly calculated for all distances R in terms of standard functions. To obtain this result nonstandard indefinite integrals of spherical Bessel functions have been calculated and the results are given in terms of recursion relations. Also, a closed form is obtained for the Cauchy principal value of the singular integral $\int_0^{\infty} dy\, y^2 j_l(y)/(y^2 - x^2)$. Results are given for the electron spin polarization due to a local spin.

I. INTRODUCTION

The long-range Ruderman–Kittel–Kasuya–Yosida (RKKY) indirect exchange interaction[1] between two localized spins S in a metal is mediated by the conduction electrons, and arises from the *s-d* or *s-f* interaction of each spin with the electrons according to the Hamiltonian[2,3]

$$\mathcal{H}_{s\text{-}d} = -\frac{1}{N_a}\sum_{\substack{\mathbf{k}\mathbf{k}'\\ \sigma\sigma'}} J(\mathbf{k},\mathbf{k}')c_{\mathbf{k}\sigma}^{+}\boldsymbol{\sigma}_{\sigma\sigma'}c_{\mathbf{k}'\sigma'}\cdot\mathbf{S}\,. \tag{1}$$

The operators $c_{\mathbf{k}\sigma}^{+}$ and $c_{\mathbf{k}'\sigma'}$ create and destruct free electrons with wave vectors $\mathbf{k},\mathbf{k}'$ in spin states σ,σ'. The sum $\Sigma_{\substack{\mathbf{k}\mathbf{k}'\\ \sigma\sigma'}}$ runs over all the states of the electrons. Direct exchange integrals give positive contributions to J, while mixing exchange, via the Schrieffer–Wolff transformation[4] of the Anderson Hamiltonian,[5] gives negative contributions to J. It is clear that some assumptions implicit in this model need not be good approximations in particular systems, like the assumption of free electrons or the existence of a local spin. Such complications are discussed by other authors[6,7] and will not be further considered here.

In the two cases of direct and mixing exchange referred to above, Watson, Freeman, and Koide[8] give a partial wave expansion[3] of J

$$J(\mathbf{k},\mathbf{k}') = \sum_{L=0}^{\infty}(2L+1)P_L(\hat{k}\cdot\hat{k}')J_L(k,k')\,, \tag{2}$$

where P_L is a Legendre polynomial. The implications of this, when *one* of the partial waves is assumed to dominate, were discussed by for example Coqblin and Schrieffer,[7] who gave the asymptotic form of the RKKY interaction for $L = 3$, and by Caroli.[9] Attempts to evaluate a *set* of these J_L values (taken as constants independent of k and k'), for example J_0, J_1, and J_2 in AgMn, were done on the basis of different experiments that depend on different combinations of the J_L by Davidov *et al.*[10] and by Walstedt and Walker.[11] One finds that the conduction electron spin polarization in the asymptotic limit of long distances from a local spin S, associated with a Mn impurity, for example, will be proportional to the combination

$$J_{\mathrm{RKKY}} = \sum_{L=0}^{\infty}(2L+1)(-1)^L J_L\,, \tag{3}$$

while for example the conduction electron g-shift, by virtue of $P_L(1) = 1$, in first order of perturbation theory in the J_L is proportional to just

$$\sum_{L=0}^{\infty}(2L+1)J_L\,. \tag{4}$$

Davidov *et al.*[10] also gave an involved expression for the resistivity in the second Born approximation (i.e., third-order perturbation theory in J_L) that thoroughly mixes the three J_L that are included. That calculation can be shown to behave better in an extension[12] of the nonperturbative theory of Larsen[13] to deal with all the partial waves present in Eq. (2). This produces expressions that, among other things, essentially confirm the original result due to Blandin[14] that the Kondo temperature is given by[3]

$$T_K = T_F \exp\left(-\frac{1}{n(E_F)|J_K|}\right), \tag{5}$$

where J_K corresponds to the *resonantly* scattering partial wave $L = K$ with $J_K < 0$, that gives rise to the Kondo effect.

Recently Lindgård, Harmon, and Freeman[15] emphasized that in pure Gd the spin wave spectrum can only be properly calculated if one uses all the dependence on $\mathbf{q} = \mathbf{k} - \mathbf{k}'$. Only direct exchange was considered in this work. For the mixing exchange, originating in the Schrieffer–Wolff transformation of the Anderson Hamiltonian, the dependence of J on $\mathbf{k}$ and $\mathbf{k}'$ comes from the $V_{\mathbf{k}}$ mixing matrix elements,[5] i.e., $J(\mathbf{k},\mathbf{k}') \propto V_{\mathbf{k}}V_{\mathbf{k}'}$.

This part is negative and larger than the corresponding direct exchange integral whenever there is a Kondo effect. The complications of the Kondo effect, however, make this question of the $\mathbf{k}$ and $\mathbf{k}'$ dependence enormously difficult to

 0022-2488/80/071925-13$1.00

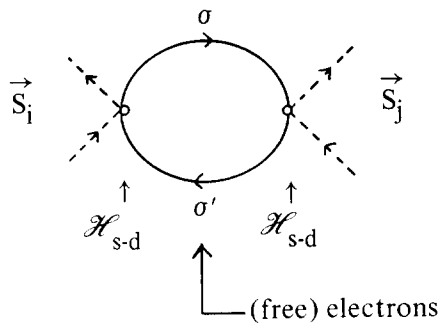


FIG. 1.

deal with because all measurable effects depend in complicated nonperturbative ways on J. In these cases the J_L are therefore always assumed constant, and since the present results are primarily intended for use in the context of RKKY modified Kondo systems,[13,16] i.e., spin glasses,[17] the approximation of constant J_L will also be done at a certain stage in the present work.

However, it has not been worked out how the partial waves combine in the RKKY interaction[3]

$$\mathscr{H}_{\mathrm{RKKY}} = -\tfrac{1}{2}\sum_{ij}\mathscr{F}(R_{ij})\mathbf{S}_i\cdot\mathbf{S}_j\,. \tag{6}$$

For only *one* partial wave it is known[9] that $\mathscr{F}\propto(2L+1)^2 J_L^2$.

The interest in this question from an experimental point of view has recently been accentuated by investigations of spin glass and Kondo dilute alloys under high pressures by Schilling and others.[16,18] In a recent comprehensive review[19] of the existing body of experimental work, Schilling demonstrates a universal increase with increasing pressure of the resonant $|J_K|$ in a variety of dilute Kondo alloys, like CuFe, based on the observation of a universal increase in the Kondo temperature, given by Eq. (5). In more concentrated alloys one may obtain information about the pressure dependence of the RKKY interaction by studying the resistance maximum[13,16-19] and the spin glass freezing temperature.[20,21] In order to analyze these results one must know how $\mathscr{F}(R_{ij})$ depends on the components J_L, not *only* on the resonant J_K. Schilling has conjectured that[19]

$$\mathscr{F}(R_{ij})\propto J_{\mathrm{RKKY}}{}^2\,, \tag{7}$$

and on the basis of this one finds good agreement between the pressure dependences determined in dilute alloys (J_K) and in more concentrated spin glass alloys (all J_L). This result supports the possibility that due to the $(-1)^L$ the positive J_0 and J_1 of nonvanishing magnitude may partially cancel out when forming the relevant J_{RKKY}, which may then be dominated by the resonant $J_2<0$ (for 3-*d* impurities) enhanced by the $(2L+1)$. The values determined by Davidov *et al.* for AgMn, $J_0=0.13$ eV, $J_1=0.09$ eV, and $J_2=-0.13$ eV are consistent with this idea.

In the present paper I *verify Eq.* (7) *as a result valid in the asymptotic limit of* $k_F R_{ij}\to\infty$. This is sufficient to deal with reasonable confidence with the dilute Kondo and spin glass alloys ($c<1\%$) of the 3-*d* type, whereas in 4-*f* type alloys (like LaCe of concentrations up to 40% Ce) and compounds (like $CeAl_2$ and $CeAl_3$) and certainly in elements (like Ce) there may be a need to consider the more complicated behavior at shorter range. In the following *the exact form of* $\mathscr{F}(R)$ *including all partial waves is given.* Next, all J_L are taken as constants (an approximation that must be justified in each case where applied), and further *exact expressions in terms of standard functions are given for the remaining integral*

$$\mathbb{J}_l(k_F R) = -\frac{1}{\pi k_F^4}\int_0^{k_F} dk k^2 j_l(kR)\times\Xint-_0^\infty dk'\,\frac{k'^2}{k'^2-k^2}j_l(k'R)\,, \tag{8}$$

as well as related integrals, as for example the inner integral over k', that are not standard integrals. This effort, which is mostly relegated to the Appendices, is motivated by mathematical completeness, but also by the fact that the integrals may occur as well in other contexts. $\mathbb{J}_l$ is in fact just the l'th partial wave of the range-dependent, static, zero-temperature susceptibility, i.e., the Fourier transform of the Lindhard function,[22] that describes the response of the degenerate free Fermion gas when coupling by all partial waves to a localized (non-point-like) disturbance. An example of this response is the conduction electron spin polarization, due to a spin localized in atomic impurity orbitals as in the systems discussed above, which displays the familiar Friedel oscillations. The full range dependence of this is given in Section 3 as a corollary to the present calculation of the indirect exchange.

The inclusion of more than one partial wave in $\mathscr{F}(R)$ has implications for properties derived from it, like for example the spin glass freezing temperature T_0 and the noise temperature Δ_c. In a recent calculation[20] of these quantities only the resonant partial wave was included and the asymptotic form of $\mathscr{F}(R)$ for $R\to\infty$ was used. It is shown here that *in general the RKKY interaction has the asymptotic form*[3]

$$\mathscr{F}(R)\sim-\frac{9\pi n^2}{E_F}J_{\mathrm{RKKY}}^2\frac{\cos 2k_F R}{(2k_F R)^3}\,. \tag{9}$$

If this is used the modifications of the theory in Ref. 20 consist in replacing $(-1)^K(2K+1)J_K$ by J_{RKKY} and one obtains

$$k_B T_0\propto k_B\Delta_c\propto(J_{\mathrm{RKKY}}^2/E_F)f(c,\lambda)\,, \tag{10}$$

where $f(c,\lambda)$ is a function of the concentration c and the electron mean free path λ, which was derived and discussed in Ref. 20.

2. GENERAL RKKY INTERACTION

The diagrammatic finite temperature Green function method used in the derivation in this section is the standard method described in the books. It differs in form but not in content from the conventional procedure. The interaction $\mathscr{H}_{\mathrm{RKKY}}$ is the Fourier transform of the static part of the diagram shown in Fig. 1.

The spin part of this becomes, using that $\mathrm{Tr}(\sigma^\alpha\sigma^\beta)$

$$= 2\delta_{\alpha\beta},$$

$$\sum_{\sigma\sigma'} (\mathbf{S}_i\cdot\boldsymbol{\sigma}_{\sigma\sigma'})(\boldsymbol{\sigma}_{\sigma'\sigma}\cdot\mathbf{S}_j) = \sum_{\alpha,\beta=x,y,z} S_i^{\alpha}\mathrm{Tr}(\sigma^{\alpha}\sigma^{\beta})S_j^{\beta} = 2\mathbf{S}_i\cdot\mathbf{S}_j.$$

Then using

$$\sum_{\mathbf{q}\mathbf{k}'} e^{i\mathbf{q}\cdot\mathbf{R}_{ij}} \circlearrowleft_{\mathbf{k}'}^{\mathbf{k}'+\mathbf{q}} = \sum_{\mathbf{k}\mathbf{k}'} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{ij}} \circlearrowleft_{\mathbf{k}'}^{\mathbf{k}},$$

and the sum over the Matsubara energies $i\omega_n$ (T = temperature)

$$\circlearrowleft_{n}^{n+p} \quad i\omega_n = i\pi k_B T(2n-1), n = 0, \pm 1, \pm 2, \cdots,$$
$$i\xi_p = i\pi k_B T 2p, \qquad p = 0, \pm 1, \pm 2, \cdots,$$

it follows that $\mathscr{H}_{\mathrm{RKKY}}$ can be written in the form (6) with

$$\mathscr{F}(R_{ij}) = -\frac{2}{N_a^2}\sum_{L,L'=0}^{\infty}(2L+1)(2L'+1)\sum_{\mathbf{k}\mathbf{k}'}P_L(\hat{k}\cdot\hat{k}')P_{L'}(\hat{k}'\cdot\hat{k}) \times J_L(k,k')J_{L'}(k',k)\exp[i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{ij}] \times k_B T\sum_n G(\mathbf{k},i\omega_n + i\xi_p)G(\mathbf{k}',i\omega_n). \tag{11}$$

The expression must be analytically continued from the points $i\xi_p$ to just above the real axis $i\xi_p \to 0 + i0^+$. The electron Green functions are

$$G(\mathbf{k}',i\omega_n) = -1/(i\omega_n - \epsilon_{\mathbf{k}'}), \tag{12}$$

where $\epsilon_{\mathbf{k}} = \epsilon_k = \hbar^2k^2/2m$ is the energy of a free electron. Also, as usual,

$$\sum_{\mathbf{k}}\cdots = V\int_{-\infty}^{\infty}\frac{d^3k}{(2\pi)^3}\cdots = \frac{V}{(2\pi)^3}\int_0^{\infty}dkk^2\int_{4\pi}d\Omega_k\cdots,$$

where V is the volume, Ω_k is the direction of $\mathbf{k}$, and $d\Omega_k$ the element of solid angle in that direction. Then by a well known theorem[22]

$$T\sum_n \frac{1}{i\omega_n + i\xi_p - \epsilon_k}\frac{1}{i\omega_n - \epsilon_{k'}} = \frac{f(\epsilon_{k'}) - f(\epsilon_k)}{\epsilon_{k'} - \epsilon_k + i\xi_p},$$

where $f(\epsilon) = (1 + e^{\epsilon/k_BT})^{-1}$ is the Fermi function at temperature T. Then letting $i\xi_p \to 0 + i0^+$ one gets the standard decomposition

$$\int dk\frac{f(\epsilon_k) - f(\epsilon_{k'})}{\epsilon_k - \epsilon_{k'} - i0^+} \to \int dk\frac{f(\epsilon_k) - f(\epsilon_{k'})}{\epsilon_k - \epsilon_{k'}} + i\pi\int dk\delta(\epsilon_k - \epsilon_{k'}), \times(f(\epsilon_k) - f(\epsilon_{k'})),$$

where the imaginary part vanishes by virtue of the δ-function. This term now only depends on the magnitudes of $\mathbf{k}$ and $\mathbf{k}'$. One gets

$$\mathscr{F}(R) = -\frac{2}{N_a^2}\sum_{L,L'=0}^{\infty}(2L+1)(2L'+1)\int_{4\pi}d\Omega_k \times\int_{4\pi}d\Omega_{k'}P_L(\hat{k}\cdot\hat{k}')P_L(\hat{k}'\cdot\hat{k})\frac{2m}{\hbar^2}\left(\frac{V}{(2\pi)^3}\right)^2 \fint_0^{\infty}dkk^2\fint_0^{\infty}dk'k'^2J_L(k,k')J_{L'}(k',k) \times\exp[i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}]\frac{f(\epsilon_k) - f(\epsilon_{k'})}{k^2 - k'^2}. \tag{13}$$

Now let $T \to 0$ and use $J_L(k,k') = J_L(k',k)$, then

$$\fint\fint = -\int_0^{k_F}dkk^2e^{i\mathbf{k}\cdot\mathbf{R}}\fint_0^{\infty}dk'\frac{k'^2}{k'^2 - k^2} \times J_L(k,k')J_{L'}(k',k)e^{-i\mathbf{k}'\cdot\mathbf{R}} - \int_0^{k_F}dk'k'^2e^{-i\mathbf{k}'\cdot\mathbf{R}}\fint_0^{\infty}dk\frac{k^2}{k^2 - k'^2} \times J_L(k,k')J_{L'}(k',k)e^{i\mathbf{k}\cdot\mathbf{R}} = 2\mathrm{Re}\int_0^{k_F}dkk^2e^{i\mathbf{k}\cdot\mathbf{R}}\fint_0^{\infty}dk'\frac{k'^2}{k'^2 - k^2}J_L(k,k') \times J_{L'}(k',k)e^{-i\mathbf{k}'\cdot\mathbf{R}}.$$

Next one expands the exponentials

$$e^{i\mathbf{k}\cdot\mathbf{R}} = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{m=+l}i^lj_l(kR)Y_l^{m*}(\Omega_k)Y_l^m(\Omega_R),$$

$$e^{-i\mathbf{k}'\cdot\mathbf{R}} = 4\pi\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{m'=+l'}(-i)^{l'}j_{l'}(k'R)Y_{l'}^{m'}(\Omega_{k'})Y_{l'}^{m'*}(\Omega_R),$$

whereby (13) becomes

$$\mathscr{F}(R) = \frac{9\pi n^2}{E_F}\frac{1}{\pi k_F^4}\sum_{L,L'=0}^{\infty}(2L+1)(2L'+1) \times\sum_{l,l'=0}^{\infty}\sum_{m,m'=-l_<-l'}^{+l_<+l'}\mathrm{Re}\int_0^{k_F}dkk^2i^lj_l(kR) \times\fint_0^{\infty}dk'\frac{k'^2}{k'^2 - k^2}(-i)^{l'}j_{l'}(k'R)J_L(k,k')J_{L'}(k',k) \times\int_{4\pi}d\Omega_k\int_{4\pi}d\Omega_{k'}P_L(\hat{k}\cdot\hat{k}')P_{L'}(\hat{k}'\cdot\hat{k})Y_l^{m*}(\Omega_k) \times Y_l^m(\Omega_R)Y_{l'}^{m'}(\Omega_{k'})Y_{l'}^{m'*}(\Omega_R). \tag{14}$$

Then by the addition theorem

$$P_L(\hat{k}\cdot\hat{k}') = \frac{4\pi}{2L+1}\sum_{M=-L}^{M=+L}Y_L^{M*}(\Omega_k)Y_L^M(\Omega_{k'}),$$

and similarly for $P_{L'}$, to calculate $\mathscr{F}(R)$ there are therefore two angular integrals to do

$$\int_{4\pi}d\Omega_kY_L^{M*}(\Omega_k)Y_l^{m*}(\Omega_k)Y_{L'}^{M'}(\Omega_k) = \left[\frac{(2L+1)(2l+1)}{4\pi(2L'+1)}\right]^{1/2}\langle Ll00|L'0\rangle\langle LlMm|L'M'\rangle$$

and

$$\int_{4\pi}d\Omega_{k'}Y_{L'}^{M'*}(\Omega_{k'})Y_{l'}^{m'}(\Omega_{k'})Y_L^M(\Omega_{k'}) = \left[\frac{(2L+1)(2l'+1)}{4\pi(2L'+1)}\right]^{1/2} \times\langle Ll'00|L'0\rangle\langle Ll'Mm'|L'M'\rangle.$$

Both integrals vanish unless L, L' and l or l' satisfy the triangular rules

$$\overset{L'}{\underset{L}{\triangleleft}}\ l \text{ or } l' \qquad \begin{matrix} M+m = M' \\ M+m' = M' \end{matrix}. \tag{15}$$

It is possible to further simplify the result by choosing for Ω_R the direction of the coordinate z axis. Then

$$Y_l^m(\Omega_R = \Omega_{\hat{z}}) = \sqrt{\frac{2l+1}{4\pi}}\,\delta_{m0},$$

$$Y_{l'}^{m'*}(\Omega_R = \Omega_{\hat{z}}) = \sqrt{\frac{2l'+1}{4\pi}}\,\delta_{m'0},$$

which also leads to the condition

$$M = M'. \tag{16}$$

One gets for the angular integrals in Eq. (14)

$$\int_{4\pi}\int_{4\pi} = \frac{(2l+1)(2l'+1)}{(2L'+1)^2}\,\delta_{m0}\delta_{m'0} \times \sum_{MM'} \langle Ll00|L'0\rangle\langle LlMm|L'M'\rangle\langle Ll'00|L'0\rangle \times \langle Ll'Mm'|L'M'\rangle .$$

Then use the expression of the vector-addition coefficients in terms of the $3j$ symbols

$$\langle j_1 j_2 m_1 m_2 | j_3 m_3\rangle = (-1)^{j_1 - j_2 + m_3}\sqrt{2j_3+1}\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & -m_3 \end{pmatrix}$$

to write

$$\int_{4\pi}\int_{4\pi} = (2l+1)(2l'+1)\,\delta_{m0}\delta_{m'0}\begin{pmatrix} L & l & L' \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} L & l' & L' \\ 0 & 0 & 0 \end{pmatrix} \times \sum_{MM'}\begin{pmatrix} L & l & L' \\ M & m & -M' \end{pmatrix}\begin{pmatrix} L & l' & L' \\ M & m' & -M' \end{pmatrix}. \tag{17}$$

The value of the M,M' double sum can be obtained from

$$\sum_{m_1 m_2}\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & -m \end{pmatrix}\begin{pmatrix} j_1 & j_2 & j' \\ m_1 & m_2 & -m' \end{pmatrix} = \frac{\delta_{jj'}\delta_{mm'}}{2j+1}$$

by putting

$$j_1 = L,\ m_1 = M,$$
$$j_2 = L',\ m_2 = -M',$$
$$j = l,\ -m \to m,$$
$$j' = l',\ -m' \to m'.$$

Then one gets

$$\sum_{MM'}\begin{pmatrix} L & L' & l \\ M & -M' & m \end{pmatrix}\begin{pmatrix} L & L' & l' \\ M & -M' & m' \end{pmatrix} = \frac{\delta_{ll'}\delta_{mm'}}{2l+1},$$

and by using the symmetry relation for column permutation, this becomes

$$\sum_{MM'}\begin{pmatrix} L & l & L' \\ M & m & -M' \end{pmatrix}\begin{pmatrix} L & l' & L' \\ M & m' & -M' \end{pmatrix} = \frac{\delta_{ll'}\delta_{mm'}}{2l+1}. \tag{18}$$

One now combines the results (14), (17), (18) and obtains the final version

$$\mathcal{F}(R) = -\frac{9\pi n^2}{E_F}\sum_{L,l,L'=0}^{\infty}(2L+1)(2l+1)(2L'+1) \times \begin{pmatrix} L & l & L' \\ 0 & 0 & 0 \end{pmatrix}^2\left\{-\frac{1}{\pi k_F^4}\int_0^{k_F} dk k^2 j_l(kR) \int_0^{\infty} dk'\,\frac{k'^2}{k'^2 - k^2} j_l(k'R) J_L(k,k') J_{L'}(k',k)\right\}. \tag{19}$$

Here the explicit form of the $3j$ symbol is

$$\begin{pmatrix} L & l & L' \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{(L+l-L')!(L-l+L')!(-L+l+L')!}{(2p+1)!} \times\left[\frac{p!}{(p-L)!(p-l)!(p-L')!}\right]^2 \tag{20}$$

for $2p = L + l + L' =$ even, and $= 0$ for $2p =$ odd. A special case is $l = 0$ which implies $L = L'$ whereby

$$\begin{pmatrix} L & 0 & L \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{1}{2L+1}. \tag{21}$$

The restrictions on L,l, and L' are then

$$L + l + L' = \text{even}, \quad |L - L'| \leqslant l \leqslant L + L'. \tag{22}$$

It is clear that one gets no further than this without knowing or assuming something about $J_L(k,k')$. However, *in the asymptotic limit of* $k_F R \to \infty$ there is a possibility of simplifications. The spherical Bessel function $j_l(kR)$ oscillates with "period" $k \simeq 2\pi/R$. Looking first at the inner integral in Eq. (19), therefore there will be a tendency to average out the integrand whenever R becomes large enough that an oscillation covers a range of k where the rest of the integrand only varies slowly. The only reason that the inner integral does not vanish entirely in that limit is the singularity at $k' = k$. This will pick out $[J_L(k,k)]^2$. A similar argument shows the emphasis on the values near $k \simeq k_F$ in the k integral. It is therefore a reasonable approximation to take $[J_L(k_F,k_F)]^2$ *in the leading terms of the asymptotic expansion for* $k_F R \to \infty$. Not because it is a good approximation to $[J_L(k,k')]^2$, which is not assured, but *because the integrations single out this part*. Therefore, in the remainder of this section put $J_L(k,k') = J_L = \text{const}$, whereby

$$\mathcal{F}(R) = -\frac{9\pi n^2}{E_F}\sum_{L,l,L'=0}^{\infty}(2L+1)(2l+1)(2L'+1) \times\begin{pmatrix} L & l & L' \\ 0 & 0 & 0 \end{pmatrix}^2 J_L J_{L'} \mathbb{J}_l(k_F R), \tag{23}$$

where $\mathbb{J}_l$ was defined in Eq. (8). One may obtain directly from the definition the asymptotic form of $\mathbb{J}_l$ which is a well-known result. Here it will be obtained as the appropriate limit of the exact solution of the integral. $\mathbb{J}_l$ may be expressed in terms of the function

$$\mathbb{I}_l(x) = \int_0^{\infty} dy\,\frac{y^2}{y^2 - x^2} j_l(y), \tag{24}$$

as follows:

$$\mathbb{J}_l(k_F R) = -\frac{1}{\pi(k_F R)^4}\int_0^{k_F R} dx\, x^2 j_l(x)\mathbb{I}_l(x). \tag{25}$$

In Appendix A is proved a recursion relation for $\mathbb{J}_l$, which is then used in Appendix B to prove that for $l =$ even

$$\mathbb{I}_l(x) = -\frac{\pi}{2}\left\{xy_l(x) + \sum_{m=2,4,\ldots}^{m=l}\frac{(l+m-1)!!}{(l-m)!!}\frac{1}{x^m}\right\}, \tag{26}$$

and for $l =$ odd

$$\mathbb{I}_l(x) = -xj_l(x)\mathrm{Ci}(x) - xy_l(x)\mathrm{Si}(x) - \sum_{\nu=0}^{(l-3)/2} \frac{b_\nu^{(l)}}{x^{l-1-2\nu}}, \tag{27}$$

where Ci and Si are the cosine and sine integrals. In both cases $\mathbb{I}_l(0)$ is finite

$$\mathbb{I}_l(0) = \frac{\sqrt{\pi}}{2}\Gamma\left(\frac{l+1}{2}\right)\Big/\Gamma\left(\frac{l+2}{2}\right) = \frac{(l-1)!!}{l!!} \times \begin{cases} \pi/2 & \text{for } l = \text{even}, \\ 1 & \text{for } l = \text{odd}, \end{cases} \tag{28}$$

but $\mathrm{Ci}(x)$ introduces a term $\propto x^{l+1}\ \ln x$ for $x \to 0$, so the function is not analytic there despite the exact cancellation of all pole-terms by the substracted sums in Eqs. (26) and (27). The numbers $b_\nu^{(l)}$ are defined for $2\nu \leqslant l-3$ by Appendix B, Eq. (B15)

$$b_\nu^{(l)} = \sum_{k=0}^{\nu} \frac{(-1)^k(2l-2\nu+2k-1)!!}{(2k+1)(2k+1)!(2\nu-2k)!!}. \tag{29}$$

As a curious corollary to the proof of these results it becomes possible to prove the validity of the following rather exotic sum of products of binomial coefficients (cf. Appendix D):

$$\sum_{p=0}^{p=\mu} (-1)^p \binom{l}{p}\binom{2l+1-2p}{2\mu+1-2p} = 2^{2\mu}\frac{(2l+1-2\mu)}{(l-2\mu)}\binom{l}{2\mu+1}, \tag{30}$$

valid for integers l and μ that satisfy $0 \leqslant 2\mu \leqslant l-1$. It is peculiar by the appearance of "doubled" integers in the second binomial coefficient of the products. It was only possible to prove this independently because the result was suggested (cf. Appendix B) by the necessity to cancel all the poles between the second and third terms in Eq. (27), to satisfy Eq. (28).

The leading term for $x \to \infty$ comes in both cases from the term containing y_l as $\mathrm{Ci}(x) \to 0$ and $\mathrm{Si}(x) \to \pi/2$, so

$$\mathbb{I}_l(x) \sim (\pi/2)\cos[x - (l\pi/2)]. \tag{31}$$

This result is also obtained directly from Eq. (24) by using the asymptotic form $j_l(y) \sim \sin[y - (l\pi/2)]/y$ whereby it becomes a standard integral. Next we obtain the asymptotic form of $\mathbb{J}_l$ by considering the asymptotic series for the derivative ($z = k_F R$)

$$\begin{aligned}(d/dz)(-\pi z^4 \mathbb{J}_l(z)) &\sim z^2 j_l(z)(\pi/2)\cos[z - (l\pi/2)] \\ &\sim (\pi/4)z\sin(2z - l\pi) \\ &= (\pi/4)(-1)^l z \sin 2z.\end{aligned}$$

The leading term in the expansion of $\mathbb{J}_l$ is obtained by integration

$$\mathbb{J}_l(k_F R) \sim (-1)^l[\cos 2k_F R/(2k_F R)^3]. \tag{32}$$

Now $L + l + L' = $ even implies $(-1)^l = (-1)^L(-1)^{L'}$. The relation

$$\sum_{jm}(2j+1)\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & -m \end{pmatrix}\begin{pmatrix} j_1 & j_2 & j \\ m_1' & m_2' & -m \end{pmatrix} = \delta_{m_1' m_1}\delta_{m_2' m_2}$$

specialized to $m_1 = m_1' = m_2 = m_2' = 0$, which implies that $m = 0$ in the sum, then yields

$$\sum_{l=|L-L'|}^{l=L+L'}(2l+1)\begin{pmatrix} L & l & L' \\ 0 & 0 & 0 \end{pmatrix}^2 = 1. \tag{33}$$

From Eq. (23) one then obtains

$$\mathcal{F}(R) \sim -\frac{9\pi n^2}{E_F}\left[\sum_{L=0}^{\infty}(-1)^L(2L+1)J_L\right]^2\frac{\cos 2k_F R}{(2k_F R)^3}, \tag{34}$$

which is the result quoted in Eqs. (3) and (9) *for the leading term in the limit* $k_F R \to \infty$.

Now consider the full $\mathbb{J}_l(k_F R)$. Introduce two new functions,

$$\mathbb{F}_m^l(\rho) = \int_0^\rho dx\, x^{2-m} j_l(x), \tag{35}$$

which is calculated in Appendix E, and

$$\mathbb{A}_l(\rho) = \int_0^\rho dx\, x^3 j_l(x) y_l(x), \tag{36}$$

which is calculated in Appendix F. One may then write for $l = $ even

$$\mathbb{J}_l(k_F R) = \frac{1}{2(k_F R)^4}\left(\mathbb{A}_l(k_F R) + \sum_{m=2,4,\ldots}^{m=l}\frac{(l+m-1)!!}{(l-m)!!}\mathbb{F}_m^l(k_F R)\right). \tag{37}$$

To obtain the result for odd l introduce another three functions,

$$\mathbb{B}_l(\rho) = \int_0^\rho dx\, x^3 j_l(x) j_l(x), \tag{38}$$

also calculated in Appendix F, and next

$$\mathbb{C}_l(\rho) = \int_0^\rho dx\, x^3 j_l(x)[j_l(x)\mathrm{Ci}(x) + y_l(x)\mathrm{Si}(x)], \tag{39}$$

$$\mathbb{D}_l(\rho) = \int_0^\rho dx\, x^3 j_l(x)[j_l(x)\cos(x) + y_l(x)\sin(x)], \tag{40}$$

both of which are calculated in Appendix G. None of the functions $\mathbb{A}_l$, $\mathbb{B}_l$, $\mathbb{C}_l$, and $\mathbb{D}_l$, are standard. $\mathbb{B}_l$, and $\mathbb{D}_l$, are introduced to obtain $\mathbb{C}_l$. One then obtains for $l = $ odd

$$\mathbb{J}_l(k_F R) = \frac{1}{\pi(k_F R)^4}\left(\mathbb{C}_l(k_F R) + \sum_{\nu=0}^{(l-3)/2} b_\nu^{(l)}\mathbb{F}_{l-1-2\nu}^l(k_F R)\right). \tag{41}$$

Even though the approximation of constant J_L is not certain to be valid outside the asymptotic limit of large distance it is still of interest to consider the short-range limit of $k_F R \to 0$. This will provide information about the way that the different partial wave terms are mixed. The limit is easy to obtain. Consider the definition equation (25). The leading term in the integrand for $x \to 0$ is obtained from Eq. (28) and $j_l(x) \simeq x^l/(2l+1)!!$ and gives

$$\mathbb{J}_l(k_F R) \simeq -\frac{1}{\pi}\frac{(l-1)!!}{l!!}\frac{(k_F R)^{l-1}}{(l+3)(2l+1)!!} \times \begin{cases} \pi/2 & \text{for } l = \text{even}, \\ 1 & \text{for } l = \text{odd}. \end{cases} \tag{42}$$

This means that at short range the $l = 0$ term will dominate. The R^{-1} singularity is integrable, and causes no difficulty. In fact, the integral over space of $\mathbb{J}_0$ is proportional to the finite Pauli spin susceptibility of the free electron gas, as is well

known. Then in Eq. (23) $l = 0$ terms have $L = L'$ and, using Eqs. (21) and (42), one obtains for $k_F R \to 0$

$$\mathcal{F}(R) \simeq \frac{9\pi n^2}{E_F} \sum_{L=0}^{\infty} (2L+1) J_L^2 \frac{1}{6k_F R}. \tag{43}$$

The combination of J_L is quite different from the J^2_{RKKY} obtained in the long distance limit. One thing to notice is that the emphasis on large L is less at short range than at long range.

In the partial wave expansion of $\mathcal{H}_{s\text{-}d}$ one usually considers only terms up to the value of $L = K$ that has a negative $J_L = J_K$, due to the mixing exchange, if such a term exists at all. This will be $L = 2$ for 3-d impurities like Fe in Au or Mn in Cu, while it will be $L = 3$ for 4-f impurities, like Ce in La. By the triangular condition (22) one needs only values of l up to twice the maximum L included. When one does not consider either of the two limits of short or long distance, the result is still rather complicated to write down. While the recursion relations of the Appendices are therefore the most convenient for practical purposes, it may be interesting to write out explicitly the first few functions $\mathbb{J}_l$. Using Eqs. (37) and (41) together with the results of the Appendices one obtains

$$\mathbb{J}_0(k_F R) = [1/(2k_F R)^4][2k_F R \cos 2k_F R - \sin 2k_F R], \tag{44}$$

which is the standard RKKY result. Next

$$\begin{aligned}\mathbb{J}_1(k_F R) &= [1/\pi(k_F R)^4]\Big\{ \tfrac{1}{8}\mathrm{Ci}(k_F R)[5\cos 2k_F R \\ &\quad + 2k_F R \sin 2k_F R - 5 + 2(k_F R)^2] \\ &\quad - \tfrac{1}{8}\mathrm{Si}(k_F R)[2k_F R \cos 2k_F R - 5\sin 2k_F R] \\ &\quad + \int_0^{k_l R} \frac{dx}{x} \sin x f(x) - \frac{\pi}{4}\mathrm{Si}(2k_F R) - \tfrac{1}{4}k_F R \sin k_F R \Big\},\end{aligned} \tag{45}$$

where

$$f(x) = \int_0^{\infty} dt \frac{\sin t}{t + x} = \mathrm{Ci}(x)\sin x - \mathrm{Si}(x)\cos x + \frac{\pi}{2}\cos x \tag{46}$$

is a standard function. The remaining integral does not seem to be reducible to standard functions, and will turn up in all the odd l functions by virtue of the recursion relation for $\mathbb{C}_l$. A few other forms of this function are given in Appendix G. Finally

$$\begin{aligned}\mathbb{J}_2(k_F R) &= [1/2(k_F R)^4]\{\mathbb{A}_2(k_F R) + 3\mathbb{F}_2^2(k_F R)\} \\ &= [1/2(k_F R)^4]\{ -\tfrac{3}{8}[4\mathrm{Si}(2k_F R) \\ &\quad + 2k_F R \cos 2k_F R - 5\sin 2k_F R] \\ &\quad - \tfrac{1}{2}(k_F R)^4[j_1(k_F R)y_1(k_F R) \\ &\quad + j_2(k_F R)y_2(k_F R)] \\ &\quad + \tfrac{3}{2}[\mathrm{Si}(k_F R) - \sin k_F R] + \tfrac{3}{2}k_F R j_2(k_F R)\Big\} \\ &= \frac{1}{(2k_F R)^4}\Big\{ 12[\mathrm{Si}(k_F R) - \mathrm{Si}(2k_F R)] \\ &\quad + 36\left[\frac{\cos k_F R}{k_F R} - \frac{\sin k_F R}{(k_F R)^2}\right] \\ &\quad + \left[2k_F R - \frac{72}{2k_F R}\right]\cos 2k_F R \\ &\quad + \left[\frac{72}{(2k_F R)^2} - 13\right]\sin 2k_F R \Big\}.\end{aligned} \tag{47}$$

3. ELECTRON SPIN POLARIZATION

From the derivation in the previous section it is clear that one does not have to use the same $\mathcal{H}_{s\text{-}d}$ at the two ends of the diagram in Fig. 1, and that rather the indirect exchange interaction applies to *any* two spins that couple to the electron gas by an interaction of the form in Eq. (1). The result therefore also include cases of nuclear spins interacting with the electrons. It is of interest to consider the electron spin polarization, which may be defined in the following way. The RKKY interaction is written in the form

$$\mathcal{H}_{\mathrm{RKKY}} = -(2I_0/N_a)\mathbf{s}(R)\cdot\mathbf{S}_2, \tag{48}$$

similar to an $\mathcal{H}_{s\text{-}d}$ type interaction with the spin $\mathbf{S}_2$ that only includes the $L' = 0$ partial wave, and where $\mathbf{s}(R)$ is the spin operator due to the electrons polarized by the spin $\mathbf{S}_1$ at the distance R, i.e., the electron spin polarization. In order that this quantity be a property solely of the spin $\mathbf{S}_1$ and the electrons, *the probe spin* $\mathbf{S}_2$ *must be pointlike*, whereby the exchange integral becomes just $I(k,k') = I_0$ as assumed in Eq. (48). This can be seen in the following way. Suppose the exchange integral is of the direct exchange type

$$I(\mathbf{k},\mathbf{k}') = N_a \int\int d^3r\, d^3r'\, \psi^*_{\mathbf{k}'}(\mathbf{r}')\phi^*(\mathbf{r})V(\mathbf{r}-\mathbf{r}')\psi_{\mathbf{k}}(\mathbf{r})\phi(\mathbf{r}'). \tag{49}$$

The probe is then an electronic spin in a state with wavefunction $\phi(\mathbf{r})$. If it has to be pointlike one must assume (i) zero size of this orbital: $|\phi(\mathbf{r})|^2 = \delta(\mathbf{r})$, and (ii) zero range of the interaction: $V(\mathbf{r}-\mathbf{r}') = (I_0/N_a)V\delta(\mathbf{r}-\mathbf{r}')$. For free electrons in states with wavefunctions $\psi_{\mathbf{k}}(\mathbf{r}) = V^{-1/2}e^{i\mathbf{k}\cdot\mathbf{r}}$ one then has

$$I(\mathbf{k},\mathbf{k}') = I_0. \tag{50}$$

Other probes could lead to similar results, like the hyperfine coupling to nuclei of the metal. *The concept of spin polarization is only useful when the probe can be considered pointlike, otherwise one must use the more general coupling* $\mathcal{H}_{\mathrm{RKKY}}$, which in any case is the fundamental quantity. From (19) one now obtains, observing that for $L' = 0$ one must have $l = L$, and the column interchange invariance of (21), $\mathbf{s}(R) = \mathcal{F}(R)(N_a/2I_0)\mathbf{S}_1$,

$$\begin{aligned}\mathbf{s}(R) = -\frac{9\pi N_e^2}{2E_F N_a}\sum_{L=0}^{\infty}(2L+1)\Big(-\frac{1}{\pi k_F^4}\int_0^{k_l} dk\, k^2 \\ \times j_L(kR)\mathcal{P}\int_0^{\infty} dk'\,\frac{k'^2}{k'^2-k^2} j_L(k'R)J_L(k,k')\Big) \\ \times\mathbf{S}_1.\end{aligned} \tag{51}$$

If in addition one takes $J_L(k,k') = J_L = \text{constant}$, as discussed in the previous section, then

$$\mathbf{s}(R) = -\frac{9\pi N_e^2}{2E_F N_a}\sum_{L=0}^{\infty}(2L+1)J_L\mathbb{J}_l(k_F R)\mathbf{S}_1, \tag{52}$$

and the asymptotic limit becomes the well-known [cf (3)]

$$\mathbf{s}(R) \sim -\frac{9\pi N_e^2}{2E_F N_a}\left[\sum_{L=0}^{\infty}(-1)^L(2L+1)J_L\right]\frac{\cos 2k_F R}{(2k_F R)^3}\mathbf{S}_1. \tag{53}$$

It is clear from this that it is impossible to generalize from the

spin polarization back to the RKKY interaction, as the calculation in the previous section showed. This is most clearly seen by comparing (52) with (23). The coupling to a *pointlike* probe S_2 does not contain the information relevant for the coupling to a probe with *structure*, whereas the reverse is true because one may shrink a structured probe to a mathematical point, with more or less justification in the physical reality.

