Shift operators for discrete representations of O(p, q)

T. Murphy
Physics Department, University of Wisconsin–Madison, Madison, Wisconsin 53706

(Received 18 March 1980; accepted for publication 23 May 1980)

Shift operators for discrete representations of O(*p*,*q*) are found in product form. Their normalization is found and conditions are derived for the weight components of such representations.

INTRODUCTION

Various noncompact orthogonal groups have arisen in the study of physical problems, such as the De Sitter group, conformal group, etc. The unitary representations of such groups are all infinite dimensional, and no general theory for them exists. The case of O(*n*,1) has been considered by Schwarz,[1] Wong,[2] and others, using an approach based on the properties of the O(*n*) subgroup. The idea of shift operators for the orthogonal groups was introduced by Pang and Hecht[3] and further developed by Wong[4] and Bincer.[5] In this paper I extend Bincer's approach to the noncompact othogonal groups with discrete representations.

1. DEFINITIONS AND NOTATION

My notation is essentially that of Bincer,[5] and the reader is referred to this paper for more complete discussion of the definitions used here.

Discrete series representations of O(*p*,*q*) can occur only if either *p* or *q* is even. Since O(*p*,*q*) is isomorphic to (*q*,*p*), there is no loss of generality in assuming *p* is always even. There are then two cases:

A. *p* even, *q* odd. Then O(*p*,*q*) is generated by operators G^a_b, $\bar{\nu}\leqslant a,b\leqslant\nu$, where $\nu=\frac{1}{2}(p+q-1)$, $\bar{\nu}=-\nu$.

B. *p* even, *q* even. O(*p*,*q*) is generated by operators G^a_b, $\bar{\nu}\leqslant a,b\leqslant\nu$, $a,b\neq 0$, where $\nu=\frac{1}{2}(p+q)$.

For the noncompact algebras the indices are divided into two blocks, symmetric under $i\to -i$, as follows:

$$\text{Block One}\quad \{i|1\leqslant|i|\leqslant p/2\}$$
$$\text{Block Two}\quad \{0\}\cup\{i|p/2+1\leqslant|i|\leqslant\nu\}. \tag{1.1}$$

Notice that the index 0 is present in the second block only if the algebra is odd-dimensional.

The generators are defined by the commutation relations

$$[G^a_b, G^c_d]=\delta^c_b G^a_d-\delta^a_d G^c_b+\delta^{\bar{b}}_d G^c_{\bar{a}}-\delta^c_{\bar{a}} G^{\bar{b}}_d \tag{1.2}$$

and have the property

$$G^a_b=-G^{\bar{b}}_{\bar{a}}. \tag{1.3}$$

For a unitary representation the generators must also satisfy

$$(G^a_b)^\dagger=\begin{cases} G^b_a & \text{if } a \text{ and } b \text{ are in the same block}\\ -G^b_a & \text{if } a \text{ and } b \text{ are in different blocks.}\end{cases} \tag{1.4}$$

Equations (1.3) and (1.4) impose the requirement that the blocks of indices be symmetric, and therefore preclude the use of this approach for O(*p*,*q*) where both *p* and *q* are odd.

The operators G^a_a, $a=1\cdots\nu$ all commute with each other and may all be taken to be diagonal. Their eigenstates will be used as the basis for the representation, and their eigenvalues are called the components of the weight of the state,

$$G^a_a|w_1w_2\cdots w_\nu\rangle=w_a|w_1\cdots w_\nu\rangle. \tag{1.5}$$

For the compact algebra O(*n*), a representation is specified completely by the values of the Casimir operators, which are functions of the *w*'s. A given representation will have both a lowest weight state and a highest weight state, and the entire representation may be constructed by the appropriate application of shift operators, starting from either state. For the noncompact algebra O(*p*,*q*) the values of the Casimir operators do not uniquely specify the represenation, and the representations obtained from the highest and lowest weight states corresponding to the same Casimir values are inequivalent. I shall therefore have to treat these two cases separately. For representations with a highest weight I introduce the ordering

$$w_{i+1}\geqslant w_i\quad i=1\cdots\nu-1. \tag{1.6}$$

For representations with a lowest weight state the ordering is

$$\tilde{w}_{i+1}\leqslant\tilde{w}_i\quad i=1\cdots\nu-1. \tag{1.7}$$

Here and elsewhere a tilde indicates a lowest weight state.

With this ordering I have in both cases that G^a_b is a

$$\begin{cases}\text{raising}\\ \text{weight operator,}\\ \text{lowering}\end{cases}\quad\text{if}\quad\begin{cases}a>b\\ a=b.\\ a<b\end{cases}$$

An O(*p*,*q*) one tensor-operator V_a is defined by the commutation relations

$$[G^a_b, V_d]=-\delta^a_d V_b+\delta^{\bar{b}}_d V_{\bar{a}}. \tag{1.8}$$

$$\text{A one-tensor is a}\begin{cases}\text{raising}\\ \text{weight operator if,}\\ \text{lowering}\end{cases}\begin{cases}d<0\\ d=0.\\ d>0\end{cases}$$

2. SUBGROUP CHAINS

The complete labelling of the states in a representation requires more than the specification of the state of highest or lowest weight. One labelling scheme is to specify the weight components for the states of highest or lowest weight of all the subgroups in a given subgroup chain, a technique introduced by Gel'fand and Tsetlin.[6] The chain of subgroups I shall consider is

 0022-2488/80/092317-04$1.00

$$O(p,q)\supset O(p,q-1)\supset O(p,q-2))\supset$$
$$\cdots\supset O(p,1)\supset O(p)\supset\cdots\supset O(2). \tag{2.1}$$

The generators of the subgroups are defined by Eqs. (1.2)–(1.4) as before, except that the range of the indices is restricted. If l is the rank of a noncompact subgroup, the range of indices is given by

$$l \text{ odd} \quad |i|\in\{0\}\cup\{1\cdots p/2\}\cup\{p/2+1+\nu-(l-1)/2,\cdots\nu\} \tag{2.2}$$

$$l \text{ even} \quad |i|\in\{1\cdots p/2\}\cup\{p/2+1+\nu-l/2,\cdots\nu\}. \tag{2.3}$$

Notice that passing down the subgroup chain from an odd rank subgroup to an even rank one simply involves deleting zero from the range of the indices. Passing from an even rank subgroup to an odd rank one requires the range of the indices to change from

$$|i|\in\{1\cdots p/2\}\cup\{\delta,\cdots\nu\} \quad \delta\equiv p/2+1+\nu-l/2 \tag{2.4}$$

to

$$|i|\in\{0\}\cup\{1\cdots p/2\}\cup\{\delta+1,\cdots\nu\}. \tag{2.5}$$

Here both the values $\pm\delta$ have been deleted, and zero has been added. A suitable definition for those operators involving the new index is

$$G^o_a=\frac{G^\delta_a+G^{\bar\delta}_a}{\sqrt{2}},\ G^0_0=\tfrac{1}{2}(G^\delta_\delta+G^{\bar\delta}_{\bar\delta}),\ V_0=\frac{V_\delta+V_{\bar\delta}}{\sqrt{2}}. \tag{2.6}$$

Eventually, as one proceeds down the subgroup chain, there will remain only compact subgroups. These may be treated using the results of Bincer,[5] with a slight modification. For the noncompact subgroups, the weight components of each subgroups label a state of highest (or lowest) weight. For the compact subgroups these labels are for the state of lowest (or highest) weight, resulting in an inversion of the ordering of the weight components. To apply Bincer's results I must define a new index by

$$a'=p/2-a+1,\ 0'=0. \tag{2.7}$$

Bincer's results can then be used with the primed indices, and the range of the indices for compact subgroups is given by

$$l \text{ odd} |i|\in\{0\}\cup\{1\ldots(l-1)/2\}$$

or

$$|i'|\in\{0\}\cup\{p/2-l-1/2+1,\ldots p/2\} \tag{2.8}$$
$$l \text{ even} |i|\in\{1\ldots l/2\}$$

or

$$|i'|\in\{p/2-l/2+1,\ldots p/2\}. \tag{2.9}$$

As an example, a state in a representation of O(4,5) with a highest weight state would be specified by a state vector with labels m^l_i representing the ith weight component of the rank l subgroup as

$$\left|\begin{array}{cccc|l} m^9_1 & m^9_2 & m^9_3 & m^9_4 & \text{highest state of O(4,5)}\\ m^8_1 & m^8_2 & m^8_3 & m^8_4 & \text{highest state of O(4,4)}\\ m^7_1 & m^7_2 & & m^7_4 & \text{highest state of O(4,3)}\\ m^6_1 & m^6_2 & & m^6_4 & \text{highest state of O(4,2)}\\ m^5_1 & m^5_2 & & & \text{highest state of O(4,1)}\\ m^4_1(m^4_{2'}) & m^4_2(m^4_{1'}) & & & \text{lowest state of O(4)}\\ m^3_1(m^3_{2'}) & & & & \text{lowest state of O(3)}\\ m^2_1(m^2_{2'}) & & & & \text{lowest state of O(2)} \end{array}\right| \tag{2.10}$$

3. SHIFT OPERATORS

All the states in a discrete representation may be generated from a single state by the repeated application of appropriate shift operators. The action of such an operator on a state is given by

$${}^lS_\mu\left|\begin{array}{l} m^{p+q}_i\\ \cdot\\ \cdot\\ m^l_i\\ m^{l-1}_i \end{array}\right\rangle \quad \alpha \quad \left|\begin{array}{l} m^{p+q}_i\\ \cdot\\ \cdot\\ m^l_i\\ m^{l-1}_i-\delta_{\mu i}+\delta_{i\bar\mu} \end{array}\right\rangle. \tag{3.1}$$

The operator ${}^lS_\mu$ changes the $|\mu|$ component of the weight of the rank $l-1$ subgroup by one unit up or down, depending on the sign of μ.

For those representations with a highest weight state, I may simply adopt the results of Bincer,[5] since these depend only on the commutation relations (1.2) and the definition (3.1). In that case I have

$${}^lS_\mu=V(\mu)_\mu \tag{3.2}$$

$$V(\mu)_a=\{V(\bar\nu)\prod_{j=\bar\nu}^{\mu-1}{}'(G-c^{l-1}_j)\}_a \tag{3.3}$$

$$V(\bar\nu)_a=G^0_a \quad l \text{ odd} \tag{3.4}$$
$$=\frac{1}{\sqrt{2}}(G^\delta_a-G^{\bar\delta}_a) \quad l \text{ even}.$$

The prime on the product indicates that the range of the index j is that of the subgroup with rank $l-1$. The constants c^{l-1}_j are determined by the condition

$$c^{l-1}_j=m^{l-1}_j+\sum_{b=\bar\nu}^{\mu-1}{}'(1-\delta^{\mu-1}_b) \tag{3.5}$$

and are given by

$$l \text{ odd} \quad c^l_j=\begin{cases} m^l_j+\nu+j-(2\nu-l+2)\theta_j & |j|>p/2\\ m^l_j+(l-1)/2+j-\theta_j & |j|\leqslant p/2 \end{cases} \tag{3.6}$$

$$l \text{ even} \quad c^l_j=\begin{cases} m^l_j+\nu+j-(2\nu-l+2)\theta_j & |j|>p/2\\ m^l_j+l/2+j-2\theta_j & |j|\leqslant p/2 \end{cases} \tag{3.7}$$

For the representations which have a lowest weight state one may follow the same iterative procedure used for the case above, except that it must be based on the annihilation of the lowest state by all lowering operators. The results in this case are

$$ {}^{l}S_{\mu} = \widetilde{V}(\mu)_{\mu} \tag{3.8} $$

$$ \widetilde{V}(\mu)_{a} = \{\widetilde{V}(\nu) \prod_{j=\mu+1}^{\nu} (G - \tilde{c}_{j}^{l-1})\}_{a} \tag{3.9} $$

$$ \widetilde{V}(\nu)_{a} = \{ G_{a}^{0} \quad l \text{ odd} $$
$$ = \frac{1}{\sqrt{2}}(G_{a}^{\delta} - G_{a}^{\bar{\delta}}) \quad l \text{ even}. \tag{3.10} $$

The constants c_j^{l-1} are given by

$$ l \text{ odd} \quad \tilde{c}_{j}^{l} = \begin{cases} \tilde{m}_{j}^{l} + \nu - j - (2\nu - l + 2)\theta_{\bar{j}} & |j| > p/2 \\ \tilde{m}_{j}^{l} + (l-1)/2 - j - \theta_{\bar{j}} & |j| \leqslant p/2 \end{cases} \tag{3.11} $$

$$ l \text{ even} \quad \tilde{c}_{j}^{l} = \begin{cases} \tilde{m}_{j}^{l} + \nu - j - (2\nu - l + 2)\theta_{\bar{j}} & |j| > p/2 \\ \tilde{m}_{j}^{l} + l/2 - j - 2\theta_{\bar{j}} & |j| \leqslant p/2 \end{cases}. \tag{3.12} $$

From the definitions (3.4) and (3.10) and Eq. (1.4) I see

$$ V^{+}(\vec{\nu})_{a} = (-1)^{l} \begin{cases} -V(\vec{\nu})_{\bar{a}} & 1 \leqslant |a| \leqslant p/2 \\ V(\vec{\nu})_{\bar{a}} & a = 0 \text{ or } |a| > p/2 \end{cases} \tag{3.13} $$

$$ \widetilde{V}^{+}(\vec{\nu})_{a} = (-1)^{l} \begin{cases} -\widetilde{V}(\nu)_{\bar{a}} & 1 \leqslant |a| \leqslant p/2 \\ \widetilde{V}(\nu)_{\bar{a}} & a = 0 \text{ or } |a| > p/2. \end{cases} \tag{3.14} $$

4. NORMALIZATION OF THE SHIFT OPERATORS

The shift operators found in Sec. 3 are defined only up to a normalization factor. The evaluation of this factor is crucial in determining the states in a representation and will impose certain "betweeness" conditions on the m_i^l.

The normalization coefficient for the shift operator ${}^{l}S_{\mu}$ is defined by

$$ {}^{l}S_{\mu} \left| \begin{matrix} m_i^{p+q} \\ \cdot \\ \cdot \\ m_i^{l} \\ m_i^{l-1} \end{matrix} \right\rangle = \left(\begin{matrix} m_i^{l-1} \\ \\ \\ \\ m_i^{l-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{matrix} \right) \left| \begin{matrix} m_i^{p+q} \\ \cdot \\ \cdot \\ m_i^{l} \\ m_i^{l-1} - \delta_{\mu i} + \delta_{\bar{\mu} i} \end{matrix} \right\rangle . \tag{4.1} $$

The evaluation of these coefficients is lengthy and follows very closely the procedure used by Bincer in Ref. 7, modified to include the different range of indices and the altered hermeticity properties of $V(\vec{\nu})$ or $\widetilde{V}(\nu)$ in the noncompact algebras. The resulting expressions for the coefficients for representations with a highest weight state are:

l odd, $l > p$

$$ \left| \binom{m_i^{l-1}}{m_i^{l-1} - \delta_{i\mu} + \delta_{i\bar{\mu}}} \right|^2 $$
$$ = \prod_{j=\nu}^{\bar{\nu}}{}'' (c_{\bar{\mu}}^{l-1} - c_j^l + 1) \prod_{j=\bar{\mu}+1}^{\nu}{}' (c_{\bar{\mu}}^{l-1} - c_j^{l-1}) $$
$$ \times \prod_{j=\bar{\nu}}^{\bar{\mu}}{}' \frac{1}{c_{\bar{\mu}}^{l-1} - c_j^{l-1} + 1 - \delta_{j\mu}} \times \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} $$
$$ \text{for} \begin{cases} 1 \leqslant |\mu| \leqslant p/2 \\ |\mu| > p/2 \end{cases} \tag{4.2} $$

l even, $l > p$, $0 < \alpha \leqslant p/2$

$$ \left| \binom{m_i^{l-1}}{m_i^{l-1} - \delta_{i\alpha} + \delta_{i\bar{\alpha}}} \right|^2 $$
$$ = \{c_{\bar{\alpha}}^l - c_{\alpha}^{l-1} + 1)(c_{\bar{\alpha}}^{l-1} - l/2 + 2 + m_{\delta}^l) $$
$$ \times (c_{\bar{\alpha}}^{l-1} - l/2 + 2 - m_{\delta}^l) - 4(c_{\bar{\alpha}}^{l-1} - l/2 + 2)^2 \} $$
$$ \times \prod_{j=\bar{\nu}}^{(\bar{p}/2)-1}{}' (c_{\bar{\alpha}}^{l-1} - c_j^l + 1) \prod_{j=\bar{p}/2}^{\bar{1}}{}' (c_{\bar{\alpha}}^{l-1} - c_j^l + 2) $$
$$ \times \prod_{j=1}^{p/2}{}' (c_{\bar{\alpha}}^{l-1} - c_j^l) \prod_{j=p/2+1}^{\nu}{}' (c_{\alpha}^{l-1} - c_j^l + 1) $$
$$ \times \prod_{j=\bar{\alpha}+1}^{\nu}{}' (c_{\bar{\alpha}}^{l-1} - c_j^{l-1}) $$
$$ \times \prod_{j=\bar{\nu}}^{\bar{\alpha}}{}' = (c_{\bar{\alpha}}^{l-1} - c_j^{l-1} + 1 + \delta_{j\alpha})^{-1} \tag{4.3} $$

l even, $l > p$, $\mu > p/2$

$$ \left| \binom{m_i^{l-1}}{m_i^{l-1} - \delta_{i\mu} + \delta_{i\bar{\mu}}} \right|^2 $$
$$ = (c_{\delta}^l - c_{\bar{\mu}}^{l-1} - p/2 - 1)(c_{\delta}^l - c_{\mu}^{l-1} - p/2) $$
$$ \times \prod_{j=\bar{\nu}}^{\bar{p}/2-1}{}' (c_{\bar{\mu}}^{l-1} - c_{j+1}^l) \times \prod_{j=\bar{p}/2}^{\bar{1}}{}' (c_{\bar{\mu}}^{l-1} - c_j^{l-1} + 2) $$
$$ \times \prod_{j=1}^{p/2}{}' (c_{\bar{\mu}}^{l-1} - c_j^l) \prod_{j=(p/2)+1}^{\nu}{}' (c_{\bar{\mu}}^{l-1} - c_j^l + 1) $$
$$ \times \prod_{j=\bar{\mu}+1}^{\nu}{}' (c_{\bar{\mu}}^{l-1} - - c_j^{l-1}) $$
$$ \times \prod_{j=\bar{\nu}}^{\bar{\mu}}{}' \frac{1}{c_{\bar{\mu}}^{l-1} - c_j^{l-1} + 1 + \delta_{j\mu}} . \tag{4.4} $$

In Eq. (4.2) the double prime on the first product indicates the range is that of the rank l group. For $l \leqslant p$, the coefficients are those given by Bincer with the substitution $i \to i'$, as defined in Eq. (2.6). The requirement that the above expressions be nonnegative yields the following conditions on the m_i^l:

$$ l > p \quad m_i^l \geqslant m^{l-1}{}_i^{-1} \geqslant m_{i-1}^l \tag{4.5} $$

$$ m_{\mu}^{l-1} \geqslant \nu - (l-1)/2 - \mu + 1 \quad \mu > p/2, \ l \text{ odd} \tag{4.6} $$

$$ m_{\mu}^l \leqslant 1 - \mu \quad \mu \leqslant p/2, \qquad l \text{ odd} \tag{4.7} $$

$$ m_{\mu}^{l-1} \geqslant \nu - l/2 - \mu + 2 + |m_{\delta}^l| \quad \mu > p/2, \ l \text{ even} \tag{4.8} $$

$$ m_{\mu}^l \leqslant -1 - \mu - |m_{\delta}^l| \quad \mu \leqslant p/2, \qquad l \text{ even} \tag{4.9} $$

$$ l \leqslant p \quad m_i^l \leqslant m_i^{l-1} \leqslant m_{i+1}^l \leqslant 0. \tag{4.10} $$

It is possible for a label to be outside the range indicated above if it is equal to the next smaller label. The value of such a weight component cannot be shifted, however, and representations including such labels are called degenerate representations, as discussed by Nikolov.[8] The most degenerate case, where all the labels are initially equal, has been considered by Niederle.[9]

The results for representations with a lowest weight state are the same as the above, with c_i replaced by $\tilde{c}_i$ in Eqs. (4.2)–(4.4); the limits on the products sent to their negatives ($\Pi_{j=\bar{\nu}}^{\bar{\alpha}} \to \Pi_{j=\alpha}^{\nu}$), and $-p/2 \to p/2$ in Eq. (4.4). The conditions (4.5)–(4.10) also apply, with m replaced by $-\tilde{m}$.

5. CONCLUSION

For odd-dimensional algebras there are two inequivalent representations with the same value of the quadratic Casimir operators, one with a lowest weight state and one with a highest weight state. For even-dimensional algebras, there is another possibility. There are two choices of subgroup chains unless $p = q$, because $O(p,q)$ is isormorphic to $O(q,p)$. Thus in general there will be four inequivalent representations of $O(p,q)$ for p,q even and $p \neq q$.

Comparison of my results with those of Schwarz for the case $O(p,1)$ shows our results to be in agreement if the indices are switched as in Eq. (2.6). The results of Wong for $O(p,1)$ seem to include othere discrete representations than the ones I have found. However, further analysis of his results[10] shows that the only allowed value of S in Eqs. (17) and (18) of Ref. 2 is $S = K$, and with this restriction our results are in agreement.

Nikolov[8,11] has given conditions, without proof, on the weight components for discrete series representations of $O(p,q)$. For the case of a representation with a lowest weight state our results are in agrement, with the following relabeling of weight components:

Nikolov Murphy

$$m_{j,i} \rightarrow \begin{cases} \tilde{m}_i^{j+1} & j < p \\ \tilde{m}_i^{j+1} + p + 1 - j & j \geqslant p, \quad i \leqslant p/2 \\ -\tilde{m}_{v+p/2+1-i}^{j+1} & j \geqslant p, \quad i > p/2 \end{cases} . \tag{5.1}$$

Nikolov does not consider representations with a highest weight state.

ACKNOWLEDGMENTS

I would like to thank Dr. A. M. Bincer for many helpful discussions and patient encouragement, and Dr. M. K. F. Wong for useful correspondence.

[1] F. Schwarz, J. Math. Phys. **12**, 131 (1971).
[2] M. K. F. Wong, J. Math. Phys. **15**, 25 (1974).
[3] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[4] M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967).
[5] A. M. Bincer, J. Math. Phys. **19**, 1173 (1978).
[6] I. Gel'fand and M. Tsetlin, Dokl. Akad. Nauk. USSR **71**, 1017 (1950).
[7] A. M. Bincer, J. Math. Phys. **19**, 1179 (1978).
[8] A. V. Nikolov, Funkt. Anal. Ego Pril. **4**, 91 (1969).
[9] J. Niederle, J. Math. Phys. **8**, 1921 (1967).
[10] M. K. F. Wong, private communication.
[11] A. V. Nikolov, Funkt. Anal. Ego Pril. **2**, 99 (1968).

The character generator of SU(n)[a)]

Richard P. Stanley
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 15 October 1979; accepted for publication 25 March 1980)

A simple combinatorial method for writing the character generator of SU(n) is described.

1. INTRODUCTION

Generating functions have proved to be a useful tool in the representation theory of continuous and discrete groups.[1] In the case of a compact semisimple Lie group G, the character generator is the starting point for obtaining many other generating functions of interest. The character generator for irreducible representations of a connected simply-connected semisimple Lie group G is defined by

$$X_A(\alpha) = \sum \chi_\lambda(\alpha) A_1^{r_1} \cdots A_l^{r_l},$$

where l is the rank of G, the summation extends over all nonnegative integers $r_1,\ldots,r_l$, and $\chi_\lambda(\alpha)$ is the character of the finite irreducible representation of G with highest weight $\lambda = r_1\lambda_1 + \cdots + r_l\lambda_l$. Here $\lambda_1,\ldots,\lambda_l$ are the fundamental weights of G. Thus the coefficient of $A_1^{r_1}\cdots A_l^{r_l} \times \alpha_1^{\mu_1}\cdots\alpha_l^{\mu_l}$ (which we abbreviate as $A^r\alpha^\mu$) in $X_A(\alpha)$ is the multiplicity of the weight $\mu = (\mu_1,\ldots,\mu_l)$ (written with respect to some basis for the weight space). It follows easily from Weyl's character formula that $X_A(\alpha)$ is a rational function of A and α. For many applications it is desirable to write $X_A(\alpha)$ as a sum of terms of the form

$$A^s\alpha^v \Big/ \prod_{i=1}^{d} (1 - A_j \alpha_{l_1}\alpha_{l_2}\cdots\alpha_{l_h}), \tag{1}$$

where j, h, and $l_1,l_2,\ldots,l_h$ depend on i, and where d is the same for all terms and is necessarily equal to $\frac{1}{2}(\dim G + \operatorname{rank} G)$. The method[1] used for computing $X_A(\alpha)$ does not directly yield a sum of terms of the form (1), and it is unknown in general whether $X_A(\alpha)$ can always be written in this form. We will describe a different method for computing $X_A(\alpha)$ when $G = \mathrm{SU}(n)$, which automatically expresses $X_A(\alpha)$ as a sum of terms (1). Each term can be read off by inspection from a certain type of tableau, and we state a formula for the total number of terms. Our derivation will be purely combinatorial, based on the well-known description of the characters of SU(n) in terms of Young tableaux.

2. BASIC CONCEPTS AND FUNDAMENTAL THEOREMS

We now introduce the necessary combinatorial concepts and terminology. Fix integers $m_1 > m_2 > \cdots > m_k > 0$, and set $\mathbf{m} = (m_1,\ldots,m_k)$. Let $\mathbf{r} = (r_1,\ldots,r_k)$ be a k–tuple of nonnegative integers, and let $Y_\mathbf{r}$ be the Young diagram with r_i columns of length i. Thus $Y_\mathbf{r}$ is a left-justified array of squares, with $r_i + r_{i+1} + \cdots + r_k$ squares in row i. Let ρ be an array obtained by inserting positive integers into the squares of $Y_\mathbf{r}$ subject to the rules: (i) Every row is non-increasing, (ii) every column is strictly decreasing, and (iii) no entry in row i exceeds m_i. For instance, if $\mathbf{m} = (5,4,2)$ and $\mathbf{r} = (4,2,3)$, then a typical ρ looks like

5 5 4 4 4 3 1 1 1
3 3 2 2 2
2 1 1.

We call ρ a *column-strict plane partition*[2] of *type* $(\mathbf{m},\mathbf{r})$. Introduce new variables $X_1,X_2,\ldots$, and set

$$M(\rho) = X_1^{a_1}X_2^{a_2}\cdots,$$

where a_i parts of ρ are equal to i. Thus, for the above example, $M(\rho) = X_1^5X_2^4X_3^3X_4^3X_5^2$. In general, $a_i = 0$ if $i > m_1$, and $\sum a_i = \sum i r_i$. Given $\mathbf{m} = (m_1,\ldots,m_k)$, define the generating function

$$F_\mathbf{m}(A,X) = \sum_\rho A^r M(\rho), \tag{2}$$

where the sum is over all column-strict plane partitions ρ of type $(\mathbf{m},\mathbf{r})$ for some $\mathbf{r} = (r_1,\ldots,r_k)$. We will give a method for computing $F_\mathbf{m}(A,X)$ as a sum of terms of the form

$$A^sX^v \Big/ \prod_{i=1}^{m} (1 - A_j X_{l_1}\cdots X_{l_j}), \tag{3}$$

where j and $l_1,\ldots,l_j$ depend on i, and where $m = m_1 + \cdots + m_k$. From this it will be easy to obtain the character generator for SU(n).

We now define the type of tableaux necessary to describe the terms (3) of $F_\mathbf{m}(A,X)$. A *shifted Young diagram* $Z_\mathbf{m}$ of shape $\mathbf{m} = (m_1,\ldots,m_k)$ consists of an array of $m = m_1 + \cdots + m_k$ squares, with m_i squares in row i, and with row $i+1$ indented one space to the right from row i. A *standard shifted Young tableau* (SSYT) of shape $\mathbf{m}$ is obtained by inserting the integers $1,2,\ldots,m$ into the squares of $Z_\mathbf{m}$ without repetition such that every row and column is increasing.[3] For instance, an example of an SSYT of shape (7,4,3,2) is given by

1 2 3 5 9 14 16
4 6 7 10
8 11 13
12 15. (4)

If π is an SSYT, define the *sub*-SSYT $\pi^{(i)}$ to be the SSYT obtained from π by deleting all entries $> i$. For instance, if π is given by (4), then $\pi^{(16)} = \pi$, $\pi^{(3)} = 123$, and

$$\pi^{(8)} = \begin{array}{cccc} 1 & 2 & 3 & 5 \\ & 4 & 6 & 7 \\ & & 8. & \end{array}$$

[a)]Partially supported by NSF Grant MCS 78-02743.

 0022-2488/80/092321-06$1.00

If π is an SSYT of shape $\mathbf{m} = (m_1,\ldots,m_k)$, define a monomial $\Gamma(\pi) = A_k X_{m_1} X_{m_2} \cdots X_{m_k}$. For instance, if π is given by (4), then $\Gamma(\pi) = A_4 X_2 X_3 X_4 X_7$ and $\Gamma(\pi^{(8)}) = A_3 X_1 X_3 X_4$. We now state the fundamental theorem which explains how a formula for $F_{\mathbf{m}}(A,X)$ can be read off from the set of all SSYT of shape $\mathbf{m}$.

Theorem: (i) We have

$$F_{\mathbf{m}}(A,X) = \sum_{\pi} \prod_{i \in K_\pi} \Gamma(\pi^{(i)}) \Big/ \prod_{i=1}^{m} [1 - \Gamma(\pi^{(i)})], \tag{5}$$

where π ranges over all SSYT of shape $\mathbf{m}$, and K_π is the set of those i for which $i+1$ appears in π in a row above i.

(ii) To obtain the character generator for SU(n) in the form (1), with respect to the basis $\lambda_1,\ldots,\lambda_{n-1}$ of fundamental weights, take $\mathbf{m} = (n,n-1,\ldots,2)$ in (5) and set $X_i = \alpha_{i-1}^{-1}\alpha_i$ for $1 \leqslant i \leqslant n$ (where we set $\alpha_0 = \alpha_n = 1$). (If one prefers the characters with respect to a different basis for the weight space, replace each α_i by an appropriate $\alpha_1^{u_1}\cdots\alpha_{n-1}^{u_{n-1}}$.) More generally, if $\lambda_1,\ldots,\lambda_{n-1}$ are the fundamental weights of SU(n) in their usual order, then to get the generating function for those characters of SU(n) corresponding to a highest weight $r_1\lambda_1 + \cdots + r_k\lambda_k$ for some fixed $k \leqslant n-1$, take $\mathbf{m} = (n,n-1,\ldots,n-k+1)$ and $X_i = \alpha_{i-1}^{-1}\alpha_i$, $1 \leqslant i \leqslant n$.

(iii) The number $g^{\mathbf{m}}$ of terms in the sum (5) (equivalently, the number of SSYT of shape $\mathbf{m}$) is given by

$$g^{\mathbf{m}} = \frac{m!}{m_1!\cdots m_k!} \prod_{1 \leqslant i < j \leqslant k} \frac{m_i - m_j}{m_i + m_j},$$

where $\mathbf{m} = (m_1,\ldots,m_k)$. In particular,

$$g^{(n,n-1,\ldots,2)} = \begin{cases} \dfrac{\binom{n}{2}+1)!2!4!\cdots(n-2)!}{(n+1)!(n+3)!\cdots(2n-1)!}, & n \text{ even} \\ \dfrac{\binom{n}{2}+1)!2!4!\cdots(n-1)!}{n!(n+2)!\cdots(2n-1)!}, & n \text{ odd}. \end{cases}$$

3. PROOF OF FUNDAMENTAL THEOREM

(i) The right-hand side of (5) may be rewritten as

$$\sum_{\pi} \sum_{b_1,\ldots,b_m} \Gamma(\pi^{(1)})^{b_1} \cdots \Gamma(\pi^{(m)})^{b_m}, \tag{6}$$

where $b_1,\ldots,b_m$ ranges over all sequences of nonnegative integers such that $b_i > 0$ if $i \in K_\pi$. To each term $\Gamma(\pi^{(1)})^{b_1}\cdots\Gamma(\pi^{(m)})^{b_m}$ of (6), associate a column-strict plane partition ρ by defining ρ to have b_i columns with entries $l_1 > \cdots > l_j$, where $\pi^{(i)}$ has shape $(l_1,\ldots,l_j)$. If ρ is of type $(\mathbf{m},\mathbf{r})$ then $\Gamma(\pi^{(1)})^{b_1}\cdots\Gamma(\pi_{(m)})^{b_m}$ is just the monomial $A^{\mathbf{r}}M(\rho)$ appearing in (2). Hence to prove (i), we need to show that the map $(\pi,\mathbf{b}) \longrightarrow \rho$ defined above between (a) ordered pairs $(\pi,\mathbf{b})$ where π is a SSYT of shape $\mathbf{m}$ and $\mathbf{b}$ is a sequence of nonnegative integers $b_1,\ldots,b_m$ such that $b_i > 0$ if $i \in K_\pi$, and (b) column-strict plane partitions ρ of type $(\mathbf{m},\mathbf{r})$ for some $\mathbf{r}$, is a one-to-one correspondence.

Given $(\pi,\mathbf{b})$ define $a_i = b_i + b_{i+1} + \cdots + b_m$. Thus $a_1 \geqslant \cdots \geqslant a_m \geqslant 0$, and $a_i > a_{i+1}$ if $i \in K_\pi$. Clearly we can recover $\mathbf{b}$ from $\mathbf{a} = (a_1,\ldots,a_m)$ by $b_i = a_i - a_{i+1}$. Now let σ be the array obtained by replacing i in π by a_i. Then σ is a *shifted plane partition*[3] of shape $\mathbf{m}$, i.e., an array obtained by inserting nonnegative integers into the squares of $Z_{\mathbf{m}}$ so that every row and column is nonincreasing.

We can recover ρ from σ by defining the ith column of ρ to be the shape of the shifted plane partition consisting of all entries of σ which are $\geqslant i$. Hence we need to show that the map $(\pi,\mathbf{a}) \longrightarrow \sigma$ just defined between (a) ordered pairs $(\pi,\mathbf{a})$ where π is a SSYT of shape $\mathbf{m}$ and $\mathbf{a}$ is a sequence $a_1 \geqslant \cdots \geqslant a_m \geqslant 0$ of integers such that $a_i > a_{i+1}$ if $i \in K_\pi$, and (b) shifted plane partitions σ of shape $\mathbf{m}$, is a one-to-one correspondence. This will follow from a general result about partially ordered sets which we now describe.

Let P be any finite partially ordered set (poset) with m elements, and let $\omega: P \to \{1,2,\ldots,m\}$ be a fixed order-preserving bijection (so $x \leqslant y$ in P implies $\omega(x) \leqslant \omega(y)$). Let $\mathscr{L}(P)$ be the set of *all* order-preserving bijections $\pi: P \to \{1,2,\ldots,m\}$. If $\pi \in \mathscr{L}(P)$, let S_π denote the set of all integer sequences $a_1 \geqslant \cdots \geqslant a_m \geqslant 0$ such that $a_i > a_{i+1}$ if $\omega\pi^{-1}(i) > \omega\pi^{-1}(i+1)$. Finally, let $\mathscr{A}(P)$ consist of all order-reversing maps $\sigma: P \to \{0,1,2,\ldots\}$ [i.e., $x \leqslant y$ in P implies $\sigma(x) \geqslant \sigma(y)$]. According to Ref. 4 or Theorem 6.2 of Ref. 5, we have:

Lemma: Define a map $\Phi(\pi,\mathbf{a}) = \sigma$ between ordered pairs $(\pi,\mathbf{a})$ where $\pi \in \mathscr{L}(P)$ and $\mathbf{a} \in S_\pi$, and the set $\mathscr{A}(P)$, by the rule $\sigma(x) = a_{\pi^{-1}(x)}$. Then Φ is a one-to-one correspondence.

We may regard the shifted Young diagram $Z_{\mathbf{m}}$ as a poset, with the elements (squares) increasing as we read left-to-right or top-to-bottom. Choose $\omega: Z_{\mathbf{m}} \to \{1,2,\ldots m\}$ to increase by unit amounts along each row. E.g., for $\mathbf{m} = (5,3,1)$, ω is given by

1 2 3 4 5
6 7 8
9.

It is clear that a map $\sigma \in \mathscr{A}(Z_{\mathbf{m}})$ is nothing more than a shifted plane partition of shape $\mathbf{m}$, and that an order-preserving bijection $\pi \in \mathscr{L}(Z_{\mathbf{m}})$ is just an SSYT. It follows from the lemma and our choice of ω that we have exactly the one-to-one correspondence $(\pi,\mathbf{a}) \longrightarrow \sigma$ needed to complete the proof of (i).

(ii) This follows immediately from (i) and the well-known description of the irreducible representations of SU(n) in terms of Young tableaux.

(iii) The number $g^{\mathbf{m}}$ of SSYT of shape $\mathbf{m}$ has been calculated implicitly by Schur,[7] and more explicitly in Refs. 3 and 8.

4. EXAMPLES

We will use the Fundamental Theorem to compute the character generators of SU(3) and SU(4). These two cases are at least implicit in Ref. 6.

For the case of SU(3), there are two SSYT π of shape (3,2). For each of these π, we need to compute (by inspection) the shape $(l_1,\ldots,l_j)$ of each of the five sub-SSYT $\pi^{(1)},\ldots,\pi^{(5)}$ and hence obtain the monomial $\Gamma(\pi^{(i)}) = A_j X_{l_1}\cdots X_{l_j}$. We also compute by inspection the set K_π of i in π such that $i+1$ appears in a higher row than i. Then π will contribute a term $\Pi_{i \in K_\pi}\Gamma(\pi^{(i)})/\Pi_{i=1}^{m}[1 - \Gamma(\pi^{(i)})]$ to $F_{\mathbf{m}}(A,X)$. Substituting $X_1 = \alpha_1$, $X_2 = \alpha_1^{-1}\alpha_2$, $X_3 = \alpha_2^{-1}$ yields the character generator $X_A(\alpha)$. The table below gives the relevant information for each SSYT π.

1. $\pi = \begin{matrix} 1 & 2 & 3 \\ 4 & 5 \end{matrix}$ $K_\pi = \phi$

i	1	2	3	4	5
$\pi^{(i)}$	1	1 2	1 2 3	1 2 3 4	1 2 3 4 5
$\Gamma(\pi^{(i)})$	A_1X_1	A_1X_2	A_1X_3	$A_2X_1X_3$	$A_2X_2X_3$

2. $\pi = \begin{matrix} 1 & 2 & 4 \\ 3 & 5 \end{matrix}$ $K_\pi = \{3\}$

i	1	2	3	4	5
$\pi^{(i)}$	1	1 2	1 2 3	1 2 4 3	1 2 4 3 5
$\Gamma(\pi^{(i)})$	A_1X_1	A_1X_2	$A_2X_1X_2$	$A_2X_1X_3$	$A_2X_2X_3$

Hence

$$F_{(3,2)}(A,X) = \frac{1}{(1-A_1X_1)(1-A_1X_2)(1-A_1X_3)(1-A_2X_1X_3)(1-A_2X_2X_3)} + \frac{A_2X_1X_2}{(1-A_1X_1)(1-A_1X_2)(1-A_2X_1X_2)(1-A_2X_1X_3)(1-A_2X_2X_3)}.$$

Thus the character generator for SU(3) is given by:

$$X_A(\alpha) = \frac{1}{(1-\alpha_1A_1)(1-\alpha_1{}^{-1}\alpha_2A_1)(1-\alpha_2{}^{-1}A_1)(1-\alpha_1\alpha_2{}^{-1}A_2)(1-\alpha_1{}^{-1}A_2)} + \frac{\alpha_2A_2}{(1-\alpha_1A_1)(1-\alpha_1{}^{-1}\alpha_2A_1)(1-\alpha_2A_2)(1-\alpha_1\alpha_2{}^{-1}A_2)(1-\alpha_1{}^{-1}A_2)}.$$

For the case of SU(4), there are 12 SSYT of shape (4, 3, 2). For each one we list the set K_π and the shapes $(l_1,\ldots,l_j)$ of each $\pi^{(i)}$, so $\Gamma(\pi^{(i)}) = A_jX_{l_1}\cdots X_{l_j}$.

(1) $\pi = \begin{matrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 \\ 8 & 9 \end{matrix}$ $K_\pi = \phi$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots l_j$	1	2	3	4	4,1	4,2	4,3	4,3,1	4,3,2

(2) $\pi = \begin{matrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 8 \\ 7 & 9 \end{matrix}$ $K_\pi = \{7\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	3	4	4,1	4,2	4,2,1	4,3,1	4,3,2

(3) $\pi = \begin{matrix} 1 & 2 & 3 & 5 \\ 4 & 6 & 7 \\ 8 & 9 \end{matrix}$ $K_\pi = \{4\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	3	3,1	4,1	4,2	4,3	4,3,1	4,3,2

(4) $\pi = \begin{matrix} 1 & 2 & 3 & 5 \\ 4 & 6 & 8 & \\ 7 & 9 & & \end{matrix}$ $K_\pi = \{4,7\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	3	3,1	4,1	4,2	4,2,1	4,3,1	4,3,2

(5) $\pi = \begin{matrix} 1 & 2 & 3 & 6 \\ 4 & 5 & 7 & \\ 8 & 9 & & \end{matrix}$ $K_\pi = \{5\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	3	3,1	3,2	4,2	4,3	4,3,1	4,3,2

(6) $\pi = \begin{matrix} 1 & 2 & 3 & 6 \\ 4 & 5 & 8 & \\ 7 & 9 & & \end{matrix}$ $K_\pi = \{5,7\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	3	3,1	3,2	4,2	4,2,1	4,3,1	4,3,2

(7) $\pi = \begin{matrix} 1 & 2 & 3 & 7 \\ 4 & 5 & 8 & \\ 6 & 9 & & \end{matrix}$ $K_\pi = \{6\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	3	3,1	3,2	3,2,1	4,2,1	4,3,1	4,3,2

(8) $\pi = \begin{matrix} 1 & 2 & 4 & 5 \\ 3 & 6 & 7 & \\ 8 & 9 & & \end{matrix}$ $K_\pi = \{3\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	2,1	3,1	4,1	4,2	4,3	4,3,1	4,3,2

(9) $\pi = \begin{matrix} 1 & 2 & 4 & 5 \\ 3 & 6 & 8 & \\ 7 & 9 & & \end{matrix}$ $K_\pi = \{3,7\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	2,1	3,1	4,1	4,2	4,2,1	4,3,1	4,3,2

(10) $\pi = \begin{matrix} 1 & 2 & 4 & 6 \\ 3 & 5 & 7 & \\ & 8 & 9 & \end{matrix}$ $K_\pi = \{3,5\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	2,1	3,1	3,2	4,2	4,3	4,3,1	4,3,2

(11) $\pi = \begin{matrix} 1 & 2 & 4 & 6 \\ 3 & 5 & 8 & \\ & 7 & 9 & \end{matrix}$ $K_\pi = \{3,5,7\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	2,1	3,1	3,2	4,2	4,2,1	4,3,1	4,3,2

(12) $\pi = \begin{matrix} 1 & 2 & 4 & 7 \\ 3 & 5 & 8 & \\ & 6 & 9 & \end{matrix}$ $K_\pi = \{3,6\}$

i	1	2	3	4	5	6	7	8	9
$l_1,\ldots,l_j$	1	2	2,1	3,1	3,2	3,2,1	4,2,1	4,3,1	4,3,2

Thus we obtain the following expression for the character generator $X_A(\alpha)$ of SU(4):

$$
\begin{aligned}
&(1-\alpha_1A_1)(1-\alpha_1^{-1}\alpha_2A_1)(1-\alpha_1\alpha_2^{-1}A_3)(1-\alpha_1^{-1}A_3)X_A(\alpha)\\
&= \frac{1}{(1-\alpha_2^{-1}\alpha_3A_1)(1-\alpha_3^{-1}A_1)(1-\alpha_1\alpha_3^{-1}A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2^{-1}A_2)}\\
&\quad+\frac{\alpha_2\alpha_3^{-1}A_3}{(1-\alpha_2^{-1}\alpha_3A_1)(1-\alpha_3^{-1}A_1)(1-\alpha_1\alpha_3^{-1}A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2\alpha_3^{-1}A_3)}\\
&\quad+\frac{\alpha_1\alpha_2^{-1}\alpha_3A_2}{(1-\alpha_2^{-1}\alpha_3A_1)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1\alpha_3^{-1}A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2^{-1}A_2)}\\
&\quad+\frac{\alpha_1A_2A_3}{(1-\alpha_2^{-1}\alpha_3A_1)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1\alpha_3^{-1}A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2\alpha_3^{-1}A_3)}\\
&\quad+\frac{\alpha_1^{-1}\alpha_3A_2}{(1-\alpha_2^{-1}\alpha_3A_1)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2^{-1}A_2)}\\
&\quad+\frac{\alpha_1^{-1}\alpha_2A_2A_3}{(1-\alpha_2^{-1}\alpha_3A_1)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2\alpha_3^{-1}A_3)}\\
&\quad+\frac{\alpha_3A_3}{(1-\alpha_2^{-1}\alpha_3A_1)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_3A_2)(1-\alpha_3A_3)(1-\alpha_2\alpha_3^{-1}A_3)}\\
&\quad+\frac{\alpha_2A_2}{(1-\alpha_2A_2)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1\alpha_3^{-1}A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2^{-1}A_2)}\\
&\quad+\frac{\alpha_2^2\alpha_3^{-1}A_2A_3}{(1-\alpha_2A_2)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1\alpha_3^{-1}A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2\alpha_3^{-1}A_3)}\\
&\quad+\frac{\alpha_1^{-1}\alpha_2\alpha_3A_2^2}{(1-\alpha_2A_2)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2^{-1}A_2)}
\end{aligned}
$$

$$+\frac{\alpha_1^{-1}\alpha_2^2 A_2^2 A_3}{(1-\alpha_2A_2)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_2\alpha_3^{-1}A_2)(1-\alpha_2\alpha_3^{-1}A_3)}$$

$$+\frac{\alpha_2\alpha_3A_2A_3}{(1-\alpha_2A_2)(1-\alpha_1\alpha_2^{-1}\alpha_3A_2)(1-\alpha_1^{-1}\alpha_3A_2)(1-\alpha_3A_3)(1-\alpha_2\alpha_3^{-1}A_3)}.$$

There seems little point in writing down the character generator of SU(5), which by part (iii) of the theorem has 286 terms. Even more impractical is the character generator of SU(6), with 33592 terms.

5. CONCLUSIONS

The generating function $F_m(A,X)$ has some additional properties of interest. If $m=(n,n-1,\ldots,2)$ then write $F_m(A,X)=F_n(A,X)$. If we set each $A_1=1$ in $F_n(A,X)$, then it follows, e.g., from Eq. (11.9;6) of Ref. 9 or Corollary 8.3 of Ref. 2 that

$$F_n(1,1,\ldots,1;X)=(1-X_1X_2\cdots X_n)\Big/\prod_{i=1}^{n}(1-X_i)\prod_{1\le i<j\le n}(1-X_iX_j).$$

If we set each $X_i=1$ and $A_i=A$ in $F_m(A,X)$, then it follows from (5) that the coefficient of A^q in $F_m(A,\ldots,A,1,\ldots,1)$ is a polynomial function $P_m(q)$ of q of degree $m-1$ and leading coefficient $g^m/(m-1)!$ When $m=(n,n-1,\ldots,n-k+1)$, this polynomial $P_{n,k}(q)$ is given by

$$P_{n,k}(q)=\sum\dim(a_1\lambda_1+\cdots+a_k\lambda_k), \tag{7}$$

where the sum is over all k-tuples of nonnegative integers $(a_1,\ldots,a_k)$ such that $a_1+\cdots+a_k=q$, and where $\dim\lambda$ denotes the dimension of the irreducible representation of SU(n) with highest weight λ. When $k=n-1$, the sum (7) can be explicitly evaluated using a result of Andrews[10] and independently Macdonald[11] (pp. 50–52). Namely,

$$P_{n,n-1}(q)=\begin{cases}\Delta^2\displaystyle\prod_{i=0}^{l}\frac{(q+n+2i-2)_{4i+1}}{(n+2i)_{4i+1}}, & \text{if } n=2l+1\\[2ex] \Delta^2\displaystyle\prod_{i=1}^{l}\frac{(q+n+2i-3)_{4i-1}}{(n+2i-1)_{4i-1}}, & \text{if } n=2l,\end{cases}$$

where $(r)_s=r(r-1)(r-2)\cdots(r-s+1)$, and where Δ^2 is the second-difference operator, defined by $\Delta^2Q(q)=Q(q+2)-2Q(q+1)+Q(q)$. Alternatively, we have $P_{n,n-1}(q)=\Delta^2\dim((q-2)\lambda_n)$, where λ_n is the highest weight of the spin representation of the Lie algebra so($2n+1,\mathbb{C}$). A theoretical explanation of this fact can be given by considering the decomposition of $gl(n,\mathbb{C})\subset$so($2n+1,\mathbb{C}$) in the representation $(q-2)\lambda_n$. We will not enter into the details here.

We have described a method for writing $F_m(A,X)$ as a sum of g^m terms of the form (3). One may wonder whether there is some alternative way to write $F_m(A,X)$ as a sum of fewer terms of the form (3). If we have any such representation of $F_m(A,X)$ then setting $A_i=A$ and $X_i=1$ as above, we obtain

$$F_m(A,\ldots,A,1,\ldots,1)=\sum_j\frac{A^{t_j}}{(1-A)^m}=\frac{\sum_jA^{t_j}}{(1-A)^m},$$

for certain integers $t_j\geqslant 0$. Hence the integers t_j are uniquely determined by $F_m(A,X)$, not by the way in which $F_m(A,X)$ is written as a sum of terms (3). In particular, the number of terms is always the same, namely, g^m.

Let us mentions that the numbers g^m were shown by Schur[5] to be the degrees of the irreducible projective representations of the symmetric group S_m. We don't know if this connection between SU(n) and S_m is just a coincidence.

It is natural to ask whether our results for SU(n) can be extended to other simple Lie groups, in particular O(n) and Sp($2n$). We have been unable to write the character generator for these groups in the form (1) because of the lack of a combinatorial description of the characters which would allow the use of the lemma on posets. Though there exist combinatorial descriptions of the characters of these groups (e.g., Ref. 9, p. 240, and Ref. 12), they seem unsuitable for the implementaion of the Lemma.

ACKNOWLEDGMENT

I am grateful to J. Patera and R. T. Sharp for encouraging me to write this paper, and to the referee for his helpful comments regarding exposition.

[1] R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978); J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978); P. E. Desmier and R. T. Sharp, J. Math. Phys. **20**, 74 (1979); J. Patera and R. T. Sharp, "Generating Function Techniques Pertinent to Spectroscopy and Crystal Physics," in *Recent Advances in Group Theory and their Application to Spectroscopy* (Plenum, New York, 1979), pp. 219-248; J. Patera and R. T. Sharp, "Generating Functions for Characters of Group Representations and Their Applications," in *Lectures Notes in Physics*, Vol. 94, (Springer, New York, 1979), pp. 175–183; J. Patera and R. T. Sharp, J. Phys. A: Math. Gen. **13**, 397 (1980).

[2] R. P. Stanley, Stud. Appl. Math. **50**, 167 (1971) and **50**, 259 (1971).

[3] E. Gansner, "Matrix Correspondences and the Enumeration of Plane Partitions," Ph.D. thesis, M.I.T., 1978.

[4] D. E. Knuth, Math. Comp. **24**, 955 (1970).

[5] R. P. Stanley, Mem. Amer. Math. Soc., No. 119 (1972).

[6] R. T. Sharp and D. Lee, Rev. Mex. Fisica **20**, 203 (1971).

[7] I. Schur, J. Reine Angew. Math. **139**, 155 (1911).

[8] R. M. Thrall, Michigan Math. J. **1**, 81 (1952); D. E. Rutherford, Proc. Royal Soc. Edinburgh A **67**, 215 (1966); D. E. Knuth, *The Art of Computer Programming*, Vol. 3 (Addison-Wesley, Reading, Massachusetts, 1973), pp. 71 and 592; B. E. Sagan, "Partially Ordered Sets with Hook-lengths--An Algorithmic Approach," Ph.D. Thesis, M.I.T., 1979.

[9] D. E. Littlewood, *The Theory of Group Characters*, Second Edition (Oxford, 1950).

[10] G. E. Andrews, Proc. Natl. Acad. Sci. USA **74**, 426 (1977).

[11] I. G. Macdonald, *Symmetric Functions and Hall Polynomials* (Oxford, 1979).

[12] C. DeConcini, Adv. Math. **34**, 1 (1979).

Classification of three-particle states according to an orthonormal SU(3)⊃SO(3) basis [a]

F. del Aguila [b]
Departament de Física Teòrica, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

(Received 22 January 1980; accepted for publication 16 May 1980)

In this paper we generalize Dragt's approach to classifying three-particle states. Using his formalism of creation and annihilation operators, we obtain explicitly a complete set of orthonormal functions $Y^{\lambda\mu RL}{}_{M}$ on S_5. This set of functions carries all the irreducible representations of the group SU(3) reduced according to SO(3). The $Y^{\lambda\mu RL}{}_{M}$, which are eigenvectors of the togetherness and angular momentum operators, have very simple properties under three-particle permutations. We obtain also explicitly the coefficients "3ν" which reduce the products of these functions.

INTRODUCTION

In the study of three-body decays,[1,2] and in any problem with three equal particles, it is useful to employ a complete set of orthonormal states which treat all the particles on the same footing and have definite angular momentum and localization. In this paper we obtain explicitly a set of states with these properties. Following two papers of Dragt[3] we prove in Sec. 1 that the three-particle states can be classified according to the group SU(3), and this classification leads to states with the announced properties. In Sec. 2 we generalize Dragt's formalism of creation and annihilation operators and obtain the general expression for such states. Their symmetry properties are given in Sec. 3. In Sec. 4 we reduce their products by means of "3ν" coefficients, whose general expression is also given explicitly.

1. CLASSIFICATION OF THREE-PARTICLE STATES

This section and a part of the following are already presented in Dragt's papers,[3] but we review them in order to make the paper self-contained.

We shall assume three spinless, nonrelativistic and noninteracting particles of mass m. If we fix the total energy and momentum, T and $\mathbf{Q}$, to classify three-particle states means in the momentum representation to obtain a complete set of functions on the phase space.

A. Phase space

The phase space of a such system is S_5, the five-dimensional surface of a sphere in a six-dimensional space. If $\mathbf{p}_A$, $\mathbf{p}_B$, $\mathbf{p}_C$ are the particle 3-momenta, energy and momentum conservation imply

$$\sum_{x=A,B,C} \mathbf{p}_x^2 = 2mT, \tag{1a}$$

$$\sum_{x=A,B,C} \mathbf{p}_x = \mathbf{Q}. \tag{1b}$$

These equations can also be written

$$\mathbf{q}^2 + \mathbf{q}'^2 = 2mT - \mathbf{Q}^2/3, \tag{2a}$$

$$\sqrt{3}\mathbf{q}'' = \mathbf{Q}, \tag{2b}$$

where

$$\mathbf{q} = (\mathbf{p}_B - \mathbf{p}_A)/\sqrt{2},$$
$$\mathbf{q}' = (2\mathbf{p}_C - \mathbf{p}_A - \mathbf{p}_B)/\sqrt{6}, \tag{3}$$
$$\mathbf{q}'' = (\mathbf{p}_A + \mathbf{p}_B + \mathbf{p}_C)/\sqrt{3}.$$

Thus, the phase space is the set of configurations $(\mathbf{q},\mathbf{q}')$ which satisfy Eq. (2a). It must be noted that if we define

$$|\mathbf{z}\rangle = (\mathbf{q}' + i\mathbf{q})/(2mT - \mathbf{Q}^2/3)^{1/2}, \tag{4}$$

the phase space is the set of vectors of a three dimensional Hilbert space H which satisfy

$$\langle \mathbf{z}|\mathbf{z}\rangle = 1 \tag{5}$$

(i.e., S_5).

B. Classification of three-particle states according to SU(3)

On the real space $(\mathbf{q},\mathbf{q}')$ let us consider the set of linear transformations

$$R = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}, \tag{6a}$$

where α, β are 3×3 matrices verifying

$$\alpha\alpha^T + \beta\beta^T = 1,$$
$$\alpha\beta^T - \beta\alpha^T = 0. \tag{6b}$$

On H the corresponding set of transformations,

$$U = \alpha + i\beta,$$

constitutes the unitary group U(3). Its subgroup SU(3) acts on the three-dimensional Hilbert space and leaves $\langle \mathbf{z}|\mathbf{z}\rangle$ fixed. Hence, S_5 is topologically equivalent to the cosets of SU(3) with respect to SU(2), the little group of an arbitrary but fixed point of S_5. For instance, if

$$|\mathbf{a}\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

its little group is $\begin{bmatrix} 1 & 0 \\ 0 & \mathrm{SU}(2) \end{bmatrix}$ and $U|a\rangle = |z\rangle$ with $z_i = U_{i1}$ is

[a] Work partially supported by the Instituto de Estudios Nucleares (Madrid).

[b] Present address: Department of Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, England.

 0022-2488/80/092327-08$1.00

any unit vector (i.e., matrices U of the same coset have identical first column).

Let E be the Hilbert space of square-integrable functions, f, on S_5. SU(3) acts on E according to the rule

$$D_U(f) = f \circ U^{-1}, \quad \forall U \in \mathrm{SU}(3), \quad f \in E. \tag{7}$$

Thus, E is a representation space of SU(3). Using the notion of induced representation,[4] we can see this representation of SU(3) on E as induced by the trivial one of a subgroup SU(2). Therefore, the Frobenius reciprocity theorem[5] can be applied, assuring in this case that the representation of SU(3) on E carries any irreducible representation Γ of SU(3) the same number of times that Γ when restricted to SU(2) contains its trivial representation. As each irreducible representation of SU(3) contains once and only once the trivial representation of SU(2),[6] any irreducible representation of SU(3) occurs once and only once in the representation of SU(3) on E. This implies that it can be found a basis of E labelled only by SU(3) labels. In fact, we should prove the completeness of the functions classified according to SU(3). This follows from the Peter–Weyl theorem.[7] For SU(3) it assures that the set of matrix elements of all its irreducible representations is complete in the Hilbert space of square-integrable functions on the manifold of the group. Then, the subset of matrix elements obtained when one of the states fixing the matrix element runs over the corresponding singlets of SU(2) is complete in E.

C. Advantages of such classification

We have proved that three-particle states can be classified according to SU(3), i.e., two Casimir operators of SU(3) which fix the irreducible representation and three operators of the algebra or enveloping algebra which specify the corresponding basis vectors constitute a complete set of observables. This classification has three important advantages, it is "democratic" and the "togetherness" operator[8] and the total angular momentum can be used as elements of a complete set of commuting observables.

This classification is "democratic" because none of the observables needed to constitute a complete set of commuting observables discriminates among the three particles, i.e., all of them commute with the elements of the alternate subgroup A (3) of the permutation group of the three particles S(3). Therefore, the corresponding basis vectors will transform under the three-particle permutations in a very simple way. It is trivial to verify that

$$[C_{\mathrm{ABC}}, R] = 0, \tag{8a}$$

where C_{ABC} is the cyclic permutation, generating A (3),

$$C_{\mathrm{ABC}} = -\tfrac{1}{2}\begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix} \tag{8b}$$

and R is a linear transformation on $(\mathbf{q},\mathbf{q}')$ of the type (6a). As the elements of SU(3) are of this type, the observables obtained from them commute with C_{ABC}, and therefore, they do not discriminate among the three particles.

The "togetherness" operator Λ^2 measures the localization of the three particles, in a similar way as the angular momentum does it for a two-particle system. The main difference is that for interacting systems the togetherness operator does not corresponds to a conserved quantity. But, its use to classify three particle states is interesting because its eigenvalues can be bounded in some cases.[2] In the next section we shall see that the togetherness operator can be used as a Casimir operator of SU(3). We shall also prove that the group of rotations about a center-of-momentum frame is a subgroup SO(3) of SU(3). Thus, the total angular momentum J^2 and its third component J_3 can be two of the three operators needed to specify the basis vectors of the irreducible representations of SU(3).

2. EXPLICIT EXPRESSION OF THE FUNCTIONS $Y^{\lambda\mu RL}{}_M$

In this section we shall solve explicitly the problem just described, to obtain the eigenvectors of Λ^2, J^2, J_3, another Casimir operator of SU(3), and a fifth operator to provide the missing label in the reduction scheme SU(3) $\supset$ SO(3) $\supset$ SO(2). This problem has been widely studied and solved.[9-12] Our approach generalizes that of Dragt.[3] His method is based on algebraic manipulations of creations and annihilation operators. A similar approach is used in Ref. 9, but in another representation of SU(3) and with different technical details. A general study of the SU(3) $\supset$ SO(3) reduction scheme can be found in Ref. 10. Different complete sets of commuting observables can be considered[10,11]; we will pick the easier one. Finally, in Ref. 12 an equivalent result is obtained using a different approach.

A. Complete set of commuting observables

Let the complete set of generators of U(3) in its representation on $(\mathbf{q},\mathbf{q}')$ [cf. Eq. (6)] be

$$\tilde{J}_{12} = \begin{bmatrix} \cdot & 1 & \cdot & & & \\ -1 & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & & & \\ & & & \cdot & 1 & \cdot \\ & & & -1 & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \end{bmatrix}, \quad \tilde{J}_{23} = \begin{bmatrix} \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & 1 & & & \\ \cdot & -1 & \cdot & & & \\ & & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & 1 \\ & & & \cdot & -1 & \cdot \end{bmatrix},$$

$$\tilde{J}_{31}=\begin{bmatrix} \cdot & \cdot & -1 & & & \\ \cdot & \cdot & \cdot & & & \\ 1 & \cdot & \cdot & & & \\ & & & \cdot & \cdot & -1 \\ & & & \cdot & \cdot & \cdot \\ & & & 1 & \cdot & \cdot \end{bmatrix},\quad \tilde{K}_{11}=\tfrac{1}{3}\begin{bmatrix} & & & 4 & \cdot & \cdot \\ & & & \cdot & -2 & \cdot \\ & & & \cdot & \cdot & -2 \\ -4 & \cdot & \cdot & & & \\ \cdot & 2 & \cdot & & & \\ \cdot & \cdot & 2 & & & \end{bmatrix},$$

$$\tilde{K}_{12}=\begin{bmatrix} & & & \cdot & 1 & \cdot \\ & & & 1 & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ & \cdot & -1 & \cdot & & \\ -1 & & \cdot & \cdot & & \\ \cdot & & \cdot & \cdot & & \end{bmatrix},\quad \tilde{K}_{13}=\begin{bmatrix} & & & \cdot & \cdot & 1 \\ & & & \cdot & \cdot & \cdot \\ & & & 1 & \cdot & \cdot \\ & \cdot & -1 & & & \\ \cdot & \cdot & \cdot & & & \\ -1 & \cdot & \cdot & & & \end{bmatrix}, \tag{9a}$$

$$\tilde{K}_{22}=\tfrac{1}{3}\begin{bmatrix} & & & -2 & \cdot & \cdot \\ & & & \cdot & 4 & \cdot \\ & & & \cdot & \cdot & -2 \\ 2 & \cdot & \cdot & & & \\ \cdot & -4 & \cdot & & & \\ \cdot & \cdot & 2 & & & \end{bmatrix},\quad \tilde{K}_{23}=\begin{bmatrix} & & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & 1 \\ & & & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & -1 & & & \\ \cdot & -1 & \cdot & & & \end{bmatrix},$$

$$\tilde{K}_{33}=\tfrac{1}{3}\begin{bmatrix} & & & -2 & \cdot & \cdot \\ & & & \cdot & -2 & \cdot \\ & & & \cdot & \cdot & 4 \\ 2 & \cdot & \cdot & & & \\ \cdot & 2 & \cdot & & & \\ \cdot & \cdot & -4 & & & \end{bmatrix},\quad \tilde{S}=\begin{bmatrix} & & & 1 & \cdot & \cdot \\ & & & \cdot & 1 & \cdot \\ & & & \cdot & \cdot & 1 \\ -1 & \cdot & \cdot & & & \\ \cdot & -1 & \cdot & & & \\ \cdot & \cdot & -1 & & & \end{bmatrix},$$

where

$$\tilde{K}_{11}+\tilde{K}_{22}+\tilde{K}_{33}=0. \tag{9b}$$

In its representation on E these generators are written

$$\begin{aligned} J_{ij} &= \sum_{\alpha,\beta=1}^{6}(\tilde{J}_{ij})_{\alpha\beta}p_\alpha r_\beta\,, \\ K_{ij} &= \sum_{\alpha,\beta=1}^{6}(\tilde{K}_{ij})_{\alpha\beta}p_\alpha r_\beta\,, \\ S &= \sum_{\alpha,\beta=1}^{6}(\tilde{S})_{\alpha\beta}p_\alpha r_\beta\,, \end{aligned} \tag{10a}$$

where

$$(p_\alpha)=(\mathbf{q},\mathbf{q}'),\quad (r_\beta)=\left(i\frac{\partial}{\partial p_\beta}\right), \tag{10b}$$

and their nonvanishing commutators are

$$\begin{aligned} &[J_k,J_m]=i\epsilon_{kml}J_l\,, \\ &[J_k,K_{mn}]=i(\epsilon_{kml}K_{ln}+\epsilon_{knl}K_{ml}), \\ &[K_{mn},K_{rs}]=i(\delta_{mr}\epsilon_{nsl}+\delta_{ms}\epsilon_{nrl}+\delta_{nr}\epsilon_{msl} \\ &\qquad +\delta_{ns}\epsilon_{mrl})J_l, \end{aligned} \tag{11}$$

where $J_k=\frac{1}{2}\epsilon_{klm}J_{lm}$. The quantities J_k, K_{ij} generate the subgroup SU(3), while J_i generate the subgroup SO(3) of rotations about a center-of-momentum frame, $\mathbf{Q}=0$

$$\mathbf{J}=\sum_{x=A,B,C}\mathbf{r}_x\wedge\mathbf{p}_x\,, \tag{12}$$

where $\mathbf{r}_x=i(\partial/\partial\mathbf{p}_x)$. The three usual Casimir operators of U(3) are

S,

$$G^{(2)}=\sum_{i=1}^{3}J_i^2+\tfrac{1}{2}\sum_{i,j=1}^{3}K_{ij}^2\,, \tag{13}$$

$$G^{(3)}=\tfrac{1}{3}\sum_{i,j,l=1}^{3}K_{ij}K_{jl}K_{li}-\sum_{i,l=1}^{3}J_iK_{il}J_l,$$

where $G^{(2)}$, $G^{(3)}$ are the corresponding Casimir operators of SU(3). It should be noted that in this representation S, $G^{(2)}$, $G^{(3)}$ are not independent

$$G^{(3)}=\tfrac{2}{3}S(4+G^{(2)}-\tfrac{4}{9}S^2), \tag{14}$$

as we could guess from the fact that each irreducible representation of SU(3) occurs once and only once in its decomposition. Equation (14) implies that we can use S and $G^{(2)}$ as Casimir operators of SU(3); thus, we change an eigenvalue problem for a cubic operator into one for a linear operator. On the other hand, the togetherness operator

$$\begin{aligned} \Lambda^2 &= \tfrac{1}{2}\sum_{\alpha,\beta=1}^{6}(p_\alpha r_\beta-p_\alpha r_\beta)^2 \\ &= \sum_{i=1}^{3}J_i^2+\tfrac{1}{2}\sum_{i,j=1}^{3}K_{ij}^2-\tfrac{1}{3}S^2 \end{aligned} \tag{15}$$

can be written [cf. Eqs. (13), (15)]

$$\Lambda^2=G^{(2)}-\tfrac{1}{3}S^2. \tag{16}$$

Therefore, to fix the irreducible representations of SU(3) we

can use the eigenvalues of the operators S and Λ^2.