APPENDIX A: RECURSION RELATION FOR $\mathbb{I}_l(x)$

In this appendix is constructed a recursion relation for the integral (a Cauchy principal value)

$$\mathbb{I}_l(x) = \fint_0^\infty dy \frac{y^2}{y^2 - x^2} j_l(y), \tag{A1}$$

where j_l is a spherical Bessel function. The following standard recursion relation holds

$$l j_{l-1}(y) - (l+1) j_{l+1}(y) = (2l+1)(d/dy) j_l(y). \tag{A2}$$

Operating with the integration in (A1) this gives

$$l\mathbb{I}_{l-1}(x) - (l+1)\mathbb{I}_{l+1}(x) = (2l+1) \fint_0^\infty dy \frac{y^2}{y^2 - x^2} \frac{d}{dy} j_l(y). \tag{A3}$$

By a partial integration, and using

$$\frac{d}{dy}\left(\frac{y^2}{y^2 - x^2}\right) = \frac{d}{dy}\left(\frac{x^2}{y^2 - x^2}\right) = -\frac{2yx^2}{(y^2 - x^2)^2} = -x \frac{d}{dx}\left(\frac{y}{y^2 - x^2}\right), \tag{A4}$$

Eq. (A3) becomes

$$l\mathbb{I}_{l-1}(x) - (l+1)\mathbb{I}_{l+1}(x) = x \frac{d}{dx} \fint_0^\infty dy \frac{y}{y^2 - x^2} (2l+1) j_l(y). \tag{A5}$$

Using another standard recursion relation

$$(2l+1) j_l(y) = y[j_{l-1}(y) + j_{l+1}(y)], \tag{A6}$$

one gets in (A5)

$$l\mathbb{I}_{l-1}(x) - (l+1)\mathbb{I}_{l+1}(x) = x \frac{d}{dx} \fint_0^\infty dy \frac{y^2}{y^2 - x^2} (j_{l-1}(y) + j_{l+1}(y)) = x \frac{d}{dx} [\mathbb{I}_{l-1}(x) + \mathbb{I}_{l+1}(x)]. \tag{A7}$$

Introducing

$$F_l \equiv \mathbb{I}_{l-1} + \mathbb{I}_{l+1} \tag{A8}$$

and $t = \ln x$ this becomes the simple differential equation

$$(d/dt) F_l + (l+1) F_l = (2l+1)\mathbb{I}_{l-1}, \tag{A9}$$

with the solution

$$F_l = \exp\left[-(l+1)\int dt\right]\left\{(2l+1)\int dt\, \mathbb{I}_{l-1} \times \exp\left[(l+1)\int dt\right] + C_l\right\}$$

$$= x^{-(l+1)}\left\{(2l+1)\int dx\, x^l \mathbb{I}_{l-1}(x) + C_l\right\}, \tag{A10}$$

where C_l is a constant. Then one has

$$\mathbb{I}_{l+1}(x) = -\mathbb{I}_{l-1}(x) + \frac{2l+1}{x^{l+1}} \int dx\, x^l \mathbb{I}_{l-1}(x) + \frac{C_l}{x^{l+1}}. \tag{A11}$$

To find the constants C_l look at the $x = 0$ values:

$$\mathbb{I}_l(0) = \int_0^\infty dy\, j_l(y) = \sqrt{\frac{\pi}{2}} \int_0^\infty dy\, y^{-1/2} J_{l+1/2}(y) \tag{A12}$$

$$= \frac{\sqrt{\pi}}{2} \Gamma\left(\frac{l+1}{2}\right) \Big/ \Gamma\left(\frac{l+2}{2}\right). \tag{A13}$$

In fact, using the recursion relation (A2) one sees immediately that

$$l\mathbb{I}_{l-1}(0) - (l+1)\mathbb{I}_{l+1}(0) = -(2l+1) j_l(0).$$

So for $l = 0$ one gets

$$\mathbb{I}_1(0) = 1, \tag{A14}$$

and for $l = 1,2,\cdots$

$$\mathbb{I}_{l+1}(0) = \frac{l}{l+1} \mathbb{I}_{l-1}(0). \tag{A15}$$

One also easily shows that (cf Appendix B)

$$\mathbb{I}_0(0) = \pi/2. \tag{A16}$$

This is equivalent to the result

$$\mathbb{I}_l(0) = \frac{(l-1)!!}{l!!} \times \begin{cases} \pi/2 & \text{for } l \text{ even}, \\ 1 & \text{for } l \text{ odd}, \end{cases} \tag{A17}$$

where (even)!! = 2·4·....·(even) and (odd)!! = 1·3·5·....·(odd), and 0!! = 1 and (−1)!! = 1. Letting $x \to 0$ in (A11) gives

$$\mathbb{I}_{l+1}(0) = -\mathbb{I}_{l-1}(0) + \frac{2l+1}{l+1} \mathbb{I}_{l-1}(0) + \lim_{x \to 0}\left(\frac{C_l}{x^{l+1}}\right) = \frac{l}{l+1} \mathbb{I}_{l-1}(0) + \lim_{x \to 0}\left(\frac{C_l}{x^{l+1}}\right).$$

Comparing with (A15) it follows that for $l = 0,1,2,\cdots$ one must have $C_l = 0$, and (A11) becomes

$$\mathbb{I}_{l+1}(x) = -\mathbb{I}_{l-1}(x) + \frac{2l+1}{x^{l+1}} \int_0^x du\, u^l \mathbb{I}_{l-1}(u). \tag{A18}$$

This is valid for $l = 1,2,3,\cdots$, while for $l = 0$ one must keep the form (A11) with $C_0 = 0$ since $\mathbb{I}_{-1}$ is logarithmically singular at $x \to 0$ (cf Appendix B).

APPENDIX B: INTEGRAL $\mathbb{I}_l(x)$

In this appendix is calculated the integral (a Cauchy principal value)

$$\mathbb{I}_l(x) = \fint_0^\infty dy \frac{y^2}{y^2 - x^2} j_l(y), \tag{B1}$$

where j_l is a spherical Bessel function of the first kind. The results are (B10) and (B13). Consider first some values of l for which the standard decomposition

$$j_l(y) = f_l(y) \sin y + (-1)^{l+1} f_{-l-1}(y) \cos y \tag{B2}$$

does not include more singular f than y^{-2}. These integrals are

relatively simple to calculate by contour integration and are tabulated. First

$$\mathbb{I}_0(x) = \fint_0^\infty dy \frac{y \sin y}{y^2 - x^2} = \frac{\pi}{2} \cos x = -\frac{\pi}{2} x y_0(x), \tag{B3}$$

where y_l is a spherical Bessel function of the second kind. Next

$$\begin{aligned}\mathbb{I}_{-1}(x) &= \fint_0^\infty dy \frac{y \cos y}{y^2 - x^2} = -\cos x \mathrm{Ci}(x) - \sin x \mathrm{Si}(x) \\ &= -x j_{-1}(x) \mathrm{Ci}(x) - x y_{-1}(x) \mathrm{Si}(x), \end{aligned} \tag{B4}$$

where $\mathrm{Ci}(x)$ and $\mathrm{Si}(x)$ are the sine and cosine integrals. Further

$$\begin{aligned}\mathbb{I}_1(x) &= \fint_0^\infty dy \frac{\sin y}{y^2 - x^2} - \fint_0^\infty dy \frac{y \cos y}{y^2 - x^2} \\ &= -x j_1(x) \mathrm{Ci}(x) - x y_1(x) \mathrm{Si}(x). \end{aligned} \tag{B5}$$

For $l = 2$ and further on the integrals will have singularities of the integrand at $y = 0$ when split in this way, and these will give an important contribution, as seen in the following. For example

$$\mathbb{I}_2(x) = -(\pi/2)[x y_2(x) + 3/x^2]. \tag{B6}$$

In Appendix A it was shown that $\mathbb{I}_l$ is finite at $x \to 0$ [cf (A17)] for $l = 0,1,2,\cdots$

$$\mathbb{I}_l(0) = \frac{\sqrt{\pi}}{2} \Gamma\left(\frac{l+1}{2}\right) \Gamma\left(\frac{l+2}{2}\right), \tag{B7}$$

while it follows from (B4) that due to $\mathrm{Ci}(x) \sim \ln x$

$$\mathbb{I}_{-1}(x) \sim -\ln x. \tag{B8}$$

According to the recursion relation of Appendix A one should have

$$\mathbb{I}_1(x) = -\mathbb{I}_{-1}(x) + \frac{1}{x} \int dx \mathbb{I}_{-1}(x)$$

as can easily be verified by a partial integration. Aside from this special case the recursion relation (A18) holds.

Consider the expansion

$$x y_l(x) = -\sum_{n=-l,-l+2,\cdots}^{\infty} \frac{(l-n-1)!!}{(l+n)!!} x^n, \tag{B9}$$

which has a number of terms that are singular for $x \to 0$, the strongest being $-(2l-1)!!/x^l$. For example, for $l = 2$ there is one singular term

$$-3/x^2,$$

which is exactly substracted out in (B6). In accordance with the general result that $\mathbb{I}_l$ is finite for $x \to 0$ [and for $l =$ even equal to $\pi/2$ times the $n = 0$ term in (B9)] it is now first to be shown that for $l = even$

$$\begin{aligned}\mathbb{I}_l(x) &= -\frac{\pi}{2}\left(x y_l(x) + \sum_{n=-l,-l+2,\cdots}^{n=-2} \frac{(l-n-1)!!}{(l+n)!!} x^n\right) \\ &= -\frac{\pi}{2}\left(x y_l(x) + \sum_{m=2,4,\cdots}^{m=l} \frac{(l+m-1)!!}{(l-m)!!} \frac{1}{x^m}\right). \end{aligned} \tag{B10}$$

This is constructed with the right behavior at $x \to 0$. By the principle of induction it is sufficient to show that if (B10) is true for some even value of l then it is also true for $l + 2$. This is done by the recursion relation proved in Appendix A, written in the form (A18)

$$\mathbb{I}_{l+2}(x) = -\mathbb{I}_l(x) + \frac{2l+3}{x^{l+2}} \int_0^x du\, u^{l+1} \mathbb{I}_l(u). \tag{B11}$$

Using the standard relation

$$\frac{d}{du}\left[u^{l+2} y_{l+1}(u)\right] = u^{l+2} y_l(u) \tag{B12}$$

one obtains by inserting (B10) in (B11) that

$$\begin{aligned}\mathbb{I}_{l+2}(x) &= -\mathbb{I}_l(x) - \frac{2l+3}{x^{l+2}} \frac{\pi}{2} \int_0^x du\, u^{l+2} y_l(u) - \frac{2l+3}{x^{l+2}} \\ &\quad \times \frac{\pi}{2} \sum_{m=2,4,\cdots}^{m=l} \frac{(l+m-1)!!}{(l-m)!!} \int_0^x du u^{l+1-m} \\ &= -\mathbb{I}_l(x) - \frac{2l+3}{x^{l+2}} \frac{\pi}{2} \left[x^{l+2} y_{l+1}(x) + (2l+1)!!\right] \\ &\quad - (2l+3) \frac{\pi}{2} \sum_{m=2,4,\cdots}^{m=l} \frac{(l+m-1)!!}{(l-m+2)!!} \frac{1}{x^m} \\ &= -\frac{\pi}{2}\left[(2l+3) y_{l+1}(x) - x y_l(x) + \frac{(2l+3)!!}{x^{l+2}} \right. \\ &\quad \left. + \sum_{m=2,4,\cdots}^{m=l} \left(\frac{2l+3}{l-m+2} - 1\right) \frac{(l+m-1)!!}{(l-m)!!} \frac{1}{x^m}\right] \\ &= -\frac{\pi}{2}\left(x y_{l+2}(x) + \sum_{m=2,4,\cdots}^{m=l+2} \frac{(l+m+1)!!}{(l-m+2)!!} \frac{1}{x^m}\right), \end{aligned}$$

where the relation similar to (A6) for y_l was used. This proves (B10).

For $l = odd$ one can generalize (B4) or (B5)

$$\mathbb{I}_l(x) = -x j_l(x) \mathrm{Ci}(x) - x y_l(x) \mathrm{Si}(x) - \sum_{\nu=0}^{(l-3)/2} \frac{b_\nu^{(l)}}{x^{l-2\nu}}. \tag{B13}$$

The last term substracts the singular terms from the second term. Using (B9) and the expansion

$$\mathrm{Si}(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)(2k+1)!} \tag{B14}$$

the term is

$$\begin{aligned}&\sum_{k=0} \sum_{n=-l,-l+2,\cdots} x^{2k+1+n} \frac{(-1)^k (l-n-1)!!}{(2k+1)(2k+1)!(l+n)!!} \\ &= \sum_{k=0} \sum_{m=0,1,2,\cdots} x^{2(k+m)+1-l} \frac{(-1)^k (2l-2m-1)!!}{(2k+1)(2k+1)!(2m)!!}. \end{aligned}$$

The sums run up to $2(k+m)+1-l = -2$, i.e., up to $\nu = k + m = (l-3)/2$ and can be written as

$$\sum_{\nu=0}^{(l-3)/2} \sum_{k=0}^{\nu} x^{2\nu+1-l} \frac{(-1)^k (2l-2\nu+2k-1)!!}{(2k+1)(2k+1)!(2\nu-2k)!!},$$

whereby (for $2\nu \leqslant l-3$)

$$b_\nu^{(l)} = \sum_{k=0}^{\nu} \frac{(-1)^k (2l-2\nu+2k-1)!!}{(2k+1)(2k+1)!(2\nu-2k)!!}. \tag{B15}$$

For example

$$b_0^{(l)} = (2l-1)!!. \tag{B16}$$

Now again use Appendix A, i.e., the recursion relation (B11), to prove (B13) by induction. First consider the integral

$$Q_l(x) = -\int_0^x du\, u^{l+2}\,[j_l(u)\mathrm{Ci}(u) + y_l(u)\mathrm{Si}(u)]\,. \quad \text{(B17)}$$

Using again (B12) which also holds for j_l one obtains

$$\begin{aligned} Q_l(x) &= -\int_0^x du\left\{\frac{d}{du}\left[u^{l+2}j_{l+1}(u)\right]\mathrm{Ci}(u)\right. \\ &\quad \left. + \frac{d}{du}\left[u^{l+2}y_{l+1}(u)\right]\mathrm{Si}(u)\right\} \\ &= -x^{l+2}\left[j_{l+1}(x)\,\mathrm{Ci}(x) + y_{l+1}(x)\,\mathrm{Si}(x)\right] \\ &\quad + \int_0^x du\, u^{l+1}\left[j_{l+1}(u)\cos u + y_{l+1}(u)\sin u\right]. \end{aligned}$$

By (C9) and (C7) then

$$\begin{aligned} Q_l(x) &= -x^{l+2}\left[j_{l+1}(x)\mathrm{Ci}(x) + y_{l+1}(x)\mathrm{Si}(x)\right] \\ &\quad + (-1)^l \int_0^x du\, u^{l+1} f_{-l-2}(u) \\ &= -x^{l+2}\left[j_{l+1}(x)\mathrm{Ci}(x) + y_{l+1}(x)\mathrm{Si}(x)\right] + 2(-1)^l \\ &\quad \times \sum_{k=0}^{l} \delta_{k,\mathrm{par}(l)}(-1)^{(l+k+2)/2}\frac{(l+1+\frac{1}{2},k)}{2^{k+1}} \\ &\quad \times \int_0^x du\, u^{l-k}, \end{aligned}$$

where par $(l)\equiv$parity of integer l = even or odd. The result therefore becomes

$$\begin{aligned} Q_l(x) = -x^{l+2}\Bigg\{ &\left[j_{l+1}(x)\mathrm{Ci}(x) + y_{l+1}(x)\mathrm{Si}(x)\right] \\ &+ \sum_{k=0}^{l}\delta_{k,\mathrm{par}(l)}(-1)^{(l-k)/2}\frac{(l+\frac{3}{2},k)}{2^k(l-k+1)}\frac{1}{x^{k+1}}\Bigg\}. \end{aligned} \quad \text{(B18)}$$

Then from (B11), (B13) and (B18) one has

$$\begin{aligned} \mathbb{I}_{l+2}(x) &= -x[-j_l(x) + (2l+3)j_{l+1}(x)/x]\mathrm{Ci}(x) \\ &\quad -x[-y_l(x) + (2l+3)y_{l+1}(x)/x]\mathrm{Si}(x) \\ &\quad + \sum_{\nu=0}^{(l-3)/2}\frac{b_\nu^{(l)}}{x^{l-1-2\nu}} \\ &\quad - \sum_{k=0}^{l}\delta_{k,\mathrm{par}(l)}(-1)^{(l-k)/2}\frac{(l+\frac{3}{2},k)}{2^k(l-k+1)} \\ &\quad \times \frac{2l+3}{x^{k+1}} - \sum_{\nu=0}^{(l-3)/2}\frac{b_\nu^{(l)}}{x^{(l-1-2\nu)}}\frac{2l+3}{2\nu+3}. \end{aligned}$$

Then using the relation (A6) on the first two terms, reorganizing the k-sum (since l = odd) as $k = l - 2\mu$, summing over μ from 0 to $(l-1)/2$, and changing variable to $\nu = \mu - 1$ [summing μ from 1 to $(l-1)/2$] in the remaining sums, one obtains

$$\begin{aligned} \mathbb{I}_{l+2}(x) &= -xj_{l+2}(x)\mathrm{Ci}(x) - xy_{l+2}(x)\mathrm{Si}(x) \\ &\quad - \sum_{\mu=0}^{(l-1)/2}(-1)^\mu\frac{(l+\frac{3}{2},l-2\mu)}{2^{l-2\mu}(2\mu+1)}\frac{2l+3}{x^{l+1-2\mu}} \\ &\quad + \sum_{\mu=1}^{(l-1)/2} b_{\mu-1}^{(l)}\frac{2(\mu-1-l)}{2\mu+1}\frac{1}{x^{l+1-2\mu}}. \end{aligned} \quad \text{(B19)}$$

This in fact proves (B13) because $\mathbb{I}_{l+2}$ must be finite when $x\to 0$, and the last two sums have the correct form to produce the required subtractions.

According to (B13) the subtraction term should be

$$-\sum_{\nu=0}^{(l-1)/2}\frac{b_\nu^{(l+2)}}{x^{l+1-2\nu}}.$$

It should then hold that in particular

$$\begin{aligned} b_0^{(l+2)} &= \frac{(l+\frac{3}{2},l)(2l+3)}{2^l} = \frac{(2l+1)!(2l+3)}{l!2^l} \\ &= (2l+3)!!\,, \end{aligned}$$

which agrees with (B16). Then for the remaining terms (for $1\leqslant\mu\leqslant(l-1)/2$) it should hold that

$$\begin{aligned} b_\mu^{(l+2)} &= -b_{\mu-1}^{(l)}\frac{2(\mu-1-l)}{(2\mu+1)} \\ &\quad + (-1)^\mu\frac{(l+\frac{3}{2},l-2\mu)(2l+3)}{2^{l-2\mu}(2\mu+1)}. \end{aligned} \quad \text{(B20)}$$

Since according to (B15) $b_\nu^{(l)} = 0$ for $\nu < 0$ one may include $\mu = 0$ in (B20). For completeness the remainder of this appendix is used to verify explicitly (B20). From the definition (B15) separate the $k = \mu$ term

$$\begin{aligned} b_\mu^{(l+2)} &= \sum_{k=0}^{\mu-1}\frac{(-1)^k(2l-2\mu+2k+3)!!}{(2k+1)(2k+1)!(2\mu-2k)!!} \\ &\quad + \frac{(-1)^\mu(2l+3)!!}{(2\mu+1)(2\mu+1)!} \end{aligned}$$

and

$$\begin{aligned} &\frac{2\mu-2-2l}{2\mu+1}b_{\mu-1}^{(l)} \\ &\quad = \sum_{k=0}^{\mu-1}\frac{(-1)^k(2l-2\mu+2k+1)!!}{(2k+1)(2k+1)!(2\mu-2k-2)!!} \\ &\quad \times \frac{(2\mu-2-2l)}{(2\mu+1)}. \end{aligned}$$

Observing that

$$\begin{aligned} &(2l-2\mu+2k+3)(2\mu+1) + (2\mu-2k)(2\mu-2-2l) \\ &\quad = (2l+3)(2k+1)\,, \end{aligned}$$

one obtains by adding the two sums

$$\begin{aligned} &\sum_{k=0}^{\mu-1}\frac{(-1)^k(2l-2\mu+2k+1)!!(2l+3)(2k+1)}{(2k+1)(2k+1)!(2\mu-2k)!!(2\mu+1)} \\ &\quad + \frac{(-1)^\mu(2l+3)!!}{(2\mu+1)(2\mu+1)!} \\ &\quad = \frac{2l+3}{2\mu+1}\sum_{k=0}^{\mu}\frac{(-1)^k(2l-2\mu+2k+1)!!}{(2k+1)!(2\mu-2k)!!}. \end{aligned}$$

Then (B20) becomes

$$\begin{aligned} &\sum_{k=0}^{\mu}\frac{(-1)^k(2l-2\mu+2k+1)!!}{(2k+1)!(2\mu-2k)!!} \\ &\quad = (-1)^\mu 2^{2\mu-l}(l+\tfrac{3}{2},l-2\mu)\,. \end{aligned} \quad \text{(B21)}$$

Using definitions $(2n)!!(2n+1)!! = (2n+1)!$, $(2n)!! = 2^n\, n!$, and (C3) (see) for the Hankel symbol, this may be written as

$$\begin{aligned} &\sum_{k=0}^{\mu}(-1)^k\frac{(2l-2\mu+2k+1)!}{(2k+1)!(l-\mu+k)!(\mu-k)!} \\ &\quad = (-1)^\mu 2^{2\mu}\frac{(2l+1-2\mu)!}{(l-2\mu)!(2\mu+1)!}. \end{aligned} \quad \text{(B22)}$$

The left-hand side can be further rewritten with $p = \mu - k$ $(0\leqslant p\leqslant\mu)$ in terms of binomial coefficients

$$\sum_{p=0}^{\mu}(-1)^p\frac{(2l-2p+1)!}{(2\mu-2p+1)!(l-p)!p!}$$
$$=(-1)^\mu\frac{(2l-2\mu)!}{l!}\sum_{p=0}^{\mu}(-1)^p\binom{l}{p}\binom{2l+1-2p}{2\mu+1-2p},$$

whereby (B22) becomes

$$\sum_{p=0}^{\mu}(-1)^p\binom{l}{p}\binom{2l+1-2p}{2\mu+1-2p}=2^{2\mu}\frac{(2l+1-2\mu)}{(l-2\mu)}\times\binom{l}{2\mu+1}. \quad \text{(B23)}$$

An equivalent version is

$$\sum_{p=0}^{\mu}\binom{p-(l+1)}{p}\binom{2l+1-2p}{2\mu+1-2p}=2^{2\mu}\frac{(2l+1-2\mu)}{l+1}\times\binom{l+1}{2\mu+1}. \quad \text{(B24)}$$

This is proved in Appendix D and therefore completes the independent proof of (B20). This was known to be a correct relation as mentioned after (B19) and therefore suggested the validity of the unusual relation (B23) and (B24). As will be seen from Appendix D, the independent proof given there would not have been possible without this prior knowledge of the right-hand side.

APPENDIX C: RELATIONS OF SPHERICAL BESSEL FUNCTIONS

In this appendix are derived relations combining the spherical Bessel functions with trigonometric functions. These relations are used in the other appendices. Standard relations are

$$j_l(x)=f_l(x)\sin x+(-1)^{l+1}f_{-l-1}(x)\cos x \quad \text{(C1)}$$

and

$$y_l(x)=(-1)^{l+1}j_{-l-1}(x). \quad \text{(C2)}$$

The functions f_l satisfy recursion relations $(l=0,\pm1,\pm2,\cdots)f_{l-1}+f_{l+1}=(2l+1)f_l/x$ and $f_0=x^{-1}, f_1=x^{-2}$. In the present context it is however more convenient to have a closed form of f_l. Using other standard relations it is relatively easy to obtain such a result. Let the Hankel symbol be given by

$$(l+\tfrac{1}{2},k)=\frac{(l+k)!}{k!(l-k)!} \quad \text{(C3)}$$

and let

$$\text{par}(n)\equiv\text{parity of integer } n=\text{even or odd}. \quad \text{(C4)}$$

Then one obtains for $l=0,1,2,\cdots$

$$f_l(x)=2\sum_{k=0}^{l}(-1)^{\text{par}(l)}\delta_{k,\text{par}(l)}(-1)^{(l+k)/2}\frac{(l+\frac{1}{2},k)}{(2x)^{k+1}}. \quad \text{(C5)}$$

For example

$$f_0=\frac{1}{x},\ f_1=\frac{1}{x^2},\ f_2=\frac{3}{x^3}-\frac{1}{x},\ f_3=\frac{15}{x^4}-\frac{6}{x^2}, \quad \text{(C6)}$$
$$f_4=\frac{105}{x^5}-\frac{45}{x^3}+\frac{1}{x}.$$

Similarly

$$f_{-l-1}(x)=2\sum_{k=0}^{l-1}\delta_{k,\text{par}(l-1)}(-1)^{(l+k+1)/2}\frac{(l+\frac{1}{2},k)}{(2x)^{k+1}}, \quad \text{(C7)}$$

for $l=0,1,2,\cdots$. For example

$$f_{-1}=0,\ f_{-2}=-\frac{1}{x},\ f_{-3}=\frac{3}{x^2},\ f_{-4}=-\frac{15}{x^3}+\frac{1}{x}, \quad \text{(C8)}$$
$$f_{-5}=\frac{105}{x^4}-\frac{10}{x^2}.$$

Furthermore, it is easily shown that

$$j_l(x)\cos x+y_l(x)\sin x=(-1)^{l+1}f_{-l-1}(x) \quad \text{(C9)}$$

and

$$j_l(x)y_l(x)=(-1)^l f_l(x)f_{-l-1}(x)\cos 2x+\tfrac{1}{2}[f_{-l-1}(x)^2-f_l(x)^2]\sin 2x. \quad \text{(C10)}$$

Also

$$j_l(x)j_l(x)=(-1)^{l+1}f_l(x)f_{-l-1}(x)\sin 2x+\tfrac{1}{2}[f_{-l-1}(x)^2-f_l(x)^2]\cos 2x+\tfrac{1}{2}[f_l(x)^2+f_{-l-1}(x)^2]. \quad \text{(C11)}$$

APPENDIX D: A SUM OF PRODUCTS OF BINOMIAL COEFFICIENTS

In this appendix is proved the relation between nonnegative integers l and μ

$$\sum_{p=0}^{\mu}(-1)^p\binom{l}{p}\binom{2l+1-2p}{2\mu+1-2p}=2^{2\mu}\frac{(2l+1-2\mu)}{(l-2\mu)}\times\binom{l}{2\mu+1}, \quad \text{(D1)}$$

for $0\leqslant\mu<(l-1)/2$. It is obviously satisfied for $\mu=0$ and any l. First is performed a series of algebraic rearrangements that aim at producing a relation between polynomials in l. Divide out the $p=\mu$ term [except $(-1)^\mu$]:

$$\sum_{p=0}^{\mu}(-1)^p\frac{\binom{l}{p}\binom{2l+1-2p}{2\mu+1-2p}}{\binom{l}{\mu}\binom{2l+1-2\mu}{1}}=\frac{2^{2\mu}(2l+1-2\mu)\binom{l}{2\mu+1}}{(l-2\mu)\binom{l}{\mu}\binom{2l+1-2\mu}{1}},$$
$$\sum_{p=0}^{\mu}(-1)^p\frac{\mu!(l-\mu)!(2l+1-2p)!}{p!(l-p)!(2l+1-2\mu)!(2\mu+1-2p)!}=\frac{2^{2\mu}\mu!(l-\mu)!}{(2\mu+1)!(l-2\mu)!}.$$

Now use, four times, $(2n+1)!=(2n+1)!!2^n n!$ and move over $(2\mu+1)!!$

$$\sum_{p=0}^{\mu}(-1)^p\frac{\mu!2^{l-p}(2l+1-2p)!!(2\mu+1)!!}{p!2^{l-\mu}(2l+1-2\mu)!!(2\mu+1-2p)!!\ 2^{\mu-p}(\mu-p)!}=\frac{2^\mu(l-\mu)!}{(l-2\mu)!},$$
$$\sum_{p=0}^{\mu}(-1)^p\binom{\mu}{p}(2\mu+1)(2\mu-1)\cdots(2\mu+3-2p)\ (p\text{ factors})\times(2l+1-2p)(2l-1-2p)\cdots(2l+3-2\mu)\ (\mu-p\text{ factors})=2^\mu(l-\mu)(l-\mu-1)\cdots(l-2\mu+1)\ (\mu\text{ factors}).$$

It is convenient to write this in terms of the factorial polynomials defined by

$$\underbrace{x^{(n)} \equiv x(x-1)\cdots(x-n+1)}_{n \text{ factors}}. \tag{D2}$$

The desired polynomial relation then becomes

$$\sum_{p=0}^{\mu} (-1)^p \binom{\mu}{p} (\mu+\tfrac{1}{2})^{(p)} (l+\tfrac{1}{2}-p)^{(\mu-p)} = 2^{\mu}(l-\mu)^{(\mu)}. \tag{D3}$$

From this it is evident that the two sides are polynomials in l of degree μ. Furthermore the right-hand side is completely characterized by the μ zeros

$$l = \mu, \mu+1, \ldots, 2\mu-1. \tag{D4}$$

It is therefore sufficient to show that the polynomial on the left-hand side has the same zeros, whereby the polynomials will be identical for all values of l and μ. Write

$$\mu = l - m$$

and then the left-hand side must vanish for $m = 0, 1, \ldots, (l-1)/2$. First

$$(\mu+\tfrac{1}{2})^{(p)} (l+\tfrac{1}{2}-p)^{(\mu-p)} = (l+\tfrac{1}{2}-m)^{(p)} \times (l+\tfrac{1}{2})^{(\mu)} / (l+\tfrac{1}{2})^{(p)}.$$

The $(l+\frac{1}{2})^{(\mu)}$ is nonzero for all integer l and μ and can go outside the sum over p. Then reorganize, using the definition (D2)

$$\frac{(l+\frac{1}{2}-m)^{(p)}}{(l+\frac{1}{2})^{(p)}} = \frac{(l+\frac{1}{2})^{(p+m)}}{(l+\frac{1}{2})^{(p)}(l+\frac{1}{2})^{(m)}} = \frac{(l+\frac{1}{2}-p)^{(m)}}{(l+\frac{1}{2})^{(m)}}.$$

Now

$$(l+\tfrac{1}{2}-p)^{(m)} = \frac{d^{(m)}}{dx^m} (x^{l+(1/2)-p})_{x=1}$$

makes the left-hand side of (D3) proportional to

$$\begin{aligned} &\frac{d^{(m)}}{dx^m}\left[x^{l+\frac{1}{2}} \sum_{p=0}^{\mu} \binom{\mu}{p} 1^{\mu-p} \left(-\frac{1}{x}\right)^p\right]_{x=1} \\ &= \frac{d^{(m)}}{dx^m}\left[x^{l+(1/2)}\left(1-\frac{1}{x}\right)^{\mu}\right]_{x=1} \\ &= \frac{d^{(m)}}{dx^m}\left[x^{m+(1/2)}(x-1)^{l-m}\right]_{x=1}. \end{aligned} \tag{D5}$$

Doing the differentiation by the Leibniz rule produces as the term having the smallest power of $x-1$:

$$\left(x^{m+(1/2)} \frac{d^{(m)}}{dx^m} (x-1)^{l-m}\right)_{x=1} \propto [(x-1)^{l-2m}]_{x=1}.$$

The smallest power possible is then obtained for $m = (l-1)/2$ and is equal to 1. Thereby all terms of (D5) vanish. This completes the proof of (D1).

APPENDIX E: INTEGRAL $\mathbb{F}^l_m(x)$.

In this appendix is calculated the integral $(0 \leqslant m \leqslant l)$

$$\mathbb{F}^l_m(\rho) = \int_0^{\rho} dx\, x^{2-m} j_l(x). \tag{E1}$$

This is an example from a class of general standard integrals of Bessel functions. Relations pertinent to the present work are given below. Using (B12) one obtains the recursion relation

$$\mathbb{F}^l_m(\rho) = \rho^{2-m} j_{l+1}(\rho) + (l+m)\mathbb{F}^{l+1}_{m+1}(\rho). \tag{E2}$$

Similarly, using the standard relation

$$x j_l(x) = (l-1) j_{l-1}(x) - x(d/dx) j_{l-1}(x) \tag{E3}$$

one obtains for $l - m + 3 > 0$

$$\mathbb{F}^l_{m-1}(\rho) = (l-m+3)\mathbb{F}^l_m(\rho) - \rho^{3-m} j_l(\rho). \tag{E4}$$

Using these two recursions one may generate all $\mathbb{F}^l_m$ starting with the following, which are obtained by direct integration

$$\mathbb{F}^0_0(\rho) = \rho^2 j_1(\rho), \tag{E5}$$

$$\mathbb{F}^1_1(\rho) = \mathrm{Si}(\rho) - \sin\rho, \tag{E6}$$

$$\mathbb{F}^1_0(\rho) = 2 - 2\cos\rho - \rho\sin\rho. \tag{E7}$$

For example, to produce $\mathbb{F}^6_2$, starting with $\mathbb{F}^1_1$ three steps with (E2) give $\mathbb{F}^4_4$ and then two steps (E4) give $\mathbb{F}^6_2$.

APPENDIX F: INTEGRALS $\mathbb{A}_l(x)$ AND $\mathbb{B}_l(x)$

This appendix calculates the integral $(l \geqslant 0)$

$$\mathbb{A}_l(\rho) = \int_0^{\rho} dx\, x^3 j_l(x) y_l(x), \tag{F1}$$

where j_l and y_l are spherical Bessel functions. The recursion relation derived in the following holds for any product of two spherical Bessel functions in the integrand, hence for example also for

$$\mathbb{B}_l(\rho) = \int_0^{\rho} dx\, x^3 j_l(x) j_l(x). \tag{F2}$$

The leading power of x in the integrand of (F1) for $x \to 0$ is $-x^2/2l + 1$ and the integral is well behaved for all ρ. For $\rho \to 0$ one has

$$\mathbb{A}_l(\rho) \simeq -\frac{\rho^3}{3(2l+1)} + O(\rho^5), \tag{F3}$$

$$\mathbb{B}_l(\rho) \simeq \frac{\rho^{(2l+4)}}{(2l+4)[(2l+1)!!]^2} + O(\rho^{2l+6}). \tag{F4}$$

The following is a standard integral

$$\begin{aligned} S(x) &= \int_0^x dt\, t^{2l+3} j_l(t) y_l(t) \\ &= \frac{x^{2l+4}}{4(l+1)} [j_l(x) y_l(x) + j_{l+1}(x) y_{l+1}(x)]. \end{aligned} \tag{F5}$$

A partial integration gives

$$\begin{aligned} \mathbb{A}_l(\rho) &= \int_0^{\rho} dx\, x^{-2l} \frac{dS(x)}{dx} \\ &= \rho^{-2l} S(\rho) + 2l \int_0^{\rho} dx\, x^{-2l-1} S(x) \\ &= \frac{\rho^4}{4(l+1)} [j_l(\rho) y_l(\rho) + j_{l+1}(\rho) y_{l+1}(\rho)] \\ &\quad + \frac{1}{4} \frac{2l}{l+1} [\mathbb{A}_l(\rho) + \mathbb{A}_{l+1}(\rho)]. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{A}_{l+1}(\rho) &= \frac{l+2}{l} \mathbb{A}_l(\rho) \\ &\quad - \frac{\rho^4}{2l} [j_l(\rho) y_l(\rho) + j_{l+1}(\rho) y_{l+1}(\rho)]. \end{aligned} \tag{F6}$$

In particular, for $l = 0$ one gets, using (C10), (C6), and (C8)

$$\mathbb{A}_0(\rho) = \tfrac{1}{4}\rho^4[j_0(\rho)y_0(\rho) + j_1(\rho)y_1(\rho)]$$
$$= \tfrac{1}{8}(2\rho\cos 2\rho - \sin 2\rho)\,. \tag{F7}$$

The case of $l = 1$ must be calculated separately. The integration is elementary and gives

$$\mathbb{A}_1(\rho) = -\tfrac{1}{8}[4\mathrm{Si}(2\rho) + 2\rho\cos 2\rho - 5\sin 2\rho\ ]\,, \tag{F8}$$

where $\mathrm{Si}(x)$ is the sine integral

$$\mathrm{Si}(x) = \int_0^x dt\,\frac{\sin t}{t}\,. \tag{F9}$$

This completes the calculation of $\mathbb{A}_l$. For $\mathbb{B}_l$ one obtains similarly

$$\mathbb{B}_{l+1}(\rho) = \frac{l+2}{l}\,\mathbb{B}_l(\rho) - \frac{\rho^4}{2l}\,[j_l(\rho)^2 + j_{l+1}(\rho)^2]\,. \tag{F10}$$

Furthermore, using (C11), (C6), and (C8), for $l = 0$

$$\mathbb{B}_0(\rho) = \tfrac{1}{4}\rho^4(j_0(\rho)^2 + j_1(\rho)^2)$$
$$= \tfrac{1}{8}(2\rho^2 + 1 - 2\rho\sin 2\rho - \cos 2\rho)\,, \tag{F11}$$

and by elementary integration of (F2) for $l = 1$

$$\mathbb{B}_1(\rho) = \tfrac{1}{8}[4\mathrm{Cin}(2\rho) + 5\cos 2\rho + 2\rho\sin 2\rho - 5 + 2\rho^2]\,, \tag{F12}$$

where Cin is the cosine integral

$$\mathrm{Cin}(x) = \int_0^x dt\,\frac{1-\cos t}{t}\,.$$

This completes the calculation of $\mathbb{B}_l$.

APPENDIX G: INTEGRALS $\mathbb{C}_l(x)$ AND $\mathbb{D}_l(x)$.

In this appendix the integral ($l \geqslant 0$) is calculated:

$$\mathbb{C}_l(\rho) = \int_0^\rho dx\, x^3 j_l(x)[j_l(x)\mathrm{Ci}(x) + y_l(x)\mathrm{Si}(x)]\,, \tag{G1}$$

where j_l and y_l are spherical Bessel functions of the first and second kinds and where Ci and Si are the cosine and sine integrals. The result is given in terms of the recursion relation (G12) and the auxiliary function $\mathbb{D}_l$ defined in (G4) and (G6) together with the starting function $\mathbb{C}_1(\rho)$ given in (G20). According to Appendix F one may express $\mathbb{C}_l$ in terms of $\mathbb{A}_l$ and $\mathbb{B}_l$ as

$$\mathbb{C}_l(\rho) = \int_0^\rho dx\left(\frac{d\mathbb{B}_l(x)}{dx}\,\mathrm{Ci}(x) + \frac{d\mathbb{A}_l(x)}{dx}\,\mathrm{Si}(x)\right). \tag{G2}$$

Then using the recursion relations for $\mathbb{A}_l$ and $\mathbb{B}_l$, (F6) and (F10), one obtains by a partial integration of the second terms

$$\mathbb{C}_{l+1}(\rho) = \frac{l+2}{l}\,\mathbb{C}_l(\rho) - \frac{\rho^4}{2l}\,[j_l(\rho)^2 + j_{l+1}(\rho)^2]\mathrm{Ci}(\rho)$$
$$- \frac{\rho^4}{2l}\,[j_l(\rho)y_l(\rho) + j_{l+1}(\rho)y_{l+1}(\rho)]\mathrm{Si}(\rho)$$
$$+ \frac{1}{2l}\,[\mathbb{D}_l(\rho) + \mathbb{D}_{l+1}(\rho)]\,, \tag{G3}$$

where

$$\mathbb{D}_l(\rho) = \int_0^\rho dx\, x^3 j_l(x)[j_l(x)\cos x + y_l(x)\sin x]\,. \tag{G4}$$

According to (C9) one may write this as

$$\mathbb{D}_l(\rho) = \int_0^\rho dx\, x^3 j_l(x)(-1)^{l+1} f_{-l-1}(x)\,, \tag{G5}$$

where, by (C7), and for $l = 0,1,2,\cdots$

$$f_{-l-1}(x) = 2\sum_{k=0}^{l-1}\delta_{k,\mathrm{par}(l-1)}(-1)^{(l+k+1)/2}\,\frac{(l+\frac{1}{2},k)}{(2x)^{k+1}}\,.$$

Using the functions $\mathbb{F}_m^l$ of Appendix E one then obtains

$$\mathbb{D}_l(\rho) = (-1)^{l+1}\sum_{k=0}^{l-1}\delta_{k,\mathrm{par}(l-1)}(-1)^{(l+k+1)/2}$$
$$\times\,\frac{(l+\frac{1}{2},k)}{2^k}\,\mathbb{F}_k^l(\rho)\,, \tag{G6}$$

where the Kronecker delta is defined in (C4) and the Hankel symbol in (C3). This completes the calculation of $\mathbb{D}_l$. For example, we have

$$\mathbb{D}_0(\rho) = 0\,, \tag{G7}$$

$$\mathbb{D}_1(\rho) = \rho\sin\rho + 2\cos\rho - 2\,, \tag{G8}$$

$$\mathbb{D}_2(\rho) = \frac{9}{\rho}\sin\rho - 3\cos\rho - 6\,. \tag{G9}$$

The leading term of $\mathbb{D}_l(\rho)$ for $\rho \to 0$ is obtained using $xj_l(x)y_l(x) \simeq -(2l+1)^{-1}$, so that for $l \geqslant 1$

$$\mathbb{D}_l(\rho) \simeq -\frac{1}{4}\,\frac{\rho^4}{2l+1} + O(\rho^6)\,. \tag{G10}$$

It is usually convenient to write the recursion relation for $\mathbb{C}_l$ in the form obtained from (G3) by the use of (F6) and (F10), in terms of a new function

$$\widetilde{\mathbb{C}}_l(\rho) = \mathbb{C}_l(\rho) - \mathbb{B}_l(\rho)\,\mathrm{Ci}(\rho) - \mathbb{A}_l(\rho)\,\mathrm{Si}(\rho)\,. \tag{G11}$$

It then becomes

$$\widetilde{\mathbb{C}}_{l+1}(\rho) = [(l+2)/l\ ]\widetilde{\mathbb{C}}_l(\rho) + \frac{1}{2l}\,[\mathbb{D}_l(\rho) + \mathbb{D}_{l+1}(\rho)]\,. \tag{G12}$$

For $l = 0$ one has

$$\widetilde{\mathbb{C}}_0(\rho) = -\tfrac{1}{4}\mathbb{D}_1(\rho)\,, \tag{G13}$$

and

$$\mathbb{C}_0(\rho) = \mathbb{B}_0(\rho)\,\mathrm{Ci}(\rho) + \mathbb{A}_0(\rho)\,\mathrm{Si}(\rho) - \tfrac{1}{4}\mathbb{D}_1(\rho)\,, \tag{G14}$$

where $\mathbb{A}_0$, $\mathbb{B}_0$, and $\mathbb{D}_1$ were given in (F7), (F11), and (G8). To start the recursion one then needs $\widetilde{\mathbb{C}}_1$ or $\mathbb{C}_1$. A direct partial integration gives

$$\widetilde{\mathbb{C}}_l(\rho) = -\int_0^\rho \frac{dx}{x}\,[\mathbb{B}_l(x)\cos x + \mathbb{A}_l(x)\sin x]\,. \tag{G15}$$

Using (F8) and (F12) this reduces for $l = 1$ to

$$\widetilde{\mathbb{C}}_1(\rho)$$
$$= \frac{1}{2}\int_0^\rho \frac{dx}{x}\,[\mathrm{Si}(2x)\sin x - \mathrm{Cin}(2x)\cos x] - \tfrac{1}{4}\rho\sin\rho\,. \tag{G16}$$

Neither this integral, nor $\mathbb{C}_1$, seem to be reducible to standard functions. In the remainder of this appendix is given various forms that may be helpful in the evaluation. First perform a partial integration of (G16)

$$\widetilde{\mathbb{C}}_1(\rho) = \tfrac{1}{2}[\mathrm{Si}(2\rho)\mathrm{Si}(\rho) - \mathrm{Cin}(2\rho)\,\mathrm{Ci}(\rho)] - \tfrac{1}{4}\rho\sin\rho$$
$$- \int_0^\rho \frac{dx}{x}\sin x[\mathrm{Si}(x)\cos x - \mathrm{Ci}(x)\sin x]\,. \tag{G17}$$

Using the standard relation

$$f(x)\equiv\int_0^\infty dt\,\frac{\sin t}{t+x}$$
$$=\mathrm{Ci}(x)\sin x-\mathrm{Si}(x)\cos x+(\pi/2)\cos x \qquad \text{(G18)}$$

one obtains

$$\tilde{C}_1(\rho)=\tfrac{1}{2}[\mathrm{Si}(2\rho)\,\mathrm{Si}(\rho)-\mathrm{Cin}(2\rho)\,\mathrm{Ci}(\rho)]-\tfrac{1}{4}\rho\sin\rho$$
$$+\int_0^\rho\frac{dx}{x}\sin x f(x)-\frac{\pi}{4}\mathrm{Si}(2\rho)\,. \qquad \text{(G19)}$$

Alternatively, using (F8) and (F12)

$$\tilde{C}_1(\rho)=\tfrac{1}{8}\mathrm{Ci}(\rho)(5\cos 2\rho+2\rho\sin 2\rho-5+2\rho^2)$$
$$-\tfrac{1}{8}\mathrm{Si}(\rho)(2\rho\cos 2\rho-5\sin 2\rho) \qquad \text{(G20)}$$
$$+\int_0^\rho\frac{dx}{x}\sin x f(x)-\frac{\pi}{4}\mathrm{Si}(2\rho)-\tfrac{1}{4}\rho\sin\rho\,.$$

The integral in (G16) may be reduced by the definitions of Si and Cin and elementary trigonometric relations

$$\int_0^\rho\frac{dx}{x}\,[\mathrm{Si}(2x)\sin x-\mathrm{Cin}(2x)\cos x]$$
$$=2\int_0^\rho\frac{dx}{x}\int_0^x\frac{du}{u}\sin u\sin(x-u) \qquad \text{(G21)}$$
$$=\int_0^\rho dx\int_0^x du\,\frac{\sin u}{u}\,\frac{\sin(x-u)}{x-u}\,.$$

For $\rho\to\infty$ one obtains from $f(x)\sim(1/x)+O(x^{-3})$

$$\int_0^\rho\frac{dx}{x}\sin x f(x)\sim\int d\rho\,\frac{\sin\rho}{\rho^2}\sim-\frac{\sin\rho}{\rho}+\mathrm{Ci}(\rho)$$
$$\sim-\frac{\cos\rho}{\rho^2}-2\,\frac{\sin\rho}{\rho^3}+O(\rho^{-4}), \qquad \text{(G22)}$$

while for $\rho\to 0$ using $f(x)\to\pi/2$ for $x\to 0$

$$\int_0^\rho\frac{dx}{x}\sin x f(x)\simeq\frac{\pi}{2}\mathrm{Si}(\rho)\simeq\frac{\pi}{2}\rho\,. \qquad \text{(G23)}$$

ACKNOWLEDGMENTS

The author thanks Professor H. Højgaard Jensen and Dr. James S. Schilling for critical reading of the manuscript and valuable comments.

[1]C. Kittel, Solid State Phys. **22**, 1 (1968)
[2]The *s-d* exchange Hamiltonian is due to C. Zener, Phys. Rev. **81**, 440 (1951) and T. Kasuya, Prog. Theor. Phys. **16**, 45 (1956). A simple derivation is given in U. Larsen, J. Phys. C **4**, 1835 (1971).
[3]The notation is: N_a = number of atoms in crystal, N_e = number of electrons, n = number of electrons per atom = N_e/N_a, $\sigma_{\alpha\alpha'}$ = Pauli spin matrices, $n(E_F)$ = electron density of states of both spin directions at E_F = Fermi energy, k_B = Boltzmann's constant, $T_F=E_F/k_B$ = Fermi temperature, R_{ij} = distance between two spins $\mathbf{S}_i$ and $\mathbf{S}_j$, $\mathbf{k}$ = electron wave vector, $k=|\mathbf{k}|$ = wave number, $\hat{k}=\mathbf{k}/k$ = unit wave vector, k_F = Fermi wave number, $\mathcal{F}(R_{ij})$ = RKKY coupling constant between spins at distance R_{ij}.
[4]J.R. Schrieffer and P.A. Wolff, Phys. Rev. **149**, 491 (1966).
[5]P.W. Anderson, Phys. Rev. **124**, 41 (1961).
[6]J.R. Schrieffer, J. Appl. Phys. **38**, 1143 (1967).
[7]B. Coqblin, J.R. Schrieffer, Phys. Rev. **185**, 847 (1969).
[8]R.E. Watson, A.J. Freeman, and S. Koide, Phys. Rev. **186**, 625 (1969).
[9]B. Caroli, J. Phys. F **5**, 1399 (1975).
[10]D. Davidov, K. Maki, R. Orbach, C. Rettori, and E.P. Chock, Solid State Commun. **12**, 621 (1973), and D. Davidov, C. Rettori, R. Orbach, A. Dixon, E.P. Chock, Phys. Rev. B **11**, 3546 (1975).
[11]R.E. Walstedt and L.R. Walker, Phys. Rev. B **11**, 3280 (1975).
[12]U. Larsen, unpublished.
[13]U. Larsen, J. Appl. Phys. **49**, 1610 (1978); Solid State Commun. **27**, 943 (1978); and Phys. Rev. B **14**, 4356 (1976).
[14]A. Blandin, J. Appl. Phys. **39**, 1285 (1968).
[15]P.A. Lindgård, B.N. Harmon, and A.J. Freeman, Phys. Rev. Lett. **35**, 383 (1975).
[16]J.S. Schilling, P.J. Ford, U. Larsen, and J.A. Mydosh, Phys. Rev. B **14**, 4368 (1976).
[17]J.A. Mydosh, J. Magn. and Magn. Mat. **7**, 237 (1978).
[18]F. Zimmer, J.S. Schilling, J. Magn. and Magn. Mat. **9**, 37 (1978), and *Proceedings of the International Conference on High Pressure and Low Temperature Physics*, edited by C.W. Chu and J.A. Woollam (Plenum, New York, 1978); and U. Larsen, H. Penkert, and J.S. Schilling, Z. Phys. B **27** (1980).
[19]J.S. Schilling, Adv. Phys., **28**, 657 (1979).
[20]U. Larsen, Solid State Commun. **22**, 311 (1977); Phys. Rev. B **18**, 5014 (1978).
[21]U. Hardebusch, W. Gerhardt, and J. S Schilling, Phys. Rev. Lett. **44**, 352 (1980).
[22]J. Lindhard, Dan. Mat. Fys. Medd. **28**, no. 8 (1954).

Symmetries and metrics of homogeneous cosmologies

M. A. Melvin
Temple University, Philadelphia, Pennsylvania 19122
T. R. Michalik
Randolph–Macon Woman's College, Lynchburg, Virginia 24503

(Received 19 July 1979; accepted for publication 21 November 1979)

It is shown that an assumption of "historical homogeneity"—or the persistence of spatial homogeneity in time—leads to the synchronous (time-orthogonal) form $\mathbf{d}s^2 = -\mathbf{d}t^2 + \gamma_{AB}(t)\epsilon^A \otimes \epsilon^B$, of the metrics for general spatially homogeneous (Bianchi) cosmologies. The ϵ^A are group-representation 1-forms associated with the spatial symmetry group of the given cosmology, and are independent of any particular space–time theory; in particular they do not depend on Einstein's equations and the local physics. Expressions for the ϵ^A in the canonical basis used by Estabrook, Wahlquist, and Behr (EWB) have been tabulated in the companion paper on "Spatially homogeneous neutrino cosmologies" (SHNC). In the present paper explicit expressions are given for the scale factors $\gamma_{AB}(t)$. These do correspond to the choice of the canonical basis to describe the symmetry but, in contrast to the ϵ^A, they also depend on the space–time physical theory. Here the $\gamma_{AB}(t)$ come out as observable geometric–kinematic quantities (generalized Hubble constants) which also appear in the solutions of Einstein's equations for a given cosmology. Central to the discussion is the matrix $\mathbf{C}(t)$ which relates the time-independent canonical basis describing the symmetry to a time-dependent orthonormalizing basis. With the imposed requirement that the Riemannian geometry of the evolving spatial hypersurfaces retain its quasi-canonical form in time ("quasi-canonical gauge") it is shown that $\mathbf{C}(t)$ is a product of a diagonal matrix $\dot{\mathbf{D}}$ and a rotation matrix $\mathbf{R}$; and a table of these forms of $\mathbf{C}(t)$ for the various Bianchi types is given. The simple Hubble-constant expressions for the metric scale factors $\gamma_{AB}(t)$ for all types come solely from the diagonal factor $\dot{\mathbf{D}}$ in $\mathbf{C}$. The metric, then, in all cases takes the form $\mathbf{d}s^2 = -\mathbf{d}t^2 + \gamma_{AB}\epsilon^A \otimes \epsilon^B = -\mathbf{d}t^2 + (\epsilon^{\mathrm{I}}/\alpha)^2 + (\epsilon^{\mathrm{II}}/\beta)^2 + (\epsilon^{\mathrm{III}}/\gamma)^2$. Thus, the specializing assumption of Misner and of Ryan and Shepley, requiring $\mathbf{C}$ to be symmetric, does not limit the form of the metric but does restrict the relation between the invariant symmetry basis and the quasi-canonical orthonormal basis to the cases where the orthogonal factor $\mathbf{R}$ reduces to the identity, i.e., to those cases where the two bases remain parallel at all times. Thus, the gauge assumed by them is not compatible with the preservation of the quasi-canonical form of the spatial geometry in almost all cases where there is degeneracy involved in the spatial curvatures.

INTRODUCTION AND SUMMARY

As is well known, requiring (a) *spatial homogeneity* and (b) *spatial isotropy* of spacetime means that it is dissectable into spacelike hypersurfaces on each of which there acts a six-parameter Lie group which has (a) a three-parameter simply transitive subgroup and (b) a three-parameter isotropy subgroup. These two requirements lead to certain restricted forms (Robertson–Walker) of metrics for spacetime. Except for a time-dependent scale factor in front of the spatial part of the metric, these forms are independent of time and of the content of spacetime, i.e., of the local physics embodied in the specific structure of the stress-energy tensor in Einstein's equations.

In accord with what one may call "the modest Copernican principle," that there is no privileged spatial *location* in the universe—"on the average things everywhere are the same"—one should require only (a) *spatial homogeneity*, and not necessarily isotropy. Spatial homogeneity alone (three-parameter simply transitive Lie group) leads to restrictions on the metric, substantially broader than those imposed by Robertson–Walker. By subsuming under the concept of spatial homogeneity not only (instantaneous) "geographic" homogeneity but also "historical" homogeneity—i.e., the requirement that there exists a global time parameter t with respect to which history evolves the same at every point of space—we can show that the broader forms of metrics for spatially homogeneous cosmologies (Bianchi types) have the general form

$$\mathbf{d}s^2 = -\mathbf{d}t^2 + \mathbf{d}\sigma^2,$$
$$\mathbf{d}\sigma^2 = \tilde{C}_A{}^d\epsilon^A \otimes \tilde{C}_B{}^d\epsilon^B \equiv \gamma_{AB}(t)\epsilon^A \otimes \epsilon^B.$$

The ϵ^A are invariant-basis 1-forms which are determined for each Bianchi type purely group theoretically from the symmetry group (structure constants $C^A{}_{BC}$) of the homogeneous space of that type. The ϵ^A and the structure constants $C^A{}_{BC}$ are tensors with respect to $\mathrm{GL}(3,\mathbb{R})$ transformations constant over the entire space. Various representations of the ϵ^A depending upon which invariant basis of the Lie algebra is chosen have appeared in the literature. In this and the companion paper (SHNC) we have adopted the canonical basis

 0022-2488/80/071938-14$1.00

of Estabrook, Wahlquist, and Behr (EWB) for the Bianchi Lie algebras, and in SHNC we have given the associated invariant basis forms ϵ^A. We establish the canonical basis of the Lie algebras as follows: By analyzing the two lower indices of $C^k{}_{nl}$ into one upper index with the help of the Levi-Civita tensor density, we define a *structure-constant matrix* A^{km} (tensor density of weight $w = -1$) which represents all the structure constants $C^k{}_{nl}$. Using the freedom allowed by GL(3,ℝ) transformations, A^{km} is put into canonical forms A^{KM}, for each type, in which only three independent elements appear. In the canonical forms the *symmetric structure-constant matrix* $A^{(KM)}$ is diagonal, with elements ± 1 or 0; the *antisymmetric structure matrix* $A^{[KM]}$ has at most only one nonvanishing element $A^{[I,II]} \equiv n$. These serve to classify the Bianchi types.

In all but two type categories—VI and VII—n may be made equal to 0 or 1, in each case defining a single type. In categories VI_h and VII_h an infinite number of types corresponding to every value of $\pm n^2$ $(0 \leqslant n \leqslant \infty)$ occurs:

$$h \equiv -n^2 \text{ (type } VI_h), \quad h \equiv +n^2 \text{ (type } VII_h).$$

The problem of canonical metric forms is now evident. The ϵ^A 1-forms and their dual vector fields $\mathbf{e}_A$, though dependent upon the spatial symmetry and the choice of basis of the Lie algebra, are independent of any particular spacetime theory; in particular, they do not depend on Einstein's equations and the local physics. In contrast, the $\gamma_{AB}(t)$ do depend on the spacetime theory and the local physics. In the present paper explicit expressions are given for these "generalized scale factors" $\gamma_{AB}(t)$ in terms of observable geometric–kinematic quantities which also appear in the solutions of Einstein's equations.

The $\widetilde{C}_A{}^d$ are time-dependent transformation coefficients relating the invariant-basis triad to a geometrically defined metric—orthonormalizing time-varying triad ("orthonormal triad") $\mathbf{l}_d$ of which the basis forms are $\lambda^d = \widetilde{C}_A{}^d \epsilon^A$. The $\widetilde{C}_A{}^d(t)$ and $\widetilde{C}_B{}^d(t)$ and their scalar product $\gamma_{AB}(t) = \widetilde{C}_A{}^d(t)\widetilde{C}_B{}^d(t)$ depend on the appropriate solution of Einstein's equations in the following manner:

Spatial homogeneity for all times implies "historical homogeneity," i.e., on a cosmological scale, history—and therefore the measure of time—is invariant under the three spatial invariant translation fields $\mathbf{e}_A$; the latter in turn are related to the orthonormal frame fields $\mathbf{l}_d$ by the reciprocals $C_d{}^A$ of the time-dependent transformation coefficients $\widetilde{C}_A{}^d$. From historical homogeneity we arrive at first order linear differential equations in the time, with the structure

$$\dot{\widetilde{C}}_D{}^a = \widetilde{C}_D{}^b(S + W)_b{}^a.$$

Here $S_b{}^a$ are the orthonormal-triad components of the rate-of-strain tensor or, geometrically, the extrinsic curvature of the hypersurface of homogeneity; and $W_b{}^a = -W_a{}^b$ are the (three independent) orthonormal-triad components of the angular velocity of the orthonormal triad relative to a Fermi–Walker transported triad, i.e., the angular velocity of the orthonormal reference triad with respect to a gyroscopically stabilized inertially guided reference frame. The choice of

$$W_{23} \equiv -\omega_1 \quad W_{31} \equiv -\omega_2 \quad W_{12} \equiv -\omega_3$$

is only partially free in some of the types; in other words, for some Bianchi type solutions there is an inescapable rotation of the reference triad with respect to the "inertial compass."

The orthonormalizing transformation matrix $\widetilde{\mathbf{C}}$ is a product of a diagonal matrix $\grave{\mathbf{D}}$ and a rotation matrix $\mathbf{R}$. (The accent above a letter indicates a diagonal matrix.) It is nonsingular. From this, and the antisymmetry of $\mathbf{W}$, there results the *Double-conjugate-pair theorem* which implies at most block diagonal structure of $\widetilde{\mathbf{C}}$ in almost all cases if the quasi-canonical (diagonal) form of the time-evolved symmetric structure matrix upon which the spatial curvature depends

$$\grave{\mathbf{F}}(t) \equiv \Delta\, \widetilde{\mathbf{C}}^T \grave{\mathbf{F}}_0 \widetilde{\mathbf{C}}$$

$[\widetilde{\mathbf{C}} \equiv \mathbf{C}^{-1}$; $\Delta \equiv \det\mathbf{C}$; $\grave{\mathbf{F}}_0$ canonical (diagonal) form of $A^{(KM)}]$ is to be preserved.

A table of the forms of $\widetilde{\mathbf{C}}$ and simple Hubble constant expressions for the metric scale factors for the various Bianchi types are given. The metric scale factors come solely from the diagonal factor in $\widetilde{\mathbf{C}}$, a fact which has tended to obscure the incompleteness of certain procedures in the literature. Thus, the specializing assumption on the form of $\widetilde{\mathbf{C}}$, made by Misner, and Ryan and Shepley, respectively, does not limit the form of the metric but does restrict the relation between the invariant basis and quasi-canonical orthonormal basis to the cases where the orthogonal factor $\mathbf{R}$ reduces to the identity, i.e., to those cases where the two bases remain parallel at all times. This gauge assumed by them is not compatible with the preservation of the diagonal form for $\mathbf{F}(t)$ in almost all cases where $\mathbf{F}$ has degenerate eigenvalues.

A. Definition of spatially homogeneous spacetime, establishment of the cosmic time-orthogonal form of the metric

Spacetime is assumed to have a (pseudo-) Riemannian structure with some (as yet unspecified) metric $\mathbf{d}s^2$ of signature (3,1). Such a spacetime is defined to be spatially homogeneous if *all* of the spacetime may be foliated into a one-parameter family of spacelike 3-surfaces—hypersurfaces—which are totally homogeneous; in other words, there exists a linearly ordered splitting of spacetime into 3-spaces which are (1) instantaneously homogeneous, and (2) historically homogeneous. A parameter for the linear order is called a "cosmic time coordinate t", and a convenient admissible choice for t will be described shortly. We must first, however, make precise the key phrases "instantaneously homogeneous" and "historically homogeneous" which together characterize "total" spatial homogeneity.

(1) *Instantaneous homogeneity*: "*Geography does not matter*": The first expression means, in accord with what we may call the "modest Copernican principle" (modest because we are assuming only homogeneity, not necessarily isotropy), that in each of the hypersurfaces—space at any one time—all spatial locations are equivalent with respect to all relevant properties. In other words, there exist within each hypersurface three linearly independent vector fields such that all well-defined scalars, vectors, tensors, etc., are unchanged under transport (Lie differentiation giving zero) along the directions of these vectors fields. In particular,

they leave the metric tensor $\mathbf{ds}^2$ within each hypersurface invariant; such isometry-generating fields are of course called "Killing-vector fields."

In the companion paper[1] (Spatially Homogeneous Neutrino Cosmologies–SHNC) and in this paper, we introduce a canonicalized (EWB) basis. In this basis the Killing vector fields are labeled $\mathbf{z}_r$ $(r = 1,2,3)$, and their dual Killing forms ζ^r [vector-, dyadic-, and exterior-form sets, without indices are boldface; all indices on boldface Latin or Greek letters are labels indicating the member of the triad or tetrad to which we are referring].

Prior to introducing the EWB basis—in an arbitrary basis—let the Killing vector fields be labeled $\mathbf{Z}_i$. Let the structure constants of the Lie algebra comprised of these Killing vector fields be $-\mathsf{G}^{\bar{k}}{}_{\bar{i}\bar{j}}$. Within the three-dimensional spacelike hypersurfaces we may choose a basis consisting of three linearly independent vector fields

$$\mathbf{E}_{\bar{j}} = a_{\bar{j}}{}^{\bar{s}}\mathbf{Z}_{\bar{s}}, \tag{1}$$

which are *invariant*, i.e., the Lie derivative—or commutator—of $\mathbf{E}_{\bar{j}}$ with respect to $\mathbf{Z}_{\bar{i}}$ equals zero:

$$[\mathbf{Z}_{\bar{i}},\mathbf{E}_{\bar{j}}] = 0. \tag{2}$$

Expanding this relation and rearranging, we have

$$(\mathbf{Z}_{\bar{i}}a_{\bar{j}}{}^{\bar{s}})\mathbf{Z}_{\bar{s}} = -a_{\bar{j}}{}^{\bar{s}}[\mathbf{Z}_{\bar{i}},\mathbf{Z}_{\bar{s}}] = a_{\bar{j}}{}^{\bar{s}}\mathsf{G}^{\bar{k}}{}_{\bar{i}\bar{s}}\mathbf{Z}_{\bar{k}}. \tag{3}$$

Applying this formula to the expansion of the commutator of $E_{\bar{i}}$ with $E_{\bar{j}}$ we find, after a cancellation,

$$[\mathbf{E}_{\bar{i}},\mathbf{E}_{\bar{j}}] = a_{\bar{i}}{}^{\bar{s}}a_{\bar{j}}{}^{\bar{t}}\mathsf{G}^{\bar{k}}{}_{\bar{s}\bar{t}}\mathbf{Z}_{\bar{k}} = \overline{\mathsf{G}}^{\bar{d}}{}_{\bar{i}\bar{j}}\mathbf{E}_{\bar{d}}, \tag{4}$$

where the $\overline{\mathsf{G}}$'s are transforms of the G's in the invariant basis. Now if we adopt for the set of first order differential equations (2) the initial condition at some point P_0:

$$\mathbf{E}_{\bar{i}}(P_0) = \mathbf{Z}_{\bar{i}}(P_0), \ \therefore\ a_{\bar{i}}{}^{\bar{s}}(P_0) = \delta_{\bar{i}}{}^{\bar{s}}, \tag{5}$$

we have that the commutation coefficients $\overline{\mathsf{G}}^{\bar{d}}{}_{\bar{i}\bar{j}}$ which result from the commutation (4) of the $\mathbf{E}$'s with each other at P_0 are the negatives of the structure constants $-\mathsf{G}^{\bar{d}}{}_{\bar{i}\bar{j}}$ in the Lie algebra of the Killing vector fields ("reciprocal relation"[2]). Since, however, the $\mathbf{E}$'s by their definition are *invariant* vector fields, the commutation coefficients of the $\mathbf{E}$'s must be independent of location within the hypersurface, and we have the reciprocal relation to the Killing structure constants holding everywhere:

$$[\mathbf{E}_{\bar{i}},\mathbf{E}_{\bar{j}}] = \mathsf{G}^{\bar{k}}{}_{\bar{i}\bar{j}}\mathbf{E}_{\bar{k}}. \tag{6}$$

It is to be noted that in general the Killing fields agree in value (same vectors) with the invariant fields only at a single point; the only exception, as may be seen in Appendix II of SHNC, is the Bianchi type I where the two fields are identical. In all other cases except for type IX, the point P_0 is at the origin of the parameter space, i.e., if vectors are specified by from one to three parameters x^1, x^2, x^3, P_0 is that point where all the x's which occur are equal to zero. In the exceptional case of type IX, P_0 is defined by $x^1 = \pi/2$, $x^2 = x^3 = 0$. These statements may be verified by substituting in the tables in Appendix II of SHNC. We also note that the invariant basis vectors may not in general be orthogonal (better: "metric-orthonormalizing"), i.e., they may not correspond to an orthogonal metric in which the coefficients $\gamma_{\bar{r}\bar{s}} = \mathbf{E}_{\bar{r}}\cdot\mathbf{E}_{\bar{s}}$ are given by $\delta_{\bar{r}\bar{s}}$.

(2) *Historical homogeneity*: *History is the same everywhere* (but it may matter, because all relevant properties, though independent of geography, are not necessarily the same "everywhen"). This second homogeneity requirement means then that the equivalence of location exists also with respect to development looking forward in time or looking backward in time. This property of historical homogeneity means that the three-dimensional isometry group generated by the three Killing vector fields acts not merely on the three-dimensional spaces but also on four-dimensional spacetime according to

$$[\mathbf{Z}_{\bar{i}},\mathbf{E}_{\bar{4}}] = 0, \quad \mathbf{E}_{\bar{4}} \equiv \partial_t, \tag{7}$$

where $\mathbf{E}_{\bar{4}}$ is the holonomic vector field (timelike congruence) associated with the time coordinate. The condition (7) expresses the principle of historical homogeneity in that it requires the time-coordinate vector field $\mathbf{E}_{\bar{4}}$ parameterizing the spatial hypersurfaces to have vanishing Lie derivatives (i.e., to be unchanged under translations) with respect to the Killing fields $\mathbf{Z}_{\bar{i}}$ (see Appendix). We note that since the two vector field sets, the three $\mathbf{Z}$ fields, and the three $\mathbf{E}$ fields each constitute a basis for the three-dimensional hypersurface in which they lie, the determinant of the transformation (1) is not zero, and we may write

$$\mathbf{Z}_{\bar{s}} = \bar{a}_{\bar{s}}{}^{\bar{k}}\mathbf{E}_{\bar{k}}, \quad a_{\bar{j}}{}^{\bar{s}}\bar{a}_{\bar{s}}{}^{\bar{k}} = \delta_{\bar{j}}{}^{\bar{k}}. \tag{8}$$

Because of the linear independence of the $\mathbf{E}$'s we may infer from Eq. (7) that

$$[\mathbf{E}_{\bar{i}},\mathbf{E}_{\bar{4}}] = 0. \tag{6'}$$

Equations (6) and (6') constitute a system of relations governing $\mathbf{E}_{\bar{\sigma}}$, $(\bar{\sigma} = 4, 1,2,3)$.

There remains a freedom to choose both direction and magnitude of $\mathbf{E}_{\bar{4}}$ (Ref. 3 and Ref. 4, pp. 106 and 147). Starting with the given basis $\mathbf{E}_{\bar{\sigma}}$ we choose another invariant basis $(\mathbf{e}_{1'},\mathbf{e}_{2'},\mathbf{e}_{3'},\mathbf{e}_{4'})$ which is a linear transform of the $\mathbf{E}$'s with time-dependent coefficients, but for which we require that the spatial vector fields lie solely in the 3-spaces:

$$\mathbf{e}_{i'} = b_{i'}{}^{\bar{s}}(t)\mathbf{E}_{\bar{s}}, \quad \mathbf{e}_{4'} = b_{4'}{}^{\bar{\sigma}}\,\mathbf{E}_{\bar{\sigma}} \tag{9}$$

and that the conditions (6) and (6') be maintained with the $\mathbf{e}$'s, with transforms of the G's which we shall now call $\mathsf{G}^{k'}{}_{i'j'}$. It turns out that these requirements can be met with the $\mathsf{G}^{k'}{}_{i'j'}$, again time-independent structure coefficients (constants),[3,4] which justifies our calling the new basis, $(\mathbf{e}_{4'},\mathbf{e}_{i'})$ "invariant." A convenient choice for the time direction, which goes with the possibility of synchronizing clocks by light rays throughout each spatial hypersurface, is to define $\mathbf{e}_{4'}$ to be "orthogonal" to the hypersurfaces, i.e., such that the time–space cross terms in the metric vanish[5] (see Appendix).