The algebra of SU(3) contains the subalgebra of SU(2) in two different ways, namely, generating a subgroup SU(2) or a subgroup SO(3).[13] The main difference is that in the first case an element can be found among the remaining ones of the algebra of SU(3) which commutes with the subalgebra of SU(2) [the usual scheme SU(2)×U(1)], while in the second case this is not possible. We are interested in the second case because the total orbital angular momentum $J^2 = \Sigma_{i=1}^3 J_i^2$ and its third component J_3 correspond to a subgroup SO(3). Thus, we must go to the enveloping algebra of SU(3) to find an operator whose eigenvalues provide the missing label.[11] The surprising thing is that none of the operators solving the missing problem has integer eigenvalues.[14] (This implies that in general we will need numerical calculation.) We will choose the easier of such operators, the one of lower order, i.e.

$$X = \sum_{i,l=1}^{3} J_i K_{il} J_l \,. \tag{17}$$

B. The trick of the harmonic oscillator

The problem is the construction of a basis $Y^{\lambda\mu RL}{}_M$ in E which diagonalizes the observables $\Lambda^2, S, J^2, J_3, X$

$$\begin{aligned}
&\Lambda^2 Y^{\lambda\mu RL}{}_M(\omega) = \lambda(\lambda+4) Y^{\lambda\mu RL}{}_M(\omega),\\
&SY^{\lambda\mu RL}{}_M(\omega) = -\mu Y^{\lambda\mu RL}{}_M(\omega),\\
&J^2 Y^{\lambda\mu RL}{}_M(\omega) = L(L+1) Y^{\lambda\mu RL}{}_M(\omega),\\
&J_3 Y^{\lambda\mu RL}{}_M(\omega) = M Y^{\lambda\mu RL}{}_M(\omega),\\
&XY^{\lambda\mu RL}{}_M(\omega) = R Y^{\lambda\mu RL}{}_M(\omega),
\end{aligned} \tag{18}$$

where λ, μ, R, L, M are the quantum numbers associated with $\Lambda^2, S, X, J^2, J_3$ and ω is a point of S_5. To this end, we generalize the algebraical approach introduced by Dragt.[3] The momentum and position operators p_α, r_β correspond to canonically conjugate variables

$$[p_\alpha, r_\beta] = -i\delta_{\alpha\beta}, \quad \alpha,\beta = 1,\ldots,6. \tag{19}$$

The U(3) operators have the same expression in the momentum representation ($r_\beta = i\partial/\partial p_\beta$) as in the position representation ($p_\alpha = -i\partial/\partial r_\alpha$) [cf. Eqs. (9), (10)]. Therefore, the eigenvectors of $\Lambda^2, S, J^2, J_3, X$ have the same expression in both representations. To obtain the explicit expression of $Y^{\lambda\mu RL}{}_M$, we shall use the position representation.

Let $r, \hat{\omega}$ be the polar coordinates in the position space,

$$r = \left(\sum_{\alpha=1}^{6} r_\alpha^2 \right)^{1/2} \tag{20}$$

and $\hat{\omega}$ the five angular variables fixing a point of $\hat{S}_5$, the surface of a sphere in this space. The Hamiltonian of a six-dimensional harmonic oscillator,

$$H = \sum_{\alpha=1}^{6} (p^2{}_\alpha + r^2{}_\alpha), \tag{21a}$$

can be written using polar coordinates

$$H = \frac{1}{2}\left(-\frac{\partial^2}{\partial r^2} - \frac{5}{r}\frac{\partial}{\partial r} + \frac{\Lambda^2}{r^2} + r^2 \right), \tag{21b}$$

where Λ^2 is the togetherness operator depending on the angular variables. The corresponding eigenvalue equation

$$H\psi(r,\hat{\omega}) = E\psi(r,\hat{\omega}) \tag{22a}$$

is equivalent to

$$\left(\frac{\partial^2}{\partial r^2} + \frac{5}{r}\frac{\partial}{\partial r} - \frac{\lambda(\lambda+4)}{r^2} - r^2 + 2E \right) F^{E\lambda}(r) = 0, \tag{22b}$$

$$\Lambda^2 Y^\lambda(\hat{\omega}) = \lambda(\lambda+4) Y^\lambda(\hat{\omega}), \tag{22c}$$

where $\psi(r,\hat{\omega}) = F^{E\lambda}(r) Y^\lambda(\hat{\omega})$. In the previous section we proved that to classify the functions $Y^\lambda(\hat{\omega})$ we can use the operators S, X, J^2, J_3. The trick of the harmonic oscillator is: i) to associate with each set of values λ, μ, R, L, M one of the values E compatible with λ; ii) to construct algebraically with creation operators acting on the vacuum or lowest energy state $|0\rangle$ the corresponding eigenfunction of the harmonic oscillator, and iii) to factorize the angular part $Y^{\lambda\mu RL}{}_M$ of the eigenfunction.

Let the creation and annihilation operator representation be

$$\begin{aligned}
\mathbf{a}^+ &= \tfrac{1}{2}[\mathbf{r}' + i\mathbf{r} - i(\mathbf{q}' + i\mathbf{q})],\\
\mathbf{b}^+ &= \tfrac{1}{2}[\mathbf{r}' - i\mathbf{r} - i(\mathbf{q}' - i\mathbf{q})],\\
\mathbf{a} &= \tfrac{1}{2}[\mathbf{r}' - i\mathbf{r} + i(\mathbf{q}' - i\mathbf{q})],\\
\mathbf{b} &= \tfrac{1}{2}[\mathbf{r}' + i\mathbf{r} + i(\mathbf{q} + i\mathbf{q})],
\end{aligned} \tag{23}$$

where $(\mathbf{r},\mathbf{r}') = (r_\alpha)$, $(\mathbf{q},\mathbf{q}') = (p_\alpha)$. Their nonvanishing commutators are

$$[a_i, a_j^+] = [b_i, b_j^+] = \delta_{ij}, \quad i,j = 1,2,3. \tag{24}$$

The vectors of the state space are polynomials in creation operators $P(\mathbf{a}^+, \mathbf{b}^+)$ acting on the vacuum. The U(3) generators are written

$$\begin{aligned}
&\mathbf{J} = i(\mathbf{a}\wedge\mathbf{a}^+ + \mathbf{b}\wedge\mathbf{b}^+),\\
&K_{ij} = b_i^+ b_j + b_j^+ b_i - (a_i^+ a_j + a_j^+ a_i)\\
&\qquad - \tfrac{2}{3}\delta_{ij}(\mathbf{b}^+\cdot\mathbf{b} - \mathbf{a}^+\cdot\mathbf{a}),
\end{aligned} \tag{25a}$$

$$S = \mathbf{b}^+\cdot\mathbf{b} - \mathbf{a}^+\cdot\mathbf{a},$$

and the Hamiltonian operator is

$$H = \mathbf{a}^+\cdot\mathbf{a} + \mathbf{b}^+\cdot\mathbf{b} + 3. \tag{25b}$$

It follows from Eq. (25a) that only four independent bilinear operators commute with U(3)

$$\begin{aligned}
N_a &= \mathbf{a}^+\cdot\mathbf{a},\\
N_b &= \mathbf{b}^+\cdot\mathbf{b},\\
\Delta_+ &= \mathbf{a}^+\cdot\mathbf{b}^+,\\
\Delta_- &= \mathbf{a}\cdot\mathbf{b},
\end{aligned} \tag{26a}$$

where N_a, N_b measure the number of exitations of type a and b, respectively, and $\Delta_\pm$ are operators of double creation and annihilation

$$[N_{a,b}, \Delta_\pm] = \pm\Delta_\pm \,. \tag{26b}$$

The Hamiltonian operator H commutes with N_a and N_b, but not with $\Delta_\pm$

$$[H, \Delta_\pm] = \pm 2\Delta_\pm \,. \tag{27}$$

Therefore, the eigenvectors of $H, \Lambda^2, S, X, J^2, J_3$ are eigenvectors of $N_{a,b}$, but not of $\Delta_\pm$

$$\Delta_{\pm}|E\lambda\mu RLM\rangle = |E\pm 2,\lambda\mu RLM\rangle, \tag{28a}$$

except when E is the lowest energy value compatible with λ, in which case

$$\Delta_{-}|E\lambda\mu RLM\rangle = 0. \tag{28b}$$

We will associate with a given value of λ the lowest energy E compatible with it.

The eigenvalue equations of Λ^2 and S can be written

$$\begin{aligned}&[(N_a+N_b)(N_a+N_b+4)-4\Delta_{+}\Delta_{-}]|E\lambda\mu RLM\rangle\\ &\quad=\lambda(\lambda+4)|E\lambda\mu RLM\rangle.\end{aligned} \tag{29}$$

$$(N_a-N_b)|E\lambda\mu RLM\rangle = \mu|E\lambda\mu RLM\rangle.$$

If $|E\lambda\mu RLM\rangle$ satisfies Eq. (28b), Eq. (29) implies that the polynomial in creation operators $P^{E\lambda\mu RL}{}_{M}(\mathbf{a}^+,\mathbf{b}^+)$ acting on the vacuum is composed of a sum of terms with n_a operators of type a and n_b of type b

$$\begin{aligned}&n_a+n_b=\lambda,\\ &n_a-n_b=\mu.\end{aligned} \tag{30}$$

It follows from Eq. (25b), that the lowest energy E compatible with λ is $\lambda+3$,

$$(N_a+N_b+3)|E\lambda\mu RLM\rangle = E|E\lambda\mu RLM\rangle. \tag{31}$$

In order to obtain the eigenstates of J^2 and J_3, it is useful to introduce a spherical basis

$$\begin{aligned}&a^{+}_{+1} = -(a^{+}_{1}+ia^{+}_{2})/\sqrt{2},\\ &a^{+}_{0} = a^{+}_{3},\\ &a^{+}_{-1} = (a^{+}_{1}-ia^{+}_{2})/\sqrt{2},\end{aligned} \tag{32a}$$

similarly for $\mathbf{b}^+$, $\mathbf{a}$, $\mathbf{b}$; whose nonvanishing commutators are

$$[a_m,a^{+}_{n}] = [b_m,b^{+}_{n}] = (-1)^m\delta_{m,n},\quad m,n=+1,0,-1. \tag{32b}$$

It should be noted that we are only interested in the state with $M=L$ because the others are obtained from it applying the operator

$$J_{-1} = J_1 - iJ_2. \tag{33}$$

We shall see that J_{-1} acts in a very simple way. In order to construct a state with total angular momentum and its third component equal to L the terms of the polynomial in creation operators must be of the form

$$\begin{aligned}&(\mathbf{a}^+\cdot\mathbf{a}^+)^{S_a}(\mathbf{b}^+\cdot\mathbf{b}^+)^{S_b}(\mathbf{a}^+\cdot\mathbf{b}^+)^{S_{ab}}\\ &\quad\times(\mathbf{a}^{+}_{+1})^{V_a}(\mathbf{b}^{+}_{+1})^{V_b}(\mathbf{a}^+\wedge\mathbf{b}^+)^{V_{ab}}_{+1},\end{aligned} \tag{34a}$$

where

$$V_a+V_b+V_{ab}=L. \tag{34b}$$

Note that $(\mathbf{a}^+\cdot\mathbf{a}^+)^{S_a}(\mathbf{b}^+\cdot\mathbf{b}^+)^{S_b}(\mathbf{a}^+\cdot\mathbf{b}^+)^{S_{ab}}$ is the most general scalar operator and $(\mathbf{a}^{+}_{+1})^{V_a}(\mathbf{b}^{+}_{+1})^{V_b}(\mathbf{a}^+\wedge\mathbf{b}^+)^{V_{ab}}$ the most general operator generating a state with angular momentum equal to its third component which can appear in a term of a polynomial. Thus, the state $|\lambda+3,\lambda\mu RLL\rangle$ has the form

$$\sum_i c_i\,|\,S^a_iS^b_iS^{ab}_iV^a_iV^b_iV^{ab}_i\rangle, \tag{35a}$$

where c_i are numerical coefficients to be determined,

$$\begin{aligned}&|\,S^a_iS^b_iS^{ab}_iV^a_iV^b_iV^{ab}_i\rangle\\ &\quad=(\mathbf{a}^+\cdot\mathbf{a}^+)^{S^a_i}(\mathbf{b}^+\cdot\mathbf{b}^+)^{S^b_i}(\mathbf{a}^+\cdot\mathbf{b}^+)^{S^{ab}_i}\\ &\qquad\times(\mathbf{a}^{+}_{+1})^{V^a_i}(\mathbf{b}^{+}_{+1})^{V^b_i}(\mathbf{a}^+\wedge\mathbf{b}^+)^{V^{ab}_i}_{+1}\,|0\rangle,\end{aligned} \tag{35b}$$

and i goes over the sets of natural numbers S^a_i, S^b_i, S^{ab}_i, V^a_i, V^b_i, V^{ab}_i satisfying

$$2(S^a_i+S^b_i+S^{ab}_i+V^{ab}_i)+V^a_i+V^b_i=\lambda, \tag{36a}$$

$$2(S^a_i-S^b_i)+V^a_i-V^b_i=\mu, \tag{36b}$$

$$V^a_i+V^b_i+V^{ab}_i=L. \tag{36c}$$

Really V^{ab}_i is fixed, it is equal to 0 or 1 depending whether $\lambda+L$ is even or odd. This is because greater powers can be reduced and the addition of Eqs. (36a) and (36c) implies that $V^{ab}_i+\lambda+L$ must be even.

To obtain the coefficients c_i, up to a global factor, we must solve the equations

$$\Delta_{-}\Big(\sum_i c_i\,|\,S^a_iS^b_iS^{ab}_iV^a_iV^b_iV_{ab}\rangle\Big)=0, \tag{37a}$$

$$(X-R)\Big(\sum_i c_i\,|\,S^a_iS^b_iS^{ab}_iV^a_iV^b_iV_{ab}\rangle\Big)=0. \tag{37b}$$

Their resolution is reduced to solve a system of linear equations if we introduce the action of Δ_{-} and X on a state $|S_aS_bS_{ab}V_aV_bV_{ab}\rangle$,

$$\begin{aligned}&\Delta_{-}|S_aS_bS_{ab}V_aV_bV_{ab}\rangle\\ &\quad=4S_aS_b|S_a-1S_b-1S_{ab}+1V_aV_bV_{ab}\rangle\\ &\qquad+2S_aV_b\,|\,S_a-1S_bS_{ab}V_a+1V_b-1V_{ab}\rangle\\ &\qquad+2S_bV_a\,|\,S_aS_b-1S_{ab}V_a-1V_b+1V_{ab}\rangle\\ &\qquad+S_{ab}[S_{ab}+V_a+V_b+2(S_a+S_b+V_{ab}+1)]\\ &\qquad\times|S_aS_bS_{ab}-1V_aV_bV_{ab}\rangle,\end{aligned} \tag{38a}$$

$$\begin{aligned}&X|S_aS_bS_{ab}V_aV_bV_{ab}\rangle\\ &\quad=4S_aV_b(2V_b+2V_{ab}-1)\\ &\qquad\times|S_a-1S_bS_{ab}+1V_a+1V_b-1V_{ab}\rangle\\ &\qquad-4S_aV_b(V_b-1)\\ &\qquad\times|S_a-1S_b+1S_{ab}V_a+2V_b-2V_{ab}\rangle\\ &\qquad-4S_bV_a(2V_a+2V_{ab}-1)\\ &\qquad\times|S_aS_b-1S_{ab}+1V_a-1V_b+1V_{ab}\rangle\\ &\qquad+[4(S_bV_a^2-S_aV_b^2)+V_a(V_a+1)-V_b(V_b+1)\\ &\qquad+V_{ab}(4(S_b(2V_a+1)-S_a(2V_b+1))\\ &\qquad+2V_a(V_a+3)-2V_b(V_b+3))\\ &\qquad+\tfrac{2}{3}(V_a+V_b+V_{ab})(V_a+V_b+V_{ab}+1)\\ &\qquad\times(2(S_a-S_b)+V_a-V_b)]\\ &\qquad\times|S_aS_bS_{ab}V_aV_bV_{ab}\rangle+4S_bV_a(V_a-1)\\ &\qquad\times|S_a+1S_b-1S_{ab}V_a-2V_b+2V_{ab}\rangle.\end{aligned} \tag{38b}$$

[These equations can be obtained using the definitions of Δ_{-} in Eq. (26a), X in Eqs. (17) and (25a), and $|S_aS_bS_{ab}V_aV_bV_{ab}\rangle$ in Eq. (35b), the commutation relations in Eq. (32b), and the fact that the annihilation operators destroy the vacuum.]

In general to solve this system of linear equations, it is necessary to resort to a computer. This is due to the fact that

in general R must be calculated numerically.[11,14] We have made a computer program to calculate the R eigenvalues and the corresponding coefficients c_i for any λ, μ, L assignment (note that the numbers c_i do not change under J_{-1}, and therefore they are independent of M). The output of this program for the lowest L, λ, μ values is given in Ref. 1, using the conventions which we shall fix in the next section. The R values are not given in this reference, but can be found in Ref. 11 with an extra factor 3/2.

The input of this program are the values of L, λ, μ and the output the values of R and c_i. It is composed essentially of four subroutines. The first one verifies that [cf. Eqs. (36a)–(36c)]

$$\begin{aligned} &\lambda \geqslant 0, \\ &\mu = \lambda, \lambda-2, \lambda-4, \ldots, -\lambda, \\ &L \leqslant \lambda, \end{aligned} \tag{39}$$

and evaluates the number of times that the representation L of SO(3) occurs in the representation λ, μ of SU(3) using Racah's formula[15]

$$\begin{aligned} N_L(\lambda,\mu) = {} & \mathrm{IP}[(\lambda+2-L)/2] - \mathrm{IP}[(\lambda+\mu+2-2L)/4] \\ & - \mathrm{IP}[(\lambda-\mu+2-2L)/4], \end{aligned} \tag{40}$$

where IP$[x]$ means the integer part of the positive part of x, i.e., IP$[x]$ vanishes for $x<0$ and IP$[x]$ is the largest integer less than or equal to x for $x \geqslant 0$. Depending on the value of $N_L(\lambda,\mu)$ we have three cases: i) $N_L(\lambda,\mu)=0$, this means that there is no eigenvector with values λ, μ, L; ii) $N_L(\lambda,\mu)=1$, in this case there is only one vector for the λ, μ, L values, and we do not need the eigenvalue R and Eq. (37b) to find it; and iii) $N_L(\lambda,\mu)>1$, then to obtain the different eigenvectors we need the different eigenvalues R and the Eq. (37b). For the nontrivial cases (ii) and (iii) the second subroutine constructs the terms which can appear in the polynomial (35a) using Eqs. (36a–c). The third one solves the Eq. (37a) for the former cases. This means that it obtains $N_L(\lambda,\mu)$ independent linear combinations verifying this equation. To this point the use of a computer program is not essential because these linear combinations, up to a normalization factor, have integer coefficients [cf. Eqs. (37a), (38a)]. They span the subspaces corresponding to the $N_L(\lambda,\mu)$ representations L of SO(3) occurring in the λ, μ representation of SU(3). The fourth subroutine, which is essential only for the case (iii), evaluates the matrix of X in these bases using Eq. (38b), then it diagonalizes this matrix, whose eigenvalues are the R eigenvalues and whose eigenvectors satisfy Eq. (37b) and give us the coefficients c_i.

C. Explicit expression of the functions $Y^{\lambda\mu RL}{}_M$

To obtain explicitly the functions $Y^{\lambda\mu RL}{}_M(\hat{\omega})$, we must write the corresponding harmonic oscillator state in the position representation and factorize the angular part. The radial part is the corresponding solution of Eq. (22b), i.e., for E equal to $\lambda+3$

$$F^{E\lambda}(r) = r^{\lambda} e^{-r^2/2}. \tag{41}$$

Therefore

$$Y^{\lambda\mu RL}{}_L(\hat{\omega}) = r^{-\lambda} e^{r^2/2} \langle r,\hat{\omega} | P^{E\lambda\mu RL}{}_L(\mathbf{a}^+ \cdot \mathbf{b}^+) | 0 \rangle \tag{42}$$

where $P^{E\lambda\mu RL}{}_L(\mathbf{a}^+ \cdot \mathbf{b}^+)$, $E=\lambda+3$, is the polynomial in creation operators obtained in the previous section. In the position representation it is a differential operator which gives a polynomial in $(\mathbf{r},\mathbf{r}')$ multiplied by $e^{-r^2/2}$ when acting on the vacuum [the corresponding solution of Eq. (22a) for $E=3$, $\langle r,\hat{\omega}|0\rangle = e^{-r^2/2}$]. This polynomial must be homogeneous of order λ according to Eq. (42). From Eq. (23) and the vacuum expression we have

$$Y^{\lambda\mu RL}{}_L(\hat{\omega}) = r^{-\lambda} P^{\lambda+3,\lambda\mu RL}{}_L(\mathbf{r}'+i\mathbf{r}, \mathbf{r}'-i\mathbf{r}), \tag{43a}$$

where $\hat{\omega}$ are the angular variables associated with $(\mathbf{r},\mathbf{r}')$. In the momentum representation we have the equation

$$Y^{\lambda\mu RL}{}_L(\omega) = p^{-\lambda} P^{\lambda+3,\lambda\mu RL}{}_L(\mathbf{q}'+i\mathbf{q}, \mathbf{q}'-i\mathbf{q}), \tag{43b}$$

where

$$p = \left(\sum_{\alpha=1}^{6} p_\alpha^2 \right)^{1/2}$$

and ω are the radial and angular variables associated with $(p_\alpha) = (\mathbf{q},\mathbf{q}')$.

We are interested[1,2] in the explicit form of $Y^{\lambda\mu RL}{}_L(\omega)$ when ω is specified by two Dalitz plot variables,[16] ρ, ϕ, and three Euler angles,[17] α, β, γ. The variables ρ, ϕ,

$$0 \leqslant \rho \leqslant 1, \quad 0 \leqslant \phi \leqslant 2\pi, \tag{44a}$$

are polar coordinates of the Dalitz plot defined by the points which are distant $\mathbf{p}_x^2/p^2, x = A,B,C$ from the sides of an equilateral triangle of unit altitude,

$$\begin{aligned} &\mathbf{p}_x^2 = p^2(1+\rho\epsilon_x)/3, \\ &\epsilon_x = \cos(\phi - 2k_x\pi/3), \quad k_x = 1,2,3, \end{aligned} \tag{44b}$$

where the three-momenta $\mathbf{p}_x$ are assumed in a center-of-momentum frame. The angles α, β, γ fix the rotation from an initial standard frame to a frame attached to the three-momenta $\mathbf{p}_x$ and defined by the orthonormal vectors [cf. Eq. (3)]

$$\begin{aligned} &\mathbf{u} = [\sin(\phi/2)\mathbf{q} + \cos(\phi/2)\mathbf{q}']/[(1+\rho)p^2/2]^{1/2}, \\ &\mathbf{v} = [-\cos(\phi/2)\mathbf{q} + \sin(\phi/2)\mathbf{q}']/[(1-\rho)p^2/2]^{1/2}, \\ &\mathbf{w} = \mathbf{u} \wedge \mathbf{v}. \end{aligned} \tag{45}$$

From these definitions it follows that

$$\begin{aligned} &(\mathbf{q}'+i\mathbf{q})^2/p^2 = e^{i\phi}\rho, \\ &(\mathbf{q}'-i\mathbf{q})^2/p^2 = e^{-i\phi}\rho, \\ &(\mathbf{q}'+i\mathbf{q})(\mathbf{q}'-i\mathbf{q})/p^2 = 1, \end{aligned} \tag{46}$$

$$\begin{aligned} (\mathbf{q}'+i\mathbf{q})_{+1}/p = {} & -e^{i\phi/2}\left[(\sqrt{1-\rho}+\sqrt{1+\rho})D^1_{11}(\alpha\beta\gamma)\right. \\ & \left. + (\sqrt{1-\rho}-\sqrt{1+\rho})D^1_{-11}(\alpha\beta\gamma)\right]/2, \end{aligned}$$

$$\begin{aligned} &(\mathbf{q}'-i\mathbf{q})_{+1}/p \\ &= -e^{-i\phi/2}\left[(\sqrt{1-\rho}+\sqrt{1+\rho})D^1_{11}(\alpha\beta\gamma)\right. \\ &\quad \left. + (\sqrt{1-\rho}-\sqrt{1+\rho})D^1_{-11}(\alpha\beta\gamma)\right]/2, \end{aligned}$$

$$[(\mathbf{q}'+i\mathbf{q}) \wedge (\mathbf{q}'-i\mathbf{q})]_{+1}/p^2 = i\sqrt{1-\rho^2}\, D^1_{01}(\alpha\beta\gamma),$$

where the D functions are the usual matrix elements of the rotation group SO(3). Using Eqs. (35), (43b), and (46), we obtain the expression

$$Y^{\lambda\mu RL}{}_L(\rho\phi,\alpha\beta\gamma)$$

$$= \sum_j c_j (e^{i\phi}\rho)^{S_j^a}(e^{-i\phi}\rho)^{S_j^b}(-e^{i\phi/2}[(\sqrt{1-\rho} + \sqrt{1+\rho})D^{1}_{11}(\alpha\beta\gamma) + (\sqrt{1-\rho} - \sqrt{1+\rho}) \times D^{1}_{-11}(\alpha\beta\gamma)]/2)^{V_j^a}(e^{-i\phi/2} \times [(\sqrt{1-\rho} - \sqrt{1+\rho})D^{1}_{11}(\alpha\beta\gamma) + (\sqrt{1-\rho} + \sqrt{1+\rho}) \times D^{1}_{-11}(\alpha\beta\gamma)]/2)^{V_j^b}(i\sqrt{1-\rho^2}D^{1}_{01}(\alpha\beta\gamma))^{V_{ab}}, \quad (47a)$$

or developing the different factors

$$Y^{\lambda\mu RL}{}_L(\rho\phi,\alpha\beta\gamma) = \sum_j c_j i^{V_{ab}}(-1)^{V_j^a}(1/2)^{V_j^a+V_j^b}e^{i\mu\phi/2}(1-\rho^2)^{V_{ab}/2}\rho^{S_j^a+S_j^b} \times \sum_{n_a,n_b=0}^{V_j^a,V_j^b} \begin{pmatrix} V_j^a \\ n_a \end{pmatrix}\begin{pmatrix} V_j^b \\ n_b \end{pmatrix}\left(\begin{matrix} n_a+n_b & V_j^a+V_j^b-n_a-n_b \\ n_a+n_b & -V_j^a-V_j^b+n_a+n_b \end{matrix}\middle| \begin{matrix} V_j^a+V_j^b \\ -V_j^a-V_j^b+2(n_a+n_b) \end{matrix}\right) \times \left(\begin{matrix} V_j^a+V_j^b & V_{ab} \\ -V_j^a-V_j^b+2(n_a+n_b) & 0 \end{matrix}\middle| \begin{matrix} L \\ -V_j^a-V_j^b+2(n_a+n_b) \end{matrix}\right)(\sqrt{1-\rho}+\sqrt{1+\rho})^{V_j^b+n_a-n_b} \times (\sqrt{1-\rho}-\sqrt{1+\rho})^{V_j^a-n_a+n_b}D^{L}_{-V_j^a-V_j^b+2(n_a+n_b),L}(\alpha\beta\gamma), \quad (47b)$$

where (|) is a Clebsch–Gordan coefficient of SO(3). This last equation gives the announced explicit expression of the $Y^{\lambda\mu RL}{}_M$ functions if we add some remarks. In Eq. (47b) the dependence on the eigenvalue L of the third component of the angular momentum is only through a subindex of the matrix elements D. Obviously J_{-1} decreases one unit in this subindex, and the general expression of $Y^{\lambda\mu RL}{}_M$ is obtained replacing L by M in such subindex. Remember that V_{ab} is 0 or 1 according to $\lambda + L$ be even or odd, and only two of the four natural numbers S_j^a, S_j^b, V_j^a, V_j^b are independent, although bounded. Thus, taking for instance, S_j^a, S_j^b or equivalently $S_j^a + S_j^b$, S_j^a we have [cf. Eq. (36)]

$$V_j^a + V_j^b = L - V_{ab},$$
$$V_j^a - V_j^b = \mu + 2(S_j^b - S_j^a), \quad (48)$$
$$S_j^a + S_j^b \leqslant (\lambda - L - V_{ab})/2.$$

3. SYMMETRY PROPERTIES OF THE FUNCTIONS $Y^{\lambda\mu RL}{}_M$

In the last section we have obtained the general expression of $Y^{\lambda\mu RL}{}_M$, but we have not fixed its normalization and phase. The coefficients c_i are normalized according to

$$\int_{S_3} \overline{Y^{\lambda\mu RL}{}_M(\omega)}\, Y^{\lambda'\mu'R'L'}{}_{M'}(\omega)\, d\omega = \delta_{\lambda\lambda'}\delta_{\mu\mu'}\delta_{RR'}\delta_{LL'}\delta_{MM'}, \quad (49a)$$

where

$$d\omega = \left(\frac{1}{\pi}\rho d\rho d\phi\right)\left(\frac{1}{8\pi^2}\, d\alpha d(\cos\beta) d\gamma\right). \quad (49b)$$

The phase is fixed imposing that the only complex factors in $Y^{\lambda\mu RL}{}_M$ are the exponential $e^{i\mu\phi/2}$ and the matrix elements D (the factor $i^{V_{ab}}$ in Eq. (47b) disappears) and taking the following sign convention. If $\mu > 0$ (or $\mu = 0$ and $R \geqslant 0$) the first coefficient c_1 must be positive, where it is assumed the order relation

$j \leqslant j'$ if $S_j^a + S_j^b < S_{j'}^a + S_{j'}^b$ or if $S_j^b \leqslant S_{j'}^b$ when $S_j^a + S_j^b = S_{j'}^a + S_{j'}^b$. If $\mu < 0$ (or $\mu = 0$ and $d < 0$) it is assumed the equation

$$\overline{Y^{\lambda\mu RL}{}_M(\omega)} = (-1)^M Y^{\lambda-\mu-RL}{}_{-M}(\omega). \quad (50)$$

The parity transformation P changes the sign of $\mathbf{q}$, $\mathbf{q}'$ [cf. Eq. (3)]. Then, as $p^{\lambda+3,\lambda}\mu^{RL}{}_M$ is a polynomial of order λ in these vectors [cf. Eq. (43b)],

$$Y^{\lambda\mu RL}{}_M(P\omega) = (-1)^\lambda Y^{\lambda\mu RL}{}_M(\omega). \quad (51)$$

The transposition of the two first particles T_{AB} interchanges the roles of $\mathbf{a}^+$ and $\mathbf{b}^+$ [cf. Eqs. (3), (23)]. Therefore, it changes the sign of S and X while Λ^2, J^2 and J_3 remain unchanged; this implies that

$$Y^{\lambda\mu RL}{}_M(T_{AB}\omega) = (-1)^{\lambda+L}Y^{\lambda-\mu-RL}{}_M(\omega), \quad (52)$$

where the extra sign arises from the sign convention. The cyclic permutation C_{ABC} can be written [cf. Eqs. (8b), (9a)]

$$C_{\mathrm{ABC}} = \cos\tfrac{2}{3}\pi 1 + \sin\tfrac{2}{3}\pi\tilde{S}, \quad (53a)$$

therefore

$$Y^{\lambda\mu RL}{}_M(C_{\mathrm{ABC}}\omega) = e^{2\pi i\mu/3}Y^{\lambda\mu RL}{}_M(\omega). \quad (53b)$$

4. "3ν" COEFFICIENTS

In some problems[1,2] it is interesting to expand the products of $Y^{\lambda\mu RL}{}_M$ functions in linear combinations of themselves. This is done using the "3ν" coefficients. [|],

$$Y^{\lambda_1\mu_1R_1L_1}{}_{M_1}(\omega)\overline{Y^{\lambda_2\mu_2R_2L_2}{}_{M_2}(\omega)} = \sum_{\lambda_3\mu_3R_3L_3M_3}\begin{pmatrix} L_1 & L_3 \\ M_1 & M_3 \end{pmatrix}\begin{vmatrix} L_2 \\ M_2 \end{vmatrix} \times \left(\begin{matrix} \lambda_1\mu_1R_1 & \lambda_3\mu_3R_3 \\ L_1 & L_3 \end{matrix}\middle| \begin{matrix} \lambda_2\mu_2R_2 \\ L_2 \end{matrix}\right)Y^{\lambda_3\mu_3R_3L_3}{}_{M_3}(\omega), \quad (54)$$

where (|) is the corresponding Clebsch–Gordan coefficient of SO(3). In order to study their properties it is more convenient to introduce the symmetrical "3ν" symbol, [ ],

$$\int_{S_3} Y^{\lambda_1\mu_1R_1L_1}{}_{M_1}(\omega)Y^{\lambda_2\mu_2R_2L_2}{}_{M_2}(\omega)Y^{\lambda_3\mu_3R_3L_3}{}_{M_3}(\omega)\, d\omega = \begin{pmatrix} L_1 & L_2 & L_3 \\ M_1 & M_2 & M_3 \end{pmatrix}\begin{pmatrix} \lambda_1\mu_1R_1 & \lambda_2\mu_2R_2 & \lambda_3\mu_3R_3 \\ L_1 & L_2 & L_3 \end{pmatrix}, \quad (55)$$

where ( ) is the usual 3J symbol. As can be easily proved,

$$\begin{bmatrix} \lambda_1\mu_1R_1 & \lambda_2\mu_2R_2 & \lambda_3\mu_3R_3 \\ L_1 & L_2 & L_3 \end{bmatrix}$$

$$= \frac{(-1)^{L_1 - L_2}}{\sqrt{2L_3+1}} \begin{bmatrix} \lambda_1\mu_1 R_1 & \lambda_2\mu_2 R_2 & \lambda_3 - \mu_3 - R_3 \\ L_1 & L_2 & L_3 \end{bmatrix}. \quad (56)$$

Using the explicit expression of $Y^{\lambda\mu RL}{}_M$ in Eq. (47b), with the corresponding conventions, and the 3ν symbol definition in Eq. (55) we obtain its explicit expression

$$\begin{bmatrix} \lambda_1\mu_1 R_1 & \lambda_2\mu_2 R_2 & \lambda_3\mu_3 R_3 \\ L_1 & L_2 & L_3 \end{bmatrix} = \delta_{0,\mu_1+\mu_2+\mu_3} \sum_{i_1 i_2 i_3} c_{i_1} c_{i_2} c_{i_3} Q_{i_1 i_2 i_3}, \quad (57a)$$

where i_x, $x = 1,2,3$ are restricted by equations of the type (48) and

$$Q_{i_1 i_2 i_3} = (-1)^{\Sigma_x L_x} \left(\tfrac{1}{2}\right)^{\Sigma_x (V^a_{i_x} + V^b_{i_x})/2} \sum_{n^a_x, n^b_x = 0,}^{V^a_{i_x}, V^b_{i_x}} (-1)^{\Sigma_x (n^a_x + n^b_x)}$$
$$\times \begin{pmatrix} L_1 & L_2 & L_3 \\ 2(n^a_1 + n^b_1) - V^a_1 - V^n_1 & 2(n^a_2 + n^b_2) - V^a_2 - V^b_2 & 2(n^a_3 + n^b_3) - V^a_3 - V^b_3 \end{pmatrix}$$
$$\times \prod_x \left[\binom{V^a_{i_x}}{n^a_x} \binom{V^b_{i_x}}{n^b_x} \binom{n^a_x + n^b_x \quad V^a_{i_x} + V^b_{i_x} - n^a_x - n^b_x \quad \Big| \quad V^a_{i_x} + V^b_{i_x}}{n^a_x + n^b_x \quad n^a_x + n^b_x - V^a_{i_x} - V^b_{i_x} \quad \Big| \quad 2(n^a_x + n^b_x) - V^a_{i_x} - V^b_{i_x}} \right.$$
$$\left. \times \begin{pmatrix} V^a_{i_x} + V^b_{i_x} & V^{ab}_x & \Big| & L_x \\ 2(n^a_x + n^b_x) - V^a_{i_x} - V^b_{i_x} & 0 & \Big| & 2(n^a_x + n^b_x) - V^a_{i_x} - V^b_{i_x} \end{pmatrix} \right]$$
$$\times \sum_{\substack{k=0 \\ k + \Sigma_x V^{ab}_x \text{even}}}^{N} \binom{N}{k} B\left(\frac{\Sigma_x V^{ab}_x + k}{2} + 1, \sum_x \left(\frac{S^a_{i_x} + S^b_{i_x}}{2} + \frac{V^a_{i_x} + V^b_{i_x}}{4} \right) - \frac{N}{2} + 1 \right), \quad (57b)$$

where

$$N = \left| \sum_x \left(\frac{V^a_{i_x} - V^b_{i_x}}{2} - n^a_x + n^b_x \right) \right|$$

and $B(x,y)$ is the usual Beta function.[18] In Ref. 1 the 3ν coefficients are tabulated for the more usual cases. They have been calculated using Eqs. (56), (57) and the computer program which evaluates the c_i coefficients.

The 3ν symbols have symmetry properties similar to those of the $3J$ symbols. From their explicit expression in Eq. (57) it follows that

$$\begin{bmatrix} \lambda_1\mu_1 R_1 & \lambda_2\mu_2 R_2 & \lambda_3\mu_3 R_3 \\ L_1 & L_2 & L_3 \end{bmatrix} = \text{real}, \quad (58)$$

$$\begin{bmatrix} \lambda_1\mu_1 R_1 & \lambda_2\mu_2 R_2 & \lambda_3\mu_3 R_3 \\ L_1 & L_2 & L_3 \end{bmatrix} = \begin{bmatrix} \lambda_2\mu_2 R_2 & \lambda_3\mu_3 R_3 & \lambda_1\mu_1 R_1 \\ L_2 & L_3 & L_1 \end{bmatrix}$$
$$= (-1)^{L_1 + L_2 + L_3} \begin{bmatrix} \lambda_2\mu_2 R_2 & \lambda_1\mu_1 R_1 & \lambda_3\mu_3 R_3 \\ L_2 & L_1 & L_3 \end{bmatrix}, \quad (59)$$

if we note that the c_i coefficients are real and the permutations of the indices 1, 2, 3 are only relevant for the $3J$ symbol in Eq. (57b). The relation

$$\begin{bmatrix} \lambda_1\mu_1 R_1 & \lambda_2\mu_2 R_2 & \lambda_3\mu_3 R_3 \\ L_1 & L_2 & L_3 \end{bmatrix}$$
$$= (-1)^{L_1 + L_2 + L_3} \times \begin{bmatrix} \lambda_1 - \mu_1 - R_1 & \lambda_2 - \mu_2 - R_2 & \lambda_3 - \mu_3 - R_3 \\ L_1 & L_2 & L_3 \end{bmatrix} \quad (60)$$

can be proved by introducing Eqs. (51), (52) into the definition (55). The indices λ, μ, L must also satisfy the equations:

$$\lambda_1 + \lambda_2 + \lambda_3 = \text{even}, \quad (61)$$

$$|\lambda_1 - \lambda_2| \leqslant \lambda_3 \leqslant \lambda_1 + \lambda_2, \quad (62)$$

$$\mu_1 + \mu_2 + \mu_3 = 0, \quad (63)$$

$$|L_1 - L_2| \leqslant L_3 \leqslant L_1 + L_2. \quad (64)$$

Equations (63) and (64) follow from the delta of Kronecker in Eq. (57a) and the $3J$ symbol in Eq. (57b), respectively. Equation (61) follows from Eqs. (39) and (63), while Eq. (62) can be obtained from the reduction of the product of two SU(3) representations.

ACKNOWLEDGMENTS

The author is indebted to Professor M. G. Doncel for suggesting the problem and many useful discussions, and to Professor A. Méndez, Professor L. Michel, and Professor P. Minnaert for helpful suggestions.

[1] F. del Aguila and M. G. Doncel, UAB preprint, UAB-FT-51, 1979.
[2] F. del Aguila, M. Aguilar-Benítez, M. Cerrada, M. G. Doncel, and Ph. Gavillet, Z. Physik C. **4**, 1 (1980).
[3] A. J. Dragt, J. Math. Phys. **6**, 533, 1621 (1965).
[4] G. W. Mackey, *The Theory of Unitary Group Representations* (University of Chicago Press, Chicago, 1976), p. 130.
[5] For instance, in Ref. 4, p. 162.
[6] Straightforward application of the Branching Theorem, H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1970), p. 175.
[7] L. Pontrjagin, *Topological Groups* (Gordon and Breach, New York, 1966), p. 225.
[8] F. T. Smith, Phys. Rev. **120**, 1058 (1960).
[9] V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961).
[10] M. Moshinsky, J. Patera, R. T. Sharp, and P. Winternitz, Ann. Phys. **95**, 139 (1975).
[11] B. R. Judd, W. Miller Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[12] V. V. Pustovalov and Ya. A. Smorodinskii, Sov. J. Nucl. Phys. **10**, 729 (1970).
[13] L. Michel and L. A. Radicati, Ann. Inst. Henri Poincaré, A **18**, 185 (1973).
[14] G. Racah, *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964), p. 33.
[15] G. Racah, Rev. Mod. Phys. **21**, 494 (1949).
[16] R. H. Dalitz, Phil. Mag. **44**, 1068 (1953); E. Fabri, Nuovo Cimento **11**, 479 (1954).
[17] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N.J. 1960).
[18] A. Erdélyi, *Higher Transcendental Functions*, I (McGraw-Hill, New York, 1953), p. 9.

On a new relation between semisimple Lie algebras [a),b)]

J. Patera
Centre de recherches de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada

R. T. Sharp
Department of Physics, McGill University, Montreal, Quebec, Canada

R. Slansky
Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545

(Received 4 February 1980; accepted for publication 20 March 1980)

A recently discovered relation between pairs of semisimple Lie algebras is further investigated. This relation, which is called subjoining and denoted by " $>$ ", is a generalization of inclusion, where a subalgebra is embedded in an algebra. Nontrivial subjoinings of two algebras of the same type are described. New chains of algebras involving proper inclusions and subjoinings can be formed. Infinite families of maximal subjoinings $C_n > B_n$ and $B_n > C_n$ are shown.

I. INTRODUCTION

The purpose of this paper is to investigate further the relation of subjoining[1] between two semisimple Lie algebras. Subjoining is a generalization of embedding a subalgebra in a semisimple algebra in the sense that every embedding is a subjoining, and it gives new relations between pairs of algebras that are not algebra–subalgebra pairs. As far as we know, this relation has not been reported in the mathematics literature and it may prove to be useful in applications such as symmetry breaking, construction of bases, classification of states, and in other situations where the analysis of subgroups is useful. It has already proved to be a valuable shortcut in lengthy computations of certain generating functions in group representation theory.[1] Our aim here is to provide many new examples, which ought to be helpful for developing a better understanding of subjoining.

Among the examples here is an infinite family of subjoinings of mutually isomorphic Lie algebras and subjoinings $B_n > C_n$ and $C_n > B_n$ for $n \geqslant 2$. We use throughout the paper the symbol $>$ for subjoining; when a subjoining is a usual embedding (inclusion) we use $\supset$.Subjoinings and inclusions can be combined into chains; this also creates new subjoinings, for instance $C_n > B_n \supset D_n$ gives $C_n > D_n$, which is not maximal.

All considerations of the paper are carried out in terms of weight systems of finite dimensional representations of semisimple Lie algebras (groups). In this way all the operations are well defined. Naturally, one would like to employ the one-to-one correspondence between a representation and its weight system and to transfer all the results into language of representations and/or the representation spaces. At present, the possibility of such a transfer is an open question.

The principal difficulty that is encountered can be illustrated as follows: suppose that W_φ is the set of weights of a representation φ of an algebra G, and W_ψ is a similar set referring to a representation ψ of G; if, furthermore,

$$W_\varphi \cap W_\psi = W_\psi, \tag{1.1}$$

i.e., the set of weights W_φ contains every weight of W_ψ, then $W_\varphi \setminus W_\psi$ can be defined to be the set of weights W_φ from which all the weights common to representations φ and ψ have been subtracted. Hence the operation of subtraction $W_\varphi \setminus W_\psi$ is well defined provided (1.1) holds. A necessary condition for (1.1) to hold is *congruence*[2] of representations. In analogy with the one-to-one correspondence,

$$\varphi + \psi \longleftrightarrow W_\varphi \cup W_\psi, \tag{1.2}$$

where on the left is the direct sum of representations φ and ψ, and on the right stands its weight system, we would like to infer a relation between $W_\varphi \setminus W_\psi$ and the "difference" φ-ψ. The difficulty is in defining the latter operation.

We say that a semisimple algebra H of rank r_H is *subjoined* to a semisimple algebra G of rank $r_G \geqslant r_H$, and write $G > H$, if there exists an $r_G \times r_H$ matrix P of rank r_H such that for *every* finite dimensional representation $\varphi(G)$ of G, there is a branching rule,

$$PW_{\varphi(G)} = W_{\psi(H)} \setminus W_{\omega(H)}, \tag{1.3}$$

where ψ(H) and ω(H) are representations of H. Clearly the matrix P is not unique. We do not need to distinguish two matrices P and P' that give the same result when acting on W_φ of (1.3), namely $W_\varphi \setminus W_\psi$. Hence P is specified up to transformation by elements of the Weyl groups of G and H. With a suitable choice of bases in the root spaces of G and H, all elements of P are non-negative integers. The algebra H is called a *hypoalgebra* of G.

If H is a subalgebra of G then clearly it is also subjoined to G. Indeed, in such a case the right side of (1.3) equals $W_{\psi(H)}$, i.e., the set $W_{\omega(H)}$ is empty. Many examples of the matrices P for subalgebras are found in Table IV of Ref. 3.

For simplicity of notation we shall write (1.3) as

$$P\varphi(G) = \psi(H) - \omega(H) \tag{1.3a}$$

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

 0022-2488/80/092335-07$1.00

or

$$\varphi(G) > \psi(H) - \omega(H), \tag{1.3b}$$

and we shall talk about representations instead of weight systems.

In order to specify a subjoining of H to G, $H < G$, we either give the matrix P or the result of its action on the weight system W_φ of a faithful representation $\varphi(G)$ of G, that is, we give the representations $\psi(H)$ and $\omega(H)$ of (1.3b).

A subjoining $G > H$ is called *maximal* if it cannot be extended into $G > H' > H$, where $G > H'$ and $H' > H$ are specified by a matrix P such that $|\det P| > 1$.

Conventions concerning the representations, numbering of simple roots, etc., adopted here coincide with those of Dynkin.[4,5]

In Sec. II an example of $C_2 \supset H$ and $C_2 > H$, where C_2 is the Lie algebra of groups Sp(4) and O(5), and H in both cases is of the type $A_1 + A_1$ [group SU(2)×SU(2)], is considered in detail in order to illustrate the distinction between a usual inclusion and a genuine subjoining. Section III contains a description of maximal subjoinings between two isomorphic algebras. With only a few exceptions these are given by $P = nI$, where I is an identity matrix and n is a prime number. Some other subjoinings that are maximal and do not involve isomorphic algebras are described in Sec. IV. Section V contains comments and remarks.

II. AN EXAMPLE

Let us consider an example where G is the Lie algebra C_2 of the group Sp(4). C_2 contains only one subalgebra S of rank two; it is $A_1 + A_1$. [In other words, Sp(4) $\supset S$ = SU(2)×SU(2).] However, there is another algebra H, which is also of type $A_1 + A_1$, that is subjoined to C_2. [Sp(4) $> H$ = SU(2)×SU(2).] In order to illustrate the difference between embedding SU(2)×SU(2) and subjoining SU(2)×SU(2) in Sp(4), we first examine the weight system of the adjoint representation of C_2. The weight system of the adjoint representation is also the root system of the algebra. Every weight Λ is a linear combination of the simple roots, which are conveniently summarized by the Dynkin diagram.[4] In terms of the simple root basis, a weight Λ is conveniently specified by the integer coordinates,

$$\lambda_i = 2(\Lambda,\alpha_i)/(\alpha_i,\alpha_i), \tag{2.1}$$

where the simple roots α_i are numbered as in Table I of Ref. 4. The C_2 algebra has the two simple roots, α_1 and α_2, shown in Fig. 1a, and the weight system of the adjoint representation in the simple root basis and the components (2.1) is

$$a_1 = 2\alpha_1 + \alpha_2 = (2,0), \quad a_2 = \alpha_1 + \alpha_2 = (0,1),$$
$$a_3 = \alpha_1 = (2,-1), \quad \text{and} \quad a_4 = \alpha_2 = (-2,2). \tag{2.2}$$

The four negative roots are $-a_1$, $-a_2$, $-a_3$, and $-a_4$, and the two zero roots correspond to the Cartan subalgebra. The a_i are indicated in Fig. 1a.

The subalgebra may be derived from the C_2 root diagram by the projection matrix[3]

$$P_S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{2.3}$$

which transforms the root system (and any weight system) of C_2 into a root (weight) system of S. If we arrange the coordinates of the weights vertically, then the explicit calculation of the positive weights of S gives

$$P_S a_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \beta_1,$$
$$P_S a_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \tfrac{1}{2}\beta_1 + \tfrac{1}{2}\beta_2,$$
$$P_S a_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \tfrac{1}{2}\beta - \tfrac{1}{2}\beta_2,$$
$$P_S a_4 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \beta_2, \tag{2.4}$$

where β_1 and β_2 are the two simple roots of S = SU(2)×SU(2) shown in Fig. 1b. The weight system is completed with the negatives of $P_S a_i$ and the two zero weights. The root system after projection is the weight system of a reducible representation of the subalgebra. There are three irreducible components corresponding to the highest weights, β_1, β_2, and $\frac{1}{2}\beta_1 + \frac{1}{2}\beta_2$. Thus, the branching rule for the adjoint of C_2 into irreducible representations of $A_1 + A_1$, written in terms of highest weights in the notation of (1.3a), is

$$P_S(20) = (2,0) + (0,2) + (1,1). \tag{2.5}$$

The root system of the subalgebra (see Fig. 1b) consists of β_1, β_2, their negatives, and two zeros. In a similar way the matrix P_S allows the reduction of the weight system of any

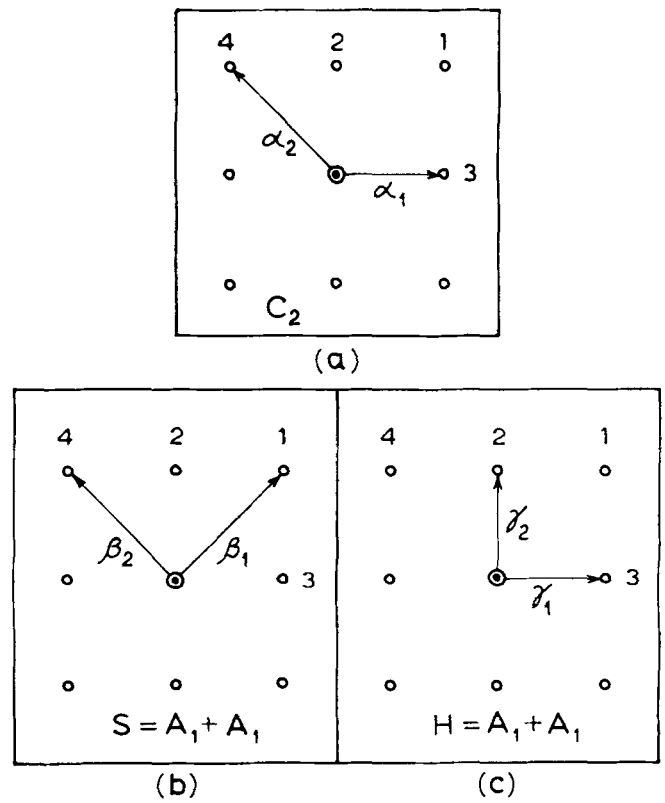


FIG. 1. Roots and simple roots for (a) C_2 which is the algebra of Sp(4) group; (b) the subalgebra $A_1 + A_1$ of C_2, where A_1 is the algebra of SU(2) group; (c) the hypoalgebra $A_1 + A_1$ of C_2. Roots are indicated by circles, simple roots by arrows.

representation of C_2 into a representation of the subalgebra S (see Ref. 6).

The procedure of subjoining allows the consideration of the projection matrix

$$P_H = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}. \tag{2.6}$$

(The origin of P_H will become clearer in the next sections.) P_H projects the root system of C_2 onto the weights, $W_{\varphi(H)}$,

$$\begin{aligned} P_H a_1 &= \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \gamma_1 + \gamma_2, \\ P_H a_2 &= \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \gamma_2, \\ P_H a_3 &= \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \gamma_1, \\ P_H a_4 &= \begin{pmatrix} -2 \\ 2 \end{pmatrix} = -\gamma_1 + \gamma_2, \end{aligned} \tag{2.7}$$

plus two zero weights and the negatives of (2.7). The simple roots of H are γ_1 and γ_2, and are shown in Fig. 1c. Equations (2.7) show that the highest weights of the irreducible representations of H are (2,2) and (0,0); the branching rule is

$$P_H(20) = (2,2) + (0,0). \tag{2.8}$$

Note that the adjoint representation of $A_1 + A_1$ does not appear on the right hand side of (2.8).

The projection of the weight system of a simple algebra onto that of a hypoalgebra has a feature that is exemplified by studying the 5-dimensional representation of C_2, which has highest weight (01). The application of P_H to the weights of this representation gives the weight system

$$P_H(01) = (2,0) + (0,2) - (0,0), \tag{2.9}$$

in the notation of (1.3a). Of course, no weight ever appears a negative number of times in the branching rule; (2,0) and (0,2) each have one zero weight.

The result of the action of the projection matrices P_S and P_H on weight systems of all representations of C_2 can be expressed in terms of the corresponding generating functions. Thus for $C_2 \supset S$, the generating function is[1]

$$1/(1 - N_1S_1)(1 - N_1S_2)(1 - N_2)(1 - N_2S_1S_2). \tag{2.10}$$

In order to retrieve information from (2.10), one has to expand it in a power series. Its general term has the form

$$m_{n_1n_2}^{s_1s_2} N_1^{n_1} N_2^{n_2} S_1^{s_1} S_2^{s_2},$$

where n_1,n_2,s_1,s_2 and $m_{n_1n_2}^{s_1s_2}$ are positive integers. The presence of such a term in the series means that the weight system of the irreducible representation (n_1,n_2) of C_2 after it has been projected by means of P_S into a weight system of a (reducible) representation of S, contains the weight system of the irreducible representation (s_1,s_2) of S exactly $m_{n_1n_2}^{s_1s_2}$ times.

In the case of $C_2 > H$ a similar interpretation is given to the generating function[1]

$$\begin{aligned} &(1 + N_1N_2H_1H_2)/(1 - N_1^2)(1 - N_2H_2^2)(1 - N_2H_1^2) \\ &\quad \times (1 - N_1H_1H_2)(1 + N_2). \end{aligned} \tag{2.11}$$

In this case the role of variables N_1 and N_2 is the same as in (2.10), i.e., in the expansion they carry the coordinates of the C_2 highest weights, the powers of variables H_1 and H_2 specify highest weights of representations of H. When a coefficient $m_{n_1n_2}^{h_1h_2}$ of the power $N_1^{n_1}N_2^{n_2}H_1^{h_1}H_2^{h_2}$ of the series is negative, it implies that the weight system of the representation (h_1,h_2) of H has to be subtracted $m_{n_1n_2}^{h_1h_2}$ times from the result of the projection of the weights of (n_1,n_2) by the matrix P_H. Condition (1.1) is automatically satisfied, i.e., a weight that is being subtracted is always there in the first place.

As an illustration of the use of the generating functions let us answer the following question. Which representations of C_2 contain the adjoint representation, (2,0) + (0,2), of the subalgebra S and the hypoalgebra H? For that one needs to find the expression multiplying $S_1^2 + S_2^2$ and $H_1^2 + H_2^2$ respectively in the power series of (2.10) and (2.11). That expression is equal to $\Sigma_{k=1}^{\infty} N_1^2 N_2^k$ in case of (2.10) and $-\Sigma_{l,k=1}^{\infty} N_1^{2k}(-N_2)^{l+1}$ in (2.11). Hence the adjoint representation of S is contained once only in representation $(2,k)$, $k = 0,1,2,\ldots$ of C_2, while the adjoint of H occurs in $(2k,l)$ (l odd) once, and is subtracted once in $(2k,l)$ (l even and nonzero). Clearly this is in agreement with the particular cases considered above.

III. SUBJOINING OF ISOMORPHIC ALGEBRAS

Let us restrict our consideration to relations of the type $G > G$. Unlike the embedding $G \supset G$, the subjoining $G > G$ turns out to be a nontrivial relation. Moreover, there are infinitely many of them that are maximal. The present section is devoted to the description of such subjoinings. In order to avoid an ambiguity, the hypoalgebra is distinguished by a prime.

A. Rank one algebras

There is only one semisimple Lie algebra of rank one corresponding to groups SU(2) and O(3), it is of type A_1. A relation $A_1 > A_1'$ is specified by the matrix P, which, in the present case, is just an integer.

Proceeding as in the previous section, one verifies that P can be any integer. Without loss of generality it suffices to consider only positive integers. Instead of going into details in this trivial case, we point out that the result of the action of $P = n$ on any irreducible representation of A_1 is easily inferred from the following generating function:

$$\begin{aligned} F(A,a) &= (1 - Aa^{n-2})/(1 - Aa^n)(1 - A^2) \\ &= 1 + A(a^n - a^{n-2}) + A^2(a^{2n} - a^{2n-2} + 1) + \cdots, \\ &\qquad (n \geq 2). \end{aligned} \tag{3.1}$$

Thus the linear term in A indicates that for the two-dimensional representation (1) one has the reduction

$$(1) > (n) - (n-2), \tag{3.2}$$

which, in terms of dimensions, means $2 = n + 1 - (n - 1)$. Similarly, from the quadratic term one concludes that

$$(2) > (2n) - (2n-2) + (0), \tag{3.3}$$

which again implies an equality of dimensions $3 = 2n + 1 - (2n - 1) + 1$. The relations (3.2) and (3.3) should always be interpreted as the corresponding relations between the weight systems.

A particularly interesting question is which representations of A_1 contain the vector representation (2) of A_1'? To find the answer, let us reorder the series (3.1) according to powers of a and find the coefficient of a^2:

$$F(A,a)=\frac{1}{1-A^2}\{[1+Aa^n+(Aa^n)^2+\cdots]-Aa^{n-2}[1+Aa^n+(Aa^n)^2+\cdots]\}.$$

Obviously a^2 occurs in the series only in two cases, namely $n=2$ or 4. In the second case it occurs with a negative sign, i.e., it is being "subtracted." In the first case the relevant term is

$$a^2\frac{A}{1-A^2}=a^2(A+A^3+A^5+\cdots).$$

Hence the representation (2) of A_1' is found once in every even dimensional representation of A_1.

Consider $A_1>A_1'>A_1''$, and suppose that $P=n_1$ in the case of $A_1>A_1'$, and $P=n_2$ for $A_1'>A_1''$. Then $P=n_1n_2$ for $A_1>A_1''$, as can easily be verified considering weights of a representation. Consequently, the indecomposable subjoinings between two algebras A_1 are those with n equal to a prime number. In that case we say that the hypoalgebra is maximal.

B. Algebras of rank two

There are four semisimple Lie algebras of rank two: A_2, C_2, G_2, and A_1+A_1.

Maximal subjoinings $A_2>A_2'$ correspond to $P=\left(\begin{smallmatrix}n&0\\0&n\end{smallmatrix}\right)$, where n is a prime number.

Maximal subjoinings $C_2>C_2'$ are those given by

$$P=\begin{pmatrix}n&0\\0&n\end{pmatrix}\quad\text{and}\quad\begin{pmatrix}0&1\\2&0\end{pmatrix},\tag{3.4}$$

where n is a prime number different from 2. Indeed, $P=\left(\begin{smallmatrix}2&0\\0&2\end{smallmatrix}\right)=\left(\begin{smallmatrix}0&1\\2&0\end{smallmatrix}\right)\left(\begin{smallmatrix}0&1\\2&0\end{smallmatrix}\right)$, hence such a subjoining is composite, i.e., not maximal.

Maximal subjoinings $G_2>G_2'$ correspond to

$$P=\begin{pmatrix}n&0\\0&n\end{pmatrix}\quad\text{and}\quad\begin{pmatrix}0&1\\3&0\end{pmatrix},\tag{3.5}$$

with n prime, $n\neq 3$.

The matrices $P=\left(\begin{smallmatrix}0&1\\2&0\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0&1\\3&0\end{smallmatrix}\right)$, acting on the single roots of C_2 and G_2, interchange the roots without changing the length of the longer one and stretch the shorter root in such a way that it becomes the longer simple root of the algebra.

The above assertions can be verified by considering the action of P on weight systems of representations of the algebras. It is again possible to summarize the action in terms of a generating function. In particular, the subjoining $A_2>A_2'$ with $P=\left(\begin{smallmatrix}2&0\\0&2\end{smallmatrix}\right)$ leads to the generating function

$$F(A,B,a,b)=\frac{1}{(1-Aa^2)(1-Bb^2)(1-AB)}\left(\frac{1}{1+Ab}-\frac{Ba}{1-Ba}\right)$$
$$=1+A(a^2-b)+B(b^2-a)+AB(a^2b^2-a^3-b^2+1)+\cdots.\tag{3.6}$$

There is no problem in finding the generating function for $P=n\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ with general n except that the computation is laborious. The subjoining $C_2>C_2'$ with $P=2\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0&1\\2&0\end{smallmatrix}\right)$ imply respectively

$$F=\frac{1}{(1-Aa^2)(1+A)(1-Bb^2)(1-B)}\times\left(\frac{1}{1+AB}-\frac{Ba^2}{1+Ba^2}\right),\tag{3.7}$$

$$F=\frac{1}{(1-Ab)(1+A)(1-Ba^2)(1+Bb)}.\tag{3.8}$$

It was pointed out earlier that the first of the two subjoinings is a result of two successive applications of $P=\left(\begin{smallmatrix}0&1\\2&0\end{smallmatrix}\right)$. The corresponding manipulation of generating functions consists of substituting (3.8) into another function of the same type. The result is (3.7). How to carry out such a substitution as has been described previously.[1]

The subjoinings $G_2>G_2'$ with $P=\left(\begin{smallmatrix}2&0\\0&2\end{smallmatrix}\right)$ and $P=\left(\begin{smallmatrix}0&1\\3&0\end{smallmatrix}\right)$ give respectively

$$F=\frac{1}{(1-Aa^2)(1-A)(1-Bb^2)(1-B)}\times\left[\frac{1}{(1+Aa)(1+Ba)}+\frac{Ab^2}{(1+Aa)(1-Ab^2)}-\frac{Ab^3}{(1-Ab^2)(1+Ab^3)}-\frac{Bb}{(1+Ab^3)(1+Bb)}\right],\tag{3.9}$$

$$F=\frac{1}{(1-Ba)(1-B^2a)(1-Ab^3)(1-A)}\times\left(\frac{1-Bb}{1-B^3b^3}+\frac{A^2a^3-Aab+ABa^2}{1-A^2a^3}\right).\tag{3.10}$$

Subjoinings $A_1+A_1>A_1'+A_1'$ are obvious combinations of $A_1>A_1'$ of the preceding subsection.

C. Higher rank algebras

In most cases of rank >2 maximal subjoinings $G>G'$ are exhausted by matrices P which are prime multiples of identity. There are however some exceptions.

The first exception is $F_4>F_4'$ with $P=2I$. Consider

$$P=\begin{pmatrix}0&0&0&2\\0&0&2&0\\0&1&0&0\\1&0&0&0\end{pmatrix}.$$

It provides a maximal subjoining $F_4>F_4'$ such that the lowest representation decomposes as

$$(0001)>(2000)-(0100)+(0010)-(0001).$$

Combining two such subjoinings, i.e., multiplying the matrix P with itself, one arrives at the conclusion above.

Other cases in which $P=2I$ is not a maximal subjoining are $C_n>C_n'$ and $B_n>B_n'$ for $n\geqslant 2$. Corresponding explanations are found in the next section.

The outer automorphisms of the algebras A_n, D_n, and E_6 could be defined as a particular case of subjoining realized by matrices

$$P(A_n > A'_n) = \begin{pmatrix} & & & & 1 \\ & & & \cdot & \\ & & \cdot & & \\ & \cdot & & & \\ 1 & & & & \end{pmatrix},$$

$$P(D_n > D'_n) = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & \cdot & & & & & \\ & & & \cdot & & & & \\ & & & & \cdot & & & \\ & & & & & 1 & & \\ & & & & & & 0 & 1 \\ & & & & & & 1 & 0 \end{pmatrix},$$

$$P(E_6 > E'_6) = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Each of them applied twice would result in a trivial subjoining given by the identity matrix. Therefore, we exclude these cases by definition in order to preserve the concept of maximal subjoining.