The vector fields $\mathbf{e}_{i'}$ have as their duals three invariant 1-forms $\boldsymbol{\epsilon}^{i'}$ which constitute a form basis, and there are bilinear combinations of the invariant basis forms, the $\boldsymbol{\epsilon}$'s, in terms of which the spatial metric may be expressed

$$\mathbf{d}\sigma^2 = \sum_{d=1}^{3} \widetilde{C}_{f'}{}^{d}\boldsymbol{\epsilon}^{f'} \otimes \widetilde{C}_{g'}{}^{d}\boldsymbol{\epsilon}^{g'} = \gamma_{f'g'}(t)\boldsymbol{\epsilon}^{f'} \otimes \boldsymbol{\epsilon}^{g'}. \tag{10}$$

By their definition as invariant forms under the symmetry group which has been prescribed for all times, the $\boldsymbol{\epsilon}$'s cannot

involve the time, but the C's can. We note that the spatial metric goes into orthonormal forms by representing it in terms of (generally independent) orthonormalizing basis forms $\lambda^{d} = \widetilde{C}_{f'}{}^{d}\epsilon^{f'}$. We also introduce a fourth basis form $\epsilon^{4'} \equiv dt \equiv \lambda^{4}$ dual to $\mathbf{e}_{4'}$. Then the metric takes the form

$$\begin{aligned} ds^2 &= -dt^2 + d\sigma^2 = -\epsilon^{4'} \otimes \epsilon^{4'} + \gamma_{i'j'}(t)\epsilon^{i'} \otimes \epsilon^{j'} \\ &= -(\lambda^4)^2 + (\lambda^1)^2 + (\lambda^2)^2 + (\lambda^3)^2 \end{aligned} \tag{10''}$$

with no cross terms involving $\epsilon^{4'} \otimes \epsilon^{i'}$. Thus, we have a global "hypersurface-orthogonal" and "synchronous" metric in which clocks at all points of space are synchronized at each time.

The matrix of spatial metric scale coefficients in Eq. (10)

$$\gamma_{i'j'}(t) = \widetilde{C}_{i'}{}^{d}(t)\widetilde{C}_{j'}{}^{d}(t) \quad \text{or} \quad \gamma(t) = \widetilde{\mathbf{C}}(t)\widetilde{\mathbf{C}}(t)^{\text{transpose}} \tag{11}$$

is one of functions of time alone; in the fully isotropic cases[6] the matrix reduces to the unit matrix multiplied by a single overall scale factor—a function of time alone.

Quite generally, for any spatially homogeneous cosmology, the spatial metric $d\sigma^2$ belongs to one of two categories:

General: Either $d\sigma^2$ is such that its isometry group is one of the nine types of simply transitive three-parameter isometry groups classified by Bianchi[7];

Particular: or else $d\sigma^2$ is a Kantowski–Sachs metric[8] where there is a four-parameter isometry group whose three-parameter subgroup does not act transitively on 3-surfaces, but acts transitively on 2-spheres. The particular Kantowski–Sachs case has been treated in Ref. 8.

We shall discuss here only the general category of the nine Bianchi types. By a purely spatial transformation from any original choice of group generators, the structure constants $-\mathsf{G}^{d'}{}_{g'h'}$ of the isometry-group Lie algebras—i.e., Killing vector algebras—of the nine types may be put in a canonical arrangement

$$\mathsf{A}^{d'f'} \equiv \tfrac{1}{2}\mathsf{G}^{d'}{}_{g'h'}\epsilon^{f'g'h'} \sim \begin{pmatrix} \mathsf{G}^{1'}{}_{2'3'} & \mathsf{G}^{1'}{}_{3'1'} & \mathsf{G}^{1'}{}_{1'2'} \\ \mathsf{G}^{2'}{}_{2'3'} & \mathsf{G}^{2'}{}_{3'1'} & \mathsf{G}^{2'}{}_{1'2'} \\ \mathsf{G}^{3'}{}_{2'3'} & \mathsf{G}^{3'}{}_{3'1'} & \mathsf{G}^{3'}{}_{1'2'} \end{pmatrix}$$

$$\Rightarrow \mathsf{A}^{DF} \sim \begin{pmatrix} \mathsf{G}^{\mathrm{I}}{}_{\mathrm{II\,III}} & \mathsf{G}^{\mathrm{I}}{}_{\mathrm{III\,I}} & 0 \\ \mathsf{G}^{\mathrm{II}}{}_{\mathrm{II\,III}} & \mathsf{G}^{\mathrm{II}}{}_{\mathrm{III\,I}} & 0 \\ 0 & 0 & \mathsf{G}^{\mathrm{III}}{}_{\mathrm{I\,II}} \end{pmatrix}$$

$$\equiv \begin{pmatrix} a_0 & n_0 & 0 \\ -n_0 & b_0 & 0 \\ 0 & 0 & c_0 \end{pmatrix} \quad (a_0 = 1,0 \;\; b_0, c_0 = \pm 1, 0)$$

$$(n_0 c_0 = 0) \tag{12}$$

used by Estabrook, Wahlquist, and Behr[9] whose notation we follow: ($\mathsf{G}^{\mathrm{II}}{}_{\mathrm{II\,III}} = -\mathsf{G}^{\mathrm{I}}{}_{\mathrm{III\,I}}$ and $\mathsf{G}^{\mathrm{II}}{}_{\mathrm{II\,III}}\mathsf{G}^{\mathrm{III}}{}_{\mathrm{I\,II}} = 0$ by the Jacobi identity). Normalizing a_0 to 0 or $+1$, and b_0 and c_0 to 0 or ± 1, and listing all possible cases leads to Tables II and III of the Bianchi–Behr types which we will discuss in detail in the sequel. The deeper meaning of the Behr parameter

$$h \equiv n_0^2/a_0 b_0 \tag{13}$$

will become clear in our later development.

B. Canonicalization of the structure constant matrix for the spatially homogeneous symmetry types

We have not found the completely explicit procedure for the canonicalization in Eq. (12) in the available literature; on the basis of references we surmise that it is contained in the unpublished Hamburg thesis of C.G. Behr (1962). Since it brings out some subtle points, and is very useful for the subsequent analysis, we find it desirable to develop in detail.

The way in which the canonicalization in Eq. (12) is accomplished is as follows: The square "structure constant matrix"

$$\mathsf{A}^{k'm'} \equiv \tfrac{1}{2}\mathsf{G}^{k'}{}_{g'h'}\epsilon^{m'g'h'}$$

can be written as the sum of a symmetric structure constant matrix

$$\mathsf{A}^{(k'm')} \equiv \mathsf{G}^{(k'}{}_{u'v'}\epsilon^{m')u'v'} \tag{14}$$

and a skew-symmetric structure constant matrix

$$\mathsf{A}^{[k'm']} \equiv \mathsf{G}^{[k'}{}_{u'v'}\epsilon^{m']u'v'}. \tag{15}$$

We may term the dual of $\mathsf{A}^{[k'm']}$ the "Bianchi vector":

$$\begin{aligned} n_{i'} &\equiv \tfrac{1}{2}\epsilon_{i'k'm'}\mathsf{A}^{[k'm']} = \tfrac{1}{2}\epsilon_{i'k'm'}\mathsf{G}^{[k'}{}_{g'h'}\epsilon^{m']g'h'} \\ &= \mathsf{G}^{k'}{}_{i'k'} \end{aligned} \tag{16}$$

and we have

$$\mathsf{A}^{[p'q']} = \epsilon^{j'p'q'}n_{j'}. \tag{16'}$$

We note also that the Jacobi identity—condition for associativity in the overlying group—telescopes to $\mathsf{A}^{k'm'}n_{m'} = \mathsf{A}^{(k'm')}n_{m'} = 0$.

That the appelation "Bianchi vector" is justified, i.e., that the $n_{j'}$ transform as covariant vector components under a general nonsingular linear transformation (space independent)

$$\begin{aligned} &\mathbf{e}_{f'} = G^{g}_{f'}\mathbf{e}_g, \quad \epsilon^{f'} = \epsilon^{h}G^{f'}_{h}, \\ &\qquad\qquad G^{f'}_{h}G^{g}_{f'} = \delta_h{}^g \\ &\mathbf{e}_g = G^{f'}_{g}\mathbf{e}_{f'}, \quad \epsilon^{h} = \epsilon^{f'}G^{h}_{f'}, \end{aligned} \tag{17}$$

of the invariant basis triads, may be seen from the tensor transformation behavior of the structure-constant array and from the fact that the $n_{j'}$ constitute the vector trace of this 3-index array:

$$\mathsf{G}^{k}_{gh} = \langle \epsilon^k, [\mathbf{e}_g, \mathbf{e}_h] \rangle = G^{k}_{u'}G^{v'}_{g}G^{w'}_{h}\mathsf{G}^{u'}{}_{v'w'}. \tag{18}$$

Setting the index $k = h$, we have

$$n_{g'} = G^{v}_{g'}n_v. \tag{19}$$

Corresponding, $\mathsf{A}^{[k'm']}$ and $\mathsf{A}^{(k'm')}$ transform as contravariant 2-index relative tensor components of weight -1 as may be verified from the expression derived from Eq. (18)

$$\mathsf{G}^{k}{}_{gh}\epsilon^{mgh} = \mathscr{D}G^{k}_{u'}G^{m}_{s'}\mathsf{G}^{u'}_{v'w'}\epsilon^{s'v'w'} \tag{20}$$

or

$$\mathsf{A}^{[km]} = \mathscr{D}G^{k}_{u'}\mathsf{A}^{[u's']} \tag{20a}$$

and

$$\mathsf{A}^{(km)} = \mathscr{D}G^{k}_{u'}G^{m}_{s'}\mathsf{A}^{(u's')}. \tag{20b}$$

where $\mathscr{D} \equiv \det G_p{}^{q'}$.

With the "Bianchi vector" $\mathbf{n}$ recognized as an invariant irreducible part of $C^{k}{}_{gh}$ there are two general classes of possibilities: Either $\mathbf{n}$ vanishes, in which case we have the "non-vector types" (called "class A" by Ellis and MacCallum[10]), or $\mathbf{n}$ does not vanish, in which case we have the "vector types" ("class B" in Ref. 10).

In all types where $A^{(km)}$ exists, i.e., in all types but I and V, reduction of $\mathbf{e}_{k'}A^{(k'm')}\mathbf{e}_{m'}$ to diagonal form while simultaneously orthonormalizing the positive definite metric form

$$d\sigma^2 = \epsilon^{i'}\gamma_{i'j'}\epsilon^{j'},$$

which transforms contragrediently relative to $A^{(km)}$, is something of a variation on the standard problem of simultaneously reducing two cogredient quadratic forms—the positive definite one becoming a sum of squares with unit coefficients and the other one becoming diagonal. In our case the canonical form can be attained in three stages as follows:

(1) A *combined rotation and pure strain*, e.g., $G_{h'}{}^{f'}$, on the form basis [Eq. (17)] transforms the metric to a sum of squares (orthonormal quadratic form with exterior form basis $\lambda^{1'}$, $\lambda^{2'}$, $\lambda^{3'}$); the matrix is transformed to the unit matrix:

$$\epsilon^{i'}\gamma_{i'j'}\epsilon^{j'} = \lambda^{h'}G_{h'}{}^{i'}\gamma_{i'j'}G_{k'}{}^{j'}\lambda^{k'} = \lambda^{k'} = \lambda^{h'}\delta_{h'k'}\lambda^{k'}. \quad \text{(Stage 1)}$$

Simultaneously, the dual *vector basis* $\mathbf{e}_{1'}$, $\mathbf{e}_{2'}$, $\mathbf{e}_{3'}$ undergoes the contragredient transformation [$G^{g'}{}_{f'}$ in Eq. (17)], and $A^{(km)}$ transforms contragrediently relatively to γ_{ij} [with the additional weight factor $\mathscr{D}$]. The form basis $\lambda^{1'}$, $\lambda^{2'}$, $\lambda^{3'}$, (and the dual vector basis: $\mathbf{l}_{1'}$, $\mathbf{l}_{2'}$, and $\mathbf{l}_{3'}$) are *metric-orthonormalizing*, but not yet aligned with the principal directions of the symmetric structure matrix.

(2) The preceding transformation, which has put the metric in orthonormal form, is now followed by a *pure rotation*

$$\mathbf{l}_{\bar{k}} = R_{\bar{k}}{}^{h'}\mathbf{l}_{h'}, \quad \mathbf{l}_{h'} = R_{h'}{}^{\bar{k}}\mathbf{l}_{\bar{k}}, \quad \text{(Stage 2)}$$

to the basis $\mathbf{l}_{\bar{1}}$, $\mathbf{l}_{\bar{2}}$, $\mathbf{l}_{\bar{3}}$ defined by the principal axes $\bar{a}$, $\bar{b}$, $\bar{c}$, of $A^{(km)}$. The symmetric structure matrix then assumes the diagonal form

$$A^{(\bar{k}\bar{m})} \equiv A(\bar{k})\delta^{\bar{k}\bar{m}} = \mathrm{diag}(\bar{a}\bar{b}\bar{c}).$$

For convenience we shall designate this form of $A^{(km)}$, and the corresponding basis, *quasi-canonical*. The contragrediently transforming metric remains orthonormal under this orthogonal transformation of the basis from $\lambda^{1'}$, $\lambda^{2'}$, $\lambda^{3'}$ to $\lambda^{\bar{1}}$, $\lambda^{\bar{2}}$, $\lambda^{\bar{3}}$.

The reduction to quasi-canonical form is of course not unique. By Sylvester's Law of Inertia for quadratic forms the numbers of positive, negative and zero elements in the diagonalized $A^{(km)}$ are respectively invariant whatever transformations are performed on the basis to reach or maintain a diagonal form of $A^{(km)}$—*and this will be true whether or not the transformation coefficients are time dependent.* The possible invariant types are thus immediately deducible and are listed in Table I. The classification is standarized by taking the signatures to be nonnegative, always, and the $A^{(11)}$ = a nonnegative.

(3) Though, after the first two stages, $A^{(km)}$ is diagonal (quasi-canonical) it is not yet in the canonical EWB form in which only $\pm$ 1's and 0's appear on the diagonal. This form which we shall write

$$A^{(KM)} \equiv A^{(K)}\delta^{KM} \equiv \mathrm{diag}(a_0 b_0 c_0), \quad \text{(Stage 3)}$$

is attained by a *final pure strain transformation* along the principal axes of $A^{(km)}$ so that the nonvanishing diagonal elements are normalized to $\pm$ 1. At the same time *the metric remains diagonal but no longer orthonormal.* We emphasize the point since the relation we infer for this final canonicalizing transformation will be useful subsequently, interpreted in a reverse manner, i.e., when, with a time-dependent transformation, we undo the canonicalization of $A^{(KM)}$—leaving it in diagonal form—while restoring the normalization of the diagonal $d\sigma^2$.

The final transformation—of the nature of a pure strain—relates the metric-orthonormalizing frame $\mathbf{l}_{\bar{1}}$, $\mathbf{l}_{\bar{2}}$, $\mathbf{l}_{\bar{3}}$ to the $A^{(KM)}$—canonicalizing frame $\mathbf{e}_{\mathrm{I}}$, $\mathbf{e}_{\mathrm{II}}$, $\mathbf{e}_{\mathrm{III}}$:

$$\mathbf{l}_{\bar{1}} = G_{\bar{1}}{}^{\mathrm{I}}\mathbf{e}_{\mathrm{I}} \equiv \alpha\mathbf{e}_{\mathrm{I}}, \quad \mathbf{l}_{\bar{2}} = G_{\bar{2}}{}^{\mathrm{II}}\mathbf{e}_{\mathrm{II}} \equiv \beta\mathbf{e}_{\mathrm{II}}, \quad \mathbf{l}_{\bar{3}} = G_{\bar{3}}{}^{\mathrm{III}}\mathbf{e}_{\mathrm{III}} \equiv \gamma\mathbf{e}_{\mathrm{III}}, \quad (21)$$

TABLE I. The possible types of simply transitive 3-parameter isometry groups which act transitively on 3-spaces—the Bianchi–Behr types. The classification comes from all possible ways of satisfying the matrix-algebraic inequalities for $A^{(km)}$:

rank $-2\times$ index $\leqslant 0 \leqslant$ rank ("index" $\equiv$ number of positive coefficients).

The left-hand inequality follows from our convention that the signature must be nonnegative and from the algebraic identity

signature $\equiv$ index minus number of negative coefficients $= 2\times$ index minus rank.

Rank of $A^{(km)}$	Index of $A^{(km)}$	n^2	SIGN a	SIGN b	SIGN c	Bianchi–Behr Type
0	0	0	0	0	0	I
1	1	0	+	0	0	II
2	1	0	+	−	0	VI_0
2	2	0	+	+	0	$VIII_0$
3	2	0	+	+	−	VIII
3	3	0	+	+	+	IX
0	0	+	0	0	0	V
1	1	+	+	0	0	IV
2	1	+	+	−	0	III
2	1	+	+	−	0	VI_0
2	2	+	+	+	0	$VIII_0$

TABLE II. Nonvector Bianchi–Behr Types. The requirements listed in the last column are those which must be satisfied by the pure strain transformation needed to go *from a quasi-canonical basis* in which the metric is orthonormal, and the symmetric structure constant matrix is diagonal, *to a canonical basis* in which that matrix has diagonal elements 0 or ± 1.

Dimension of the derived Lie algebra = rank of $A^{(km)}$	a_0	b_0	c_0	Bianchi–Behr Type	Nonvacuous requirements on pure strain reduction coefficients
0	0	0	0	I	none
1	1	0	0	II	$\beta/\alpha = \bar{a}/\gamma$
2	1	−1	0	VI_0	$\beta/\alpha = (-\bar{a}/\bar{b})^{1/2} \quad \gamma = (-\bar{a}\bar{b})^{1/2}$
2	1	1	0	VII_0	$\beta/\alpha = (\bar{a}/\bar{b})^{1/2} \quad \gamma = (\bar{a}\bar{b})^{1/2}$
3	1	1	−1	VIII	$\alpha = (-\bar{b}\bar{c})^{1/2} \quad \beta = (-\bar{c}\bar{a})^{1/2} \quad \gamma = (\bar{a}\bar{b})^{1/2}$
3	1	1	1	IX	$\alpha = (\bar{b}\bar{c})^{1/2} \quad \beta = (\bar{c}\bar{a})^{1/2} \quad \gamma = (\bar{a}\bar{b})^{1/2}$

where we have introduced convenient abbreviations for the pure strain components. Because $A^{(km)}$ is a relative tensor the transformation involves the determinant

$$\Delta \equiv G_{\bar{1}}{}^{I} G_{\bar{2}}{}^{II} G_{\bar{3}}{}^{III} \equiv \alpha\beta\gamma. \tag{22}$$

The transformation coefficients satisfy the requirements

$$\bar{a} = \Delta\,(G_{I}{}^{\bar{1}})^2 a_0 = a_0\alpha^{-1}\beta\gamma, \tag{23}$$

$$\bar{b} = \Delta\,(G_{II}{}^{\bar{2}})^2 b_0 = b_0\beta^{-1}\gamma\alpha, \tag{24}$$

$$\bar{c} = \Delta\,(G_{III}{}^{\bar{3}})^2 c_0 = c_0\gamma^{-1}\alpha\beta, \tag{25}$$

in those types where these equations are not vacuous because of the vanishing of $\bar{a}$, $\bar{b}$, or $\bar{c}$ (and a_0, b_0, or c_0).

Quite generally, for both vector and nonvector cases, the reduction is to a canonical form in which $\bar{a} \to a_0 = 0$ or 1, and $\bar{b} \to b_0 = 0$, -1, or $+1$ (likewise with the values of c_0 when $n_0 = 0$). In all cases reduction is accomplished by a rotation-strain which we need not deal with explicitly followed by a rotation and finally a pure strain transformation. The latter is in the form of Eq. (21), with conditions (23) and (24), and—for nonvector types—Eq. (25). For vector types Eq. (25) is vacuous, and instead we have the condition

$$\gamma = \bar{n}/n_0. \tag{26}$$

We consider first the nonvector class. As a consequence of its transformation behavior, $A^{(km)}$, which is now totally symmetric, may be reduced by the sequence of nonsingular linear transformations indicated above to canonical form $\mathrm{diag}(A_1,A_2,A_3) \equiv \mathrm{diag}(a_0,b_0,c_0)$, ($a_0 = 0,1$; b_0, $c_0 = 0, \pm 1$). The possible cases are listed in Table II. The last column in Table II corresponds to the applicability of Eqs. (23) to (25). In Type I all the equations are vacuous. In Type II only Eq. (23) applies, in Type VI_0 and VII_0 only Eqs. (23) and (24) apply, and consequently only the ratio

$$\frac{\beta}{\alpha} = \left|\frac{\bar{a}}{\bar{b}}\right|^{1/2}, \quad \text{and the coefficient} \quad \gamma = |\bar{a}\bar{b}|^{1/2}, \tag{27}$$

are uniquely determined. Finally in Types VIII and IX all three coefficients are determined. In all these types the dimension of the derived Lie algebra is the same as the rank of $A^{(km)}$.

We now consider the vector class. Upon multiplying Eqs. (23) and (24) together and substituting from Eq. (26) we find that the combination

$$\frac{n_0^2}{a_0 b_0} = \frac{\bar{n}^2}{\bar{a}\bar{b}} \equiv h \quad \text{(vector types)} \tag{28}$$

is an invariant under transformation. Each value of h will characterize a different type. There are thus $2\times\infty + 3$ types, with Type III a singled out special case of VI_h ($h = -1$). The transformation coefficients are restricted only by the Eq. (23) requirement

$$\beta/\alpha = (\bar{a}/a_0)(n_0/\bar{n}) \tag{29}$$

and that of Eq. (26). These are equivalent to Eqs. (27) but are stronger: Eq. (29) is nonvacuous in all vector types except V and Eq. (26) is nonvacuous in all vector types without exception. All types with the range $-\infty < h < 0$ are designated types VI_h. Those with range $0 < h < \infty$ are designated VII_h. Correspondingly for both types VI_h and VII_h

$$n_0 = (h a_0 b_0)^{1/2} = (h b_0)^{1/2} \tag{30}$$

takes values in the range $0+$ to ∞. The special case $h = -1$, $n_0 = 1$ is designated type III in the original Bianchi classification. Though it is not conventional, we may even regard types IV and V as limiting cases where, fixing $n_0 = 1$, by a suitable choice of $G_{3'}{}^{3}$ (Eq. 26), and letting $a_0 b_0 \to \pm 0$ we have

$$\lim h = \lim_{(a_0,b_0)\to \pm 0} 1/(a_0 b_0) = \pm\infty.$$

In this context, contrary to usual practice, we could introduce a unified notation

$$S^{n_0,h}_{a_0,b_0,c_0} \tag{31}$$

for all the Bianchi types—in particular the vector types; in the latter n_0 is not zero and c_0, which is equal to zero, need not be indicated. In summary we have the following possibilities for vector types given in Table III.

C. Conversion from the canonical invariant basis to a metric-orthonormalizing basis whose time dependence is governed by Einstein's equations

We have denoted the *invariant-basis* 1-forms in the canonical reduction by ϵ^A; they and their duals, the *invariant-basis* vector fields $\mathbf{e}_A$ corresponding to the canonical representation of the Lie algebra, satisfy

$$\langle \epsilon^A, \mathbf{e}_B \rangle = \delta^A{}_B. \tag{32}$$

TABLE III. Vector Bianchi–Behr types. The coefficients of the transformation effecting the reducton from quasicanonical to canonical frames in all vector cases except type V satisfy the equations

$$\frac{\alpha}{\beta} \equiv \frac{G_{II}{}^{\bar{2}}}{G_{I}{}^{\bar{1}}} = \frac{a_0\bar{n}}{\bar{a}n_0} = \frac{\bar{n}}{\bar{a}n_0}, \quad \gamma \equiv G_3{}^{III} = \frac{\bar{n}}{n_0}.$$

In type V only the equation for γ applies.

Dimension of derived Lie algebra = 1 + rank A	n_0	a_0	b_0	c_0	Limiting value or range of $h = n_0^2/a_0b_0$	Bianchi–Behr type	Our symbol $S^{n_0,h}_{a_0,b_0}$
2	1	1	0	0	$\pm\infty$	IV	$S^{1,\pm\infty}_{1,0}$
1	1	0	0	0	$\pm\infty$	V	$S^{1,\pm\infty}_{0,0}$
3	$n_0 = \lvert h\rvert^{1/2}$	1	−1	0	$-\infty < h < 0$	VI_h	$S^{n_0,-n_0^2}_{1,-1}$
3	1	1	−1	0	−1	III	$S^{1,-1}_{1,-1}$
3	$n_0 = \lvert h\rvert^{1/2}$	1	1	0	$0 < h < \infty$	VII_h	$S^{n_0,n_0^2}_{1,1}$

The invariant-basis fields are generally nonholonomic and, corresponding their dual forms are generally nonclosed; i.e., the commutators of the former and the exterior derivatives of the latter are not zero and define coefficients which are the structure constants of the isometry group. Explicitly, the invariance under the isometry group generated by the Killing vectors with structure constants $-\mathsf{G}^M{}_{JK}$ means that

$$[\mathbf{e}_J, \mathbf{e}_K] = \mathsf{G}^M{}_{JK}\mathbf{e}_M, \quad d\boldsymbol{\epsilon}^M = -\mathsf{G}^M{}_{JK}\boldsymbol{\epsilon}^J \wedge \boldsymbol{\epsilon}^K. \tag{33}$$

The right-hand relation here is the *Maurer–Cartan equation*. One can use orthogonal holonomic coordinate triads ∂_i and their dual forms $\mathbf{d}x^i$ as bases, and the $\mathbf{e}$'s and $\boldsymbol{\epsilon}$'s are determined purely group theoretically in terms of these. Explicit representations of the $\mathbf{e}$'s and the $\boldsymbol{\epsilon}$'s as functions of up to three variables in such a basis, corresponding to the canonical reduction of Estabrook, Wahlquist, and Behr[8] (and therefore different from Taub's[11]), are given in Appendix II of the companion paper "Spatially Homogeneous Neutrino Cosmologies (SHNC)."

As described also in SHNC, for greater ease in applying the basic Cartan structure formulas, it is advantageous, to make a real (generally time-dependent) affine transformation, $C_j{}^K(t)$, substituting instead of the invariant basis $\boldsymbol{\epsilon}$'s a metric-orthonormalizing basis of forms $\boldsymbol{\lambda}^i$ so that with

$$\boldsymbol{\lambda}^i = \boldsymbol{\epsilon}^J\widetilde{C}_J{}^i, \quad \boldsymbol{\epsilon}^J = \boldsymbol{\lambda}^k C_k{}^J \quad (C_g{}^J\widetilde{C}_J{}^d = \delta_g^d), \tag{34}$$

we have

$$\mathbf{d}\sigma^2 = (\boldsymbol{\lambda}^1)^2 + (\boldsymbol{\lambda}^2)^2 + (\boldsymbol{\lambda}^3)^2 = \gamma_{LN}C_k{}^L C_m{}^N \boldsymbol{\lambda}^k \otimes \boldsymbol{\lambda}^m. \tag{35}$$

Thus we have

$$\gamma_{LN}C_k{}^L C_m{}^N = \delta_{km} \quad \text{or} \quad \gamma_{LN}(t) = \delta_{km}\widetilde{C}_L{}^k\widetilde{C}_N{}^m \tag{36}$$

and the scale factors $\gamma_{LN}(t)$ for the Bianchi-type metrices are determined as soon as the transformation coefficients $\widetilde{C}_B{}^a(t)$, relating the time-dependent *metric-orthonormalizing triad* to the invariant basis triad are known. From Eq. (10) and from the invariant forms listed in Appendix II of SHNC the metric

$$\mathbf{d}\sigma^2 = \gamma_{KM}(t)\boldsymbol{\epsilon}^K \otimes \boldsymbol{\epsilon}^M \tag{37}$$

would then be fully determined as a function of time and three holonomic coordinate variables.

We can establish the relation between the canonical invariant basis $\mathbf{e}_{\mathrm{I}}$, $\mathbf{e}_{\mathrm{II}}$, $\mathbf{e}_{\mathrm{III}}$ and a general admissible metric-orthonormalizing basis $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$ in two stages by formally reversing the last two stages of the canonicalization procedure—i.e., first applying a pure strain described by α, β, γ (Functions of time whose relation to the post-strain values, $\bar{a}, \bar{b}, \bar{c}$, are indicated in the last columns of Tables II and III.)—and second performing a rotation. The admissible strain-rotations are determined by a set of conditions which relate to maintaining the canonical array (indicated in the upper left-hand corner of Table IV) for the time evolvent of the structure constant matrix in all cases where it does not vanish identically—these conditions together with the requirement that the metric takes the orthonormal form (35) determines the possible families of orthonormalizing bases for the different types. Technically, what is determined is an admissible subgroup of the full gauge (automorphism) group of the canonical structure matrix. The specific particular structure of the appropriate family in any given physical case is then determined by solving Einstein's equations on the level of the connection.

Specifically, how shall the transformation coefficients $\widetilde{C}_B{}^a(t)$ be determined? Corresponding to the orthonormal basis forms there are the dual orthonormal vector fields

$$\mathbf{l}_i = C_i{}^K\mathbf{e}_K, \quad \mathbf{e}_K = \widetilde{C}_K{}^m\mathbf{l}_m, \tag{37'}$$

which define an orthonormal triad at every point of space and at different times. While independent of position in space within each hypersurface, these triads are functions of time: Upon going from the invariant frame to the orthonormal frame the invariant-basis *structure constants*, $\mathsf{G}^C{}_{AB}$ of any given isometry group are transformed into *time-dependent commutation* coefficients $\mathsf{C}^f{}_{gh}(t)$ for the metric orthonormalizing triad:

$$[\mathbf{l}_g, \mathbf{l}_h] = \mathsf{C}^f{}_{gh}\mathbf{l}_f. \tag{38}$$

The commutation coefficients are functions only of time according to the transformation

$$\mathsf{C}^f{}_{gh} \equiv \langle \boldsymbol{\lambda}^f, [\mathbf{l}_g, \mathbf{l}_h] \rangle = \widetilde{C}_K{}^f C_g{}^L C_h{}^M \mathsf{G}^K{}_{LM} \tag{39}$$

and inversely,

$$G^{D}{}_{FG} = C_m{}^{D}\widetilde{C}_F{}^{p}\widetilde{C}_G{}^{q}C^{m}{}_{pq}. \tag{40}$$

Since everything is to be considered in spacetime we also introduce a timelike vector field $\mathbf{l}_4$ orthogonal to the spacelike hypersurfaces. (This can be the normal gradient of any well defined scalar in the hypersurfaces.) It and its dual 1-form, λ^4 are given by the holonomic time quantites

$$\lambda^4 \equiv \mathbf{d}t, \quad \mathbf{l}_4 \equiv \partial_t. \tag{41}$$

So that we now have an orthonormal tetrad field which is generally time-dependent. In spacetime we also have commutation coefficients between the timelike vector field $\mathbf{l}_4$ and $\mathbf{l}_k$. By the Maurer–Cartan equation these commutation coefficients appear in the exterior derivatives of the orthonormal tetrad basis 1-forms:

$$\mathbf{d}\lambda^{\alpha} = -C^{\alpha}{}_{|\mu\nu|}\lambda^{\mu}\wedge\lambda^{\nu} \quad (|\mu\nu| \text{ means } \mu<\nu). \tag{42}$$

With $\lambda^4 = \mathbf{d}t$ we have

$$\mathbf{d}\lambda^4 = -C^4{}_{|\mu\nu|}\lambda^{\mu}\wedge\lambda^{\nu} = 0 \quad \text{or} \quad C^4{}_{\mu\nu} = 0. \tag{43}$$

The vanishing of these six coefficients means that the hypersurface-orthogonal vector field $\mathbf{l}_4$ is *geodesic* and *nonvortical* (If $\mathbf{l}_4$ represents the 4-velocity of matter world lines then these have zero acceleration and zero rotation relative to a Fermi–Walker transported triad, i.e., relative to an inertially guided gyroscopically stabilized frame of reference). This interpretation is seen from the relation of the $C^4{}_{\mu\nu}$ to the corresponding Ricci rotation coefficients (See Appendix)

$$C_{4k4} = \Gamma_{k44} \equiv l_k\cdot(l_{4;4}), \quad C_{4km} = -2\Gamma_{4[km]} = l_4\cdot[l_k,l_m], \tag{44}$$

which have an immediate kinematic meaning as acceleration and vorticity.

As a second application of Eq. (42) we can obtain the differential equations governing $\widetilde{C}_J{}^{s}$ by comparing the exterior derivative of λ^4 represented in two different ways. This is done in SHNC. Here we shall present the dual derivation. We substitute the expressions

$$\mathbf{l}_4 = C_4{}^{\mathrm{IV}}\mathbf{e}_{\mathrm{IV}}, \quad \mathbf{l}_i = C_i{}^{K}\mathbf{e}_K \tag{45}$$

into the commutation relation for $\mathbf{l}_4$ and $\mathbf{l}_i$ and obtain

$$[\mathbf{l}_4,\mathbf{l}_i] = [C_4{}^{\mathrm{IV}}\mathbf{e}_{\mathrm{IV}}, C_i{}^{K}\mathbf{e}_K] = C_4{}^{\mathrm{IV}}\dot{C}_i{}^{K}\mathbf{e}_K = C^{u}{}_{4i}\mathbf{l}_u = C^{u}{}_{4i}C_u{}^{K}\mathbf{e}_K,$$

where we have used the condition of Eq. (43) that the time-measuring vector field $\mathbf{l}_4$ is geodesic and nonvortical. Making the immaterial specialization that the time coordinate is not changed at all when we go from the invariant $\mathbf{e}_A$ basis to the orthonormal $\mathbf{l}_\mu$ basis ($C_4{}^{\mathrm{IV}} = 1$), we infer

$$\dot{C}_i{}^{K} = C^{u}{}_{4i}C_u{}^{K} = -(\mathsf{S}+\mathsf{W})_i{}^{u}C_u{}^{K} \sim -(\mathbf{S}+\mathbf{W})\mathbf{C} = \dot{\mathbf{C}} \tag{46}$$

and, correspondingly,

$$\dot{\widetilde{C}}_K{}^{r} = -\widetilde{C}_K{}^{u}C^{r}{}_{4u} = (\mathsf{S}-\mathsf{W})^{r}{}_{u}\widetilde{C}_K{}^{u} \sim \widetilde{\mathbf{C}}(\mathbf{S}+\mathbf{W}) = \dot{\widetilde{\mathbf{C}}}. \tag{46'}$$

Here the right-hand sides come from the definition by the coefficients C_{fgh} of kinematic quantities **S** and **W** in accord with

$$-C_{(gh)4} \equiv \Gamma_{4(gh)} \equiv -S_{gh} = -S_{hg},$$
$$-C_{[gh]4} = \Gamma_{gh4} \equiv W_{gh} = -W_{hg}. \tag{47}$$

The S_{gh} are the orthonormal triad components of the rate-of-strain tensor or, geometrically, the extrinsic curvature of the hypersurface of homogeneity; and

$$W_{fg} = -W_{gf} \equiv \epsilon_{fgh}\omega^{h} \tag{48}$$

represent the three independent orthonormal components of the angular velocity of the orthonormal spatial triad relative to a Fermi–Walker transported triad, i.e., the angular velocity of the orthonormal reference triad with respect to a gyroscopically stabilized inertially guided reference frame ("compass of inertia" in Weyl's term). The choice of $W_{ab} \equiv -\omega^{c}$ $(a,b,c = 1,2,3$ *et cycl.*) is partially free only in some of the types and solutions; in other words for many Bianchi type solutions there is an inescapable rotation of a naturally chosen reference triad with respect to the "compass of inertia."

In solving Eqs. (46), assuming that **S** and **W** are known—say from solving Einstein's equations for given physical conditions (see Appendix I) we have to assign initial values $C_u{}^{K}(t_0)$ at some time $t = t_0$. This transformation from the generally nonorthonormalizing but invariant basis to an initial orthonormalizing basis at $t = 0$ obeys the relations listed in the last columns of Tables II and III interpreted as an inverse transformation. For even if the initial $C_u{}^{s}(t_0)$ can be chosen diagonal, meaning that initially the orthonormal triad is parallel to the invariant triad, from then on an evolution of the orientation and magnitude of the members of the orthonormal triad takes place and the $C_u{}^{K}(t)$ need not be diagonal. They will evolve according to Eqs. (46) and will remain diagonal if and only if $(\mathsf{S}+\mathsf{W})_f{}^{g} = 0$ for all $f \neq g$. But then

$$(\mathsf{S}+\mathsf{W})_{fg} = (\mathsf{S}+\mathsf{W})_{gf} = (\mathsf{S}-\mathsf{W})_{fg} = 0 \quad (f \neq g).$$

Thus the condition for the **C**'s to be diagonal is that **S** be diagonal and $\mathbf{W} = 0$. Actually, this result may be considered to be a corollary of a general theorem which we shall now establish.

The **C**-matrix may not be fully diagonal—only block diagonal. When Eqs. (46) are written out one establishes immediately the following *Double-conjugate-pair Theorem*: *If two pairs of conjugate off-diagonal elements of the* **C***-matrix are known to be zero then the corresponding pairs of conjugate elements of* **S** *and* **W** *are zero.*

For instance, set $C_1{}^{\mathrm{III}}$, $C_2{}^{\mathrm{III}}$, $C_3{}^{\mathrm{I}}$, $C_3{}^{\mathrm{II}}$ equal to zero in that subset of Eq. (46) which is relevant. We then have

$$\begin{aligned}
&(S_{13}+\omega_2)C_3{}^{\mathrm{III}} = 0,\\
&(S_{23}-\omega_1)C_3{}^{\mathrm{III}} = 0,\\
&(S_{13}-\omega_2)C_1{}^{\mathrm{I}} + (S_{23}+\omega_1)C_2{}^{\mathrm{I}} = 0,\\
&(S_{13}-\omega_2)C_1{}^{\mathrm{II}} + (S_{23}+\omega_1)C_2{}^{\mathrm{II}} = 0.
\end{aligned} \tag{46a}$$

Because the matrix of **C**'s is nonsingular,

$$C_3{}^{\mathrm{III}}(C_1{}^{\mathrm{II}}C_2{}^{\mathrm{II}} - C_1{}^{\mathrm{II}}C_2{}^{\mathrm{I}}) \neq 0,$$

we have that

$$S_{13}+\omega_2=0,\quad S_{13}-\omega_2=0,\quad \therefore S_{13}=0,\quad \omega_2=0,$$
$$S_{23}-\omega_1=0,\quad S_{23}+\omega_1=0,\quad \therefore S_{23}=0,\quad \omega_1=0. \tag{46b}$$

That the converse is also true, i.e., that the block diagonal structure of the **C** matrix is both necessary and sufficient for the vanishing of the corresponding elements of **S** and **W**, follows immediately from the same differential equations which led to (46a). Assuming the conditions of Eq. (46b) we have

$$\dot{C}_1^{\,III}=-s_1C_1^{\,III},\quad \dot{C}_2^{\,III}=-s_2C_2^{\,III},$$
$$\dot{C}_3^{\,I}=-s_3C_3^{\,I},\quad \dot{C}_3^{\,II}=-s_3C_3^{\,II}$$
$$(s_1\equiv S_{11},\ s_2\equiv S_{22},\ s_3\equiv S_{33}).$$

The initial vanishing of these off-diagonal elements then implies that they remain zero thereafter.

Under the conditions of the Double-conjugate-pair theorem the remaining four equations of the total set (46) break up into two pairs of coupled equations, one pair for $C_1^{\,I}$ and $C_2^{\,I}$, and the other pair for $C_1^{\,II}$ and $C_2^{\,II}$. The coefficients are identical and so likewise are the forms of the solutions for the two pairs. We need find the solution then for only one pair, e.g., $C_1^{\,I}$ and $C_2^{\,I}$. The equations then read

$$-\dot{C}_1^{\,I}=s_1C_1^{\,I}+(S_{12}-\omega)C_2^{\,I},$$
$$-\dot{C}_2^{\,I}=(S_{12}+\omega)C_1^{\,I}+s_2C_2^{\,I} \qquad (\omega\equiv\omega_3). \tag{46c}$$

First we give a preliminary survey of results. In the following, in the context of solving the differential equation system (46), we shall again prove (in matrix notation) the Double-conjugate-pair theorem; we will then make considerable application of the theorem because in almost all types the block diagonal structure is the most general that can occur if the following basic condition is satisfied: *The orthonormalizing basis is chosen to coincide at all times with the characteristic directions of a symmetric tensor* **F** *and its eigenvector* **n** *which are the evolvents in time of* $A^{(KM)}$ *and* n_0; **F** *and* **n** *also completely determine the intrinsic curvature of the spatial hypersurfaces.*

Nondiagonal structure (in particular, proper block diagonal) of the **C** matrix occurs only if a degree of symmetry exists in the geometry of the hypersurface—degeneracy in the eigenvalues of **F**. Contrapositively, nondegeneracy in **F** implies diagonal structure of the **C** matrix. This situation of non-degeneracy can occur in particular solutions of all types except in I, II, IV, and V where degeneracy is built in the type structure.

Though non-degeneracy in **F** is *sufficient* for diagonal **S** and vanishing ω, it is *not necessary*. In fact it will turn out that invariance of the diagonal structure—and degeneracies when they occur—of **F** implies that **S** is also diagonal in all cases (*a fortiori* in Types I and V when it is chosen diagonal:). With **F** degenerate **S** is always diagonal but we may or may not have ω vanish. We may have solutions with vanishing ω—and therefore diagonal **C** matrix—also in cases where degeneracy is inescapable, e.g., in types I, II, IV, and V. The general situation however when degeneracy in **F** occurs is that ω does not vanish and the **C** matrix does not remain diagonal but develops into a product of a diagonal matrix $\grave{\mathbf{D}}$ and an orthogonal matrix **R** with the rotation angles given by the time integrals of ω. The fact that **C** is representable as a product $\mathbf{R}\grave{\mathbf{D}}$ is a specialized form of the general polar decomposition of an arbitrary nonsingular matrix. Other than in type I and totally symmetrical IX, where general rotations are involved, the orthogonal matrix has at most block diagonal structure (rotation about one axis only). But in general, in cases of degenerate **F**, **C** is certainly not symmetric contrary to Ryan and Shepley's assumption.[4] Evidently they are assuming a gauge which does not preserve the quasi-canonical form of **F**. In our gauge the **C**'s are symmetric only when diagonal. Restricting the orthonormalizing transformations in this way limits the possible solutions to those where ω vanishes and **S** remains diagonal at all times. While the latter condition is generally no restriction (see Table IV) the requirement of vanishing ω arbitrarily restricts solutions in almost all types in cases where there is some symmetry in the intrinsic geometry. Fortunately, however, this restriction does not affect the possible canonical forms of metrics for all the types because these metrics depend only on the diagonal factor $\grave{\mathbf{D}}$ in the product $\mathbf{C}=\mathbf{R}\grave{\mathbf{D}}$ or $\widetilde{\mathbf{C}}=\grave{\mathbf{D}}^{-1}\mathbf{R}^{-1}\equiv\grave{\mathbf{D}}^{-1}\mathbf{R}^{T}$. This can be seen from Eqs. (11) or (36) which become

$$\gamma=\widetilde{\mathbf{C}}\mathbf{C}^{T}=\grave{\mathbf{D}}^{-1}\mathbf{R}^{T}\mathbf{R}(\grave{\mathbf{D}}^{-1})^{T}=(\grave{\mathbf{D}}^{-1})^{2}. \tag{36*}$$

However, **R** (and ω) play a role in the physics and geometry.

We proceed now to the detailed analysis. The two other sets of quantities, besides **S** and **W**, which play an important role in the theory are then the symmetric "dyadic"

$$F^{(km)}=C^{(k}{}_{de}\epsilon^{m)de} \tag{49}$$

and the "Bianchi vector"

$$n_k=C^{m}{}_{km}. \tag{50}$$

It is clear that **F** and **n** bear the same relation to the commutation coefficients that a_0, b_0, c_0, and n_0 bear to the structure constants. In fact we see that they are time evolvents of: (1) the symmetrized, and (2) the (dualized) antisymmetric parts of the structure constant matrix introduced in Eqs. (14) and (15) respectively. They transform from any initial values they may have by the transformation formulas (19) and (20), with the transformation coefficients now being functions of time which satisfy the differential equations (46). Substituting the definitions (49) and (50) in the transformation formulas and taking for the initial values the canonical values

$$F_0^{\,KM}=\mathrm{diag}F_0(K)\equiv\mathrm{diag}(F_0(I),F_0(II),F_0(III))$$
$$\equiv\mathrm{diag}(a_0,b_0,c_0), \tag{51}$$

we have

$$F^{ef}(t)=\Delta\widetilde{C}_K{}^{e}\widetilde{C}_M{}^{f}F_0(K)\delta^{KM}, \tag{52}$$
$$n_g(t)=C_g^{\,III}(t)n_0. \tag{53}$$

The differential equations (46) govern the transformation coefficients $C_i^{\,K}(t)$. There is a certain amount of freedom in the choice of the orthonormalizing frame, i.e., one may choose the ω components so that certain conditions are maintained. These together with the differential equations (46) restrict the transformation coefficients $C_i^{\,K}$.

Two very useful results may be established.

(1) *By proper choice of* ω (and therefore of the metric-orthonormalizing **C**'s) **F** *may be kept diagonal* (quasi-canonical) *at all times*

$$\mathbf{F}\sim\mathrm{diag}(a,b,c)\equiv\mathrm{diag}(\mathsf{F}(1),\mathsf{F}(2),\mathsf{F}(3))\sim\grave{\mathbf{F}}.$$

In the types I and V where $\mathbf{F}$ is identically zero the $\mathbf{C}$'s may be chosen to diagonalize $\mathbf{S}$ instead.

(2) The general intrinsically defined vector in the pure differential-geometric structure of the vector types is the "Bianchi vector" $\mathbf{n}$. There are of course also three principal axis directions of the symmetric quadric $\mathbf{F}$ (modulo possible degeneracies). Fortunately the directions picked out coincide here: *In all vector types, the third direction of the standard orthonormalizing frame* (i.e., the frame which also maintains $\mathbf{F}$, or $\mathbf{S}$, diagonal), *may be chosen in the direction of* $\mathbf{n}$ which then has only the one non-vanishing component n.

That $\mathbf{F}$ or $\mathbf{S}$ may be diagonalized simultaneously with $\boldsymbol{\gamma}$ is evident, since this can always be done with two symmetric quadratic forms if one of them ($\boldsymbol{\gamma}$ in this case) is positive definite. What is the bonus here is that, due to the persistence of the Jacobi identity

$$\mathbf{F}\cdot\mathbf{n}=0$$

—condition for associativity in the commutation of the evolving basis-vector fields—the eigenvector $\mathbf{n}$ may be maintained in the third principal direction of $\mathbf{F}$.

As has been remarked, $\mathbf{F}$ and $\mathbf{n}$ also have a geometrical meaning. They are related to the intrinsic curvature of the hypersurfaces of homogeneity. In fact, the Riemann curvature form of the 3-geometry on the spatial hypersurfaces is entirely composed of $\mathbf{F}$ and $\mathbf{n}$.[12] The nonvanishing elements in the matrix of Riemann curvature 2-forms are

$$\begin{aligned}
\mathbf{R}_{12} &= -(n^2-cD+AB)\boldsymbol{\lambda}^1\wedge\boldsymbol{\lambda}^2,\\
\mathbf{R}_{13} &= (-n^2+bB-AD)\boldsymbol{\lambda}^1\wedge\boldsymbol{\lambda}^3+n(B-D+a)\boldsymbol{\lambda}^2\wedge\boldsymbol{\lambda}^3,\\
\mathbf{R}_{23} &= -n(A-D+b)\boldsymbol{\lambda}^1\wedge\boldsymbol{\lambda}^3\\
&\quad+(-n^2+aA-BD)\boldsymbol{\lambda}^2\wedge\boldsymbol{\lambda}^3.
\end{aligned}$$

The Ricci curvature tensor is

$$\mathbf{R}=\begin{bmatrix}\mathsf{R}-4n^2+2aA & n(b-a) & 0\\ n(b-a) & \mathsf{R}-4n^2+2bB & 0\\ 0 & 0 & \mathsf{R}-4n^2+2CD\end{bmatrix}$$

where the Ricci curvature scalar is

$$\mathsf{R}=6n^2+\tfrac{1}{2}(a^2+b^2+c^2)-(bc+ca+ab)$$

and the abbreviations A, B, and D have been introduced:

$$A\equiv\tfrac{1}{2}(b+c-a)\quad B\equiv\tfrac{1}{2}(c+a-b),\quad D\equiv\tfrac{1}{2}(a+b-c).$$

$\mathbf{F}$ and $\mathbf{n}$ vanish identically for Bianchi type I. Thus, all spatial hypersurfaces for type I solutions are flat, in contrast to the general situation for the five "*spatially curved*" nonvector types II, VI_0, VII_0, VIII, IX and the $3+2\infty$ spatially curved vector types III, IV, V, VI_h, VII_h.

With the quasi-canonical choice for the orthonormal triads, Eqs. (52) and (53) become

$$\begin{aligned}
&\mathsf{F}^p(t)=\Delta\sum_{K=\mathrm{I}}^{\mathrm{III}}(\widetilde{C}_K{}^p)^2\mathsf{F}_0(K),\quad 0=\Delta\sum_{K=\mathrm{I}}^{\mathrm{III}}\widetilde{C}_K{}^p\widetilde{C}_K{}^q\mathsf{F}_0(K)\quad(p\neq q),\\
&n_g(t)=C_g{}^{\mathrm{III}}n_0\qquad(=0\text{ for }g=1,2\text{ except in type V}).
\end{aligned}\tag{54}$$

In matrix notation Eq. (54) may be rewritten

$$\grave{\mathbf{F}}=\Delta(\grave{\mathbf{F}}_0\widetilde{\mathbf{C}})^T\widetilde{\mathbf{C}}=\Delta\widetilde{\mathbf{C}}^T\grave{\mathbf{F}}_0\widetilde{\mathbf{C}}.\tag{54'}$$

As explained in the preceding the most general form of $\mathbf{C}$ or $\widetilde{\mathbf{C}}$ is that of a product of a diagonal matrix and a rotation matrix (special case of polar decomposition) $\mathbf{C}=\mathbf{R}\grave{\mathbf{D}}$ and $\widetilde{\mathbf{C}}=\grave{\mathbf{D}}^{-1}\mathbf{R}^{-1}=\grave{\mathbf{D}}^{-1}\mathbf{R}^T$. Here the diagonal matrix $\mathbf{D}$ represents the pure strain—the inverse of the transformation given by (21)—needed to go from the canonical to the quasi-canonical but orthonormal form of the spatial metric

$$\begin{aligned}
\mathbf{d}\sigma^2&=(\boldsymbol{\epsilon}^{\mathrm{I}}/\alpha)^2+(\boldsymbol{\epsilon}^{\mathrm{II}}/\beta)^2+(\boldsymbol{\epsilon}^{\mathrm{III}}/\gamma)^2\\
&=(\boldsymbol{\lambda}^1)^2+(\boldsymbol{\lambda}^2)^2+(\boldsymbol{\lambda}^3)^2
\end{aligned}$$

and the rotation $\mathbf{R}$ represents an additional gauge transformation, i.e., one which preserves the orthonormal form of $\mathbf{d}\sigma^2$ and the quasi-canonical form of $\grave{\mathbf{F}}$. We may write Eq. (54')

$$\grave{\mathbf{F}}=\Delta\,\mathbf{R}\grave{\mathbf{D}}^{-1}\grave{\mathbf{F}}_0\grave{\mathbf{D}}\mathbf{R}^{-1}\equiv\mathbf{R}\grave{\overline{\mathbf{F}}}_0\mathbf{R}^{-1},\tag{55}$$

where we have set, in accord with Eqs. (23)–(25)

$$\Delta\grave{\mathbf{D}}^{-1}\grave{\mathbf{F}}_0\grave{\mathbf{D}}\equiv\grave{\overline{\mathbf{F}}}_0\sim\mathrm{diag}(a_0\beta\gamma/\alpha,\ b_0\gamma\alpha/\beta,\ c_0\alpha\beta/\gamma).$$

Eq. (55) can only be satisfied as follows for various canonical structure matrix degeneracies, with which are listed the associated admissible rotation matrices and types:

1. $a\neq b\neq c\neq a$: only $\mathbf{R}=\mathbf{R}_0$ (the identity).

Nonvector: All VI_0. Axially unsymmetric VII_0, VIII and completely unsymmetric IX.
Vector: All VI_h. Axially unsymmetric VII_h.

2. $a=b\neq c$: $\mathbf{R}=\mathbf{R}_z$ rotation about the z axis

Nonvector: Axially symmetric VII_0, VIII and z-rotational IX.
Vector: Axially symmetric VII_h.

3. $b=c\neq a$: $\mathbf{R}=\mathbf{R}_x$ rotation about the x axis

Nonvector: All II, x-rotational IX.
Vector: All IV.

4. $c=a\neq b$: $\mathbf{R}=\mathbf{R}_y$ rotation about the y axis

Nonvector: y-rotational IX.
Vector: None.

5. $a=b=c=0$: $\mathbf{R}_x,\mathbf{R}_y,\mathbf{R}_z$

Nonvector: extrinsically anisotropic I extrinsically isotropic I: $\mathbf{R}_x\mathbf{R}_y\mathbf{R}_z$ if $s_1=s_2=s_3$

if $s_1\neq s_2$, no $\mathbf{R}_z$;

if $s_2\neq s_3$, no $\mathbf{R}_x$;

if $s_3\neq s_1$, no $\mathbf{R}_y$;

Vector: V with quasi-canonical $\mathbf{n}=(0,0,n)$, level $\mathbf{R}_z$ only if $s_1=s_2$

6. $a=b=c\neq 0$: $\mathbf{R}_x\mathbf{R}_y\mathbf{R}_z$.

Nonvector: isotropic IX.

All these cases, which comprise all possible instances of Bianchi–Behr types, are listed together with the asociated $\mathbf{S}$, $\boldsymbol{\omega}$ and $\mathbf{C}$ quantities in Table IV.

If is of interest first to derive the conditions on $\mathbf{S}$, $\boldsymbol{\omega}$ and $\mathbf{C}$ for the completely isotropic cases (isotropic IX and assumed extrinsically isotropic I). Under these conditions we have that the $\mathbf{C}$ matrix can only have the form of a uniform

TABLE IV.

Degeneracies of diagonal elements of the canonical structure matrix $\begin{bmatrix} a & n & 0 \\ -n & b & 0 \\ 0 & 0 & c \end{bmatrix}$	Possible Bianchi types and admissible R_x, R_y, R_z (+ identity R_0) preserving canonical form of the structure matrix	Nonvanishing elements of the extrinsic curvature $S_{ij} = \begin{bmatrix} S_1 & S_{12} & S_{13} \\ S_{12} & S_2 & S_{23} \\ S_{13} & S_{23} & S_3 \end{bmatrix}$	Nonvanishing elements of $W_{ij} \equiv \omega_k$	Gauge subgroup of admissible pure strain-rotation matrices $\mathbf{C} \equiv \mathbf{R}\dot{\mathbf{D}}$ ($\dot{D} \equiv$ diagonal strain matrix). These are metric-orthonormalizing while preserving the quasi-canonical form of the structure matrix	Parameters in C matrices expressed in terms of basic intrinsic and extrinsic quantities. The lower limit in the integrals is the orthonormal triad is aligned with the invariance triad, i.e., when $a = a_0$, $b = b_0$, $c = c_0$
None	*Nonvector*:				
$a \neq b \neq c \neq a$	All VI_0 R_0	S_1 S_2 S_3	None	$\begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha \left\lvert \frac{a}{b} \right\rvert^{1/2} & 0 \\ 0 & 0 & \lvert ab \rvert^{1/2} \end{bmatrix}$	$\alpha = e^{\int S_1 dt}$
	VII_0 R_0	S_1 S_2 S_3	None		
	VIII R_0	S_1 S_2 S_3	None	$\begin{bmatrix} \lvert bc \rvert^{1/2} & 0 & 0 \\ 0 & \lvert ca \rvert^{1/2} & 0 \\ 0 & 0 & \lvert ab \rvert^{1/2} \end{bmatrix}$	
	IX R_0	S_1 S_2 S_3	None		
	Vector:				
	All VI_h (III) R_0	S_1 S_2 S_3	None	$\begin{bmatrix} \alpha & 0 & 0 \\ 0 & \lvert h \rvert^{1/2} \frac{a}{n} \alpha & 0 \\ 0 & 0 & n/\lvert h \rvert^{1/2} \end{bmatrix}$	$\alpha = e^{\int S_1 dt}$
	VII_h R_0	S_1 S_2 S_3			
$a = b \neq c$	*Non-Vector*:				$\alpha = e^{\int S_1 dt}$
	VII_0 R_z	$S_1 = S_2$ S_3	ω_3	$\alpha \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & a/\alpha \end{bmatrix}$	$\theta = \int \omega_3 dt$
	VIII R_z	$S_1 = S_2$ S_3	ω_3	$\lvert ac \rvert^{1/2} \begin{bmatrix} \cos & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & \lvert a/c \rvert^{1/2} \end{bmatrix}$	$\theta = \int \omega_3 dt$
	IX R_z	$S_1 = S_2 \neq S_3$	ω_3		
	Vector:				$\alpha = e^{\int S_1 dt}$
	VII_h R_z	$S_1 = S_2$ S_3	ω_3	$\alpha \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & a/\alpha \end{bmatrix}$	$\theta = \int \omega_3 dt$
$b = c \neq a$	*Nonvector*:				$\alpha = e^{\int S_1 dt}$
	All type II R_x	S_1 $S_2 = S_3$	ω_1	$\begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta\cos\theta & \frac{\alpha}{\beta} a\sin\theta \\ 0 & -\beta\sin\theta & \frac{\alpha}{\beta} a\cos\theta \end{bmatrix}$	$\beta = e^{\int S_1 dt}$ $\theta = \int \omega_1 dt$
	IX	$S_1 \neq S_2 = S_3$	ω_1	$\lvert ba \rvert^{1/2} \begin{bmatrix} \lvert b/a \rvert^{1/2} & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$	$\theta = \int \omega_1 dt$
	Vector:				
	IV R_x	$S_1 = S_2$ S_3	ω_1	$a \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$	$\theta = \int \omega_1 dt$
	R_0	$S_1 \neq S_2$ S_3	None	$\begin{bmatrix} \alpha & 0 & 0 \\ 0 & \frac{a}{n}\alpha & 0 \\ 0 & 0 & n \end{bmatrix}$	$\alpha = e^{\int S_1 dt}$
$b \neq c = a$	*NonVector*:				$\lvert ba \rvert^{1/2} = e^{\int S_1 dt}$
	IX R_y	$S_3 = S_1 \neq S_2$	ω_2	$\lvert ba \rvert^{1/2} \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & \lvert a/b \rvert^{1/2} & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$	$a = e^{\int S_2 dt}$ $\theta = -\int \omega_2 dt$

TABLE IV.

Degeneracies of diagonal elements of the canonical structure matrix $\begin{bmatrix} a & n & 0 \\ -n & b & 0 \\ 0 & 0 & c \end{bmatrix}$	Possible Bianchi types and admissible R_x, R_y, R_z (+ identity R_0) preserving canonical form of the	Nonvanishing elements of the extrinsic curvature $S_{ij} = \begin{bmatrix} S_1 & S_{12} & S_{13} \\ S_{12} & S_2 & S_{23} \\ S_{13} & S_{23} & S_3 \end{bmatrix}$	Nonvanishing elements of $W_{ij} \equiv \omega_k$	Gauge subgroup of admissible pure strain-rotation matrices $\mathbf{C} \equiv \mathbf{R}\dot{\mathbf{D}}$ $\dot{D} \equiv$ diagonal strain matrix). These are metric-orthonormalizing while preserving the quasi-canonical form of the structure matrix	Parameters in C matrices expressed in terms of basic intrinsic and extrinsic quantities The lower limit in the integrals is the orthonormal triad is aligned with the invariance triad, i.e., when $a = a_0$, $b = b_0$, $c = c_0$
$a = b = c$	*Nonvector:*				
	Unsymmetric I R_0	$S_1\ S_2\ S_3$	None	$\mathbf{R}_x$	$\alpha = e^{\int S_1 dt}$
	Axial I R_x	$S_1 = S_2 \neq S_3$	ω_3	$\alpha \mathbf{R}_x \mathbf{R}_y \mathbf{R}_z$	
	Isotropic I and IX R_x, R_y, R_z	$S_1 = S_2 = S_3 \equiv S$	$\omega_1, \omega_2, \omega_3$		
	Vector:				
	Axial or isotropic V R_z	$S_1 = S_2\ S_3$	ω_3	$\alpha \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & n/\alpha \end{bmatrix}$	$\alpha = e^{\int S_1 dt}$ $\theta = \int \omega_3\, dt$

dilation multiplied by a general rotation

$$\mathbf{C} = \alpha \mathbf{R}, \quad (\mathbf{R} = \mathbf{R}_x \mathbf{R}_y \mathbf{R}_z). \tag{56}$$

From the properties of the rotation matrix we then have

$$(\mathbf{C}/\alpha)^{-1} = (\mathbf{C}/\alpha)^T = \alpha \mathbf{C}^{-1} \equiv \alpha \tilde{\mathbf{C}}$$

or

$$\mathbf{C}^T = \alpha^2 \tilde{\mathbf{C}},$$

Applying this to the transpose of Eq. (46) we get

$$(-\alpha^2 \tilde{\mathbf{C}})^{\cdot} = \mathbf{C}^T(\mathbf{S} - \mathbf{W}) = \alpha^2 \tilde{\mathbf{C}}(\mathbf{S} - \mathbf{W})$$

or

$$\alpha^2 \dot{\tilde{\mathbf{C}}} = \alpha^2 \tilde{\mathbf{C}}[\mathbf{W} - \mathbf{S} - (\ln\alpha^2)^{\cdot}\, \mathbf{I}].$$

Comparing with Eq. (46), and noting that $\mathbf{C}$ is nonsingular, we find that $\mathbf{S}$ is a scalar multiple of the identity matrix with the multiplier, or diagonal element s, equal to $-(\ln\alpha)^{\cdot}$. It follows that

$$\alpha = \alpha_0 e^{-\int s\, dt}. \tag{57}$$

This is the specially totally symmetrical case. Similar forms for the dilation factors hold in the more general cases of lesser symmetry when the matrix $\mathbf{C}$ is only block-diagonal or diagonal, as will now be seen.

The diagonal case is self-evident because then $\mathbf{S} - \mathbf{W}$ must be a diagonal matrix which can only hold if $\mathbf{W} = 0$ and $\mathbf{S}$ is diagonal. In a typical (e.g., z-rotational) block diagonal case we can write the $\mathbf{C}$ matrix in the form

$$\mathbf{C} = \begin{pmatrix} \dot{\alpha} & 0 \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \mathfrak{R} & 0 \\ 0 & 1 \end{pmatrix}$$
$$\left(\mathfrak{R} \equiv \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad \dot{\alpha} \equiv \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \right), \tag{58}$$

where $\mathfrak{R}^{-1} = \mathfrak{R}^T$ and where we assume that

$$\alpha \neq \gamma \quad \text{or} \quad \tilde{\alpha}\gamma \neq 1 \quad (\tilde{\alpha} \equiv \alpha^{-1})$$

in order not to be in the previous completely isotropic case. We then have

$$\mathbf{C}^T = \begin{pmatrix} \mathfrak{R}^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{\alpha} & 0 \\ 0 & \gamma \end{pmatrix} = \tilde{\mathbf{C}} \begin{pmatrix} \dot{\alpha}^2 & 0 \\ 0 & \gamma^2 \end{pmatrix}.$$

From the transpose of Eq. (46), into which we substitute this expression, we get

$$\dot{\tilde{\mathbf{C}}} \begin{pmatrix} \dot{\alpha}^2 & 0 \\ 0 & \gamma^2 \end{pmatrix} + \tilde{\mathbf{C}} \begin{pmatrix} \dot{\alpha}^2 & 0 \\ 0 & \gamma^2 \end{pmatrix}^{\cdot}$$
$$= \tilde{\mathbf{C}} \begin{pmatrix} \dot{\alpha}^2 & 0 \\ 0 & \gamma^2 \end{pmatrix} (\mathbf{W} - \mathbf{S}).$$

Comparing with Eq. (46′) which can be written

$$\dot{\tilde{\mathbf{C}}} \begin{pmatrix} \dot{\alpha}^2 & 0 \\ 0 & \gamma^2 \end{pmatrix} = \tilde{\mathbf{C}}(\mathbf{S} + \mathbf{W}) \begin{pmatrix} \dot{\alpha}^2 & 0 \\ 0 & \gamma^2 \end{pmatrix}$$

and noting that $\mathbf{C}$ is nonsingular, we find

$$\mathbf{S} + \mathbf{W} = \begin{pmatrix} \dot{\alpha}^2 & 0 \\ 0 & \gamma^2 \end{pmatrix} (\mathbf{W} - \mathbf{S}) \begin{pmatrix} \tilde{\dot{\alpha}}^2 & 0 \\ 0 & \tilde{\gamma}^2 \end{pmatrix}$$
$$- \begin{pmatrix} \ln\dot{\alpha}^2 & 0 \\ 0 & \ln\gamma^2 \end{pmatrix}^{\cdot}.$$

The antisymmetric and symmetric parts of the matrices on the two sides of this equation must be separately equal and, as $\tilde{\alpha}\gamma \neq 1$, this holds iff $W_{13} = W_{23} = S_{13} = S_{23} = 0$ (conjugate elements likewise), which is again a proof of the Double-conjugate-pair theorem. The remaining equation is

$$\mathbf{S} = -\begin{pmatrix} \ln\dot{\alpha} & 0 \\ 0 & \ln\gamma \end{pmatrix}^{\cdot}. \tag{59}$$

Thus $\mathbf{S}$ is again a diagonal matrix here, with elements

$$s_1 = s_2,$$

so that the intrinsic axial symmetry is reflected in an extrinsic axial symmetry as well. The diagonal factor in $\mathbf{C}$ is:

$$\begin{pmatrix} \dot{\alpha} & 0 \\ 0 & \gamma \end{pmatrix} = \begin{pmatrix} \alpha_0 e^{-\int s_1\,dt} & 0 & 0 \\ 0 & \alpha_0 e^{-\int s_1\,dt} & 0 \\ 0 & 0 & \gamma_0 e^{-\int s_3\,dt} \end{pmatrix}. \quad (60)$$

To find an expression for the only remaining time-dependent parameter in **C**, i.e., θ, we substitute Eq. (59) for **S** into the differential equation (46) with **C** given by Eq. (58) and we find

$$\begin{pmatrix} \dot{\mathcal{A}} & 0 \\ 0 & 0 \end{pmatrix} = -\mathbf{W}\begin{pmatrix} \mathcal{A} & 0 \\ 0 & 1 \end{pmatrix}$$

or

$$(\cos\theta)^{\cdot} = -\omega_3 \sin\theta,$$

$$\therefore \theta = \int \omega_3\,dt. \quad (61)$$

$$(\sin\theta)^{\cdot} = \omega_3 \cos\theta,$$

Differential equations for **F** and **n** may be derived from the finite evolution Eqs. (54) and (55) by differentiating and using Eqs. (46) and (46′). We obtain (setting the rate of volume expansion $-\dot{\Delta}/\Delta \equiv \theta$):

$$\dot{a} = (2s_1 - \theta)a, \quad \dot{b} = (2s_2 - \theta)b, \quad \dot{c} = (2s_3 - \theta)c \text{ (alltypes)}, \quad (62)$$

$$(a - b)\mathsf{W}_{21} = (a + b)\mathsf{S}_{21}, \quad \textit{et cycl.} \text{ (all types)}. \quad (63)$$

In particular

$$a(\mathsf{S} - \mathsf{W})_{31} = 0, \quad b(\mathsf{S} - \mathsf{W})_{32} = 0 \text{ (all except VIII and IX)}. \quad (64)$$

Using Eq. (46) we obtain for all vector types

$$\dot{n}_1 = 0 = (\mathsf{S} - \mathsf{W})_{31} C_3^{\mathrm{III}} n_0 = -(\mathsf{S} - \mathsf{W})_{31} n \quad (n \equiv n_3(t)),$$

$$\dot{n}_2 = 0 = -(\mathsf{S} - \mathsf{W})_{32} C_3^{\mathrm{III}} n_0 = (\mathsf{S} - \mathsf{W})_{32} n, \quad (65)$$

$$\dot{n} = -s_3 C_3^{\mathrm{III}} n_0 = -s_3 n.$$

In the one vector type, V, where $\mathbf{F} = 0$, and **S** is chosen diagonal instead, $\boldsymbol{\omega} \| \mathbf{n}$ and **n** is an eigenvector of **S**; thus, again the direction of **n** may be chosen as the third direction of the standard orthonormalizing frame. This can be seen as follows: Consider any well-defined spatial vector appearing in a structure enlarged relative to the homogeneous spatial hypersurfaces, e.g., an embedding spacetime or, more specifically, Einstein's theory specifying a coupling of this spacetime with the local physics. On general symmetry grounds (spelled out for instance by applying the commutation algebra—time evolvent of the isometry Lie algebra—to the components of the vector) such a vector will have to be parallel to the common direction of $\boldsymbol{\omega}$ and **n** whenever a nonvanishing **n** exists. But such a well-defined vector is the energy flux, **t**, which satisfies one of the Einstein equations [Eq. (2.23) of SHNC]:

$$\mathbf{t} - \dot{\mathbf{F}} \times \mathbf{S} = -\theta \mathbf{n} + 3\mathbf{S}\cdot\mathbf{n}$$

or (Type V)

$$\mathbf{t} = -\theta \mathbf{n} + 3\mathbf{S}\cdot\mathbf{n}.$$

It is evident that **t** being parallel to **n** implies that **n** is an eigenvector of **S** in any type V solution, and can be associated with the third principal-axis direction of **S**. On the same general symmetry grounds we may expect that only the third component, ω_3, of the angular velocity vector $\boldsymbol{\omega}$ may be nonvanishing in the standard orthonormalizing basis for vector types, and this is confirmed by looking at the specific equations. [Actually these considerations should be examined in more detail—and will be, elsewhere—because while the "vectors" **n** and $\boldsymbol{\omega}$ originate in the same irreducible tensor symmetry type—that of antisymmetric tensors—**t** originates in a different symmetry type].

Leaving aside Type V where no problems arise, since Eqs. (63) and (64) are vacuous in that case, vector types are the only ones where the question of consistency could arise between Eqs. (64) and (65). We see however that Eqs. (65) require

$$\omega_1 = \mathsf{S}_{23}, \quad \omega_2 = -\mathsf{S}_{13}, \quad (66)$$

which is consistent with Eqs. (64). Both of these components must vanish by the preceding argument. We note that because of Einstein's equations this implies vanishing of the corresponding components of the stress tensor.

The explicit equation in the system (63)—valid for all types—gives

$$-\mathsf{W}_{12} \equiv \omega_3 = \frac{a + b}{a - b}\mathsf{S}_{12} \quad (67)$$

or, in all cases where $a = b \neq 0$, i.e., in axially symmetric cases of VII, VIII, and IX, $\mathsf{S}_{12} = 0$.

We note also from Eqs. (62) that in this case the first two diagonal elements of **S** are equal:

$$s_1 = s_2.$$

A corresponding conclusion follows for the other cases of degeneracy: $b = c$, $a = c$, etc.

Combining the foregoing requirements with the analysis indicated above of the gauge subgroup, i.e., the subgroup of pure strain-rotations which leave quasi-canonical forms of the structure matrix invariant, we again arrive at the results tabulated in Table IV. By virtue of the conditions imposed, the solutions for differential Eqs. (46) for the various types reduce to simple quadratures and block-diagonal form in almost all cases. The final canonical form for the metrics which results from this analysis is

$$\mathbf{ds}^2 = -\mathbf{d}t^2 + \gamma_{AB}\boldsymbol{\epsilon}^A \otimes \boldsymbol{\epsilon}^B = -\mathbf{d}t^2 + \left(\frac{\boldsymbol{\epsilon}^{\mathrm{I}}}{\alpha}\right)^2 + \left(\frac{\boldsymbol{\epsilon}^{\mathrm{II}}}{\beta}\right)^2 + \left(\frac{\boldsymbol{\epsilon}^{\mathrm{III}}}{\gamma}\right)^2, \quad (68)$$

where, for each of the Bianchi types in turn, expressions for α, β, γ in terms of the Einstein solution-quantities $e^{-\int s\,dt}$, a, b, c, n (or h) may be read out of the last two columns in Table IV.

ACKNOWLEDGMENTS

One of us (M.A.M.) acknowledges with pleasure the hospitality and very valuable scientific insights provided by Dr. J. F. Plebañski at the Centro de Investigación y Estudios Avanzados, I.P.N. in Mexico City and the stimulating discussions there with Dr. Plebański, Dr. Roy Kerr, and Dr. Andrzej Krasinski. He also wishes to express his appreciation to Dr. Paul Melvin and Dr. Stanislaw Bazanski for stimulating discussions, and to Dr. John Milnor and Dr. John Stachel for reading this manuscript. Travel support by Temple University is also gratefully acknowledged.