IV. FURTHER EXAMPLES OF MAXIMAL HYPOALGEBRAS

It is natural to ask about all maximal subjoinings among semisimple Lie algebras. At present we do not have an answer to that question. Therefore, in this section we list the maximal subjoinings that we have encountered so far. In each case we present the reduction of the defining and the adjoint representations to that of the hypoalgebra, and the projection matrix P. In some cases we have also calculated the generating function for the reduction.

1. $C_3 > B_3$, $P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$,

$$(100) > (100) - (0), \tag{4.1}$$
$$(200) > (200) - (100) + (0),$$
$$F = 1/(1 - Aa)(1 + A)(1 - Bb) \times (1 + Ba)(1 - Cc^2)(1 + Cb).$$

2. $B_3 > C_3$, $P = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$,

$$(100) > (200) - (010),$$
$$(010) > (020) - (101) + (0), \tag{4.2}$$
$$(001) > (001) - (100),$$

$$F = \frac{1}{(1 - Aa^2)(1 - Bb^2)(1 - Cc)(1 - B)} \times \left(\frac{1}{(1 + Ca)(1 - ABc^2)(1 + Bac)} + \frac{ACc}{(1 + Ca)(1 - ABc^2)(1 - ACc)} + \frac{AC^2b^2}{(1 + Ca)(1 - ACc)(1 - AC^2b^2)} - \frac{Ab}{(1 - ABc^2)(1 - ACc)(1 + Ab)} + \frac{A^2C^3b^2c}{(1 + Ab)(1 - ACc)(1 - AC^2b^2)} \right).$$

3. $B_n > C_n$, $n > 3$,

$$P = \begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 2 \\ & & & & & 1 \end{pmatrix},$$

$$(10\cdots0) > (20\cdots0) - (010\cdots0), \tag{4.3}$$
$$(010\cdots0) > (020\cdots0) - (1010\cdots0) + (00010\cdots0) + (0), \tag{4.3a}$$
$$(0\cdots01) > (0\cdots01) - (0\cdots0100) + (0\cdots010000) - \cdots, \tag{4.3b}$$
$$(20\cdots0) > (40\cdots0) - (210\cdots0) + (020\cdots0). \tag{4.3c}$$

Clearly this case is a generalization of the subjoining of case 2 above. Similarly the next case generalizes case 1.

4. $C_n > B_n$, $n > 3$,

$$P = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 1 \\ & & & & & 2 \end{pmatrix},$$

$$(10\cdots0) = (10\cdots0) - (0),$$
$$(20\cdots0) = (20\cdots0) - (10\cdots0) + (0).$$

Combining the subjoinings $C_n > B_n$ and $B_n > C_n$ and multiplying the appropriate matrices P, one arrives at the conclusion that neither $C_n > C_n$ with $P = 2I$, nor $B_n > B_n$ with the same P, is a maximal subjoining.

V. REMARKS

(1) Combinations of subjoinings naturally lead to a great variety of subjoinings which are not maximal. One example: The maximal subjoining $C_n > B_n$ of Sec. IV can be combined with the proper inclusion $B_n \supset D_n$ into the subjoining $C_n > D_n$. In terms of matrices P one has

$$P(C_n > D_n) = P(B_n \supset D_n) P(C_n > B_n)$$
$$= \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & 1 & 0 \\ & & & & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 1 \\ & & & & & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & 1 & 0 \\ & & & & 1 & 2 \end{pmatrix}.$$

The low-dimensional representations of C_n then branch into representations of D_n as follows

$$(10\cdots0) > (10\cdots0),$$

$$(20\cdots0) > (20\cdots0) + (0),$$

$$(0\cdots01) > (0\cdots02) + (0\cdots020) - (0\cdots0100).$$

A similar example, where inclusion and subjoining are done in the other order, is $D_n > C_{n-1}$, with

$$P(D_n > C_{n-1}) = P(B_{n-1} > C_{n-1})\, P(D_n \supset B_{n-1}).$$

(2) Our second example of combining of subjoinings and inclusions is an application to computing generating functions. Let our objective be to find the generating function for reduction of representations of

$$SO(7) \supset SO(5) \times SO(2), \tag{5.1}$$

which, in terms of algebras, is

$$B_3 \supset B_2 + P, \tag{5.2}$$

where P stands for the compact 1-parameter subalgebra. If subjoinings are avoided, the inclusion (5.1) is maximal. Consider the chain of groups

$$SO(7) > Sp(6) \supset Sp(4) \times SU(2) > Sp'(4) \times U(1), \tag{5.3}$$

or alternatively the chain of subalgebras

$$B_3 > C_3 \supset C_2 + A_1 > C_2' + P \equiv B_2 + P, \tag{5.4}$$

where in the last step we have used the isomorphism of C_2 and B_2; the subjoining $Sp(4) > Sp'(4)$, or $C_2 > C_2'$, corresponds to the generating function (3.8). In order to verify the equivalence of subalgebras $B_2 + P$ and $C_2 + P$ of (5.2) and (5.4) one has to verify that, for instance, the reduction of the 7-dimensional representation of SO(7) results in the same representation in both cases. Having confirmed that, we can now combine known generating functions for $B_3 > C_3$ (cf. (4.2a)), with that for $C_3 \supset C_2 + A_1$ (it is readily found from the results of Ref. 9), the generating function (3.8) for $C_2' > C_2$, and the trivial generating function for the step $A_1 \supset P$. The composition of generating functions is done according to the prescription (3.2) of Ref. 1. Proceeding step by step along the longer chain (5.4) instead of the short one (5.1) results in a considerable economy of work. The required generating function is then:

$$[(1 - Aq^2)(1 - Aq^{-2}(1 - Bbq^2)(1 - Bbq^{-2}) \times (1 - Caq)(1 - Caq^{-1})]^{-1} \times \left[\frac{1}{(1 - Ab)(1 - C^2b)} + \frac{B}{(1 - C^2b)(1 - B)} + \frac{Ba^2}{(1 - Ab)(1 - Ba^2)} + \frac{B^2a^2}{(1 - Ba^2)(1 - B)}\right],$$

where the Dynkin labels of the B_3 representation are the exponents of A, B, and C, while power of a and b label the representations of C_2 and the power of q labels the representation of P.

(3) Tensor products of "reducible representations with minus signs" can be used for deriving the branching rules for subjoining in the same way as for inclusions. There is one subtlety which we now examine in detail on the example of $B_n > C_n$.

The decomposition of the product of two natural representations of B_n is

$$(10\cdots0) \times (10\cdots0) = (20\cdots0)_s + (010\cdots0)_a,$$

where s and a refer to the symmetric and antisymmetric parts of the tensor product. Replacing each (10···0) by the branching rule (4.3), we have

$$\begin{aligned}
&[(20\cdots0) - (010\cdots0)] \times [(20\cdots0) - (010\cdots0)] \\
&\quad = (20\cdots0) \times (20\cdots0) + (010\cdots0) \times (010\cdots0) \\
&\qquad - (20\cdots0) \times (01\cdots0) - (010\cdots0) \times (200\cdots0) \\
&\quad = [(40\cdots0)_s + (020\cdots0)_s + (010\cdots0)_s \\
&\qquad + (00\cdots0)_s + (210\cdots0)_a + (20\cdots0)_a] \\
&\qquad + [(020\cdots0)_s + (010\cdots0)_s + (00\cdots0)_s \\
&\qquad + (1010\cdots0)_a + (20\cdots0)_a] \\
&\qquad - [(210\cdots0)_s + (1010\cdots0)_s + (20\cdots0)_s + (010\cdots0)_s] \\
&\qquad - [(210\cdots0)_a + (1010\cdots0)_a + (20\cdots0)_a + (010\cdots0)_a].
\end{aligned}$$

The cancellations between representations of the same kind but appearing with opposite signs takes place *regardless* of subscript a or s. The results then are right sides of (4.3a) and (4.3b). All other branching rules can be derived in a similar fashion from the branching rules (4.3) and (4.3b).

(4) Higher indices of representations,[7] which prove to be very useful for computing complicated tensor products and branching rules[6,7] can be used also for the subjoining and related operations without any change. Indices $I^{(k)}(\Lambda_1)$ and $I^{(k)}(\Lambda_2)$ of arbitrary order k of representations (Λ_1) and (Λ_2) give an index of the difference $(\Lambda_1) - (\Lambda_2)$ as $I^{(k)}((\Lambda_1) - (\Lambda_2)) = I^{(k)}(\Lambda_1) - I^{(k)}(\Lambda_2)$, etc.

(5) The usual rules restricting branchings of selfcontragredient representations[5,6] apply also to subjoining. Thus in $G > H$ any selfcontragredient representation of G reduces to selfcontragredient representations of H. However the distinction between orthogonal and symplectic representations[5,6] disappears as can be seen for instance from the example $C_n > D_n$ above. That is, orthogonal (symplectic) representations of G may reduce to both orthogonal and symplectic representations of the hypoalgebra H.

(6) Most interesting of all the questions related to subjoining $G > H$ is undoubtedly the problem of constructing the generators of one in terms of generators of the other. It is clear that such a relation cannot be a linear one.

(7) The definition (1.3) of subjoining implies that lengths in weight space are increased (or in trivial cases remain constant) in passing from algebra weights to hypoalgebra weights. If the definition is relaxed somewhat, new subjoinings appear, in which the scale in weight space is reduced; the representations of the parent group are constrained to belong to certain congruence classes. Thus for $A_1 > A_1'$, with $P = \frac{1}{2}$, one is led to the generating function $(1 + A^2)/(1 - A^2a)$ for branching rules. The parent group tensors must belong to even representations (integer angular

momentum). Similarly for $C_2 > C_2'$, with parent group tensors restricted to the non-spinor representations, in which the first label is even, we find the generating function $(1 + B + A^2a + A^2)/(1 - Ba)(1 - A^2b)$ for branching rules. Similar subjoinings exist for the pairs $B_n > C_n$ and $C_n > B_n$.

(8) The generating functions for branching rules were found by matching dimensions and second indices for low representations. They were checked by converting them to generating functions for characters[8] and giving random numerical values to the dummy variables.

ACKNOWLEDGMENT

Two of us (J.P. and R.S.) are grateful for the hospitality extended to them at the Aspen Center for Physics, where the present work was started.

[1] J. Patera and R. T. Sharp, J. Phys. A **13,** 397 (1980).
[2] F. Lemire and J. Patera, J. Math. Phys. **21,** 2026 (1980).
[3] W. McKay, J. Patera, and D. Sankoff, in *Computers in Nonassociative Rings and Algebras,* edited by R. Beck and B. Kolman (Academic, New York, 1977).
[4] E. B. Dynkin, Am. Math. Soc. Trans. Ser. 2 **6,** 111 (1957).
[5] E. B. Dynkin, Am. Math. Soc. Trans. Ser. 2 **6,** 245 (1957).
[6] W. McKay and J. Patera, *Tables of Dimensions, Second and Fourth Indices and Branching Rules of Representations of Simple Lie Algebras* (Dekker, New York, 1980); also J. Patera and D. Sankoff, *Tables of Branching Rules for Representations of Simple Lie Algebras* (Presses de l'Université de Montréal, Montréal, Québec, 1973), p. 235.
[7] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17,** 1972 (1976); J. Math. Phys. **18,** 1579 (1977).
[8] J. Patera and R. T. Sharp, "Generating functions for characters of group representations," in *Lecture Notes in Physics Vol.* 94 (Springer, New York, 1979).
[9] R. T. Sharp, Proc. Camb. Philos. Soc. **68,** 571 (1970).

Schrödinger spectral problems with energy-dependent potentials as sources of nonlinear Hamiltonian evolution equations [a)]

L. Martínez Alonso
Departmento de Métodos Matemáticos, Facultad de Físicas, Universidad Complutense, Madrid-3, Spain

(Received 7 February 1980; accepted for publication 2 May 1980)

We develop a method to derive infinite families of completely integrable nonlinear Hamiltonian evolution equations associated with Schrödinger spectral problems whose potential functions depend on the spectral parameter.

1. INTRODUCTION

Since the discovery of the inverse scattering method by Gardner, Green, Kruskal, and Miura,[1] it has become clear that there is a wide class of linear spectral problems from which infinite families of completely integral nonlinear evolution equations may be deduced. Included in this class are the Schrödinger problem,[1] the Zakharov–Shabat problem[2] and its generalization[3] which lead to important nonlinear evolution equations of mathematical physics. An infinite subset of this class of spectral problems is provided by the matrix Schrödinger equations proposed by Wadati and Kamijo.[4] In this paper we present a new infinite subset given by the spectral problems

$$\left(-\partial_{xx}+\sum_{r=0}^{N-1}\lambda^{r}\cdot v_r(x)\right)f=\lambda^{N}\cdot f. \tag{1.1}$$

One of the most intriguing aspects of the inverse scattering method is the relationship between complete integrability and the existence of an infinite set of constants of motion. It is suspected that this empirical fact means that the statement of Liouville's theorem[5] in classical finite dimensional Hamiltonian mechanics is, in some sense, generalizable to the infinite dimensional context. Indeed, many of the evolution equations arising in the inverse scattering method have been found to be completely integrable infinite-dimensional Hamiltonian systems.[6-9] This suggests that the main task for finding the completely integrable evolution equations associated with a given spectral problem consists in deriving an infinite family of functionals which are in involution with respect to some Poisson bracket operation. In what follows, these families will be referred to as "Liouville families." A common feature of several known Liouville families is that they are generated by the coefficients of an asymptotic expansion for the function $\ln a(k)$ where $a(k)$ is the inverse of a transmission coefficient.[7,8,9] Nevertheless, there is not a clear common prescription to find the relevant Poisson bracket operation in the cases just mentioned. On the other hand, in a series of papers[10,11] Gel'fand and Dikii have developed a method for finding Liouville families which applies to an infinite class of spectral problems including the Schrödinger one. Their method is strongly based on the properties of the function $R(k,x)$, defined as the restriction to the diagonal of a resolvent kernel from which a choice for a Poisson bracket operation is suggested. But the proof of the involutiveness of the Liouville family requires a complicated procedure involving a method for constructing Lax's pairs.[12]

In this article we propose a new method for obtaining Liouville families which applies to the spectral problems (1.1) in a very simple form. It uses both functions $a(k)$ and $R(k,x)$ and its strategy is based on the following properties:

(1) The gradient of the function $\ln a(k)$ admits a simple expression in terms of $R(k,x)$.

(2) The function $R(k,x)$ satisfies a linear differential equation which leads us to deduce another one which involves the gradient of $\ln a(k)$ and two symplectic operators.

(3) This last equation implies that the coefficients of an asymptotic expansion for $\ln a(k)$ form a family of functionals with a twofold Hamiltonian structure.[13] As an immediate consequence, they generate a Liouville family.

Properties 1) and 2) are studied in Sec. 2. The asymptotic expansion for $\ln a(k)$ and the twofold Hamiltonian structure are derived in part A of Sec. 3. In the rest of Sec. 3 a recursion relation is deduced which enables one to find the explicit form of the elements of the Liouville family and their associated Hamiltonian evolution equations. As must be expected, for $N=1$ the nonlinear evolution equations obtained from our analysis are the well-known set[7] of Hamiltonian systems associated with the Schrödinger equation. The case $N=2$ is particularly intersting since it provides the Jaulent–Miodek[14] family of nonlinear evolution equations for which a transformation identifying these equations with those arising in the generalized Zakharov-Shabat spectral problem has been found.[15] Finally, Sec. 4 is devoted to investigating the evolution laws of the scattering data and it reveals that the simplicity of these laws is strongly related to the existence of a twofold Hamiltonian structure for the Liouville family.

2. TWO RELEVANT SYMPLECTIC STRUCTURES

A. The functions $a(k)$ and $R(k,x)$

Let V be a functional space whose elements are of the form

$$v(x)=\begin{pmatrix} v_0(x) \\ \cdot \\ \cdot \\ v_{N-1}(x) \end{pmatrix},$$

where $v_r:\mathbb{R}\to\mathbb{R}$ $(r=0,..,N-1)$ are regular functions such that they and their derivatives vanish sufficiently fast as $x\to\pm\infty$.[16] Every $v\in V$ defines a spectral problem of the class

[a)]Partially supported by the Junta de Energía Nuclear, Madrid.

 0022-2488/80/092342-08$1.00

(1.1), which may be written as

$$(-\partial_{xx}+\sigma(\lambda)\cdot v(x))f=\lambda^N f, \tag{2.1}$$

where the following notational conventions have been introduced

$$\sigma(\lambda)\equiv\begin{pmatrix}1\\ \lambda\\ \vdots\\ \lambda^{N-1}\end{pmatrix},\quad \sigma(\lambda)\cdot v(x)\equiv\sum_{r=0}^{N-1}\lambda^r\cdot v_r(x). \tag{2.2}$$

We define the Jost functions $f_\pm$ as the solutions of (2.1) verifying

$$k\equiv\lambda^{N/2},\quad f_+(k,x)\underset{x\to+\infty}{\sim}e^{ikx},\quad f_-\underset{x\to-\infty}{\sim}e^{-ikx}. \tag{2.3}$$

They satisfy the integral equations

$$f_+(k,x)=e^{ikx}-\int_x^\infty\frac{\sin k(x-y)}{k}\sigma(\lambda)\cdot v(y)f_+(k,y)\,dy, \tag{2.4a}$$

$$f_-(k,x)=e^{-ikx}+\int_{-\infty}^x\frac{\sin k(x-y)}{k}\sigma(\lambda)\cdot v(y)f_-(k,y)\,dy. \tag{2.4b}$$

Under the minimal condition $v_r\in L^1(\mathbb{R})$ $(r=0,\ldots,N-1)$, the method of successive approximations shows the existence and the continuity of $f_\pm$ in the upper half-plane $\operatorname{Im}k\geqslant 0$ and their analyticity in the region $\operatorname{Im}k>0$. The proof is identical to that required for the usual Schrödinger problem. In the same way, one finds that for real k there are functions $a(k)$ and $b(k)$ such that

$$f_+(k,x)=a(k)f_-^*(k,x)+b(k)f_-(k,x). \tag{2.5}$$

In addition, since the Wronskian $W(f,g)\equiv f\partial_x g-g\partial_x f$ of two solutions of (2.1) is independent of x, it follows that the function $a(k)$ can be defined as an analytic function in the upper half-plane by means of the equation

$$a(k)=(1/2ik)W(f_-(k,x),f_+(k,x)). \tag{2.6}$$

It is possible to write down the asymptotic forms of f_+ for $x\to-\infty$ and f_- for $x\to+\infty$ in terms of $a(k)$, namely

$$\operatorname{Im}k>0,\quad f_+(k,x)\underset{x\to-\infty}{\sim}a(k)e^{ikx},$$

$$f_-(k,x)\underset{x\to+\infty}{\sim}a(k)e^{-ikx}. \tag{2.7}$$

On the basis of these properties it is found that the kernel of the resolvent operator $(-\partial_{xx}+\sigma(\lambda)\cdot v(x)-\lambda^N)^{-1}$ is given by

$$R(k,x,y)=\begin{cases}\dfrac{i}{2ka(k)}f_+(k,x)f_-(k,y), & x\geqslant y\\ \dfrac{i}{2ka(k)}f_+(k,y)f_-(k,x), & y\geqslant x.\end{cases} \tag{2.8}$$

We shall denote by $R(k,x)$ the restriction to the diagonal $x=y$ of this kernel, that is

$$R(k,x)=\frac{i}{2ka(k)}f_+(k,x)f_-(k,x) \tag{2.9}$$

An elementary calculation shows that a quadratic product $f(\lambda,x)g(\lambda,x)$ of two solutions of Eq. (2.1) verifies

$$\left(-\tfrac{1}{4}\partial_{xxx}+\sum_{r=0}^{N-1}\lambda^r\cdot j(v_r)\right)(fg)=\lambda^N\partial_x(fg), \tag{2.10}$$

where $j(v_r)$ are the operators

$$j(v_r)\equiv v_r\partial_x+\tfrac{1}{2}v_{r,x}.$$

This equation will play, as we shall see, a central role in all the subsequent analysis. In particular, we observe the according to (2.9) and (2.10) we have

$$\left(-\tfrac{1}{4}\partial_{xxx}+\sum_{r=0}^{N-1}\lambda^r\cdot j(v_r)\right)R(k,x)=\lambda^N\partial_x R(k,x). \tag{2.11}$$

B. The Poisson bracket

In what follows the maps $F:V\to\mathbb{C}$ will be called functionals and they will be denoted by $F=F[v]$. Let us define in V the bilinear form

$$\langle v,w\rangle\equiv\int_{-\infty}^{\infty}v(x)\cdot w(x)\,dx,$$

$$v(x)\cdot w(x)\equiv\sum_{r=0}^{N-1}v_r(x)w_r(x).$$

A functional F is said to be differentiable at $v\in V$ if there is a unique N-component function, which will be denoted by $\nabla F(v)$, such that[17]

$$\frac{d}{d\epsilon}F[v+\epsilon w]\Big|_{\epsilon=0}=\langle\nabla F(v),w\rangle \tag{2.12}$$

for all $w\in W$. In that case $\nabla F(v)$ is called the gradient of $F\in V$. In terms of the variational derivatives $\delta F/\delta v_r(x)$, the gradient of F can be written as

$$\nabla F(v)=\begin{pmatrix}\delta F/\delta v_0(x)\\ \vdots\\ \delta F/\delta v_{N-1}(x)\end{pmatrix}.$$

It is very easy to calculate $\nabla F(v)$ when F is of the form

$$F[v]=\int_{-\infty}^{\infty}f(v_r(x),\partial_x v_r(x),\ldots,\partial_x^n v_r(x),..)\,dx, \tag{2.13}$$

where f is a polynomial depending on the functions v_r and their derivatives. Indeed, from (2.12) we have that the variational derivatives coincide with the Euler–Lagrange operators

$$\frac{\delta F}{\delta v_r(x)}=\sum_n(-1)^n\partial_x^n\frac{\partial f}{\partial(\partial_x^n v_r)}.$$

For the general case, in order to calculate gradients of functionals it is helpful to use the following formula valid to first order in the parameter ϵ

$$\delta v\equiv\epsilon w,\quad \delta F[v]\equiv F[v+\delta v]-F[v]=\langle\nabla F(v),\delta v\rangle. \tag{2.14}$$

Let us return to our spectral problem (2.1) and let us consider $a(k)$ as a functional defined on V. We prove

Proposition 2.1:

$$\nabla\ln a(k)=R(k,x)\sigma(\lambda)\ [\text{i.e.},\delta\ln a(k)/\delta v_r(x)=R(k,x)\lambda^r].$$

Proof:

By Eqs. (2.1) and (2.3) we have that under a variation δv of the potential the corresponding variation of the Jost function f_+ verifies

$$(-\partial_{xx}+\sigma(\lambda)\cdot v-\lambda^N)\delta f_+=-f_+\,\sigma(\lambda)\cdot\delta v,$$

$$\delta f_+(k,x) \underset{x\to+\infty}{\sim} 0.$$

Thus we deduce

$$\delta f_+(k,x_0) = -\int_{-\infty}^{\infty} dx\, R_+(k,x_0,x) f_+(k,x)\sigma(\lambda)\cdot\delta v(x), \tag{2.15}$$

where R_+ is the kernel

$$R_+(k,x_0,x) = \frac{i\theta(x-x_0)}{2ka(k)} \times (f_+(k,x)f_-(k,x_0) - f_+(k,x_0)f_-(k,x)). \tag{2.16}$$

Clearly, form (2.14) and (2.15) it follows that

$$\nabla f_+(k,x_0) = -R_+(k,x_0,x) f_+(k,x)\sigma(\lambda). \tag{2.17}$$

If $\operatorname{Im} k > 0$, then taking the limit $x_0 \to -\infty$ in this equation and using (2.3) and (2.7) we find

$$e^{ikx}\nabla a(k) \underset{x_0\to-\infty}{\sim} e^{ikx_0}(i/2k) f_-(k,x) f_+(k,x)\sigma(\lambda).$$

Therefore

$$\nabla \ln a(k) = \nabla a(k)/a(k) = i f_+(k,x) f_-(k,x)\sigma(\lambda)/2ka(k) = R(k,x)\sigma(\lambda)$$

for $\operatorname{Im} k > 0$. By the continuity of $a(k)$ and $R(k,x)$ the result is also true if $\operatorname{Im} k \geqslant 0$. (QED)

Given $v \in V$, let us define the operators

$$L_v \equiv \begin{pmatrix} -j(v_1) & -j(v_2) & \cdots & -j(v_{N-1}) & \partial_x \\ -j(v_2) & & & \partial_x & 0 \\ \vdots & & \iddots & & \vdots \\ -j(v_{N-1}) & \partial_x & & & \\ \partial_x & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{2.18}$$

$$M_v \equiv \left(\begin{array}{c|ccccc} -\tfrac{1}{4}\partial_{xxx} + j(v_0) & 0 & \cdots & & 0 \\ \hline 0 & -j(v_2) & \cdots & -j(v_{N-1}) & \partial_x \\ 0 & -j(v_3) & & \iddots & 0 \\ \vdots & \vdots & & & \vdots \\ & -j(v_{N-1}) & \partial_x & 0 \cdots & 0 \\ 0 & \partial_x & 0 & \cdots & \end{array}\right).$$

It follows directly from (2.11) that

$$M_v R(k,x)\sigma(\lambda) = \lambda L_v R(k,x)\sigma(\lambda),$$

which, because of Proposition 2.1, allows us to conclude that

Proposition 2.2:

$$M_v \cdot \nabla \ln a(k) = \lambda \cdot L_v \nabla \ln a(k). \tag{2.19}$$

Suppose given for every $v \in V$ a linear operator J_v acting on N-component functions and let us define for any pair of functionals $F_i (i = 1,2)$ the new functional

$$\{F_1,F_2\}[v] \equiv \langle \nabla F_1(v), J_v \cdot \nabla F_2(v)\rangle. \tag{2.20}$$

Clearly $\{,\}$ is bilinear and satisfies

$$\{F_1,F_2\cdot F_3\} = \{F_1,F_2\}\cdot F_3 + F_2\cdot\{F_1,F_3\}.$$

It is said that $\{,\}$ is a Poisson bracket if it has the following two properties

(1) $\{F_1,F_2\} = -\{F_2,F_1\}$,

(2) $\{F_1,\{F_2,F_3\}\} + \{F_2,\{F_3,F_1\}\} + \{F_3,\{F_1,F_2\}\} = 0$ (Jacobi identity).

In this case J_v is called a symplectic field of operators[13] and the pair (V,J_v) is said to be a phase space.

Proposition 2.3: L_v and M_v are symplectic fields of operators.

Proof:

Since both L_v and M_v are antisymmetric operators with respect to the bilinear form $\langle,\rangle$, then (2.20) implies that property (1) is verified by the operations $\{,\}$ associated with L_v and M_v. The proof that property (2) is also satisfied requires a nontrivial analysis and therefore it is given apart in Appendix A. (QED)

By a Liouville family in a phase space (V,J_v) we shall mean an infinite dimensional linear space of functionals of the form (2.13) such that the Poisson bracket of any two of them vanishes identically. In the next section it will be showed how Eq. (2.19) leads us to obtain a Liouville family.

3. LIOUVILLE FAMILY AND HAMILTONIAN SYSTEMS

A. Twofold Hamiltonian systems

Let J_v be a symplectic field of operators and let us consider the phase space (V,J_v). By a Hamiltonian system (V,J_v) we shall mean an evolution equation which can be written as

$$\partial_t v = J_v \cdot \nabla H(v), \tag{3.1}$$

where H is a functional of the form (2.13). In this case H is called the Hamiltonian functional of this evolution equation and one finds easily from (2.12) and (3.1) that the time evolution of an arbitrary functional $F = F[v]$ verifies

$$\partial_t F = \{F,H\}, \tag{3.2}$$

where $\{,\}$ is the Poisson bracket in (V,J_v). In particular, F will be a constant of the motion if and only if $\{F,H\} = 0$.

We are now going to derive a Liouville family in both phase spaces (V,L_v) and (VM_v).

Theorem 3.1: The function $\ln a(k)$ admits an asymptotic expansion of the form

$$\ln a(k) = (i/2\lambda)\lambda^{N/2} \sum_{n=0}^{\infty} H_n/\lambda^n, \quad \operatorname{Arg} k = \pi/2, \tag{3.3}$$

such that the coefficients H_n are functionals of the form (2.13) which verify

$$L_v \cdot \nabla H_{n+1} = M_v \cdot \nabla H_n \tag{3.4}$$

for all n.

Proof:

In Appendix B it is proved that the function $R(k,x)$, defined in (2.9), has an asymptotic expansion

$$R(k,x) = \frac{i}{2\lambda^{N/2}} \sum_{n=0}^{\infty} R_n/\lambda^n, R_0 = 1, \operatorname{Arg} k = \pi/2, \tag{3.5}$$

where the coefficients R_n are polynomials depending on v_r $(r = 0,\ldots,N-1)$ and their derivatives such that $R_n \to 0$ as $|x| \to \infty$ for $n > 0$. On the other hand, as consequence of (2.1) and (2.9) one finds

$$(\partial_\lambda(\sigma(\lambda)\cdot v - \lambda^N))R$$

$$= \frac{i}{2ka(k)} \partial_x (f_- \partial_{x\lambda} f_+ - \partial_x f_- \partial_\lambda f_+).$$

Then, taking $\mathrm{Im}\, k > 0$ and using (2.3) and (2.7), we deduce by integrating that

$$\partial_\lambda \ln a = \int_{-\infty}^{\infty} dx \left(i \frac{N}{2} \lambda^{\frac{1}{2}N-1} + \left(\sum_{r=1}^{N-1} r\lambda^{r-1} v_r(x) - N\lambda^{N-1} \right) R(k,x) \right). \quad (3.6)$$

By introducing (3.5) into (3.6) we conclude that $\ln a$ admits an asymptotic expansion of the form (3.3) with the coefficients H_n given by

$$H_n = \frac{1}{(N/2) - n - 1} \int_{-\infty}^{\infty} dx \times \left(\sum_{r=N-(n+1)}^{N-1\prime} r v_r R_{n+r+1-N} - N R_{n+1} \right) \quad (3.7)$$

where the sum in the integrand is extended to all positive r between $N-(n+1)$ and $N-1$. Finally, if we introduce (3.3) in the identity (2.19) and we identify powers of λ in the resulting equation, then (3.4) follows at once. (QED)

It is interesting and important to analyze the relation (3.4) satisfied by the functionals H_n. According to (3.4) we conclude that the evolution equations

$$\partial_t v = L_v \cdot \nabla H_{n+1} = M_v \cdot \nabla H_n \quad (3.8)$$

have a twofold Hamiltonian structure.[13] That is, they are Hamiltonian systems in both phase spaces (V, L_v) and (V, M_v). This implies the following fundamental property.

Theorem 3.2: Let $\{,\}_L$ and $\{,\}_M$ be the Poisson brackets associated with the phase spaces (V, L_v) and (V, M_v), respectively. Then, we have

$$\{H_n, H_m\}_L = \{H_n, H_m\}_M = 0 \quad (3.9)$$

for all $n, m \geqslant 0$.

Proof:

To prove this we follow a method due to F. Magri.[13] Suppose $n < m$, then

$$\begin{aligned} \{H_n, H_m\}_L &= \langle \nabla H_n, L_v \nabla H_m \rangle = \langle \nabla H_n, M_v \nabla H_{m-1} \rangle \\ &= \{H_n, H_{m-1}\}_M \\ &= -\langle M_v \nabla H_n, \nabla H_{m-1} \rangle \\ &= -\langle L_v \nabla H_{n+1}, \nabla H_{m-1} \rangle \\ &= \{H_{n+1}, H_{m-1}\}_L. \end{aligned}$$

Therefore, by iterating we find

$$\begin{aligned} \{H_n, H_m\}_L &= \{H_n, H_{m-1}\}_M = \{H_{n+2}, H_{m-2}\}_L \\ &= \cdots = \{H_m, H_n\}_L, \end{aligned}$$

which, from the skew symmetry of the Poisson bracket, implies (3.9). (QED)

Let us denote by $\mathcal{F}$ the linear space of functionals generated by the family $\{H_n, n \geqslant 0\}$

$$\mathcal{F} \equiv \left\{ \sum_{m=0}^{M} c_m H_m : c_m \in \mathbb{R}, M \geqslant 0 \right\}. \quad (3.10)$$

As a consequence of (3.9) we conclude that $\mathcal{F}$ is a Liouville family. Moreover, from (3.2) we have

Theorem 3.3: All the elements of $\mathcal{F}$ are constants of motion for each Hamiltonian system $\partial_t v = L_v \cdot \nabla H$ with Hamiltonian $H \in \mathcal{F}$.

In this way, by means of the properties of $\ln a(k)$ and $R(k,x)$, we have been able to deduce an infinite family of Hamiltonian systems with a common infinite set of constants of motion.

B. Recursion relations and evolution equations

The equation (3.7) provides us an expression of the functionals H_n in terms of the coefficients R_m of the asymptotic expansion of $R(k,x)$. But the method used to prove (3.5) is not appropriate as a practical rule to find the explicit form of such coefficients. Nevertheless, we are able to do it by means of a consequence of (2.11). Indeed, substituting (3.5) into (2.11) and identifying equal powers of λ, we obtain the recusion relation

$$\begin{aligned} &R_0 = 1, \partial_x R_n = \sum_{r=1}^{n} j(v_{N-r}) R_{n-r}, n = 1, \ldots, N-1, \\ &\partial_x R_{N+n} = (-\tfrac{1}{4}\partial_{xxx} + j(v_0)) R_n + \sum_{r=1}^{N-1} j(v_r) R_{n+r}, n \geqslant 0. \end{aligned} \quad (3.11)$$

It allows us to calculate the functionals H_n by using (3.7).

In order to get the explicit form of the Hamiltonian systems defined by the elements of $\mathcal{F}$ it is not necessary to know the functionals H_n. To see it, let us note that inserting the asymptotic expansion (3.3) and (3.5) into the equation $\nabla \ln a(k) = R(k,x)\sigma(\lambda)$ we obtain

$$\begin{aligned} \nabla H_n &= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ R_0 \\ \vdots \\ R_n \end{pmatrix} \quad n = 0, 1, \ldots, N-1, \\ \nabla H_{N+n} &= \begin{pmatrix} R_{n+1} \\ R_{n+2} \\ \vdots \\ R_{n+N} \end{pmatrix} \quad n \geqslant 0. \end{aligned} \quad (3.12)$$

In particular, from (2.18) and (3.11) we deduce

$$L_v \nabla H_0 = L_v \nabla H_1 = \cdots = L_v \nabla H_{N-1} = 0 \quad (3.13)$$

and

$$L_v \cdot \nabla H_N = \tfrac{1}{2} v_x = \tfrac{1}{2} \begin{pmatrix} v_{0,x} \\ \vdots \\ v_{N-1,x} \end{pmatrix}. \quad (3.14)$$

At this point we notice that, in view of the polynomial nature of the coefficients R_n and the form of L_v, the components of ∇H_n and $L_v \nabla H_n$ are elements of the polynomial ring A generated by the variables $\partial_x^n v_r (n \geqslant 0, r = 0, 1, \ldots, N-1)$. In the subset $\mathrm{Ran}\partial_x$ of A it is possible to give a convenient meaning to the operator ∂_x^{-1}. Ideed, given $f \in \mathrm{Ran}\partial_x$ there is a unique $g \in A$ verifying $f = \partial_x g$ and such that g vanishes when all the variables $\partial_x^n v_r$ $(n \geqslant 0, r, 0, \ldots, N-1)$ are set up equal to zero. This is what we mean by ∂_x^{-1} in the following. Consider the operator

$$K_v \equiv \left(\begin{array}{c|c} 0\cdots 0 & (-\tfrac{1}{4}\partial_{xxx} + j(v_0))\partial_x^{-1} \\ \hline & j(v_1)\partial_x^{-1} \\ \mathbf{1}_{N-1} & j(v_2)\partial_x^{-1} \\ & \vdots \\ & j(V_{N-1})\partial_x^{-1} \end{array}\right). \tag{3.15}$$

It is elementary to verify that

$$M_v = K_v \cdot L_v. \tag{3.16}$$

Then (3.4) and (3.14) lead to

$$\begin{aligned} L_v \cdot \nabla H_{N+n} &= M_v \cdot \nabla H_{N+n-1} = K_v (L_v \cdot \nabla H_{N+n-1}) \\ &= K_v^2 L_v \nabla H_{N+n-2} = \cdots = K_v^n L_v \cdot \nabla H_N \\ &= \tfrac{1}{2} K_v^n \partial_x v, \quad n \geqslant 0. \end{aligned} \tag{3.17}$$

Therefore, Eqs. (3.13) and (3.17) enable us to conclude that the nontrivial Hamiltonian systems generated by the family of functionals $\{H_m, m \geqslant 0\}$ are given by

$$\begin{aligned} \partial_t v &= L_v \cdot \nabla \left(\sum_{n=0}^{M} c_n H_{N+n} \right) = M_v \cdot \nabla \left(\sum_{n=0}^{M} c_n H_{N+n-1} \right) \\ &= \tfrac{1}{2} \Omega (K_v) \partial_x v, \end{aligned} \tag{3.18}$$

where Ω denotes the polynomial function $\Omega(\lambda) = \Sigma_{n=0}^{M} c_n \lambda^n$.

For instance, the choice $\Omega(\lambda) = 2\lambda$ yields the evolution equation

$$\left\{\begin{array}{l} \partial_t v_0 = -\tfrac{1}{4} v_{N-1,xxx} + v_0 v_{N-1,x} + \tfrac{1}{2} v_{N-1} v_{0,x}, \\ \partial_t v_r = v_{r-1,x} + v_r v_{N-1,x} + \tfrac{1}{2} v_{N-1} v_{r,x} \\ \qquad (r = 1,..,N-1), \\ \partial_t v_{N-1} = v_{N-1,x} + \tfrac{3}{2} v_{N-1} v_{N-1,x}. \end{array}\right. \tag{3.19}$$

We end this section with the application of our analysis to the cases $N = 1,2$.

(i) Consider $N = 1$; then the spectral problem (2.1) reduces to the Schrödinger eigenvalue problem

$$(-\partial_{xx} + v(x)) f = \lambda f, \tag{3.20}$$

with an scalar potential v. According to (3.18), the associated evolution equations are given by

$$\partial_t v = \tfrac{1}{2} \Omega (K_v) v_x, \quad K_v = (-\tfrac{1}{4}\partial_{xxx} + v\partial_x + \tfrac{1}{2} v_x)\partial_x^{-1}. \tag{3.21}$$

They form the well-known family of nonlinear equations related to the Schrödinger equation.[18] In particular, for the choice $\Omega(\lambda) = 8\lambda$ we obtain

$$\partial_t v = -v_{xxx} + 6 v v_x,$$

which is the KdV equation. From (2.18) we have that the symplectic fields which enable us to write this family in Hamiltonian form are

$$L_v = \partial_x, \quad M_v = -\tfrac{1}{4}\partial_{xxx} + v\partial_x + \tfrac{1}{2} v_x. \tag{3.22}$$

To find the infinite set of constants of motion $\{H_n, n \geqslant 0\}$ we note that (3.7) and (3.11) become

$$H_n = \frac{2}{2n+1} \int_{-\infty}^{\infty} R_{n+1}\, dx. \tag{3.23}$$

and

$$R_0 = 1, \quad \partial_x R_{n+1} = (-\tfrac{1}{4}\partial_{xxx} + v\partial_x + \tfrac{1}{2} v_x) R_n, \tag{3.24}$$

respectively. Thus, the first few R_n are found to be

$$R_1 = \tfrac{1}{2} v, \quad R_2 = \tfrac{1}{8}(-v_{xx} + 3v^2),$$

$$R_3 = \tfrac{1}{32}(\partial_x^4 v - 10 v v_{xx} - 5 v_x^2 + 10 v^3)$$

and they lead us to

$$H_0 = \int_{-\infty}^{\infty} v\, dx, \quad H_1 = \int_{-\infty}^{\infty} \tfrac{1}{4} v^2\, dx,$$

$$H_2 = \int_{-\infty}^{\infty} \tfrac{1}{16}(v_x^2 + 2v^3)\, dx.$$

(ii) We now take $N = 2$, then (2.1) reduces to

$$(-\partial_{xx} + v_0(x) + \lambda v_1(x)) f = \lambda^2 f. \tag{3.25}$$

This spectral problem has been extensively studied by Jaulent and Jean.[19] In the radial case[20] ($x \geqslant 0$) it applies to scattering problems in absorbing media.[21] From (3.18) we see that its associated evolution equations are

$$\partial_t v = \tfrac{1}{2} \Omega (K_v) \partial_x v, \quad K_v = \left(\begin{array}{c|c} 0 & (-\tfrac{1}{4}\partial_{xxx} + j(v_0))\partial_x^{-1} \\ \hline 1 & j(v_1)\partial_x^{-1} \end{array}\right). \tag{3.26}$$

They coincide with the family of nonlinear equations found by Jaulent and Miodek[14] by applying techniques of inverse scattering transform to (3.25). This is a very important family of evolution equations since, as has been shown by Jaulent and Miodek,[15] there is a map $v \to \bar{v}$ which transforms these equations into the family of evolution equations associated with the generalized Zakharov–Shabat spectral problem

$$\begin{pmatrix} \partial_x & -\bar{v}_0 \\ -\bar{v}_1 & \partial_x \end{pmatrix} f = -i\lambda \sigma_3 f, \quad f = \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}. \tag{3.27}$$

As is well known,[3] this family includes several important equations such as the modified KdV, the Sine-Gordon and the nonlinear Schrödinger ones.

As a consequence of our analysis we conclude that the Jaulent–Miodek equations are Hamiltonian systems with respect to the following symplectic fields:

$$\begin{aligned} L_v &= \begin{pmatrix} -j(v_1) & \partial_x \\ \partial_x & 0 \end{pmatrix}, \\ M_v &= \begin{pmatrix} -\tfrac{1}{4}\partial_{xxx} + j(v_0) & 0 \\ 0 & \partial_x \end{pmatrix}. \end{aligned} \tag{3.28}$$

It is worth mentioning that starting from this Hamiltonian formulation one can prove[22] that the Jaulent–Miodek transformation is a canonical map.

By (3.7) and (3.11) it follows that

$$H_{n+1} = \frac{1}{n+1} \int_{-\infty}^{\infty} (2R_{n+2} - v_1 R_{n+1})\, dx, \tag{3.29}$$

where

$$\begin{aligned} R_1 &= \tfrac{1}{2} v_1, \\ \partial_x R_{n+2} &= (-\tfrac{1}{4}\partial_{xxx} + j(v_0)) R_n + j(v_1) R_{n+1}. \end{aligned} \tag{3.30}$$

Thus, the first few H_n are given by

$$H_1 = \int_{-\infty}^{\infty} (\tfrac{1}{4} v_1^2 + v_0)\, dx, \quad H_2 = \int_{-\infty}^{\infty} (\tfrac{1}{8} v_1^3 + \tfrac{1}{2} v_0 v_1)\, dx,$$

$$H_3 = \int_{-\infty}^{\infty} (-\tfrac{1}{6} v_1 v_{1xx} - \tfrac{5}{48} v_{1x}^2 + \tfrac{5}{64} v_1^4 + \tfrac{3}{8} v_1^2 v_0 + \tfrac{1}{4} v_0^2)\, dx.$$

4. SCATTERING DATA AND COMPLETE INTEGRABILITY

Given a Hamiltonian system with a finite number n of degrees of freedom, if we know n constants of motion H_i $(i = 1,...,n)$ such that $\{H_i,H_j\} = 0$ for all i,j, then we can find a set of coordinates in the phase space for which Hamilton equations take a trivially integrable form. This is essentially the content of Liouville's Theorem in classical Hamiltonian mechanics.[5] There are reasons to think that this result is also true, in some sense, for infinite-dimensional Hamiltonian system having an infinite number of functionals such that the Poisson bracket of any two of them vanishes identically. In this later context, the role of the set of coordinates in phase space for which Hamilton equations are trivially integrable seems to be played by the scattering data of the associated spectral probelm, when it exists. The task of finding a set of scattering data from which the potential functions of spectral problems can be univocally determined is called the inverse scattering problem. The solution of this problem is well known for the cases $N = 1,2$ of Eq. (2.1)[19,23] and it can be extended to cover the general case.[24]

In this section we shall show how the methods of Hamiltonian formalism may be applied in order to find the evolution law of the scattering data of Eq. (2.1) under the evolution equations (3.18). We look for the set of scattering data $(k_l,b_l,a(k),b(k))$ where $k_l (l = 1,2,...)$ denote the zeros of the function $a(k)$ in the region $\mathrm{Im}\, k > 0$, the parameters b_l are defined by $f_+(k_l,x) = b_l f_-(k_l,x)$ and the function $b(k)$ was defined in (2.5).

Proposition 4.1: The following equations are satisfied:

$$M_v \nabla k_l = \lambda_l L_v \nabla k_l, \tag{4.1}$$

$$\langle M_v \nabla f_+(k,x_0),w\rangle = \lambda \langle L_v \nabla f_+(k,x_0),w\rangle + (l_+ w)(x_0), \tag{4.2}$$

where l_+ is the operator

$$l_+ \equiv \tfrac{1}{4} f_+(k,x)\partial_x - \tfrac{1}{2}\partial_x f_+(k,x). \tag{4.3}$$

Proof:

Let k_l be a zero of $a(k)$ such that $\mathrm{Im} k_l > 0$, then from (2.7) we have that $f_+(k_l,x)$ and $f_-(k_l,x)$ must be proportional and therefore, by the boundary conditions (2.3), we deduce that both functions decrease exponentially as $|x| \to \infty$. Moreover, we have

$$(-\partial_{xx} + \sigma(\lambda_l)\cdot v(x)) f_+(k_l,x) = \lambda_l^N f_+(k_l,x), \tag{4.4}$$

where $\lambda_l \equiv k_l^{2/N} (0 < \mathrm{Arg}\lambda_l < 2\pi/n)$. Then, under a variation δv of the potential, we deduce

$$(-\partial_{xx} + \sigma\cdot v)\delta f_+ + (\partial_\lambda \sigma\cdot v \delta\lambda_l + \sigma\cdot\delta v) f_+ = \lambda_l^N \delta f_+ + N\lambda_l^{N-1}\delta\lambda_l \cdot f_+ . \tag{4.5}$$

Multiplying by $f_+(k_l,x)$ and integrating, it implies that

$$\delta\lambda_l = C_l \int_{-\infty}^{\infty} f_+(k_l,x)^2 \sigma(\lambda_l)\cdot\delta v(x)\, dx,$$

where

$$C_l \equiv \left(\int_{-\infty}^{\infty} (N\lambda_l^{N-1} - \partial_\lambda \sigma(\lambda_l)\cdot v(x)) f_+(k_l,x)^2 dx \right)^{-1}.$$

Therefore

$$\nabla k_l = \tfrac{1}{2} N \lambda_l^{N/2-1} \nabla\lambda_l = (\tfrac{1}{2} N \lambda_l^{N/2-1} c_l) f_+(k_l,x)^2 \sigma(\lambda_l), \tag{4.6}$$

and from Eq. (2.10) we deduce (4.1) at once. In what concerns (4.2), let us observe the form (2.17) of ∇f_+ and the equation satisfied by R_+

$$(-\partial_{xx} + \sigma(\lambda)\cdot v(x) - \lambda^N) R_+(k,x_0,x) = \delta(x - x_0).$$

This equation and (2.1) imply that the product $R_+(k,x_0,x) f_+(k,x)$ verifies

$$\left(-\tfrac{1}{4}\partial_{xxx} + \sum_{r=0}^{N-1} \lambda^r j(v_r) \right) R_+ f_+ = \lambda^N \partial_x (R_+ f_+) + \tfrac{1}{4}(3\partial_x f_+ + f_+ \partial_x)\delta(x - x_0). \tag{4.7}$$

The rest of the proof is an immediate consequence of (4.7) and (2.17). QED

Theorem 4.1: If $v = v(x)$ evolves in time according to Eq. (3.18), then the evolution law of the scattering data is given by

$$k_l(t) = k_l(0), \tag{4.8a}$$

$$b_l(t) = b_l(0)\exp(-ik_l \Omega(\lambda_l)t), \tag{4.8b}$$

$$a(k,t) = a(k,0), \tag{4.8c}$$

$$b(k,t) = b(k,0)\exp(-ik\Omega(\lambda)t). \tag{4.8d}$$

Proof: Let us begin by calculating $k_l(t)$; from Eqs. (3.4), (3.13), and (4.1) we have

$$\begin{aligned} \{k_l,H_{N+n}\}_L &= \langle \nabla k_l, L_v \nabla H_{N+n}\rangle = \langle \nabla k_l, M_v \nabla H_{N+n-1}\rangle \\ &= -\langle M_v \nabla k_l, \nabla H_{N+n-1}\rangle = \lambda_l \langle \nabla k_l, L_v \nabla H_{N+n-1}\rangle \\ &= \lambda_l \{k_l,H_{N+n-1}\}_L = \cdots = \lambda_l^{n+1}\{k_l,H_{N-1}\}_L \\ &= \lambda_l^{n+1}\langle \nabla k_l, L_v \nabla H_{N-1}\rangle = \lambda_l^{n+1}\langle \nabla k_l, 0\rangle = 0. \end{aligned} \tag{4.9}$$

Then using (3.2) we deduce (4.8a). In the same form, with the help of (3.4), (3.12), and (4.2) it follows that

$$\begin{aligned} \{f_+(k,x),H_{N+n}\}_L &= \lambda \{f_+(k,x),H_{N+n-1}\}_L - l_+ R_n \\ &= \cdots = \lambda^n \{f_+(k,x),H_N\}_L - l_+ \sum_{r=0}^{n-1} \lambda^r R_{n-r}. \end{aligned} \tag{4.10}$$

On the other hand, (2.17), (3.14), and (4.7) imply

$$\begin{aligned} &\{f_+(k,x),H_N\} \\ &= -\tfrac{1}{2}\int_{-\infty}^{\infty} dy\, R_+(k,x,y) f_+(k,y)\sigma(\lambda)\cdot\partial_y v(y) \\ &= \lambda^N \int_{-\infty}^{\infty} dy\, \partial_y (R_+(k,x,y) f_+(k,y)) + \tfrac{1}{2}\partial_x f_+(k,x) \\ &= \tfrac{1}{2}(\partial_x f_+(k,x) - ikf_+(k,x)). \end{aligned} \tag{4.11}$$

Then, since $R_m \to 0$ $(m > 0)$ as $x \to -\infty$, we have

$$\{f_+(k,x),H_{N+n}\} \underset{x\to-\infty}{\sim} \tfrac{1}{2}\lambda^n(\partial_x f_+(k,x) - ikf_+(k,x)). \tag{4.12}$$

Now if we take we take $k = k_l$ and we introduce into (4.12) the asymptotic behavior

$$f_+(k_l,x) = b_l f_-(k_l,x) \underset{x\to-\infty}{\sim} b_l e^{-k_l x},$$

then it follows that

$$\{b_l,H_{N+n}\} = -ik_l \cdot \lambda_l^n b_l. \tag{4.13}$$

In the same form, taking $k\in\mathbb{R}$ and using

$$f_+(k,x) \underset{x\to-\infty}{\sim} a(k)e^{ikx} + b(k)e^{-ikx},$$

Eq. (4.12) yields

$$\{a(k),H_{N+n}\} = 0, \quad \{b(k),H_{N+n}\} = -ik\lambda^n b(k). \tag{4.14}$$

From (4.13), (4.14), and (3.2) the rest of the proof follows at once. QED

APPENDIX A

In this Appendix we shall complete the proof of Proposition 2.3 by showing that the operations $\{,\}$ associated with L_v and M_v satisfy the Jacobi identity. In terms of the operator J_v of (2.20) this identity can be written as[13]

$$\langle\varphi,J_v'(\psi,J_v\xi)\rangle + \langle\psi,J_v'(\xi,J_v\varphi)\rangle + \langle\xi,J_v'(\varphi,J_v\psi)\rangle = 0 \tag{A1}$$

for all $\varphi,\psi,\xi\in V$, where J_v' denotes the bilinear operator

$$J_v'(,):V\times V\to V, \quad J_v'(\varphi,\psi) = \frac{d}{d\epsilon}J_{v+\epsilon\psi}\varphi\Big|_{\epsilon=0}. \tag{A2}$$

Clearly, in order to satisfy (A1) it is sufficient to prove that for all $\varphi,\psi,\xi\in V$ the function

$$(\varphi|\psi|\xi)_{J_v} \equiv \varphi\cdot J_v'(\psi,J_v\xi) + \psi\cdot J_v'(\xi,J_v\varphi) + \xi\cdot J_v'(\varphi,J_v\psi) \tag{A3}$$

is the derivative with respect to x of a function which vanishes at infinity.

From (2.18) we observe that M_v is a direct sum of two operators; the first one, given by $-\frac{1}{4}\partial_{xxx} + j(v_0)$ is well known as a symplectic operator in the analysis of the KdV equation[13] and the second one is of the form $L_{\bar v}$ with $\bar v$ being an $(N-1)$-component function. Therefore, all we have to prove is that L_v verifies (A3). To do it, we start by calculating L_v', which an immediate application of (A2) shows to be

$$L'(\varphi,\psi) = -N_\phi, \quad N_{\psi\varphi} \equiv \begin{pmatrix} j(\psi_1) & j(\psi_2) & \cdots & j(\psi_{N-1}) & 0 \\ j(\psi_2) & & & & 0 \\ \vdots & & & & \vdots \\ j(\psi_{N-1}) & & & & \vdots \\ 0 & 0 & \cdots & & 0 \end{pmatrix}. \tag{A4}$$

Then, by means of a straightforward computation one finds

$$\begin{aligned}\varphi\cdot L_v'(\psi,L_v\xi) = &\sum_{r+s+q<N-1}\varphi_r[v_{r+s+q+2}(\psi_{s,x}\xi_{q,x} + \tfrac{1}{2}\psi_s\xi_{q,xx}) \\ &+ \tfrac{1}{2}v_{r+s+q+2,x}(\psi_{s,x}\xi_{q,x} + \tfrac{3}{2}\psi_s\xi_{q,x}) \\ &+ \tfrac{1}{4}v_{r+s+q+2,xx}\psi_s\xi_q],\end{aligned} \tag{A5}$$

where we denote $v_N \equiv -1$. This leads us to the expression

$$(\varphi|\psi|\xi)_{L_v} = \partial_x\Big[\sum_{r+s+q<N-1}(\tfrac{1}{2}v_{r+s+q+2}\partial_x + \tfrac{3}{4}v_{r+s+q+2,x})\varphi_r\psi_s\xi_q\Big], \tag{A6}$$

which proves that L_v verifies (A1).

APPENDIX B

We give here the derivation of the asymptotic expansion (3.5) for the function $R(k,x)$. Our proof will be based on a method applied to the usual Schrödinger problem by Gel'fand and Dikii.[10] We begin by considering the Born asymptotic series for the kernel $R(k,x,y)$ of the resolvent operator of (2.1). Its restriction to the diagonal $x=y$ gives

$$R(k,x) = \sum_{M=0}^{\infty} R_M(k,x) \tag{B1}$$

$$R_M(k,x) = \frac{i}{2k}\Big(\frac{-i}{2k}\Big)^M\int\exp[ik(|x-x_1| + |x_1-x_2| + \cdots + |x_M-x|)]\prod_{l=1}^{M}\sigma(\lambda)\cdot v(x_l)\,dx_l. \tag{B2}$$

Then, we have

$$R_M(k,x) = \frac{i}{2k}\Big(\frac{-i}{2k}\Big)^M\sum_{r_1,\ldots,r_M=0}^{N-1}\lambda^{r_1+\cdots+r_M}I(k)_{r_1,\ldots,r_M}, \tag{B3}$$

where

$$I(k)_{r_1,\ldots,r_M} = \int\exp[ik(|x-x_1| + \cdots + |x_M-x|)]\times\prod_{l=1}^{M}v_{r_l}(x_l)\,dx_l. \tag{B4}$$

Let us take k such that

$$\operatorname{Arg}k = \pi/2 \tag{B5}$$

Then, by introducing the charge of variables

$$(x_l)\to(\eta_l), \quad \eta_l = -ik(x_l - x), \tag{B6}$$

and the Taylor series

$$v_{r_l}(x_l) = \sum_{n_l=0}^{\infty}\frac{1}{n_l!}v_{r_l,n_l}(x)(x_l-x)^{n_l}, \tag{B7}$$

the integrals (B4) take the asymptotic form

$$I(k)_{r_1,\ldots,r_M} = \Big(\frac{i}{k}\Big)^M\sum_{r_1,\ldots,n_M=0}^{\infty}\Big(\frac{i}{k}\Big)^{n_1+\cdots+n_M}c_{n_1,\ldots,n_M}\times\prod_{l=1}^{M}v_{r_l,n_{r_l}}(x), \tag{B8}$$

where

$$c_{n_1,\ldots,n_M} \equiv \int\exp[-(|\eta_1| + |\eta_2-\eta_1| + \ldots + |\eta_M|)]\times\prod_{l=1}^{M}(n_l!)^{-1}\eta_l^{n_l}\,d\eta_l. \tag{B9}$$

From (B3) and (B8) we have

$$R_M(k,x) = \frac{i}{2\lambda^{N/2}}\frac{1}{(2\lambda^N)^M}\times\sum_{n_1,\ldots,n_M=0}^{\infty}\sum_{r_1,\ldots,r_M=0}^{N-1}\lambda^{r_1+\cdots+r_M}\times\Big(\frac{i}{\lambda^{N/2}}\Big)^{n_1+\cdots+n_M}c_{n_1,\ldots,n_M}\prod_{l=1}^{M}v_{r_l,n_l}(x). \tag{B10}$$

By making $\eta_r\to-\eta_r$ in (B9) it follows that

$$c_{n_1,\ldots,n_M} = (-1)^{n_1,\ldots,n_M}c_{n_1,\ldots,n_M}. \tag{B11}$$

Therefore, only the terms for which $n_1 + \ldots + n_M$ is even contribute to (B10). So we obtain an expansion for $R_M(k,x)$ of the form

$$R_M(k,x) = \frac{i}{2\lambda^{N/2}} \sum_{n=0}^{\infty} P_n/\lambda^{M+n}, \tag{B12}$$

where the coefficients P_n are polynomials depending on the variables v_{r_i,n_i}, which vanish when these variables are set equal to zero. In this way, from (B1) and (B12) the asymptotic expansion (3.5) follows.

ACKNOWLEDGMENTS

The author is greatly indebted to Professor M. Jaulent for many halpful discussions on the subject of this paper. He is also grateful to Professor P.C. Sabatier for his hospitality at the University of Montpellier, where this work was done.

[1]C. S. Gardner, J. M. Green, M. D. Kruskal and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[2]V. E. Zakharov and A. B. Shabat, Zh. Eks. Teor. Fiz. **61**, 118 (1971).
[3]M. J. Ablowitz, D. J. Kamp, A. C. Newell and H. Segur, Phys. Rev. Lett. **31**, 125 (1973); and Stud. Appl. Math. **53**, 294 (1974).
[4]M. Wadati and T. Kamijo, Prog. Theor. Phys. **52**, 397 (1974).
[5]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1978), p. 401.
[6]C. Gardner, J. Math. Phys. **12**, 1548 (1971).
[7]L. Faddeev and V. Zakharov, Funk. Anal. Priloz. **5**, 18 (1971).
[8]V. Zakharov and S. Manakov, Teor. Math. Fiz. **19**, 332 (1974).
[9]H. Flashka and A. C. Newell, "Integrable systems of nonlinear evolution equations", edited by J. Moser in *Lecture Notes in Physics*, Vol. 38 (Berlin, 1975).
[10]I. M. Gel'fand and L. A. Dikii, Usp. Mat. Mat. Nauk **30**, 67 (1975).
[11]I. M. Gel'fand and L. A. Dikii, Funk. Anal. Priloz. **10**, No. 1, 18 (1976); **10**, No. 4, 13 (1976); **11**, No. 2, 11 (1977); **12**, No. 2, 8 (1978).
[12]P. D. Lax, Comm. Pure Appl. Math. **21**, 467 (1968).
[13]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[14]M. Jaulent and I. Miodek, Lett. Math. Phys. **1**, 243 (1976).
[15]M. Jaulent and I. Miodek, Lett. Nuovo Cimento, **20**, 655 (1977).
[16]For instance, the functions v_r may be assumed to be Schwartz functions on **R**.
[17]If the functional space V admits a dual space V^* then we may require $\nabla F(v) \in V^*$. In that case $\nabla F(v)$ may be a distribution.
[18]C. S. Gardner, J. M. Green, and M. D. Kruskal, Comm. Pure Appl. Math. **27**, 97 (1974).
[19]M. Jaulent and C. Jan, Ann. Hist. Henri Poincaré **25**, 105 (1976).
[20]M. Jaulent and C. Jean, Comm. Math. Phys. **28**, 177 (1972).
[21]M. Jaulent, J. Math. Phys. **17**, 1351 (1976).
[22]F. Guil and L. M. Alonso, Lett. Nuovo Cimento **27**, 85 (1980).
[23]L. D. Faddeev, Tr. Mat. V. A. Steklova **73**, 314 (1964).
[24]M. Jaulent, private communication, in preparation.

Gauge invariance and the Helmholtz conditions

Robert Cawley
Naval Surface Weapons Center White Oak, Silver Spring, Maryland 20910

(Received 22 October 1979; accepted for publication 6 June 1980)

In a previous work I have outlined a new formalism of theoretical mechanics, extending the traditional subject in a new way to elementary systems whose second-order dynamical equations do not satisfy the classical Helmholtz conditions for the existence of a Lagrange function. In the present paper I determine the role of a new gauge invariance ("dynamical" gauge invariance), satisfied by the equations of motion of systems covered by the formalism, and the relationship this bears to the Helmholtz conditions. A certain subgroup, of "kinematic" gauge transformations, is singled out: the kinematic gauge transformations correspond for the Lorentz force law to the usual gauge transformation of electromagnetism, general dynamical gauge transformations correspond to minimal substitutions. An analogue for the nonalbelian gauge force law is discussed briefly. An implication of the present work is the result, in a sense made explicit within, that (the usual) gauge invariance is an *intrinsic property* of classical Hamiltonian systems.

1. INTRODUCTION

Gauge theories enjoy wide application and success as models for elementary particle and radiation processes; and, from the standpoint of quantum field theory, their renormalizability makes them especially attractive. Equally, from the same standpoint, it appears either strange or at least arbitrary that nature should grant such serendipity as to make these so successful for us, and ubiquitous, as they have been. On the other hand, an old discovery of Jauch (see below) suggests another viewpoint, that gauge invariance, in some sense, may be "implicit" already elsewhere in our basic theoretical formalism. In the present paper I present some developments of classical theoretical mechanics which suggest the same sort of thing.

In a previous work[1] I have presented outlines of a generalized classical Hamilton-Lagrange mechanics, for application to systems whose second-order equations of motion need not satisfy the Helmholtz conditions for existence of a Lagrangian. Systems for the nonHelmholtz case are referred to more commonly as nonHamiltonian, and typically they are open systems; some simple examples are a damped oscillator, the multivibrator (van der Pol equation), and a radiating charge.

There are other approaches to Lagrange-Hamilton formulations for equations of open systems. The best known, but which suffers from difficulties of physical interpretation that stem from the failure of the Hamiltonian to admit an interpretation as system energy, uses integrating factors to the equations of motion.[2,3] Another proposal is based upon a nonLie algebraic (Lie admissible) time evolution law for the Hamiltonian variables.[4] And E.H Kerner has introduced a "canonicalization" of first order systems for biological applications.[5] The methods introduced by H. Dekker[6] appear to bear the closest relationship to mine, however, despite apparent dissimilarities, and differences of motivation. Further references, more detailed physical discussion, and my motivation are given in Ref. 1. Nevertheless, I will add a few further remarks here to indicate some aspects of the direction of my efforts.

As the physically useful form of theoretical mechanics has been restricted to closed systems, elementary (i.e. "intrinsically macroscopic") system representation of nonequilibrium processes has been able to draw only indirectly or not at all from methods grounded in Hamiltonian theory. Some of the kinds of problems of physical interest for which this shortcoming seems to me likely to be limiting to our capabilities for theoretical modeling, and our basic understanding, are collective nonequilibrium processes in plasmas, macroscopic dissipation phenomena in certain nuclear processes (fission, and heavy ion scattering at high energies), and "dressing" (renormalization) effects in radiation and fundamental particle processes.

Another kind of problem to which the present formalism of mechanics appears likely to find useful application is in the construction of guiding center models for plasma applications, where Hamiltonian formulations are desirable, along with proper physical interpretation.[7]

The proposed generalization for theoretical mechanics applies presently to systems of equations having the form $f_{2n}(x,\dot{x},t) - f_{1n}(x,\dot{x},t) = 0$, $n = 1$ to N [Eqs. (2.8)], subject only to the restriction given by [(2.11)]. I plan later to extend it to cover more general forms, such as $G_n(x,\dot{x},\ddot{x},t) = 0$, with the G_n arbitrary. In this paper I continue the development of the present scheme, and as a by-product obtain general results concerning gauge invariance in classical mechanics. Although the main purpose here is to carry forward the development of the new formalism by securing its relationship to the old, I wish also to draw attention to the rather general character of the role played by gauge transformations in classical theory, not obvious heretofore but which emerges naturally from the new scheme. In addition, the classical analogue of Jauch's result, referred to in the first paragraph, is immediate in the present context, and is set into a wider, and changed perspective.

Jauch showed[8] that Galilei invariance in nonrelativistic quantum mechanics, implemented as a symmetry in the standard way, by projective representations, ψ, of the inhomogeneous Galilei group assumed to satisfy a Schrödinger

 0022-2488/80/092350-10$1.00

equation, uniquely fixes the Hamiltonian of an elementary system to the form

$$H = \frac{1}{2\mu}(\mathbf{p} - \mathbf{A})^2 + V, \tag{1.1}$$

where $\mathbf{A}$ and V are arbitrary functions of the position and the time, μ is the mass (parameter for the projective representation) and $\mathbf{p}$ is the momentum operator. As Eq. (1.1) is invariant to the phase transformation of ψ,

$$\psi \rightarrow \exp(i\phi(\mathbf{x},t))\cdot\psi, \tag{1.2}$$

if also we transform $\mathbf{A}$ and V,

$$\mathbf{A} \rightarrow \mathbf{A} + \nabla\phi, \quad V \rightarrow V - \partial\phi/\partial t, \tag{1.3}$$

Jauch concluded that Galilei invariance may be said to imply gauge invariance—a result having nothing to do with electromagnetism.