APPENDIX: SYNCHRONOUS FORM OF THE METRIC FOR HOMOGENEOUS COSMOLOGIES—RELATION BETWEEN STRUCTURE AND RICCI COEFFICIENTS

The necessity and sufficiency of the requirement of historical homogeneity, i.e., that each successive hypersurface constructed by measuring off equal parameter intervals along a timelike geodesic congruence normal to an initial homogeneous hypersurface should maintain spatial homogeneity, is fundamental, and should be proved rigorously:

Let t be a parameter along a timelike curve, i.e., the tangent vector to the curve is $\mathbf{d}/\mathbf{d}t$. Historical homogeneity is formally expressed by the existence of an *entire congruence* of timelike curves $\xi^{\alpha}(t;y_1,y_2,y_3)$ parametrized by y_1, y_2, y_3 such that the Killing vector fields $\mathbf{Z}_k$ operating in successive spacelike hypersurfaces of homogeneity connect points of equal t. The analytical expression of this property is the transport law

$$\mathscr{L}_{\mathbf{e}_{\mathrm{iv}}}(\mathbf{Z}_k)=0 \quad \text{or} \quad [\mathbf{e}_{\mathrm{iv}},\mathbf{Z}_k]=0, \tag{A1}$$

where $\mathbf{e}_{\mathrm{iv}} \equiv \partial/\partial t$ is the tangent vector field to the timelike congruence. [An excellent diagram illustrating this property of a vanishing Lie derivative of one vector field like $\mathbf{Z}_k$, with respect to another, like $\mathbf{e}_{\mathrm{iv}}$, is given in problem 8.14 Ref. 13.]

By the antisymmetry property of the commutators, the "historical homogeneity" or "conservation-of-homogeneity" equation (A1) implies the "global-time existence" equation:

$$\mathscr{L}_{\mathbf{Z}_k}(\mathbf{e}_{\mathrm{iv}})=[\mathbf{Z}_k,\mathbf{e}_{\mathrm{iv}}]=0.$$

As argued in the text of the present article immediately following Eq. (7)—by the completeness of the **e** and **Z** bases, respectively—we also have the vanishing of the commutators

$$-[\mathbf{e}_K,\mathbf{e}_{\mathrm{iv}}]=[\mathbf{e}_{\mathrm{iv}},\mathbf{e}_K]=\mathscr{L}_{\mathbf{e}_{\mathrm{iv}}}(\mathbf{e}_K)=0.$$

This equation guaranties that the metric, initially in the synchronous form (with $g_{\mathrm{iv}\,K}=0$), will retain this property from one homogeneous hypersurface to the next. For we have

$$\mathscr{L}_{\mathbf{l}_4}(g_{4k})=\mathscr{L}_{\mathbf{l}_4}(\mathbf{l}_4\cdot\mathbf{l}_k)=\mathscr{L}_{C^{\mathrm{iv}}{}_4\mathbf{e}_{\mathrm{iv}}}(C^{\mathrm{iv}}{}_4\mathbf{e}_{\mathrm{iv}}\cdot C_k{}^J\mathbf{e}_J)$$
$$=\dot{C}_k{}^J(\mathbf{e}_{\mathrm{iv}}\cdot\mathbf{e}_J)=g_{\mathrm{iv}\,J}\dot{C}_k{}^J=0,$$

where again we have made the immaterial specialization [Eq. (46)] that $C^{\mathrm{iv}}{}_4=1$. This establishes that historical homogeneity is *sufficient* to maintain a synchronous metric. That it is also *necessary* follows by tracing the steps in reverse order.

The Ricci coefficients defined by

$$\Gamma_{\mu\alpha\beta}\equiv\mathbf{l}_\mu\cdot\mathbf{l}_{\alpha;\beta} \quad \text{or} \quad \Gamma^{\mu}{}_{\alpha\beta}\equiv\langle\lambda^{\mu},l_{\alpha;\beta}\rangle$$

appear in expressions for covariant derivatives. In Riemannian geometry the Γ's are related to the structure coeficients by the zero torsion condition:

$$\mathbf{l}_{\alpha;\beta}-\mathbf{l}_{\beta;\alpha}\equiv-[\mathbf{l}_\alpha,\mathbf{l}_\beta]=-C^{\mu}{}_{\alpha\beta}\mathbf{l}_\mu,$$

$$\Gamma^{\mu}{}_{\alpha\beta}-\Gamma^{\mu}{}_{\beta\alpha}=-C^{\mu}{}_{\alpha\beta}.$$

With μ set equal to 4, this gives the vanishing of $\Gamma_{k\,44}$ and $\Gamma_{4[km]}$ with the kinematic significance of zero acceleration and zero vorticity as described in the text.

[1]T. R. Michalik and M. A. Melvin, "Spatially homogeneous neutrino cosmologies" (SHNC), J. Math. Phys. **21**, 1952 (1980). Note that the time coordinate is labeled 4 here and 0 in SHNC. The general equations for connection forms and curvature forms, without restricting the structure constant matrix to have the canonical structure, are given in Appendix I of SHNC. The reduction to the canonical structures by Behr does not seem to have ever been published, so that we have had to work it out for ourselves and believe others will find the analysis useful. Also we do not know if any one previously has arrived at the canonical structures for the metric-orthonormalizing transformations listed in Table IV and the metric structure—generalization of Robertson–Walker to all Bianchi Types—listed in Table V. [Note by M.A.M.: Upon acceptance of this paper for publication attention was drawn to the paper by M.A.H. MacCallum "On diagonal Bianchi cosmologies", Phys. Lett. A **40**, 385 (1972). The development there is based on the Ellis–MacCallum basis which differs from our EWB basis. I apologize to my good friend Malcolm MacCallum (a) for not having discovered his paper earlier, and (b) for having to point out that a specialization in his assumptions has limited his results. By assuming that for a diagonal metric "obviously the shear is diagonal and the rotation of the chosen axes relative to axes Fermi-Propagated along $\{x^{\nu}=\text{const};\ \nu=1,2,3\}$ is zero," he has found diagonal metrics only in certain specialized "Class B" {vector type} cases. As our analysis shows, while the zero-rotation condition is sufficient for a diagonal metric, it is not necessary. With our choice of gauge, preserving the quasi-canonical form of **F** while allowing for dilation-rotation of the quasi-canonical basis with respect to the invariant symmetry basis, *the metric may be maintained diagonal in all cases.*]

Added in proof: A propos of the diagonal-shear question and the Double-conjugate-pair Theorem, in a more general treatment than the one given in the present paper, I have shown that the relation $\mathbf{S}=-\mathbf{R}(\ln\mathbf{D})\,\mathbf{R}^{-1}$ holds under all conditions. This yields immediately the results detailed in Table IV. S is automatically diagonal if the orthonormal triad remains lined up with the invariant triad ($R=R_0$), e.g., for all intrinsically anisotropic cases. Further, if for instance $\alpha=\beta$, S remains diagonal in the $x-y$ plans for all R_z rotations, etc. These considerations will be discussed elsewhere.

[2]L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, New Jersey, 1926).

[3]E. M. Lifshitz and I. M. Khalatnikov, "Investigations in relativistic cosmology," Adv. Phys. **12**, 185 (1963).

[4]M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, New Jersey, 1975).

[5]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, (Addison–Wesley, Reading, Massachusetts, 1962, 1971, 1975), 2nd, 3rd, and 4th eds.

[6]H. P. Robertson, "Kinematics and world structure," Astrophys. J. **82**, 284 (1935); A. G. Walker, "On Riemannian spaces with spherical symmetry about a line and the conditions for isotropy in general relativity,"Quart. J. Math. Oxford Ser. **6**, 81 (1935).

[7]L. Bianchi "On Three-dimensional spaces which admit a group of motions," Societa Italiana Della Scienze Roma, Memorie de matematica e difisica (Dei. XL) (3) **11**, 267 (1897). A modern derivation independent of knowledge of Bianchi's work is given by J. Milnor, "Curvatures of Left Invariant Metrics on Lie Groups," Adv. Math. **21**, 239 (1976). A glossary of identification of the respective groups is given in Milnor's paper.

[8]R. Kantowski and R. K. Sachs, "Some spatially homogeneous anisotropic relativist cosmological models," J. Math. Phys. **7**, 443 (1966).

[9]F. B. Estabrook, H. D. Wahlquist, and C. G. Behr, "Dyadic analysis of spatially homogeneous world models," J. Math. Phys. **9**, 497 (1968).

[10]G. F. R. Ellis and M. A. H. MacCallum "A class of homogeneous cosmological models," Commun. Math. Phys. **12**, 108 (1969).

[11]A. H. Taub, "Empty space-times admitting a three parameter group of motions," Ann. Math. **53**, 472 (1951).

[12]M. A. Melvin, "Cylindrical symmetries and isotropies in the spatially homogeneous Bianchi cosmologies," *Proceedings of the 8th International Conference on General Relativity and Gravitation*, Waterloo, Ontario [GR8], 380, 1977.

[13]Lightman, Press, Price, and Teukolsky, *Problem Book in Relativity and Gravitation* (Princeton U.P., Princeton, New Jersey, 1975).

Spatially homogeneous neutrino cosmologies

T. R. Michalik
Randolph–Macon Woman's College, Lynchburg, Virginia 24504

M. A. Melvin
Temple University, Philadelphia, Pennsylvania 19122

(Received 3 August 1977; accepted for publication 21 November 1979)

Gravitational effects of a c number neutrino field in spatially homogeneous geometries are analyzed. At the basis of the description of the geometries is an arbitrary timelike congruence with aligned tetrad reference frames, i.e., one basis vector of the tetrad frame at each point of space–time is aligned along the congruence. The remaining three basis vectors of the tetrad lie in orthogonal spacelike hypersurfaces, and for these there are three alternative choices: (1) *invariant reference triads*, i.e., which conform to the isometry group of the space with structure constants C^c_{gh}; (2) *orthonormal reference triads* simulating a Euclidean structure at each point; (3) *holonomic coordinate triads* in terms of which metrics are usually given. The relations between (1), (2), and (3) for all nine Bianchi–Behr types are reviewed. The relatively compact vector–dyadic formalism used by Estabrook, Wahlquist, and Behr is introduced; it has the advantage that it gives directly the Bianchi–Behr classification of cosmologic solutions, and represents these solutions in terms of the orthonormal frame components of interesting physical quantities such as shear rates, volume expansion rates, and rotation with respect to inertially stabilized directions. The coupled Einstein–Dirac equations for the gravitational neutrino field and the neutrino stress-energy tensor are derived, and the coupling equations are replaced by a set in which the neutrino amplitude is eliminated, and instead there appear the neutrino current $(\mathbf{J}, J_0 = |\mathbf{J}|)$ (vector bilinear covariant) and the "eikonal momentum" $\mathbf{P} = \nabla f$, the gradient of the phase function $f(x)$ which appears in the neutrino spinor amplitude $\psi = \chi(t)e^{if(x)}$. Solutions for all nine Bianchi types are classified. Further, with the two assumptions: (1) axial symmetry of the geometry, (2) alignment of the neutrino momentum density $\mathbf{t}$ with the Bianchi vector $\mathbf{n}$ (in the symmetry direction), a new exact solution is found explicitly. It is of type VII_h. This solution goes over into a solution for type V when a parameter in the solution is set equal to zero. This is to be expected from the fact that type V may be regarded as a limiting case of type VII_h when $a = b \to 0$.

1. INTRODUCTION

This investigation of the neutrino in general relativity takes place on the "third level" [1] of approach which treats the neutrino as a classical ("c-number") field satisfying the coupled Einstein–Dirac equations. Here, the Dirac equation is a field equation which can be derived by varying the field Lagrangian. Variation of the general relativistic Lagrangian density $(\sqrt{-g})[\mathbf{R} + \mathbf{L}_{\text{field}}]$ with respect to the metric gives the neutrino stress-energy tensor as a source in Einstein's equations.[2]

Essentially, this approach studies the effect of single neutrino states on the geometry of spacetime. Such a neutrino state would need to be one of tremendous energy ($\sim 10^{78}$ GeV) to generate even the small cosmologic curvature of the present universe (radius $\sim 10^{10}$ light years). It is certainly of more physical interest to examine the gravitational effects of a collection of less energetic neutrinos as manifested by a many-neutrino state and stress-energy tensor taking account of statistics. Unfortunately, an exact many-neutrino treatment in general relativity is difficult.

The role of the c-number field theory is to provide a single-particle wave equation whose solutions are eigenfunctions which should form a complete set of single-particle states into which the general many-particle state is expanded. Statistics can be made an inherent part of the theory by choosing the expansion coefficients in such a way that the general many-particle wavefunction has the proper symmetries under the interchange of pairs of particles. Second quantization puts the theory into operator language. It generalizes from the many-particle wavefunction to a *quantum field*, with the expansion coefficients now becoming operators which satisfy commutation or anticommutation relations appropriate to the statistics being considered. Thus, in the context of a second quantized theory, the role of the c-number field equation is to provide the eigenfunctions which multiply the creation and annihilation operators in the expansion of the general many-particle field.

A quantized field theory for fermions has been developed for one particular geometry, that of flat space. There are infinitely many other geometries to consider, and the quantization procedure may be different for each. In particular, the symmetries or Killing vector fields of the spacetime play an important role in determining conserved quantities, and these, in turn, give information about which quantum numbers are relevant. It is also conceivable that the spacetime under consideration may not possess any Killing vectors. In the nonstatic spatially homogeneous cosmologies, for example, there are, in general, no timelike Killing vectors. The global energy associated with a localized stress-energy–momentum distribution is therefore not a conserved quantity, and the general time dependent state cannot be

 0022-2488/80/071952-13$1.00

expanded in energy eigenstates. There are, however, other symmetries. This investigation considers the neutrino field only in those spacetimes which possess the symmetries of the spatially homogeneous cosmologies.

Construction of an exact many-particle solution of the Einstein–Dirac equations out of single particle solutions is seriously hindered by the nonlinearity of Einstein's equations. The situation is not entirely hopeless, however. In some cases it is possible to treat the general relativistic many-fermion system with success in a self-consistent manner.[3]

Although the work here is not a many-neutrino theory, the study of the gravitational effects of single neutrino states is a first step towards a more relevant many-neutrino theory.

2. BIANCHI COSMOLOGIES

A. Invariant Basis to Orthonormal Tetrad Field

It is well known[4] that the geometry of the Bianchi cosmologies can be described by the metric

$$\mathbf{d}s^2 = -\mathbf{d}t^2 + \mathbf{d}\sigma^2, \tag{2.1}$$

where the infinitesimal squared distance within the hypersurfaces of homogeneity is given by

$$\mathbf{d}\sigma^2 = \gamma_{AB}(t)\boldsymbol{\epsilon}^A \otimes \boldsymbol{\epsilon}^B \tag{2.2}$$

(Roman letters 1–3, Greek 0–3). The metric coefficients γ_{AB} are purely time dependent, and the 1-forms $\boldsymbol{\epsilon}^A$ constitute an invariant form basis while their duals, the vectors $\mathbf{e}_A$, constitute an invariant vector basis.[5]

The $\mathbf{e}_A$, viewed as operators, satisfy a Lie algebra with structure constants which are the *negatives* of those of the isometry (Killing vector) Lie algebra of the different Bianchi–Behr types:

$$[\mathbf{e}_A, \mathbf{e}_B] = \mathsf{G}^R{}_{AB}\mathbf{e}_R, \quad \mathbf{e}_A \equiv \mathbf{e}_A^i \frac{\partial}{\partial x^i}, \tag{2.3}$$

where the holonomic coordinate tetrad and its dual

$$\left(\frac{\partial}{\partial x^0} \equiv \frac{\partial}{\partial t}, \frac{\partial}{\partial x^a}\right) \quad (\mathbf{d}x^0 \equiv \mathbf{d}t, \mathbf{d}x^a) \tag{2.4}$$

are not specified initially but are to be found later as suitable bases for the e's and $\boldsymbol{\epsilon}$'s. Following the Estabrook, Wahlquist, and Behr convention,[6] the $\mathbf{e}_A$ are chosen to give the canonical group algebra:

$$\begin{aligned}&[\mathbf{e}_1, \mathbf{e}_{II}] = c_0\mathbf{e}_{III}, \quad [\mathbf{e}_{II}, \mathbf{e}_{III}] = a_0\mathbf{e}_1 - n_0\mathbf{e}_{II},\\ &[\mathbf{e}_{III}, \mathbf{e}_1] = n_0\mathbf{e}_1 + b_0\mathbf{e}_{II},\end{aligned} \tag{2.5}$$

where the constants a_0, b_0, c_0, and n_0 have the values for the different Bianchi–Behr types given in Table I.

Instead of the invariant frames, it is advantageous to introduce a field of orthonormal frames (i.e., frames which put $\mathbf{d}\sigma^2$ in diagonal form). This consists of a homogeneous field of spacelike triads of orthogonal basis vectors $\mathbf{l}_a$ which lies within the successive hypersurfaces of homogeneity and a timelike vector field $\mathbf{l}_0$ which is orthogonal to these hypersurfaces. The *orthonormal basis vectors and their dual 1-forms* $\boldsymbol{\lambda}^a$ and $\boldsymbol{\lambda}^0$ are related to the invariant basis quantities by

$$\boldsymbol{\lambda}^0 \equiv \mathbf{d}t, \quad \boldsymbol{\lambda}^a \equiv \widetilde{C}_B{}^a\boldsymbol{\epsilon}^B, \quad \mathbf{l}_a \equiv C_a{}^B\mathbf{e}_B, \quad \mathbf{l}_0 \equiv \frac{\partial}{\partial t}, \tag{2.6}$$

TABLE I. Values of structure constants for the different Bianchi–Behr types
($h \equiv n_0^2/a_0b_0$).

Bianchi–Behr type	n_0	a_0	b_0	c_0
I	0	0	0	0
II	0	1	0	0
VI_0	0	1	−1	0
VII_0	0	1	1	0
VIII	0	1	1	−1
IX	0	1	1	1
V	1	0	0	0
IV	1	1	0	0
III or VI_{-1}	1	1	−1	0
$VI_h (0 > h > -\infty)$	$n_0 \neq 1$	1	−1	0
$VII_h (0 < h < \infty)$	n_0	1	1	0

where $C_b{}^A$ and $\widetilde{C}_B{}^a$ are matrices of a real affine transformation. They depend only on time and satisfy

$$\widetilde{C}_R{}^a C_b{}^R = \delta^a_b. \tag{2.7}$$

The orthonormality requires

$$\gamma_{AB}C_c{}^A C_d{}^B = \delta_{cd} \quad \text{or} \quad \gamma_{AB}(t) = \delta_{cd}\widetilde{C}_A{}^c\widetilde{C}_B{}^d. \tag{2.8}$$

Note that there is no reason to impose symmetry conditions on $C_b{}^A$ at this point. In fact, such an imposition as occurs, for instance, in the treatise of Ryan and Shepley[5] substantially reduces the generality of the analysis as is demonstrated in the companion paper (SMHC) where a classification of the possible cases is given.[4]

Upon going over to the orthonormal frame, the invariant basis structure constants of the different isometry groups ($\mathsf{G}^C{}_{AB}$) are transformed into commutation coefficients for the spatial subset of the orthonormal tetrad. These commutation coefficients are functions only of time according to the transformation

$$\mathsf{C}^f{}_{gh} \equiv \langle \boldsymbol{\lambda}^f, [\mathbf{l}_g, \mathbf{l}_h] \rangle = \mathsf{G}^R{}_{AB}\widetilde{C}_R{}^f C_g{}^A C_h{}^B. \tag{2.9}$$

(following Misner, Thorne and Wheeler, $\langle\ \rangle$ denotes the contraction of a 1-form with a vector.)

The remaining commutation coefficients (those involving $\mathbf{l}_0$) can be computed by taking the exterior derivatives of the orthonormal basis 1-forms

$$\mathbf{d}\boldsymbol{\lambda}^\alpha = -\mathsf{C}^\alpha{}_{|\mu\nu|}\boldsymbol{\lambda}^\mu \wedge \boldsymbol{\lambda}^\nu, \tag{2.10}$$

where $|\mu\ \nu|$ means an ordering of indices so that $\mu < \nu$. Because $\boldsymbol{\lambda}^0 = \mathbf{d}t$

$$\mathbf{d}\boldsymbol{\lambda}^0 = -\mathsf{C}^0{}_{|\mu\nu|}\boldsymbol{\lambda}^\mu \wedge \boldsymbol{\lambda}^\nu = 0. \tag{2.11}$$

Also, $\boldsymbol{\lambda}^a = \widetilde{C}_B{}^a\boldsymbol{\epsilon}^B = \widetilde{C}_B{}^a\boldsymbol{\epsilon}^B_i\mathbf{d}x^i$ so that

$$\begin{aligned}\mathbf{d}\boldsymbol{\lambda}^a &= \dot{\widetilde{C}}_D{}^a C_b{}^D\boldsymbol{\lambda}^0 \wedge \boldsymbol{\lambda}^b - \mathsf{C}^a{}_{|cb|}\boldsymbol{\lambda}^c \wedge \boldsymbol{\lambda}^b\\ &= -\mathsf{C}^a{}_{0b}\boldsymbol{\lambda}^0 \wedge \boldsymbol{\lambda}^b - \mathsf{C}^a{}_{|cb|}\boldsymbol{\lambda}^c \wedge \boldsymbol{\lambda}^b.\end{aligned} \tag{2.12}$$

The overdot indicates differentiation with respect to time. The result of (2.11) and (2.12) is that $\mathsf{C}^0{}_{\mu\nu} = 0$ and $\mathsf{C}^a{}_{0b} = -\dot{\widetilde{C}}_D{}^a C_b{}^D$. So all commutation coefficients are either zero or functions of time alone. Any quantity capable of be-

ing expressed in terms of the commutation coefficients will therefore be constant on a given hypersurface of homogeneity.

B. The Dyadic Formalism

The geometry of spacetime will be described with the dyadic formalism of Estabrook, Wahlquist, and Behr.[4,6] This formalism, although generally applicable, is particularly useful when a preferred timelike congruence exists. It is advantageous here because it manifests the spatial homogeneity, permits the immediate Bianchi-Behr classification of cosmologic solutions, and represents these solutions in terms of the orthonormal frame components of interesting physical quantities such as shear and volume expansion. It will be introduced here in a way that makes contact with more familiar quantities.

In general, with a timelike congruence, there are six dyadics: **a**, **Ω**, **ω**, **S**, **n**, and **F**. We define these in three equivalent ways, in terms of (1) the Ricci rotation coefficients ($\mathbf{l}_\mu\cdot\nabla_{\mathbf{l}_\nu}\mathbf{l}_\nu = \Gamma_{\mu\nu\alpha} = -\Gamma_{\nu\mu\alpha}$), (2) the commutation coefficients $C^\alpha{}_{\mu\nu}$, or (3) contractions of the connection 1-forms $\bar{\omega}_{\mu\nu} = \Gamma_{\mu\nu\alpha}\lambda^\alpha$ with the orthonormal tetrad $\mathbf{l}_\mu$.

a = the triad components of the acceleration of the timelike congruence (or world lines of matter with 4-velocity $\mathbf{l}_0$):

$$\begin{aligned} a_a &= \Gamma_{a00} = -\Gamma_{0a0} \quad \text{(Ricci)},\\ a_a &= -C^0{}_{a0} = C^0{}_{0a} \quad \text{(commutation)},\\ a_a &= -\langle\bar{\omega}_{0a},\mathbf{l}_0\rangle \quad \text{(connection)}. \end{aligned} \tag{2.13}$$

Ω (or $\mathbf{V}\equiv\mathbf{I}\times\mathbf{\Omega}$) = the triad components of the angular velocity of the congruence (or world lines of matter with 4-velocity $\mathbf{l}_0$) relative to a Fermi–Walker transported triad. **Ω** (or **V**) is called the vorticity (measures deviation from 3-normality of the timelike congruence):

$$\begin{aligned} \Omega_a &= \tfrac{1}{2}\epsilon_{abc}\Gamma_{0bc} \quad \text{(Ricci)},\\ \Omega_a &= -\tfrac{1}{4}\epsilon_{abc}C_{0cb} \quad \text{(commutation)},\\ V_{ab} &= \begin{bmatrix} 0 & -\Omega_3 & \Omega_2\\ \Omega_3 & 0 & -\Omega_1\\ -\Omega_2 & \Omega_1 & 0 \end{bmatrix} = -\langle\bar{\omega}_{0[a},\mathbf{l}_{b\,]}\rangle \end{aligned}$$

(connection). (2.14)

It follows from (2.11) that $\mathbf{a} = \mathbf{\Omega} = 0$ for Bianchi cosmologies. The timelike congruence is therefore geodesic and does not rotate with respect to Fermi–Walker transport. [From here on **a** and **Ω** will be omitted from our consideration except for brief appearances in (3.5) and Appendix III where they are included for completeness.]

ω (or $\mathbf{W}\equiv\mathbf{I}\times\boldsymbol{\omega}$) = the triad components of the angular velocity of the orthonormal triad $\mathbf{l}_a$ relative to a Fermi–Walker transported triad:

$$\begin{aligned} \omega_a &= \tfrac{1}{2}\epsilon_{abc}\Gamma_{cb0} \quad \text{(Ricci)},\\ \omega_a &= \tfrac{1}{2}\epsilon_{abc}C_{bc0} \quad \text{(commutation)},\\ W_{ab} &= \begin{pmatrix} 0 & -\omega_3 & \omega_2\\ \omega_3 & 0 & -\omega_1\\ -\omega_2 & \omega_1 & 0 \end{pmatrix} = \langle\bar{\omega}_{ab},\mathbf{l}_0\rangle \end{aligned}$$

(connection). (2.15)

S = the triad components of the rate of strain tensor. **S** is not traceless and therefore includes the shear **σ** and the volume expansion $\theta = \mathrm{Tr}\mathbf{S}$:

$$\begin{aligned} S_{ab} &= -\Gamma_{0(ab)} \quad \text{(Ricci)},\\ S_{ab} &= C_{(ab)0} \quad \text{(commutation)},\\ S_{ab} &= -\langle\bar{\omega}_{0(a},\mathbf{l}_{b)}\rangle \quad \text{(connection)}. \end{aligned} \tag{2.16}$$

The dyadic **S** describes the extrinsic curvature of the hypersurface of homogeneity. In particular, if the separation of two nearby members of the timelike congruence is described by the separation vector $\mathbf{r} = r^a\mathbf{l}_a$, then the time rate of change of this separation is given by[7]

$$\dot{\mathbf{r}} = -\mathbf{r}\cdot\mathbf{S} + \mathbf{r}\times\boldsymbol{\omega}, \tag{2.17}$$

where the dot and the cross operator have the vector-dyadic meanings. From (2.17) one immediately gets the generalized Hubble equation for the relative rate of change of the magnitude of **r** and an equation giving the rate of change of the unit vector in the direction of **r** ($\hat{\mathbf{r}}$):

$$-\frac{\dot{r}}{r} = \hat{\mathbf{r}}\cdot\boldsymbol{\sigma}\cdot\hat{\mathbf{r}} + \tfrac{1}{3}\theta, \tag{2.18}$$

$$-\dot{\hat{\mathbf{r}}} = \hat{\mathbf{r}}\cdot\boldsymbol{\sigma} + \frac{\dot{r}}{r}\hat{\mathbf{r}} + \tfrac{1}{3}\theta\hat{\mathbf{r}} - \hat{\mathbf{r}}\times\boldsymbol{\omega}, \tag{2.19}$$

where θ is defined above, $\boldsymbol{\sigma}\equiv\mathbf{S} - \tfrac{1}{3}\theta\mathbf{I}$ is the shear dyadic, and **I** is the unit dyadic.

The two remaining dyadics **F** and **n** are defined as follows ($\Phi\equiv\mathrm{Tr}\mathbf{F}$):

$$\begin{aligned} [\mathbf{F} - \tfrac{1}{2}\Phi\mathbf{I}]_{ad} &= \tfrac{1}{2}\epsilon_{cb(d}\Gamma_{a)}{}^{cb} \quad \text{(Ricci)},\\ F_{ad} &= \tfrac{1}{2}C_{(a}{}^{cb}\epsilon_{d)cb} \quad \text{(commutation)},\\ n_a &= \tfrac{1}{2}\Gamma_{ab}{}^b \quad \text{(Ricci)},\\ n_a &= \tfrac{1}{2}C^b{}_{ab} \quad \text{(commutation)}. \end{aligned} \tag{2.20}$$

$\mathbf{F} - \tfrac{1}{2}\Phi\mathbf{I}$ and $\mathbf{n}\times\mathbf{I}$ (or $\epsilon_{ced}n^e$) are the respective symmetric and antisymmetric parts of a dyadic

$$\mathbf{A} - \tfrac{1}{2}\Phi\mathbf{I} \equiv \mathbf{F} + \mathbf{n}\times\mathbf{I} - \tfrac{1}{2}\Phi\mathbf{I}.$$

The commutation triadic C'_{abd} is then

$$\mathbf{C}'_{abd} = [\mathbf{I}\times\mathbf{A}]_{abd} = A_{af}\epsilon_{fbd},$$

which is related to the connection and to **F** and **n** by

$$C'_{abd} = -\langle\mathbf{l}_a,\bar{\boldsymbol{\omega}}_{bd}\rangle = F_{af}\epsilon_{fbd} + \delta_{ab}n_d - \delta_{ad}n_b. \tag{2.21}$$

It will be seen that **F** and **n** are time evolutions of the structure constants a_0, b_0, c_0, and n_0 appearing in the canonical form (2.5) of the Lie algebras. The Jacobi identity satisfied by the structure constants evolves into the relation between commutation coefficients $\mathbf{F}\cdot\mathbf{n} = 0$.

F and **n** are related to the intrinsic curvature of the hypersurfaces of homogeneity. In fact, the Riemann curvature tensor of the 3-geometry on the hypersurfaces is entirely composed out of **F** and **n** as is shown in Appendix I, where all curvature quantities and the Ricci and Einstein tensors are derived.

From the analysis in Appendix I it is clear that all quantities of interest can be computed from a knowledge of **F**, **S**, **n**, and **ω**. The freedom to choose a convenient orthonormal frame exists, and a particular choice of frame will determine **ω**. Thus, the variables **F**, **S**, and **n** are all that are needed to

specify the geometry and the geometrical quantities in Einstein's equations.

The components of Einstein's equations ($G_{\mu\nu} = -2T_{\mu\nu}$ with $c = 1$ and $G = 1/4\pi$) become (see Appendix I)

$$\rho = \tfrac{1}{4}[\theta^2 - \mathbf{S}:\mathbf{S} - \mathbf{F}:\mathbf{F} + \tfrac{1}{2}\Phi^2 - 6n^2] \quad (-\tfrac{1}{2}G_{00} = \rho), \tag{2.22}$$

$$\mathbf{t} = \mathbf{F}\times\mathbf{S} - \theta\mathbf{n} + 3\mathbf{n}\cdot\mathbf{S} \quad (\tfrac{1}{2}G_{0f} = t_f), \tag{2.23}$$

$$\begin{aligned}\dot{\mathbf{S}} = {} & 2\mathbf{T} + \mathbf{S}\times\boldsymbol{\omega} - \boldsymbol{\omega}\times\mathbf{S} \\ & - \theta\mathbf{S} - 2\mathbf{F}\times\mathbf{F} + \Phi\mathbf{F} + \mathbf{F}\times\mathbf{n} - \mathbf{n}\times\mathbf{F} \\ & + [-(\mathrm{Tr}\mathbf{T}) + \tfrac{1}{4}\theta^2 - \tfrac{1}{4}\mathbf{S}:\mathbf{S} \\ & + \tfrac{3}{4}\mathbf{F}:\mathbf{F} - \tfrac{3}{8}\Phi^2 + \tfrac{1}{2}n^2]\mathbf{I} \quad (\mathbf{G} = -2\mathbf{T}).\end{aligned} \tag{2.24}$$

Here $\rho \equiv T_{00}$ is the energy density, $(\mathbf{t})^a \equiv T^{0a}$ is the momentum density, and $\mathbf{T}$ represents the dyadic of the components of $T_{\mu\nu}$ with spacelike indices. By differentiating (2.22) and (2.23) one arrives at the contracted Bianchi identities

$$\dot{\rho} = -\theta\rho - \mathbf{T}:\mathbf{S} + 2\,\mathbf{n}\cdot\mathbf{t}, \tag{2.25}$$

$$\dot{\mathbf{t}} = -\theta\mathbf{t} + \mathbf{t}\times\boldsymbol{\omega} - \mathbf{t}\cdot\mathbf{S} + 3\mathbf{n}\cdot\mathbf{T} - (\mathrm{Tr}\mathbf{T})\mathbf{n} + \mathbf{F}\times\mathbf{T}. \tag{2.26}$$

From (2.12) the following differential equations for the time-dependent transformation matrix $C_b{}^A$ and its inverse $\tilde{C}_B{}^a$ are obtained:

$$\dot{\tilde{C}}_R{}^a = -C^a{}_{0b}\tilde{C}_R{}^b = \tilde{C}_R{}^b[\mathbf{S} + \mathbf{W}]_b{}^a, \tag{2.27}$$

$$\dot{C}_d{}^A = -[\mathbf{S} + \mathbf{W}]_d{}^b C_b{}^A. \tag{2.28}$$

Solution of (2.27) allows the metric γ_{ab} to be computed from (2.8). Then, using the canonical basis triads of Appendix II, the metric may be written in a coordinate (holonomic) frame. As mentioned earlier, the imposition of symmetry on $C_b{}^A$ reduces the generality of the analysis because such an imposition must be consistent with (2.28). Symmetry of $C_b{}^A$ and its consistency with (2.28) implies conditions on $\mathbf{S}$ and $\boldsymbol{\omega}$ which vary with Bianchi–Behr type, but are specializations away from a completely general analysis. Such a specialization is made implicitly in the book by Ryan and Shepley[5] who follow Misner. A detailed discussion is given in the companion paper (SMHC).[4]

It is possible to relate $\mathbf{F}$ and $\mathbf{n}$ to the canonical set of ones and zeros in Table I because in the orthonormal frame $\mathbf{F}$ and $\mathbf{n}$ are expressed in terms of the commutation coefficients by (2.20). Using (2.9) one can express the orthonormal frame commutation coefficients $C^a{}_{gh}$ in terms of the invariant frame commutation coefficients $G^A{}_{GH}$ (group structure constants a_0, b_0, c_0, n_0):

$$F^{ab}(t) = F_{ab}(t) = \Delta\tilde{C}_K{}^a\tilde{C}_M{}^b F_0^{KM}, \tag{2.29}$$

$$n_g(t) = C_g{}^A n_{0A}, \tag{2.30}$$

where $\Delta \equiv \det C_b{}^A$, $F_0^{KM} \equiv \mathrm{diag}\,(a_0, b_0, c_0)$, and $n_{0A} \equiv (0, 0, n_0)$. It is now clear from (2.29) and (2.30) that $\mathbf{F}$ and $\mathbf{n}$ are simply time evolutions of the group structure constants $G^a{}_{gh}$. Differential equations for $\mathbf{F}$ and $\mathbf{n}$ may be derived by simply differentiating (2.29) and (2.30) and using (2.27) and (2.28):

$$\dot{\mathbf{F}} = -\theta\mathbf{F} + \mathbf{S}\cdot\mathbf{F} + \mathbf{F}\cdot\mathbf{S} + \mathbf{F}\times\boldsymbol{\omega} - \boldsymbol{\omega}\times\mathbf{F}, \tag{2.31}$$

$$\dot{\mathbf{n}} = \mathbf{n}\times\boldsymbol{\omega} - \mathbf{n}\cdot\mathbf{S}. \tag{2.32}$$

Equations (2.31) and (2.32) are independent of Einstein's equations and do not involve the stress-energy tensor at all (except through the coupling to Einstein's equations).

The solution for a given cosmologic model, then, involves the specification of the energy content of the universe in the stress-energy dyadics, and the solution of Eqs. (2.22)–(2.28), and (2.31) and (2.32).

Finally, the Bianchi–Behr classification can be reproduced in the orthonormal frame. From (2.30), the vanishing or nonvanishing of $\mathbf{n}_0$ implies the vanishing or nonvanishing of $\mathbf{n}(t)$ so that the Bianchi cosmologies are readily classified into vector ($\mathbf{n} \neq 0$) and nonvector ($\mathbf{n} = 0$) types. Also, if $F_0^{KM} = 0$ (types I and V), then $\mathbf{F} = 0$ from (2.29).

It is convenient and allowable for types other than I and V to choose the orthonormal frame such that $\mathbf{F}$ is diagonal and $\mathbf{n}$ is in the third direction. (Recall that the evolving Jacobi identity is written $\mathbf{F}\cdot\mathbf{n} = 0$.) For types I and V the orthonormal frame is chosen to make $\mathbf{S}$ diagonal because $\mathbf{F} = 0$ for these types. Let the special choices for $\mathbf{F}$ and $\mathbf{n}$ be written: $\mathbf{F} = \mathrm{diag}\,(a, b, c)$ and $\mathbf{n} = (0, 0, n_3)$.

From (2.31) and (2.32) the signs of a, b, c and n are *conserved*[6] as the cosmologic model evolves unless the solution passes through a singularity. Thus, using Table I one arrives at the Bianchi–Behr (Table II) classification in the orthonormal frame based on the conserved signs of a, b, c, and n^2.