In the example of electromagnetism it is not hard to show that the effect of adding the time derivative of a function of the position and the time to the Lagrangian is equivalent to a guage transformation. This result cannot be further generalized without some more general notion of gauge transformations for other classes of examples. On the other hand, the Helmholtz conditions assure the general form, Eq. (2.8), for the equations of motion—for this is the form of the Euler–Lagrange equations—and as we shall see this form is invariant to a large group of transformations (dynamical gauge group) containing the usual gauge transformations as a subgroup (kinematic, or integrable subgroup). Now the formalism introduced in Ref. 1 extends, except for the mild restriction (2.11) noted earlier, to any system of the form, Eqs. (2.8), including, of course, the subclass for the Helmholtz case. The integrable (gauge) subgroup then is realized as an invariance group for the equations of motion whenever a Lagrangian exists in the sense of Helmholtz. In addition, the above result for the Lorentz force of electromagnetism, of the equivalence of gauge transformations to the addition of time derivatives to the Lagrangian, becomes completely general. These results have nothing to do with either electromagnetism or with Galilei invariance, but are simply a feature of the form of Eqs. (2.8). So, instead, the matter may be put this way when the context is the formalism of Ref. 1: (the usual, i.e. kinematic) gauge invariance is a consequence of the Helmholtz conditions. In other words, *gauge invariance is an intrinsic property of the "usual" domain of classical theoretical mechanics, of the existence of a Lagrangian, and therewith, of a Hamiltonian.*

There is a non abelian field analog to these results and which I discuss briefly in Sec. 3.

In the present work I will derive results concerning "principal part" Lagrange and Hamilton functions, L and H, objects which arise in the generalized formalism. These quantities, as I shall show, are the same as the usual Lagrange and Hamilton functions of mechanics when the system satisfies the Helmholtz conditions (and for choice of an "integrable dynamical gauge".) In addition, every system has a "universal" Lagrangian and Hamiltonian, $\mathscr{L}$ and $\mathscr{H}$, defined over "doubled" spaces of twice the normal number of dimensions; thus, the "universal" phase space of an N dimensional system is $4N$-dimensional, etc. The system motion is executed on a (universal) constraint hypersurface of half this number of dimensions, and this is the old configuration space, or phase space, of the proper dimensionality. The generalized structure for non–Helmholtz systems over the constraint submanifold(s) is then determined by the usual (constrained) formalism over the big space(s).

With these orienting remarks I turn now to the business of the paper. In Sec. 2 I give an overview (and brief review) of the new mechanics, summarizing key results of Ref. 1, and present a few new results (including, incidentally, a curious generalization of the Legendre transformation); in Sec. 3 I establish necessary and sufficient conditions on the functions, $f_{1n}(x,\dot{x},t), f_{2n}(x,\dot{x},t)$ in Eqs. (2.8) for the Helmholtz conditions to be met; in Sec 4 I establish the correspondences of gauge transformations to canonical transformations for the new formalism; in Sec. 5 I establish the kinematic gauge invariance of integrability for the universal Hamiltonian; and in Sec. 6 I conclude with a discussion of the significance of the results derived in the paper.

2. OVERVIEW OF THE FORMALISM

The universal Lagrangian and Hamiltonian $\mathscr{L}$ and $\mathscr{H}$, in the central coordinates[1] for the problem are

$$\mathscr{L} \simeq f_{1n}(x,\dot{x},t)\,z^n + f_{2n}(x,\dot{x},t)\,\dot{z}^n, \tag{2.1}$$

$$\mathscr{H} \simeq F_{1n}(x,p_z,t)\,z^n + F_2^n(x,p_z,t)\,p_{x^n}, \tag{2.2}$$

where the equality symbol denotes strong equality in the sense of Dirac.[9-12] I have introduced a slight change of notation from that previously used: F_{2n} has been replaced by F_2^n, x_n by x^n, etc., and the Einstein summation convention is introduced for repeated upper and lower indices.

The transformation to principal coordinates[1] is given by the equations

$$\xi^n = x^n + \tfrac{1}{2}z^n, \tag{2.3a}$$

$$p_n = p_{z^n} + \tfrac{1}{2}p_{x^n}, \tag{2.3b}$$

$$\xi'^n = x^n - \tfrac{1}{2}z^n, \tag{2.3c}$$

$$-p'_n = p_{z^n} - \tfrac{1}{2}p_{x^n}, \tag{2.3d}$$

and the principal decomposition forms for $\mathscr{L}$ and $\mathscr{H}$ are

$$\mathscr{L} \simeq L(\xi,\dot{\xi},t) - L(\xi',\dot{\xi}',t) + \delta\mathscr{L}(\xi,\dot{\xi};\xi',\dot{\xi}',t), \tag{2.4}$$

$$\mathscr{H} \simeq H(\xi,p,t) - H(\xi',-p',t) + \delta\mathscr{H}(\xi,p;\xi',-p',t). \tag{2.5}$$

The universal constraints of the canonical formalism are

$$z^n \approx 0, \tag{2.6a}$$

$$p_{x^n} \approx 0, \tag{2.6b}$$

in central coordinates, and

$$\xi^n - \xi'^n \equiv \delta\xi^n \approx 0, \tag{2.7a}$$

$$p_n + p'_n \equiv \delta p_n \approx 0, \tag{2.7b}$$

in principal coordinates.

The significance of the quantities appearing in the foregoing list of equations is as follows. The Euler–Lagrange equations for the system motion are

$$\dot{f}_{2n}(x,\dot{x},t) - f_{1n}(x,\dot{x},t) = 0, \tag{2.8}$$

so that the f_{2n} are "momenta" and the f_{1n} can be interpreted as "forces." Passage to the Hamiltonian is accomplished by substituting into the right side of

$$F_{1n}(x,p_z,t) = -f_{1n}(x,\dot{x},t) \tag{2.9a}$$

for the $\dot{x}^n$, the solution to the system of equations

$$p_{z^n} = \frac{\partial \mathscr{L}}{\partial \dot{z}^n} = f_{2n}(x,\dot{x},t), \tag{2.10}$$

viz.,

$$\dot{x}^n = F_2^n(x,p_z,t), \tag{2.9b}$$

and the last equation itself, to realize Eq. (2.2). These steps can always be carried out in the standard case, defined as that for which Eqs. (2.8) can be solved for the accelerations, viz.

$$\det\| \partial f_{2n}/\partial \dot{x}^m \| \neq 0, \tag{2.11}$$

since the latter condition is that for the invertibility of Eqs. (2.10). The Lagrangian, Eq. (2.1), results from the Hamiltonian, Eq. (2.2), by carrying out the inverse Legendre transformation using

$$\dot{x}^n = \frac{\partial \mathscr{H}}{\partial p_{x^n}} = F_2^n(x,p_z,t), \tag{2.12}$$

to eliminate the p_{z^n} in favor of the $\dot{x}^n$. The solution of the system, Eqs. (2.12), obviously is

$$p_{z^n} = f_{2n}(x,\dot{x},t), \tag{2.13}$$

and Eq. (2.1) follows immediately upon substituting (2.13) and (2.9a) into

$$\begin{aligned}\mathscr{L} &\simeq p_{x^n}\dot{x}^n + p_{z^n}\dot{z}^n - \mathscr{H} \\ &\simeq -F_{1n}(x,p_z,t)z_n + p_{z^n}\dot{z}^n .\end{aligned} \tag{2.14}$$

Evidently, we need to have

$$\det\| \partial F_2^n/\partial p_{z^m} \| \neq 0; \tag{2.15}$$

but this is equivalent simply to the (tacit) assumption that the left-hand side of (2.11) be finite. The remarkable symmetry between Lagrangian and Hamiltonian forms goes even further.

Since the decompositions of $\mathscr{L}$ and $\mathscr{H}$ can be written down arbitrarily and hence independently, in the general case, it is evident that L and H need not be connected by Lengendre transformation. Nonetheless, if we assume a decomposition for $\mathscr{L}$, there is a natural corresponding decomposition of $\mathscr{H}$, in which

$$H = H_L = p_n\dot{\xi}^n - L(\xi,\dot{\xi},t). \tag{2.16}$$

However, this is a *generalized* Legendre transformation from L induced by the (ordinary) transformation from $\mathscr{L}$, for the velocities are to be eliminated by means of

$$p_n = \frac{\partial \mathscr{L}}{\partial \dot{\xi}^n} \approx \frac{\partial L}{\partial \dot{\xi}^n} + \left.\frac{\partial \delta \mathscr{L}}{\partial \dot{\xi}^n}\right|_{\xi'=\xi,\dot{\xi}'=\dot{\xi}}, \tag{2.17}$$

which differs from the normal case in the presence of the second term of the right-hand member. When this is done, Eq. (2.5) results, H being given by (2.16) and the residual part by

$$\delta \mathscr{H} \simeq -\delta \mathscr{L}. \tag{2.18}$$

Furthermore, the right side of Eq. (2.17) must reduce to $f_{2n}(x,\dot{x},t)$, so the uniqueness of the transformation, i.e., the uniqueness of H_L, is guaranteed by (2.11), given the f_{2n}.

Conversely, if a principal decomposition is assumed for $\mathscr{H}$, there is a natural, induced, decomposition for $\mathscr{L}$, defined from the inverse transformations. In exactly the manner of the ordinary Legendre transformation, the inverse is realized in

$$L = L_H = \dot{\xi}^n p_n - H(\xi,p,t), \tag{2.19}$$

with the p_n to be eliminated by inverting the system

$$\dot{\xi}^n = \frac{\partial \mathscr{H}}{\partial p_n} \approx \frac{\partial H}{\partial p_n} + \left.\frac{\partial \delta \mathscr{H}}{\partial p_n}\right|_{\xi'=\xi,\,-p'=p}. \tag{2.20}$$

It is easy to show from the canonical equations that the right side of Eq. (2.20) is $F_2^n(\xi,p,t)$. So Eq. (2.20) is the same as Eq. (2.12) and the solution is given by (2.10) or (2.17). Evidently, Eq. (2.19) is supplemented by the analogue of Eq. (2.18), viz.

$$\delta \mathscr{L} \simeq -\delta \mathscr{H}. \tag{2.21}$$

Equations (2.17) and (2.20) do not depend upon the choice of principal decomposition employed to express them, and their solutions for the elimination of the active variables of the transformation are independent of this choice. Consequently, the relationship of L_H and H_L for any choice of principal decomposition (of either L or H), depends upon the function set, $\{f_{2n}(\xi,\dot{\xi},t)\}$, or its "inverse", $\{F_2^n(\xi,p,t)\}$. One can see the full situation in the diagram

$$H_L = H[L,\{f_{2n}\}] \longleftrightarrow L_H = L[H,\{F_2^n\}]. \tag{2.22}$$

Thus a principal decomposition of L induces a principal decomposition of H in a natural way and *vice versa*. I shall assume henceforth that L and H in Eqs. (2.4) and (2.5) are connected by generalized Legendre transformation, and drop the identifying subscripts from L_H and H_L.

An immediate corollary of the foregoing is that if either of δL or δH vanishes, so does the other, by Eq. (2.18) or Eq. (2.21).

3. HELMHOLTZ CONDITIONS

Given a set of functions, $\{G_n\}$, to be used to express the motion of a system, viz.

$$G_n(x,\dot{x},\ddot{x},t) = 0, \quad n = 1,...,N, \tag{3.1}$$

a Lagrange function[13] $L_H = L_H(x,\dot{x},t)$ having the property,

$$\frac{d}{dt}\frac{\partial L_H}{\partial \dot{x}^n} - \frac{\partial L_H}{\partial x^n} = G_n, \quad n = 1,...,N, \tag{3.2}$$

exists if and only if the functions G_n satisfy the identities,[2]

$$\frac{\partial G_n}{\partial \ddot{x}^m} \equiv \frac{\partial G_m}{\partial \ddot{x}^n}, \tag{3.3}$$

$$\frac{\partial G_n}{\partial \dot{x}^n} + \frac{\partial G_m}{\partial \dot{x}^n} \equiv \frac{d}{dt}\left(\frac{\partial G_n}{\partial \ddot{x}^m} + \frac{\partial G_m}{\partial \ddot{x}^n}\right), \tag{3.4}$$

$$\frac{\partial G_n}{\partial x^m} - \frac{\partial G_m}{\partial x^n} \equiv \frac{1}{2}\frac{d}{dt}\left(\frac{\partial G_n}{\partial \dot{x}^m} - \frac{\partial G_m}{\partial \dot{x}^n}\right). \tag{3.5}$$

These are the Helmholtz conditions for the second-order system, Eqs. (3.1).

Equations (3.3) and (3.4) require the G_n to be linear functions of the $\ddot{x}^n$, so we may write

$$G_n(x,\dot{x},\ddot{x},t) = G_{nm}(x,\dot{x},t)\,\ddot{x}^m + g_n(x,\dot{x},t)\,. \tag{3.6}$$

In the standard case, for which the Hamiltonian formalism from L_H is free from constraints, the G_{nm} satisfy the condition of nonvanishing determinant,

$$\det\|G_{nm}(x,\dot{x},t)\| \neq 0\,. \tag{3.7}$$

The Helmholtz conditions can be reexpressed as restrictions of the $G_{nm}(x,\dot{x},t)$ and the $g_n(x,\dot{x},t)$ by substituting (3.6). Parallelling Havas,[2] I write these as

$$G_{nm} \equiv G_{mn}\,, \tag{3.7a}$$

$$\frac{\partial G_{nm}}{\partial \dot{x}^l} \equiv \frac{\partial G_{lm}}{\partial \dot{x}^n}\,, \tag{3.7b}$$

$$\frac{\partial g_n}{\partial \dot{x}^m} + \frac{\partial g_m}{\partial \dot{x}^n} \equiv \left(\dot{x}^k \frac{\partial}{\partial x^k} + \frac{\partial}{\partial t}\right)(G_{nm} + G_{mn})\,, \tag{3.7c}$$

$$\frac{\partial G_{nm}}{\partial x^l} - \frac{\partial G_{lm}}{\partial x^n} \equiv \frac{1}{2}\frac{\partial}{\partial \dot{x}^m}\left(\frac{\partial g_n}{\partial \dot{x}^l} - \frac{\partial g_l}{\partial \dot{x}^n}\right), \tag{3.7d}$$

$$\frac{\partial g_n}{\partial x^m} - \frac{\partial g_m}{\partial x^m} \equiv \frac{1}{2}\left(\dot{x}^k \frac{\partial}{\partial x^k} + \frac{\partial}{\partial t}\right)\left(\frac{\partial g_n}{\partial \dot{x}^m} - \frac{\partial g_m}{\partial \dot{x}_n}\right). \tag{3.7e}$$

Equations (3.7a)–(3.7e) are *necessary and sufficient conditions* for existence of a classical Lagrange function, in the usual sense, and I shall refer to systems for which they are satisfied as the *Helmholtz case*. Note that (3.7b) implies the existence of a function $\gamma_m(\dot{x})$ for every m, such that $G_{nm} = \partial\gamma_m/\partial\dot{x}^n$. From the symmetry expressed in (3.7a), *the Helmholtz conditions therefore guarantee the form, Eq.* (3.8) *below, assumed for the* G_n *in the present theory.*

The present formalism is defined for systems of equations for which

$$G_n(x,\dot{x},\ddot{x},t) = \dot{f}_{2n}(x,\dot{x},t) - f_{1n}(x,\dot{x},t)\,, \tag{3.8}$$

which, expanded, assumes the form of Eqs. (3.6), with the unique identifications:

$$G_{nm} \equiv \frac{\partial f_{2n}}{\partial \dot{x}^m}\,, \tag{3.9}$$

$$g_n \equiv \left(\dot{x}^l \frac{\partial}{\partial x^l} - \frac{\partial}{\partial t}\right) f_{2n} - f_{1n}\,. \tag{3.10}$$

The restriction to the standard case reduces to (2.11) but the f_{1n}, f_{2n} otherwise are arbitrary. The $2N$ functions are not uniquely specified by the system, Eqs. (3.8), however, but only up to equivalence under the replacements,

$$f_{1n} \to f'_{1n} = f_{1n} + \frac{d}{dt}A_n(x,t)\,, \tag{3.11a}$$

$$f_{2n} \to f'_{2n} = f_{2n} + A_n(x,t)\,, \tag{3.11b}$$

where the N functions, $A_n(x,t)$, are arbitrary and mutually independent. These "dynamical gauge transformations" form a group, D, and comprise a wider class than the usual gauge transformations of the potentials for the electromagnetic field, which here correspond to an "integrable" subgroup, as we shall see.

I define functions, A_{nm}, B_{nm} and C_{nm}, whose vanishing expresses integrability relations for the f_{1n}, f_{2n}. These quantities and the integrability relations, are

$$A_{nm} \equiv \frac{\partial f_{1n}}{\partial x^m} - \frac{\partial f_{1m}}{\partial x^n} = 0, \tag{3.12a}$$

$$B_{nm} \equiv \frac{\partial f_{1n}}{\partial \dot{x}^m} - \frac{\partial f_{2m}}{\partial x^n} = 0, \tag{3.12b}$$

$$C_{nm} \equiv \frac{\partial f_{2n}}{\partial \dot{x}^m} - \frac{\partial f_{2m}}{\partial \dot{x}^n} = 0. \tag{3.12c}$$

I will prove the following for the system, Eqs. (3.8):

(i) *If the integrability relations, Eqs.* (3.12a)–(3.12c), *are met, the Helmholtz conditions* (3.7a)–(3.7e), *are satisfied identically.*

(ii) *Conversely, if the Helmholtz conditions* (3.7a)–(3.7e) *hold, the existence of an integrable gauge,* f'_{1n}, f'_{2n}, *for which Eqs.* (3.12a)–(3.12c) *are satisfied, is guaranteed.*

Under (3.11a) *and* (3.11b) *we have*

$$A_{nm} \to \mathbf{A}_{nm} = A_{nm} + \frac{d}{dt}\left(\frac{\partial A_n}{\partial x^m} - \frac{\partial A_m}{\partial x^n}\right), \tag{3.13a}$$

$$B_{nm} \to \mathbf{B}_{nm} = B_{nm} + \left(\frac{\partial A_n}{\partial x^m} - \frac{\partial A_m}{\partial x^n}\right), \tag{3.13b}$$

$$C_{nm} \to \mathbf{C}_{nm} = C_{nm}\,, \tag{3.13c}$$

so that the integrability relations are gauge invariant only for the subgroup $K \subset D$ for which the A_n satisfy

$$A_n(x,t) = \partial W(x,t)/\partial x^n\,. \tag{3.14}$$

K will be called the *kinematic* or *integrable* subgroup. I will also prove the following:

(iii) *Each gauge transformation from* K corresponds exactly to the addition of a time derivative of a function, $W(x,t)$, *of the coordinates and the time, to principal part Lagrangians,* L. *Moreover, in the Helmholtz case for integrable dynamical gauges,* L *and* L_H *are kinematic gauge equivalent.*

As I have already noted in Ref. 1, the addition to $\mathscr{L}$ of the *most general form* of a time derivative of the coordinates and the time causes a change of $\mathscr{L}$ corresponding to the transformation (3.11a) and (3.11b). Thus, for the effects of gauge transformations in the Lagrange functions, we have

$$D\colon \mathscr{L} \to \tilde{\mathscr{L}} + \frac{d}{dt}z^n A_n(x,t)\,, \tag{3.15a}$$

$$K\colon L \to \tilde{L} = L + \frac{d}{dt}W(\xi,t)\,. \tag{3.15b}$$

The situation may be summarized as follows: (i) *Modulo* a dynamical gauge transformation (K-equivalence class)., integrability of $[(f_{1n}(x,\dot{x},t), f_{2n}(x,\dot{x},t)]$ and the Helmholtz conditions for a second-order system are equivalent; and (ii) the Lagrangian potentials for a set of equations satisfying the Helmholtz conditions have the line integral expression[14]

$$\mathscr{L} \simeq L(\xi,\dot{\xi},t) - L(\xi',\dot{\xi}',t) = \int_{(\xi',\dot{\xi}')}^{(\xi,\dot{\xi})} (dx^n f_{1n} + d\dot{x}^n f_{2n})\,, \tag{3.16}$$

for integrable dynamical gauge choices for $[f_{1n}(x,\dot{x},t), f_{2n}(x,\dot{x},t)]$. Under (3.15b) residual part specification of principal decompositions of a universal Lagrangian may be tak-

en as kinematic gauge invariant, and the corresponding gauge covariance group for $\delta\mathscr{L}$ is then the factor group[15] $D' = D/K$.

To establish (i), (ii), and (iii) we can reexpress Eqs. (3.7a)–(3.7e), substituting Eqs. (3.9), (3.10), and (3.12). Due to the form of the G_n in Eqs. (3.8), the velocity-integrability of the C_{nm} for each m is assured, so Eq. (3.7b) can be dropped. Respectively, the other four conditions are

$$C_{nm} \equiv 0, \tag{3.16a}$$

$$B_{nm} + B_{mn} \equiv 0, \tag{3.16b}$$

$$\frac{\partial}{\partial \dot{x}^m}(B_{nl} - B_{ln}) - \left(\dot{x}^k \frac{\partial}{\partial x^k} + \frac{\partial}{\partial t}\right)\frac{\partial C_{nl}}{\partial \dot{x}^m} \equiv 0, \tag{3.16c}$$

$$-A_{nm} + \left(\dot{x}^k \frac{\partial}{\partial x^k} + \frac{\partial}{\partial t}\right)[\tfrac{1}{2}(B_{nm} - B_{mn})] - \tfrac{1}{2}\left(\dot{x}^k \frac{\partial}{\partial x^k} + \frac{\partial}{\partial t}\right) C_{nm} \equiv 0. \tag{3.16d}$$

Since Eqs. (3.16a)–(3.16d) are homogeneous in the A_{nm}, B_{nm}, and C_{nm}, assertion (i) is immediate. To prove assertion (ii) we must show there exists a gauge in which Eqs. (3.16a)–(3.16d) guarantee the vanishing of the A_{nm}, B_{nm}, and C_{nm}. Equation (3.16a) already states the vanishing of C_{nm}. Using (3.16a) the last terms in Eq. (3.16c) and (3.16d) are zero and we may rewrite the equations as

$$C_{nm} \equiv 0, \tag{3.17a}$$

$$B_{nm} + B_{mn} \equiv 0, \tag{3.17b}$$

$$\partial B_{nm}/\partial \dot{x}^l \equiv 0, \tag{3.17c}$$

$$dB_{nm}/dt \equiv A_{nm}, \tag{3.17d}$$

where Eq. (3.17d) qualifies as the expression of (3.16d) owing to the freedom of B_{nm} from $\dot{x}^l$-dependence implied by Eq. (3.17c). Equation (3.17d) shows that we need only prove that B_{nm} can be made to vanish by a gauge transformation, for then $A_{nm} = 0$ also will hold. By Eq. (3.17b) this means that the B_{nm} must be line integrable in the x^n, which is assured if they satisfy the equations

$$\beta_{nml} \equiv \partial B_{nm}/\partial x^l + \partial B_{ln}/\partial x^m + \partial B_{ml}/\partial x^n = 0, \tag{3.18}$$

identically.

Equation (3.12c) shows that $C_{nm} = 0$ implies that there is a function $\phi(x,\dot{x},t)$ such that

$$f_{2m} = \partial\phi/\partial \dot{x}^m, \tag{3.19}$$

so we may write

$$B_{nm} = \frac{\partial}{\partial \dot{x}^m}\left(f_{1n} - \frac{\partial \phi}{\partial x^n}\right). \tag{3.20}$$

Using (3.17b) we have

$$2B_{nm} = B_{nm} - B_{mn}, = \frac{\partial f_{1n}}{\partial \dot{x}^m} - \frac{\partial f_{1m}}{\partial \dot{x}^n} - \left[\frac{\partial}{\partial x^n}\left(\frac{\partial \phi}{\partial \dot{x}^m}\right) - \frac{\partial}{\partial x^m}\left(\frac{\partial \phi}{\partial \dot{x}^n}\right)\right]. \tag{3.21}$$

Now form the sum in (3.18) using (3.21); the quantity in brackets drops out as it is already in the form of a "curl", and the rest can be reexpressed in terms of the A_{nm}, giving

$$2\beta_{nml} = \frac{\partial A_{ml}}{\partial \dot{x}^n} + \frac{\partial A_{nm}}{\partial \dot{x}^l} + \frac{\partial A_{ln}}{\partial \dot{x}^m}. \tag{3.22}$$

Using Eq. (3.17d) in the form

$$A_{nm} = \frac{\partial B_{nm}}{\partial t} + \dot{x}^k \frac{\partial B_{nm}}{\partial x^k} \tag{3.23}$$

and substituting into (3.22), the right hand side, owing to the independence of the B_{nm} of the $\dot{x}^k$, reduces to β_{nml} and Eq. (3.18) then follows immediately. This proves assertion (ii).

I will prove (iii) next. Suppose first that

$$L \to L + \frac{d}{dt} W(\xi,t). \tag{3.24}$$

To determine the effect in (f_{1n}, f_{2n}) do the following: substitute (3.24) into Eq. (2.4), use Eqs. (2.3a) and (2.3c) and expand in a Taylor series about $z^n = 0, \dot{z}^n = 0$, retaining only first order terms to arrive at the (strong equation) equivalent of Eq. (2.1). Upon comparison, the effect is seen to be the addition of the gradient, $(\partial/\partial x^n, \partial/\partial \dot{x}^n)$, of dW/dt, viz.

$$f_{1n} \to f_{1n} + \frac{\partial}{\partial x^n}\left(\frac{dW}{dt}\right) = f_{1n} + \frac{d}{dt}\left(\frac{\partial W}{\partial x^n}\right), \tag{3.25}$$

$$f_{2n} \to f_{2n} + \frac{\partial}{\partial \dot{x}^n}\left(\frac{dW}{dt}\right) = f_{2n} + \frac{\partial W}{\partial x^n}, \tag{3.26}$$

which is a gauge transformation from K by Eq. (3.14). Conversely, if we perform the gauge transformation given in (3.25) and (3.26), Eq. (2.1) gives for the change in $\mathscr{L}$,

$$\Delta\mathscr{L} = z^n \frac{d}{dt}\left(\frac{\partial W}{\partial x^n}\right) + \dot{z}^n \frac{\partial W}{\partial x^n}, \tag{3.27}$$

$$= \left(z^n \frac{\partial}{\partial x^n} + \dot{z}^n \frac{\partial}{\partial \dot{x}^n}\right)\frac{dW}{dt}. \tag{3.28}$$

The last line is obtained with the help of the identity,

$$\frac{dW}{dt} = \left(\frac{\partial}{dt} + \dot{x}^k \frac{\partial}{\partial x^k}\right) W(x,t). \tag{3.29}$$

Using Taylor expansion methods again, Eq. (3.28) can be reexpressed in principal coordinates as

$$\Delta\mathscr{L} \simeq \frac{dW(\xi,t)}{dt} - \frac{dW(\xi',t)}{dt}, \tag{3.30}$$

displaying the role of the operator on the right side of (3.28) as a total differential on the $(x,\dot{x})$ space. Thus the transformation results in

$$\mathscr{L} \to \mathscr{L}' = \mathscr{L} + \Delta\mathscr{L} \simeq (L + dW/dt) - (L' + dW'/dt), \tag{3.31}$$

to which we may give the interpretation

$$L \to L + dW(\xi,t)/dt. \tag{3.32}$$

Equation (3.16) shows the equations of motion from L to be Eqs. (3.8), which are assumed to be those from L_H. So L and L_H can differ only by a time derivative of a function of the coordinates and the time, as that is the most general solution of the set,

$$\left(\frac{d}{dt}\frac{\partial}{\partial \dot{x}^n} - \frac{\partial}{\partial x^n}\right)(L_H - L) \equiv 0, \quad n = 1,\ldots,N, \tag{3.33}$$

regarded as identities, and with L_H-L assumed to be a func-

tion of the x, $\dot{x}$, and t. This proves the second part of (3), that L and L_H are kinematic gauge equivalent.

I will illustrate the material of this section by examining the problem of a charge q experiencing a Lorentz force, for which

$$\mathbf{f}_1 = q(\mathbf{E} + c^{-1}\dot{\mathbf{x}}\times\mathbf{B}), \tag{3.34a}$$

$$\mathbf{f}_2 = \mathbf{p}_{\text{particle}} = \mathbf{p}(\dot{\mathbf{x}}). \tag{3.34b}$$

Necessary and sufficient conditions for the equations of motion,

$$\dot{\mathbf{f}}_2 - \mathbf{f}_1 = 0, \tag{3.35}$$

whether $\mathbf{p} = \mathbf{p}(\dot{\mathbf{x}})$ is the relativistic or nonrelativistic momentum function, to satisfy the Helmholtz conditions are easily shown to be

$$\nabla\times\mathbf{E} + c^{-1}\partial\mathbf{B}/\partial t = 0, \tag{3.36a}$$

$$\nabla\cdot\mathbf{B} = 0. \tag{3.36b}$$

These are also the conditions for the existence of the usual set of potentials, $\phi = \phi(\mathbf{x},t)$, $\mathbf{a} = \mathbf{a}(\mathbf{x},t)$, of the electromagnetic field: $\mathbf{E} = -\nabla\phi - c^{-1}\partial\mathbf{a}/\partial t$, $\mathbf{B} = \nabla\times\mathbf{a}$. The Lagrangian

$$\mathscr{L} \simeq \mathbf{f}_1\cdot\mathbf{z} + \mathbf{f}_2\cdot\dot{\mathbf{z}}, \tag{3.37}$$

cannot be expressed in the form of Eq. (3.16), as $(\mathbf{f}_1,\mathbf{f}_2)$ is not integrable due to things like [cf. Eq. (3.12a)],

$$\begin{aligned}\nabla\times\mathbf{f}_1 &= q(\nabla\times\mathbf{E} + c^{-1}\nabla(\dot{\mathbf{x}}\times\mathbf{B})),\\ &= q(\nabla\times\mathbf{E} + c^{-1}\dot{\mathbf{x}}(\nabla\cdot\mathbf{B}) - c^{-1}(\dot{\mathbf{x}}\cdot\nabla)\mathbf{B}),\\ &= -qc^{-1}\,d\mathbf{B}(\mathbf{x},t)/dt \neq 0.\end{aligned} \tag{3.38}$$

The dynamical gauge transformation corresponding to the choice in (3.11b)

$$\mathbf{A}(\mathbf{x},t) = qc^{-1}\,\mathbf{a}(\mathbf{x},t), \tag{3.39}$$

results in the integrable set

$$\begin{aligned}\mathbf{f}_1\to\mathbf{f}_1' &= q(\mathbf{E} + c^{-1}\dot{\mathbf{x}}\times\mathbf{B}) + qc^{-1}\,d\mathbf{a}/dt\\ &= \nabla_{\mathbf{x}}(-q\phi + qc^{-1}\dot{\mathbf{x}}\cdot\mathbf{a}),\end{aligned} \tag{3.40a}$$

$$\mathbf{f}_2\to\mathbf{f}_2' = \mathbf{p}(\dot{\mathbf{x}}) + qc^{-1}\mathbf{a}. \tag{3.40}$$

If we define

$$L_0 = \int^{\dot{\mathbf{x}}} d\dot{\mathbf{x}}\cdot\mathbf{p}(\dot{\mathbf{x}}), \tag{3.41}$$

for the kinetic term, Eqs. (3.40) take the form

$$\mathbf{f}_1' = \nabla_{\mathbf{x}}(L_0(\dot{\mathbf{x}}) - q\phi + qc^{-1}\dot{\mathbf{x}}\cdot\mathbf{a}), \tag{3.42a}$$

$$\mathbf{f}_2' = \nabla_{\dot{\mathbf{x}}}(L_0(\dot{\mathbf{x}}) - q\phi + qc^{-1}\dot{\mathbf{x}}\cdot\mathbf{a}), \tag{3.42b}$$

vindicating Eq. (3.16), the quantity in parenthesis being the usual Lagrangian for a charged particle in Lorentz force field.

Observe that the dynamical gauge transformation, (3.40b), to the integrable gauges is what in other language would be termed a "minimal substitution." There the idea is to introduce a gauge field interaction into a formalism; here the equations of motion, or the force law, are assumed given and Eq. (3.40) constitutes a dynamical gauge fixing in order to send δL to zero.

The kinematic gauge transformations all are given by

$$\begin{aligned}\mathbf{f}_1\to\mathbf{f}_1' &= \mathbf{f}_1 + qc^{-1}\frac{d}{dt}\nabla\Lambda\\ &\equiv \mathbf{f}_1 + qc^{-1}\nabla\left(\frac{\partial\Lambda}{\partial t} + \dot{\mathbf{x}}\cdot\nabla\Lambda\right),\end{aligned} \tag{3.43a}$$

$$\mathbf{f}_2\to\mathbf{f}_2' = \mathbf{f}_2 + qc^{-1}\nabla\Lambda, \tag{3.43b}$$

which corresponds by (3.40a) and (3.40b) to transformations of $(\phi,\mathbf{a})$

$$\phi\to\phi' = \phi - c^{-1}\,\partial\Lambda/\partial t, \tag{3.44a}$$

$$\mathbf{a}\to\mathbf{a}' = \mathbf{a} + \nabla\Lambda, \tag{3.44b}$$

the usual gauge transformations of the electromagnetic potentials.

I will close with a brief remark; it is possible to extend many, if not all, of the results of this section to the classical nonAlbelian gauge (Yang–Mills) force law equations derived by Wong.[16] In the extension the "precession" equations[16] of the $I^a(\tau)$, quantities corresponding to the gauge group generators, are treated as subsidiary conditions; and gauge covariant derivatives replace ordinary derivatives. Thus, for example, in Eq. (3.8) $d/d\tau$ gets replaced by $\delta/\delta\tau$, where τ is the particle's proper time and the connection for the covariant derivative pertains (only) to the x^λ– variation, viz.

$$\frac{\delta f_{2\mu}}{\delta\tau} \equiv \dot{x}^\lambda\frac{\partial f_{2\mu}}{\partial x^\lambda} + \dot{x}^\lambda\nabla_\lambda f_{2\mu}, \tag{3.45}$$

where ∇_λ is the gauge covariant derivative operator and the dot denotes differentiation by τ. In this method, Eqs. (3.2) are satisfied *after* the subsidiary conditions have been used to eliminate the $I^a(\tau)$ dependence entering in from the first term. It is this procedure which makes it possible to achieve the form of Eq. (3.45). The modified form of Eq. (3.8), viz.

$$\frac{\delta f_{2\mu}}{\delta\tau} - f_{1\mu} = 0, \tag{3.46}$$

is invariant to the generalized dynamical gauge transformation,

$$f_{1\mu}\to f_{1\mu}' = f_{1\mu} + \frac{\delta}{\delta\tau}\mathscr{A}_\mu(x), \tag{3.47a}$$

$$f_{2\mu}\to f_{2\mu}' = f_{2\mu} + \mathscr{A}_\mu(x), \tag{3.47b}$$

while the kinematic subgroup is defined by the restriction of the $\mathscr{A}_\mu(x)$ to the form

$$\mathscr{A}_\mu(x) = \nabla_\mu\,\mathscr{W}(x), \tag{3.48}$$

which may be compared with Eq. (3.14). Evidently, the generalized dynamical gauge group, for the nonalbelian gauge force law, is still abelian. The Lagrangian giving the Wong equations by this procedure is

$$L(x,\dot{x},\tau) = -m(-\dot{x}^2)^{1/2} + g\dot{x}^\lambda A_{\lambda a}(x)\,I^a(\tau), \tag{3.49}$$

where g is the coupling strength of the particle, of mass m, to the field, and the $A_{\lambda a}(x)$ are the gauge field potential functions.

4. GAUGE TRANSFORMATIONS AND THE HAMILTONIAN FORMALISM

I shall show that (i) the effect in $\mathscr{H}$ of a dynamical gauge transformation can be realized in a canonical transformation on the full phase space; while (ii) that of a kinematic gauge transformation can be understood to correspond to a

canonical transformation of H in the space of the (ξ, p) and the $(\xi' - p')$, the two being identical. Specifically, I shall find transformations of the forms of the Hamiltonians,

$$D: \mathscr{H} \to \bar{\mathscr{H}} \simeq \mathscr{H} - \frac{\partial}{\partial t} z^n A_n(x,t), \tag{4.1a}$$

$$K: \mathscr{H} \to \bar{\mathscr{H}} = H - \frac{\partial}{\partial t} W(\xi,t), \tag{4.1b}$$

where in the momentum variables in the arguments of $\mathscr{H}$ and H one makes also the substitutions

$$\left.\begin{aligned} D: p_{x^m} &\to p_{x^m} - \frac{\partial}{\partial x^m} z^n A_n(x,t), \\ p_{z^m} &\to p_{z^m} - \frac{\partial}{\partial z^m} z^n A_n(x,t) \end{aligned}\right\} \tag{4.2a}$$

$$K: p_m \to p_m - \frac{\partial}{\partial \xi^m} W(\xi,t). \tag{4.2b}$$

Equations (4.1) and (4.2) may be compared with Eqs. (3.15). Note again that L and H are not the Lagrangian and the Hamiltonian, but rather *principal part functions for (arbitrary) principal decompositions.*

In addition, I shall obtain the result (iii) that the transformations of principal coordinates by the canonical transformations for the gauge group have the property that primed and unprimed variables transform the same to strong equality *only for the kinematic subgroup.* In consequence of this, residual parts, $\delta\mathscr{H}$, of principal decompositions of the Hamiltonians can be assumed kinematic gauge invariant up to equivalence by the duplicate (joint) canonical transformation of the primed and unprimed variables, indicated in (4.2b), but are then dynamical gauge dependent. The corresponding gauge covariance group is again the factor group D', in parallel with the situation in Sec. 3, for $\delta\mathscr{L}$. Proceeding to the proofs, I take up (i) first.

From Eqs. (2.10) and (3.11b), the effect of a *gauge* transformation on H is determined from

$$p_{z^n} - A_n(x,t) = f_{2n}(x,\dot{x},t), \tag{4.3}$$

whose solution, by Eq. (2.9b) is

$$\dot{x}^m = F_2^m(x, p_z - A(x,t), t). \tag{4.4}$$

From Eqs. (2.9a) and (3.11a), together with Eqs. (4.3) and (4.4), we have also

$$F_{1n}(x, {}_z, t) \to F_{1n}(x, p_z - A, t) - \frac{dA_n}{dt}(x, p_z - A, t), \tag{4.5}$$

where the last term stands for

$$\frac{dA_n}{dt}(x, p_z - A, t) \equiv \frac{\partial A_n}{\partial t} + F_2^m(x, p_z - A, t)\frac{\partial A_n}{\partial x^m}, \tag{4.6}$$

with the help of Eq. (4.4) for the coefficient of $\partial A_n/\partial x^m$. Substituting into Eq. (2.2) we find

$$\mathscr{H} \to \bar{\mathscr{H}} = \left[F_{1n}(x, p_z - A, t) - \frac{dA_n}{dt}(x, p_z - A, t)\right] z^n + F_2^n(x, p_z - A, t)\, p_{x^n}. \tag{4.7}$$

On the other hand, consider the effect produced by the *canonical* transformation generated by means of

$$\mathscr{R}_1 = \mathscr{R}_1(x, z, P_X, P_Z, t), \tag{4.8}$$

$$= [P_{Z^n} - A_n(x,t)]\, z^n + x^n P_{X^n}. \tag{4.9}$$

The equations of the transformation are

$$\frac{\partial \mathscr{R}_1}{\partial z^n} = p_{z^n}, \quad \frac{\partial \mathscr{R}_1}{\partial P_{X^n}} = X^n, \tag{4.10}$$

$$\frac{\partial \mathscr{R}_1}{\partial x_n} = p_{x^n}, \quad \frac{\partial \mathscr{R}_1}{\partial_{P_{Z^n}}} = Z^n. \tag{4.11}$$

These equations determine the transformation, T_1, as

$$T_1: x^n \to X^n = x^n, \tag{4.12a}$$

$$z^n \to Z^n = z^n, \tag{4.12b}$$

$$p_{x^n} \to P_{X^n} = p_{x^n} + \frac{\partial A_m(x,t)}{\partial x^n} z^m, \tag{4.12c}$$

$$p_{z^n} \to P_{Z^n} = p_{z^n} + A_n(x,t). \tag{4.12d}$$

Note that if $A_n = 0$ for all n, R_1 generates the identity transformation. The transformation of $\mathscr{H}$ is given by

$$\mathscr{H}(x,z,p_x,p_z,t) \to \bar{\mathscr{H}}(x,z,p_x,p_z,t), = \mathscr{H}(T_1^{-1}(x,z,p_x,p_z,t) + \partial\mathscr{R}_1/\partial t. \tag{4.13}$$

Inverting the transformation, Eqs. (4.12), one finds

$$\begin{aligned} \bar{\mathscr{H}} &= \mathscr{H}(x,z,p_x - z^m\,\partial A_m/\partial x, p_z - A, t) - \partial(z^n A_m)/\partial t \\ &\simeq [F_{1n}(x, p_z - A, t) - \partial A_n/\partial t\,]\, z^n \\ &\quad + F_2^n(x, p_z - A, t)[p_{x^n} - (\partial A_m/\partial x^n)\, z^m] \\ &\simeq [F_{1n}(x, p_z - A, t) - \{\partial A_n/\partial t + F_2^m(x, p_z - A, t) \\ &\quad \times (\partial A_n/\partial x^m)\}]\, z^n + F_2^n(x, p_z - A, t)\, p_{x^n}, \end{aligned} \tag{4.14}$$

which, upon substitution of Eq. (4.6) for the bracketed quantity, reduces to the same expression as that on the right side of Eq. (4.7). Thus, finally,

$$\bar{\mathscr{H}}(x,z,p_x,p_z,t) \simeq \bar{\mathscr{H}}(x,z,p_x,p_z,t), \tag{4.15}$$

the promised result. I turn next to the proof of (iii).

Equations (4.12) can be rewritten in a slightly different way. If we denote the $2N$ coordinates, (x^n, z^n) by the single symbol, q^r, and the (p_{x^n}, p_{z^n}) by π_r, we have

$$T_1: q^r \to q^r, \tag{4.16a}$$

$$\pi_r \to \pi_r + \frac{\partial}{\partial q^r}[z^n A_n(x,t)], \tag{4.16b}$$

and the quantity in parenthesis is the object whose time derivative, when added to $\mathscr{L}$, corresponds to the gauge transformation to which the canonical transformation T_1 is equivalent [compare Eqs. (4.2a)]. Let us examine the behavior of principal coordinates under this transformation now. The definitions of ξ and ξ' in Eqs. (2.3), are untouched, but those of p and $-p'$ are transformed; from Eqs. (2.3) and (4.12)

$$\begin{aligned} p_n \to \tilde{p}_n &= p_n + A_n(x,t) + \tfrac{1}{2} z^n\,\partial A_m/\partial x_n \\ &\simeq p_n + A_n(\xi,t) + \tfrac{1}{2}\delta\xi^m(\partial A_m/\partial \xi^n - \partial A_n/\partial \xi^m), \end{aligned} \tag{4.17a}$$

$$-p'_n \to -\tilde{p}'_n = -p'_n + A_n(x,t) - \tfrac{1}{2}z^m\,\partial A_m/\partial x^n$$
$$\simeq -p'_n + A_n(\xi',t) - \tfrac{1}{2}\delta\xi^m(\partial A_m/\partial\xi'^n - \partial A_n/\partial\xi'^m), \quad (4.17b)$$

where I have used things like

$$A_n(x,t) = A_n(\xi - \tfrac{1}{2}\delta\xi,t) \simeq A_n(\xi,t) - \tfrac{1}{2}\delta\xi^m\,\partial A_n(\xi,t)/\partial\xi^m. \quad (4.18)$$

The "strongly specified" transformations of p_n and $-p'_n$ are identical if and only if

$$\partial A_m/\partial x^n - \partial A_n/\partial x^m = 0, \quad (4.19)$$

but otherwise asymmetrical.

To obtain (ii) I shall only have to express the transformed $\mathscr{H}$ in an appropriate way in principal coordinates under the assumption of Eq. (4.19), or equivalently, Eq. (3.14). Using Eq. (3.14), the canonical transformation corresponding to (4.17) becomes

$$T_2: \xi^n \to \xi^n, \quad (4.20a)$$
$$p_n \to p_n + \partial W(\xi,t)/\partial\xi^n, \quad (4.20b)$$
$$\xi'^n \to \xi'^n, \quad (4.20c)$$
$$-p'_n \to -p'_n + \partial W(\xi',t)/\partial\xi'^n. \quad (4.20d)$$

The generating function for T_2 in Eqs. (4.20) is

$$\mathscr{R}_2 = \mathscr{R}_2(\xi,\xi',P,P',t) \quad (4.21)$$
$$= P_n\,\xi^n - W(\xi,t) + P'_n\,\xi'^n + W(\xi',t), \quad (4.22)$$

which can be verified from Eqs. (4.10) and (4.11) specialized to this case. So, from Eqs. (4.13) and (2.5), the corresponding transformed Hamiltonian is given as

$$\tilde{\mathscr{H}}(\xi,\xi',p,p',t) \simeq \mathscr{H}(T_2^{-1}(\xi,\xi',p,p',t)) + \partial\mathscr{R}_2/\partial t$$
$$\simeq [H(\xi,p-\partial W,t) - \partial W/\partial t] - [H(\xi',-p'-\partial' W',t) - \partial W'/\partial t] + \delta H(\xi,p-\partial W;\xi',-p'-\partial' W',t). \quad (4.23)$$

Of course $\tilde{\mathscr{H}}$ must be the same, by Eq. (4.15). But it is worthwhile to display this explicitly. Substituting Eq. (3.14) into Eq. (4.7) we have

$$\tilde{\mathscr{H}} \simeq F_{1n}(x,p_z - \partial W,t)\,z^n + F_2^n(x,p_z-\partial W,t) \times [p_{x^n} - z^m(\partial/\partial x^m)\,\partial W/\partial x^n] - z^m(\partial/\partial x^m)\,\partial W/\partial t, \quad (4.24)$$

which follows after a little manipulation and some help from Eq. (4.6). Equation (4.24) can be reexpressed by applying the Taylor series expansion trick, as in Eq. (4.18); we have

$$z^m(\partial/\partial x^m)\,\partial W/\partial t \simeq \partial W(\xi,t)/\partial t - \partial W(\xi',t)/\partial t, \quad (4.25a)$$
$$z^m(\partial/\partial x^m)\,\partial W/\partial x^n \simeq \partial W(\xi,t)/\partial\xi^n - \partial W(\xi',t)/\partial\xi'^n. \quad (4.25b)$$

Noting also Eqs. (2.3) and (2.7) and introducing $\bar{\xi} = 1/2(\xi+\xi') = x$, $\bar{p} = 1/2(p-p') = p_z$, Eq. (4.24) becomes

$$\tilde{\mathscr{H}} \simeq F_{1n}(\bar{\xi},\bar{p} - \bar{\partial}\bar{W},t)\,\delta\xi^n + F_2^n(\bar{\xi},\bar{p}-\bar{\partial}\bar{W},t) \times\delta(p_n - \partial_n W) - (\partial W/\partial t - \partial W'/\partial t). \quad (4.26)$$

Equation (2.2) can be written with the same notation as Eq. (4.26); Eqs. (2.5) and (4.26) then give the expected result

$$\tilde{\mathscr{H}} \simeq [H(\xi,p-\partial W,t) - \partial W/\partial t] - [H(\xi',-p'-\partial' W',t) - \partial W'/\partial t] + \delta\mathscr{H}(\xi,p-\partial W;\xi',-p'-\partial' W',t). \quad (4.27)$$

From Eqs. (4.23) and (4.27) evidently, as claimed in (ii), we can interpret the effect of the canonical transformation of $\mathscr{H}$ for a kinematic gauge transformation corresponding to W, as corresponding to a transformation of H:

$$H(\xi,p,t) \to H(\xi,p-\partial W,t) - \partial W/\partial t, \quad (4.28)$$

i.e. the result of a canonical transformation on the phase space of the (ξ,p). Indeed, the generating function in Eq. (4.22) obviously has the form,

$$\mathscr{R}_2(\xi,\xi',P,P',t) = R(\xi,P,t) - R(\xi',-P',t), \quad (4.29a)$$

with

$$R(\xi,P,t) = P_n\,\xi^n - W(\xi,t), \quad (4.29b)$$

and R generates the transformation Eqs. (4.20a,b) and (4.28).

I will reexamine the relationship between R_1 and R_2 from another standpoint in the next paper of the present series.[14]

5. INTEGRABILITY FOR THE HAMILTONIAN

I have fixed the relationship between principal decompositions of $\mathscr{L}$ and $\mathscr{H}$ by the generalized Legendre transformation and shown the equivalence of the Helmholtz conditions to integrability. The latter refers to the existence of an integrable dynamical gauge, for which $\delta\mathscr{L}$, and, by Eq. (2.18), $\delta\mathscr{H}$ as well, vanish. Integrability is a kinematic gauge invariant property of the $f_{1n}(x,\dot{x},t)$, $f_{2n}(x,\dot{x},t)$, expressed in Eqs. (3.12) and realized in Eq. (3.16) in the determination of the Lagrange function L, for the Helmholtz case.

We have seen the complete correspondence of properties of $\mathscr{H}$, H and $\delta\mathscr{H}$ to those of $\mathscr{L}$, L and $\delta\mathscr{L}$ in Secs. 3 and 4, in the behavior of these quantities under gauge transformations. For the Helmoltz case, as I shall now show, when $\delta\mathscr{H} \simeq 0$, the F_{1n}, F_2^n, also satisfy integrability relations, analogous to Eqs. (3.12), viz.

$$\mathscr{A}_{nm} \equiv \frac{\partial F_{1n}}{\partial x^m} - \frac{\partial F_{1m}}{\partial x^n} = 0, \quad (5.1a)$$
$$\mathscr{B}^m_n \equiv \frac{\partial F_{1n}}{\partial p_{z^m}} - \frac{\partial F_2^m}{\partial x^n} = 0, \quad (5.1b)$$
$$C^{nm} \equiv \frac{\partial F_2^n}{\partial p_{z^m}} - \frac{\partial F_2^m}{\partial p_{z^n}} = 0. \quad (5.1c)$$

This may be seen by using Eqs. (2.3) in Eq. (2.5); expanding about $z^n \simeq 0$, $p_{x^n} \simeq 0$, and keeping only first-order terms to obtain a strong equation; equating the result then to the central coordinate form in Eq. (2.2); and finally identifying the coefficients of the constraints. This gives F_{1n} as $\partial H/\partial x^n$ and F_2^n as $\partial H/\partial p_n$, in analogy with the results for the Lagrangian L, as in Eq. (3.16); equating cross-partials of H then produces Eqs. (5.1). In exact parallel with Eq. (3.16) we have[14]

$$\mathscr{H} \simeq H(\xi,p,t) - H(\xi',-p',t) = \int_{(\xi',-p')}^{(\xi,p)} [dx^n F_{1n}(x,p_z,t) + dp_{z^n} F_2^n(x,p_z,t)], \tag{5.2}$$

where the line integration is along an arbitrary path $\sigma \to [x^n(\sigma), p_{z^n}(\sigma)]$ between the points $(\xi' - p')$ and (ξ, p). Now H is the "potential" for the "vector" (F_{1n}, F_2^n) over the crossed-pair "phase space" of (x^n, p_{z^n}).

Equations (3.13) give the gauge transformation properties of the quantities appearing in Eqs. (3.12), from which the kinematic gauge invariance of integrability for a set of f_{1n}, f_{2n} was noted previously. Similarly, we expect that integrability for a set of F_{1n}, F_2^n is also a kinematic gauge invariant property. Equations (4.6) determine the behavior of the quantities in Eqs. (5.1) under a gauge transformation; I find, after a little labor

$$\mathscr{A}_{nm} \to \bar{\mathscr{A}}_{nm} = \left[\mathscr{A}_{nm} - \left(\frac{\partial}{\partial t} + F_2^k \frac{\partial}{\partial x^k}\right)\left(\frac{\partial A_n}{\partial x^m} - \frac{\partial A_m}{\partial x^n}\right) + \frac{\partial A_k}{\partial x^n}\frac{\partial A_l}{\partial x^m} C^{kl}\right]' + \left[\left\{\left(\frac{\partial F_{1n}}{\partial p_{z^k}} - \frac{\partial A_l}{\partial x^n}\frac{\partial F_2^k}{\partial p_{z^l}}\right) \times \left(\frac{\partial A_m}{\partial x^k} - \frac{\partial A_k}{\partial x^m}\right) + \frac{\partial A_m}{\partial x^k}\mathscr{B}^k_{\ m}\right\} - \{n \rightleftarrows m\}\right]', \tag{5.3}$$

$$\mathscr{B}^m_{n\cdot} \to \bar{\mathscr{B}}^m_{n\cdot} = \left[\mathscr{B}^m_{n\cdot} + \frac{\partial F_2^k}{\partial p_{z^m}}\left(\frac{\partial A_k}{\partial x^n} - \frac{\partial A_n}{\partial x^k}\right) + \frac{\partial A_k}{\partial x^n} C^{nk}\right]', \tag{5.4}$$

$$C^{nm} \to \bar{C}^{nm} = [C^{nm}]', \tag{5.5}$$

where the primes on the right sides express the convention, for a function $f(x,p_z,t)$,

$$[f(x,p_z,t)]' \equiv f(x,p_z - A(x,t),t). \tag{5.6}$$

Equations (5.3)–(5.5) display the desired result, since Eqs. (5.1) imply the vanishing of the right hand sides if A_n has the form of Eq. (3.14).

6. DISCUSSION AND CONCLUSION

The Helmholtz conditions, Eqs. (3.3)–(3.5), are integrability relations on the G_n for existence of a potential function L. The G_n, in turn, in the Helmholtz case have the form of Eq. (2.8); while the $[f_{1n}(x,\dot{x},t), f_{2n}(x,\dot{x},t)]$ array forms a $2N$-dimensional vector over an "expanded" configuration space of coordinates and velocities. This vector is the gradient of L, as may be inferred, for instance, from Eqs. (3.12) or from Eq. (3.16). In the nonHelmholtz case where the G_n yet have the form of Eq. (2.8), the vector $[f_{1n}(x,\dot{x},t), f_{2n}(x,\dot{x},t)]$, no longer a gradient, still can be expressed as the sum of a gradient part and a nonconservative part. This corresponds directly to the principal decomposition of $\mathscr{L}$ when it is expressed in principal coordinates: The gradient of L, plus the analogous array of partial derivatives of $\delta\mathscr{L}$ evaluated on the constraint hypersurface (in the expanded configuration space of $4N$ dimensions), equals the vector (f_{1n}, f_{2n}). The separation of a gradient piece from this vector is arbitrary when this vector is not itself a gradient; correspondingly the decomposition of $\mathscr{L}$ is arbitrary.

So the first two theorems of Sec. 3, above Eqs. (3.13), serve to characterize the scope of the traditional subject of mechanics, where L exists in the sense of Eq. (3.2), as defined in terms of $2N$–dimensional conservative vector fields on the space of $(x,\dot{x})$. The new formalism can be understood as an extension to the nonconservative case.

The corresponding characterization exists for the phase space. The generalized Legendre transformation in Sec. 2 provides the basis of uniqueness for the correspondences of principal part functions L and H, while Eqs. (2.9a) (2.9b) and (2.10) specify the relationship of the vectors (f_{1n}, f_{2n}) and (F_{1n}, F_2^n). The corollary stated at the end of Sec. 2 shows the correspondence of integrability of these two vector fields on their respective (and distinct constraint submanifold) spaces.

The precision of the foregoing correspondences has to be understood against the backdrop of two well-known versions of arbitrariness present in Hamilton–Lagrange theory, however. Equation (3.33) expresses the first; its general solution is $dW(x,t)/dt$, where W is arbitrary. On the other hand, Eq. (2.8) is invariant against the general gauge transformation given in (3.11a) and (3.11b). The third theorem of Sec. 3 ties down these "freedoms" in terms of the properties of the dynamical gauge field functions, $A_n(x,t)$, in complete generality: those parts of A_n—arrays that form a (configuration) gradient of some function $W(x,t)$ may be regarded as generating t—derivative contributions to the principal part function L. That the corresponding effects in H can be understood as canonical transformations, in the right way, was detailed in Secs. 4 and 5. Thus correspondences between Lagrangian and Hamiltonian quantities survive, and in addition receive natural extensions.

It is an interesting and very significant feature of the present theory of mechanics, to my mind, that one does not have to choose an integrable dynamical gauge (assuming one exists) in order to have a theoretical mechanics description for a system. Thus it is not *a priori* necessary to introduce potentials for the problem of a particle experiencing a Lorentz force, as it is in the usual mechanics, which does not even get off the ground without them. This is a satisfying feature of the present formalism because Eq. (2.8) contains no potentials.

In Sec. 4 two generating functions were discussed, $\mathscr{R}_1$ and $\mathscr{R}_2$, in Eqs. (4.9) and (4.22). When $A_n = \partial W/\partial x^n$ holds, $\mathscr{R}_1 \simeq \mathscr{R}_2$ also does. But the last relation is extraordinary in that it pertains *not* to functions of canonical variables but to *generating* functions, of *mixed old and new variables*. Also, Eqs. (4.17) showed that nonintegrable gauge transformations destroy the separation of primed and unprimed principal coordinates. These circumstances are symptomatic of a more general situation that will find a formulation and resolution in the next paper of this series,[14] where, from the Hamilton-Jacobi theory on the big space, a complete time-dependent general coordinate transformation theory is constructed for the constraint submanifold.

ACKNOWLEDGMENT

This work was supported by Naval Surface Weapons Center Independent Research Program.

[1]Robert Cawley, Phys. Rev. A **20**, 2370 (1979).
[2]P. Havas, Nuovo Cimento Supp. **5** (Ser. X) **3**, 363 (1957). The stress of the physical difficulties with the integrating factor approach appears to be due originally to this author.
[3]P. Havas, Acta Phys. Austriaca **38**, 14 5(1973).
[4]R. M. Santilli, Hadronic J. **1**, 223 (1978). This author also gives an extensive list of references, older as well as more recent, to the general problem of Hamilton and Lagrange formulations, along with an exposition of the history of the subject.
[5]E. H. Kerner, Adv. Chem. Phys. **19**, 325 (1971).
[6]H. Dekker, Z. Physik B **21**, 295 (1975); *ibid.* B **24**, 211 (1976); and Phys. Rev. A **16**, 2126 (1977).
[7]A recent reference is C. Grebogi, A. N. Kaufman, and R. G. Littlejohn, Phys. Rev. Lett. **43**, 1668 (1979).
[8]J. M. Jauch, Helv. Phys. Acta **37**, 284 (1964).
[9]P. A. M. Dirac. Can. J. Math. **2**, 147 (1950); and Proc. Roy. Soc. (London) A **246** 326 (1958); see also P. A. M. Dirac, "Lectures on Quantum Mechanics," Yeshiva U., New York, 1964 Lectures No. 1 and No. 2.
[10]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974), Chap. 8.
[11]The use of strong equality in the Lagrangian formalism is introduced in Ref. 1. The hypersurface needed to define it lies in the $4N$-dimensional "expanded configuration space" of coordinates and velocities, and is defined by $z^n = 0$, $\dot{z}^n = 0$, $n = 1$ to N.
[12]The possibility of the Hamiltonian form for an arbitrary system of equations by appending extra variables is a general theorem due to S. Lie and G. Koenigs [E. T. Whittaker, *Analytical Dynamics* (Cambridge, 1937), p. 275]. The Hamiltonian due to G. D. Birkoff ("Dynamical Systems," Amer. Math. Soc., New York, 1927, pp. 57 and 58) closely resembles the function appearing for $\mathscr{K}$, on the right side of Eq. (2.2). $\mathscr{K}$, however, specified only to strong equality, is not a function but an equivalence class of functions. The relationship to the Birkhoff Hamitonian is discussed in Ref. 1.
[13]In this section, the subscript H stands for Helmholtz rather than Hamiltonian, as in the last section.
[14]This formula has interesting significance and generalization for the present theory, to a line element space symplectic formalism [Robert Cawley, Phys. Rev. D (in press)].
[15]Since the dynamical gauge group is Abelian, every subgroup is invariant so the factor group of D by K is defined.
[16]S. K. Wong, Nuovo Cimento A **65**, 689 (1970).

Relativistic covariance and rotational electrodynamics

James F. Corum
Department of Electrical Engineering, West Virginia University, Morgantown, West Virginia 26506

(Received 13 September 1979; accepted for publication 16 May 1980)

The following article demonstrates how the logical coherence of relativistic electrodynamics is maintained for a particular family of rotational paradoxes. The internal computational unity, for rotation, is preserved through the manifestation of a commonly unrecognized geometrical property of tensor calculus.

INTRODUCTION

An interesting family of paradoxes frequently discussed in electromagnetics classes concerns the fields produced by rotating charge distributions.[1] That a rotating spherical shell of charge, for example, produces a magnetic field in the frame of a laboratory observer is readily accepted by many students. "However", a student will query, "with respect to an observer whose system of reference is co-rotating with the sphere, the charges are at rest and hence, in this system, no magnetic fields ought to exist."

A similar paradox occurs with rotating cylindrical distributions of charge.[2] Once again, a laboratory observer perceives an axial magnetic field whose source is the rotating cylinder of charge. For a co-rotating observer, the charges are at rest and therefore should produce no magnetic fields. Even worse, for the rotating observer inside the cylinder, by Gauss's Law, there should exist no electric fields. How then can we have a nonzero field tensor inside, in the inertial system (laboratory frame), and a vanishing field tensor in the rotating system, since if a tensor vanishes in any frame it must vanish in all other systems of reference at that point?

In both of these paradoxical examples the reader is cautioned not to accept the conclusions unquestioningly. The logic may be impeccable—but the presuppositions are erroneous. Why do we resurrect these historical paradoxes? Because we believe that they illustrate the computational beauty and conceptual richness of relativity theory as manifested through the inherent presence of the object of anholonomity.

We parenthetically comment that our physics is transpiring on an underlying manifold that has zero curvature (in the limit—i.e., we assume that we can paste charges onto a flat manifold and not disturb the geometrical structure of the manifold). Even though we are doing non-inertial physics, we are properly within the realm of what is *traditionally* called special relativity. (Our approach works equally well on curved spacetimes, of course.)

1. THE ROTATING SHELL OF CHARGE

The presentation of the magnetic field arising from a rotating charged spherical shell appears in many textbooks on electromagnetism. The computation is performed in the frame of an inertial observer and, with the aid of calculational conveniences, smoothly proceeds from a specification of the current density

$$J_\phi(r') = \frac{Q\omega}{4\pi a}\sin\theta'\,\delta(r'-a) \tag{1}$$

to the computation of the vector potential

$$A_\phi(r,\theta) = \begin{cases} \dfrac{Q\omega r}{3ca}\sin\theta & (r\leqslant a) \quad (2a) \\ \dfrac{Q\omega a^2}{3cr^2}\sin\theta & (r\geqslant a) \quad (2b) \end{cases}.$$

The magnetic induction follows from the curl of the vector potential as

$$B_\theta(r,\theta) = \begin{cases} -\dfrac{2Q\omega}{3ca}\sin\theta & (r<a) \quad (3a) \\ \dfrac{Q\omega a^2}{3cr^3}\sin\theta & (r>a) \quad (3b) \end{cases}.$$

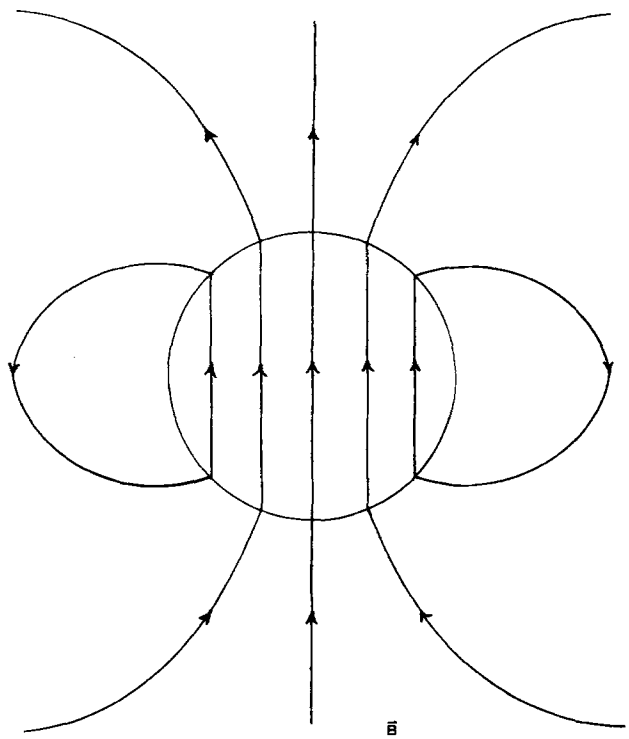


FIG. 1. Magnetic flux density arising from a rotating spherical shell which is uniformly charged as perceived by a nonrotating observer. (Compare figure 5-11a of Ref. 7)

 0022-2488/80/092360-05$1.00

and

$$B_r(r,\theta)=\begin{cases}\dfrac{2\omega Q}{3ca}\cos\theta & (r<a) \quad (3c)\\ \dfrac{2\omega Qa^2}{3cr^3}\cos\theta & (r>a) \quad (3d)\end{cases}.$$

These familar fields are shown for reference in Fig. 1. The prescription followed has been to specify the current distribution and follow the path $J^\mu \to A^\mu \to F^{\mu\nu}$. The inertial electromagnetic field tensor is taken as

$$F_{\mu\nu}=\frac{\partial A_\nu}{\partial x^\mu}-\frac{\partial A_\mu}{\partial x^\nu}.$$

One wonders if a similar path could be followed in the rest frame of the rotating charges: Say, $J^a \to A^a \to F^{ab}$, where the rotating quantities and inertial quantities, by their tensorial nature, would be related by some Lorentz–like transformation. In this article, we propose to pedagogically demonstrate that this covariant nature of Maxwellian electrodynamics, under relativistic rotation, is only attainable with the inclusion of the object of anholonomity. (This remarkable nontensorial object does not modify the theory of relativity in anyway, but rather is a commonly unrecognized[3] inherent pre-supposition of tensor calculus on manifolds[4].)

2. ROTATION IN SPHERICAL COORDINATES

Pirani[5] and later Irvine[6] have discussed how a rotating observer may let his world line provide a time-like direction and employ the Frenet–Serret Formulas to obtain a *field* of orthogonal reference frames, $\mathbf{e}_a$.[4] The result for spherical coordinates, is the *field* of frames ($x^4=\tau=ct$)

$$\mathbf{e}_1=\mathbf{e}_r \quad (4a)$$

$$\mathbf{e}_2=\mathbf{e}_\theta \quad (4b)$$

$$\mathbf{e}_3=\gamma\mathbf{e}_\phi+\gamma\frac{r^2\omega\sin^2\theta}{c}\mathbf{e}_\tau \quad (4c)$$

$$\mathbf{e}_4=\gamma\frac{\omega}{c}\mathbf{e}_\phi+\gamma\mathbf{e}_\tau. \quad (4d)$$

These are orthogonal, and are related to the natural basis vectors of the inertial observer's field of reference frames, at every observable spacetime event, by the Lorentz–like transformation

$$\mathbf{e}_a=h^\mu_a\mathbf{e}_\mu, \quad (5)$$

where

$$h^\mu_a=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&\gamma&\gamma\dfrac{\omega}{c}\\0&0&\gamma\dfrac{r^2\omega\sin^2\theta}{c}&\gamma\end{pmatrix} \quad (6)$$

and

$$\gamma=\left(1-\frac{r^2\omega^2\sin^2\theta}{c^2}\right)^{-\frac{1}{2}}. \quad (7)$$

Furthermore, for all observers in the rotating frame, the metric tensor $g_{ab}=\mathbf{e}_a\cdot\mathbf{e}_b$, will then have the orthogonal form

$$g_{ab}=\begin{pmatrix}1&0&0&0\\0&r^2&0&0\\0&0&r^2\sin^2\theta&0\\0&0&0&-1\end{pmatrix}. \quad (8)$$

Notice that this is consistent with $g_{ab}=h^\mu_a h^\nu_b g_{\mu\nu}$, where $g_{\mu\nu}$ is the metric of the inertial (or non-rotating) frame.

In the inertial frame, one employs the usual spherical coordinate Christoffel Symbols

$$\begin{Bmatrix}\mu\\ \nu\alpha\end{Bmatrix}=\tfrac{1}{2}g^{\mu\beta}\left[g_{\alpha\beta,\nu}+g_{\beta\nu,\alpha}-g_{\nu\alpha,\beta}\right]. \quad (9)$$

These familiar non-tensorial objects are symmetric in their lower indices, and their spherical coordinate values appear in many textbooks. Given the components with respect to one frame, one transforms them with respect to any other frame according to

$$\Gamma^a_{bc}=h^a_\mu h^\alpha_b h^\nu_c\begin{Bmatrix}\mu\\ \nu\alpha\end{Bmatrix}-h^\mu_b h^\nu_c\frac{\partial h^a_\nu}{\partial x^\mu}. \quad (10)$$

Employing Eq. 6, one finds that even though the inertial frame $\begin{Bmatrix}\mu\\ \nu\alpha\end{Bmatrix}$ is symmetric in the lower indices, the Γ^a_{bc} are not. This is because the set of frames, Eq.(4)and the transformation, Eq.(6)are *anholonomic*. The concept of anholonomity has been discussed elsewhere and one should note that it arises from a choice of the field of reference frames and is not a tensorial quantity on the underlying manifold as, for example, torsion would be. (Torsion can't be transformed away over an extended region.) How does this asymmetry affect electrodynamics?

An invariant form of Maxwell's Equations may be arrived at from a variational principle as

$$\nabla_\nu F^{\mu\nu}=\frac{4\pi}{c}J^\mu, \quad (11)$$

where

$$J^\mu=\rho v^\mu \quad (12)$$

and

$$F_{\mu\nu}=\nabla_\mu A_\nu-\nabla_\nu A_\mu. \quad (13)$$

The four-vector potential has the covariant components given by:

$$A_\mu=(A_r, rA_\theta, r\sin\theta A_\phi; -\Phi). \quad (14)$$

In inertial frames of reference, Eq. (13) reduces to the simple expression

$$F_{\mu\nu}=\frac{\partial A_\nu}{\partial x^\mu}-\frac{\partial A_\mu}{\partial x^\nu}. \quad (15)$$

However, in the rotating frames of Eq. (4), Eq. (13) must be written as the tensor

$$F_{ab}=\frac{\partial A_b}{\partial x^a}-\frac{\partial A_a}{\partial x^b}+2\,\Omega^c_{ab}A_c, \quad (16)$$

where

$$\Omega^c_{ba}\triangleq\tfrac{1}{2}\left[\Gamma^c_{ab}-\Gamma^c_{ba}\right]. \quad (17)$$

In inertial frames, this last component vanishes, but in rotating frames, Eq. (15) is inappropriate for the description of

electrodynamics because it leaves out anholonomic effects (Eq. (15) is *not* a tensor unless $\Omega^{\alpha}_{\nu\mu} = 0$). The objects of anholonomity may be computed by a variety of techniques,[4] and since they are needed here, we present their computation by the Cartan Calculus in the Appendix.

3. SPECIFICATION OF THE SOURCE DISTRIBUTION

In the inertial frame, let us specify the components of the four–current density

$$J^{\mu} = \rho v^{\mu} = (J_r, \frac{J_\theta}{r}, \frac{J_\phi}{r\sin\theta}; c\rho) \tag{18}$$

so that

$$J^{\mu} = (0,0,\omega\rho_0; c\rho_0)\,. \tag{19}$$

The charge density is taken as *uniform when observed from the inertial frame*, say

$$\rho_0 = \frac{Q}{4\pi a}\,\delta(r' - a)\,.$$

In particular, for computational convenience, this means that we are assuming that *in the proper frame of the shell* the density of charge varies continuously with latitude in such a way that the charge distribution just compensates for the relativistic increase in density and hence, in the inertial frame, is perceived as a uniform charge distribution. [This assumption was really made back in Eq. (1).] In the inertial frame, the differential equation represented by 11, 12 and 13 may be solved as

$$A_4 = -\Phi = \begin{cases} -\dfrac{Q}{a} & r \leqslant a \\ -\dfrac{Q}{r} & r \geqslant a \end{cases} \tag{20}$$

and

$$A_3 = A_\phi r\sin\theta = \begin{cases} \dfrac{Q\omega r^2}{3ca}\sin^2\theta & r \leqslant a \\ \dfrac{Q\omega a^2}{3rc}\sin^2\theta & r \geqslant a\,. \end{cases} \tag{21}$$

One may quickly form the inertial frame field tensor from Eq. (15) (since $\Omega^{\alpha}_{\nu\mu} = 0$). Further, the separate $F^{\mu\nu}$ for $r < a$ and $r > a$ satisfy the point-wise boundary conditions across the shell discontinuity. We have followed the prescription given in Sec. 1: $J^{\mu} \to A^{\mu} \to F^{\mu\nu}$. We now shift to the rotating frame and pursue the suggestion $J^{a} \to A^{a} \to F^{ab}$. If our analysis is acceptable, we should have a completely covariant formulation of the problem, and have resolved any paradoxes along the way.

4. COMPUTATIONS IN THE ANHOLONOMIC FRAME

Employing the dual to the transformation of Eq. (6), we write down the current density in the proper frame of the charges (the rotating frame)

$$J^{a} = h^{a}_{\mu} J^{\mu} = \rho v^{a} \tag{22}$$

or, more explicitly

$$J^{a} = \left(0,0,0; c\frac{\rho_0}{\gamma}\right). \tag{23}$$

Again, J^4 reflects our choice of having ρ_0 specified as uniform in the inertial frame. This form for the charge density is acceptable because the rotating observer perceives no moving charges (and hence no apparent magnetic field-producing currents.)

Remembering that in the rotating frame $dV = \gamma dV_0$, we note that charge invariance is satisfied:

$$\frac{1}{c}\int\int\int J^{a} dS_a = Q\,. \tag{24}$$

Since Eq(11) is form-invariant, the vector potential must satisfy the differential equation

$$\nabla_b\left[g^{ac}g^{bd}\left(\frac{\partial A_d}{\partial x^c} - \frac{\partial A_c}{\partial c^d} + 2\,\Omega^{e}_{cd}A_e\right)\right] = \frac{4\pi}{c}J^{a}, \tag{25}$$

where the objects of anholonomity are given in the Appendix. One may find the transformed components of the vector potential by Eq. (6)

$$A_a = h^{\mu}_{a} A_{\mu}$$

as

$$A_1 = A_2 = 0, \tag{26a}$$

$$A_3 = \begin{cases} -\dfrac{2\gamma Q r^2\omega\sin^2\theta}{3ca} & (r \leqslant a) \\ \dfrac{\gamma Q\omega a^2\sin^2\theta}{3cr} - \dfrac{\gamma Q r\omega\sin^2\theta}{c} & (r \geqslant a) \end{cases} \tag{26b}$$

$$A_4 = \begin{cases} -\dfrac{\gamma Q}{a} + \dfrac{\gamma Q r^2\omega^2\sin^2\theta}{3c^2 a} & (r \leqslant a) \\ -\dfrac{\gamma Q}{r} + \dfrac{\gamma Q a^2\omega^2\sin^2\theta}{3rc^2} & (r \geqslant a). \end{cases} \tag{26c}$$

Now, by Eq. (26), (8), and (23) and the tabulated Ω^{a}_{bc}, one may readily verify that Eq. (26) is indeed the solution of Eq.

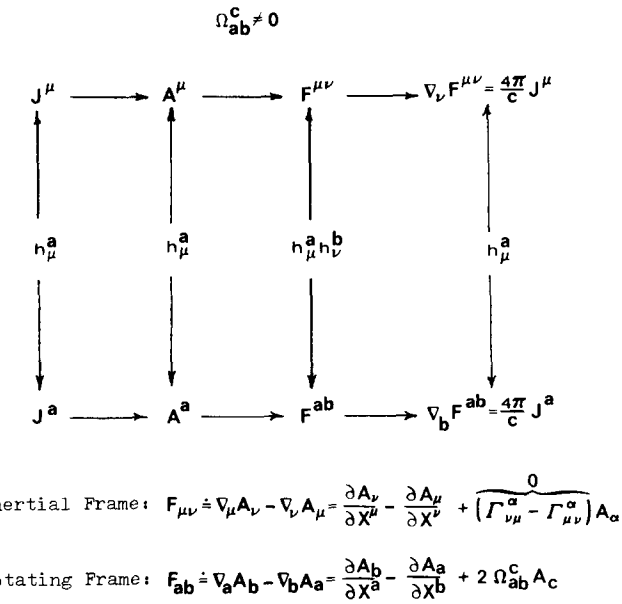


FIG. 2. This diagram indicates what is demanded of any covariant formulation of electrodynamics. A relativistically covariant formulation is possible for rotation by virtue of the intrinsic nature of the anholonomic object.

(25). What would have happened if the rotating observer had neglected the anholonomic contribution? Figure 2 would then have to display an internal inconsistency.

Next, one might desire the field tensor in the rotating frame. The reader is encouraged to transform the inertial $F_{\mu\nu}$ and compare with the results computed by Eq. (16) and the tabulated Ω^{c}_{ab}. For completeness, we list the nonzero results (as obtained by either method).

$$F_{14} = E_r = \begin{cases} \dfrac{2\gamma Q r\omega^2 \sin^2\theta}{3ac^2} & (r<a) \\ \dfrac{\gamma Q}{r^2}\left(1 - \dfrac{a^2\omega^2\sin^2\theta}{3c^2}\right) & (r>a) \end{cases} \tag{27a}$$

$$F_{24} = rE_\theta = \begin{cases} \dfrac{2\gamma Q r^2\omega^2 \sin\theta\cos\theta}{3ac^2} & (r<a) \\ \dfrac{2\gamma Q a^2\omega^2\sin\theta\cos\theta}{3rc^2} & (r>a) \end{cases} \tag{27b}$$

$$F_{31} = r\sin\theta B_\theta = \begin{cases} -\dfrac{2\gamma Q r\omega\sin^2\theta}{3ac} & (r<a) \\ -\dfrac{\gamma Q\omega\sin^2\theta}{c}\left(1 - \dfrac{a^2}{3r^2}\right) & (r>a) \end{cases} \tag{27c}$$

$$F_{23} = r^2\sin\theta B_r = \begin{cases} \dfrac{2\gamma Q r^2\omega\sin\theta\cos\theta}{3ac} & (r<a) \\ \dfrac{2\gamma Q a^2\omega\sin\theta\cos\theta}{3rc} & (r>a)\,. \end{cases} \tag{27d}$$

One may verify that Eqs. 27 really do satisfy the Maxwell Equations (if, by now, he is not already convinced of the built in anholonomity in relativity for non-inertial frames). This exercise is particularly illuminating since the left-hand side of Maxwell's source equations [or Eq. (25)] vanishes for $r \neq a$, and the jump conditions are satisfied across the shell discontinuity.

5. CYLINDRICAL SHELL OF CHARGE

When discussing the rotating cylinder, Fig. 3, Feynman[2] makes the provocative comment, "There is no 'relativity of rotation'. A rotating system is not an inertial frame, and the laws of physics are different. We must be sure to use equations of electromagnetism only with respect to inertial coordinate systems." To which we agree wholeheartedly. But, after making this seductive and tantalizing statement, he passes on to another topic without hinting how one does do electromagnetism in noninertial systems. (At a similar point in their analysis of the sphere, Panofsky and Phillips[1] appeal to General Relativity, even though they are working in a flat spacetime.) Surely one may proceed as $J^a \to A^a \to F^{ab}$ since all are tensor quantities.

Let us formulate the problem in the inertial system in the following manner: we specify the current density in cylindrical coordinates.

$$J^\mu = \rho v^\mu = (J_r, J_\theta/r, J_z; c\rho) \tag{28}$$

as

$$J^\mu = (0, \omega\sigma_0, 0, c\sigma_0), \tag{29}$$

where σ_0 is the surface charge density, assumed to be uniform over the thin cylindrical shell as perceived by the inertial observer. We write

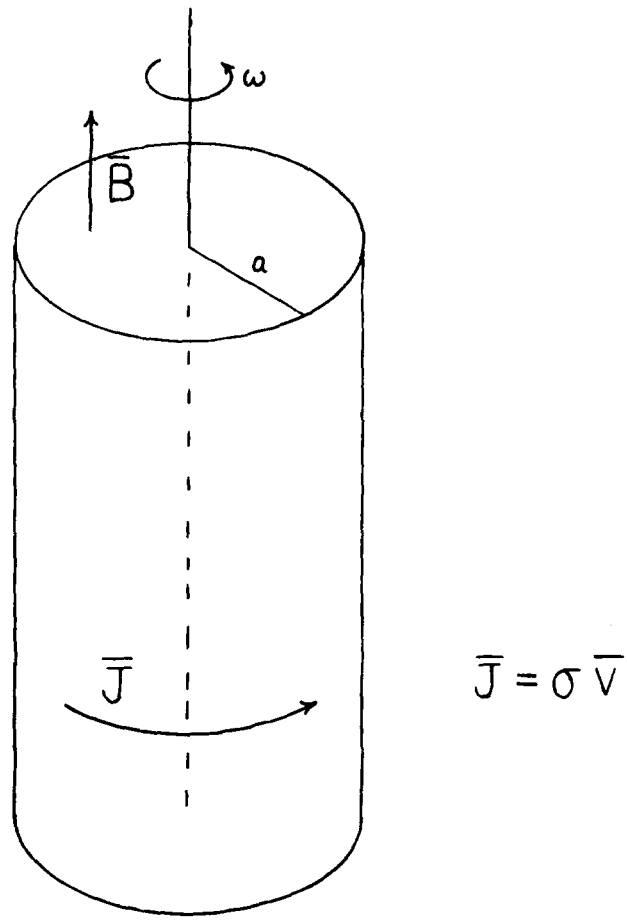


FIG. 3. Magnetic flux density arising from a rotating charged cylindrical shell. (Compare Fig. 14-5 of Ref. 2)

$$\sigma_0 = \frac{Q}{2\pi aL}\delta(r'-a) = \frac{\lambda}{2\pi a}\delta(r'-a)\,. \tag{30}$$

Equations (11), (12), and (13) are then solved as

$$A_2 = rA_\theta = \begin{cases} \dfrac{2\pi\sigma_0 a\omega r^2}{c} & (r \leqslant a) \\ \dfrac{2\pi\sigma_0 a^3\omega}{c} & (r \geqslant a) \end{cases} \tag{31a}$$

$$A_4 = -\Phi = \begin{cases} \Phi_0 & (r \leqslant a) \\ \Phi_0 + \dfrac{\lambda}{2\pi}\ln\left(\dfrac{a}{r}\right), & (r \geqslant a) \end{cases} \tag{31b}$$

where Φ_0 is a suitably chosen constant. The field tensor has nonzero components:

$$F_{14} = E_r = \begin{cases} 0 & r<a \\ \dfrac{\lambda}{2\pi r} & r>a \end{cases} \tag{32}$$

$$F_{12} = rB_z = \begin{cases} \dfrac{4\pi\sigma_0 a\omega r}{c} & r<a \\ 0 & r>a\,. \end{cases} \tag{33}$$

We now turn to the analysis $J^a \to A^a \to F^{ab}$ in the rotating frame. In this system

$$J^a = \rho v^a = (0,0,0,c\sigma_0/\gamma). \tag{34}$$

Here again, we have assumed that in the proper frame of the cylinder the density of charge varies in such a way that the charge distribution just compensates for the relativistic increase in density and is consequently perceived *by the inertial observer* as uniform. (In this case a naive application of Gauss's Law would surely lead a rotating observer to conclude that F^{ab} for $r < a$ vanishes entirely.) The Ω^{c}_{ab} for rotating cylindrical coordinates have been included in the Appendix. One may readily verify that a solution to Eq.(25) for the distribution of Eq. (34) is

$$A_2 = \frac{2\pi\gamma\sigma a\omega r^2}{c} + \frac{\gamma\omega\Phi_0 r^2}{c} \quad (r \leqslant a) \tag{35a}$$

$$A_4 = \gamma\Phi_0 + \frac{2\pi\gamma\sigma_0 a\omega^2 r^2}{c^2} \quad (r \leqslant a)\,. \tag{35b}$$

From Eq. (16) we compute the nonzero components of the field tensor as

$$F_{12} = E_r = \frac{4\pi\sigma_0 a\omega^2 r}{c^2} \quad (r < a) \tag{36a}$$

$$F_{12} = rB_z = \frac{4\pi\gamma\sigma_0 ra\omega}{c} \quad (r < a)\,. \tag{36b}$$

We note that these satisfy the Maxwell Equations. As a check, we also see that

$$J^a = h^a_\mu J^\mu \tag{37a}$$

$$A^a = h^a_\mu A^\mu \tag{37b}$$

$$F_{ab} = h^\mu_a h^\nu_b F_{\mu\nu}\,. \tag{37c}$$

The internal consistancy of Fig. 2 is again demonstrated. We are now in a position to analyze Feynman's students' query, " 'What if I put myself in the frame of reference of the rotating cylinder? Then there is just a charged cylinder at rest, and I know that the electrostatic equations say there will be *no* electrostatic fields inside . . . . Something must be wrong.' "

Our response is to reecho our opening comments: the logic is unquestionable, but the presuppositions (concerning Ω^{c}_{ab}) are unsound.

ACKNOWLEDGMENTS

The author wishes to acknowledge conversations held early in the formation of this analysis with H.C. Ko, C.V. Heer, and U.H. Gerlach, all of the Ohio State University.

APPENDIX

For reasons of completeness, we briefly sketch one of several techniques for obtaining the spherical coordinate objects of anholonomity for the rotating observer. (The cylindrical coordinate objects are derived in Ref. 4 .) In order to obtain the Ω^{c}_{ab}, one may actually perform the laborious calculation indicated by Eq. (10) of this article. Alternatively, he might employ the field frames given by Eq.(4) and the duality relation:

$$\langle \omega^a, \mathbf{e}_b \rangle = \delta^a_b \tag{A1}$$

to find the natural 1–forms for the rotating observer.

$$\omega^1 = dr \tag{A2a}$$

$$\omega^2 = d\theta \tag{A2b}$$

$$\omega^3 = \gamma(d\phi - \omega dt) \tag{A2c}$$

$$\omega^4 = \gamma c\left(dt - \frac{r^2\omega\sin^2\theta}{c^2}\,d\phi\right). \tag{A2d}$$

Given these forms, one may compute the exterior derivative

$$d\omega^a = 2\Omega^a_{bc}\omega^b \wedge \omega^c \tag{A3}$$

and read off the non-zero components of the spherical objects of anholonomity:

$$\Omega^{3}_{13} = -\Omega^{3}_{31} = \tfrac{1}{2}\frac{\gamma^2 r\omega^2\sin^2\theta}{c^2} \tag{A4a}$$

$$\Omega^{3}_{23} = -\Omega^{3}_{32} = \tfrac{1}{2}\frac{\gamma^2 r^2\omega^2\sin\theta\cos\theta}{c^2} \tag{A4b}$$

$$\Omega^{4}_{41} = -\Omega^{4}_{14} = \tfrac{1}{2}\frac{\gamma^2 r\omega^2\sin^2\theta}{c^2} \tag{A4c}$$

$$\Omega^{4}_{42} = -\Omega^{4}_{24} = \tfrac{1}{2}\frac{\gamma^2 r^2\omega^2\sin\theta\cos\theta}{c^2} \tag{A4d}$$

$$\Omega^{4}_{31} = -\Omega^{4}_{13} = \gamma^2\frac{r\omega\sin^2\theta}{c} \tag{A4e}$$

$$\Omega^{4}_{32} = -\Omega^{4}_{23} = \gamma^2\frac{r^2\omega\sin\theta\cos\theta}{c} \tag{A4f}$$

Similarly, the nonzero cylindrical coordinate components of the anholonomic objects are:

$$\Omega^{2}_{12} = -\Omega^{2}_{21} = +\frac{\gamma^2 r\omega^2}{2c^2} \tag{A5a}$$

$$\Omega^{4}_{21} = -\Omega^{4}_{12} = +\frac{\gamma^2 r\omega}{c} \tag{A5b}$$

$$\Omega^{4}_{41} = -\Omega^{4}_{14} = +\frac{\gamma^2 r\omega^2}{2c^2} \tag{A5c}$$

[1]W.K.H. Panofsky and M. Phillips, *Classical Electricity and Magnetism* (Addison-Wesley, Reading, Mass., 1962), 2nd edition, p. 339.
[2]R.P. Feynman, R.B. Leighton, and M. Sands, *The Feynman Lecture on Physics* Vol. II (Addison-Wesley, Reading, Mass., 1964), pp. 14-7.
[3]J. Ise and J.L. Uretsky, Am. J. Phys. **26**, 431–35 (1958).
[4]J.F. Corum, J. Math. Phys. **18**, 770 (1977).
[5]F.A.E. Pirani, Bull. Acad. Pol. Sci. **5**, 143 (1957).
[6]W.M. Irvine, Physica **30**, 1160 (1964).
[7]J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd edition.

Fock–Tani representation for composite particles in a soluble model

M. D. Girardeau [a]
Physique Théorique, Université de Nice, [b] *Nice, France*

(Received 29 January 1980; accepted for publication 6 June 1980)

The transformation from the usual Fock representation to the Fock–Tani representation for composite particles is carried out for a simplified model of the composite which allows closed-form expressions for the transformed quantities. A "statistical renormalization", in which single particle energies and interactions become dependent on the composite particle occupation number, plays an essential rôle in the solution, allowing absorption of unlinked terms into the definition of the renormalized energies.

I. INTRODUCTION

The Fock–Tani representation for composite particles has been shown to be a convenient representation for the states and observables of systems of composite particles such as atoms, molecules, ions, and nuclei, allowing explicit incorporation of the effects of internal structure and excitation, rearrangement collisions, etc... into the algebra of states and observables.[1-4] There are, however, some formidable mathematical problems involved in the transformation from the standard Fock representation to the new Fock–Tani representation, and as a result only partial results are available for the explicit terms in the Hamiltonian in the new representation. It is therefore useful to examine a model in which the structure of the system is so simplified that all steps of the change of representation can be carried out explicitly, in hopes that the insight thus gained into the structure of the representation can be extended to real physical systems.

The simplest imaginable quantum-mechanical model of a composite particle is one composed of two types of fermions, distinguished by two values 1,2 of a fermion species subscript, such that each type of fermion can be present in only a single state. Denote the annihilation and creation operators for the fermion of type 1 by ψ_1 and $\psi_1^\dagger$, and those for the second fermion by ψ_2 and $\psi_2^\dagger$. As usual, the dagger denotes the Hermitian conjugate. These are assumed to satisfy anticommutation relations of the usual form[5]

$$\psi_1^2 = \psi_2^2 = 0, \quad [\psi_1,\psi_1^\dagger]_+ = [\psi_2,\psi_2{}^\dagger]_+ = 1,$$
$$[\psi_1,\psi_2]_+ = [\psi_1,\psi_2^\dagger]_+ = 0. \tag{1}$$

The Hamiltonian is taken to be

$$H = \epsilon_1\psi_1^\dagger\psi_1 + \epsilon_2\psi_2^\dagger\psi_2 + v\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1. \tag{2}$$

It has only four eigenstates, consisting of the vacuum state $|0\rangle$ with eigenvalue zero, the one-fermion states

$$|1\rangle = \psi_1^\dagger|0\rangle, \quad |2\rangle = \psi_2^\dagger|0\rangle \tag{3}$$

with eigenvalues ϵ_1 and ϵ_2 respectively, and a two-fermion state

$$|a\rangle = \psi_1^\dagger\psi_2^\dagger|0\rangle \equiv A^\dagger|0\rangle, \tag{4}$$

with eigenvalue

$$\epsilon_a = \epsilon_1 + \epsilon_2 + v. \tag{5}$$

If $v < -(\epsilon_1 + \epsilon_2)$, then this state can be thought of as a simplified model of a bound atom (composite particle), and for this reason the notation $|a\rangle$ will be employed. The "physical atom" creation operation $A^\dagger = \psi_1^\dagger\psi_2^\dagger$ and its corresponding annihilation operator $A = \psi_2\psi_1$ are not Bose operators even though built from fermions pairs, but instead satisfy the nontrivial commutation relations

$$A^2 = 0, \quad [A,A^\dagger]_- = 1 - N_1 - N_2,$$
$$[A,\psi_1]_- = [A,\psi_2]_- = A\psi_1 = A\psi_2 = 0, \tag{6}$$
$$[A,\psi_1^\dagger]_- = \psi_2, \quad [A,\psi_2^\dagger]_- = -\psi_1,$$

where $N_1 = \psi_1^\dagger\psi_1$ and $N_2 = \psi_2^\dagger\psi_2$ are the fermion occupation numbers. The model may be described as a "zero-dimensional" model since the composite particle has neither translational nor internal degrees of freedom.

The Fock space spanned by the eigenstates of H is only four-dimensional and the quantum mechanics is utterly trivial. Nevertheless, and in fact for this very reason, the model is very useful for investigation and comparison of various procedures for carrying out the transformation to the Fock–Tani representation. In the following sections we shall investigate two methods, the "d-matrix" method[3] based on normally ordered operator basis expansions, and a new method involving a "statistical renormalization" of the d-matrix and of the matrix elements of the Hamiltonian, leading to completely explicit expressions in closed form. This latter method may have important implications for both the mathematical structure and the physical implications of this representation in realistic models; the investigation of such generalizations is best deferred until the simplest model is thoroughly understood.

2. IDEAL STATE SPACE

Introduce ideal Bose operators a, $a^\dagger$ satisfying the elementary Bose communication relation

$$[a,a^\dagger]_- = 1, \tag{7}$$

and acting on an "ideal atom" Fock space $\mathscr{A}$ spanned by the states $(n!)^{-1/2}(a^\dagger)^n|0\rangle$, where $|0\rangle$ is the vacuum state of $\mathscr{A}$. This space is initially completely independent of the physical Fock space $\mathscr{F}$ spanned by the four eigenstates $|0\rangle$, $|1\rangle$, $|2\rangle$, $|a\rangle$ of H. The ideal state space $\underline{\mathscr{F}}$ is defined to be the direct product $\underline{\mathscr{F}} = \mathscr{F} \otimes \mathscr{A}$ of the physical and ideal atom spaces.

[a] On leave from Dept. of Physics, University of Oregon, Eugene, Oregon 97043. (present address).
[b] Laboratoire de Physique Théorique, Université de Nice, Parc Valrose 06034 Nice Cedex, France. Equipe de Recherche Associée au C.N.R.S. N°. 128.

 0022-2488/80/092365-11$1.00

The operators $\psi_1, \psi_1^\dagger, \psi_2, \psi_2^\dagger$ on $\mathcal{F}$ and those $a, a^\dagger$ on $\mathcal{A}$ are extended to $\underline{\mathcal{F}}$ in the usual way,

$$\psi_j(\underline{\mathcal{F}}) = \psi_j(\mathcal{F}) \otimes 1(\mathcal{A}), \quad \psi_j^\dagger(\underline{\mathcal{F}}) = \psi_j^\dagger(\mathcal{F}) \otimes 1(\mathcal{A})$$
$$a(\underline{\mathcal{F}}) = 1(\mathcal{F}) \otimes a(\mathcal{A}), \quad a^\dagger(\underline{\mathcal{F}}) = 1(\mathcal{F}) \otimes a^\dagger(\mathcal{A}) \tag{8}$$

and the vacuum $|0)$ now is interpreted as the direct product of that of $\mathcal{F}$ and that of $\mathcal{A}$ satisfying

$$\psi_1|0) = \psi_2|0) = a|0) = 0, \quad (0|0) = 1. \tag{9}$$

On this extended state space $\underline{\mathcal{F}}$, the annihilation and creation operators satisfy the anticommunication and commutation relations

$$\begin{aligned} &\psi_1^2 = \psi_2^2 = 0, \quad [\psi_1, \psi_1^\dagger]_+ = [\psi_2, \psi_2^\dagger]_+ = 1, \\ &[\psi_1, \psi_2]_+ = [\psi_1, \psi_2^\dagger]_+ = 0, \\ &[a, a^\dagger]_- = 1, \\ &[a, \psi_1]_- = [a, \psi_2]_- = [a, \psi_1^\dagger]_- = [a, \psi_2^\dagger]_- = 0. \end{aligned} \tag{10}$$

Note the simplicity of the relations involving a compound with those involving A [Eq. (6)]. This is the main advantage of ideal state space representations in more realistic cases where bound composite states cause difficulties with quantum field theory methods in the usual Fock representation.

There is an isometry between the physical state space $\mathcal{F}$ and the subspace $\underline{\mathcal{F}}'$ of $\underline{\mathcal{F}}$ spanned by the four states $|0)$, $|1)$, $|2)$, and $|a)$, where

$$|1) = \psi_1^\dagger|0), \quad |2) = \psi_2^\dagger|0), \quad |a) = a^\dagger|0). \tag{11}$$

The mapping $|0\rangle \to |0)$, $|1\rangle \to |1)$, $|2\rangle \to |2)$ is trivial, merely involving replacement of the vacuum $|0\rangle$ of $\mathcal{F}$ by that $|0)$ of $\underline{\mathcal{F}}$. On the other hand, the mapping $|a\rangle \to |a)$ from the physical to ideal atom state is nontrivial, involving replacement of the fermion pair operator $A^\dagger$ by the ideal boson operator $a^\dagger$ satisfying simpler commutation relations. Nevertheless, the mapping is an isometry in that norms and inner products between the four basis states $|0\rangle$, $|1\rangle$, $|2\rangle$, $|a\rangle$ of $\mathcal{F}$ are the same as those involving the corresponding ideal states $|0)$, $|1)$, $|2)$, $|a)$ which span $\underline{\mathcal{F}}'$. One can also define an ideal Hamiltonian

$$\underline{H} = \epsilon_1|1)(1| + \epsilon_2|2)(2| + \epsilon_a|a)(a| + \underline{B}(\underline{1} - \underline{R}), \tag{12}$$

where $|1)(1|$, $|2)(2|$, and $|a)(a|$ are the projectors onto the corresponding ideal states, $\underline{R}$ is the projector onto $\underline{\mathcal{F}}'$,

$$\underline{R} = |0)(0| + |1)(1| + |2)(2| + |a)(a|, \tag{13}$$

and $\underline{B}$ is an arbitrary operator on $\underline{\mathcal{F}}$. The ideal Hamiltonian $\underline{H}$ has the same eigenvalues on $\underline{\mathcal{F}}'$ as H does on $\mathcal{F}$, with eigenstates which are the images of those of H according to the previously defined mapping; however, $\underline{H}$ also has physically spurious eigenstates lying in $\underline{\mathcal{F}} \ominus \underline{\mathcal{F}}'$. The projected Hamiltonian

$$\underline{\mathcal{H}} = \underline{R}\underline{H}\underline{R} = \epsilon_1|1)(1| + \epsilon_2|2)(2| + \epsilon_a|a)(a|, \tag{14}$$

has the same eigenstates and eigenvalues as does $\underline{H}$ on the subspace $\underline{\mathcal{F}}'$ isometric with $\mathcal{F}$, whereas it annihilates the spurious states (they all have eigenvalue zero). These properties are all special cases of a general ideal state space formulation of the quantum mechanics of systems of composite particles, but the definition of $\underline{H}$, $\underline{R}$, and $\underline{\mathcal{H}}$ in terms of projectors is rather abstract. A number of methods are known for constructing explicit expressions for $\underline{H}$ and $\underline{\mathcal{H}}$ in terms of annihilation and creation operators.[6] The particular method which will be employed in the remainder of this paper is that of the generalized Tani transformation.[1-4]

TABLE I. Notation for the operator basis.

i	B_i
$n, 1$	$(a^\dagger)^n a^n$
$n, 2$	$(a^\dagger)^n \psi_1^\dagger \psi_1 a^n$
$n, 3$	$(a^\dagger)^n \psi_2^\dagger \psi_2 a^n$
$n, 4$	$(a^\dagger)^n \psi_1^\dagger \psi_2^\dagger a^{n+1}$
$n, 5$	$(a^\dagger)^{n+1} \psi_2 \psi_1 a^n$
$n, 6$	$(a^\dagger)^n \psi_1^\dagger \psi_2^\dagger \psi_2 \psi_1 a^n$

3. GENERALIZED TANI TRANSFORMATION

Consider the unitary operator U on $\underline{\mathcal{F}}$, defined by

$$U = \exp\left(\frac{\pi}{2} F\right), \quad F = A^\dagger a - a^\dagger A = \psi_1^\dagger \psi_2^\dagger a - a^\dagger \psi_2 \psi_1, \tag{15}$$

If follows from the properties

$$\begin{aligned} &F|0) = 0, \quad F\psi_1^\dagger|0) = F\psi_2^\dagger|0) = 0, \\ &F\psi_1^\dagger\psi_2^\dagger|0) = -a^\dagger|0), \quad Fa^\dagger|0) = \psi_1^\dagger\psi_2^\dagger|0), \end{aligned} \tag{16}$$

and the power series expansion of the exponential that

$$\begin{aligned} &U^{-1}|0) = |0), \quad U^{-1}|1) = |1), \quad U^{-1}|2) = |2), \\ &U^{-1}|a\rangle = |a), \end{aligned} \tag{17}$$

where $|1) = \psi_1^\dagger|0)$, $|2) = \psi_2^\dagger|0)$, and $|a\rangle = \psi_1^\dagger\psi_2^\dagger|0)$. Thus this transformation effects the desired mapping from $\mathcal{F}$ to $\underline{\mathcal{F}}'$. The ideal Hamiltonian $\underline{H}$, which has the same eigenvalues and matrix elements on $\underline{\mathcal{F}}'$ as H does on $\mathcal{F}$, is taken to be

$$\underline{H} = U^{-1} H U. \tag{18}$$

The projector $\underline{R}$ on $\underline{\mathcal{F}}'$ is

$$\underline{R} = U^{-1} \underline{P}_0 U, \tag{19}$$

where $\underline{P}_0$ is the projector onto the subspace of $\underline{\mathcal{F}}$ spanned by eigenstates of N_a with eigenvalue zero; here $\underline{N}_a$ is the ideal boson number operator $N_a = a^\dagger a$. $\underline{P}_0$ has a number of representations of which two convenient ones are[2]

$$\underline{P}_0 = (2\pi)^{-1} \int_0^{2\pi} d\vartheta \, e^{i\underline{N}_a \vartheta}, \tag{20}$$

and

$$\underline{P}_0 = :e^{-\underline{N}_a}: = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} (a^\dagger)^n a^n, \tag{21}$$

where the colons denote the normal-ordering operation. The evaluation of $\underline{H}$, $\underline{R}$, and $\underline{\mathcal{H}}$ is a nontrivial mathematical problem. Use of the multiple commutator expansion for the unitary transformation (18) is not productive since $\pi/2$ is not a small expansion parameter and the infinite operator series are difficult to sum into closed-form expressions. Equation of motion techniques are much more powerful and

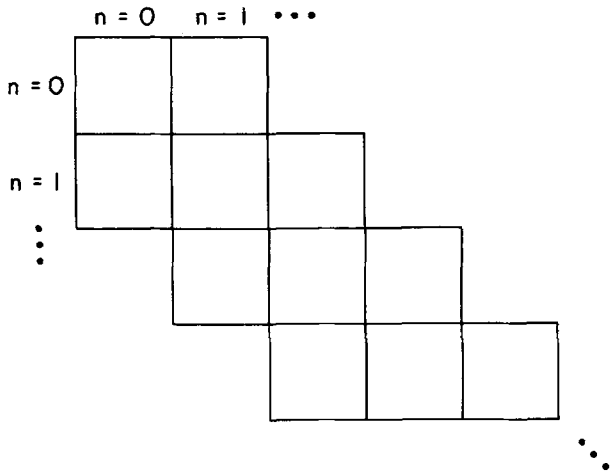


FIG. 1. Tridiagonal block structure of the d-matrix.

can, in fact, be made to yield exact closed-form expressions in this simple case. Define

$$B(t) = e^{-tF} B e^{tF}, \tag{22}$$

where B is any operator on $\mathcal{F}$. Then

$$B(0) = B, \quad B\left(\frac{\pi}{2}\right) = U^{-1} B U. \tag{23}$$

$B(t)$ satisfies the "equation of motion"

$$\dot{B}(t) = [B(t), F]_-, \tag{24}$$

where the dot denotes the derivative with respect to t. The equation of motion techniques are based on systematic methods of solving these equations of motion subject to the initial conditions $B(0) = B$. Evaluation of the solutions at "time" $t = \pi/2$ then yields the desired transforms.

4. d-MATRIX METHOD AND NORMAL EXPANSIONS

This method[3] is based on operator basis expansions of the equations of motion. Define an *operator basis* $\{B_i\}$ to be a set of linearly independent operators on the state space, i.e., a set of operators such that no linear combination of these operators with c-number coefficients is the zero operator on the given state space. It is furthermore assumed that this basis is sufficiently complete to allow expansion (with c-number coefficients) of every operator of physical interest. In practice this implies that the set $\{B_i\}$ must be closed under commutation with F. The operators $B_i(t)$ involved in the equations of motion (24) are expanded as

$$B_i(t) = \sum_j c_{ij}(t) B_j, \tag{25}$$

with c-number coefficients c_{ij} to be determined. The d-matrix (d_{ij}) with respect to the basis B_i is defined by

$$[B_i, F]_- = \sum_j d_{ij} B_j. \tag{26}$$

The d-matrix elements are "structure constants" of the composite particle which exhibit the kinematical relations between the composite and its constituents. Substitution into the equations of motion (24) and use of the linear independence of the different B_i leads to the following coupled equations of motion for the $c(t)$:

$$\dot{c}_{ij}(t) = \sum_k c_{ik}(t) d_{kj}, \tag{27}$$

or in matrix notation $\dot{\mathsf{c}}(t) = \mathsf{c}(t)\mathsf{d}$, with the formal solution

$$\mathsf{c}(t) = \mathsf{c}(0) e^{t\mathsf{d}}. \tag{28}$$

However, exponentiation of the d-matrix is difficult, and it is easier to proceed by direct solution of the coupled differential equations.

A convenient choice of basis elements B_i, sufficiently complete to expand all operators involved in the Hamiltonian, the transformation U, and the projector, is given by the normally ordered products $(a^\dagger)^n a^n$, $(a^\dagger)^n \psi_1^\dagger \psi_1 a^n$, $(a^\dagger)^n \psi_2^\dagger \psi_2 a^n$, $(a^\dagger)^n \psi_1^\dagger \psi_2^\dagger a^{n+1}$, $(a^\dagger)^{n+1} \psi_2 \psi_1 a^n$, and $(a^\dagger)^n \psi_1^\dagger \psi_2^\dagger \psi_2 \psi_1 a^n$. Let i stand for the pair (n,ν) whree $n = 1,2,...$ is the boson index and $\nu = 1,...,6$ the fermion index. The notation is then defined by Table I. The method of evaluation of the d-matrix elements is illustrated by the following example:

$$\begin{aligned}[(a^\dagger)^n a^n, F]_- &= \sum_{n'\nu} d_{n\,1;n'\nu} B_{n'\nu} \\ &= -n(a^\dagger)^{n-1} \psi_1^\dagger \psi_2^\dagger a^n - n(a^\dagger)^n \psi_1 \psi_1 a^{n-1},\end{aligned} \tag{29}$$

which implies

$$\begin{aligned} &d_{n1;n-1,4} = d_{n1;n-1,5} = -n, \\ &d_{n1;n'\nu} = 0, \quad (n'\nu) \neq (n-1,4) \text{ or } (n-1,5). \end{aligned} \tag{30}$$

In evaluating the commutators it is convenient to write

$$\begin{aligned}[(a^\dagger)^n a^n, F]_- &= (a^\dagger)^n a^n \psi_1^\dagger \psi_2^\dagger a - \psi_1^\dagger \psi_2^\dagger a (a^\dagger)^n a^n \\ &\quad - (a^\dagger)^n a^n a^\dagger \psi_2 \psi_1 + a^\dagger \psi_2 \psi_1 (a^\dagger)^n a^n,\end{aligned} \tag{31}$$

and to make use of the identities

$$\begin{aligned} a(a^\dagger)^n &= n(a^\dagger)^{n-1} + (a^\dagger)^n a, \\ a^n a^\dagger &= n a^{n-1} + a^\dagger a^n, \end{aligned} \tag{32}$$

leading immediately to (30). Evaluating the other elements similarly, one finds the d-matrix shown in Figs. 1 and 2. The blocks shown in Fig. 1 are the 6×6 submatrices of $d_{n\nu;n'\nu'}$ with fixed values of n and n' but with ν and ν' running from 1 to 6. Note the tridiagonal block structure; only the blocks $d_{n\nu;n'\nu'}$, $d_{n\nu;n+1,\nu'}$, and $d_{n\nu;n-1,\nu'}$ are nonzero. The elements of the nonzero blocks are exhibited in Fig. 2.

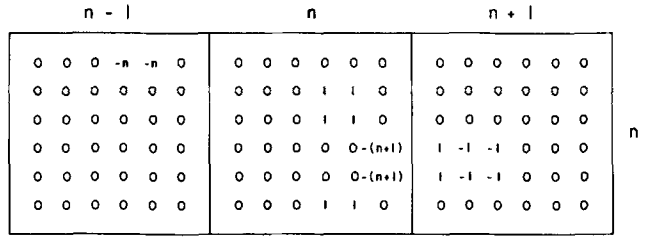


FIG. 2. The $(n\nu)$ rows of the d-matrix for $\nu = 1,...,6$ and a single value of n. All nonzero elements lie in the columns with $n' = n-1$, n or $n+1$.

In terms of the $(n\nu)$ notation the equations of motion (27) are

$$\dot{c}_{n\nu;n'\nu'}(t) = \sum_{n''\nu''} c_{n\nu;n''\nu''}(t) d_{n''\nu'';n'\nu'}. \tag{33}$$

The initial condition is obtained from (22), which implies $B_{n\nu}(0) = B_{n\nu}$ and hence, by (25),

$$c_{n\nu;n'\nu'}(0) = \delta_{nn'}\delta_{\nu\nu'}, \tag{34}$$

where δ is the Kronecker delta function. It follows from Fig. 2 that

$$\dot{c}_{n\nu;01}(t) = \dot{c}_{n\nu;02}(t) = \dot{c}_{n\nu;03}(t) = 0,$$

$$c_{n\nu;01}(t) = \delta_{n0}\delta_{\nu 1}, \quad c_{n\nu;02}(t) = \delta_{n0}\delta_{\nu 2}, \tag{35}$$
$$c_{n\nu;03}(t) = \delta_{n0}\delta_{\nu 3}.$$

Similarly, Fig. 2 and Eqs. (35) imply

$$\begin{aligned}
\dot{c}_{n\nu;04}(t) &= \dot{c}_{n\nu;05}(t) \\
&= c_{n\nu;02}(t) + c_{n\nu;03}(t) + c_{n\nu;06}(t) - c_{n\nu;11}(t) \\
&= \delta_{n0}\delta_{\nu 2} + \delta_{n0}\delta_{\nu 3} + c_{n\nu;06}(t) - c_{n\nu;11}(t), \\
\dot{c}_{n\nu;06}(t) &= -c_{n\nu;04}(t) - c_{n\nu;05}(t) \\
\dot{c}_{n\nu;11}(t) &= c_{n\nu;04}(t) + c_{n\nu;05}(t).
\end{aligned} \tag{36}$$

These four equations can be written in matrix form as follows:

$$\frac{d}{dt}\begin{bmatrix} c_{n\nu;04}(t) \\ c_{n\nu;05}(t) \\ c_{n\nu;06}(t) \\ c_{n\nu;11}(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_{n\nu;04}(t) \\ c_{n\nu;05}(t) \\ c_{n\nu;06}(t) \\ c_{n\nu;11}(t) \end{bmatrix} + \begin{bmatrix} \delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3}) \\ \delta_{n0}(\delta_{\nu_2 + \nu 3}) \\ 0 \\ 0 \end{bmatrix}. \tag{37}$$

The general solution of this inhomogeneous equation is the sum of the general solution of the homogeneous equation and a particular solution of the inhomogeneous equation. One choice of the particular solution is

$$\begin{aligned}
&c_{n\nu;04}{}^{P}(t) = c_{n\nu;05}{}^{P}(t) = c_{n\nu;06}{}^{P}(t) = 0, \\
&c_{n\nu;11}{}^{P}(t) = \delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3}).
\end{aligned} \tag{38}$$

The four linearly independent solutions of the homogeneous equation can be taken to be the eigenvectors of the secular equation

$$\begin{vmatrix} -i\omega & 0 & 1 & -1 \\ 0 & -i\omega & 1 & -1 \\ -1 & -1 & -i\omega & 0 \\ 1 & 1 & 0 & -i\omega \end{vmatrix} = 0 \tag{39}$$

with roots $\omega = 0$ (with multiplicity two) and $\omega = \pm 2$. The solutions are linear combinations of the functions $e^{i\omega t}$ with these "frequencies". It is convenient to choose $\cos(2t)$ and $\sin(2t)$ as basis functions rather than $e^{\pm i2t}$. Then one has

$$\begin{aligned}
c_{n\nu;04}(t) &= A_{04} + B_{04}\cos(2t) + C_{04}\sin(2t), \\
c_{n\nu;05}(t) &= A_{05} + B_{05}\cos(2t) + C_{05}\sin(2t), \\
c_{n\nu;06}(t) &= A_{06} + B_{06}\cos(2t) + C_{06}\sin(2t), \\
c_{n\nu;11}(t) &= A_{11} + B_{11}\cos(2t) + C_{11}\sin(2t).
\end{aligned} \tag{40}$$

The particular solution (38) may be considered to be included in A_{11}. The initial conditions on the $c_{n\nu;n'\nu'}$ and their derivatives imply

$$\begin{aligned}
&A_{04} + B_{04} = \delta_{n0}\delta_{\nu 4}, \quad 2C_{04} = \delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}, \\
&A_{05} + B_{05} = \delta_{n0}\delta_{\nu 5}, \quad 2C_{05} = \delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}, \\
&A_{06} + B_{06} = \delta_{n0}\delta_{\nu 6}, \quad 2C_{06} = -\delta_{n0}(\delta_{\nu 4} + \delta_{\nu 5}), \\
&A_{11} + B_{11} = \delta_{n1}\delta_{\nu 1}, \quad 2C_{11} = \delta_{n0}(\delta_{\nu 4} + \delta_{\nu 5}),
\end{aligned} \tag{41}$$

whereas the differential equations imply

$$\begin{aligned}
&A_{06} - A_{11} + \delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3}) = 0, \quad 2B_{04} + C_{06} - C_{11} = 0, \\
&-2C_{04} + B_{06} - B_{11} = 0, \quad 2B_{05} + C_{06} - C_{11} = 0, \\
&-2C_{05} + B_{06} - B_{11} = 0, \quad A_{04} + A_{05} = 0, \\
&2B_{06} - C_{04} - C_{05} = 0, \quad 2C_{06} + B_{04} + B_{05} = 0, \\
&2B_{11} + C_{04} + C_{05} = 0, \quad 2C_{11} - B_{04} - B_{05} = 0.
\end{aligned} \tag{42}$$

Equations (41) and (42) are all satisfied by

$$\begin{aligned}
&A_{04} = -A_{05} = \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4} - \delta_{\nu 5}), \\
&A_{06} = \tfrac{1}{2}\delta_{n0}(\delta_{\nu 6} - \delta_{\nu 2} - \delta_{\nu 3}) + \tfrac{1}{2}\delta_{n1}\delta_{\nu 1}, \\
&A_{11} = \tfrac{1}{2}\delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6}) + \tfrac{1}{2}\delta_{n1}\delta_{\nu 1}, \\
&B_{04} = B_{05} = \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4} + \delta_{\nu 5}), \\
&B_{06} = -B_{11} = \tfrac{1}{2}\delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6}) - \tfrac{1}{2}\delta_{n1}\delta_{\nu 1}, \\
&C_{04} = C_{05} = \tfrac{1}{2}\delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6}) - \tfrac{1}{2}\delta_{n1}\delta_{\nu 1}, \\
&C_{11} = -C_{06} = \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4} + \delta_{\nu 5}).
\end{aligned} \tag{43}$$

The solutions (40) are then

$$\begin{aligned}
C_{n\nu;04}(t) &= \delta_{n0}\delta_{\nu 4}\cos^2 t - \delta_{n0}\delta_{\nu 5}\sin^5 t \\
&\quad + \tfrac{1}{2}[\delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}]\sin(2t), \\
C_{n\nu;05}(t) &= \delta_{n0}\delta_{\nu 5}\cos^2 t - \delta_{n0}\delta_{\nu 4}\sin^2 t \\
&\quad + \tfrac{1}{2}[\delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}]\sin^2(2t), \\
C_{n\nu;06}(t) &= \delta_{n0}\delta_{\nu 6}\cos^2 t + [\delta_{n1}\delta_{\nu 1} - \delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3})]\sin^2 t \\
&\quad - \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4} + \delta_{\nu 5})\sin(2t), \\
C_{n\nu;11}(t) &= \delta_{n1}\delta_{\nu 1}\cos^2 t + \delta_{n0}(\delta_{\nu 2} + \delta_{\nu 3} + \delta_{\nu 6})\sin^2 t \\
&\quad + \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4} + \delta_{\nu 5})\sin(2t).
\end{aligned} \tag{44}$$

This method of solution can be extended to determine, successively, the $c_{n\nu;n'\nu'}$ for $n' = 1,2,\ldots$. The differential equations for $n' \geqslant 1$ are found to be

$$\begin{aligned}
\dot{c}_{n\nu;n'1}(t) &= c_{n\nu;n'-1,4}(t) + c_{n\nu;n'-1,5}(t), \\
\dot{c}_{n\nu;n'2}(t) &= \dot{c}_{n\nu;n'3}(t) = -c_{n\nu;n'-1,4}(t) - c_{n\nu;n'-1,5}(t), \\
\dot{c}_{n\nu;n'4}(t) &= \dot{c}_{n\nu;n'5}(t) \\
&= c_{n\nu;n'2}(t) + c_{n\nu;n'3}(t) + c_{n\nu;n'6}(t) - (n'+1) \\
&\quad c_{n\nu;n'+1,1}(t), \\
\dot{c}_{n\nu;n'6}(t) &= -(n'+1)c_{n\nu;n'4}(t) - (n'+1)c_{n\nu;n'5}(t).
\end{aligned} \tag{45}$$

The second and third of these equations are solved by quadrature using the previously determined solutions for

$c_{n\nu;n'-1,4}$ and $c_{n\nu;n'-1,5}$. The fourth, fifth, sixth, and first (with n' replaced by $n'+1$ in the first) can be solved simultaneously by the matrix method as were Eqs. (36), taking $c_{n\nu;n'2}$ and $c_{n\nu;n'3}$ as known inhomogeneous terms. The secular equation for the homogeneous system is

$$\begin{vmatrix} -i\omega & 0 & 1 & -(n'+1) \\ 0 & -i\omega & 1 & -(n'+1) \\ -(n'+1) & -(n'+1) & -i\omega & 0 \\ 1 & 1 & 0 & -i\omega \end{vmatrix} = 0, \tag{46}$$

with roots $\omega = 0, \pm 2(n'+1)^{1/2}$. In addition, the inhomogeneous terms contribute all the "frequencies" generated previously for smaller values of n', i.e. $\omega = 0, 2, 2\cdot 2^{1/2}, 2\cdot 3^{1/2}, \ldots, 2\cdot(n')^{1/2}$. Hence the solutions for $c_{n\nu;n'4}$, $c_{n\nu;n'5}$, $c_{n\nu;n'6}$, and $c_{n\nu;n'+1,1}$ have "frequencies" $0, 2, 2\cdot 2^{1/2}, 2\cdot 3^{1/2}, \ldots, 2(n'+1)^{1/2}$.

We shall exhibit the solutions obtained in this way for the case $n'=1$, without the details of the derivation. One finds

$$\begin{aligned}
c_{n\nu;12}(t) &= \delta_{n1}\delta_{\nu 2} - \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4}+\delta_{\nu 5})\sin(2t) \\
&\quad - [\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}]\sin^2 t, \\
c_{n\nu;13}(t) &= \delta_{n1}\delta_{\nu 3} - \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4}+\delta_{\nu 5})\sin(2t) \\
&\quad - [\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}]\sin^2 t, \\
c_{n\nu;14}(t) &= \tfrac{1}{2}\delta_{n1}(\delta_{\nu 4}-\delta_{\nu 5}) + \tfrac{1}{2}(\delta_{n0}+\delta_{n1})(\delta_{\nu 4}+\delta_{\nu 5}) \\
&\quad \times\cos(2^{3/2}t) + [2^{-3/2}\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) + 2^{-3/2}\delta_{n1}(\delta_{\nu 2} \\
&\quad + \delta_{\nu 3}+\delta_{\nu 6}-\delta_{\nu 1}) - 2^{-1/2}\delta_{n2}\delta_{\nu 1}]\sin(2^{3/2}t) \\
&\quad - \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4}+\delta_{\nu 5})\cos(2t) \\
&\quad - \tfrac{1}{2}[\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}]\sin(2t), \\
c_{n\nu;15}(t) &= C_{n\nu;14}(t) + \delta_{n1}(\delta_{\nu 5}-\delta_{\nu 4}), \\
c_{n\nu;16}(t) &= \tfrac{1}{2}\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) + \tfrac{1}{2}\delta_{n1}(\delta_{\nu 6}-\delta_{\nu 1}-\delta_{\nu 2}-\delta_{\nu 3}) \\
&\quad + \delta_{n2}\delta_{\nu 1} + [\tfrac{1}{2}\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) \\
&\quad + \tfrac{1}{2}\delta_{n1}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}-\delta_{\nu 1}) - \delta_{n2}\delta_{\nu 1}]\cos(2^{3/2}t) \\
&\quad - 2^{-1/2}(\delta_{n0}+\delta_{n1})(\delta_{\nu 4}+\delta_{\nu 5})\sin(2^{3/2}t) \\
&\quad - [\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}]\cos(2t) \\
&\quad + \delta_{n0}(\delta_{\nu 4}+\delta_{\nu 5})\sin(2t) \\
c_{n\nu;21}(t) &= -\tfrac{1}{4}\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) \\
&\quad + \tfrac{1}{4}\delta_{n1}(\delta_{\nu 1}+\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) + \tfrac{1}{2}\delta_{n2}\delta_{\nu 1} \\
&\quad + [-\tfrac{1}{4}\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) - \tfrac{1}{4}\delta_{n1} \\
&\quad \times(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}-\delta_{\nu 1}) + \tfrac{1}{2}\delta_{n2}\delta_{\nu 1}]\cos(2^{3/2}t) \\
&\quad + 2^{-3/2}(\delta_{n0}+\delta_{n1})(\delta_{\nu 4}+\delta_{\nu 5})\sin(2^{3/2}t) \\
&\quad + \tfrac{1}{2}[\delta_{n0}(\delta_{\nu 2}+\delta_{\nu 3}+\delta_{\nu 6}) - \delta_{n1}\delta_{\nu 1}]\cos(2t) \\
&\quad - \tfrac{1}{2}\delta_{n0}(\delta_{\nu 4}+\delta_{\nu 5})\sin(2t).
\end{aligned} \tag{47}$$

One notes that there are no secular terms in the solutions for the $c_{n\nu;n'\nu'}$ for $n'=0$ and $n'=1$, i.e., t appears only inside trigonometric functions and not algebraically. This result is in fact quite general, i.e., it holds for all of the $c_{n\nu;n'\nu'}$, not merely for those with $n'=0$ and 1. Although a proof could probably be constructed by thd d-matrix method of this section, the proof is simpler in terms of the more powerful method of Secs. 6 and 7 and will therefore be deferred until then. The absence of secular terms in t implies that when we evaluate the transformed Hamiltonian by putting $t=\pi/2$, factors of π will occur only inside trigonometric functions and not algebraically. This is an important result, since other iterative methods of solution can lead to secular terms[2] which would require complication rearrangements and resummations to remove.

5. HAMILTONIAN, PROJECTOR, AND PROJECTED HAMILTONIAN

It follows from Eqs. (2), (23), and (25) and Table I that the Fock–Tani Hamiltonian $\underline{H}$ of Eq. (18) is given, in terms of the solutions $c_{n\nu;n'\nu'}(t)$ evaluated at "time" $\pi/2$, by

$$\underline{H} = \sum_{n'\nu'}\left[\epsilon_1 c_{02;n'\nu'}\left(\frac{\pi}{2}\right) + \epsilon_2 c_{03;n'\nu'}\left(\frac{\pi}{2}\right) + v c_{06;n'\nu'}\left(\frac{\pi}{2}\right)\right]B_{n'\nu'}. \tag{48}$$

The solutions (35), (44), and (47) allow explicit evaluation of all of the terms with $n'=0$ and 1, and also those with $n'=2$, $\nu'=1$, yielding

$$\begin{aligned}
\underline{H} &= \underline{H}_0 + \underline{V} \\
\underline{H}_0 &= \epsilon_a a^\dagger a + \epsilon_1\psi_1^\dagger\psi_1 + \epsilon_2\psi_2^\dagger\psi_2, \\
\underline{V} &= (v-\epsilon_a)\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1 - \epsilon_a a^\dagger\psi_1^\dagger\psi_1 a - \epsilon_a a^\dagger\psi_2^\dagger\psi_2 a \\
&\quad + 2^{-3/2}\sin(2^{1/2}\pi)\epsilon_a\,[a^\dagger\psi_1^\dagger\psi_2^\dagger a^2 + (a^\dagger)^2\psi_2\psi_1 a] \\
&\quad + \tfrac{1}{2}[3+\cos(2^{1/2}\pi)]\epsilon_a a^\dagger\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1 a \\
&\quad - \tfrac{1}{4}[3+\cos(2^{1/2}\pi)]\epsilon_a(a^\dagger)^2 a^2 + \cdots,
\end{aligned} \tag{49}$$

where the terms "..." not exhibited involve the operators $B_{n'\nu'}$ (Table I) with $n'\geqslant 2$, $\nu'=2-6$, as well as the $B_{n'1}$ [i.e., $(a^\dagger)^{n'}a^{n'}$] with $n'\geqslant 3$. Note the decomposition into a free-composite and free-constituent Hamiltonain $\underline{H}_0$ plus an interaction term $\underline{V}$, as in the previous work[1-4] on the Fock–Tani representation. Recall that ϵ_a is the composite particle energy, Eq. (5). Note also that the states $|1)$, $|2)$, and $|a)$ of Eq. (11) are exact eigenstates of $\underline{H}$ [including the omitted terms "..." in (49)] with eigenvalues ϵ_1, ϵ_2, and ϵ_a, respectively. The composite particle is stable (an energy eigenstate) and hence there are no spontaneous breakup and recombination terms $\psi_1^\dagger\psi_2^\dagger a$ and $a^\dagger\psi_2\psi_1$ in $\underline{V}$, a property expressed algebraically by

$$\underline{V}|a) = 0 \tag{50}$$

Again, this is as in the previous work[4]; however, we now have exact, explicit expressions (in this simplified model) for some terms which were previously only evaluated approximately.

The projector $\underline{R}$ onto the physical subspace $\underline{\mathcal{F}}'$ of the Fock–Tani state space $\underline{\mathcal{F}}$ is given, according to (19), (21), (23), (25), and Table I, by

$$\underline{R} = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\sum_{n'\nu'}C_{n1;n'\nu'}\left(\frac{\pi}{2}\right)B_{n'\nu'}. \tag{51}$$

Again evaluating all the terms with $n'=0$, 1 and those with $n'=2$, $\nu'=1$, one finds

$$\begin{aligned}
\underline{R} &= 1 - \psi_1^\dagger\psi_2^\dagger\psi_2\psi_1 - a^\dagger\psi_1^\dagger\psi_1 a - a^\dagger\psi_2^\dagger\psi_2 a \\
&\quad + 2a^\dagger\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1 a - \tfrac{1}{2}(a^\dagger)^2 a^2 + \cdots,
\end{aligned} \tag{52}$$

where the terms "..." not exhibited involve the same operators as in (49). Note that the terms proportional to $\sin(2^{1/2}\pi)$ and $\cos(2^{1/2}\pi)$ have cancelled. The projected Hamiltonian (14) is

$$\underline{\mathscr{H}} = \underline{R}\underline{H} = \underline{H}\underline{R}, \tag{53}$$

since the operator P_0 of Eq. (19) commutes with H, therefore $\underline{R}$ commutes with $\underline{H}$, and $\underline{R}$ is a projector and hence idempotent, $\underline{R}^2 = \underline{R}$. Multiplication of the expressions (52) and (49) for $\underline{R}$ and $\underline{H}$, using Wick's theorem or the commutation and anticommutation relations, leads to the following expression for the leading terms in the normal expansion of $\underline{\mathscr{H}}$:

$$\begin{aligned}\underline{\mathscr{H}} &= \underline{H}_0 + \underline{\mathscr{V}} \\ \underline{\mathscr{V}} &= (\upsilon - \epsilon_a)\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1 - (\epsilon_a + \epsilon_1)a^\dagger\psi_1^\dagger\psi_1 a \\ &\quad - (\epsilon_a + \epsilon_2)a^\dagger\psi_2^\dagger\psi_2 a + (2\epsilon_a - \upsilon)a^\dagger\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1 a \\ &\quad - \epsilon_a(a^\dagger)^2a^2 + \cdots,\end{aligned} \tag{54}$$

in which $\underline{H}_0$ is the same unperturbed Hamiltonian as in (49), whereas $\underline{\mathscr{V}}$ has terms of the same structure as $\underline{V}$ but with different matrix elements. It is interesting to note that $\underline{R}\underline{V}$ vanishes by cancellations between various contributions; all the contributions to both $\underline{H}_0$ and $\underline{\mathscr{V}}$ in (54) come from $\underline{R}\underline{H}_0$.

It is easy to demonstrate explicitly that the expression (54) has the desired projection properties

$$\underline{\mathscr{H}}|1,2) = \underline{\mathscr{H}}|1,a) = \underline{\mathscr{H}}(2,a) = \underline{\mathscr{H}}|1,2,a) = \underline{\mathscr{H}}|a,a) = 0, \tag{55}$$

for the states

$$\begin{aligned}&|1,2) = \psi_1^\dagger\psi_2^\dagger|0), |1,a) = \psi_1^\dagger a^\dagger|0), |2,a) = \psi_2^\dagger a^\dagger|0), \\ &\quad |1,2,a) = \psi_1^\dagger\psi_2^\dagger a^\dagger|0), \quad |a,a) = 2^{-1/2}(a^\dagger)^2|0),\end{aligned} \tag{56}$$

lying in the unphysical subspace $\underline{\mathscr{F}}\Theta\underline{\mathscr{F}}'$ of the Fock–Tani state space $\underline{\mathscr{F}}$. This unphysical subspace is of infinite dimensionality, there being an infinity of other unphysical states also annihilated by $\underline{\mathscr{H}}$ in addition to the ones (55); however, an explicit demonstration would require the terms in $\underline{\mathscr{H}}$ beyond those exhibited in (54). $\underline{\mathscr{H}}$ has the same eigenvalues as H on the physical subspace $\underline{\mathscr{F}}'$ spanned by the four Fock–Tani states $|0)$, $|1)$, $|2)$, ad $|a)$ [Eq. (11)]. In fact, one finds

$$\begin{aligned}&\underline{\mathscr{H}}|0) = \underline{H}|0) = \underline{H}_0|0) = 0, \quad \underline{\mathscr{H}}|1) = \underline{H}|1) \\ &= \underline{H}_0|1) = \epsilon_1|1), \\ &\underline{\mathscr{H}}|2) = \underline{H}|2) = \underline{H}_0|2) = \epsilon_2|2), \quad \underline{\mathscr{H}}|a) = \underline{H}|a) \\ &= \underline{H}_0|a) = \epsilon_a|a).\end{aligned} \tag{57}$$

The physical states are all annihilated by both $\underline{V}$ and $\underline{\mathscr{V}}$:

$$\begin{aligned}&\underline{V}|0) = \underline{\mathscr{V}}|0) = 0, \quad \underline{V}|1) = \underline{\mathscr{V}}|1) = 0, \\ &\underline{V}|2) = \underline{\mathscr{V}}|2) = 0, \quad \underline{V}|a) = \underline{\mathscr{V}}|a) = 0,\end{aligned} \tag{58}$$

since $\underline{H}_0$ incorporates the internal interaction energy of the composite as well as the free-constituent energies.