3. THE DIRAC EQUATION AND NEUTRINO CURRENT

The neutrino will be described by the generalized Dirac equation

$$\gamma^\mu\nabla_\mu\Psi + m\Psi = 0, \tag{3.1}$$

where the rest mass m of the neutrino will be taken to be zero and the γ^μ are a set of 4×4 matrices satisfying $\{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu}$. Orthonormal frames are used exclusively here so that $g^{\mu\nu} = \eta^{\mu\nu} = (-1, 1, 1, 1)$ and the γ^μ are just the same as they are in special relativity. ∇_μ is the spinor covariant differentiation operator which when acting on the column spinor Ψ gives

TABLE II. The possible types of simply transitive 3-parameter isometry groups which act transitively on 3-spaces—the Bianchi–Behr types.

n^2	SIGN a	SIGN b	SIGN c	Bianchi–Behr type
0	0	0	0	I
0	+	0	0	II
0	+	−	0	VI_0
0	+	+	0	VII_0
0	+	+	−	VIII
0	+	+	+	IX
+	0	0	0	V
+	+	0	0	IV
+	+	−	0	III
+	+	−	0	VI_h
+	+	+	0	VII_h

$$\nabla_\mu \Psi = \langle \mathbf{d}\Psi, \mathbf{l}_\mu \rangle - \Gamma_\mu \Psi \tag{3.2}$$

and when acting on the row spinor $\bar{\Psi} = i\Psi^\dagger \gamma^0$ (adjoint spinor) gives

$$\nabla_\mu \bar{\Psi} = \langle \mathbf{d}\bar{\Psi}, \mathbf{l}_\mu \rangle + \bar{\Psi}\Gamma_\mu. \tag{3.3}$$

The spinor connection Γ_μ is defined so that the spinor covariant derivative of γ^μ (and hence $g^{\mu\nu}$) is zero. In the orthonormal frame Γ_μ is given by[8]

$$\Gamma_\mu = -\tfrac{1}{4}\Gamma_{\alpha\beta\mu}\gamma^\alpha\gamma^\beta + Y_\mu I, \tag{3.4}$$

where the $\Gamma_{\alpha\beta\mu}$ are the Ricci rotation coefficients, I is the unit matrix, and Y_μ is an arbitrary imaginary vector which is zero for neutral particles but is proportional to the electromagnetic vector potential for charged particles.

If a timelike congruence is introduced, and the dyadic formalism is applied, the spinor connection may be written in terms of the dyadic quantities and the Dirac equation in a general spacetime becomes

$$\gamma^\mu \langle \mathbf{d}\Psi, \mathbf{l}_\mu \rangle + \{\gamma^0[\tfrac{1}{2}\theta + \tfrac{1}{4}\Phi\gamma^5] - \boldsymbol{\gamma}\cdot[\tfrac{1}{4}\gamma^0(\boldsymbol{\omega}\times\boldsymbol{\gamma}) + \mathbf{n}]\}\Psi$$
$$+ \{\tfrac{1}{2}\boldsymbol{\gamma}\cdot[\mathbf{a} + \gamma^0(\boldsymbol{\Omega}\times\boldsymbol{\gamma})] - Y_\mu\gamma^\mu + m\}\Psi = 0. \tag{3.5}$$

The ordering of the γ^μ is important since not all of the γ^μ commute.

Specializing to the cases of a zero rest mass neutrino, subject to and generating the geometry of the Bianchi cosmologies, we have $m = \mathbf{a} = \boldsymbol{\Omega} = Y_\mu = 0$. Also

$$\gamma^\mu \langle \mathbf{d}\Psi, \mathbf{l}_\mu \rangle = \gamma^0\dot{\Psi} + (\boldsymbol{\gamma}\cdot{}^{(3)}\mathbf{d}\Psi), \tag{3.6}$$

where ${}^{(3)}\mathbf{d}\Psi$ is the exterior derivative of Ψ intrinsic to the hypersurface of homogeneity. The Dirac equation together with the equation for the adjoint spinor become

$$\dot{\Psi} = \gamma^0(\boldsymbol{\gamma}\cdot{}^{(3)}\mathbf{d}\Psi) + [-\gamma^0\boldsymbol{\gamma}\cdot\mathbf{n} + \tfrac{1}{2}(\boldsymbol{\omega}\cdot\boldsymbol{\Sigma}) - \tfrac{1}{2}\theta - \tfrac{1}{4}\Phi\gamma^5]\Psi, \tag{3.7}$$

$$\dot{\bar{\Psi}} = ({}^{(3)}\mathbf{d}\bar{\Psi}\cdot\boldsymbol{\gamma})\gamma^0 + \bar{\Psi}[-\tfrac{1}{2}(\boldsymbol{\omega}\cdot\boldsymbol{\Sigma}) + \gamma^0\boldsymbol{\gamma}\cdot\mathbf{n} - \tfrac{1}{2}\theta - \tfrac{1}{4}\Phi\gamma^5], \tag{3.8}$$

where $\gamma^5 \equiv \gamma^0\gamma^1\gamma^2\gamma^3$, and where we have introduced the matrices

$$\Sigma_k \equiv (\gamma_i\gamma_j - \gamma_j\gamma_i)/2 \quad (i,j,k = 1,2,3 \text{ } et \text{ } cycl).$$

The neutrino (Ψ_ν) and antineutrino ($\Psi_{\bar{\nu}}$) wavefunctions are obtained from Ψ through the use of projection operators:

$$\Psi_\nu = \tfrac{1}{2}(1 + i\gamma_5)\Psi, \quad \text{when} \quad (1 - i\gamma_5)\Psi_\nu = 0, \tag{3.9}$$

$$\Psi_{\bar{\nu}} = \tfrac{1}{2}(1 - i\gamma_5)\Psi, \quad \text{when} \quad (1 + i\gamma_5)\Psi_{\bar{\nu}} = 0, \tag{3.10}$$

Of course, this has the effect of reducing the four equations of (3.7) to two equations for the two independent components of Ψ_ν and $\Psi_{\bar{\nu}}$.

Whenever an explicit representation of the γ^μ is called for, the following real representation will be used:

$$\gamma^0 = \begin{bmatrix} 0 & i\sigma_2 \\ i\sigma_2 & 0 \end{bmatrix}, \quad \gamma^1 = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \quad \gamma^2 = \begin{bmatrix} 0 & \sigma_1 \\ \sigma_1 & 0 \end{bmatrix},$$

$$\gamma^3 = \begin{bmatrix} 0 & -\sigma_3 \\ -\sigma_3 & 0 \end{bmatrix}, \quad \gamma^5 = -\gamma_5 = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}, \tag{3.11}$$

where the σ_k are the Pauli matrices and $I = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. In this representation, Ψ_ν and $\Psi_{\bar{\nu}}$ are given by

$$\Psi = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ i\hat{\nu}\Psi_1 \\ i\hat{\nu}\Psi_2 \end{bmatrix}, \tag{3.12}$$

where $\hat{\nu} = +1$ for the neutrino wavefunction and $\hat{\nu} = -1$ for the antineutrino wavefunction. The neutrino–antineutrino subscripts on Ψ can be dropped from now on as it is understood that Ψ is represented by (3.12).

It is useful to obtain a set of differential equations for the bilinear covariants composed of $\bar{\Psi}$, γ^μ, and Ψ because the neutrino stress-energy tensor will contain only products of dyadics and bilinear covariants. If (3.9)–(3.11) are used, the only nonzero bilinear covariants are the vector ($V^\mu \equiv i\bar{\Psi}\gamma^\mu\Psi$) and the axial vector ($A^\mu = \bar{\Psi}\gamma^5\gamma^\mu\Psi$). Furthermore, V^μ and A^μ are not independent, but by virtue of (3.9) and (3.10) satisfy $A^\mu = -\hat{\nu}V^\mu$ so that A^μ can be entirely eliminated. V^μ is a pure imaginary lightlike vector

$$\mathbf{V}\cdot\mathbf{V} = (V_0)^2, \tag{3.13}$$

and if $J^\mu \equiv -iV^\mu$, then J^μ is a real current density ($J^0 \geqslant 0$) whose covariant divergence is zero and which is in principle measurable.

Now, if the spacetimes and their contents are to be homogeneous, then J^μ must depend only on time so that a measurement of J^μ will not provide a means of distinguishing one point from another on a hypersurface of homogeneity. The pure time dependence of J^μ can be accomplished by having $\Psi = \chi(t)e^{if(x,t)}$ where χ is a purely time-dependent spinor and f is a real function of space and time which will be determined following the discussion of the neutrino stress-energy tensor below.

To find the differential equations satisfied by J^μ, simply take the time derivative of $\bar{\Psi}\gamma^\mu\Psi$ and use (3.7) and (3.8) to get

$$\dot{J}^0 = -\theta J^0 + 2\mathbf{n}\cdot\mathbf{J}, \tag{3.14}$$

$$\dot{\mathbf{J}} = -\theta\mathbf{J} - \boldsymbol{\omega}\times\mathbf{J} + 2\mathbf{n}J^0 - 2\hat{\nu}(\mathbf{P}\times\mathbf{J}), \tag{3.15}$$

where ${}^{(3)}\mathbf{d}f \equiv \mathbf{P}$, the "eikonal momentum." The four equations (3.14) and (3.15) are not independent because of (3.13), but they will replace the four equivalent equations (also not independent) contained in (3.7). When a solution of (3.14) and (3.15) is known Ψ may be recovered if desired:

$$\Psi = \tfrac{1}{2}Q^{1/2}\begin{bmatrix} 1 \\ -U \\ i\hat{\nu} \\ -i\hat{\nu}U \end{bmatrix} e^{i\phi}e^{if(x,t)}, \tag{3.16}$$

where $Q \equiv J^0 - J^2$; $U \equiv (J^3 + i\hat{\nu}J^1)/Q$ and ϕ is a time-dependent phase factor which satisfies the following equation:

$$\dot{\phi} = -\dot{f} + P_2 - \frac{1}{Q}[P_1J_1 + P_3J_3]$$
$$- \frac{\hat{\nu}}{Q}[\tfrac{1}{2}(\omega_1J_1 + \omega_3J_3) + (\mathbf{n}\times\mathbf{J})_{13}] + \frac{\hat{\nu}}{2}\omega_2$$
$$- \frac{\hat{\nu}}{4}\Phi. \tag{3.17}$$

When a cosmologic model has been solved, the right-hand side of (3.17) will be a known function of time and ϕ can be found by integration of (3.17). Equation (3.16) is derived by solving $J^{\mu} = \bar{\Psi}\gamma^{\mu}\Psi$ for the components of the spinor χ in terms of the components J^{μ} and (3.17) is derived by requiring that (3.16) satisfy the Dirac equation.

The Stress-Energy Tensor: The neutrino stress-energy tensor is obtained by a standard variation of the neutrino field Lagrangian[9]:

$$T_{\mu\nu} = \frac{i}{2}\left[\bar{\Psi}\gamma_{(\mu}\nabla_{\nu)}\Psi - \{\nabla_{(\mu}\bar{\Psi}\}\gamma_{\nu)}\Psi\right] \quad (\hbar = 1). \tag{3.18}$$

This stress-energy tensor is derived and applied in the context of standard Riemannian geometry. We do not interpret spin density in a geometry with torsion here.

If the spinor covariant derivatives are removed with the aid of the Dirac equations (3.7) and (3.8) (specialized for the Bianchi cosmologies), and if the previously defined stress-energy dyadics are introduced, (3.18) becomes

$$\rho = \mathbf{P}\cdot\mathbf{J} + \frac{\hat{v}}{4}\Phi J^0, \tag{3.19}$$

$$\mathbf{t} = \mathbf{P}J^0 + \frac{\hat{v}}{2}(\mathbf{n}\times\mathbf{J}) - \frac{\hat{v}}{2}(\mathbf{F}'\cdot\mathbf{J}), \tag{3.20}$$

$$\mathbf{T} = \tfrac{1}{2}[\mathbf{P}\mathbf{J} + \mathbf{J}\mathbf{P}] + \frac{\hat{v}}{4}[\mathbf{J}\times\mathbf{S} - \mathbf{S}\times\mathbf{J}] - \frac{\hat{v}}{2}\mathbf{F}'J^0, \tag{3.21}$$

where $\mathbf{F}' \equiv \mathbf{F} - \tfrac{1}{2}\Phi\mathbf{I}$.

In the flat space limit, $\mathbf{P}$ becomes the space part of the lightlike momentum 4-vector $P^{\mu} = (E, \mathbf{P})$. The stress-energy dyadics in flat space become

$$\rho = \mathbf{P}\cdot\mathbf{J} = EJ^0, \tag{3.22}$$

$$\mathbf{t} = \mathbf{P}J^0, \tag{3.23}$$

$$\mathbf{T} = \tfrac{1}{2}[\mathbf{P}\mathbf{J} + \mathbf{J}\mathbf{P}]. \tag{3.24}$$

If both the spacetime and its contents are to be homogeneous, then (3.19)–(3.21) must depend only on time so that a measurement of $T_{\mu\nu}$ will yield no information about position on a hypersurface of homogeneity. Because J^{μ} and the dyadic quantities are purely time dependent, it is clear that $T_{\mu\nu}$ will be purely time dependent if $\mathbf{P}$ is a real function of time only. $\mathbf{P}$ is a 1-form whose orthonormal, invariant and holonomic components are given by

$$^{(3)}\mathbf{d}f \equiv \mathbf{P} \equiv P_b\boldsymbol{\lambda}^b = \tilde{C}_A{}^b P_b \boldsymbol{\epsilon}^A = \tilde{C}_A{}^b P_b \epsilon^A_j \mathbf{d}x^j = f_{,j}\mathbf{d}x^j. \tag{3.25}$$

In terms of the partial derivatives of f, the orthonormal (P_b), invariant (H_A), and holonomic (p_j) components of $\mathbf{P}$ are respectively

$$P_b = C_b{}^A e^j_A f_{,j}, \tag{3.26}$$

$$H_A = e^j_A f_{,j} \equiv \mathbf{e}_A f, \tag{3.27}$$

$$p_j = f_{,j}. \tag{3.28}$$

Now, a necessary (but not sufficient) condition that H_A (and hence P_b) be a function only of time is

$$[\mathbf{e}_A, \mathbf{e}_B]f = 0. \tag{3.29}$$

[The operator $\mathbf{e}_A$ which contains only partial derivatives with respect to spatial coordinates must give zero when applied to $H_B(t)$.] In general, the group algebra for the nine Bianchi types is given by (2.5) so that (3.29) becomes

$$c_0 H_{\mathrm{III}} = 0, \tag{3.30}$$

$$-n_0 H_{\mathrm{I}} - b_0 H_{\mathrm{II}} = 0, \tag{3.31}$$

$$a_0 H_{\mathrm{I}} - n_0 H_{\mathrm{II}} = 0. \tag{3.32}$$

By evaluating (3.30)–(3.32) for each Bianchi–Behr type according to Table I, and using the canonical triads of Appendix II to relate H_A and p_j for each type, the first column of Table III results. The possibly nonvanishing H_A and p_j are yet to be determined, but they can be at most functions of time. If this is so, then the general solution of (3.28) is

$$f = p_j X_j. \tag{3.33}$$

and the second column of Table III results. [The arbitrary function of time which could be added to (3.33) is assumed to be absorbed into the time dependent spinor $\chi(t)$.]

When $\Psi = \chi(t)e^{if}$ with f given by (3.33) is substituted into the Dirac equation, one gets an equation equivalent to (3.17) but in a more general form:

$$\dot{p}_j(t)X^j = i\frac{\dot{\chi}_1}{\chi_1} + P_2 + \frac{\hat{v}}{2}\omega_2 + \frac{i}{2}\theta + in_2 - \frac{\hat{v}}{4}\Phi + \frac{\chi_2}{\chi_1}\left[P_3 - i\hat{v}P_1 - \frac{i}{2}\omega_1 + \frac{\hat{v}}{2}\omega_3 + \hat{v}n_1 + in_3\right], \tag{3.34}$$

where χ_1 and χ_2 are the two independent components of the spinor χ. Because the rhs of (3.34) is a function of time alone, we must set $\dot{p}_j = 0$ in order to satisfy (3.34) for all x^j. Thus, the p_j and H_A are not functions of time but *constants*.

The vanishing or nonvanishing of the components of $\mathbf{P}$ in the invariant basis is related to the Lie derivatives of the invariant basis triad as follows: For all types *except type* III, if $[\mathbf{e}_A, \mathbf{e}_B] = 0$, then there exists a possibly nonzero, constant, invariant cth component of $\mathbf{P}$ where the indices a, b, and c are all different and undergo cyclic permutations. For type III, the Lie derivatives of $\mathbf{e}_{\mathrm{III}}$ with respect to $\mathbf{e}_{\mathrm{I}}$ and $\mathbf{e}_{\mathrm{II}}$ are not independent. In fact, $[\mathbf{e}_{\mathrm{I}}, \mathbf{e}_{\mathrm{III}}] = -[\mathbf{e}_{\mathrm{II}}, \mathbf{e}_{\mathrm{III}}]$. This additional symmetry not only admits the existence of possibly nonzero first and second invariant components of $\mathbf{P}$, but also implies the equality of these components.

TABLE III. Representations of H_A and p_j and the functional form of f for each Bianchi–Behr type.

Bianchi–Behr type	Possibly nonvanishing H_A and p_j	Functional form of f
I	$H_{\mathrm{I}} = p_1$; $H_{\mathrm{II}} = p_2$; $H_{\mathrm{III}} = p_3$	p_jX_j
II	$H_{\mathrm{II}} = p_3$; $H_{\mathrm{III}} = -p_1$	$p_1X_1 + p_3X_3$
III	$H_{\mathrm{I}} = H_{\mathrm{II}} = p_3$; $H_{\mathrm{III}} = 2p_1$	$p_1X_1 + p_3X_3$
IV	$H_{\mathrm{III}} = p_1$	p_1X_1
V	$H_{\mathrm{III}} = p_1$	p_1X_1
VI_0,VI_h ($h \neq -1$)	$H_{\mathrm{III}} = (n_0 + 1)p_1$	p_1X_1
VII_0,VII_h	$H_{\mathrm{III}} = (n_0^2 + 1)^{1/2}p_1$	p_1X_1
VIII	None	0
IX	None	0

A differential equation for $\mathbf{P}$ may be derived by taking the time derivative of (3.26) and using (2.28):

$$\dot{\mathbf{P}} = -\mathbf{S}\cdot\mathbf{P} - \boldsymbol{\omega}\times\mathbf{P}. \tag{3.35}$$

Two algebraic identities satisfied by $\mathsf{T}_{\mu\nu}$ are noted briefly[10]:

$$\mathsf{T}_{\mu\nu}J^{\mu}J^{\nu} \equiv \rho(J^0)^2 - 2(\mathbf{t}\cdot\mathbf{J})J^0 + \mathbf{J}\cdot\mathbf{T}\cdot\mathbf{J} = 0, \tag{3.36}$$

$$\mathsf{T}^{\mu}{}_{\mu} = \mathrm{Tr}\mathbf{T} - \rho = 0. \tag{3.37}$$

$\mathsf{T}_{\mu\nu}$ has several properties which make its physical interpretation difficult. First is the fact that the energy density ρ is not always positive definite. Of course, this problem exists even in the unquantized flat space theory. Examination of (3.19) for type IX ($\mathbf{P} = 0$, $\Phi > 0$) shows that $\rho > 0$ for the neutrino, but $\rho < 0$ for the antineutrino. For type VIII, $\mathbf{P} = 0$. If $|c| > a + b$, then $\rho > 0$ for the antineutrino, but $\rho < 0$ for the neutrino.

Next, the neutrino momentum density 4-vector $\mathsf{T}^{0\mu}$ is not always lightlike. In general,

$$\begin{aligned}\mathsf{T}^{0\mu}\mathsf{T}^{0}{}_{\mu} = & -\rho^2 + \mathbf{t}\cdot\mathbf{t} = P^2(J^0)^2 \\ & - (\mathbf{P}\cdot\mathbf{J})^2 + \hat{v}[\,\mathbf{P}\cdot(\mathbf{n}\times\mathbf{J})]J^0 \\ & + \tfrac{1}{4}(\mathbf{n}\times\mathbf{J})\cdot(\mathbf{n}\times\mathbf{J}) - \tfrac{1}{2}(\mathbf{n}\times\mathbf{J})\cdot(\mathbf{F}\cdot\mathbf{J}) \\ & - \hat{v}[\,\mathbf{P}\cdot\mathbf{F}\cdot\mathbf{J}]J^0 - \tfrac{1}{4}\Phi\,[\mathbf{J}\cdot\mathbf{F}\cdot\mathbf{J}] \\ & + \tfrac{1}{4}(\mathbf{F}\cdot\mathbf{J})\cdot(\mathbf{F}\cdot\mathbf{J}).\end{aligned} \tag{3.38}$$

Consider types VIII and IX where $\mathbf{P} = \mathbf{n} = 0$ and $\mathbf{F} = \mathrm{diag}\,(a, b, c)$. Equation (3.38) becomes

$$\begin{aligned}\mathbf{t}\cdot\mathbf{t} - \rho^2 = & -\tfrac{1}{4}[a(b+c)(J_1)^2 + b(c+a)(J_2)^2 \\ & + c(a+b)(J_3)^2].\end{aligned} \tag{3.39}$$

For type IX, $\mathsf{T}^{0\mu}$ is timelike ($\mathsf{T}^{0\mu}\mathsf{T}^{0}{}_{\mu} < 0$) because a, b, and c are all positive, but for type VIII, $\mathsf{T}^{0\mu}$ can be spacelike if $c \leqslant -a$ and $c \leqslant -b$.

The simple relation (3.23) between $\mathbf{t}$ and $\mathbf{P}$ is lost in the general expression (3.20) so although $\mathbf{P}$ becomes the space part of the neutrino momentum vector in flat space, its curved space interpretation is not clear in general.

Unlike the electromagnetic field where $\rho_{em} = 0$ implied $\mathsf{T}_{\mu\nu_{em}} = 0$, the vanishing of the neutrino energy density ρ does not, in general, imply the vanishing of either J^0 or $\mathsf{T}_{\mu\nu}$. Cases can arise where $\rho = 0$, but $J^0 \neq 0$ and $\mathsf{T}_{\mu\nu} \neq 0$. The curious occurence of the "ghost" neutrino[11,12] is also possible when $\mathsf{T}_{\mu\nu} = 0$ but $J^0 \neq 0$. Then, the neutrino does not act as a source of gravity because $\mathsf{T}_{\mu\nu} = 0$, yet the neutrino responds via the generalized Dirac equation to the background geometry which is the solution of the vacuum Einstein equations. Thus, the ghost neutrino provides a way to study neutrino behavior in a background geometry without making a weak field approximation.

Can the properties mentioned above be ruled out? By constructing the most general first order, Riemannian spacetime, spinor field equations with their associated Lagrangian, Anderson[13] has shown that if the general relativistic neutrino field equation is to be obtained from a variational principle, then the use of the Weyl equation as the neutrino field equation "forces" the use of (3.18) as the neutrino stress-energy tensor. Thus, if the neutrino is treated with the approach used here, the difficulty in the physical interpretation of $\mathsf{T}_{\mu\nu}$ appears to be unavoidable. Letelier[14] has shown, however, that in a geometry with torsion the ghost neutrino does not occur.

4. SOLUTIONS

Now specialize to the vector cosmologies ($\mathbf{n} \neq 0$): types III, IV, V, VI_h, and VII_h. For these types in general there are three dyadics $\mathbf{F}$, $\mathbf{S}$, and $\mathbf{T}$, five vectors $\mathbf{n}$, $\mathbf{t}$, $\boldsymbol{\omega}$, $\mathbf{J}$, $\mathbf{P}$, and two scalars ρ and J^0, which must be determined as a solution of the structure equations (2.31) and (2.32) and the coupled Einstein–Dirac equations. A general simplification is obtained immediately by choosing the orthonormal frame to diagonalize $\mathbf{F}$ (or $\mathbf{S}$ when $\mathbf{F} = 0$ the case of type V). The Einstein–Dirac equations are still quite complex after this initial simplification.

In analogy with the example provided by the vector cosmologies with electromagnetic fields[15] (where a general theorem states that the Poynting vector is aligned with the vector $\mathbf{n}$), we specialize the alignment of the neutrino. That is, we look only at those cases among the vector cosmologies where the momentum density vector $\mathbf{t}$ is aligned with $\mathbf{n}$ ($\mathbf{n}\times\mathbf{t} = 0$). [See also the argument[4] suggesting that this is the general situation, SMHC following Eq. (65).]

For all vector types it is possible to choose $\mathbf{n}$ to be in the third direction. This follows from the Jacobi identity $\mathbf{F}\cdot\mathbf{n} = 0$ for types III, IV, VI_h and VII_h and from $\mathbf{S}\cdot\mathbf{n} = k\mathbf{n}$ ($k = \mathrm{const}$) for type V where $\mathbf{F} = 0$ and $\mathbf{S}$ is diagonal; $\mathbf{n}\times\mathbf{t} = 0$, so that $\mathbf{t}$ must be in the third direction also. It then follows for all vector types except type III and, except for the special case VI_h with $h = -\frac{1}{9}$, that $\boldsymbol{\omega}$, $\mathbf{P}$, and $\mathbf{J}$ are in the third direction and that $\mathsf{S}_{13} = \mathsf{S}_{23} = \mathsf{T}_{13} = \mathsf{T}_{23} = 0$. For type III, the preceding statement would be true if the additional assumption $p_3 = 0$ were made.

The alignment of all vectors in the third direction is a great simplification, and results almost always from the alignment of $\mathbf{t}$ with $\mathbf{n} = (0, 0, n_3)$.

Exact solutions are still difficult to find unless one examines cases of axial symmetry (about the third direction) in the rate of strain. This implies $\mathsf{S}_{11} = \mathsf{S}_{22}$ and $\mathsf{S}_{12} = 0$ and is consistent with types V and VII_h because $\mathsf{S}_{11} = \mathsf{S}_{22}$ requires $a = b$. In this case $\omega_1 = \omega_2 = 0$ can be chosen for types V and VII_h [See Table IV in SMHC.]

The following differential and algebraic equations are obtained for type VII_h from (2.22)–(2.24), (2.31) and (2.32), (3.14) and (3.15), (3.19)–(3.21), and (3.35) for the $\mathbf{n}\times\mathbf{t} = 0$ axially symmetric case:

$$\dot{\mathsf{S}}_{11} = -2\,\mathsf{S}_{11}^2 - \mathsf{S}_{11}\mathsf{S}_{33} + 2\,n_3^2, \tag{4.1}$$

$$\dot{\mathsf{S}}_{33} = \mathsf{S}_{11}^2 - \mathsf{S}_{33}^2 - n_3^2, \tag{4.2}$$

$$\dot{n}_3 = -\mathsf{S}_{33}n_3, \tag{4.3}$$

$$\dot{J}^0 = -\theta J^0 + 2n_3J_3, \tag{4.4}$$

$$\dot{J}_3 = -\theta J_3 + 2n_3J^0, \tag{4.5}$$

$$\dot{P}_3 = -\mathsf{S}_{33}\,P_3, \tag{4.6}$$

$$\dot{a} = -\mathsf{S}_{33}a, \tag{4.7}$$

$$\begin{aligned}\rho = & \tfrac{1}{2}(\mathsf{S}_{11}^2 + 2\mathsf{S}_{11}\mathsf{S}_{33}) - \tfrac{3}{2}n_3^2 = P_3J_3 + (\hat{v}/2)aJ^0 \\ & = \mathsf{T}_{33},\end{aligned} \tag{4.8}$$

$$t_3 = n_3(S_{33} - S_{11}) = P_3 J^0 + (\hat{v}/2) a J_3. \tag{4.9}$$

to which may be adjoined (3.16) and (3.17).

An exact solution[16] of (4.1)–(4.9), comprises the following quantities which depend on the time t:

Bianchi vector **n** [whose square measures the sectional curvature of the type VII_h space]:

$$\hat{e}\,\delta\,(0, 0, t^{-1}); \tag{4.10}$$

Neutrino current time-component:

$$J^0 = \bar{\Psi}\gamma^0\Psi\text{: } \bar{J}^0 t^{-1};$$

Neutrino current:

$$\mathbf{J} = \bar{\Psi}\gamma\Psi\text{: } \hat{e}\bar{J}^0(0, 0, t^{-1}); \tag{4.11}$$

Eikonal momentum:

$$\mathbf{P} = \nabla f\text{: } \hat{e}\left\{\frac{\delta(1-\delta)}{\bar{J}^0} - \frac{\hat{v}}{2}\bar{a}\right\}\ (0,0,t^{-1}),$$

where

$$\psi = \chi(t)e^{if(x)}; \tag{4.12}$$

Momentum density **t**:

$$\hat{e}\delta(1-\delta)\ (0,0,t^{-2}); \tag{4.13}$$

Stress tensor density **T**:

$$\delta(1-\delta)\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & t^{-2}\end{bmatrix}, \tag{4.14}$$

Energy density $\rho = T_{00}$:

$$\delta(1-\delta)t^{-2}; \tag{4.15}$$

Strain rate = extrinsic curvature **S**:

$$t^{-1}\begin{bmatrix}\delta & 0 & 0\\ 0 & \delta & 0\\ 0 & 0 & 1\end{bmatrix}, \tag{4.16}$$

Bianchi-classifying dyadic $F^{cd} = \frac{1}{2}C^{(c}{}_{gh}\epsilon^{d)gh}$,
(here the commutation coefficients $C^c{}_{gh}$ are the time evolutions of the structure constants $G^C{}_{GH}$) **F**:

$$\begin{bmatrix}a & & \\ & a & \\ & & 0\end{bmatrix} = \bar{a}t^{-1}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}; \tag{4.17}$$

Neutrino spinor amplitude Ψ:

$$\tfrac{1}{2}\sqrt{\frac{\bar{J}^0}{t}}\begin{bmatrix}1\\ -\hat{e}\\ i\hat{v}\\ -i\hat{v}\hat{e}\end{bmatrix} \times\exp\left[i\left(p_1 X_1 - \frac{\delta(1-\delta)}{\bar{J}^0}\ln t - \frac{\hat{v}\hat{e}\omega}{2}t\right)\right]. \tag{4.18}$$

Here

$$\hat{e} = \pm 1, \text{ according to } J_3 = \hat{e}J^0.$$

A dimensionless time has been used, i.e., time is measured in units of some initial time $\bar{t} = 1/\bar{S}_{33}$ which defines the constant parameters $\bar{J}^0$, $\bar{a}$, δ, p_1, and $\bar{\alpha}$ as the values taken by these physical-geometric quantities at this initial time.

δ is the dimensionless strain-rate ratio

$$\delta \equiv \bar{S}_{11}/\bar{S}_{33} = \bar{S}_{11},$$

where $\bar{S}_{33} = 1/\bar{t}$ has been taken as the unit strain rate. The eikonal momentum coefficient p_1 has the value

$$p_1 \equiv \frac{\delta(1-\delta)}{\bar{J}^0} - \frac{\hat{v}}{2}\bar{a}$$

in the metric-simplifying coordinate basis introduced in the next paragraph.

To find the metric we use the type VII_h invariant basis forms (Appendix II, this paper). We substitute into the canonical formula (68) at the end of SMHC, with the Hubble type time-dependent parameters (Table IV, SMHC) evaluated for the present solution ($\bar{\alpha} = 1$ with $\bar{t}$ chosen properly):

$$\alpha = \beta = \bar{\alpha}\exp\left[-\int_{\tau}^{t} S_1\, dt\right] = \bar{\alpha}(\bar{t}/t)^{\delta} = \bar{\alpha}/t^{\delta} = 1/t^{\delta},$$

$$\gamma = n/n_0 = \delta/(n_0 t).$$

The result of the calculations is

$$ds^2 = -dt^2 + t^2\frac{A^2}{\delta^2}dX_1^2 + t^2 e^{-2AX_1}(dX_2^2 + dX_3^2 + 2A\,dX_2 dX_3).$$

The metric may then be simplified by an affine transformation, and takes the form

$$\text{Metric: } ds^2 = -dt^2 + t^2 dx^2 + (e^{-\delta x}t^{\delta})^2[dy^2 + dz^2]. \tag{4.19}$$

This significant simplification of the metric has been obtained by transforming the coordinates used in Appendix II for type VII_h by a rotation about the X_1 axis (e_{III} direction) followed by a change in scale:

$$\begin{pmatrix}x\\ y\\ z\end{pmatrix}_{\text{new}} = \begin{bmatrix}\delta^{-1}A & 0 & 0\\ 0 & \delta^{-1}\sqrt{1-A} & 0\\ 0 & 0 & \delta^{-1}\sqrt{1+A}\end{bmatrix} \times\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2}\\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}\end{bmatrix}\begin{pmatrix}X_1\\ X_2\\ X_3\end{pmatrix}_{\text{old}} \tag{4.20}$$

(for the definition of A see Appendix II, type VII). Thus, the explicit coordinate representation of the metric (4.19) and the definition of p_1 differ from what one would obtain using the untransformed basis vectors of Appendix II.

The differential and algebraic equations for type V with $\mathbf{n}\times\mathbf{t} = 0$ and with axial symmetry in the rate of strain are just (4.1)–(4.9) with $a = 0$. An exact solution of these type V equations can also be found,[16] and this solution is just (4.10)–(4.18) with $\bar{a} = 0$. Although the type V solution is similar in form to the type VII_h solution, these are really two distinct solutions. Each corresponds to a different geometry invariant under different isometry groups. Also the relations

$$\gamma = \hat{e}\bar{a}/t, \quad \delta = \bar{S}_{11} = \hat{e}n_0\bar{a},$$

which hold for Type VII_h, do not apply to Type V.

By evaluating the Weyl tensor in the NP formalism one finds that both solutions are of Petrov type D. (See Appendix III.)

In both solutions $\delta = 1$ means an isotropic solution with a ghost neutrino field, but these solutions cannot ap-

TABLE IV. Comparision of the early time limit of a radiation dominated Friedmann model with the neutrino solutions.

	Radiation dominated Friedmann (early time limit)	Type V and VII_h neutrino solutions
Avg. Hubble expansion $\equiv \frac{\theta}{3}$	$\frac{1}{2t}$	$\frac{2\delta+1}{3t}$
Energy density ρ	$\frac{3}{8t^2}$	$\frac{\delta(1-\delta)}{t^2}$

proach isotropy from an initially anisotropic state.

The behavior of all quantities is to have singularities at $t=0$, have some finite value at $t=\bar{t}$, and decrease continuously to a limit of zero as $t\to\infty$.

It is interesting to compare the early time limit of a radiation dominated Friedmann model[17] with the neutrino solutions given by (4.10)–(4.20) [Table IV].

The neutrino Hubble expansion is greater than the Friedmann rate when $\delta > \frac{1}{4}$ and because the neutrino energy density is maximum when $\delta = \frac{1}{2}$ it is always less than the Friedmann energy density.

Now examine the motion of a single neutrino in the third (orthonormal) direction. With the aid of Appendix II it is easily seen that motion in the third orthonormal direction ($\mathbf{l}_3$) corresponds for types V and VII_h to motion in the first holonomic direction (∂_1). [This occurs because the triad of invariant basis vectors e_A was chosen to put (2.5) and Table I in a canonical form according to the Estabrook, Wahlquist, and Behr convention.] In the holonomic frame, the third and second components of the neutrino 4-momentum Π^μ vanish. Π^μ is lightlike and thus,

$$\Pi_\mu \Pi^\mu = \Pi_0\Pi^0 + \Pi_1\Pi^1 = 0.$$

The energy of the neutrino as measured by a stationary observer with 4-velocity U^μ is

$$E = -\Pi_\mu U^\mu = -\Pi_0 U^0 = \Pi^0. \tag{4.21}$$

Therefore, with the metrics of either solution

$$E^2 = g^{11}(\Pi_1)^2. \tag{4.22}$$

Along any geodesic $\Pi_\mu z^\mu = \text{const}$. Evaluating $\Pi_\mu z^\mu$ for types V and VII_h with the Killing vectors z^μ from Appendix II will show that Π_1 is a conserved quantity. Thus (4.22) gives the expression for the red-shifting energy of a single neutrino traveling in the $\mathbf{l}_3$ direction:

$$E(t) = \Pi_1/t. \tag{4.23}$$

In the wavefunctions for the two solutions, the constant p_1 appears in a position that would be occupied by Π_1 in a flat space limit. If Π_1 and p_1 are equated, then

$$E(t) = p_1/t = |P_3|. \tag{4.24}$$

This implies that the magnitude of P_3 is just the red-shifting energy of a single neutrino. The energy density ρ, however, is not always expressed entirely in terms of $E(t)$:

$$\rho = \left[E(t) + \frac{\hat{v}\bar{a}}{2t}\right]J^0. \tag{4.25}$$

For type VII_h, ρ contains a contribution from the intrinsic curvature parameter a as well as from the red-shifting energy.

J^0 as a Probability Density?: The interpretation of J^0 as a probability density is certainly suggested by the fact that $J^0 \geqslant 0$ and by the fact that J^μ is the conserved probability current in a flat space limit. For the solutions found here, J^0 is a function of time alone so that J^0 is constant over each hypersurface. If J^0 is interpreted as a probability density, then the position of the neutrino on any hypersurface is totally uncertain just like the case of a plane wave in flat space. Can the wavefunction be normalized so that the 3-volume integral of J^0 equals one?

J^0 depends only on time so it may be taken outside of an integral over 3-space:

$$\int J^0\, d\mathscr{V} = J^0\mathscr{V}. \tag{4.26}$$

The product $J^0\mathscr{V}$ will be constant and normalization will be maintained only if the time variation of J^0 is inversely proportional to the time variation of $\mathscr{V}$. But, J^0 and $\mathscr{V}$ satisfy

$$\dot{\mathscr{V}} = \theta\mathscr{V}, \tag{4.27}$$

$$\dot{J}^0 = -\theta J^0 + 2\mathbf{n}\cdot\mathbf{J}. \tag{4.28}$$

It is easy to see that the time variations of J^0 and $\mathscr{V}$ are inversely proportional for all nonvector cosmologies ($\mathbf{n}=0$) and for those vector cosmologies where $\mathbf{n}\cdot\mathbf{J}=0$. (Thus, Brill and Cohen are able to normalize the wavefunction for their type I solution.[8]) Otherwise, the normalization will not be conserved. Unfortunately, the solutions presented here have $\mathbf{n}\cdot\mathbf{J}\neq 0$ and thus $J^0\mathscr{V}$ is not constant.