6. STATISTICAL RENORMALIZATION

The form of the solutions for $\underline{H}$, $\underline{R}$, and $\underline{\mathscr{H}}$, involving increasing powers $(a^\dagger)^n$ to the left and a^n to the right of the six operators $B_{0\nu}$, suggests a different choice of operator basis than the fully normally-ordered one of Table I. The products $(a^\dagger)^n a^n$ are related to the boson number operator $N_a = a^\dagger a$ by the identity

$$\begin{aligned}(a^\dagger)^n a^n &= N_a!/(N_a - n)! \\ &= N_a(N_a - 1)\cdots(N_a - n + 1).\end{aligned} \tag{59}$$

Furthermore, one has the identities

$$aG(N_a) = G(N_a + 1)a, \quad a^\dagger G(N_a) = G(N_a - 1)a^\dagger, \tag{60}$$

for any function G of the operator N_a. It follows that $\underline{H}$, $\underline{R}$, and $\underline{\mathscr{H}}$ can each be represented as a linear combination of the operators $B_{0\nu}$ of Table I, but with coefficients which are functions of N_a rather than c-numbers. We can define a *reduced operator basis* $\{B_\nu\}$ to be a set of operators which are linearly independent in the generalized sense that an equation of the form

$$\sum_\nu G_\nu(N_a)B_\nu = 0, \tag{61}$$

implies that each $G_\nu(N_a)$ vanishes separately. The basis is also required to be sufficiently complete that all operators of interest can be written as linear combination, *with coefficients which may depend upon* N_a, of the B_ν. This process of allowing the coefficients to depend upon the occupation number operator N_a is what is meant by the terminology statistical renormalization; the physical significance will become clearer after we have evaluated the closed-form expressions for $\underline{H}$ and $\underline{\mathscr{H}}$. The above statement that the basis $\{B_\nu\}$ must be "sufficiently complete..." means, in our application to the evaluation of Tani transforms (22), that the set $\{B_\nu\}$ must be closed under commutation with F, but again in the generalized sense that the coefficients in the linear combinations are allowed to depend upon N_a. Such bases are easy to construct. Starting with a set of operators B_ν whose Tani transforms are desired and which are linearly independent in the renormalized sense (61), one evaluates their commutators with F, adjoining each new linearly independent operator which arises to the set $\{B_\nu\}$, and continuing until the set closes.

The transforms $\psi_1(t)$ and $\psi_2(t)$ of the constituent field operators are of interest in themselves since they determine the transformation of states from the original Fock space $\underline{\mathscr{F}}$ to the Fock–Tani space $\underline{\mathscr{F}}$. Furthermore, it is clear from (2) that the transformed Hamiltonian is expressible in terms of these fields and their hermitian conjugates $\psi_1^\dagger(t)$ and $\psi_2^\dagger(t)$. Let us therefore first consider the bases appropriate to evaluation of $\psi_1(t)$ and $\psi_2(t)$. For the case of ψ_1, evaluation of commutators with F shows that the four-dimensional basis of Table II is both linearly independent in the sense (61) and complete with respect to $\psi_1(t)$ in the sense that $\psi_1(t)$ is expressible as a linear combination of the given B_ν with coefficients depending upon N_a. The transforms (22) of these basis elements can be expanded as

$$B_\nu(t) = \sum_{\nu'} G_{\nu\nu'}(N_a, t)B_{\nu'}, \tag{62}$$

with coefficients $G_{\nu\nu'}$ to be determined. Note that the $G_{\nu\nu'}$ do not commute with B_2 and B_4, so it is necessary to pay attention to the order. Using the identities

$$B_2G(N_a) = G(N_a + 1)B_2, \quad B_4G(N_a) = G(N_a + 1)B_4, \tag{63}$$

which follow from (60), one can always write operator ex-

TABLE II. Reduced operator basis for $\psi_1(t)$.

ν	B_ν
1	ψ_1
2	$\psi_2^\dagger a$
3	$\psi_2^\dagger\psi_2\psi_1$
4	$\psi_1^\dagger\psi_2^\dagger\psi_1 a$

pressions so that the basis elements B_ν are to the right and their N_a-dependent coefficients to the left, and we shall take this as the standard order, as in (62).

The $B_\nu(t)$ satisfy the equations of motion

$$\dot{B}_\nu(t) = \sum_{\nu'}\dot{G}_{\nu\nu'}(N_a,t)B_{\nu'} = \sum_{\nu'}[G_{\nu\nu'}(N_a,t)B_{\nu'},F]_-. \tag{64}$$

This commutator has to be rewritten as a linear combination of the $B_{\nu''}$ with the $B_{\nu''}$ on the right and their N_a-dependent coefficients on the left. It follows from (15) and (60) that

$$\begin{aligned}[G_{\nu\nu'}(N_a,t)B_{\nu'},F]_- &= G_{\nu\nu'}(N_a,t)B_{\nu'}(\psi_1^\dagger\psi_2^\dagger a - a^\dagger\psi_2\psi_1)\\ &\quad - G_{\nu\nu'}(N_a+1,t)\psi_1^\dagger\psi_2^\dagger aB_{\nu'}\\ &\quad + G_{\nu\nu'}(N_a-1,t)a^\dagger\psi_2\psi_1B_{\nu'},\end{aligned} \tag{65}$$

and the factors to the right of the $G_{\nu\nu'}$ can be written as linear combinations of the $B_{\nu''}$ by use of Wick's theorem or the commutation and anticommutation relations. Define three D-matrices $(D_{\nu\nu'}{}^0)$, $(D_{\nu\nu'}{}^+)$, and $(D_{\nu\nu'}{}^-)$ by

$$B_\nu(\psi_1^\dagger\psi_2^\dagger a - a^\dagger\psi_2\psi_1) = \sum_{\nu'}D_{\nu\nu'}{}^0(N_a)B_{\nu'}, \tag{66}$$

$$-\psi_1^\dagger\psi_2^\dagger aB_\nu = \sum_{\nu'}D_{\nu\nu'}{}^+(N_a)B_{\nu'},$$

$$a^\dagger\psi_2\psi_1B_\nu = \sum_{\nu'}D_{\nu\nu'}{}^-(N_a)B_{\nu'}.$$

Then (64) can be written as

$$\begin{aligned}\sum_{\nu'}\dot{G}_{\nu\nu'}(N_a,t)B_{\nu'} = \sum_{\nu'\nu''}[&G_{\nu\nu'}(N_a,t)D_{\nu'\nu''}{}^0(N_a)\\ &+ G_{\nu\nu'}(N_a+1,t)D_{\nu'\nu''}{}^+(N_a)\\ &+ G_{\nu\nu'}(N_a-1,t)D_{\nu'\nu''}{}^-(N_a)]B_{\nu''}.\end{aligned} \tag{67}$$

Interchanging the summation indices ν' and ν'' on the right and equating like basis elements $B_{\nu'}$, one finds the equations of motion

$$\begin{aligned}\dot{G}_{\nu\nu'}(N_a,t) = \sum_{\nu''}[&G_{\nu\nu''}(N_a,t)D_{\nu''\nu'}{}^0(N_a)\\ &+ G_{\nu\nu''}(N_a+1,t)D_{\nu''\nu'}{}^+(N_a)\\ &+ G_{\nu\nu''}(N_a-1,t)D_{\nu''\nu'}{}^-(N_a)]\end{aligned} \tag{68}$$

or in matrix notation

$$\dot{\mathsf{G}}(N_a,t) = \mathsf{G}(N_a,t)\mathsf{D}^0(N_a) + \mathsf{G}(N_a+1,t)\mathsf{D}^+(N_a) + \mathsf{G}(N_a-1,t)\mathsf{D}^-(N_a). \tag{69}$$

These equations of motion are to be solved subject to the initial conditions

$$G_{\nu\nu'}(N_a,0) = \delta_{\nu\nu'} \tag{70}$$

which follow from (62) and (22).

The explicit expressions for the D-matrices relative to the basis of Table II are easily worked out and are exhibited in Eqs. (71)–(73):

$$(D_{\nu\nu'}{}^0) = \begin{bmatrix}0 & 1 & 0 & 1\\ 0 & 0 & -(N_a+1) & 0\\ 0 & 1 & 0 & 1\\ 0 & 0 & 0 & 0\end{bmatrix}, \tag{71}$$

$$(D_{\nu\nu'}{}^+) = \begin{bmatrix}0 & 0 & 0 & -1\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}, \tag{72}$$

$$(D_{\nu\nu'}{}^-) = \begin{bmatrix}0 & 0 & 0 & 0\\ -N_a & 0 & N_a & 0\\ 0 & 0 & 0 & 0\\ N_a & 0 & -N_a & 0\end{bmatrix}. \tag{73}$$

The explicit forms of the equations of motion (68) are then

$$\begin{aligned}\dot{G}_{\nu1}(N_a,t) &= -N_aG_{\nu2}(N_a-1,t) + N_aG_{\nu4}(N_a-1,t),\\ \dot{G}_{\nu2}(N_a,t) &= G_{\nu1}(N_a,t) + G_{\nu3}(N_a,t),\\ \dot{G}_{\nu3}(N_a,t) &= -(N_a+1)G_{\nu2}(N_a,t) + N_aG_{\nu2}(N_a-1,t)\\ &\quad - N_aG_{\nu4}(N_a-1,t),\\ \dot{G}_{\nu4}(N_a,t) &= G_{\nu1}(N_a,t) + G_{\nu3}(N_a,t) - G_{\nu1}(N_a+1,t).\end{aligned} \tag{74}$$

Differentiating the first equation and substituting from the other, one finds

$$\ddot{G}_{\nu1}(N_a,t) = -N_aG_{\nu1}(N_a,t), \tag{75}$$

of which the solution satisfying the initial conditions

$$\begin{aligned}G_{\nu1}(N_a,0) &= \delta_{\nu1},\\ \dot{G}_{\nu1}(N_a,0) &= -N_aG_{\nu2}(N_a-1,0) + N_aG_{\nu4}(N_a-1,0),\\ &= N_a(\delta_{\nu4}-\delta_{\nu2})\end{aligned} \tag{76}$$

is

$$G_{\nu1}(N_a,t) = \delta_{\nu1}\cos(N_a^{1/2}t) + (\delta_{\nu4}-\delta_{\nu2})N_a^{1/2}\sin(N_a^{1/2}t). \tag{77}$$

Similarly differentiating the equation for $G_{\nu2}$, one finds

$$\ddot{G}_{\nu2}(N_a,t) = -(N_a+1)G_{\nu2}(N_a,t), \tag{78}$$

of which the solution satisfying the initial conditions is

$$\begin{aligned}G_{\nu2}(N_a,t) = \delta_{\nu2}&\cos[(N_a+1)^{1/2}t]\\ &+ (\delta_{\nu1}+\delta_{\nu3})(N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}t].\end{aligned} \tag{79}$$

The equations for $\dot{G}_{\nu1}$ and $\dot{G}_{\nu2}$ can then be solved directly for $G_{\nu3}$ and $G_{\nu4}$, giving

$$\begin{aligned}G_{\nu3}(N_a,t) = &-\delta_{\nu2}(N_a+1)^{1/2}\sin[(N_a+1)^{1/2}t]\\ &+ (\delta_{\nu1}+\delta_{\nu3})\cos[(N_a+1)^{1/2}t]\\ &- \delta_{\nu1}\cos(N_a^{1/2}t) + (\delta_{\nu2}-\delta_{\nu4})N_a^{1/2}\sin(N_a^{1/2}t),\\ G_{\nu4}(N_a,t) = &\,\delta_{\nu3}(N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}t]\\ &+ \delta_{\nu4}\cos[(N_a+1)^{1/2}t].\end{aligned} \tag{80}$$

TABLE III. Reduced operator basis for $\psi_2(t)$.

ν	B_ν
1	ψ_2
2	$\psi_1^\dagger a$
3	$\psi_1^\dagger\psi_2\psi_1$
4	$\psi_1^\dagger\psi_2^\dagger\psi_2 a$

The desired Tani transform $\psi_1(t)$ is, in the notation of Table II, equal to $B_1(t)$, and one then has by Eq. (61)

$$\psi_1(t) = \cos(N_a^{1/2}t)\psi_1 + (N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}t\,]\times\psi_2^\dagger a + \{\cos[(N_a+1)^{1/2}t\,] - \cos(N_a^{1/2}t)\}\psi_2^\dagger\psi_2\psi_1. \tag{81}$$

The derivation of the explicit expression for $\psi_2(t)$ proceeds in the same way. The appropriate choice of basis is that of Table III. The derivation differs from that of (81) only in notation and a few sign changes, and we shall only give the result:

$$\psi_2(t) = \cos(N_a^{1/2}t)\psi_2 - (N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}t\,]\psi_1^\dagger a - \{\cos[(N_a+1)^{1/2}t\,] - \cos(N_a^{1/2}t)\}\psi_1^\dagger\psi_2\psi_1. \tag{82}$$

The power of the method of statistical renormalization should now be evident. We have found closed-form expressions for $\psi_1(t)$ and $\psi_2(t)$ by calculations which are almost trivial (once the appropriate mathematical framework has been erected), whereas the method of Sec. 4, based on completely normally ordered expansions, would lead to an infinite series of terms whose complexity increases rapidly with order, such that the general term could probably not be found at all. Indeed, if one attempts to derive the normal expansions by reordering of the functions of N_a appearing in Eqs. (81) and (82), one rapidly finds that it is a formidable algebraic problem.

The transformed Hamiltonian $\underline{H}$ of Eq. (18) is given according to (2) by

$$\underline{H} = \epsilon_1\underline{\psi}_1^\dagger\underline{\psi}_1 + \epsilon_2\underline{\psi}_2^\dagger\underline{\psi}_2 + v\underline{\psi}_1^\dagger\underline{\psi}_2^\dagger\underline{\psi}_2\underline{\psi}_1, \tag{83}$$

with

$$\underline{\psi}_1 = U^{-1}\psi_1 U, \quad \underline{\psi}_2 = U^{-1}\psi_2 U. \tag{84}$$

According to (22) and (23) these transforms are obtained by evaluating $\psi_1(t)$ and $\psi_2(t)$ at "time" $\pi/2$, yielding

$$\underline{\psi}_1 = \cos\left(N_a^{1/2}\frac{\pi}{2}\right)\psi_1 + (N_a+1)^{-1/2}\times\sin\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]\psi_2^\dagger a + \left\{\cos\left[(N_a+1)^{1/2}\frac{\pi}{2}\right] - \cos\left(N_a^{1/2}\frac{\pi}{2}\right)\right\}\psi_2^\dagger\psi_2\psi_1,$$

$$\underline{\psi}_2 = \cos\left(N_a^{1/2}\frac{\pi}{2}\right)\psi_2 - (N_a+1)^{-1/2}\times\sin\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]\psi_1^\dagger a - \left\{\cos\left[(N_a+1)^{1/2}\frac{\pi}{2}\right] - \cos\left(N_a^{1/2}\frac{\pi}{2}\right)\right\}\psi_1^\dagger\psi_2\psi_1. \tag{85}$$

It is amusing to note that the operators $\underline{\psi}_1$, $\underline{\psi}_2$, $\underline{\psi}_1^\dagger$, and $\underline{\psi}_2^\dagger$ are indeed Fermi operators, i.e., they satisfy the elementary fermion anticommutation relations (1) even though they involve boson operators a, $a^\dagger$ via $N_a = a^\dagger a$; the proof involves use of (60) and trigonometric identities. The expression for $\underline{H}$ is found by substitution of (85) into (83) and use of (60) and the commutation and anticommunication relations; we shall only exhibit the result:

$$\begin{aligned}\underline{H} = {}& \epsilon_a\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right) + \left[\epsilon_1 - \epsilon_a\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)\right]\psi_1^\dagger\psi_1 \\ & + \left[\epsilon_2 - \epsilon_a\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)\right]\psi_2^\dagger\psi_2 \\ & + \tfrac{1}{2}\epsilon_a[N_a^{-1/2}\sin(N_a^{1/2}\pi)a^\dagger\psi_2\psi_1 \\ & + \psi_1^\dagger\psi_2^\dagger a N_a^{-1/2}\sin(N_a^{1/2}\pi)] \\ & + \left\{v + \epsilon_a\cos^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]\right. \\ & \left. - \epsilon_a\cos^2\left(N_a^{1/2}\frac{\pi}{2}\right)\right\}\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1.\end{aligned} \tag{86}$$

This is the exact expression of which (49) is the first few terms of the normal expansion. It is clear (at least in this model) that it is great advantage to leave the functions of the occupation number in their given forms and not to reorder them into normal expansions.

To evaluate the projected Hamiltonian $\mathscr{H}$ we need the projector $\underline{R}$ of Eqs. (19) and (20). In principle, $\exp(i\underline{N}_a\vartheta)$ could be evaluated by the D-matrix method (using a different choice of operator basis) but it is more efficient to make use of the previously obtained results in the evaluation. It is easy to show either algebraically or from the physical interpretation that the operators $N_a + \psi_1^\dagger\psi_1$ and $N_a + \psi_2^\dagger\psi_2$ commute with F; hence

$$\begin{aligned}\underline{N}_a &= N_a\left(\frac{\pi}{2}\right) = N_a + \psi_1^\dagger\psi_1 - \underline{\psi}_1^\dagger\underline{\psi}_1 \\ &= N_a + \psi_2^\dagger\psi_2 - \underline{\psi}_2^\dagger\underline{\psi}_2.\end{aligned} \tag{87}$$

Using the expression (85) for $\underline{\psi}_1$ or $\underline{\psi}_2$, one finds after algebraic reductions

$$\begin{aligned}\underline{N}_a = {}& N_a - \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right) + \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)(\psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2) \\ & - \tfrac{1}{2}N_a^{-1/2}\sin(N_a^{1/2}\pi)a^\dagger\psi_2\psi_1 \\ & - \tfrac{1}{2}(N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}\pi]\psi_1^\dagger\psi_2^\dagger a \\ & + \tfrac{1}{2}\{\cos(N_a^{1/2}\pi) - \cos[(N_a+1)^{1/2}\pi]\}\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1.\end{aligned} \tag{88}$$

We must next exponentiate this expression to find $\exp(i\underline{N}_a\vartheta)$. This can again be done by an equation of motion method. By forming products and powers of the various

TABLE IV. Reduced operator basis for $\exp(i\underline{N}_a\vartheta)$.

ν	β_ν
2	$\psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2$
3	$a^\dagger\psi_2\psi_1$
4	$\psi_1^\dagger\psi_2^\dagger a$
5	$\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1$

terms in (88), it is easy to show that $\exp(i\underline{N}_a\vartheta)$ is expressible in the form

$$\exp(i\underline{N}_a\vartheta) = \sum_{\nu=1}^{5} G_\nu(N_a,\vartheta)B_\nu, \tag{89}$$

where the basis elements B_ν are given in Table IV. Differentiating (89) with respect to ϑ, one finds

$$\begin{aligned}\sum_\nu \dot{G}_\nu(N_a\vartheta)B_\nu &= i\underline{N}_a\exp(i\underline{N}_a\vartheta)\\ &= \sum_\nu i\underline{N}_a G_\nu(N_a,\vartheta)B_\nu,\end{aligned} \tag{90}$$

where the dot now represents differentiation with respect to ϑ. Upon substituting (88), performing the algebraic reductions necessary to represent the right side as a linear combination (with N_a-dependent coefficients) of the basis elements B_ν, and equating coefficients of like basis elements, one obtains the following differential equations for the G_ν:

$$\begin{aligned}
\dot{G}_1(N_a,\vartheta) &= i\left[N_a - \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)\right]G_1(N_a,\vartheta)\\
&\quad - \tfrac{1}{2}iN_a^{1/2}\sin(N_a^{1/2}\pi)G_4(N_a-1,\vartheta),\\
\dot{G}_2(N_a,\vartheta) &= iN_aG_2(N_a,\vartheta) + i\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)G_1(N_a,\vartheta)\\
&\quad + \tfrac{1}{2}iN_a^{1/2}\sin(N_a^{1/2}\pi)G_4(N_a-1,\vartheta),\\
\dot{G}_3(N_a,\vartheta) &= i\left[N_a - \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)\right]G_3(N_a,\vartheta)\\
&\quad - \tfrac{1}{2}iN_a^{-1/2}\sin(N_a^{1/2}\pi)[G_1(N_a-1,\vartheta)+G_5(N_a-1,\vartheta)]\\
&\quad - iN_a^{-1/2}\sin(N_a^{1/2}\pi)G_2(N_a-1,\vartheta),\\
\dot{G}_4(N_a,\vartheta) &= i\left\{N_a + \sin^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]\right\}G_4(N_a,\vartheta)\\
&\quad - \tfrac{1}{2}i(N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}\pi]G_1(N_a+1,\vartheta),\\
\dot{G}_5(N_a,\vartheta) &= i\left\{N_a + \sin^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]\right\}G_5(N_a,\vartheta)\\
&\quad + 2i\sin^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]G_2(N_a,\vartheta)\\
&\quad + \tfrac{1}{2}i\{\cos(N_a^{1/2}\pi) - \cos[(N_a+1)^{1/2}\pi]\}G_1(N_a,\vartheta)\\
&\quad - \tfrac{1}{2}iN_a^{1/2}\sin(N_a^{1/2}\pi)G_4(N_a-1,\vartheta)\\
&\quad - \tfrac{1}{2}i(N_a+1)^{1/2}\sin[(N_a+1)^{1/2}\pi]G_3(N_a+1,\vartheta).
\end{aligned} \tag{91}$$

These are to be solved subject to the initial conditions

$$G_1(N_a,0) = 1; \quad G_\nu(N_a,0) = 0, \quad \nu \geqslant 2, \tag{92}$$

following from (89) and Table IV. Addition of the first two equations gives

$$\begin{aligned}\frac{d}{d\vartheta}[G_1(N_a,\vartheta)+G_2(N_a,\vartheta)]\\ = iN_a[G_1(N_a,\vartheta)+G_2(N_a,\vartheta)],\end{aligned} \tag{93}$$

whose solution subject to the initial conditions is

$$G_1(N_a,\vartheta) + G_2(N_a,\vartheta) = e^{iN_a\vartheta}, \tag{94}$$

which can be used to eliminate G_2 from the third and fifth equations. The first and fourth equations can be combined into a single matrix equation

$$\frac{d}{d\vartheta}\begin{bmatrix}G_1(N_a,\vartheta)\\ G_4(N_a-1,\vartheta)\end{bmatrix} = i\mathsf{M}(N_a)\begin{bmatrix}G_1(N_a,\vartheta)\\ G_4(N_a-1,\vartheta)\end{bmatrix}, \tag{95}$$

where the elements of the matrix M are

$$\begin{aligned}
M_{11} &= N_a - \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right),\\
M_{12} &= -\tfrac{1}{2}N_a^{1/2}\sin(N_a^{1/2}\pi),\\
M_{21} &= -\tfrac{1}{2}N_a^{-1/2}\sin(N_a^{1/2}\pi),\\
M_{22} &= N_a - 1 + \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right).
\end{aligned} \tag{96}$$

The solution vector is a linear combination of the two eigenvectors of M. The eigenvalues are the solutions of the secular equation

$$[M_{11}(N_a)-\lambda\,][M_{22}(N_a)-\lambda\,] - M_{12}(N_a)M_{21}(N_a) = 0, \tag{97}$$

of which the roots are

$$\Lambda_+(N_a) = N_a, \quad \Lambda_-(N_a) = N_a - 1. \tag{98}$$

The determination of the solution satisfying the initial conditions (92) is then straightforward; we give only the result:

$$\begin{aligned}
G_1(N_a,\vartheta) &= \cos^2\left(N_a^{1/2}\frac{\pi}{2}\right)e^{iN_a\vartheta}\\
&\quad + \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)e^{i(N_a-1)\vartheta},\\
G_4(N_a,\vartheta) &= \tfrac{1}{2}(N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}\pi]\\
&\quad \times[e^{iN_a\vartheta} - e^{i(N_a+1)\vartheta}\,].
\end{aligned} \tag{99}$$

Then by (94)

$$G_2(N_a,\vartheta) = \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)[e^{iN_a\vartheta} - e^{i(N_a-1)\vartheta}\,]. \tag{100}$$

The third and fifth equations (91) can also be solved by the matrix method, taking G_1 and G_2 as known inhomogeneous terms; we again give only the result:

$$\begin{aligned}
G_3(N_a,\vartheta) &= \tfrac{1}{2}N_a^{-1/2}\sin(N_a^{1/2}\pi)[e^{i(N_a-1)\vartheta} - e^{iN_a\vartheta}\,],\\
G_5(N_a,\vartheta) &= \sin^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]e^{i(N_a+1)\vartheta}\\
&\quad - \left\{\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right) + \sin^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]\right\}\\
&\quad \times e^{iN_a\vartheta} + \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)e^{i(N_a-1)\vartheta}.
\end{aligned} \tag{101}$$

The final expression for $\exp(i\underline{N}_a\vartheta)$ given by Eq. (89) and Table IV is then

$$\begin{aligned}
&\exp(i\underline{N}_a\vartheta) = \cos^2\left(N_a^{1/2}\frac{\pi}{2}\right)e^{iN_a\vartheta} + \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)e^{i(N_a-1)\vartheta}\\
&+ \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)[e^{iN_a\vartheta} - e^{i(N_a-1)\vartheta}\,](\psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2)\\
&+ \tfrac{1}{2}N_a^{-1/2}\sin(N_a^{1/2}\pi)[e^{i(N_a-1)\vartheta} - e^{iN_a\vartheta}\,]a^\dagger\psi_2\psi_1\\
&+ \tfrac{1}{2}(N_a+1)^{-1/2}\sin[(N_a+1)^{1/2}\pi]\,[e^{iN_a\vartheta} - e^{i(N_a+1)\vartheta}\,]\psi_1^\dagger\psi_2^\dagger a\\
&+ \left\{\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)e^{i(N_a-1)\vartheta} + \sin^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]e^{i(N_a+1)\vartheta}\right.
\end{aligned}$$

$$- \sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)e^{iN_a\vartheta} - \sin^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]e^{iN_a\vartheta}\Big\}\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1. \tag{102}$$

The projectors $\underline{R}$ can now be evaluated by performing the integrals over ϑ, using the representation

$$\underline{R} = (2\pi)^{-1}\int_0^{2\pi} d\vartheta\, \exp(i\underline{N}_a\vartheta) \tag{103}$$

following from (19) and (20). Equation (20) generalizes to

$$\underline{P}_n = (2\pi)^{-1}\int_0^{2\pi} d\vartheta\, e^{i(N_a - n)\vartheta}, \tag{104}$$

where $\underline{P}_n$ is the projector onto the subspace of $\mathcal{F}$ spanned by eigenstates of N_a with eigenvalue n. One has the obvious identity

$$\underline{P}_n G(N_a) = G(N_a)\underline{P}_n = G(n)\underline{P}_n. \tag{105}$$

Furthermore, there exist on eigenstates of N_a with negative eigenvalues, so the terms in (102) involving $e^{i(N_a+1)\vartheta}$ integrate to zero. One then finds a very simple expression

$$\underline{R} = \underline{P}_0 + \underline{P}_1 - \underline{P}_1(\psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2) + (\underline{P}_1 - \underline{P}_0)\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1. \tag{106}$$

Upon multiplying this by (85), applying the identities (105), one finds the following expression for the projected Hamiltonian

$$\underline{\mathscr{H}} = \epsilon_a\underline{P}_1 a^\dagger a + (\epsilon_1\underline{P}_0 - \epsilon_a\underline{P}_1)\psi_1^\dagger\psi_1 + (\epsilon_2\underline{P}_0 - \epsilon_a\underline{P}_1)\psi_2^\dagger\psi_2 + [(v-\epsilon_a)\underline{P}_0 + \epsilon_a\underline{P}_1]\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1. \tag{107}$$

In view of Eq. (105) the factor $a^\dagger a = N_a$ in the first term could have been omitted; however, it is included in order to facilitate the physical interpretation which will be discussed in the next section. The very simple expression (107) for $\underline{\mathscr{H}}$ could have been obtained without all the elaborate machinery of the Tani transformation, D-matrix, etc... in this simple model. However, for realistic systems trivial representations such as (14) are not available, and the Fock–Tani representation has definite physical and mathematical advantages which have already been discussed.[1-4] The methods that have been illustrated here ought to be applicable to realistic systems as well, although one can hardly expect closed-form solutions in the general case.

7. STATISTICALLY RENORMALIZED HAMILTONIAN

The physical interpretation and generalization to realistic systems require a more explicit expression than (107), in which the projectors $\underline{P}_0$ and $\underline{P}_1$ are expressed as functions of N_a. The most convenient representation for our purposes is obtained by changing the limits of integration in (104) to $[-\pi,\pi]$, taking advantage of the fact that N_a has only integral eigenvalues. The integration is then trivial and yields the representation

$$\underline{P}_n = j_0((N_a - n)\pi), \tag{108}$$

where j_0 is the spherical Bessel function of order zero, $j_0(x) = \sin x/x$. Since

$$j_0(0) = 1;\quad j_0(n\pi) = 0,\quad n = \pm 1, \pm 2, \cdots, \tag{109}$$

the desired projection properties of the representation (108) are evident. One then finds the following expression for the projected Hamiltonian (107)

$$\underline{\mathscr{H}} = \epsilon_a(N_a)a^\dagger a + \epsilon_1(N_a)\psi_1^\dagger\psi_1 + \epsilon_2(N_a)\psi_2^\dagger\psi_2 + v(N_a)\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1, \tag{110}$$

where $\epsilon_a(N_a)$, $\epsilon_1(N_a)$, and $\epsilon_2(N_a)$ are statistically renormalized single-particle energies and similarly $v(N_a)$ is a statistically renormalized interaction, defined by

$$\begin{aligned}
\epsilon_a(N_a) &= \epsilon_a j_0((N_a-1)\pi),\\
\epsilon_1(N_a) &= \epsilon_1 j_0(N_a\pi) - \epsilon_a j_0((N_a-1)\pi),\\
\epsilon_2(N_a) &= \epsilon_2 j_0(N_a\pi) - \epsilon_a j_0((N_a-1)\pi),\\
v(N_a) &= (v-\epsilon_a)j_0(N_a\pi) + \epsilon_a j_0((N_a-1)\pi).
\end{aligned} \tag{111}$$

It was noted in Secs. 4 and 5 that secular terms (powers $\pi,\pi^2,\cdots$) do not appear in the normal expansions for either $\underline{H}$ or $\underline{\mathscr{H}}$ to the orders exhibited in Eqs. (44) and (54). This result is in fact valid in arbitrary order, as can be seen by noting how the normal expansions can be generated starting from the statistically renormalized expressions (86) and (110). The proof is simpler for $\underline{\mathscr{H}}$. It follows from (110) and (111) that the normal expansion for $\underline{\mathscr{H}}$ can be obtained immediately once those of $j_0(N_a\pi)$ and $j_0((N_a-1)\pi)$ are known. That of $j_0(N_a\pi)$ is already known from Eq. (21)

$$j_0(N_a\pi) = \underline{P}_0 = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}(a^\dagger)^n a^n. \tag{112}$$

The normal expansion of $\underline{P}_1$ can be derived by the same method[2] as was (112). One has

$$j_0((N_a-1)\pi) = \underline{P}_1 = \sum_{n=0}^{\infty} c_n (a^\dagger)^n a^n, \tag{113}$$

with coefficients c_n to be determined. Letting $|n)$ be the eigenstate const. $(a^\dagger)^n|0)$ of N_a with eigenvalue n, one has successively, with (59),

$$\begin{aligned}
&\underline{P}_1|0) = c_0|0) = 0,\\
&\underline{P}_1|1) = c_1|1) = |1),\quad c_1 = 1,\\
&\underline{P}_1|2) = (2+2c_2)|2) = 0,\quad c_2 = -1,\\
&\underline{P}_1|3) = (3-6+6c_3)|3) = 0,\quad c_3 = \tfrac{1}{2},\\
&\cdot\\
&\cdot\\
&\cdot\\
&c_n = 1(n-1)!,
\end{aligned} \tag{114}$$

and hence

$$j_0((N_a-1)\pi) = \underline{P}_1 = \sum_{n=1}^{\infty}\frac{(-1)^n}{(n-1)!}(a^\dagger)^n a^n. \tag{115}$$

Substitution of (112) and (115) into (110) and (111), noting that $j_0((N_a-1)\pi)a^\dagger a = j_0((N_a-1)\pi)$, yields the complete normal expansion of $\underline{\mathscr{H}}$, of which only the first few terms were exhibited in Eq. (54)

$$\underline{\mathscr{H}} = \underline{\mathscr{H}}_0 + \underline{\mathscr{V}},$$

$$\underline{\mathscr{V}} = \sum_{n=0}^{\infty}\frac{(-1)^{n+1}}{n!}[(n+1)\epsilon_a - v](a^\dagger)^n\psi_1^\dagger\psi_2^\dagger\psi_2\psi_1 a^n + \sum_{n=1}^{\infty}\frac{(-1)^n}{n!}(\epsilon_1 + n\epsilon_a)(a^\dagger)^n\psi_1^\dagger\psi_1 a^n$$

$$+\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}(\epsilon_2+n\epsilon_a)(a^\dagger)^n\psi_2^\dagger\psi_2 a^n$$
$$+\epsilon_a\sum_{n=2}^{\infty}\frac{(-1)^{n-1}}{(n-1)!}(a^\dagger)^n a^n. \tag{116}$$

Here H_0 is the unperturbed Fock–Tani Hamiltonian (49). No secular terms (powers of π) appear in the expression (116). Note how easy it is to obtain the general term in the normal expansion once the (nontrivial) task of determination of the statistically renormalized expression for $\mathscr{H}$ has been accomplished. Note also that the statistically renormalized expression is much simpler, both in mathematical form and in physical interpretation, than the completely normally ordered expression (116).

The determination of the normal expansion for the unprojected Hamiltonian H can be done in the same way, although the algebra is more complicated. According to (86) one needs the normal expansions of $\sin^2(N_a^{1/2}\pi/2)$, $\cos^2(N_a^{1/2}\pi/2)$, $N_a^{-1/2}\sin(N_a^{1/2}\pi)$, and $\cos^2[(N_a+1)^{1/2}\pi/2]$. These can be determined as in (114), (115); one finds

$$\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right)=a^\dagger a+\left[-\tfrac{3}{4}-\tfrac{1}{4}\cos(2^{1/2}\pi)\right](a^\dagger)^2a^2$$
$$+\left[\tfrac{1}{3}+\tfrac{1}{24}\cos(2^{1/2}\pi)-\tfrac{1}{12}\cos(3^{1/2}\pi)\right](a^\dagger)^3a^3+\cdots,$$
$$\cos^2\left(N_a^{1/2}\frac{\pi}{2}\right)=1-\sin^2\left(N_a^{1/2}\frac{\pi}{2}\right),$$
$$N_a^{-1/2}\sin(N_a^{1/2}\pi)=2^{-3/2}\sin(2^{1/2}\pi)(a^{\dagger 2}a^2)$$
$$+\left[-2^{-3/2}\sin(2^{1/2}\pi)+\tfrac{1}{2}\cdot 3^{-3/2}\sin(3^{1/2}\pi)\right]$$
$$\times(a^\dagger)^3a^3+\cdots,$$
$$\cos^2\left[(N_a+1)^{1/2}\frac{\pi}{2}\right]=\left[\tfrac{1}{2}+\tfrac{1}{2}\cos(2^{1/2}\pi)\right]a^\dagger a$$
$$+\left[-\tfrac{1}{4}-\tfrac{1}{2}\cos(2^{1/2}\pi)+\tfrac{1}{4}\cos(3^{1/2}\pi)\right](a^\dagger)^2a^2$$
$$+\left[\tfrac{1}{12}+\tfrac{1}{2}\cos(2^{1/2}\pi)-\tfrac{1}{4}\cos(3^{1/2}\pi)\right.$$
$$\left.+\tfrac{1}{12}\cos(4^{1/2}\pi)\right](a^\dagger)^3a^3+\cdots. \tag{117}$$

Substitution into (86) gives terms in the normal expansion of H beyond those exhibited in (49). It is clear without attempting to evaluate the general term that no secular terms appear in any order. The statistically renormalized expression (86) is much simpler and preferable to the completely normally ordered expression.

8. LINKED VERSUS UNLINKED EXPRESSIONS

Consider the action of the various terms in the normal expansion (116) of $\mathscr{V}$ upon a state $(a^\dagger)^m|0)$ with a specified number m of bosons. In view of the identity (59), it is clear that for $m\to\infty$, the terms in (116) with index n give energy contributions of order m^n, assuming n fixed as $m\to\infty$. The contributions for $n>2$ are "hyperextensive", i.e., increase more rapidly than linearly with m as $m\to\infty$. This type of unphysical behavior is typical of the unlinked clusters in Rayleigh–Schrödinger perturbation theory of a many-particle system. In fact, if one introduces a diagram notation for the various terms in $\mathscr{V}$, then each term with a factor $(a^\dagger)^n$ on the left and a^n on the right is represented by a diagram containing n free boson lines disconnected from the rest of the diagram. Actually, the unlinked terms all cancel when combined properly so as to pass to the statistically renormalized expression for $\mathscr{H}$. This is clear from (110), in which there are no unlinked terms. Some of the unlinked terms of (116) are absorbed into the definition of the renormalized single-particle energies $\epsilon_a(N_a)$, $\epsilon_1(N_a)$, and $\epsilon_2(N_a)$, while the rest are included in the definition of the renormalized interaction vertex $v(N_a)$. The statistically renormalized expression (110) is therefore analogous to the expressions obtained in linked cluster forms of many-particle perturbation theory. This is still another reason for the superiority of the statistically renormalized representation as compared with the usual normal series. Indeed, it shows that the procedure of statistical renormalization is essential for obtaining physically correct results in the Fock–Tani representation for a system of many composite particles as soon as one goes beyond the lowest-order terms in the normal expansion. The simplified model investigated here does not really allow many-boson states, but these observations should be relevant to more realistic models.

ACKNOWLEDGMENTS

This work was begun while I was a visitor in the Laboratoire de Physique Quantique, Université Paul Sabatier, Toulouse, France and concluded at the Laboratoire de Physique Théorique, Université de Nice, France. I am grateful to these institutions for Visisting Professorships which made this work possible, and particularly to Y. Soulet of the former institution and F. Rocca of the latter.

[1]M. D. Girardeau, Phys. Rev. Lett. **27**, 1416 (1971).
[2]M. D. Girardeau, J. Math. Phys. **16**, 1901 (1975).
[3]J. D. Gilbert, J. Math. Phys. **18**, 791 (1977).
[4]M. D. Girardeau and J. D. Gilbert, Physica A **97**, 42 (1979).
[5]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964), pp. 146 ff.
[6]Y. Soulet, R. Fleckinger, and M. D. Girardeau, Adv. Chem. Phys. (to be published).

A class of explicitly soluble, local, many-center Hamiltonians for one-particle quantum mechanics in two and three dimensions. I

A. Grossmann
Centre de Physique Théorique 2, CNRS, Marseille, France

R. Hoegh-Krohn
University of Oslo and Université d'Aix-Marseille II, Luminy, France

M. Mebkhout
Université d'Aix-Marseille II, Luminy and Centre de Physique Theorique 2, CNRS, Marseille, France

(Received 18 May 1979; accepted for publication 29 August 1979)

We derive an explicit formula for the resolvent of a class of one-particle, many-center, local Hamiltonians. This formula gives, in particular, a full description of a model molecule given by point interactions at n arbitrarily placed fixed centers in three dimensions. It also gives a three-dimensional analog of the Kronig-Penney model.

1. INTRODUCTION

There are, broadly speaking, two kinds of methods for constructing explicitly soluble one-particle quantum-mechanical models. One method is the reduction of the problems to questions involving soluble ordinary differential equations (e.g., by partial wave expansion). This method almost never gives closed explicit formulas in many-center problems, which arise naturally in molecular and solid-state physics.

The second method is essentially finite-dimensional linear algebra. It works if the potentials are separable (i.e., of finite-dimensional range as operators; such operators are also called of finite rank, or degenerate). The drawback of such potentials is that they are nonlocal and have no clear physical meaning.

A well-known, explicitly soluble, local one-dimensional Hamiltonian appears in the Kronig–Penney model[1] (potential consisting of a periodic array of δ-functions in one dimension). The reasons for its solubility are basically

(i) periodicity reduces it to a one-center problem; this is easily seen in momentum representation.[2]

(ii) The "Operator of multiplication" by a δ-function has one-dimensional range (the map $\phi \to \delta_x \phi$ can be written as $\phi \to |x\rangle \langle x|\phi\rangle$).

(iii) In one degree of freedom, the expression $p^2 + |x\rangle \langle x|$ defines a self-adjoint operator in Hilbert space.

In more than one dimension, any attempt to write down a local, separable potential V for the Schrödinger operator must fail, as is shown by the following argument: the combination of separability and locality forces V to be a finite linear combination of δ-functions (and perhaps derivatives of δ-functions). However, an expression of the from $p^2 + |x\rangle \langle x|$ cannot define a reasonable operator in Hilbert space if the number of degrees of freedom is greater than one. (The reason for this difference between dimension one and higher dimensions is well understood; it will be discussed below.)

We shall show in this paper that it is nevertheless possible to construct, in two or three dimensions, a class of quantum-mechanical Hamiltonians which correspond exactly to the idea of one-particle feeling point interactions at n fixed centers, placed arbitrarily (or at an infinite array of such centers). These Hamiltonians are explicitly soluble, in the sense that all questions about bound states, resonances, scattering, etc., are answered by explicit formulas involving only finite-dimensional linear algebra and zeros of known functions.

The class of models that we study includes

(i) "molecules" in two and three dimensions,

(ii) two- and three-dimensional analogs of the Kronig–Penney model, and their perturbations by impurities (see Ref. 3),

(iii) dipoles in one dimension.

There is no contradiction between these results, and the statements about nonexistence of local separable potentials. The Hamiltonians that we study cannot be written in the form $p^2 + V$. (At least not in standard analysis and present-day distribution theory. See Ref. 4, and remarks below.) They will be defined here by an explicit expression for their resolvent or, equivalently, by a "parametric" description of their graph.

This procedure is not new by any means. It can be viewed as a special case of general results of extensions of symmetric operators,[5,6] and abstract boundary value problems.[7] We shall not use these results, but give instead a self-contained elementary presentation, involving nothing more than linear algebra and properties of Sobolev spaces.

The point interactions that we study are, in the one-center situation, a limit of boundary value models.[8] They have been used in the study of two- and three-particle systems (see the excellent review article by Flamand[9]) and are implicit already in early discussions of low-energy nucleon–nucleon scattering.[10]

Some of the Hamiltonians in this paper have been previously studied by methods of nonstandard analysis,[4] and by other means.[11–14] They have been the subject of independent work by L. Thomas.[15]

The remainder of this introduction contains a description of our results in a particular case. Secs. 2 and 3 contain general proofs; Sec. 4 discusses locality, spectral properties,

 0022-2488/80/092376-10$1.00

and some applications. Subsequent papers will deal in detail with periodic potentials and with limits of regular local potentials.

Let x be any point in $\mathbb{R}^3$, and α any real number. Then there exists in the one-particle Hilbert space $\mathscr{H}_0 = L^2(\mathbb{R}^3)$ a self-adjoint operator H_α^x with resolvent $Q_\alpha^x(E) = (H_\alpha^x - E)^{-1}$ given by the formula

$$Q_\alpha^x(E) = R(E) + (\kappa/4\pi + \alpha)^{-1} R(E)|x\rangle \langle x|R(E), \tag{1.1}$$

$$E \in \mathbb{C} \setminus [0\,\infty), \quad E^2 = -\kappa \quad \mathrm{Re}\kappa \geqslant 0, \quad \kappa + 4\pi\alpha \neq 0.$$

Here $R(E)$ is the resolvent of the kinetic energy operator: $R(E) = (T - p^2)^{-1}$, where $(T\psi)(p) = p^2\psi(p)$ in momentum representation. The dyadic $|x\rangle \langle x|$ is, again in momentum representation, the map $\psi(p) \to (2\pi)^{-3} \times (\int e^{-ip'x}\psi(p')d^3p')e^{ipx}$. Notice that $R(E)|x\rangle \langle x|R(E)$ is a bounded operator in $\mathscr{H}_0$, since $(p^2 - E)^{-1}e^{-ipx}$ is square-integrable. The operator can also be defined as the Laplacian with appropriate boundary condition at x.

The negative denominator in (1.1), i.e.,

$$\Gamma_\alpha(E) = -\kappa/4\pi - \alpha \quad (E = -\kappa^2), \tag{1.2}$$

was obtained as the solution of

$$\frac{d\Gamma(E)}{dE} = \langle x|R(E)^2|x\rangle, \tag{1.3}$$

i.e.,

$$\Gamma(E) = \int^E \langle x|R(E')^2|x\rangle \, dE',$$

$-\alpha$ being the arbitrary integration constant. The rhs of (1.3) is given by the absolutely convergent integral $(2\pi)^{-3} \int (p^2 - E)^{-2} d^3p$. Equation (1.3) follows from the fact that, as a resolvent, Q must satisfy $dQ/dE = Q^2$.

The operator H_α^x is the limit as $\omega \to \infty$ (in the sense of norm resolvent convergence) of the family of operators

$$H_\alpha^{x,\omega} = T - \lambda_\alpha^\omega |\Phi_x^\omega\rangle \langle \Phi_x^\omega|, \tag{1.4}$$

where the dyadic $|\Phi_x^\omega\rangle \langle \Phi_x^\omega|$ is the bounded operator given, in momentum representation, by

$$\psi(p) \to (2\pi)^{-3}\left(\int_{|p'|<\omega} e^{-ip'x}\psi(p')\, d^3p'\right)\Theta_\omega(p)e^{ipx} \tag{1.5}$$

($\Theta_\omega(\xi) = 1$ if $\xi \leqslant \omega$, and $\Theta_\omega(\xi) = 0$ if $\xi > \omega$), and where

$$\lambda_\alpha^\omega = [\omega/2\pi^2 - \alpha]^{-1}. \tag{1.6}$$

Notice that (1.6) tends to zero as $\omega \to \infty$ and $|\Phi_x^\omega\rangle \to |x\rangle$. So H_α should not be thought of as a Hamiltonian with a δ-function potential, but rather as a Hamiltonian with an infinitesimal coupling constant in front of a δ-function potential. This description can be made rigorous in the language of nonstandard analysis,[16] and the operator H_α^x has been treated in this way in a preceding paper.[4]

Remark: The classical functions corresponding to the operators (1.4) tend to p^2 in the sense of distributions. It should be interesting to endow spaces of functions in phase space with topologies better adapted to operator convergence and to study the "classical" meaning of the operator H_α^x.

Remark: The operator H_α^x can also be obtained as a strong resolvent limit of Hamiltonians with attractive local short-range potentials: $H_\alpha^0 = \lim_{\epsilon \to 0}[T + \lambda_0\epsilon^{-2} \times (1 + \alpha\epsilon)V_\epsilon]$ where $V_\epsilon(x) = V(\epsilon^{-1}x)$, and where λ_0 is suitably chosen. This limit will be discussed in a forthcoming paper. Notice, again, that $\epsilon^{-2}V_\epsilon = \epsilon\epsilon^{-3}V_\epsilon$ and that $\epsilon^{-3}V_\epsilon$ tends to a δ-function.

If $\alpha < 0$, then the operator H_α^x has a bound state at

$$\kappa_\alpha = -4\pi\alpha, \tag{1.7}$$

i.e., at the energy

$$E_\alpha = -(4\pi)^2\alpha^2. \tag{1.8}$$

If $\alpha > 0$, then H_α^x has a resonance at $\kappa_\alpha = -4\pi\alpha$, since $-4\pi\alpha$ is then in the "unphysical half-plane" $\mathrm{Re}\kappa < 0$.

The operator H_α^x is local, in the following sense: If ϕ is a smooth function (in the x-representation) and if ϕ vanishes in an open set $\mathscr{O}$, then the function $H_\alpha^x\phi$ also vanishes in $\mathscr{O}$.

Two-center Hamiltonians: Let x_1 and x_2 be two points in $\mathbb{R}^3$, and let

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \tag{1.9}$$

be any Hermitian 2×2 matrix. Then there exists in the one-particle Hilbert space $\mathscr{H}_0 = L^2(\mathbb{R}^3)$ a self-adjoint local operator $H_\alpha^{x_1,x_2}$, with resolvent $Q_\alpha^{x_1,x_2}(E) = (H_\alpha^{x_1,x_2} - E)^{-1}$ given by the formula

$$Q_\alpha^{x_1,x_2}(E) = R(E) - \sum_{j=1}^{2}\sum_{l=1}^{2} [\Gamma(E)^{-1}]_{jl} \times R(E)|x_j\rangle \langle x_l|R(E), \tag{1.10}$$

where $\Gamma(E)^{-1}$ is the matrix inverse of the matrix $\Gamma(E) = \Gamma_\alpha^{x_1,x_2}(E)$ given by the formula

$$\Gamma(E) = (4\pi)^{-1} \times \begin{pmatrix} -\kappa & |x_1 - x_2|^{-1}e^{-\kappa|x_1 - x_2|} \\ |x_1 - x_2|^{-1}e^{-\kappa|x_1 - x_2|} & -\kappa \end{pmatrix} - \alpha. \tag{1.11}$$

The entries of $\Gamma(E)$ in (1.11) have been obtained as solutions of

$$\frac{d}{dE}\Gamma_{jl}(E) = \langle x_j|R(E)^2|x_l\rangle, \tag{1.12}$$

where

$$\langle x_j|R(E)^2|x_l\rangle = (2\pi)^{-3}\int (p^2 - E)^{-2}e^{ip(x_l - x_j)}\, d^3p \tag{1.13}$$

is absolutely convergent, and the α's are arbitrary integration constants. If the matrix α in (1.11) is diagonal, we say that the two centers are independent. Again, (1.12) is a consequence of the fact that Q is a resolvent.

The spectral properties of H_{x_1,x_2}^α, discussed in detail in Sec. 4, correspond perfectly to one's expectations based on the intuitive image of a two-center problem: For instance, if the distance between the two centers goes to infinity, and if the two centers are independent, the bound states tend to the corresponding "atomic" states. The resolvent (1.11) exhibits

an infinite set of resonances, that can be visualized as due to the particle's bouncing back and forth between the centers, before escaping to infinity.

The above statements are all special cases of results derived in Secs. 2–4. The basic formulas (1.1), (1.11) follow from the fact that the states $|x\rangle$ belong to the Sobolev space $\mathscr{H}_{-2}$. Locality is made possible by the fact that off-diagonal matrix elements $\langle x_1|R(E)|x_2\rangle$ (with $x_1 \neq x_2$) exist (as convergent, but not absolutely convergent integrals), "as if" $|x\rangle$ belonged to $\mathscr{H}_{-1}$.

The arbitrary constants α_{jl} can be viewed as generalizations of the inverse scattering length of low-energy scattering theory. Here they are not derived from a Hamiltonian, but appear as an input.

2. PERTURBATION OF T BY MILDLY SINGULAR DYADICS

The aims of this section are mainly pedagogical. We introduce the Sobolev spaces $\mathscr{H}_{-1}$ and $\mathscr{H}_{-2}$ and define the natural partial inner product in them. This allows us to use singular dyadics which are a convenient language in our problem. We discuss the relationships with operators in the physical Hilbert space and stress the difference between $\mathscr{H}_{-1}$ and $\mathscr{H}_{-2}$, which is the reason that one-dimensional δ-function "potentials" behave differently from two- and three-dimensional ones.

Consider a quantum-mechanical system of ν degrees of freedom in momentum representation. Its space of states $\mathscr{H}_0$ is the Hilbert space of square-integrable functions

$$\mathscr{H}_0(\mathbb{R}^\nu) = \left\{\phi \,\middle|\, \int |\phi(p)|^2 \, d^\nu p < \infty\right\}. \tag{2.1}$$

The kinetic energy operator T is a multiplication operator by a positive unbounded function $t(p)$. (In most cases of interest, $t(p) = p_1^2 + \cdots + p_\nu^2$ after suitable choice of units.)

$$(T\phi)(p) = t(p)\phi(p) \tag{2.2}$$

Define $\mathscr{H}_1(\mathbb{R}^\nu)$ as the space of states that give finite expectation value to kinetic energy:

$$\mathscr{H}_1(\mathbb{R}^\nu) = \left\{\phi \,\middle|\, \int (t(p)+1)|\phi(p)|^2 \, d^\nu p < \infty\right\}, \tag{2.3}$$

and $\mathscr{H}_2(\mathbb{R}^\nu)$ as the space of states such that the expectation value of the *square* of kinetic energy is finite:

$$\mathscr{H}_2(\mathbb{R}^\nu) = \left\{\phi \,\middle|\, \int (t(p)+1)^2|\phi(p)|^2 \, d^\nu p < \infty\right\}, \tag{2.4}$$

They are called first and second Sobolev space, respectively.

The duals of these spaces will play an important role in our discussions. We define $\mathscr{H}_{-1}$ by

$$\mathscr{H}_{-1}(\mathbb{R}^\nu) = \left\{\phi \,\middle|\, \int (t(p)+1)^{-1}|\phi(p)|^2 \, d^\nu p < \infty\right\}, \tag{2.5}$$

and $\mathscr{H}_{-2}$ by

$$\mathscr{H}_{-2}(\mathbb{R}^\nu) = \left\{\phi \,\middle|\, \int (t(p)+1)^{-2}|\phi(p)|^2 \, d^\nu p < \infty\right\}. \tag{2.6}$$

Clearly

$$\mathscr{H}_2 \subset \mathscr{H}_1 \subset \mathscr{H}_0 \subset \mathscr{H}_{-1} \subset \mathscr{H}_{-2}.$$

The norm and scalar product in the Hilbert spaces $\mathscr{H}_r$ ($-2 \leqslant r \leqslant 2$) will be denoted by $\| \ \|_r$ and $(\ ,\ )_r$, respectively.

Remark: In order to treat the Kronig–Penney model and its two- and three-dimensional analogues, we shall need a slight modification of the definitions (2.1) to (2.5). Let Λ be a lattice in momentum space (the Bragg, or reciprocal, lattice) and let k be a point in $\mathbb{R}^3$. Consider the displaced lattice $k + \Lambda$, and define $\mathscr{H}_r(k)$ to consist of complex-valued functions on the discrete set $k + \Lambda$ satisfying

$$\sum_{p \in k+\Lambda} (t(p)+1)^r |\psi(p)|^2 < \infty. \tag{2.7}$$

So the integral $\int d^\nu p$ is replaced by a sum over a displaced lattice. All our general statements about the spaces $\mathscr{H}_r(\mathbb{R}^\nu)$ remain valid for the spaces $\mathscr{H}_r(k)$. Clearly k and $k + K$ (with $K \in \Lambda$) give the same $\mathscr{H}_r(k)$.

Examples

(a) It is particularly important to place exponentials (which correspond to δ-functions in x-representation) into the right Sobolev spaces. Here the dimension ν plays a crucial role. Let $t(p) = p_1^2 + \cdots + p_\nu^2$, and let

$$e_\nu^x(p) = (2\pi)^{-\nu/2} e^{ix\cdot p} \quad (x \in \mathbb{R}^\nu, p \in \mathbb{R}^\nu). \tag{2.8}$$

Then, for all x,

$$\begin{aligned} &e_1^x \in \mathscr{H}_{-1}(\mathbb{R}) \\ &e_2^x \notin \mathscr{H}_{-1}(\mathbb{R}^2) \quad e_2^x \in \mathscr{H}_{-2}(\mathbb{R}^2) \\ &e_3^x \notin \mathscr{H}_{-1}(\mathbb{R}^3) \quad e_3^x \in \mathscr{H}_{-2}(\mathbb{R}^3) \\ &e_\nu^x \notin \mathscr{H}_{-2}(\mathbb{R}^\nu) \quad (\text{if } \nu \geqslant 4). \end{aligned} \tag{2.9}$$

(b) It is interesting to consider, in one-dimension, derivatives of the δ-function ("dipoles") at the points $x_1 \cdots x_n$. In momentum representation, they are given by the functions $(2\pi)^{-1/2} p e^{ipx_j}$. They belong to $\mathscr{H}_{-2}(\mathbb{R})$.

The free resolvent is the operator

$$R(E) = (T - E)^{-1}, \tag{2.10}$$

i.e.,

$$(R(E)\phi)(p) = (t(p) - E)^{-1}\phi(p)$$

We denote by $\rho(T)$ the resolvent set of T, i.e., the set of E such that $R(E)$ is bounded as an operator in $\mathscr{H}_0$. If $t(p) = p^2$, then $\rho(T)$ is the slit plane $\mathbb{C}\setminus[0\ \infty)$. In $\mathscr{H}_0(k)$, $\rho(T)$ is the infinitely punctured plane $\mathbb{C}\setminus(k + \Lambda)$.

The square root $R^{1/2}(E)$ will be defined [in $\rho(T)$] by requiring it to be a positive operator if E is negative.

We shall be constantly using the following facts:

(i) $R(E)$ satisfies the identities

$$R(E) - R(E') = (E - E')R(E)R(E'),$$

$$\frac{d}{dE} R(E) = R(E)^2. \tag{2.11}$$

(ii) $R(E)$ is a topological isomorphism (i.e., bounded onto map with bounded inverse) from $\mathscr{H}_r$ to $\mathscr{H}_{r+2}$.

Similarly, $R^{1/2}(E)$ is a topological isomorphism from $\mathscr{H}_r$ to $\mathscr{H}_{r+1}$.

The scalar product in $\mathscr{H}_0$ can be extended to some

pairs of vectors in $\mathscr{H}_{-2}$, such that one member of the pair does not belong to $\mathscr{H}_0$. This makes $\mathscr{H}_{-2}$ a *partial inner product* space.[17] If $\Phi\in\mathscr{H}_{-2}$ and $\Psi\in\mathscr{H}_{-2}$ are such that $\Phi\in\mathscr{H}_r$ and $\Psi\in\mathscr{H}_{-r}$ for some $r\ (-2\leqslant r\leqslant 2)$, then the integral $\int \bar{\Phi}(p)\Psi(p)\,d^\nu p$ converges absolutely and defines the partial inner product

$$\langle\Phi|\Psi\rangle = \int \bar{\Phi}(p)\Psi(p)d^\nu p. \tag{2.12}$$

Notice that:

$$\langle\Phi_1|R(E)\Phi_2\rangle \text{ exists for all } \Phi_1\in\mathscr{H}_{-1},\ \Phi_2\in\mathscr{H}_{-1}, \tag{2.13}$$

while

$$\langle\Psi_1|R(E)^2\Psi_2\rangle \text{ exist for all } \Psi_1\in\mathscr{H}_{-2},\ \Psi_2\in\mathscr{H}_{-2}. \tag{2.14}$$

These remarks, together with (2.11), will be essential later on.

For any $\Psi'\in\mathscr{H}_{-2}$, $\Psi''\in\mathscr{H}_{-2}$, introduce the *dyadic operator*

$$|\Psi'\rangle\langle\Psi''|, \tag{2.15}$$

as follows: Let $\mathscr{H}_{r'}$ ($\mathscr{H}_{r''}$) be the smallest space $\mathscr{H}_r$ that contains Ψ' (Ψ''). (Remember that a small space $\mathscr{H}_r$ corresponds to a big index r.)

Then $|\Psi'\rangle\langle\Psi''|$ is the bounded map from $\mathscr{H}_{-r''}$ to $\mathscr{H}_{r'}$, defined by

$$(|\Psi'\rangle\langle\Psi''|)(\phi) = \langle\Psi''|\phi\rangle\Psi' \quad (\phi\in\mathscr{H}_{-r''}). \tag{2.16}$$

It is important to notice that, for any $\Psi'\in\mathscr{H}_{-2}$, $\Psi''\in\mathscr{H}_{-2}$, that symbol

$$R(E)|\Psi'\rangle\langle\Psi''|R(E), \tag{2.17}$$

is well defined as a bounded operator in $\mathscr{H}_0$. Indeed, if $\phi\in\mathscr{H}_0$, then $R(E)\phi\in\mathscr{H}_2$. Consequently $\langle\Psi''|R(E)\phi\rangle$ is defined. On the other hand, $R(E)\Psi'\in\mathscr{H}_0$. So (2.17) applied to $\phi\in\mathscr{H}_0$, gives $\langle\Psi''|R(E)\phi\rangle R(E)\Psi'\in\mathscr{H}_0$.

There is no difficulty in defining sums of dyadics and the product of a dyadic by a number.

The following shorthand notation will be useful: If

$$\Psi' = \{\Psi'_1,...,\Psi'_n\} \tag{2.18}$$

and

$$\Psi'' = \{\Psi''_1,...,\Psi''_n\} \tag{2.19}$$

are n-tuples of vectors in $\mathscr{H}_{-2}$ (which we write $\Psi'\in\mathscr{H}^n_{-2}$ and $\Psi''\in\mathscr{H}^n_{-2}$), and if $\mathbf{B}=\{B_{jl}\}$ is any $n\times n$ matrix, then $|\Psi'\rangle\mathbf{B}\langle\Psi''|$ is defined by

$$|\Psi'\rangle\mathbf{B}\langle\Psi''| = \sum_{j=1}^n\sum_{l=1}^n |\Psi'_j\rangle B_{jl}\langle\Psi''_l| = \sum_{j=1}^n\sum_{l=1}^n B_{jl}|\Psi'_j\rangle\langle\Psi''_l|. \tag{2.20}$$

We can finally combine the notations: If $\mathbf{B}$ is any $n\times n$ matrix, then

$$R(E)|\Psi'\rangle\mathbf{B}\langle\Psi''|R(E) \tag{2.21}$$

is shorthand for

$$\sum_{j=1}^n\sum_{l=1}^n R(E)|\Psi'_j\rangle B_{jl}\langle\Psi''_l|,$$

i.e., for the map that sends $\phi\in\mathscr{H}_0$ to

$$\sum_{j=1}^n\sum_{l=1}^n \langle\Psi_l|R(E)\phi\rangle B_{jl}R(E)\Psi_j.$$

Operators *between* spaces $\mathscr{H}_r$ are only a tool for the study of operators *in* the physical space $\mathscr{H}_0$.

To every bounded linear map A_{sr} from $\mathscr{H}_r\subseteq\mathscr{H}_0$ to $\mathscr{H}_s\supseteq\mathscr{H}_0$ we associate the *natural restriction*[17] of A which is an operator (in general unbounded) in $\mathscr{H}_0$. It is defined as the restriction of A to the domain

$$\mathscr{D} = \{\phi\,|\,\phi\in\mathscr{H}_r, A\phi\in\mathscr{H}_0\}. \tag{2.22}$$

In other words, the original domain $\mathscr{H}_r\subseteq\mathscr{H}_0$ is cut down by striking out exactly the vectors that are mapped beyond $\mathscr{H}_0$. For instance, let T_{-11} be T as a map from $\mathscr{H}_1$ onto $\mathscr{H}_{-1}$. Then the natural restriction of T_{-11} is T considered on $\mathscr{H}_{-2}$, its operator domain.

The natural restriction of a perturbation of T by a symmetric dyadic $|\Phi\rangle\langle\Phi|$ *with* $\Phi\in\mathscr{H}_{-1}$, is self-adjoint. More generally

Proposition: Let $\Phi\in\mathscr{H}^n_{-1}$, and let $\mathbf{B}$ be an $n\times n$ invertible matrix. Then the natural restriction of $T-|\Phi\rangle\mathbf{B}\langle\Phi|$ is a closed operator in $\mathscr{H}_0$. The resolvent of this operator is given by

$$R(E) - R(E)|\Phi\rangle\Gamma(E)^{-1}\langle\Phi|R(E), \tag{2.23}$$

with $\Gamma(E)$ defined by

$$\Gamma(E) = \langle\Phi|R(E)\Phi\rangle - \mathbf{B}^{-1}. \tag{2.24}$$

If $\mathbf{B}$ is Hermitian, then the operator $T\doteq|\Phi\rangle\mathbf{B}\langle\Phi|$, defined by (2.23), (2.24), is self-adjoint.

Remember that a symmetric operator A defined on a domain $\mathscr{D}\subseteq\mathscr{H}_0$ is self-adjoint if, for every nonreal E, $A-E$ maps $\mathscr{D}$ onto $\mathscr{H}_0$.

We start the proof with a lemma that involves only $\mathscr{H}_0$, and can be traced back to the beginning of this century.

Lemma: Let $\phi=\{\phi_1,...,\phi_n\}$ ($\phi_j\in\mathscr{H}_0$) and $\psi=\{\psi_1,...,\psi_n\}$ ($\psi_j\in\mathscr{H}_0$). Let $\mathbf{A}$ be an invertible $n\times n$ matrix. Assume also that $\mathbf{A}^{-1}-\langle\psi|\phi\rangle$ is an invertible matrix. Then the inverse of the operator $1-|\phi\rangle\mathbf{A}\langle\psi|$ is

$$(1-|\phi\rangle\mathbf{A}\langle\psi|)^{-1} = 1+|\phi\rangle[\mathbf{A}^{-1}-\langle\psi|\phi\rangle]^{-1}\langle\psi|. \tag{2.25}$$

Proof: Direct verification.

Proof of Proposition: Notice that $T-E$ is a topological isomorphism (i.e., bounded with bounded inverse) from $\mathscr{H}_1$ to $\mathscr{H}_{-1}$, and that $|\Phi\rangle\mathbf{B}\langle\Phi|$ is bounded from the same $\mathscr{H}_1$, to $\mathscr{H}_{-1}$. (This is the place where the proof would break down if we tried $\Phi\in\mathscr{H}^n_{-2}$.) Write

$$T-|\Phi\rangle\mathbf{B}\langle\Phi|-E = R^{-1/2}(E)[1-R^{1/2}(E)|\Phi\rangle\mathbf{B}\langle\Phi|R^{1/2}(E)]R^{-1/2}(E)$$

and use the lemma to invert the term in the square bracket. The inverse of $T-|\Phi\rangle\mathbf{B}\langle\Phi|-E$, (as an operator from $\mathscr{H}_{-1}$ to $\mathscr{H}_1$), is then

$$R^{1/2}(E)[1+R^{1/2}(E)|\Phi\rangle\times[\mathbf{B}^{-1}-\langle\Phi|R(E)\Phi\rangle]^{-1}\langle\Phi|R^{1/2}(E)]R^{1/2}(E) = R(E)-R(E)|\Phi\rangle[\langle\Phi|R(E)\Phi\rangle-\mathbf{B}^{-1}]\times\langle\Phi|R(E).$$

Restriction to $\mathscr{H}_0$ gives the assertion.

Things are very different if we perturb T by a dyadic $|\Psi\rangle\langle\Psi|$ where $\Psi\in\mathscr{H}_{-2}$ and $\Psi\notin\mathscr{H}_{-1}$. In that case, the natural restriction of $T+|\Psi\rangle\langle\Psi|$ is not self-adjoint but has a family of self-adjoint extensions, which are not determined by Ψ. All of Sec. 3 can be viewed as the study of this family.

In order to understand this, it is useful to keep in mind the proposition that follows.

Proposition: Let $\Psi\in\mathscr{H}_{-2}$ and $\Psi\notin\mathscr{H}_{-1}$. Then the natural restriction of $T+|\Psi\rangle\langle\Psi|$ is the restriction of T to the domain $\{\phi\,|\phi\in\mathscr{H}_2, \langle\phi|\Psi\rangle=0\}$ (which can be dense in $\mathscr{H}_0$).

Proof: Consider the condition

$$T\phi+\langle\Psi|\phi\rangle\Psi=\psi\in\mathscr{H}_0, \tag{2.26}$$

which defines the domain of the natural restriction. Since $\Psi\notin\mathscr{H}_{-1}$, we have necessarily $\phi\in\mathscr{H}_2$, and consequently $T\phi\in\mathscr{H}_0$. So $T\phi-\psi\in\mathscr{H}_0$, while (2.26) says that $T\phi-\psi=\langle\Psi|\phi\rangle\Psi$. This is a contradiction, unless $\langle\Psi|\phi\rangle=0$.

Notice the difference with the case $\Phi\in\mathscr{H}_{-1}$. There both $T\phi$ and $\langle\Phi|\phi\rangle\Phi$ were in $\mathscr{H}_{-1}$, and could sum up to a nonzero element of $\mathscr{H}_0$.

Nevertheless, if, for fixed $\Phi\in\mathscr{H}^n_{-1}$, we consider the *family* of all operators $T-|\Phi\rangle\mathbf{B}\langle\Phi|$ with $\mathbf{B}$ ranging over all invertible matrices, we obtain statements which remain valid if $\Phi\in\mathscr{H}^n_{-2}$. This will be the subject of the next section.

3. THE OPERATORS H_Γ^Ψ

Given n vectors $\Psi_1,...,\Psi_n$ in $\mathscr{H}_{-2}$, we shall define in this section a *family* of closed operators $\{H_\Gamma^\Psi\}$ in $\mathscr{H}_0$, labelled by n^2 complex parameters. In the mildly singular case ($\Psi\in\mathscr{H}^n_{-1}$), H_Γ^Ψ is a perturbation of T by a dyadic of the form $-|\Psi\rangle\mathbf{B}^{-1}\langle\Psi|$. In the general case, H_Γ^Ψ is defined by an explicit expression for its resolvent. All the examples of physical interest will be special cases of this definition.

A. The vector Ψ

Throughout the rest of this paper, we denote by

$$\Psi=\{\Psi_1,...,\Psi_n\} \tag{3.1}$$

an n-tuple of vectors in $\mathscr{H}_{-2}$, and make the assumption that

(A) the vectors $\Psi_1,...,\Psi_n$ are linearly independent over $\mathscr{H}_0$, i.e., such that no nontrivial linear combination $\Sigma^n\lambda_j\Psi_j$ lies in $\mathscr{H}_0$.

This assumption is made in order to avoid inessential complications in the statements of results. It rules out the case $\Psi\in\mathscr{H}^n_0$ (bounded degenerate perturbations of T) but *allows* $\Psi\in\mathscr{H}^n_{-1}$ (mildly singular degenerate perturbations) which were studied in the preceding section.

The examples of interest will be:

$$\Psi_j(p)=(2\pi)^{-\nu/2}e^{ipx_j}$$
$$(\nu=1,2,3;\ p\in\mathbb{R}^\nu,\ x_j\in\mathbb{R}^\nu,\ j=1...n) \tag{3.2}$$

(exponentials in dimensions less than three), and

$$\Psi_j(p)=(2\pi)^{-1}pe^{ipx_j}\quad(p\in\mathbb{R},\ x_j\in\mathbb{R},\ j=1,...n) \tag{3.3}$$

(exponential times a linear function in one dimension), corresponding to a derivative of the δ-function in x-representation. We see that the assumption (A) is satisfied.

B. The matrix family $\Gamma(E)$

The n vectors $\Psi_j\in\mathscr{H}_{-2}$, define a family of n^2 functions $\Gamma_{jl}(E)$, each analytic in $\rho(T)$ and determined up to an additive constant. These functions will be needed in the definition of H_Γ^Ψ.

Define $\Gamma_{jl}(E)$ as any solution of

$$\frac{d}{dE}\Gamma_{jl}(E)=\langle\Psi_j|R(E)^2\Psi_l\rangle\quad(E\in\rho(T);j,l=1,...,n). \tag{3.4}$$

The rhs of (3.4) is well defined by Sec. 2. In the notations of Sec. 2, (3.4) is

$$\frac{d}{dE}\Gamma(E)=\langle\Psi|R(E)^2\Psi\rangle. \tag{3.5}$$

If $\Gamma(E)$ is any solution of (3.5), then the most general solution of (3.5) is of the form

$$\Gamma(E)-\alpha, \tag{3.5'}$$

where α is an arbitrary $n\times n$ complex matrix.

The particular solution of (3.5) that vanishes at a given point $E_0\in\rho(T)$ is given by

$$\Gamma_{E_0}(E)=\int_{E_0}^{E}\langle\Psi|R(E')^2\Psi\rangle dE'$$
$$=(E-E_0)\langle\Psi|R(E)R(E_0)\Psi\rangle. \tag{3.6}$$

In the special case that $\Psi\in\mathscr{H}^n_{-1}$ (3.5) has a privileged solution $\Gamma_0(E)=\langle\Psi|R(E)\Psi\rangle$, and so there is an "absolute reference point" for the subtraction constant α.

The following properties of any $\Gamma(E)$ that satisfies (3.5) (with $\Psi\in\mathscr{H}^n_{-2}$) are easily verified and will be needed in the study of spectral properties of H_Γ^Ψ:

(i) If $E\in\rho(T)$ and $E'\in\rho(T)$, then

$$\Gamma(E)-\Gamma(E')=(E-E')\langle\Psi|R(E)R(E')\Psi\rangle. \tag{3.7}$$

(ii) On the negative real axis, $\Gamma(E)$ is a monotonically increasing matrix-valued function; that is, if $-\infty<E<E'<0$, then $\Gamma(E')-\Gamma(E)$ is a positive definite matrix.

(iii) On the negative real axis, the matrix

$$\mathrm{Im}\Gamma(E)=\frac{1}{2i}[\Gamma(E)-\Gamma^\dagger(E)] \tag{3.8}$$

is constant (independent of E).

(iv) $\Gamma(E)$ is analytic in $\rho(T)$. Its "unphysical sheet" continuations will be discussed in Section 4.

(v) If $\Gamma(E)$ is a solution of (3.5), then $\Gamma(E)=\Gamma^\dagger(\bar{E})$ is also a solution of (3.5). Here $\Gamma^\dagger$ is the Hermitian conjugate of Γ and $\bar{E}$, the complex conjugate of E.

C. Examples

Introduce the variable κ by

$$\kappa^2=-E,\quad \mathrm{Re}\kappa\geqslant 0. \tag{3.9}$$

So κ is related to the variable k of scattering theory by $k=i\kappa$. The "physical half-plane" $\mathrm{Im}k\geqslant 0$ corresponds to the "physical half-plane" $\mathrm{Re}\kappa\geqslant 0$. The axis $-\infty<E<0$ corresponds to $0<\kappa<\infty$. A function which increases monotonically on the negative real E-axis defines a decreasing function in the positive real κ-axis.

(a) One-dimension δ-functions: If Ψ is given by

$$\Psi_j(p) = (2\pi)^{-1/2} e^{ipx_j} \quad (p\in\mathbb{R}, x_j\in\mathbb{R}).$$

Then the general solution of (3.5) is

$$\Gamma_{jl}(E) = \frac{1}{2\kappa} e^{-|x_j - x_l|\kappa} - \alpha_{jl}\,. \tag{3.10}$$

Notice that here $\Psi_j\in\mathscr{H}_{-1}$, and consequently the privileged solution:

$$\langle \Psi_j | R(E)\Psi_l\rangle = \frac{1}{2\kappa} e^{-|x_j - x_l|\kappa},$$

is well-defined for $j,l = 1,...,n$.

(b) One dimension, derivatives of δ-functions: Let

$$\Psi_j(p) = (2\pi)^{-1/2} p e^{ipx_j}.$$

The general solution of (3.5) is

$$\Gamma_{jl}(E) = -\tfrac{1}{2}\kappa e^{-|x_j - x_l|\kappa} - \alpha_{jl}\,. \tag{3.11}$$

(c) Two dimensions, δ-functions: Consider

$$\Psi_j(p) = (2\pi)^{-1} e^{ipx} \quad (p\in\mathbb{R}^2, x\in\mathbb{R}^2).$$

We have now

$$\Gamma_{jl}(E) = \frac{1}{2\pi} K_0(|x_j - x_l|\kappa) - \alpha_{jl} \quad (j\neq l) \tag{3.12}$$

$$\Gamma_{jj}(E) = -\ln\kappa - \alpha_{jj}\,. \tag{3.12'}$$

Here K_0 is the modified Bessel function.

(d) Three dimensions, δ-functions: Let

$$\Psi_j(p) = (2\pi)^{-3/2} e^{ipx_j} \quad (p\in\mathbb{R}^3, x_j\in\mathbb{R}^3). \tag{3.13}$$

Then

$$\Gamma_{jl}(E) = \frac{1}{4\pi|x_j - x_l|} e^{-|x_j - x_l|\kappa} \quad (j\neq l), \tag{3.14}$$

$$\Gamma_{jj}(E) = -\frac{1}{4\pi}\kappa - \alpha_{jj}\,. \tag{3.15}$$

(e) Three dimensions, lattice: Consider the restriction of the function $(2\pi)^{-3/2} e^{-ipx_j}$ to the set $k + \Lambda$. This restriction belongs to $\mathscr{H}_{-2}(k)$. We have

$$\langle \Psi_j | R(E)^2 | \Psi_l\rangle = \sum_{p\in k+\Lambda} \frac{e^{ip(x_l - x_j)}}{[(p+k)^2 - E]^2} \tag{3.16}$$

which is absolutely convergent. The corresponding $\Gamma(E)$ and Hamiltonians will be studied in detail in another paper of this series.

D. Definition of H_Γ^Ψ

Given Ψ and Γ, we shall now define an operator H_Γ^Ψ in $\mathscr{H}_0$ by writing down an explicit formula for its resolvent.

Theorem: Let Ψ be an n-tuple of vectors in $\mathscr{H}_{-2}$ satisfying condition (A) of Sec. 3A. Let $\Gamma(E)$ be any solution of (3.5). Consider in $\mathscr{H}_0$ the family of operators $Q_\Gamma^\Psi(E)$, defined by

$$Q_\Gamma^\Psi(E) = R(E) - R(E)|\Psi\rangle\Gamma(E)^{-1}\langle\Psi|R(E), \tag{3.17}$$

where $\Gamma(E)^{-1}$ is the matrix inverse of $\Gamma(E)$ and $R(E) = (T - E)^{-1}$ is the resolvent of the kinetic energy operator T.

Then $Q_\Gamma^\Psi(E)$ is the resolvent family of a closed densely defined operator H_Γ^Ψ.

$$Q_\Gamma^\Psi(E) = (H_\Gamma^\Psi - E)^{-1}. \tag{3.18}$$

Before proving the theorem, we write out (3.17) without shorthand notation. If ϕ is any vector in $\mathscr{H}_0$, if E is such that $\det\Gamma(E)\neq 0$, then

$$Q_\Gamma^\Psi(E) = R(E)\phi - \sum_{j=1}^n \sum_{l=1}^n \langle \Psi_l | R(E)\phi\rangle \times (\Gamma(E)^{-1})_{jl} R(E)\Psi_j\,. \tag{3.19}$$

Proof of Theorem: For simplicity, we write $Q(E)$ instead of $Q_\Gamma^\Psi(E)$, wherever there is no ambiguity.

(a) The operator (3.19) is bounded whenever $E\in\rho(T)$ and $\det\Gamma(E)\neq 0$, since $R(E)\Psi\in\mathscr{H}_0^n$.

(b) The family $Q(E)$ satisfies the first resolvent equation

$$Q(E)Q(E') = \frac{1}{E - E'}[Q(E) - Q(E')]. \tag{3.20}$$

Indeed,

$$\begin{aligned} &Q(E)Q(E') \\ &= R(E)R(E') - R(E)R(E')|\Psi\rangle\Gamma(E')^{-1}\langle\Psi|R(E') \\ &\quad - R(E)|\Psi\rangle\Gamma(E)^{-1}\langle\Psi|R(E)R(E') + R(E)|\Psi\rangle \\ &\quad \times\Gamma(E)^{-1}\langle\Psi|R(E)R(E')\Psi\rangle\Gamma(E')^{-1}\langle\Psi|R(E'). \end{aligned}$$

The matrix in the fourth term is, by (3.7)

$$\begin{aligned} &\Gamma(E')^{-1}\langle\Psi|R(E)R(E')\Psi\rangle\Gamma(E')^{-1} \\ &\quad = \frac{1}{E - E'}\Gamma(E)^{-1}[\Gamma(E) - \Gamma(E')]\Gamma(E')^{-1} \\ &\quad = \frac{1}{E - E'}[\Gamma(E')^{-1} - \Gamma(E)^{-1}]. \end{aligned}$$

By using (2.11) in the first three terms, we obtain (3.17).

(c) $Q(E)$ is invertible in $\mathscr{H}_0$, i.e., $Q(E)\phi = 0$ and $\phi\in\mathscr{H}_0$ implies $\phi = 0$. Indeed, $Q(E)\phi = 0$ means

$$R(E)\phi = R(E)|\Psi\rangle\Gamma(E)^{-1}\langle\Psi|R(E)\phi\rangle,$$

which means that $R(E)\phi$ is in the linear span of the vectors $R(E)\Psi_j$ $(j = 1,...,n)$. Now $R(E)\phi\in\mathscr{H}_2$ by Sec. 2. On the other hand, no nontrivial linear combination of the $R(E)\Psi_j$ lies in $\mathscr{H}_2$, by the assumption (A) that the Ψ_j are linearly independent over $\mathscr{H}_0$ and the fact that $R(E)$ is a topological isomorphism from $\mathscr{H}_0$ to $\mathscr{H}_2$. It follows that $R(E)\phi = 0$. Since $R(E)$ is invertible in $\mathscr{H}_0$, we have $\phi = 0$.

(d) we have

$$[Q_\Gamma(E)]^* = Q_{\Gamma_1}(E), \tag{3.21}$$

where $\Gamma_1(E) = \Gamma^\dagger(\bar{E})$. Proof by direct verification.

(e) The range of $Q_\Gamma(E)$ (which is independent of E and is, by definition, the domain of H_Γ) is dense in $\mathscr{H}_0$. Let ψ be orthogonal to the range of $Q_\Gamma(E)$, i.e., $(\psi, Q(E)\phi) = 0$ for all $\phi\in\mathscr{H}_0$. Then, by (d) $Q_{\Gamma_1}(E)\psi = 0$, which, by (c) gives $\psi = 0$.

We can now define H_Γ^Ψ as

$$H_\Gamma^\Psi = [Q_\Gamma(E)]^{-1} + E\,.$$

It is independent of E, closed (as the inverse of a bounded invertible operator) and densely defined. Q.E.D.

E. Remarks on the definition of H_Γ^Ψ

The above definition gives H_Γ^Ψ only indirectly, by (3.17). In particular, it does not allow us to write $H_\Gamma^\Psi = T + V$ with some operator V in $\mathscr{H}_0$. We shall indeed see that such a decomposition of H_Γ^Ψ does not exist in any reasonable sense, unless Ψ is only mildly singular ($\Psi \in \mathscr{H}^n_{-1}$). However, for a spectral analysis of an operator it is much more important to have an explicit expression for the resolvent than for the operator itself. We shall see that, in concrete cases, all spectral properties of H_Γ^Ψ can be obtained by very simple computations.

F. General properties of H_Γ^Ψ : parametric representation

Proposition: The adjoint of H_Γ^Ψ is $H_{\Gamma_1}^\Psi$ where $\Gamma_1(E) = \Gamma^\dagger(\bar{E})$. [or simply $\Gamma_1(E) = \Gamma^\dagger(E)$ if E is real.] The operators H_Γ^Ψ and $(H_\Gamma^\Psi)^*$ do not commute in general. If $\Gamma(E)$ is Hermitian on the real axis, then H_Γ^Ψ is self-adjoint.

Proposition: If $\mathrm{Im}(\Gamma(E)) \geqslant 0$ on the negative real axis (i.e., if $(1/2i)[\Gamma(E) - \Gamma^\dagger(E)]$ is a positive definite matrix), then H_Γ^Ψ is absorptive, in the sense that $\|e^{-iH_\Gamma^\Psi t}\| \leqslant 1$ for $t \geqslant 0$.

Proposition: Let $\mathbf{A}$ be any invertible $n \times n$ matrix; denote by $\bar{\mathbf{A}}$ the complex conjugate and by $\mathbf{A}^T$ the transpose of $\mathbf{A}$. Then

$$H_\Gamma^\Psi = H_{\bar{A}\Gamma A^T}^{A\Psi}. \tag{3.22}$$

In particular, if λ is real and different from zero, then

$$H_\Gamma^\Psi = H_{\lambda^2\Gamma}^{\lambda\Psi}. \tag{3.23}$$

The above proposition are proved by direct verification.

Even though we cannot give an explicit form of H_Γ^Ψ, the following "parametric representation" describes explicitly the graph of H_Γ^Ψ, and is quite useful in applications.

Proposition: Let E_0 be any point in $\rho(T)$. The domain $\mathscr{D}(H_\Gamma^\Psi)$ consists exactly of vectors ψ of the form

$$\psi = \phi - R(E_0)\,|\,\Psi\rangle\Gamma(E_0)^{-1}\langle\Psi|\phi\rangle, \tag{3.24}$$

where $\phi \in \mathscr{H}_2$. For given $\psi \in \mathscr{D}$, there is only one $\phi \in \mathscr{H}_2$ satisfying (3.24). If $\psi \in \mathscr{D}(H_\Gamma^\Psi)$ is given by (3.24), then the action of H_Γ^Ψ on ψ is described by

$$(H_\Gamma^\Psi - E_0)\psi = (T - E_0)\phi. \tag{3.25}$$

Proof: By definition, $\mathscr{D}(H_\Gamma^\Psi) = Q_\Gamma(E_0)\mathscr{H}_0 = Q_\Gamma(E_0)(T - E_0)\mathscr{H}_2$. Now $Q_\Gamma(E_0)(T - E_0) = 1 - R(E_0)|\Psi\rangle\Gamma(E)^{-1}\langle\Psi|$, which proves (3.24). If $\psi = 0$, then $\phi = R(E_0)|\Psi\rangle\Gamma(E^{-1}\langle\Psi|\phi\rangle$. By assumption (A), Sec. 3A, this is possible only if $\phi = 0$. The assertion (3.25) means $\psi = Q(E_0)(T - E_0)\phi$, which is directly verified.

Corollary: Define $\mathscr{D}^0$ as the set

$$\mathscr{D}^0 = \{\psi \mid \psi \in \mathscr{H}_2, \quad \langle\psi \mid \Psi_j\rangle = 0, \quad j = 1,...,n\}.$$

Then

(i) $\mathscr{D}(H_\Gamma^\Psi) \cap \mathscr{H}_2 = \mathscr{D}^0$.

(ii) The restriction of H_Γ^Ψ to $\mathscr{D}^0$ coincides with the restriction of T to $\mathscr{D}^0$:

$$H_\Gamma^\Psi\big|_{\mathscr{D}^0} = T\big|_{\mathscr{D}^0}.$$

In other words, H_Γ^Ψ is an extension of a restriction of T.

Proof: If $\psi \in \mathscr{D}(H_\Gamma^\Psi)$ belongs to $\mathscr{H}_2$, then by (3.24) $\psi = \phi$, $\psi \in \mathscr{D}^0$, and by (3.25), $H\psi = T\psi$.

Remark: $\mathscr{D}^0$ is not the orthogonal complement of a set in $\mathscr{H}_0$, and so can be dense in $\mathscr{H}_0$.

Example: Let $\Psi_j(p) = (2\pi)^{-3/2}e^{ipx_j}$ ($j = 1,...,n$; $x_j \in \mathbb{R}^3$, $p \in \mathbb{R}^3$). Then $\mathscr{D}^0$ consists of functions $\phi(p)$ such that $\int(p^2 + 1)^2|\phi(p)|^2 d^3p < \infty$, and that $\int e^{-ipx_j}\phi(p)d^3p = 0$ ($j = 1,...,n$). In terms of the Fourier transform (x-representation) this means that the wave function $\tilde{\phi}(x)$ vanishes at x_1, ..., x_n.

The parametric representation (3.24) (3.25) gives also a convenient expression for expectation values of H_Γ^Ψ.

If $\psi \in \mathscr{D}(H_\Gamma^\Psi)$, then

$$\langle\psi|(H_\Gamma - E_0)\psi\rangle = \langle\phi\,|(T - E_0)\phi\rangle - \langle\phi\,|\Psi\rangle\Gamma(E_0)^{-1}\langle\Psi|\phi\rangle.$$

The normalization $\langle\psi|\psi\rangle$ can be computed from (3.24).

G. H_Γ^Ψ as limit of operators of the form $T + V$

Definition: Let $\Psi \in \mathscr{H}^n_{-2}$ satisfy condition (A) of Sec. 3A. For each $\omega > 0$, let $\Phi_j^\omega \in \mathscr{H}_{-1}$ ($j = 1,...,n$). Assume that Φ_j^ω converges to Ψ_j in the topology of $\mathscr{H}_{-2}$ (so Φ_j may "converge out" of $\mathscr{H}_{-1}$). That is, assume

$$\lim\|\Phi_j^\omega - \Psi_j\|_{-2} = 0 \quad (j = 1,...,n). \tag{3.26}$$

We shall say that a family $\mathbf{M}^\omega$ of $n \times n$ matrices is a counterterm for $\mathbf{\Phi}^\omega$ if the family

$$\langle\mathbf{\Phi}^\omega|R(E)\mathbf{\Phi}^\omega\rangle - \mathbf{M}^\omega \tag{3.27}$$

has a finite limit as $\omega \to \infty$, for all $E \in \rho(T)$.

Remark: It is enough to require (3.27) to have a finite limit for *one* $E_0 \in \rho(T)$, since

$$\langle\mathbf{\Phi}^\omega|R(E)\mathbf{\Phi}^\omega\rangle - \langle\mathbf{\Phi}^\omega|R(E_0)\mathbf{\Phi}^\omega\rangle = (E - E_0)\langle\Phi^\omega|R(E)R(E_0)\Phi^\omega\rangle$$

has a finite limit as $\omega \to \infty$.

Example: The above remark shows that, for any $E_0 \in \rho(T)$, we could take $\mathbf{M}^\omega = \langle\mathbf{\Phi}^\omega|R(E_0)\mathbf{\Phi}^\omega\rangle$ as the counterterm. We shall see, however, that in the interesting examples, it is preferable and possible to choose counterterms that are diagonal matrices.

We can now describe H_Γ^Ψ as a limit of operators of the form $T + V$.

Recall the definition of norm resolvent convergence. Let A and A_n ($n = 1,2...$) be closed operators in $\mathscr{H}_0$, with nonempty resolvent sets. We say that A_n converges to A in the norm resolvent sense, if any $E \in \rho(T_n)$ belongs to all $\rho(T_n)$ for sufficiently large n, and if

$$\lim_{n\to\infty}\|(A_n - E)^{-1} - (A - E)^{-1}\| = 0$$

(see e.g., Ref. 18, IV, §2, 6, and Ref. 19, Sec. VIII. 7). Norm resolvent convergence is the "stronger" of two concepts that are used in perturbation theory of operators.

Theorem: Let $\Psi \in \mathscr{H}^n_2$, satisfy (A) of Sec. 3A. For any $\omega > 0$, let $\mathbf{\Phi}^\omega \in \mathscr{H}^n_{-1}$. Assume that

$$\lim:\|\Phi_j^\omega - \Psi_j\|_{-2} = 0 \quad (j = 1,...,n).$$

Let $\mathbf{M}^\omega$ be a counterterm for $\mathbf{\Phi}^\omega$, such that the matrices $\mathbf{M}^\omega$ are invertible if ω is sufficiently large. Define

$$\Gamma(E) = \lim_{\omega\to\infty}(\langle\mathbf{\Phi}^\omega|R(E)\mathbf{\Phi}^\omega\rangle - \mathbf{M}^\omega). \tag{3.28}$$

Then

(i) $\Gamma(E)$ satisfies (3.5), i.e., defines an operator H_Γ^Ψ, and

(ii) this operator is a limit

$$H_\Gamma^\Psi = \lim_{\omega\to\infty} (T \dot{-} |\Phi^\omega\rangle(\mathbf{M}^\omega)^{-1}\langle\Phi^\omega|)$$

in the sense of norm resolvent convergence.

Proof: The resolvent of $T - |\Phi^\omega\rangle(\mathbf{M}^\omega)^{-1}\langle\Phi^\omega|$ is, by (2.23), equal to

$$R(E) - R(E)|\Phi^\omega\rangle[\langle\Phi^\omega|R(E)\Phi^\omega\rangle - \mathbf{M}^\omega]^{-1}\langle\Phi^\omega|R(E).$$

The resolvent of H_Γ^Ψ is, by (3.17),

$$R(E) - R(E)|\Psi\rangle\Gamma(E)^{-1}\langle\Psi|R(E).$$

Since $\lim R(E)\Phi^\omega = R(E)\Psi$ (in $\mathscr{H}_0$), and since $\Gamma(E)$ is assumed invertible, the convergence follows from $\lim_{\omega\to\infty}(\langle\Phi^\omega|R(E)\Phi^\omega\rangle - \mathbf{M}^\omega) = \Gamma(E)$.

Example: Let $\Psi_j(p) = (2\pi)^{-3/2}e^{ipx_j}$, with $p\in\mathbb{R}^3$, $x_j\in\mathbb{R}^3$, $j = 1,...,n$. Define Φ^ω by a "cutoff":

$$\Phi_j^\omega(p) = \Theta_\omega(|p|)\Psi_j(p),$$

where $\Theta_\omega(\xi) = 1$ if $\xi < \omega$, and $\Theta_\omega(\xi) = 0$ if $\xi > \omega$. Introduce $\mathbf{M}^\omega$ by

$$M_{jl}^\omega = 0 \quad \text{for } j\neq l,$$

$$M_{jj}^\omega = \int |\Phi_j^\omega(p)|^2 p^2\, d^3p - \alpha_j = (2\pi^2)^{-1}\omega - \alpha_j,$$

where α_j is real. Then $\mathbf{M}^\omega$ is a counterterm for Ψ. Consequently, the family

$$T - \sum_{j=1}^{n} |\Phi_j^\omega\rangle\lambda_j^\alpha(\omega)\langle\Phi_j^\omega|,$$

with $(\lambda_j^\alpha(\omega))^{-1} = (2\pi^2)^{-1}\omega - \alpha_j$, converges to H_Γ^Ψ.

4. SOME SPECTRAL PROPERTIES: LOCALITY: INDEPENDENT CENTERS

A. Eigenvalues and bound states

Proposition: The points E that belong to the spectrum of H_Γ^Ψ and do not belong to the spectrum of T are the solutions of the equation

$$\det\Gamma(E) = 0. \tag{4.1}$$

The number of such points does not exceed n (the number of centers, i.e., the number of components of $\Psi = \{\Psi_1 ... \Psi_n\}$).

Proof: The first statement is obvious by (3.17). Solutions of (4.1) give poles of the resolvent, i.e., bound-states of H_Γ^Ψ. The assertion on the number of such points can be obtained either from the limits of Sec. 3G, or more directly, from the fact that the matrix-valued function $E\to\Gamma(E)$ increases monotonically on the real axis whenever it is defined:

$$\frac{d\Gamma(E)}{dE} = \langle\Psi|R(E)^2\Psi\rangle > 0, \tag{4.2}$$

since $\langle\Psi|R(E)^2\Psi\rangle$ is a positive matrix. Consequently the eigenvalues of $\Gamma(E)$ are strictly increasing functions $\lambda_j(E)$ $(j = 1,...,n)$. There are at most n real points where one of the functions $\lambda_j(E)$ has a zero. These are the eigenvalues of H_Γ^Ψ.

For the sake of simplicity, consider now the one-center situation ($n = 1$). If E_1 is a zero of $\det\Gamma(E)$, then the projection on the corresponding eigenstate is

$$P_1(E) = -\frac{1}{2\pi i}\oint Q(E)\,dE = \frac{1}{\langle\Psi|R(E_1)^2\Psi\rangle} R(E_1)|\Psi\rangle\langle\Psi|R(E_1). \tag{4.3}$$

B. Resonances

In all the examples that we have considered and where T has a continuous spectrum, the matrix-valued function $\Gamma(E)$ can be analytically continued to a Riemann surface over the E-plane. [By (3.5), it is easy to give general conditions on $t(p)$ which will insure this]. The zeros of this extension will be said to constitute the extended discrete spectrum of H_Γ^Ψ. If a point in the extended discrete spectrum is not an eigenvalue it is called a resonance. It is easy to see that this definition is equivalent to looking at poles of the resolvent between states which are localized in the x-representation and so analytic in the p-representation.

In the example of the diatomic molecule we shall see that most resonances show no inclination of becoming bound states under natural changes of the Hamiltonian.

C. Infrared-divergence

Let now E be real again, and assume that T has continuous spectrum in $[0,\infty)$. The fact that the eigenvalues $\lambda_j(E)$ of $\Gamma(E)$ are monotonically increasing on the real axis allows an immediate discussion of the possibility of "destroying" a bound state by changing the arbitrary subtraction constant α in $\Gamma(E)$. We shall say that Ψ is *infrared divergent* if $\lim_{E\to 0}\lambda_j(E) = +\infty$ $(j = 1,...,n)$. (In the case of one center, infrared-divergence means $\int[t(p)]^{-1}|\psi(p)|^2 = +\infty$, with $\psi = (T+1)^{-1}\Psi$. This is a "small momentum" divergence.)

If Ψ is infrared divergent, then no bound state can disappear into the threshold $E = 0$ to become a resonance.

Example of infrared divergence

(1) Two dimensions, one center, $\Psi(p) = (2\pi)^{-1}e^{ipx}$ $(p\in\mathbb{R}^2, x\in\mathbb{R}^2)$: We have seen that there

$$\Gamma(E) = -\ln\kappa - \alpha \quad (\kappa^2 = -E,\ \mathrm{Re}\kappa\geqslant 0). \tag{4.4}$$

The condition $\ln\kappa = -\alpha$, $\kappa_\alpha = e^{-\alpha}$ can be satisfied for any real α, by a κ in the physical half-plane. For regular, local potentials in two dimensions this "indestructibility of bound states" has been discussed by Simon.[20]

(2) One dimension, one center, $\Psi(p) = (2\pi)^{-1/2}e^{ipx}$ $(p\in\mathbb{R}, x\in\mathbb{R})$: Here the condition for $-E = \kappa^2$ to be a bound state energy is

$$\Gamma(E) = \frac{1}{2\kappa} - \alpha = 0, \tag{4.5}$$

i.e., $\kappa = 1/2\alpha$. As α tends to zero, κ_α takes a trip to infinity and returns as a resonance. This can also be seen from the explicit form $H_\Gamma = T \dot{-} |\Psi\rangle\alpha^{-1}\langle\Psi|$. (Remember that, in this case, $\Psi\in\mathscr{H}_{-1}$.)

We now consider the case where Ψ is *not* infrared divergent.

Examples of infrared convergence

(1) Three dimensions, one center: Here (4.1) becomes

$$-\frac{1}{4\pi}\kappa-\alpha=0. \tag{4.6}$$

The bound state $\kappa_\alpha=-4\pi\alpha$ becomes a resonance when α goes through zero, starting with a negative value.

(2) Dipole in one dimension: $\Psi(p)=(2\pi)^{-1/2}pe^{ipx}$ ($p\in\mathbb{R}$, $x\in\mathbb{R}$): Here the condition (4.1) is

$$\kappa_\alpha=-2\alpha. \tag{4.7}$$

D. Locality; independent centers

A Hamiltonian of the form $H=p^2+V(x)$ is *local*, in the following sense. Consider (in x-representation) a function ψ that belongs to the domain of H and that vanishes in a ball $\mathcal{O}$. Then $H\psi$ also vanishes in $\mathcal{O}$.

The Hamiltonian H_Γ^Ψ with arbitrary $\Psi_j\in\mathscr{H}_{-2}$ will not be local in general. However, if $\Psi_j(p)=(2\pi)^{-\nu/2}e^{ipx_j}$, then the intuitive picture of H_Γ^Ψ is that of a particle feeling point interactions, and we expect locality. In the proof given below, it is important to remember the following:

If Ψ_1 and Ψ_2 are arbitrary elements of $\mathscr{H}_{-2}$, then the matrix element $\langle\Psi_1|R(E)\Psi_2\rangle$ is not defined, since the integral $\int(p^2-E)\bar{\Psi}_1(p)\Psi_2(p)\,d^\nu p$ is, in general, divergent. If, however, we take $\Psi_1(p)=(2\pi)^{-\nu/2}e^{ipx_1}$ and $\Psi_2(p)=(2\pi)^{-\nu/2}e^{ipx_2}$, *with* $x_1\neq x_2$, then $(2\pi)^{-\nu}\int e^{p(x_2-x_1)}(p^2-E)^{-1}d^\nu p$ converges (as limit of integrals over bounded sets, not as an absolutely convergent integral if $\nu=2$ or 3). Consequently the "improper ;matrix element"

$$\langle x_1|R(E)|x_2\rangle=(2\pi)^{-\nu}\int e^{ip(x_2-x_1)}(p^2-E)^{-1}d^\nu p = G_E(x_1-x_2), \tag{4.8}$$

exists for $x_1\neq x_2$. We have

$$G_E(x_1-x_2)=\frac{1}{2\pi}K_0(\kappa|x_1-x_2|)\quad(\nu=2), \tag{4.9}$$

$$G_E(x_1-x_2)=[4\pi|x_1-x_2|]^{-1}e^{-\kappa|x_1-x_2|}\quad(\nu=3). \tag{4.10}$$

Theorem: Let $H_\Gamma^{x_1,\dots,x_n}$ be the operator H_Γ^Ψ with $T=p^2$, with $\Psi_j(p)=(2\pi)^{-\nu/2}e^{ipx_j}$ ($j=1,\dots,n$; $p\in\mathbb{R}^\nu$, $x_j\in\mathbb{R}^\nu$, $\nu=2,3$), and Γ any solution of (3.5). Then $H_\Gamma^{x_1,\dots,x_n}$ is local in the following sense: if ψ belongs to the domain of $H_\Gamma^{x_1,\dots,x_n}$ (in x-representation) and if ψ vanishes in a ball $\mathcal{O}$, then $H_\Gamma^{x_1,\dots,x_n}\psi$ also vanishes in $\mathcal{O}$.

Proof: We shall use the parametric representation introduced in Sec. 3F. It now reads as follows: a function $\psi(x)$ belongs to the domain if and only if it is of the form

$$\psi(x)=\phi(x)-\sum_{j=1}^n A_j^\phi G_{E_0}(x-x_j), \tag{4.11}$$

where $\phi\in\tilde{\mathscr{H}}_2$ (Fourier transforms of the space introduced in Sec. 2), and where

$$A_j^\phi=\sum_{l=1}^n(\Gamma(E_0)^{-1})_{jl}\phi(x_l). \tag{4.12}$$

Furthermore

$$(H_\Gamma^{x_1,\dots,x_n}-E_0)\psi=(-\Delta-E_0)\phi = (T-E_0)\phi \tag{4.13}$$

[by (3.24) and (3.25)].

If ψ is regular at the point x_j, then necessarily $A_j=0$, since $\phi(x)$ is regular and $G(x-x_j)$ is singular.

Now let $\mathcal{O}$ be a ball containing, say, the point x_1 and not $x_2,\dots,x_n$, and such that the restriction of ψ to $\mathcal{O}$ is equal to zero. Then, first

$$\psi(x)=\phi(x)+A_2^\phi G(x-x_2)+\cdots+A_n^\phi G(x-x_n),$$

for all x (since zero is regular). Secondly, in $\mathcal{O}$, we have

$$\phi(x)=-A_2^\phi G(x-x_2)-\cdots-A_n^\phi G(x-x_n).$$

By (4.13), then $(H_\Gamma^{x_1,\dots,x_n}-E_0)\psi$ is zero in $\mathcal{O}$, since G is the elementary solution for $-\Delta-E_0$.

All Hamiltonians $H_\Gamma^{x_1,\dots,x_n}$ are local, but some are more local than others. We shall say that $H_\Gamma^{x_1,\dots,x_n}$ has *independent* centers if the *off-diagonal* entries of the matrix $\Gamma(E)$ are

$$\Gamma_{jl}(E)=\langle x_j|R(E)|x_l\rangle=G_E(x_l-x_j)\quad(j\neq l)$$

From Sec. 3G, it follows that such $H_\Gamma^{x_1,\dots,x_n}$ are limits, in the sense of norm resolvent convergence, of operators of the form

$$T-\sum_{j=1}^n|\Phi^\omega\rangle\lambda_j^\omega\langle\Phi^\omega|,$$

where $|\Phi^\omega\rangle\to|x_j\rangle$ in $\mathscr{H}_{-2}$; there are no "cross terms".

We shall now illustrate this on the example of the "diatomic molecule" in three dimensions.

E. Energy surface and resonance surfaces; independent centers

The bound-state energies are solutions of $\det\Gamma(E)=0$ lying in the "physical half-plane" $E<0$. The condition $\det\Gamma(E)=0$ can be written as

$$\frac{\kappa}{2\pi}=-(\alpha_1+\alpha_2)\pm\left((\alpha_1-\alpha_2)^2+\frac{1}{(4\pi)^2}\frac{e^{-2\kappa|x_1-x_2|}}{|x_1-x_2|^2}\right)^{1/2}, \tag{4.14}$$

which has to be solved for κ. Even without solving (4.14), we see that, if $|x_1-x_2|\to\infty$, the solutions tend to $\kappa_1=-4\pi\alpha_1$, and $\kappa_2=-4\pi\alpha_2$. These are the "atomic", i.e., the "one-center" values; the particle spends most of its time at one of the centers.

For the sake of simplicity, let $\alpha_1=\alpha_2=\alpha$. Introduce the dimensionless quantities

$$\gamma=4\pi\alpha|x_1-x_2|, \tag{4.15}$$

$$\zeta=i\kappa|x_1-x_2|. \tag{4.16}$$

So ζ is proportional to $k=E^{1/2}$, and

$$E=\frac{\zeta^2}{|x_1-x_2|^2}. \tag{4.17}$$

The "physical half-plane" is $\mathrm{Im}\zeta>0$.

The extended discrete spectrum [or rather the set of ζ_n that correspond to the extended discrete spectrum by (4.17)], is the set of zeros of the entire function $(\zeta+\gamma)^2-e^{-2\zeta\gamma}$. In order to have a closer look at it, set $\zeta=\xi+i\eta$ ($\xi\in\mathbb{R}$, $\eta\in\mathbb{R}$).

The points ζ_n can lie either

(i) on the imaginary axis $\xi=0$,

(ii) on the curve

$$\eta = \ln\left|\frac{\sin\xi}{\xi}\right|. \tag{4.18}$$

The curve (4.18) is symmetric with respect to the imaginary axis, and lies in the half-plane $\eta < 0$. The lines $\xi = \pm\pi, \pm 2\pi, \cdots$ are asymptotes to (4.18). The branch in the strip $|\xi| < \pi$ goes through zero. The other branches go through maxima at points ξ that satisfy $\xi = \tan\xi$. For a given γ, the points ζ_n are found as follows:

(iii) The points $\zeta_n = i\eta_n$ on the imaginary axis are found as the solutions of the equation

$$\gamma + \eta = \pm e^{-\eta}. \tag{4.19}$$

(iv) The points ζ_n on the curve (4.18) are found as the intersection of (4.18) with the curve

$$\eta = \xi\cot\xi - \gamma. \tag{4.20}$$

Let us first examine the strip $|\xi| < \pi$ of the ζ-plane. This strip always contains three points of the extended discrete spectrum. If $\gamma < -1$, all these points are on the imaginary axis; two bound states ($\eta > 0$) and one resonance above the negative energy axis ($\eta < 0$). At $\gamma = -1$, the bound states collide with the resonance and they go into the first branch of the curve (4.18). At $\gamma = 1$ the remaining bound state moves to the negative imaginary axis.

Every strip $n\pi < \xi < (n+1)\pi$ with $n \geqslant 1$, contains exactly one resonance.

All the resonances can be viewed intuitively as corresponding to various modes of bouncing back and forth before escaping to infinity.

F. Correlated centers

As $|x_1 - x_2| = r \to 0$, the bound-state solution approaches $\kappa(r) \simeq 1/r$, giving $E(r) \simeq -1/r^2$. This means that the point interaction, independent center model is not suitable for the discussion of the limit where the two centers merge, as could have been predicted on general grounds.

It is easy to obtain models with "physical" behaviour at $r \to 0$, without impairing explicit solubility or locality.

Proposition: Let $\rho(r)$ be a smooth real-valued function on $[0, \infty)$, such that $\rho(0) = 1$. Write $(d\rho/dr)_{r=0} = 4\pi\beta$. For any $x_1 \in \mathbb{R}^3$, $x_2 \in \mathbb{R}^3$, consider the Hamiltonian H_Γ^Ψ, where $\Psi_j(p) = (2\pi)^{-3/2} e^{ipx_j}$, and where Γ is given by (3.11).

Choose the subtraction matrix $\alpha(x_1, x_2)$ to be

$$\alpha(x_1, x_2) = \begin{pmatrix} \alpha_1 & \dfrac{1}{4\pi r}\rho(r) \\ \dfrac{1}{4\pi r}\rho(r) & \alpha_2 \end{pmatrix},$$

with $r = |x_1 - x_2|$. Let $\kappa(r)$ be a bound state solution of $\det\Gamma(E) = 0$. Then

$$\lim\left(-\frac{\kappa(r)}{4\pi}\right) = \frac{\alpha_1\alpha_2 - \beta^2}{\alpha_1 + \alpha_2 - 2\beta}.$$

Remark: $\rho(r)$ can be chosen so as to vanish (exactly or asymptotically) for r greater than some $r_c > 0$; so the correlation between centers can be made to disappear when the centers are apart from each other.

[1] G.H. Wannier, *Elements of Solid State Theory* (Cambridge University, Cambridge, England, 1959).
[2] J.E. Avron, A. Grossmann, and R. Rodriguez, Ann. Phys. **103**, 47 (1977).
[3] W. Thirring, in *Preludes in Theoretical Physics. Essays in honor of V. Weisskopf*, edited by de Shalit, H. Feschbach, and L. Van Hove (North Holland, Amsterdam, 1960).
[4] S. Albeverio, J.E. Fenstadt, and R. Hoegh-Krohn, "Singular Perturbations and Nonstandard Analysis," CNRS, CPT Marseille preprint, to appear in Trans. Am. Math. Soc.
[5] N. Akhiezer and I.M. Glazman, *Theory of Linear Operators in Hilbert Space* (Vishtcha Shkola, Kharkov, 1978), Chapter VIII.
[6] M. Krein, Mat. Sb. 20 (62) **3**, 431 (1947).
[7] N. Dunford and J.T. Schwartz, *Linear Operators*, Part II (Interscience, New York, 1963), Chapter XII.
[8] H. Feshbach and E.L. Lomon, Ann. Phys. **29**, 19 (1964).
[9] G. Flamand, in *Cargèse Lectures in Theoretical Physics*, edited by F. Lurçat (Gordon and Breach, New York, 1967).
[10] L.H. Thomas, Phys. Rev. **47**, 903 (1935).
[11] S. Albeverio, R. Hoegh-Krohn, and L. Streit, J. Math. Phys. **18**, 907 (1977).
[12] A. Alonso y Coria, "Shrinking Potentials in the Schrödinger Equation," Princeton Univ., Ph.D. Dissertation, 1978 (Department of Mathematics).
[13] S. Albeverio, R. Hoegh-Krohn, and L. Streit "Regularization of Hamiltonians and Processes," Preprint Bielefeld.
[14] C.N. Friedman, J. Funct. Anal. **10**, 346 (1972).
[15] L. Thomas, "Birman–Schwinger Bounds for the Laplacian with Point Interactions" (submitted to J. Math. Phys.).
[16] B. Simon, Ann. Phys. **97**, 279–88 (1976).
[17] J.P. Antoine and A. Grossmann, J. Funct. Anal. **23**, 369 (1976).
[18] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976).
[19] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. I, Functional Analysis (Academic, New York, 1972).
[20] K. Stroyan and W. Luxemburg, *Introduction to the Theory of Infinitesimals* (Academic, New York, 1976).

Representations and properties of para-Bose oscillator operators. I. Energy position and momentum eigenstates

N. Mukunda and E. C. G. Sudarshan
Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560012 India
J. K. Sharma and C. L. Mehta
Department of Physics, Indian Institute of Technology, New Dalhi 110029, India

(Received 7 August 1979; accepted for publication 26 October 1979)

Para-Bose commutation relations are related to the SL(2,R) Lie algebra. The irreducible representation $\mathscr{D}_\alpha$ of the para-Bose system is obtained as the direct sum $D_\beta \oplus D_{\beta+1/2}$ of the representations of the SL(2,R) Lie algebra. The position and momentum eigenstates are then obtained in this representation $\mathscr{D}_\alpha$, using the matrix mechanical method. The orthogonality, completeness, and the overlap of these eigenstates are derived. The momentum eigenstates are also derived using the wave mechanical method by specifying the domain of the definition of the momentum operator in addition to giving it a formal differential expression. By a careful consideration in this manner we find that the two apparently different solutions obtained by Ohnuki and Kamefuchi in this context are actually unitarily equivalent.

1. INTRODUCTION

The kinematical basis of nonrelativistic quantum mechanics is supplied by the Heisenberg commutation relations. Using these relations, one can work out the equations of motion for different dynamical systems described by different Hamiltonians. Almost three decades ago, Wigner posed the interesting question:"Do the equations of motion determine the commutation relations of quantum mechanics?" One would expect the answer to depend on the particular system considered. The simplest system one can examine in this context is the harmonic oscillator. Wigner[1] found in this case that there is a one parameter continuous family of inequivalent operator representations for position and momentum, each of which leads to the standard oscillator equations of motion. This parameter, hereafter denoted by α, is the minimum eigenvalue of the Hamiltonian and is restricted to be real and strictly positive.

In 1953, Green introduced[2] a new method of quantization of free fields leading to the concept of parastatistics for identical particles. Here one specifies the commutation relations of bilinear expressions in the free field creation and annihilation operators with the creation or annihilation operator. These commutators are to have the same values as in the normal case. Depending on whether the bilinear expressions are basically symmetric or antisymmetric in creation and annihilation operators, one speaks of para-Bose or para-Fermi systems. The para-Bose problem of one degree of freedom is identical with the oscillator problem of the preceding paragraph and can moreover be elegantly formulated in the language of the group SL(2, R).

Many interesting mathematical properties connected with the Heisenberg commutation relation are known. Among them we may recall the following: there is the matrix mechanical representation of Heisenberg and the wave mechanical one of Schrödinger[3]; there is the unitary operation of Fourier transformation that interchanges the operators of position and momentum[3]; there is the overcomplete family of coherent states—the eigenstates of the annihilation operator,[4] connected on the one hand with attaining a minimum value for the product of uncertainties in position and momentum and on the other hand with the Bargmann realization of the commutation relation in a Hilbert space of analytic functions[5]; there is the Weyl representation of operators, similar in form to the Fourier integral representation for classical functions of two variables; there is a variety of closely related forms of operator description suggested by various rules of ordering noncommuting operators, prominent among these being the normal ordered and the antinormal ordered forms[6]; and there is the diagonal coherent state representation of operators.[7,8] The existence of this wealth and variety of results is due largely to the relative simplicity of the Heisenberg commutation relation, and one must be prepared to find that those results capable of generalization to the para-Bose case show an increase in complexity.

The purpose of this investigation is to provide a study of the representations of the para-Bse oscillator from many points of view suggested by the catalog of known results for the usual case mentioned above. The construction of the matrix representations of the para-Bose system with the Hamiltonian diagonal is a fairly elementary algebraic exercise.[9-11] The question of a Schrödinger description was examined by Yang[12] in 1951. More recently a thorough analysis of this question has been made by Ohnuki and Kamefuchi.[13] While setting right some of the incorrect conclusions drawn by Yang, certain of the results of Ohnuki and Kamefuchi strike one as quite surprising and unexpected. Thus, these authors find it necessary to invoke the theory of generalized functions and even the notion of a superselection rule in order to interpret their results connected with the wave function descriptions of the momentum and energy operators. One of our tasks will be to clarify this situation completely and show that there is no real need to bring in these ideas—the mathematics can be understood at a much more elementary level.

The results of this investigation have for convenience been divided into two parts. The contents of the present pa-

 0022-2488/80/092386-09$1.00

per (Part I) are as follows: We start in Sec. 2 with a statement of the para-Bose commutation relations and then show how these may be reinterpreted in terms of the group SL(2, R). A relevant class of Hermitian representations of the SL(2, R) Lie algebra is described and these are then used to obtain the matrix representations of the para-Bose system. This recapitulation of known material allows us to introduce basic definitions and notations to be used in the rest of this work. In Sec. 3 the question of representing the para-Bose algebra in a basis where the position operator, rather than the energy, is diagonal is taken up. It turns out that there are two rather natural Schrödinger descriptions of this type and we will follow the route of obtaining them directly from the matrix description. This method allows for a clear and careful discussion of some delicate phase ambiguity questions. The eigenfuctions of momentum and energy, and their normalization and completeness properties, will be obtained. In Sec. 4 we discuss the two Schrödinger descriptions of the momentum operator, and give a detailed comparison with the work of Ohnuki and Kamefuchi. It will be shown that the key to the proper understanding of the para-Bose momentum operator is the specification of its domain of definition in addition to giving a formal differential operator expression to it. Two appendices contain proofs of some of the statements made in Secs. 3 and 4.

In Part II of this investigation we consider the question of para-Bose coherent states and related matters.

2. PARA-BOSE COMMUTATION RELATIONS, AND THE GROUP SL(2, R)

The para-Bose oscillator for one degree of freedom is described by an irreducible pair of operators[14] $\hat{a}, \hat{a}^\dagger$ (adjoints of one another) obeying the commutation relations

$$[\tfrac{1}{2}(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}), \hat{a}] = -\hat{a}, \tag{2.1a}$$

or equivalently

$$[\tfrac{1}{2}(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}), \hat{a}^\dagger] = \hat{a}^\dagger. \tag{2.1b}$$

The bilinear expression here defines the oscillator Hamiltonian:

$$\hat{H} \equiv \tfrac{1}{2}(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}). \tag{2.2}$$

Rearrangement of the terms in Eq. (2.1) allows us to derive therefrom the results

$$[\hat{a}^2, \hat{a}^\dagger] = 2\hat{a}, \quad [\hat{a}^{\dagger 2}, \hat{a}] = -2\hat{a}^\dagger. \tag{2.3}$$

As a matter of fact, from any one of the four relations (2.1) and (2.3) the other three can be derived.

If instead of the annihilation and creation operators we use position and momentum operators[15] $\hat{q}$ and $\hat{p}$ via

$$\hat{a} = 2^{-1/2}(\hat{q} + i\hat{p}), \tag{2.4a}$$
$$\hat{a}^\dagger = 2^{-1/2}(\hat{q} - i\hat{p}), \tag{2.4b}$$

then the expressions for $\hat{H}$ and the para-Bose relations become

$$\hat{H} = \tfrac{1}{2}(\hat{q}^2 + \hat{p}^2), \tag{2.5}$$
$$[\hat{q}^2, \hat{p}] = 2i\hat{q}, \tag{2.6a}$$
$$[\hat{p}^2, \hat{q}] = -2i\hat{p}. \tag{2.6b}$$

If we define three Hermitian operators $\hat{J}_0, \hat{J}_1, \hat{J}_2$ by

$$\hat{J}_0 = \tfrac{1}{2}\hat{H}, \hat{J}_1 = \tfrac{1}{4}(\hat{a}^2 + \hat{a}^{\dagger 2}), \hat{J}_2 = \tfrac{1}{4}i(\hat{a}^2 - \hat{a}^{\dagger 2}), \tag{2.7}$$

then using Eq. (2.1) one can show that they obey the commutation relations corresponding to the SL(2, R) Lie algebra[16]

$$[\hat{J}_0, \hat{J}_1] = i\hat{J}_2, [\hat{J}_0, \hat{J}_2] = -i\hat{J}_1, [\hat{J}_1, \hat{J}_2] = -i\hat{J}_0. \tag{2.8}$$

The para-Bose relations themselves take the following form:

$$\begin{aligned} &[\hat{J}_0, \hat{a}^\dagger] = \tfrac{1}{2}\hat{a}^\dagger, \quad [\hat{J}_0, \hat{a}] = -\tfrac{1}{2}\hat{a}, \\ &[\hat{J}_1 + i\hat{J}_2, \hat{a}^\dagger] = 0, \quad [\hat{J}_1 + i\hat{J}_2, \hat{a}] = -\hat{a}^\dagger, \\ &[\hat{J}_1 - i\hat{J}_2, \hat{a}^\dagger] = \hat{a}, \quad [\hat{J}_1 - i\hat{J}_2, \hat{a}] = 0. \end{aligned} \tag{2.9}$$

Thus, under the unitary representation of (the covering group of) SL(2, R) provided by the $\hat{J}$'s as generators, the doublet of operators $\hat{a}^\dagger, \hat{a}$ behaves as a tensor operator belonging to the nonunitary two dimensional spinor representation of SL(2, R). The operator $\hat{a}^\dagger$ ($\hat{a}$) is the $+\frac{1}{2}$ ($-\frac{1}{2}$) component. Thus, the search for the irreducible representations of the para-Bose system may be viewed as follows: it is a search for Hermitian (reducible) representations of the SL(2, R) Lie algebra which can support a two component spinor operator and in which moreover the SL(2, R) generators are symmetric bilinears in the components of the spinor operator, corresponding to two spin $\frac{1}{2}$ entities being coupled to spin 1. Recalling that locally the group SL(2, R) is the same as the three dimensional Lorentz group[16] SO(2, 1), this situation reminds one of the class of relativistic wave equations discovered by Bhabha,[17] wherein the six generators of homogeneous Lorentz transformations in four dimensional space–time are required to be the commutators among the components of the four-vector operator in the wave equation.

The Hermitian representations (irreducible) of the SL(2, R) Lie algebra (2.8) relevant for our analysis may be labeled by a real positive parameter β, $0 < \beta < \infty$. We write D_β for these representations; their essential feature is that $\hat{J}_0$ is positive definite in them, and in fact β is the lowest eigenvalue of $\hat{J}_0$ in D_β. The eigenvalues of $\hat{J}_0$ are integrally spaced and the corresponding eigenvectors form a complete, orthonormal basis for the space of the representation:

$$\hat{J}_0|n;\beta\rangle = (n+\beta)|n;\beta\rangle, \tag{2.10}$$
$$\langle n';\beta|n;\beta\rangle = \delta_{n'n}. \tag{2.11}$$

It is convenient to require the phases of these vectors to be so chosen that the nonvanishing matrix elements of the raising and lowering operators

$$\hat{J}_\pm \equiv \hat{J}_1 \pm i\hat{J}_2 \tag{2.12}$$

are all real positive. The description of D_β is then completed by adjoining to Eq. (2.10) the equations

$$\hat{J}_+|n;\beta\rangle = [(n+1)(n+2\beta)]^{1/2}|n+1;\beta\rangle, \tag{2.13a}$$
$$\hat{J}_-|n;\beta\rangle = [n(n+2\beta-1)]^{1/2}|n-1;\beta\rangle. \tag{2.13b}$$

The only remaining phase ambiguity is the freedom to change the phases of all the basis vectors by one common factor. For $\beta = \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$, we obtain single-valued unitary SL(2, R) representations on exponentiating the generators; otherwise we obtain representations of the covering group of SL(2, R).

The irreducible representations of the para-Bose algebra will be written $\mathscr{D}_\alpha$, indexed by the minimum eigenvalue

of $\hat{H}$ in the representation. The eigenvalues of $\hat{H}$ are integrally spaced and the corresponding eigenvectors form a complete orthonormal set:

$$\hat{H}\,|n;\alpha\rangle = (n+\alpha)|n;\alpha\rangle, \tag{2.14}$$
$$\langle n';\alpha|n;\alpha\rangle = \delta_{n'n}. \tag{2.15}$$

The appearance of $\alpha(\beta)$ in a ket or bra vector will indicate that it belongs to the space of a para-Bose [SL(2, R)] representation. The space of the representation $\mathscr{D}_\alpha$ is obtained as the direct sum of two spaces carrying the SL(2, R) representations D_β and $D_{\beta+(1/2)}$ for $\beta = \frac{1}{2}\alpha$. We may loosely indicate this as

$$\mathscr{D}_\alpha = D_\beta \oplus D_{\beta+1/2},$$
$$\beta = \tfrac{1}{2}\alpha. \tag{2.16}$$

The eigenvectors of $\hat{H}$ are obtained by adjoining the eigenvectors of $\hat{J}_0$ in D_β with those in $D_{\beta+1/2}$:

$$|2l;\alpha\rangle = |l;\beta\rangle,$$
$$|2l+1;\alpha\rangle = |l;\beta+\tfrac{1}{2}\rangle,\ l = 0, 1, 2, \ldots. \tag{2.17}$$

We have of course the condition

$$\langle n';\beta+\tfrac{1}{2}|n;\beta\rangle = 0. \tag{2.18}$$

Then, with phases chosen so that all nonzero matrix elements of $\hat{a}$ and $\hat{a}^\dagger$ are real positive, the para-Bose representation $\mathscr{D}_\alpha$ is fully determined by the relations

$$\hat{a}|\hat{n};\beta\rangle = (2n)^{1/2}|n-1;\beta+\tfrac{1}{2}\rangle, \tag{2.19a}$$
$$\hat{a}|\hat{n},\beta+\tfrac{1}{2}\rangle = (2n+4\beta)^{1/2}|n;\beta\rangle, \tag{2.19b}$$
$$\hat{a}^\dagger|n;\beta\rangle = (2n+4\beta)^{1/2}|n;\beta+\tfrac{1}{2}\rangle, \tag{2.19c}$$
$$\hat{a}^\dagger|n,\beta+\tfrac{1}{2}\rangle = (2n+2)^{1/2}|n+1,\beta\rangle. \tag{2.19d}$$

These relations can be easily rewritten in terms of $|n;\alpha\rangle$. The matrix elements then have different analytic forms for odd and even n (cf. Ref. 11):

$$\hat{a}|2n;\alpha\rangle = (2n)^{1/2}|2n-1;\alpha\rangle, \tag{2.20a}$$
$$\hat{a}|2n+1;\alpha\rangle = (2n+2\alpha)^{1/2}|2n;\alpha\rangle, \tag{2.20b}$$
$$\hat{a}^\dagger|2n;\alpha\rangle = (2n+2\alpha)^{1/2}|2n+1;\alpha\rangle, \tag{2.20c}$$
$$\hat{a}^\dagger|2n+1;\alpha\rangle = (2n+2)^{1/2}|2n+2;\alpha\rangle. \tag{2.20d}$$

Since $\hat{a}^\dagger$ and $\hat{a}$ alter the eigenvalues of $\hat{H}$ by ± 1, there are two, and just two, choices for a unitary operator $\hat{R}$ which will reverse the sign of $\hat{a}^\dagger$ and $\hat{a}$, and whose square is unity:

$$\hat{R}\hat{a}\hat{R} = -\hat{a},\ \hat{R}\hat{a}^\dagger\hat{R} = -\hat{a}^\dagger,\ \hat{R}^2 = \hat{R}^\dagger\hat{R} = \hat{1}. \tag{2.21}$$

We may either take $\hat{R}$ to be $\hat{R}_1$ defined by

$$\hat{R}_1|n;\alpha\rangle = (-1)^n|n;\alpha\rangle \tag{2.22}$$

or we may take it to be $\hat{R}_2$, the negative of $\hat{R}_1$:

$$\hat{R}_2|n;\alpha\rangle = (-1)^{n+1}|n;\alpha\rangle. \tag{2.23}$$

If $\hat{R}$ is interpreted as the parity, then the first choice corresponds to $\hat{H}$ having an even parity ground state and the second to an odd parity ground state. The commutator between $\hat{a}$ and $\hat{a}^\dagger$ can now be calculated in $\mathscr{D}_\alpha$. The result is an operator expressible in terms of $\hat{R}_1$ or $\hat{R}_2$:

$$[\hat{a},\hat{a}^\dagger] = 1 + (2\alpha-1)\hat{R}_1$$
$$= 1 + (1-2\alpha)\hat{R}_2. \tag{2.24}$$

Stated in terms of $\hat{q}$ and $\hat{p}$ this reads

$$[\hat{q},\hat{p}] = i + i(2\alpha-1)\hat{R}_1$$
$$= i + i(1-2\alpha)\hat{R}_2. \tag{2.25}$$

For $\alpha = \frac{1}{2}$, the right hand sides of these commutators reduce to c numbers, and the para-Bose representation $\mathscr{D}_{1/2}$ becomes the normal Bose representation.

3. POSITION AND MOMENTUM EIGENSTATES—MATRIX MECHANICAL SOLUTION

In the previous section, we discussed the para-Bose representation $\mathscr{D}_\alpha$ in a basis made up of the discrete sequence of eignevectors of $\hat{H}$. It is important to stress that once we make the convention that all nonzero matrix elements of $\hat{a}$ and $\hat{a}^\dagger$ are real and positive, the only freedom left in the choice of these eigenvectors is to alter the phase of all of them by a single common factor. Now we show how we may give Schrödinger descriptions of $\mathscr{D}_\alpha$ with $\hat{q}$ diagonal.

One expects that $\hat{q}$ (as well as $\hat{p}$) has a continuous spectrum running from $-\infty$ to $+\infty$. This will indeed be so. We use letters $x, x', \ldots$ to denote eigenvalues of $\hat{q}$ and letters k, k' to denote those of $\hat{p}$. We expand an eigenvector of $\hat{q}$ as

$$|x;\alpha\rangle = \sum_{n=0}^{\infty} C_n(x;\alpha)|n;\alpha\rangle, \tag{3.1}$$

and attempt to find the expansion coefficients by equating terms in the eigenvalue equation

$$\hat{q}|x;\alpha\rangle = 2^{-1/2}(\hat{a}+\hat{a}^\dagger)|x;\alpha\rangle$$
$$= x|x;\alpha\rangle. \tag{3.2}$$

Using Eq. (2.20), we obtain the system of recursion relations

$$\alpha^{1/2}C_1(x;\alpha) = xC_0(x;\alpha), \tag{3.3a}$$
$$(n+\alpha)^{1/2}C_{2n+1}(x;\alpha) + n^{1/2}C_{2n-1}(x;\alpha)$$
$$= xC_{2n}(x;\alpha), \quad n = 1, 2, \ldots, \tag{3.3b}$$
$$(n+1)^{1/2}C_{2n+2}(x;\alpha) + (n+\alpha)^{1/2}C_{2n}(x;\alpha)$$
$$= xC_{2n+1}(x;\alpha), \quad n = 0, 1, 2, \ldots. \tag{3.3c}$$

It is immediately evident that each C_n is unambiguously determined as some definite multiple of C_0. They turn out to be expressible in terms of associated Laguerre polynomials:

$$C_{2l}(x;\alpha) = C_0(x;\alpha)(-1)^l \times [l!\Gamma(\alpha)/\Gamma(l+\alpha)]^{1/2}L_l^{\alpha-1}(x^2).$$
$$C_{2l+1}(x;\alpha) = C_0(x,\alpha)(-1)^l \times [l!\Gamma(\alpha)/\Gamma(l+1+\alpha)]^{1/2}xL_l^{\alpha}(x^2). \tag{3.4}$$

The following two relations among these polynomials [Ref. 18 p. 1037, formulas (8.971.5) and (8.971.4)] compared, respectively, with Eqs. (3.3b) and (3.3c) lead to the solution (3.4):

$$L_n^{\alpha-1}(z) = L_n^\alpha(z) - L_{n-1}^\alpha(z),$$
$$zL_n^\alpha(z) = (n+\alpha)L_n^{\alpha-1}(z) - (n+1)L_{n+1}^{\alpha-1}(z). \tag{3.5}$$

At this stage, therefore, the eigenvectors of $\hat{q}$ appear as

$$|x;\alpha\rangle = C_0(x;\alpha)[\Gamma(\alpha)]^{1/2} \times \sum_{l=0}^{\infty}(-1)^l[l!/\Gamma(l+\alpha)]^{1/2} \times \Big\{L_l^{\alpha-1}(x^2)|2l;\alpha\rangle + \frac{x}{(l+\alpha)^{1/2}}L_l^\alpha(x^2)|2l+1;\alpha\rangle\Big\}, \tag{3.6}$$

with $C_0(x;\alpha)$ yet to be chosen.

It is possible to evaluate the inner product of these eigenvectors of $\hat{q}$ with one another. The details are given in Appendix A. One finds that

$$\langle x';\alpha|x;\alpha\rangle = |C_0(x;\alpha)|^2\Gamma(\alpha)e^{x^2}|x|^{1-2\alpha}\delta(x'-x). \quad (3.7)$$

We choose C_0 so that the coefficient of the delta function becomes unity. This still leaves the eigenvectors of $\hat{q}$ arbitrary up to an x-dependent phase; denoting this phase by $\xi(x)$, we have the most general solution for delta-function normalized eigenvectors of $\hat{q}$ as

$$|x;\alpha\rangle = \exp[i\xi(x)]\exp(-\tfrac{1}{2}x^2)|x|^{\alpha-1/2} \times \sum_{l=0}^{\infty}(-1)^l[l!/\Gamma(l+\alpha)]^{1/2} \times\Big\{L_l^{\alpha-1}(x^2)|2l;\alpha\rangle + \frac{x}{(l+\alpha)^{1/2}}L_l^{\alpha}(x^2)|2l+1;\alpha\rangle\Big\}, \quad (3.8)$$

$$\langle x';\alpha|x;\alpha\rangle = \delta(x'-x).$$

(A possible dependence of ξ on α is, for simplicity, not explicitly indicated.)