It is possible, however, to interpret J^0 as a one-dimensional probability density and require normalization over a length L only in the x ($\mathbf{l}_3$) direction. (The spatial dependence of Ψ is exactly like that of a plane wave traveling in the x direction.) Let

$$\int_{-L/2}^{L/2} J^0 (g_{11})^{1/2} dx = 1. \tag{4.29}$$

This normalization will be conserved for both solutions presented here.

APPENDIX I

The purpose of this appendix is to derive the Riemann, Ricci, and Einstein tensors in terms of the dyadic quantities.

Starting with the connection 1-forms $\bar{\omega}_{\mu\nu} = \Gamma_{\mu\nu\alpha}\lambda^\alpha$, we have for the Bianchi cosmologies

$$\bar{\omega}_{0a} = -\mathsf{S}_{ad}\lambda^d, \tag{I1}$$

$$\bar{\omega}_{ab} = \mathsf{W}_{ab}\lambda^0 - \mathsf{C}'_{abd}\lambda^d. \tag{I2}$$

The curvature 2-form $\mathscr{R}^\mu{}_\nu$ is then computed in a direct manner from the exterior derivative of the connection 1-form:

$$\mathscr{R}^\mu{}_\nu = d\bar{\omega}^\mu{}_\nu + \bar{\omega}^\mu{}_\alpha \wedge \bar{\omega}^\alpha{}_\nu. \tag{I3}$$

The result is

$$\mathscr{R}^0{}_a = [\dot{\mathbf{S}} + \mathbf{S}\cdot\mathbf{S} + \mathbf{W}\cdot\mathbf{S} - \mathbf{S}\cdot\mathbf{W}]_{ad}\lambda^0\wedge\lambda^d - [(\mathbf{S}\times\mathbf{A})_{afd} + (\mathbf{A}\times\mathbf{S})_{dfa}]\lambda^f\wedge\lambda^d, \tag{I4}$$

$$\mathscr{R}^a{}_b = [\{\dot{\mathbf{C}}' + \mathbf{C}'\cdot\mathbf{S} - \mathbf{C}'\cdot\mathbf{W}\}_{abd} + 2(\mathbf{W}\times\mathbf{A})_{[ba]d}]\lambda^0\wedge\lambda^d + [\mathsf{S}_{af}\mathsf{S}_{bd} + \mathsf{C}'_{afg}\mathsf{C}'^{g}{}_{bd} - \mathsf{C}'_{ab}{}^{g}\mathsf{C}'_{gdf}]\lambda^f\wedge\lambda^d, \tag{I5}$$

where, for example, $[\mathbf{C}'\cdot\mathbf{S}]_{abd} = C'_{abc}S^{c}{}_{d}$ and $[\mathbf{S}\times\mathbf{A}]_{afd} = \epsilon_{agc}S^{g}_{f}A^{c}_{d}$.

The components of the Riemann tensor may now be read from

$$\mathscr{R}^{\mu}{}_{\nu} = R^{\mu}{}_{\nu|\alpha\beta|}\lambda^{\alpha}\wedge\lambda^{\beta}. \tag{I6}$$

The result is

$$R^{0a}{}_{od} = [\dot{\mathbf{S}} + \mathbf{S}\cdot\mathbf{S} + \mathbf{W}\cdot\mathbf{S} - \mathbf{S}\cdot\mathbf{W}]_{ad}, \tag{I7}$$

$$R^{0a}{}_{fd} = 2[(\mathbf{S}\times\mathbf{A})_{a[df]} + (\mathbf{A}\times\mathbf{S})_{[fd]a}], \tag{I8}$$

$$R^{ab}{}_{0d} = 2(\mathbf{W}\times\mathbf{A})_{[ba]d} + [\dot{\mathbf{C}}' + \mathbf{C}'\cdot(\mathbf{S}-\mathbf{W})]_{abd}, \tag{I9}$$

$$R^{ab}{}_{fd} = 2S^{a}{}_{[f}S^{b}{}_{d]} + {}^{(3)}R^{ab}{}_{fd}, \tag{I10}$$

where the Riemann tensor intrinsic to the hypersurface of homogeneity is

$${}^{(3)}R^{ab}{}_{fd} = 2[C'^{a}{}_{g[f}C'^{gb}{}_{d]} - C'^{abg}C'_{g[df]}]. \tag{I11}$$

The convenience of having the commutation triadic $\mathbf{C}'$ appear explicitly in (I2)–(I11) reflects the relative complexity of the 4-index Riemann tensor components as compared to the Ricci tensor ($R^{\mu}_{\nu} = R^{\alpha\mu}{}_{\nu\alpha}$).

The Ricci and Einstein tensors ($G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$) and the Ricci scalar ($R = R^{\mu}{}_{\mu}$) can be expressed explicitly in terms of the dyadic quantities

$$R_{00} = \dot{\theta} + \mathbf{S}:\mathbf{S}, \tag{I12}$$

$$R_{0f} = G_{0f} = [\mathbf{F}\times\mathbf{S} - \theta\mathbf{n} + 3\,\mathbf{n}\cdot\mathbf{S}]_{f}, \tag{I13}$$

$$\mathbf{R} = -\dot{\mathbf{S}} - \boldsymbol{\omega}\times\mathbf{S} + \mathbf{S}\times\boldsymbol{\omega} - \theta\mathbf{S} - 2\mathbf{F}\cdot\mathbf{F} + \Phi\mathbf{F} + \mathbf{F}\times\mathbf{n} - \mathbf{n}\times\mathbf{F} + [\mathbf{F}:\mathbf{F} - \tfrac{1}{2}\Phi^{2} + 2n^{2}]\mathbf{I}. \tag{I14}$$

$$R = -2\dot{\theta} - \theta^{2} - \mathbf{S}:\mathbf{S} + \mathbf{F}:\mathbf{F} - \tfrac{1}{2}\Phi^{2} + 6n^{2}, \tag{I15}$$

$$G_{00} = \tfrac{1}{2}[-\theta^{2} + \mathbf{S}:\mathbf{S} + \mathbf{F}:\mathbf{F} - \tfrac{1}{2}\Phi^{2} + 6n^{2}], \tag{I16}$$

$$\mathbf{G} = -\dot{\mathbf{S}} - \boldsymbol{\omega}\times\mathbf{S} + \mathbf{S}\times\boldsymbol{\omega} - \theta\mathbf{S} - 2\mathbf{F}\cdot\mathbf{F} + \Phi\mathbf{F} + \mathbf{F}\times\mathbf{n} - \mathbf{n}\times\mathbf{F} + [\dot{\theta} + \tfrac{1}{2}\theta^{2} + \tfrac{1}{2}\mathbf{S}:\mathbf{S} + \tfrac{1}{2}\mathbf{F}:\mathbf{F} - \tfrac{1}{4}\Phi^{2} - n^{2}]I, \tag{I17}$$

where $n^{2} = \mathbf{n}\cdot\mathbf{n}$, $(\mathbf{F}\times\mathbf{S}) \equiv \epsilon_{fcg}F^{c}{}_{a}S^{ga}$, $\mathbf{S}:\mathbf{S} = S^{ag}S_{ga}$, and $\mathbf{R}$ and $\mathbf{G}$ are the dyadics of components of $R_{\mu\nu}$ and $G_{\mu\nu}$ with spacelike indices.

Equations (I10) and (I8) are seen to be just the equations of Gauss and Codazzi[18] as applied to the Bianchi cosmologies. The Ricci tensor and the Ricci scalar intrinsic to the hypersurfaces of homogeneity are

$${}^{(3)}\mathbf{R} = -2\mathbf{F}\cdot\mathbf{F} + \Phi\mathbf{F} + \mathbf{F}\times\mathbf{n} - \mathbf{n}\times\mathbf{F} + [\mathbf{F}:\mathbf{F} - \tfrac{1}{2}\Phi^{2} + 2n^{2}]\mathbf{I}, \tag{I18}$$

$${}^{(3)}R = \mathbf{F}:\mathbf{F} - \tfrac{1}{2}\Phi^{2} + 6n^{2}, \tag{I19}$$

respectively.

APPENDIX II

The purpose of this appendix is to present for each Bianchi type a canonical set of reciprocal group generators (invariant basis vectors) $\mathbf{e}_{A}$ and Killing vectors $\mathbf{z}_{B}$ along with their respective duals ϵ^{A} and ζ^{B}. Taub[19] has given a set of such vectors, but his vectors do not always put the group algebra in the canonical form chosen by Estabrook, Wahlquist, and Behr. The relation of the canonical triads given here to those of Taub will be made explicit for each type by giving the matrix of transformation (B^{b}_{A}) between the two sets of triads ($\mathbf{e}_{A_{\text{here}}} = B^{b}_{A}\mathbf{e}_{b_{\text{Taub}}}$).

Let ∂_{i}, $\mathbf{e}_{A}$, and $\mathbf{l}_{a}$ be, respectively, a triad of holonomic, invariant, and orthonormal basis vectors. Let $\mathbf{d}X^{i}$, ϵ^{A} and λ^{a} be, respectively, a triad of holonomic, invariant, and orthonormal basis 1-forms dual to the various vector bases. The relations between these bases are

$$\partial_{i} = \epsilon^{B}_{i}\mathbf{e}_{B} = \tilde{C}_{B}{}^{a}\epsilon^{B}_{i}\mathbf{l}_{a}, \qquad \mathbf{d}X^{i} = e^{i}_{A}\epsilon^{A} = C_{b}{}^{A}e^{i}_{A}\lambda^{b},$$
$$\mathbf{e}_{A} = e^{i}_{A}\partial_{i} = \tilde{C}_{A}{}^{b}\mathbf{l}_{b}, \qquad \epsilon^{A} = \epsilon^{A}_{i}\mathbf{d}X^{i} = C_{b}{}^{A}\lambda^{b},$$
$$\mathbf{l}_{a} = C_{a}{}^{B}\mathbf{e}_{B} = C_{a}{}^{B}e^{i}_{B}\partial_{i}, \qquad \lambda^{a} = \tilde{C}_{B}{}^{a}\epsilon^{B} = \tilde{C}_{B}{}^{a}\epsilon^{B}_{i}\mathbf{d}X^{i},$$

where $C_{b}{}^{A}$ and $\tilde{C}_{B}{}^{a}$ are defined with (2.6) and (2.7) and satisfy (2.27) and (2.28).

The anholonomic components are projected, so that e^{i}_{II} is the ith holonomic component of the second invariant basis vector, etc. The group Lie algebra for each type is given by (2.5) and Table I as follows:

Type I:

$\mathbf{e}_{\text{I}} = (1,0,0)$, $\mathbf{e}_{\text{II}} = (0,1,0)$, $\mathbf{e}_{\text{III}} = (0,0,1)$,
$\epsilon^{\text{I}} = (1,0,0)$, $\epsilon^{\text{II}} = (0,1,0)$, $\epsilon^{\text{III}} = (0,0,1)$,
$\mathbf{z}_{\text{I}} = (1,0,0)$, $\mathbf{z}_{\text{II}} = (0,1,0)$, $\mathbf{z}_{\text{III}} = (0,0,1)$,
$\zeta^{\text{I}} = (1,0,0)$, $\zeta^{\text{II}} = (0,1,0)$, $\zeta^{\text{III}} = (0,0,1)$,

$$B_{R}{}^{b} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

Type II:

$\mathbf{e}_{\text{I}} = (0,1,0)$, $\mathbf{e}_{\text{II}} = (0,X_{1},1)$, $\mathbf{e}_{\text{III}} = (-1,0,0)$,
$\epsilon^{\text{I}} = (0,1,-X_{1})$, $\epsilon^{\text{II}} = (0,0,1)$, $\epsilon^{\text{III}} = (-1,0,0)$,
$\mathbf{z}_{\text{I}} = (0,1,0)$, $\mathbf{z}_{\text{II}} = (0,0,1)$, $\mathbf{z}_{\text{III}} = (-1,-X_{3},0)$,
$\zeta^{\text{I}} = (-X_{3},1,0)$, $\zeta^{\text{II}} = (0,0,1)$, $\zeta^{\text{III}} = (-1,0,0)$,

$$B_{R}{}^{b} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix};$$

Type III:

$\mathbf{e}_{\mathrm{I}} = (0,e^{X_1},1)$, $\mathbf{e}_{\mathrm{II}} = (0,-e^{X_1},1)$, $\mathbf{e}_{\mathrm{III}} = (2,0,0)$,
$\epsilon^{\mathrm{I}} = \frac{1}{2}(0,e^{-X_1},1)$, $\epsilon^{\mathrm{II}} = \frac{1}{2}(0,-e^{-X_1},1)$, $\epsilon^{\mathrm{III}} = (\frac{1}{2},0,0)$,
$\mathbf{z}_{\mathrm{I}} = (0,1,1)$, $\mathbf{z}_{\mathrm{II}} = (0,-1,1)$, $\mathbf{z}_{\mathrm{III}} = (2,2X_2,0)$,
$\zeta^{\mathrm{I}} = \frac{1}{2}(-X_2,1,1)$, $\zeta^{\mathrm{II}} = \frac{1}{2}(X_2,-1,1)$, $\zeta^{\mathrm{III}} = (\frac{1}{2},0,0)$,

$$B_R{}^b = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix};$$

Type IV:

$\mathbf{e}_{\mathrm{I}} = (0,-e^{X_1},0)$, $\mathbf{e}_{\mathrm{II}} = (0,X_1e^{X_1},e^{X_1})$, $\mathbf{e}_{\mathrm{III}} = (1,0,0)$,
$\epsilon^{\mathrm{I}} = (0,-e^{-X_1},X_1e^{-X_1})$, $\epsilon^{\mathrm{II}} = (0,0,e^{-X_1})$, $\epsilon^{\mathrm{III}} = (1,0,0)$,
$\mathbf{z}_{\mathrm{I}} = (0,-1,1)$, $\mathbf{z}_{\mathrm{II}} = (0,0,1)$, $\mathbf{z}_{\mathrm{III}} = (1,X_2+X_3,X_3)$,
$\zeta^{\mathrm{I}} = (X_2+X_3,-1,0)$, $\zeta^{\mathrm{II}} = (-X_3,0,1)$, $\zeta^{\mathrm{III}} = (1,0,0)$,

$$B_R{}^b = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

Type V:

$\mathbf{e}_{\mathrm{I}} = (0,e^{X_1},0)$, $\mathbf{e}_{\mathrm{II}} = (0,0,e^{X_1})$, $\mathbf{e}_{\mathrm{III}} = (1,0,0)$,
$\epsilon^{\mathrm{I}} = (0,e^{-X_1},0)$, $\epsilon^{\mathrm{II}} = (0,0,e^{-X_1})$, $\epsilon^{\mathrm{III}} = (1,0,0)$,
$\mathbf{z}_{\mathrm{I}} = (0,1,0)$, $\mathbf{z}_{\mathrm{II}} = (0,0,1)$, $\mathbf{z}_{\mathrm{III}} = (1,X_2,X_3)$,
$\zeta^{\mathrm{I}} = (-X_2,1,0)$, $\zeta^{\mathrm{II}} = (-X_3,0,1)$, $\zeta^{\mathrm{III}} = (1,0,0)$,

$$B_R{}^b = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

Type VI_h *and* VI_0:

$\mathbf{e}_{\mathrm{I}} = (0,e^{X_1},e^{kX_1})$, $\mathbf{e}_{\mathrm{II}} = (0,-e^{X_1},e^{kX_1})$, $\mathbf{e}_{\mathrm{III}} = (1+n_0,0,0)$,
$\epsilon^{\mathrm{I}} = \frac{1}{2}(0,e^{-X_1},e^{-kX_1})$, $\epsilon^{\mathrm{II}} = \frac{1}{2}(0,e^{-X_1},e^{-kX_1})$, $\epsilon^{\mathrm{III}} = \left(\frac{1}{1+n_0},0,0\right)$,
$\mathbf{z}_{\mathrm{I}} = (0,1,1)$, $\mathbf{z}_{\mathrm{II}} = (0,-1,1)$, $\mathbf{z}_{\mathrm{III}} = (1+n_0)(1,X_2,kX_3)$,
$\zeta^{\mathrm{I}} = \frac{1}{2}(-X_2-kX_3,1,1)$, $\zeta^{\mathrm{II}} = \frac{1}{2}(X_2-kX_3,-1,1)$, $\zeta^{\mathrm{III}} = \left(\frac{1}{1+n_0},0,0\right)$,

$$B_R{}^b = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & (1+n_0) \end{pmatrix}.$$

For type VI, the constant k (called h in Taub) cannot equal one or zero. If $k = 0$; then type VI becomes type III. The values of k and n_0 are related by: $k = (n_0 - 1)/(n_0 + 1)$ or, $n_0 = (k+1)/(1-k)$. Therefore, if $k \neq 0$, then $n_0 \neq 1$ as indicated in Table I. Type VI_0 arises when $k = -1$ so that $n_0 = 0$.

Types VII_h *and* VII_0:

$$\mathbf{e}_{\mathrm{I}} = e^{AX_1}\left(0,n_0\sin DX_1 - \cos DX_1, \frac{-\sin DX_1}{D}\right), \qquad \epsilon^{\mathrm{I}} = -e^{-AX_1}(0,\cos DX_1, D\,[\sin DX_1 + n_0\cos DX_1]),$$

$$\mathbf{e}_{\mathrm{II}} = e^{AX_1}\left(0,n_0\cos DX_1 + \sin DX_1, \frac{-\cos DX_1}{D}\right), \qquad \epsilon^{\mathrm{II}} = -e^{-AX_1}(0,-\sin DX_1, D\,[\cos DX_1 - n_0\sin DX_1]),$$

$\mathbf{e}_{\mathrm{III}} = (D^{-1},0,0)$, $\epsilon^{\mathrm{III}} = (D,0,0)$,
$\mathbf{z}_{\mathrm{I}} = (0,-1,0)$, $\mathbf{z}_{\mathrm{II}} = (0,n_0,-D^{-1})$, $\mathbf{z}_{\mathrm{III}} = D^{-1}(1,-X_3,X_2+2AX_3)$,
$\zeta^{\mathrm{I}} = ([2A^2-1]X_3 + AX_2, -1, -A)$, $\zeta^{\mathrm{II}} = D(X_2+2AX_3,0-1)$, $\zeta^{\mathrm{III}} = (D,0,0)$,

$$B_R{}^b = \begin{pmatrix} -1 & n_0 & 0 \\ 0 & -D^{-1} & 0 \\ 0 & 0 & D^{-1} \end{pmatrix},$$

where

$$D \equiv \sqrt{1-A^2} \qquad (A^2 < 1).$$

($2A$ is called h in Taub.) From the requirement that the matrix $B_R{}^b$ be real, one finds $A = n_0D$. This implies $D = 1/(n_0^2+1)^{1/2}$. When $A = 0$, $n_0 = 0$, and type VII_0 is generated.

Type VIII:

$$\mathbf{e}_{\mathrm{I}} = (A, BX_2 - 1, B), \qquad \mathbf{e}_{\mathrm{II}} = \left(\frac{B}{2}, -X_2, -1\right), \qquad \mathbf{e}_{\mathrm{III}} = (A - \tfrac{1}{2}, BX_2 - 1, B),$$

$$\boldsymbol{\epsilon}^{\mathrm{I}} = (2, -2A, 2AX_2 + B), \quad \boldsymbol{\epsilon}^{\mathrm{II}} = (0, -B, BX_2 - 1), \quad \boldsymbol{\epsilon}^{\mathrm{III}} = (-2, 2A - 1, X_2 - 2AX_2 - B),$$

$$\mathbf{z}_{\mathrm{I}} = (0, -e^{X_3}, -1), \quad \mathbf{z}_{\mathrm{II}} = \tfrac{1}{2}(-e^{-X_3}, C, 2D), \quad \mathbf{z}_{\mathrm{III}} = \tfrac{1}{2}(-e^{-X_3}, C - 2e^{X_3}, 2D),$$

$$\boldsymbol{\zeta}^{\mathrm{I}} = (E, 0, -1), \quad \boldsymbol{\zeta}^{\mathrm{II}} = ([D-1]X_2, e^{-X_3}, -1), \quad \boldsymbol{\zeta}^{\mathrm{III}} = (-E - C, -e^{-X_3}, 1),$$

$$B_R{}^b = \begin{pmatrix} 0 & -\frac{1}{2} & -\frac{1}{2} \\ -1 & -1 & -1 \\ -1 & 0 & -1 \end{pmatrix},$$

where

$$A \equiv X_1(1 - X_1), \quad B \equiv 2X_1 - 1, \quad C \equiv (X_2)^2 e^{-X_3}, \quad D \equiv X_2 e^{-X_3} - 1, \quad E \equiv 2(e^{X_3} - X_2);$$

Type IX:

$$\mathbf{e}_{\mathrm{I}} = A(\sin X_3 \sin X_1, -\cos X_3, \cos X_1 \cos X_3), \qquad \boldsymbol{\epsilon}^{\mathrm{I}} = (\sin X_3, -\sin X_1 \cos X_3, 0),$$

$$\mathbf{e}_{\mathrm{II}} = A(-\cos X_3 \sin X_1, -\sin X_3, \cos X_1 \sin X_3), \qquad \boldsymbol{\epsilon}^{\mathrm{II}} = (-\cos X_3, -\sin X_1 \sin X_3, 0),$$

$$\mathbf{e}_{\mathrm{III}} = (0, 0, -1), \qquad \boldsymbol{\epsilon}^{\mathrm{III}} = (0, -\cos X_1, -1),$$

$$\mathbf{z}_{\mathrm{I}} = (0, -1, 0), \qquad \boldsymbol{\zeta}^{\mathrm{I}} = (0, -1, -\cos X_1),$$

$$\mathbf{z}_{\mathrm{II}} = A(-\cos X_2 \sin X_1, \cos X_1 \sin X_2, -\sin X_2), \qquad \boldsymbol{\zeta}^{\mathrm{II}} = (-\cos X_2, 0, -\sin X_1 \sin X_2),$$

$$\mathbf{z}_{\mathrm{III}} = A(\sin X_2 \sin X_1, \cos X_1 \cos X_2, -\cos X_2), \qquad \boldsymbol{\zeta}^{\mathrm{III}} = (\sin X_2, 0, -\sin X_1 \cos X_2),$$

$$B_R{}^b = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

where $A \equiv \csc X_1$.

APPENDIX III

Much of the work on neutrinos in general relativity makes use of a null tetrad and the NP formalism,[20] or its updated GHP form.[21] The relation between the orthonormal tetrad used in this paper and the null tetrad of the NP formalism is given by

$$\mathbf{l}_\mu = B_\mu{}^\nu \mathbf{L}_\nu, \quad \mathbf{L}_\mu = A_\mu{}^\nu \mathbf{l}_\nu, \tag{III1}$$

where the $\mathbf{l}_\mu$ are the four orthonormal basis vectors, and the $\mathbf{L}_\mu$ are the four null basis vectors ($\mathbf{L}_0 \equiv \mathbf{l}$, $\mathbf{L}_1 \equiv \mathbf{n}$, $\mathbf{L}_2 \equiv \mathbf{m}$, $\mathbf{L}_3 \equiv \bar{\mathbf{m}}$). Also

$$A_\mu{}^\nu = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -i \\ 0 & 0 & 1 & +i \end{bmatrix}, \qquad B_\mu{}^\nu = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & i & -i \end{bmatrix}. \tag{III2}$$

Of course, we have $B_\mu{}^\nu A_\nu{}^\sigma = \delta_\mu{}^\sigma$. The relation between the dyadic quantities (Ricci rotation coefficients ${}_{\mathrm{EWB}}\Gamma^\sigma{}_{\rho\epsilon}$) and the NP spin coefficients (${}_{\mathrm{NP}}\Gamma^\xi{}_{\beta\alpha}$) is found by the transformation

$${}_{\mathrm{NP}}\Gamma^\xi{}_{\beta\alpha} = B_\rho{}^\xi A_\beta{}^\rho{}_{,\alpha} + B_\sigma{}^\xi A_\beta{}^\rho A_\alpha{}^\epsilon {}_{\mathrm{EWB}}\Gamma^\sigma{}_{\rho\epsilon} = B_\sigma{}^\xi A_\beta{}^\rho A_\alpha{}^\epsilon {}_{\mathrm{EWB}}\Gamma^\sigma{}_{\rho\epsilon}. \tag{III3}$$

The results of (III3) will be written in terms of the GHP formalism[21] which we shall abbreviate slightly. In this formalism, the six NP spin coefficients $\kappa, \sigma, \rho, \tau, \beta$, and ϵ remain unchanged. The other six spin coefficients are obtained by a prime operation ($\nu = -\kappa', \mu = -\rho', \lambda = -\sigma', \pi = -\tau', \alpha = -\beta', \gamma = -\epsilon'$), which corresponds to the following interchange of the null basis vectors: $\mathbf{l} \rightleftharpoons \mathbf{n}$ and $\mathbf{m} \rightleftharpoons \bar{\mathbf{m}}$. We then introduce the abbreviations

$$\kappa_\pm \equiv \kappa \pm \kappa', \quad \rho_\pm \equiv \rho \pm \rho', \quad \sigma_\pm \equiv \sigma \pm \sigma', \quad \tau_\pm \equiv \tau \pm \tau', \quad \beta_\pm \equiv \beta \pm \beta', \quad \epsilon_\pm \equiv \epsilon \pm \epsilon',$$

and we have

$$\mathbf{a} + i\boldsymbol{\omega} = \sqrt{2}\{\epsilon_- \mathbf{l}_1 - \tfrac{1}{2}(\kappa_+ + \tau_+)\mathbf{l}_2 - (i/2)(\kappa_- + \tau_-)\mathbf{l}_3\}, \tag{III4}$$

$$\mathbf{n} + i\boldsymbol{\Omega} = (\sqrt{2}/2)\{\rho_- \mathbf{l}_1 + \tfrac{1}{2}(\tau_+ - \kappa_+ - 2\beta_+)\mathbf{l}_2 + (i/2)(\tau_- - \kappa_- - 2\beta_-)\mathbf{l}_3\}, \tag{III5}$$

$$\mathscr{F} \equiv \mathbf{F} - \tfrac{1}{2}\Phi\mathbf{I} + i\mathbf{S} = \frac{\sqrt{2}}{2}\begin{bmatrix} 2i\epsilon_+ & \frac{i}{2}(\tau_- - \kappa_- + 2\beta_-) & -\frac{1}{2}(\tau_+ - \kappa_+ + 2\beta_+) \\ \frac{i}{2}(\tau_- - \kappa_- + 2\beta_-) & -i(\rho_+ + \sigma_+) & \sigma_- \\ -\frac{1}{2}(\tau_+ - \kappa_+ + 2\beta_+) & \sigma_- & i(\sigma_+ - \rho_+) \end{bmatrix}. \tag{III6}$$

The inverse relations are

$$\kappa_+ = (\sqrt{2}/2)[\mathcal{F}_{31} - (n_2 + i\Omega_2) - (a_2 + i\omega_2)], \quad \text{(III7)}$$

$$\kappa_- = (\sqrt{2}/2)\, i[\mathcal{F}_{12} + (n_3 + i\Omega_3) + (a_3 + i\omega_3)], \quad \text{(III8)}$$

$$\rho_+ = (\sqrt{2}/2)\, i[\mathcal{F}_{11} - i\theta + \tfrac{1}{2}\Phi] = (\sqrt{2}/2)\, i[\mathcal{F}_{11} - \mathrm{Tr}\mathcal{F}], \quad \text{(III9)}$$

$$\rho_- = \sqrt{2}(n_1 + i\Omega_1), \quad \text{(III10)}$$

$$\sigma_+ = (\sqrt{2}/2)\, i(\mathcal{F}_{22} - \mathcal{F}_{33}), \quad \text{(III11)}$$

$$\sigma_- = \sqrt{2}\mathcal{F}_{23}, \quad \text{(III12)}$$

$$\tau_+ = (\sqrt{2}/2)[-\mathcal{F}_{31} + (n_2 + i\Omega_2) - (a_2 + i\omega_2)], \quad \text{(III13)}$$

$$\tau_- = (\sqrt{2}/2)\, i[-\mathcal{F}_{12} - (n_3 + i\Omega_3) + (a_3 + i\omega_3)], \quad \text{(III14)}$$

$$\beta_+ = (\sqrt{2}/2)[-\mathcal{F}_{31} - (n_2 + i\Omega_2)], \quad \text{(III15)}$$

$$\beta_- = i\,(\sqrt{2}/2)[-\mathcal{F}_{12} + (n_3 + i\Omega_3)], \quad \text{(III16)}$$

$$\epsilon_+ = -(\sqrt{2}/2)\, i\mathcal{F}_{11}, \quad \text{(III17)}$$

$$\epsilon_- = (\sqrt{2}/2)(a_1 + i\omega_1). \quad \text{(III18)}$$

One can now proceed to evaluate the components of the Weyl tensor. For the homogeneous cosmologies, the only nonzero intrinsic derivatives are those taken along the timelike direction. Thus

$$\Delta\phi = D'\phi = D\phi = (1/\sqrt{2})\dot{\phi}, \quad \delta\phi = \bar{\delta}\phi = \delta'\phi = 0. \quad \text{(III19)}$$

The Weyl tensor components for the homogeneous cosmologies become

$$\Psi_0 = (1/\sqrt{2})\dot{\sigma} - (\rho + \bar{\rho} + 3\epsilon - \bar{\epsilon})\sigma + (\tau + \bar{\tau}' - \bar{\beta}' + 3\beta)\kappa, \quad \text{(III20)}$$

$$\Psi_1 = (1/\sqrt{2})\dot{\beta} - (\bar{\rho} - \bar{\epsilon})\beta + (\beta' + \tau')\sigma - (\rho' + \epsilon')\kappa + (\bar{\tau}' - \bar{\beta}')\epsilon, \quad \text{(III21)}$$

$$\Psi_2 = \tfrac{1}{3}\{-(1/\sqrt{2})(\dot{\epsilon}' + \dot{\rho}' + \dot{\epsilon}) + (\tau - \bar{\tau}' + \bar{\beta}' + 2\beta)\beta' - (\bar{\tau} - \tau' - \bar{\beta})\beta - (\epsilon' + \rho')(\epsilon + \bar{\epsilon} + \rho - \bar{\rho}) + (\bar{\rho}' - \rho' - \bar{\epsilon}' - \epsilon')\epsilon + (\bar{\beta}' + \beta + \tau - \bar{\tau}')\tau' - 2\kappa'\kappa + 2\sigma'\sigma\}, \quad \text{(III22)}$$

$$\Psi_3 = (1/\sqrt{2})\dot{\beta}' - (\rho + \epsilon)\kappa' + (\tau + \beta)\sigma' + (\bar{\epsilon}' - \bar{\rho}')\beta' + (\bar{\tau} - \bar{\beta})\epsilon', \quad \text{(III23)}$$

$$\Psi_4 = (1/\sqrt{2})\dot{\sigma}' - (\rho' + \bar{\rho}')\sigma' + (\bar{\epsilon}' - \epsilon')\sigma' + [3\beta' - \bar{\beta} + \tau' + \bar{\tau})\kappa'. \quad \text{(III24)}$$

The spin coefficients evaluated for the type VII_h solution are

$$\epsilon = \epsilon' = -i\hat{e}\kappa = i\hat{e}\tau = -i\hat{e}\tau' = i\hat{e}\kappa' = -i\hat{e}\beta = i\hat{e}\beta' = (\sqrt{2}/4)\delta t^{-1}, \quad \text{(III25)}$$

$$\sigma = \sigma' = (\sqrt{2}/4)t^{-1}\{(1 - \delta) + i\bar{a}\},$$

$$\rho = \rho' = -(\sqrt{2}/4)t^{-1}\{(1 + \delta) + i\bar{a}\}. \quad \text{(III25)}$$

The spin coefficients for the type V solution are just (III25) with $\bar{a} = 0$. Substitution of (III25) into (III20)–(III24) results in $\Psi_0 = \Psi_1 = \Psi_3 = \Psi_4 = 0$ for both solutions. For the type VII_h solution $\Psi_2 = (\delta t^{-2}/12)\{(\delta - 1) - i\bar{a}\}$. Ψ_2 for the type V solution is again obtained by setting $\bar{a} = 0$. Thus, both solutions are of Petrov Type *D*.

[1] B. Kuchowicz, "Neutrinos in general relativity: four (?) levels of approach," Gen. Rel. Grav. **5**, p. 201 (1974).
[2] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 504.
[3] R. Ruffini and S. Bonazzola, "Systems of self-gravitating particles in general relativity and the concept of an equation of state," Phys. Rev. **187**, 1767 (1969).
[4] M. A. Melvin and T. R. Michalik "Symmetries and metrics of homogeneous cosmologies," J. Math. Phys. **21**, 1938 (1980), referred to as SMHC.
[5] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies*, (Princeton, U. P., Princeton, New Jersey, 1975), p. 149.
[6] F. B. Estabrook, H. D. Wahlquist, and C. G. Behr, "Dyadic Analysis of Spatially Homogeneous World Models," J. Math. Phys. **9**, 497 (1968).
[7] G. F. R. Ellis, "Relativistic cosmology," Enrico Fermi International School of Physics, Course XLVII, (AP 1971), p. 111.
[8] D. R. Brill and J. M. Cohen, "Cartan frames and the general relativistic Dirac equation," J. Math. Phys. **7**, 238 (1966).
[9] D. R. Brill and J. A. Wheeler, "Interaction of neutrino and gravitational fields," Rev. Mod. Phys. **29**, 465 (1957).
[10] O. Bergmann, "Problem of Rainich for two component spinors," J. Math. Phys. **1**, 172 (1960).
[11] T. M. Davis and J. R. Ray, "Ghost neutrinos in general relativity," Phys. Rev. D **9**, 331 (1974).
[12] T. M. Davis and J. R. Ray, "Ghost neutrinos in plane-symmetric spacetimes," J. Math. Phys. **16**, 75 (1975); "Neutrinos in cylindrically-symmetric spacetimes," J. Math. Phys. **16**, 80 (1975).
[13] I. M. Anderson, "The neutrino energy–momentum tensor and the Weyl equations in curved spacetime," Gen. Rel. Grav. **5**, 621 (1974).
[14] P. S. Letelier, "Ghost neutrinos in the Einstein–Cartan theory of gravitation," Phys. Lett. A **54**, 351 (1975).
[15] M. A. Melvin, "Homogeneous axial cosmologies with electromagnetic fields and dust," Ann. N.Y. Acad. Sci. **262**, 253 (1975).
[16] T. R. Michalik, "Homogeneous cosmologies with strong neutrino fields," Ph.D. thesis, Temple University, Philadelphia, Pennsylvania, 1976.
[17] Reference 2, p. 735.
[18] Reference 2, p. 515.
[19] A. H. Taub, "Empty spacetimes admitting a three parameter group of motions," Ann. Math. **53**, 472 (1951).
[20] E. Newman and R. Penrose, "An approach to gravitational radiation by a method of spin coefficients," J. Math. Phys. **3**, 566 (1962).
[21] R. Geroch, A. Held, and R. Penrose, "A space–time calculus based on pairs of null directions," J. Math. Phys. **14**, 874 (1973).

ERRATA

Erratum: Phase space path integrals, without limiting procedure [J. Math. Phys. 19, 298 (1978)]

Maurice M. Mizrahi
Center for Naval Analyses, University of Rochester, 2000 N. Beauregard St., Alexandria, Virginia 22311

(Received 19 March 1980; accepted for publication 28 March 1980)

1. In (68) and (73), replace $Y(t-t')$ by $Y(t'-t)$ and $Y(t'-t)$ by $Y(t-t')$.

2. In (26) and below (68), replace "$\lim_{t\to t'} - \lim_{t'\to t}$" by "$(\lim_{t-t'\to 0^+}) - (\lim_{t-t'\to 0^-})$".

3. In (92), put "$du_1\cdots du_n dv_1\cdots dv_m$" before $\mathscr{F}^{-1}$ and, in last line, replace $(\xi^i u_i, \eta^j v_j)$ by $(\xi^i \mu_i, \eta^j \nu_j)$.

4. In (104), put parentheses around $|c|\sqrt{\det A}$.

5. In (120), replace "$\lim_{t\to t'}$" by "$\lim_{t-t'\to 0^+}$" and, in last line, replace m by $1/m$.

6. In (3), open a curly bracket before $i/\hbar$ and close it after dt.

7. Delete the "note" before (23). It duplicates that below (22).

8. In (39), replace f by F.

9. In (54), next to last line, insert "dt" before $B(t)$ and before $f(t)$.

10. In (130), insert "$\int_T dt$" before "$\frac{1}{2}$".

11. Before "V. CONCLUSION", p. 306, insert: "If H has terms coupling p and q, then the corresponding path integral must be treated as described in Maurice M. Mizrahi, "Correspondence rules and path integrals," Springer-Verlag Lecture Notes in Physics **106**, 234 (1979).

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates Erratum.*

 0022-2488/80/071965-01$0.00