Before considering ways of resolving the remaining phase ambiguity, we use the results so far obtained to define a special sequence of functions of x and to set up eigenvectors for $\hat{p}$. We define, on the basis of Eq. (3.8),

$$\psi_n(x;\alpha) = \exp(-\tfrac{1}{2}x^2)|x|^{\alpha-(1/2)}\begin{cases}(-1)^l[l!/\Gamma(l+\alpha)]^{1/2}L_l^{\alpha-1}(x^2), & n=2l,\\ (-1)^l[l!/\Gamma(l+1+\alpha)]^{1/2}xL_l^{\alpha}(x^2), & n=2l+1,\end{cases} \quad l=0,1,2,\ldots, \quad (3.9)$$

so that

$$|x;\alpha\rangle = \exp[i\xi(x)]\sum_{n=0}^{\infty}\psi_n(x;\alpha)|n;\alpha\rangle. \quad (3.10)$$

By definition, $\psi_n(x;\alpha)$ is an even (odd) function of x for n even (odd). We can exploit the known orthogonality properties of associated Languerre polynomials [Ref. 18, p. 844, formula (7.414.3)] to obtain

$$\int_{-\infty}^{\infty}\psi_{n'}^*(x;\alpha)\psi_n(x;\alpha)\,dx = \delta_{n'n}, \quad (3.11a)$$

while the last line of Eq. (3.8) could be written as

$$\sum_{n=0}^{\infty}\psi_n(x;\alpha)\psi_n^*(x';\alpha) = \delta(x'-x). \quad (3.11b)$$

Next the eigenvectors $|k;\alpha\rangle$ of $\hat{p}$ can be easily obtained by using the unitary connection[19] between $\hat{q}$ and $\hat{p}$:

$$\hat{p} = \exp(i\frac{\pi}{2}\hat{H})\hat{q}\exp(-i\frac{\pi}{2}\hat{H}). \quad (3.12)$$

Allowing for an arbitrary phase factor $\eta(k)$, the most general delta function normalized eigenvectors of $\hat{p}$ have the form

$$|k;\alpha\rangle = \exp[i\eta(k)]\exp(-\tfrac{1}{2}k^2)|k|^{\alpha-1/2} \times\sum_{l=0}^{\infty}[l!/\Gamma(l+\alpha)]^{1/2} \times\Big\{L_l^{\alpha-1}(k^2)|2l;\alpha\rangle + \frac{ik}{(l+\alpha)^{1/2}}L_l^{\alpha}(k^2)|2l+1;\alpha\rangle\Big\},$$

$$\langle k';\alpha|k;\alpha\rangle = \delta(k'-k). \quad (3.13)$$

Using expansions (3.8) and (3.13) for the eigenvectors of $\hat{q}$ and $\hat{p}$, we may evaluate their overlap. As shown in Appendix A we find that

$$\langle x;\alpha|k;\alpha\rangle = \exp[i\eta(k)-i\xi(x)]\tfrac{1}{2}|xk|^{1/2} \times[J_{\alpha-1}(|xk|)+i\epsilon(xk)J_{\alpha}(|xk|)]. \quad (3.14)$$

Here J_α is the usual Bessel function and $\epsilon(x)$ is the sign of x. We may rewrite Eq. (3.14) in the form

$$\langle x;\alpha|k;\alpha\rangle = \tfrac{1}{2}|xk|^{\alpha-1/2}\exp\{i\eta(k)-i\xi(x)\}\mathcal{F}_\alpha(ixk), \quad (3.15)$$

where $\mathcal{F}_\alpha$ is an entire analytic function

$$\mathcal{F}_\alpha(z) = z^{1-\alpha}[I_{\alpha-1}(z)+I_\alpha(z)], \quad (3.16)$$

and I_α is the modified Bessel Function. Equation (3.15) readily follows from Eq. (3.14) on making use of the relation $I_\alpha(ixk) = i^\alpha J_\alpha(xk)$ and observing that $x^{-\alpha}J_\alpha(x)$ is an even function of x. We shall find in Paper II that $\mathcal{F}_\alpha(z)$ also appears while considering the normalization of the para-Bose coherent states.

On account of the delta-function normalization of the state $|x;\alpha\rangle$ we may therefore write

$$|k;\alpha\rangle = e^{i\eta(k)}\int_{-\infty}^{\infty}e^{-i\xi(x)}|xk|^{\alpha-1/2} \times\frac{1}{2}\mathcal{F}_\alpha(ixk)|x;\alpha\rangle dx, \quad (3.17a)$$

and equivalently

$$|x;\alpha\rangle = e^{i\xi(x)}\int_{-\infty}^{\infty}e^{-i\eta(k)}|xk|^{\alpha-1/2} \times\frac{1}{2}\mathcal{F}_\alpha(-ixk)|k;\alpha\rangle dk. \quad (3.17b)$$

The position and momentum representations of any state $|\varphi\rangle$, viz.,

$$|\varphi\rangle = \int\varphi(x;\alpha)|x;\alpha\rangle dx = \int\Phi(k;\alpha)|k;\alpha\rangle dk,$$

are then related by

$$\phi(x;\alpha) = e^{-i\xi(x)}\int_{-\infty}^{\infty}|xk|^{\alpha-1/2} \times\frac{1}{2}\mathcal{F}_\alpha(ixk)e^{i\eta(k)}\Phi(k;\alpha)dk, \quad (3.18a)$$

$$\Phi(k;\alpha) = e^{-i\eta(k)}\int_{-\infty}^{\infty}|xk|^{\alpha-1/2} \times\frac{1}{2}\mathcal{F}_\alpha(-ixk)e^{i\xi(x)}\phi(x;\alpha)dx, \quad (3.18b)$$

We now consider natural choices for the hitherto arbitrary phase factor ξ. If we demand that under the action of $\hat{R}_1$ the eigenvector of $\hat{q}$ with eigenvalue x is changed into the one with eigenvalue $-x$, it is easy to see that we may set $\xi = 0$. With this phase convention, we shall write $|x; \alpha; 1\rangle$ for the eigenvectors of $\hat{q}$; we shall say that this leads to the first Schrödinger description of $\mathscr{D}_\alpha$. In this description the Hilbert space carrying the representation $\mathscr{D}_\alpha$ appears as the space of all (Lebesgue) square integrable functions of x in the range $-\infty$ to ∞, with $\hat{q}$ being the operator of multiplication by x, and $\hat{R}_1$ being the parity operator reversing the sign of the argument x of a general wave function [cf. Eqs. (3.10) and (3.11)]. Thus, $|x; \alpha; 1\rangle$ is given by the right hand side of Eq. (3.8) with $\xi = 0$, and

$$\hat{R}_1|x; \alpha; 1\rangle = |-x; \alpha; 1\rangle. \tag{3.19}$$

Moreover, in this description the wave functions corresponding to eigenvectors of $\hat{p}$ and of $\hat{H}$ take the forms [cf. Eqs. (3.14) and (3.10)]

$$\begin{aligned}\langle x; \alpha; 1|k; \alpha\rangle &\equiv \psi_k^{(1)}(x; \alpha) \\ &= \tfrac{1}{2}|xk|^{1/2}\{J_{\alpha-1}(|xk|) + i\epsilon(xk)J_\alpha(|xk|)\};\end{aligned} \tag{3.20a}$$

$$\begin{aligned}\langle x; \alpha; n; \alpha\rangle &\equiv \psi_n^{(1)}(x; \alpha) = \psi_n(x; \alpha) \\ &= e^{-(1/2)x^2}|x|^{\alpha-1/2}\begin{cases}(-1)^l[l!/\Gamma(l+\alpha)]^{1/2}L_l^{\alpha-1}(x^2), & n = 2l, \\ (-1)^l[l!/\Gamma(l+1+\alpha)]^{1/2}xL_l^\alpha(x^2), & n = 2l+1,\end{cases} \quad l = 0, 1, 2, \ldots .\end{aligned} \tag{3.20b}$$

Here, we have set $\eta = 0$. Recalling the condition $0 < \alpha < \infty$, we see that the above wave functions are exactly what Ohnuki and Kamefuchi have called case (i). In this case, their parameter $c = \alpha - \frac{1}{2} > -\frac{1}{2}$.

Instead of the above convention, we can consider the alternative one wherein the operator $\hat{R}_2$ plays the role of parity. We demand that ξ be so chosen that an equation like (3.19) be obeyed with $\hat{R}_2$ in place of $\hat{R}_1$. Such a choice for ξ exists, but it is not a continuous function of x: we must set $\xi(x) = 0$ for $x > 0$ and $\xi(x) = \pi$ for $x < 0$. With this choice, we shall write $|x; \alpha; 2\rangle$ for the right hand side of Eq. (3.8), which then obeys

$$\hat{R}_2|x; \alpha; 2\rangle = |-x; \alpha; 2\rangle. \tag{3.21}$$

This leads to what we shall call the second Schrödinger description of $\mathscr{D}_\alpha$. As in the first Schrödinger description, here again the Hilbert space of the representation $\mathscr{D}_\alpha$ consists of all (Lebesgue) square integrable functions of x in the range $-\infty$ to ∞, because of Eq. (3.10) and (3.11). Moreover, $\hat{q}$ again is a multiplication by x. The difference lies in the fact that is is now $\hat{R}_2$, not $\hat{R}_1$, whose effect on a general wave function is given as reversing the sign of the argument. To emphasize this difference between these two Schrödinger descriptions, it may help to put the matter as follows: Suppose we take a vector $|\varphi\rangle$ lying in the space of the representation $\mathscr{D}_\alpha$ and describe in the basis of eigenvectors of $\hat{H}$ by the sequence $\{\phi_n\}$:

$$|\varphi\rangle = \sum_{n=0}^{\infty}\phi_n|n; \alpha\rangle, \ \Sigma|\varphi_n|^2 < \infty. \tag{3.22}$$

When we pass to the first Schrödinger description of $\mathscr{D}_\alpha$ we assign to $|\varphi\rangle$ an L_2 function $\phi^{(1)}(x)$ by the correspondence

$$\varphi^{(1)}(x) \equiv \langle x; \alpha; 1|\varphi\rangle = \sum_{n=0}^{\infty}\varphi_n\psi_n^{(1)}(x; \alpha). \tag{3.23}$$

Then the actions of $\hat{R}_1$ and $\hat{R}_2$ on these wave functions are

$$(\hat{R}_1\varphi)^{(1)}(x) = \varphi^{(1)}(-x), \ (\hat{R}_2\varphi)^{(1)}(x) = -\varphi^{(1)}(-x). \tag{3.24}$$

On the other hand, in the second Schrödinger description of $\mathscr{D}_\alpha$, we assign to $|\varphi\rangle$ determined by the same l_2 sequence $\{l_n\}$ a new L_2 function $\varphi^{(2)}(x)$:

$$\varphi^{(2)}(x) \equiv \langle x; \alpha; 2|\varphi\rangle = \sum_{n=0}^{\infty}\phi_n\psi_n^{(2)}(x; \alpha). \tag{3.25}$$

The $\psi_n^{(2)}$ is defined below [Eq. (3. 29)]. In place of Eq. (3.24) we now have

$$(\hat{R}_2\varphi)^{(2)}(x) = \varphi^{(2)}(-x); \ (\hat{R}_1\varphi)^{(2)}(x) = -\varphi(-x), \tag{3.26}$$

and moreover for a fixed $\{\varphi_n\}$,

$$\varphi^{(2)}(x) = \epsilon(x)\varphi^{(1)}(x). \tag{3.27}$$

As $\{\varphi_n\}$ varies over all l_2 sequences, the set of all $\phi^{(1)}(x)$ we get coincides with the set of all $\phi^{(2)}(x)$, and this is the L_2 space over the real line. Thus, the difference between the two Schrödinger descriptions is merely the difference between two equally valid unitary correspondences to go from l_2 to $L_2(-\infty, \infty)$. The two Schrödinger descriptions are unitarily related via the transformation

$$\psi(x) \to \epsilon(x)\psi(x). \tag{3.28}$$

As replacement for Eqs. (3.20), we have the following wave functions for eigenvectors of $\hat{p}$ and $\hat{H}$ in the second Schrödinger description:

$$\begin{aligned}&\langle x; \alpha; 2|k; \alpha\rangle \\ &\quad \equiv \psi_k^{(2)}(x; \alpha) = \epsilon(x)\psi_k^{(1)}(x; \alpha) \\ &\quad = i\epsilon(k)\tfrac{1}{2}|xk|^{1/2}\{J_\alpha(|xk|) - i\epsilon(xk)J_{\alpha-1}(|xk|)\};\end{aligned} \tag{3.29a}$$

$$\begin{aligned}\langle x; \alpha; 2|n, \alpha\rangle &\equiv \psi_n^{(2)}(x, \alpha) = \epsilon(x)\psi_n^{(1)}(x; \alpha) \\ &= e^{-(1/2)x^2}|x|^{\alpha-1/2}\begin{cases}(-1)^l[l!/\Gamma(l+\alpha)]^{1/2}\epsilon(x)L_l^{\alpha-1}(x^2), & n = 2l, \\ (-1)^l[l!/\Gamma(l+1+\alpha)]^{1/2}|x|L_l^\alpha(x^2), & n = 2l+1,\end{cases} \quad l = 0, 1, 2, \ldots .\end{aligned} \tag{3.29b}$$

The freedom in the choice of η has not been used in Eq. (3.29a); it has in fact been set equal to zero there. If instead we use this freedom to remove the factor $i\epsilon(k)$, we then see that the above momentum and energy eigenfuctions are ex-

actly what Ohnuki and Kamefuchi have called case (ii), their prarameter c being now given by $c = \frac{1}{2} - \alpha$ and so obeying $c < \frac{1}{2}$.

For the sake of completeness it is shown in Appendix A that the momentum eigenfunctions are properly orthonormalized:

$$\int_{-\infty}^{\infty} dx\, \psi_k^{(1)}(x;\alpha)^* \psi_{k'}^{(1)}(x;\alpha) = \delta(k' - k). \tag{3.30}$$

The same equation holds with 2 in place of 1 as superscript. Since these wave functions depend only on the product xk, orthonormality and completeness mean the same thing.

It is worth observing at this stage that if we set $\alpha = \frac{1}{2}$, we obtain the well known results of the normal oscillator. Thus, for instance, Eq. (3.9) for $\alpha = \frac{1}{2}$ reduces to

$$\psi_n(x;\tfrac{1}{2}) = (\pi^{1/2} 2^n n!)^{-1/2} H_n(x) e^{-1/2x^2}, \tag{3.31}$$

whereas Eq. (3.15) gives (ignoring the phases ξ and η)

$$\langle x;\tfrac{1}{2}|k;\tfrac{1}{2}\rangle = (2\pi)^{-1/2} \exp(ixk), \tag{3.32}$$

$H_n(x)$ being the Hermite polynomial. Equations (3.31) and (3.32) give the well known energy and momentum eigenfunctions. Equations (3.18) reduce to ordinary Fourier transform relationships.

4. WAVE MECHANICAL SOLUTION FOR MOMENTUM EIGENFUNCTIONS

The Schrödinger descriptions of $\mathscr{D}_\alpha$ and the eigenfunctions of $\hat{p}$ and $\hat{H}$ have so far been obtained by working directly with the matrix representation of $\mathscr{D}_\alpha$. In particular these eigenfunctions have not been obtained by solving (coupled) differential equations arrived at by first obtaining a formal solution of the $\hat{q}, \hat{p}$ commutation relation (2.25). We now address ourselves to the question whether in each of the Schrödinger descriptions of $\mathscr{D}_\alpha$, the action of $\hat{p}$ on a general wave function could be given by a formal differential operator.

We begin with the first Schrödinger description in which the momentum eigenfunctions in the representation $\mathscr{D}_\alpha$ are given by Eq. (3.20a). Proceeding purely formally, Eq. (2.25) in terms of $\hat{R}_1$ suggests that the action of $\hat{p}$ on a general wave function $\psi(x)$ might be given by

$$\begin{aligned}(\hat{p}\psi)(x) &= \left(-i\frac{d}{dx} + i\,\frac{\alpha - \frac{1}{2}}{x}\hat{P}_x\right)\psi(x) \\ &= -i\frac{d\psi(x)}{dx} + i\,\frac{\alpha - \frac{1}{2}}{x}\psi(-x).\end{aligned} \tag{4.1}$$

Here $\hat{P}_x$ is an operator that changes the sign of the argument of the function standing to its right, and we remember that in the first Schrödinger description $\hat{R}_1$ is realized as $\hat{P}_x$. Now it is a fact that the momentum eigenfunctions $\psi_k^{(1)}(x;\alpha)$ are indeed formally the eigenfunctions of the operator appearing in Eq. (4.1.); thus, the following holds:

$$\begin{aligned}&-i\frac{d\psi_k^{(1)}(x;\alpha)}{dx} + i\,\frac{\alpha - \frac{1}{2}}{x}\psi_k^{(1)}(-x;\alpha) \\ &\quad = k\psi_k^{(1)}(x;\alpha).\end{aligned} \tag{4.2}$$

[This is shown in Appendix B; incidentally, it is not necessary for us to discuss whether Eq. (4.2) is actually obeyed at $x = 0$; its validity for $x \neq 0$ is all we need]. We can write Eq. (4.2) as

$$\begin{aligned}&-i\frac{d}{dx}\langle x;\alpha;1|k;\alpha\rangle + i\frac{\alpha - \frac{1}{2}}{x}\langle -x,\alpha;1|k;\alpha\rangle \\ &\quad = \langle x;\alpha;1|\hat{p}|k;\alpha\rangle.\end{aligned} \tag{4.3}$$

Since the set of eigenvectors $|k;\alpha\rangle$ of $\hat{p}$ in $\mathscr{D}_\alpha$ is complete, we can "peel off" the ket $|k;\alpha\rangle$ in Eq. (4.3) and put in its place a general ket $|\psi\rangle$ having a Schrödinger wave function $\psi(x)$; we then see that the action of $\hat{p}$ is indeed as expressed by Eq. (4.1) in the first Schrödinger description.

However, the formal statement

$$\hat{p} = -i\frac{d}{dx} + i\,\frac{\alpha - \frac{1}{2}}{x}\hat{P}_x \tag{4.4}$$

does not suffice to completely characterize the action of $\hat{p}$ in the first Schrödinger description of $\mathscr{D}_\alpha$. As with any formal operator expression intended to determine some unbounded operator, one must specify the domain of definition of the operator. It is the formal expression together with its domain that combine to determine a definite operator; indeed it can happen (as we will see) that an operator expression can be taken with several distinct domains and then each choice of domain goes to determine one definite operator. In our case, we must find all possible choices of domian for the formal operator (4.4) such that (i) the domain is dense in the Hilbert space and (ii) the above operator is Hermitian (symmetric) on this domain. Furthermore, the domain must not be unnecessarily small for otherwise one could lose the self-adjointness property for the operator one is trying to define. After all this, one must pick just the right domain so that one is then describing the momentum operator "really belonging to $\mathscr{D}_\alpha$. We now see how to do all this.

Let $\psi(x)$ be a general square integrable function defined over the range $-\infty$ to ∞. Knowledge of $\psi(x)$ for all real x is completely equivalent to the knowledge of

$$\psi_\pm(x) = \tfrac{1}{2}\{\psi(x) \pm \psi(-x)\}, \tag{4.5}$$

For all real positive x, and moreover

$$\int_{-\infty}^{\infty} dx |\psi(x)|^2 = 2\int_0^{\infty} dx\{|\psi_+(x)|^2 + |\psi_-(x)|^2\}. \tag{4.6}$$

In particular, it is not necessary to define negative argument. Now the action of the formal operator (4.4) on $\psi_\pm$ is clearly as follows:

$$\begin{aligned}(\hat{p}\psi)_+(x) &= -ix^{(1/2)-\alpha}\frac{d}{dx}(x^{\alpha-1/2}\psi_-(x)), \\ (\hat{p}\psi)_-(x) &= -ix^{\alpha-1/2}\frac{d}{dx}(x^{(1/2)-\alpha}\psi_+(x)), \qquad x > 0.\end{aligned} \tag{4.7}$$

Thus, the first condition for ψ to be in the domain of $\hat{p}$, viz., that along with ψ, $\hat{p}\psi$ also be an L_2 function, leads to the following requirements: (i) ψ should be everywhere differentiable and its first derivative should be square integrable at infinity; (ii) the behaviors of $\psi_\pm$ near $x = 0$ must be such as to ensure

$$\int_0 dx\; x^{1-2\alpha}\left|\frac{d}{dx}(x^{\alpha-1/2}\psi_-(x))\right|^2 < \infty,$$

$$\int_0 dx\ x^{2\alpha-1}\left|\frac{d}{dx}(x^{(1/2)-\alpha}\psi_+(x))\right|^2<\infty. \tag{4.8}$$

We shall hereafter assume the necessary natural conditions at infinity are obeyed. Near $x=0$, the requirements (4.8) restrict ψ_+ and ψ_- to such behaviors:

$$\psi_+(x)=\text{constant}\cdot x^{\alpha-1/2}+\text{ powers of } x>\tfrac{1}{2},$$
$$\psi_-(x)=\text{constant}\cdot x^{(1/2)-\alpha}+\text{ powers of } x>\tfrac{1}{2}. \tag{4.9}$$

For $0<\alpha<1$, such behaviors do not conflict with ψ, itself being L_2; for $\alpha\geqslant 1$ the constant accompanying $x^{(1/2)-\alpha}$ in ψ_- must clearly be zero.

Let us now take two L_2 functions $\phi(x)$ and $\psi(x)$, each behaving near $x=0$ as described by Eq. (4.9):

$$\psi_+\sim ax^{\alpha-1/2}+bx^{\lambda},\quad \psi_-\sim cx^{(1/2)-\alpha}+d\ x^{\lambda},$$
$$\phi_+\sim a'x^{\alpha-1/2}+b'x^{\lambda},\quad \varphi_-\sim c'x^{(1/2)-\alpha}+d'x^{\lambda}, \tag{4.10}$$

The λ's need not all be the same, but symbolically denote terms where the exponent of x is greater than $\frac{1}{2}$. Also, if $\alpha\geqslant 1$ we understand $c=c'=0$. We now ask under what conditions we can have the equality

$$(\phi,\hat{p}\psi)=(\hat{p}\varphi,\psi). \tag{4.11}$$

Assuming a nice behavior at infinity we have

$$\begin{aligned}&(\phi,\hat{p}\psi)-(\hat{p}\varphi,\psi)\\ &=2\int_0^\infty dx[\varphi^*_+(\hat{p}\psi)_+ +\phi^*_-(\hat{p}\psi)_-\\ &\quad-(\hat{p}\varphi)^*_+\psi_+-(\hat{p}\varphi)^*_-\psi_-]\\ &=-2i\int_0^\infty dx\frac{d}{dx}[x^{(1/2)-\alpha}\varphi^*_+x^{\alpha-1/2}\psi_-\\ &\quad+x^{\alpha-1/2}\varphi^*_-x^{(1/2)-\alpha}\psi_+]\\ &=2i[(a'+b'x^{\lambda+(1/2)-\alpha})^*(c+d\ x^{\lambda+\alpha-1/2})\\ &\quad+(c'+d'x^{\lambda+\alpha-1/2})^*(a+bx^{\lambda+(1/2)-\alpha})]\\ &=2i\begin{cases}a'^*c+c'^*a, & \text{if } 0<\alpha<1,\\ 0, & \text{if } \alpha\geqslant 1.\end{cases}\end{aligned} \tag{4.12}$$

From here we draw the following conclusions: If $\alpha\geqslant 1$, there is only one way in which to choose a domain for the formal operator (4.4), so that a unique self-adjoint operator may be defined; the domain must consist of all L_2 functions $\psi(x)$ which are once differentiable, the derivative of ψ must be square integrable at infinity, and near $x=0$, $\psi_\pm$ must behave as

$$\psi_+\sim ax^{\alpha-1/2}+bx^{\lambda},\ \psi_-\sim d\ x^{\lambda},\ \lambda>\tfrac{1}{2}. \tag{4.13}$$

For $\alpha>1$ the first term in ψ_+ here need not be distinguished from the "λ" terms). If $0<\alpha<1$, then for each choice of a real parameter s we have a possible domain M_s in the Hilbert space that can be chosen as the domain of definition for the formal operator (4.4), and that will then lead to a definite self-adjoint unbounded operator. A change in s entails a change in this last operator. Apart from conditions at infinity and of differentiability, the elements of M_s are characterized by the fact that, near $x=0$,

$$\psi_+\sim ax^{\alpha-1/2}+bx^{\lambda},\ \psi_-\sim isax^{(1/2)-\alpha}+d\ x^{\lambda}. \tag{4.14}$$

If both ϕ and ψ belong to M_s, the right hand side of Eq. (4.12) vanishes and Eq. (4.11) is secured.

We are now in a position to give the complete specification of the operator $\hat{p}$ of the para-Bose algebra when we choose the first Schrödinger description of $\mathscr{D}_\alpha$. The situation for $\alpha\geqslant 1$ is very simple: $\hat{p}$ acts as stated in Eqs. (4.1) and (4.7) and has as domain all wave functions which, apart from differentiability and suitable behavior at infinity, look like Eq. (4.13) near $x=0$. It is easily seen that the energy eigenfunctions and momentum eigenfunctions given in Eq. (3.20) indeed conform to Eq. (4.13) when $\alpha\geqslant 1$.

On the other hand, for $0<\alpha<1$, we shall use the fact that the energy eignfunctions (3.20b) lie in the domain of the para-Bose momentum operator $\hat{p}$ and thereby determine the proper domain M_s needed to specify this operator in the first Schrödinger description. The eigenfunctions (3.20b) look near $x=0$ like

$$\psi^{(1)}_{n,+}\sim x^{\alpha-1/2},\quad \psi^{(1)}_{n,-}=0,\quad n=2l,$$
$$\psi^{(1)}_{n,+}=0,\quad \psi^{(1)}_{n,-}\sim x^{\alpha+1/2},\quad n=2l+1,\quad l=0,1,2,\cdots. \tag{4.15}$$

These functions belong to M_s for $s=0$. Thus, the complete specification of the para-Bose momentum operator $\hat{p}$ in the first Schrödinger description of $\mathscr{D}_\alpha$, $0<\alpha<1$, runs as follows: It acts in the manner given in Eqs. (4.1) and (4.7) and its domain consists of all wave functions which, apart from differentiability and proper behavior at infinity, behave near $x=0$ as

$$\psi_+\sim ax^{\alpha-1/2}+bx^{\lambda},\quad \psi_-\sim d\ x^{\lambda},\quad \lambda>\tfrac{1}{2}. \tag{4.16}$$

One checks that the momentum eigenfunctions (3.20a) do obey Eq. (4.16).

In all the above arguments we have omitted logarithmic factors and assumed that the exponents λ are real. Actually all we need is $\mathrm{Re}\lambda>\frac{1}{2}$ everywhere.

In a strictly similar manner one can show that in the second Schrödinger description of $\mathscr{D}_\alpha$ the formal action of the para-Bose momentum operator $\hat{p}$ is given by

$$\begin{aligned}(\hat{p}\psi)(x)&=\left(-i\frac{d}{dx}+i\,\frac{\frac{1}{2}-\alpha}{x}\hat{P}_x\right)\psi(x)\\ &=-i\frac{d\psi}{dx}+i\,\frac{\frac{1}{2}-\alpha}{x}\psi(-x),\\ (\hat{p}\psi)_+(x)&=-ix^{\alpha-1/2}\frac{d}{dx}(x^{(1/2)-\alpha}\psi_-(x)),\\ (\hat{p}\psi)_-(x)&=-ix^{(1/2)-\alpha}\frac{d}{dx}(x^{\alpha-1/2}\psi_+(x)),\end{aligned} \tag{4.17}$$

Corresponding to the fact that $\hat{R}_2$ is now realized as $\hat{P}_x$, and its definition is completed by stating that its domain consists of all ψ which apart from differentiability and good behavior at infinity behave near $x=0$ like

$$\psi_+\sim bx^{\lambda},\ \psi_-\sim cx^{\alpha-1/2}+d\ x^{\lambda}, \tag{4.18}$$

This is the proper definition of the domain for all α. Notice that in the first Schrödinger description too, finally the domain of definition of $\hat{p}$ turned out to be the same in form for all α, i.e., Eq. (4.13) for $\alpha\geqslant 1$ and Eq. (4.16) for $0<\alpha<1$. One checks that the wave functions (3.29) indeed obey Eq. (4.18).

This analysis shows that the effort to arrive at the Schrödinger (or wave-mechanical) descriptions of the para-Bose oscillator by first formally solving the $\hat{q}$-$\hat{p}$ commutation relations (2.25) via Eqs. (4.1) and (4.17) and then setting up (coupled) differential equations for momentum and energy

eigenfunctions can yield misleading results. Indeed, Ohnuki and Kamefuchi formally solve the commutation relation (2.25) by

$$\hat{p} = -i\frac{d}{dx} + i\frac{c}{x}\hat{P}_x, \tag{4.19}$$

where c is a real constant, and state explicitly that the theory allows c to be any real number. Their paper does not make it clear that the range $-\frac{1}{2} < c < \infty$ is unitarily equivalent to the range $-\infty < c < \frac{1}{2}$. It is difficult to demonstrate this equivalence directly working with the formal expression (4.19) and no domains specified. Furthermore, while for $c \leqslant -\frac{1}{2}$ or $c \geqslant \frac{1}{2}$ they obtain a unique set of momentum eigenfunctions imposing square integrability conditions alone (and these are just the second and first Schrödinger descriptions, respectively, for $\alpha \geqslant 1$), for $-\frac{1}{2} < c < \frac{1}{2}$ they obtain two possible sets of momentum eigenfunctions. Also, to artificially separate them, a superselection rule is invoked. Our work shows that the proper specification of domains takes care of this problem and there is no need for superselection rules, as indeed there cannot be since always the wave machanical description uses the set of *all* L_2 functions over $-\infty < x < \infty$ as the representation space. Stated in the notation of this paper, the problem arising with the formal differential operator (4.4) is that in the range $0 < \alpha < 1$, *both* $\psi_k^{(1)}(x;\alpha)$ and $\psi_k^{(2)}(x;\alpha)$ are solutions of the eigenvalue equation with k as eigenvalue; so also is any linear combination. However, while $\psi_k^{(1)}(x;\alpha)$ obeys the restriction (4.16), $\psi_k^{(2)}(x;1-\alpha)$ does not. It is therefore the domain specification that leads, for each α and for each Schrödinger description, to an unambiguous set of momentum eigenfunctions, were we to obtain these by solving (coupled) differential equations rather than by the methods of Sec. 3.

APPENDIX A: ORTHOGONALITY AND OVERLAP OF POSITION AND MOMENTUM EIGENSTATES

In this Appendix we give the proofs of Eqs. (3.7), (3.14), and (3.30). We begin with Eq. (3.14). From Eqs. (3.8) and (3.13), we find that the inner product of a momentum eigenvector with a position eigenvector is given by

$$\langle x;\alpha|k;\alpha\rangle = \exp(i\eta(k) - i\xi(x))\exp[-\tfrac{1}{2}(x^2+k^2)]\,|xk|^{\alpha-1/2} \times \sum_{l=0}^{\infty}(-1)^l l!\left[\{L_l^{\alpha-1}(x^2)L_l^{\alpha-1}(k^2)/\Gamma(l+\alpha)\} + ixk\{L_l^{\alpha}(x^2)L_l^{\alpha}(k^2)/\Gamma(l+1+\alpha)\}\right]. \tag{A1}$$

To perform the l summations we make use of the relation [(Ref. 18, p. 1038, formula (8.976.1)]

$$\sum_{l=0}^{\infty} l!L_l^{\alpha}(x^2)L_l^{\alpha}(k^2)z^l/\Gamma(l+1+\alpha) = |xk|^{-\alpha}(z^{-\alpha/2}/(1-z))\exp[-z(x^2+k^2)/(1-z)]I_\alpha(2|xk|z^{1/2}/(1-z)),\quad |z|<1, \tag{A2}$$

where I_α is the modified Bessel function of order α. Here we wish to take the limit $z \to -1$. Using the power series developement for I_α [Ref. 18, p. 961, formula (8.445.1)], we get

$$z^{1/2}I_\alpha(2|xk|z^{1/2}/(1-z)) \underset{z\to -1}{\to} J_\alpha(|xk|). \tag{A3}$$

In this limit, Eq. (A2) then gives

$$\sum_{l=0}^{\infty}(-1)^l l!L_l^{\alpha}(x^2)L_l^{\alpha}(k^2)/\Gamma(l+1+\alpha) = \tfrac{1}{2}|xk|^{-\alpha}\exp(\tfrac{1}{2}(x^2+k^2))J_\alpha(|xk|). \tag{A4}$$

Using this result, and a similar one with $\alpha - 1$ in place of α, in Eq. (A1) we readily obtain Eq. (3.14) of the text.

Next we give the proofs of Eqs. (3.7) and (3.30). We first observe that showing Eq. (3.30) to be valid is the same as establishing the relation

$$\int_0^{\infty} dx\, x[J_{\alpha-1}(|k'|x)J_{\alpha-1}(|k|x) + \epsilon(k'k)J_\alpha(|k'|x)J_\alpha(|k|x) = 2\delta(k'-k)/|k|. \tag{A5}$$

This is because Eq. (3.20a) gives the momentum eigenfunctions $\psi_k^{(1)}(x;\alpha)$ explicitly as an even function of x puls an odd function. Now it turns out that the proof of Eq. (3.7) can also be reduced to that of Eq. (A5). Let us therefore show this and then give the proof of Eq. (A5). From Eq. (3.6) the inner product of two position eigenvectors is

$$\langle x';\alpha|x;\alpha\rangle = \Gamma(\alpha)C_0(x')^*C_0(x)\sum_{l=0}^{\infty} l!\left[L_l^{\alpha-1}(x'^2)L_l^{\alpha-1}(x^2)/\Gamma(l+\alpha) + x'xL_l^{\alpha}(x'^2)L_l^{\alpha}(x^2)/\Gamma(l+1+\alpha)\right]. \tag{A6}$$

If we use the result [Ref. 18, p.721, formula (6.643.4)]

$$l!L_l^{\alpha}(x^2) = |x|^{-\alpha}e^{x^2}\int_0^{\infty} dt\, t^{l+(1/2)\alpha}e^{-t}J_\alpha(2|x|t^{1/2}),\quad l+1+\alpha>0, \tag{A7}$$

then the righthand side of Eq. (A6), omitting the factors before the summation signs, reads

$$|x|^{1-\alpha}e^{x^2}\int_0^{\infty} dt\, t^{(\alpha-1)/2}e^{-t}\,J_{\alpha-1}(2|x|t^{1/2})\sum_{l=0}^{\infty} t^l L_l^{\alpha-1}(x'^2)/\Gamma(l+\alpha) + x'x|x|^{-\alpha}e^{x^2}\int_0^{\infty} dt\, t^{\alpha/2}e^{-t}\,J_\alpha(2|x|t^{1/2})\sum_{l=0}^{\infty} t^l L_l^{\alpha}(x'^2)/\Gamma(l+1+\alpha). \tag{A8}$$

The l sums can now be performed [Ref. 18, p. 1038, formula (8.975.3)] and Eq. (A8) becomes

$$|x|^{1-\alpha}e^{x^2}\int_0^\infty dt\,|x|^{1-\alpha}J_{\alpha-1}(2|x|t^{1/2})J_{\alpha-1}(2|x'|t^{1/2}) + x'x|x|^{-\alpha}e^{x^2}\int_0^\infty dt\,|x'|^{-\alpha}J_\alpha(2|x|t^{1/2})J_\alpha(2|x'|t^{1/2})$$
$$= |x'x|^{1-\alpha}e^{x^2}\int_0^\infty dt\,[J_{\alpha-1}(2|x|t^{1/2})J_{\alpha-1}(2|x'|t^{1/2}) + \epsilon(x'x)J_\alpha(2|x|t^{1/2})J_\alpha(2|x'|t^{1/2})]. \tag{A9}$$

By using this result on the right hand side of Eq. (A6) and changing from t to $2t^{1/2}$ as integration variable one readily sees that the proof of Eq. (3.7) is reduced to that of Eq. (A5).

However, the proof of Eq. (A5) is very simple. We use the classical theorem concerning Hankel transforms. For any real number $\nu > -1$, it is known that if $f(x)$ is a function defined on the positive real line and $g(y)$ is defined as

$$g(y) = \int_0^\infty dx\,xJ_\nu(xy)f(x), \quad y>0, \tag{A10}$$

then f can be recovered from g by

$$f(x) = \int_0^\infty dy\,yJ_\nu(xy)g(y). \tag{A11}$$

For positive y and y', with $\nu > -1$, we therefore have

$$\int_0^\infty dx\,xJ_\nu(xy)J_\nu(xy') = \delta(y-y')/y. \tag{A12}$$

Since $\alpha > 0$, this result can be used for each term on the left hand side of Eq. (A5); this expression then becomes

$$[1+\epsilon(k'k)]\delta(|k'|-|k|)/|k| \equiv 2\delta(k'-k)/|k|,$$

thus proving Eq. (A5), and hence Eqs. (3.7) and (3.30).

APPENDIX B: PROOF OF EQ. (4.2)

In this Appendix we show that the momentum eigenfunctions $\psi_k^{(1)}(x;\alpha)$ obey Eq. (4.2). Since this is not a local differential equation, it is covenient to follow Eq. (4.5) and define the "even" and "odd" parts of $\psi_k^{(1)}(x;\alpha)$

$$\psi_{k,\pm}^{(1)}(x;\alpha) = \tfrac{1}{2}[\psi_k^{(1)}(x;\alpha) \pm \psi_k^{(1)}(-x;\alpha)]$$
$$= \tfrac{1}{2}|k|^{1/2}x^{1/2}\begin{cases} J_{\alpha-1}(|k|x), \\ i\epsilon(k)J_\alpha(|k|x), \end{cases} \quad x>0. \tag{B1}$$

As noted in Sec. 4, it suffices to define and deal with these parts for $x>0$ alone. Validity of Eq. (4.2) is now equivalent to the validity of two coupled first order local differential equations:

$$\left[\frac{d}{dx} \pm \frac{\frac{1}{2}-\alpha}{x}\right]\psi_{k,\pm}^{(1)}(x;\alpha) = ik\psi_{k,\mp}^{(1)}(x;\alpha). \tag{B2}$$

This can be rewritten as

$$\frac{d}{dx}\{x^{(1/2)-\alpha}\psi_{k,+}^{(1)}(x;\alpha)\} = ikx^{(1/2)-\alpha}\psi_{k,-}^{(1)}(x;\alpha), \tag{B3}$$

$$\frac{d}{dx}\{x^{\alpha-(1/2)}\psi_{k,-}^{(1)}(x;\alpha)\} = ikx^{\alpha-1/2}\psi_{k,+}^{(1)}(x;\alpha), \tag{B4}$$

and in view of Eq. (B1) we have

$$\frac{d}{dx}(x^{1-\alpha}J_{\alpha-1}(|k|x)) = -|k|x^{1-\alpha}J_\alpha(|k|x), \tag{B5}$$

$$\frac{d}{dx}(x^\alpha J_\alpha(|k|x)) = -|k|x^\alpha J_{\alpha-1}(|k|x). \tag{B6}$$

However, these equations indeed hold [cf. standard recursion relations among Bessel functions, Ref. 18, pp. 967,968, formulas (8.472.1) and (8.472.2)]. Thus, Eq. (4.2) is established.

[1] E.P. Wigner Phys. Rev. **77**, 711 (1950).
[2] H.S. Green, Phys. Rev. **90**, 270 (1953).
[3] P.A.M.Dirac, *The Principles of Quantum Mechanics* (Oxford University, New York, 1958), 4th edition.
[4] J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).
[5] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[6] C.L. Mehta, J. Math. Phys. **18**, 404 (1977).
[7] E.C.G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).
[8] C.L. Mehta, Phys. Rev. Lett. **18**, 752 (1967).
[9] L.O'Raifeartaigh and C. Ryan, Proc. R. Irish Acad. Sect. A **62**, 93 (1963).
[10] T.F. Jordan, N.Mukunda, and S.V. Pepper, J. Math. Phy. **4**, 1089 (1963).
[11] J.K. Sharma, C.L. Mehta, and E.C.G. Sudarshan, J. Math. Phys. **19**, 2089 (1978).
[12] L.M. Yang, Phys. Rev. **84**, 788 (1957).
[13] Y. Ohnuki and S. Kamefuchi J. Math. Phys. **19**, 67 (1978).
[14] All operators are denoted by a circumflex.
[15] Position, momentum, energy, etc. are all measured in proper units so that $\hat{q}, \hat{p}, \hat{H}$, etc. are all dimensionless operators.
[16] V. Bargmann, Ann. Math. **48**, 568 (1947).
[17] H.J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945).
[18] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965) 4th edition.
[19] That Eq. (3.12) holds for the para-Bose case as well is seen from the relation

$$e^{i\lambda\hat{H}}\hat{q}\,e^{-i\lambda\hat{H}} = \hat{q}\cos\lambda + \hat{p}\sin\lambda,$$

which is a consequence of the commutation relations

$$[\hat{q},\hat{H}] = i\hat{p}, \quad [\hat{p},\hat{H}] = -i\hat{q}.$$

A new proof of absence of positive discrete spectrum of the Schrödinger operator

A. G. Ramm and B. A. Taylor
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109

(Received 11 December 1979; accepted for publication 1 February 1980)

A new method to prove the absence of positive discrete spectrum of the Schrödinger operator is given.

1. INTRODUCTION

Consider the equation

$$(\nabla^2 + \lambda^2 - p(x))u = 0, \quad x \in \mathbb{R}^n, \lambda^2 > 0. \tag{1}$$

Let us assume that

$$u \in L^2(\equiv L^2(\mathbb{R}^n)). \tag{2}$$

One of the most important steps in spectral analysis of the Schrödinger operator (1) is to prove that the only solution of (1) and (2) under some assumptions about $p(x)$ is $u \equiv 0$. The first result of this kind in potential scattering was obtained by T. Kato.[1] It was generalized by many authors. In A.G. Ramm[2] the boundary value problem was considered in domains with infinite boundaries. In M. Reed and B. Simon[3] the result is obtained under more general assumptions about $p(x)$. Then, in Saito,[4] and in P. Mishnaevsky,[5] the case of the operator-valued Sturm–Liouville equation was treated; some results are known in case ∇^2 is replaced by an operator $Lu = -\Sigma_{i,j=1}^{3}(\partial/\partial x_i)(a_{ij}(x)(\partial u/\partial x_j))$, with $a_{ij}(x) = \delta_{ij}(x)$, for $|x| > R$ and which satisfies some other assumptions (M. Shifrin[6]). It is not our purpose to mention all the papers which deal with the question under discussion. For us it is important that all known proofs of the above mentioned uniqueness theorem are rather complicated and technical. It is our purpose to outline a new approach to this question and to present a new proof of the known result. We shall not present the result in the most general form but try to explain a new idea in our approach. It should be mentioned that in Ref. 1 the absence of positive eigenvalues of the Schrödinger operator was proved under the assumption $|p(x)| \leqslant C(1 + |x|)^{-a}$, $a > 1$, whereas here we handle only the case $|p(x)| \leqslant Ce^{-\epsilon|x|}$.

2. THE SIMPLEST CASE

First let us consider the simplest case when $p(x)$ is a bounded continuous function with compact support. Suppose that $u \in L^2(\mathbb{R}^n)$ and satisfies Eq. (1). Taking the Fourier transform of both sides of the equation, we find

$$(\lambda^2 - \zeta^2)\hat{u}(\zeta) = (\widehat{pu})(\zeta), \quad \zeta \in \mathbb{R}^n, \tag{3}$$

where $\zeta^2 = \zeta_1^2 + \cdots + \zeta_n^2$ and the $\hat{}$ denotes the Fourier transform

$$\hat{u}(\zeta) = \left(\frac{1}{2\pi}\right)^n \int_{\mathbb{R}^n} e^{-i\zeta \cdot x} u(x)\, dx. \tag{4}$$

Since $pu \in L^2(\mathbb{R}^n)$ and has support in some compact set in $\mathbb{R}^n$, its Fourier transform $\widehat{pu}(\zeta)$ is an entire function of $\zeta \in \mathbb{C}^n$. Further, $\widehat{pu}$ is an entire function of exponential type and its restriction to each of the real subspaces of $\mathbb{C}^n$, $\mathrm{Im}\zeta = \text{constant}$, is in $L^2(\mathbb{R}^n)$. The class of all Fourier transforms of L^2 functions with compact support in $\mathbb{R}^n$ will be denoted $\hat{L}_0^2$. We then claim that Eq. (3) implies $\hat{u} \in \hat{L}_0^2$.

Lemma 1: *If* $u \in L^2(\mathbb{R}^n)$ *and satisfies* (1) *with* $\lambda^2 > 0$, *and if* $p(x)$ *is a bounded measureable function with compact support in* $\mathbb{R}^n$, *then* $\hat{u} \in \hat{L}_0^2$. *That is, u has compact support in* $\mathbb{R}^n$.

If we assume Lemma 1, then the conclusion $u \equiv 0$ follows from the unique continuation theorem for Schrödinger equation [3].

Proof of Lemma 1: We will check first that $\hat{u}$ is an entire function and, second, that $\hat{u} \in \hat{L}_0^2$. To see this note that $(\widehat{pu})(\zeta)/(\lambda^2 - \zeta^2) = \hat{u} \in L^2(\mathbb{R}^n)$. It follows that $(\widehat{pu})(\zeta) = 0$ on the sphere $S_\lambda = \{\zeta \in \mathbb{R}^n : \zeta^2 = \lambda^2\}$, since $1/(\lambda^2 - \zeta^2)$ is not square integrable over any neighborhood in $\mathbb{R}^n$ of a point in S_λ. Then, since $(\widehat{pu})(\zeta)$ vanishes on $\{\zeta \in \mathbb{R}^n : \zeta^2 = \lambda^2\}$, it follows by analytic continuation that $(\widehat{pu})(\zeta)$ vanishes on the analytic variety in $\mathbb{C}^n$, $V = \{\zeta \in C^n : \zeta^2 = \lambda^2\}$. But, the gradient of $\zeta^2 - \lambda^2$ does not vanish on V, so $(\widehat{pu})(\zeta)/(\zeta^2 - \lambda^2)$ is a smooth function and, therefore, analytic on $\mathbb{C}^n$. Thus, $\hat{u}(\zeta) = (\widehat{pu})(\zeta)/(\zeta^2 - \lambda^2)$ is an entire function.

To prove that $\hat{u} \in \hat{L}_0^2$, we will use a version of the Paley–Wiener theorem (Ref. 7, p. 20, 21). It is not hard to verify that each $\hat{v} \in \hat{L}_0^2$ satisfies for some $C, R > 0$,

$$\begin{aligned} &\text{(i)} \quad |\hat{v}(\zeta)| \leqslant C\|v\|_{L^2(\mathbb{R}^n)} \exp(R\,|\mathrm{Im}\zeta|), \\ &\text{(ii)} \quad \text{the restriction of } \hat{v} \text{ to } \mathbb{R}^n \text{ belongs to } L^2(\mathbb{R}^n). \end{aligned} \tag{5}$$

The Paley–Wiener theorem implies that the converse also holds.

Theorem 1: *If* $\hat{v}$ *is an entire function on* $\mathbb{C}^n$ *satisfying the growth conditions* (i) *and* (ii), *then* $\hat{v} \in \hat{L}_0^2$. *In fact,* (i) *can be replaced by the weaker growth condition,*

$$\text{(i')} \quad |\hat{v}(\zeta)| \leqslant C\exp(R\,|\zeta|).$$

Further, v has support in $\{|x| \leqslant R\}$.

To conclude the proof of Lemma 1, we will prove that $\hat{u}$ satisfies (i) and (ii), assuming that $\Phi = \widehat{pu}$ satisfies (i) and (ii). Since $u \in L^2(\mathbb{R}^n)$, we already know that (ii) holds. We also know that $(\lambda^2 - \zeta^2)\hat{u}(\zeta)$ satisfies an estimate of the form (i). It is then an immediate consequence of the "division lemma," (see Corollary 1.3, p. 8 of L. Ehrenpreis[8]), that $\hat{u}$ also satisfies an estimate of the form (i). This completes the proof.

Remark: The hypothesis that p is bounded can be considerably relaxed in Lemma 1. All that is needed is that

 0022-2488/80/092395-03$1.00

$\Phi = (\widehat{pu})$ is an entire function of exponential type. This will be true if pu is a distribution with compact support; for example, if $p \in L^{\epsilon}$ for some $\epsilon > 0$. However, to conclude the result of the theorem, $u \equiv 0$, we apply the unique continuation theorem which requires stronger hypotheses. See remark 1 of Sec. 5.

3. EXPONENTIALLY DECREASING POTENTIALS

In this section we show how the idea of Sec. 2 can be used to prove $u \equiv 0$ when the assumption on the potential $p(x)$ is relaxed to

$$|p(x)| \leqslant C \exp(-a|x|), \quad x \in \mathbb{R}^n \tag{6}$$

for some $C, a > 0$. The steps are the same except that an additional induction argument is required. First, recall that the main point was to prove $\hat{u} \in \hat{L}^2_0$, and the first step in this argument was to prove that $\hat{u}$ could be analytically continued from $\mathbb{R}^n$ to $\mathbb{C}^n$. To handle potentials of the form (6), we need a local version of this result. A more general version will be stated and proved, since it isolates the crucial property satisfied by the polynomial $\zeta^2 - \lambda^2$ and is therefore perhaps of independent interest.

Lemma 2: Let Ω be an open set in $\mathbb{C}^n$. Suppose Q is a polynomial on $\mathbb{C}^n$, Φ is analytic on Ω, f is locally square integrable on $\mathbb{R}^n \cap \Omega$, and

$$Qf = \Phi, \text{ on } \mathbb{R}^n \cap \Omega, \tag{7}$$

if Q satisfies the following condition:

Each irreducible component V of $\{\zeta \in \Omega : Q(\zeta) = 0\}$ intersects $\mathbb{R}^n \cap \Omega$ in a real analytic set of dimension $n-1$, (8)

then f has an analytic continuation to all of Ω.

We will postpone the proof of Lemma 2 to the next section. Let us remark, however, that Q need not be a polynomial—Q analytic on Ω is all that is needed. Further, the property (8) satisfied by Q is exactly what is needed to conclude that f is analytic, at least when Ω is a domain of holomorphy which is homeomorphic to a ball. That is, if Ω satisfies these conditions, if $\Omega \cap \mathbb{R}^n$ is not empty, and if each f satisfying the hypotheses of Lemma 2 is also analytic on Ω, then Q must also satisfy (8). We will not include a proof here, since it does not seem relevant to the problem.

The other ingredient needed for the proof is a version of the Paley–Wiener theorem for functions with exponential decay. However, we need to have fairly explicit estimates of the constants appearing in the theorem. The form stated below is adequate. We will not give the proof, since any standard proof of the Paley–Wiener theorem will give the result just by explicitly carrying through the estimates.

Lemma 3: Let $\epsilon > 0$, $a \geqslant 0$, and suppose that $e^{(a+2\epsilon)|x|} u(x) \in L^2(\mathbb{R}^n)$ has L^2 norm at most K. Then $\hat{u} \in L^2(\mathbb{R}^n)$, $\hat{u}$ is analytic in the strip $|\mathrm{Im}\,\zeta| < a + 2\epsilon$ and there is a constant $C = C(\epsilon, n)$ dependending only on ϵ and n, and independent of a, such that

$$|\hat{u}(\zeta)| \leqslant K \cdot C$$

for all ζ with $|\mathrm{Im}\,\zeta| < a + \epsilon$. In the converse direction, if $\hat{u} \in L^2(\mathbb{R}^n)$, $\hat{u}$ is analytic in the strip $|\mathrm{Im}\,\zeta| < a + 2\epsilon$ and satisfies $|\hat{u}(\zeta)| \leqslant K$ in that strip, then $e^{(a+\epsilon)|x|} u(x) \in L^2(\mathbb{R}^n)$ and has L^2 norm at most $K \cdot C_1$, where C_1 is a constant depending only on n and ϵ.

We can now prove the theorem.

Theorem 2: Suppose $p(x)$ is a locally bounded, measurable function such that for some $\epsilon > 0$,

$$|p(x)| \leqslant K e^{-4\epsilon|x|}.$$

Suppose further that $u \in L^2(\mathbb{R}^n)$ satisfies for some $\lambda > 0$,

$$\Delta u(x) + \lambda^2 u(x) - p(x)u(x) = 0, \quad x \in \mathbb{R}^n.$$

Then $u \equiv 0$.

Proof: Taking the Fourier transform of the equation yields $(\lambda^2 - \zeta^2)\hat{u}(\zeta) = (\widehat{pu})(\zeta)$. Because of the estimate on p, we have from Lemma 3, with $a = 0$ that (pu) is analytic in the strip $|\mathrm{Im}\,\zeta| < 4\epsilon$ and bounded by $KC\|u\|_{L^2}$ in the strip $|\mathrm{Im}\,\zeta| < 3\epsilon$. Because $\lambda^2 - \zeta^2$ satisfies (8), it follows that $\hat{u}$ is analytic in $|\mathrm{Im}\,\zeta| < 4\epsilon$. Further, from a standard division lemma (see Corollary 1.3, p. 8 of L. Ehrenpreis[8]) it follows that $\hat{u}(\zeta)$ is bounded by $C_1 K \|u\|_{L^2}$ in the strip $|\mathrm{Im}\,\zeta| < 2\epsilon$, where C_1 is a constant depending only on n and ϵ. Finally, from the converse part of Lemma 3, we deduce that $e^{\epsilon|x|}u(x) \in L^2(\mathbb{R}^n)$ and, further, has L^2-norm at most $C_2 K \|u\|_{L^2(\mathbb{R}^n)}$; where C_2 is a constant depending only on n and ϵ.

We can then repeat the argument, except using $a = \epsilon$ in Lemma 3, to deduce

$$e^{2\epsilon|x|}u(x) \in L^2(\mathbb{R}^n),$$

and has L^2 norm at most $(C_2)^2 K \|u\|_{L^2(\mathbb{R}^n)}$. Continuing in this fashion, we find that $e^{k\epsilon|x|}u(x) \in L^2(\mathbb{R}^n)$ and has L^2-norm at most equal to $(C_2)^k K \|u\|_{L^2(\mathbb{R}^n)}$, $k = 1,2,\ldots$. It therefore follows from Lemma 3 that

$$|\hat{u}(\zeta)| \leqslant K \|u\|_{L^2(\mathbb{R}^n)} \exp(C_3 |\mathrm{Im}\,\zeta|),$$

for some constant $C_3 > 0$. Thus, by the Paley-Wiener theorem, u must have compact support in the ball $|x| < C_3$. Therefore, the problem has been reduced to the case treated in Sec. 2, and we conclude that $u \equiv 0$.

4. PROOF OF LEMMA 2

We will prove that if Φ/Q is locally in L^2 (or even in $L^{1+\epsilon}$) on $\Omega \cap \mathbb{R}^n$ and if Q satisfies (8), then Φ/Q is analytic in Ω. By factoring Q into a product of powers of prime factors (in the ring of functions analytic on Ω—not in the polynomial ring), it is seen that it is enough to consider the case when Q is irreducible. In this case, Φ/Q is analytic on Ω if and only if the variety $\{\zeta \in \Omega : \Phi(\zeta) = 0\}$ contains $V = \{\zeta \in \Omega : Q(\zeta) = 0\}$. Now, since $\dim_{\mathbb{R}}(V \cap \Omega \cap \mathbb{R}^n) = n - 1$, we must have $\Phi = 0$ on $V \cap \Omega$. Otherwise, there exists $x_0 \in V \cap \Omega$ and $\epsilon > 0$ such that

$$\int_{|x - x_0| \leqslant \epsilon} \frac{dx}{|Q(x)|^2} < +\infty.$$

But, the integral $\int_{|x-x_0|\leqslant\epsilon} dx/|Q(x)|^2$ must diverge. To see this, choose linear coordinates $(t_1,\ldots,t_n)$ near x_0 so that $t = 0$ corresponds to x_0 and for each choice of $(t_2,\ldots,t_n)$, $t_1 \to Q(t_1,\ldots,t_n)$ has a zero near $t_1 = 0$. This is possible since $\dim_{\mathbb{R}}(V \cap \mathbb{R}^n \cap \Omega) = n - 1$. Then

$$\int \frac{dt}{|Q(t)|^2} = \int dt_2 \ldots dt_n \int \frac{dt_1}{|Q(t_1,t_2,\ldots t_n)|^2} = +\infty.$$

Consequently, $\Phi = 0$ on $V \cap \mathbb{R}^n \cap \Omega$.

Next, since V is irreducible, to prove $V \subset \{\Phi(\zeta) = 0\}$, all we have to show is that $\{\Phi(\zeta) = 0\} \cap U \supset V \cap U$, where U is a neighborhood of a point $z_0 \in V$. Thus, let $x_0 \in V \cap \mathbb{R}^n \cap \Omega$. Then the germ of the real analytic set $\{\Phi(\zeta) = 0\} \cap \mathbb{R}^n$ at x_0 contains the germ of $V \cap \mathbb{R}^n$ at x_0, which is of dimension $n-1$. It follows that the complex germ of the real analytic set $\{\Phi(\zeta) = 0\} \cap \mathbb{R}^n$ at x_0 contains the complex germ of $Q = 0$ at x_0 (a consequence of Proposition 2, p. 92 of R. Narasimhan,[9] which is also a good reference for the other facts about analytic sets which have been used in the proof). This completes the proof.

5. REMARKS

1. The method given in this paper can be used in many cases when the potential $p(x)$ has singular points. Our arguments require only that pu be a distribution of exponential rate of decrease and of finite order. For example, this will be the case if for some ϵ, $a > 0$, $p(x)$ is measurable and $\int_{|x|>t} |p(x)|^{\epsilon}\, dx \leqslant e^{-a|t|}$. However, the proof also uses the unique continuation theorem for the Schrödinger operator in $\mathbf{R}^n$, which is only known for $p(x) \in L^q_{\mathrm{loc}}(\mathbb{R}^3)$ for $q \geqslant 3/2$, and in $\mathbf{R}^n$ for $p(x) \in L^q$, $q = \frac{2}{3}(n-1)$, $n \geqslant 5$; $q = 2$ for $n = 4$; $q > 1$ for $n = 1,2$. (See Refs. 3, 7, and 10).

2. It would be interesting to extend the ideas of this paper in order to prove absence of the positive eigenvalues of Schrödinger operator $|p(x)| \leqslant c(1+|x|)^{-a}$, $a > 1$. The Wigner–von Neumann example shows that there exists a potential $|p(x)| \leqslant c(1+|x|)^{-1}$, which has a positive eigenvalue [Ref. 3, p. 223]. It would be interesting to understand from the standpoint of our technique why $a = 1$ plays the role of frontier.

APPENDIX

We give an example which shows the importance of condition (8). If $f \geqslant 0, f \in C_0^{\infty}(\mathbb{R}^3)$, $k > 0$ and $u(x) = \int \exp(-k|x-y|)(4\pi|x-y|)^{-1} f(y)\, dy$, then $u(x) > 0$, $u \in L^2(\mathbb{R}^3)$, $-\Delta u + k^2 u = f(x)$, $\Delta^2 u + q(x)u - k^4 u = 0, q(x) \equiv (k^2 f + \Delta f) u^{-1}(x)$. Thus $k^4 > 0$ is the eigenvalue of $\Delta^2 + q(x)$, $q(x) \in C_0^{\infty}$. In this case $Q(\zeta) = \zeta^4 - k^4 = (\zeta^2 - k^2)(\zeta^2 + k^2)$ and $V = V_1 \cup V_2$, $V_1 = \{\zeta : \zeta \in \mathbb{C}^3, \zeta_1^2 + \zeta_2^2 + \zeta_3^2 - k^2 = 0\}$. Hence V_2 does not intersect $\mathbb{R}^3$, condition (8) is not fulfilled and hte operator $\Delta^2 + q(x)$ has a positive eigenvalue. The potential $q(x)$ was used also in the review article, D. Eidus, "The principal of limit amplitude," Russ. Math. Surv. **24**, N3, 97 (1969).

[1] T. Kato, Commun. Pure. Appl. Math. **12**, 403 (1959).
[2] A.G. Ramm, Sov. Math. Dokl. **4**, 1303 (1963); Math USSR, Sbornik **66**, 321 (1965); **72**, 638 (1967).
[3] M. Reed, B. Simon, *Analysis of Operators* (Academic, New York, 1978).
[4] Y. Saito, Publ. Res. Inst. Math. Sci. **7**, 581 (1971/1972).
[5] P.A. Mishnaevsky, Math. USSR, Izvestija. **10**, 145 (1978).
[6] M.A. Šifrin, Math. USSR, Sbornik **89**, 616 (1972).
[7] L. Hörmander, *Linear Partial Differential Operators* (Springer-Verlag, New York, 1964).
[8] L. Ehrenpreis, *Fourier Analysis in Several Complex Variables* (Wiley, New York, 1970).
[9] R. Narasimhan, *Introduction to the theory of analytic spaces*, Lecture notes in Math., Vol. 25 (Springer-Verlag, New York, 1966).
[10] M. Schecter and B. Simon, "Unique continuation for Schrödinger operators with unbounded potentials," to appear in J. Math. Anal. Appl.

Upper and lower bounds in nonrelativistic scattering theory

J. W. Darewych [a)]
Fachbereich Physik, Universität Kaiserslautern, 6750 Kaiserslautern, W. Germany

R. Pooran
Physics Department, York University, Toronto, Ontario M3J 1P3 Canada

(Received 22 January 1980; accepted for publication 16 May 1980)

We consider the problem of determining rigorous upper and lower bounds to the difference between the exact and approximate scattering phase shift, for the case of central potential scattering. The present work is based on the Kato identities and the phase-amplitude formalism of potential scattering developed by Calogero. For nonstationary approximations, a new first-order (in small quantities) bound is established which is particularly useful for partial waves other than s waves. Similar, but second-order, bounds are established for approximations which are stationary. Some previous results, based on the use of the Lippman–Schwinger equation are generalized, and some new bounds are established. These are illustrated, and compared to previous results, by a simple example. We discuss the advantages and disadvantages of the present results in comparison to those derived previously. Finally, we present the generalization of some of the present formalism to the case of many-channel scattering involving many-particle systems, and discuss some of the difficulties of their practical implementation.

I. INTRODUCTION

We are concerned, in this paper, with the simultaneous determination of upper and lower bounds to scattering parameters, assuming that their exact determination is impossible and that only an approximate calculation can be done. Of course, for central potential scattering the numerical determination of scattering phaseshifts to essentially arbitrary accuracy is a straightforward matter. Thus the interest in this problem lies in the possibility that it may point the way to generalizations for the case of scattering by many-particle targets, where the unambiguous and accurate determination of scattering parameters is far from trivial.

For scattering by a central potential, the exact phase shifts are determined from the solution of

$$Lu = u''(r) + [k^2 - U(r) - l(l+1)/r^2]u(r) = 0\,, \tag{1}$$

subject to the condition that

$$u(r=0) = 0\,, \tag{2}$$

and having the asymptotic form

$$u(r\to\infty) = A\sin(kr - l\pi/2 + \eta)\,. \tag{3}$$

where A is an arbitrary factor usually taken to be $\sec\eta$, $\csc\eta$ or 1. The choice of A is immaterial for exact solutions of (1). This is not the case, however, for approximate solutions, in which case an appropriate choice of A may be crucial in avoiding some spurious, nonphysical singularities in the approximate phaseshifts. For a discussion of this point see Refs. 1 and 2. We shall assume that the potential, $U(r) = (2m/\hbar^2)V(r)$, is such that $r^2U(r)$ vanishes both at the origin and at infinity.

In two previous papers[3,4] we discussed bounds to the exact phaseshift, based on the so-called Kato identities[5] (these papers also contain references to earlier work on this subject), viz.

$$kAA_T\sin(\eta - \eta_T) = \int_0^\infty uLu_T\,dr = R_1 \tag{4}$$

and

$$kAA_T\sin(\eta - \eta_T) - \int_0^\infty u_T Lu_T\,dr = -\int_0^\infty (u_T - u)Lu_T\,dr = R_2\,, \tag{5}$$

where u_T is any approximation to the exact solution of (1) which has properties like (2) and (3), namely

$$u_T(r=0) = 0 \quad\text{and}\quad u_T(r\to\infty) = A_T\sin(kr - l\pi/2 + \eta_T)\,.$$

Thus, if the potential is sufficiently weak so that

$$g_M = \text{maximum of}\int_0^\infty |G_0(r,r')U(r')|\,dr' < 1\,,$$

where $G_0(r,r')$ is the free-particle Green's function, then

$$|R_1| \leqslant \frac{A(0)|s_l(kr)|_M \mathscr{L}_1}{1 - g_M} = b_1\,, \tag{6}$$

where $|s_l(kr)|_M$ is the maximum value of $|krj_l(kr)|$, j_l being the usual spherical Bessel function.

We note that

$$\mathscr{L}_1 = \int_0^\infty |Lu_T|\,dr \leqslant \mathscr{L}_2 = \left[\int_0^\infty W^{-2}\,dr \int_0^\infty (WLu_T)^2\,dr\right]^{1/2}, \tag{7}$$

W being any suitable weight function. For the particular case where

$$K^2(r) = k^2 - U(r) - l(l+1)/r^2 > 0 \quad\text{for all } r\,, \tag{8}$$

we have shown[4] that a different bound applies, namely

[a)]Permanent address: Physics Department, York University, Toronto, Ontario M3J 1P3, Canada.

 0022-2488/80/092398-08$1.00

$$|R_1| \leqslant A \frac{k^2}{K^2(r_1)} \frac{K^2(r_2)}{K^2(r_3)} \cdots \frac{K^2(r_{2n})}{K^2(r_{2n+1})} \mathscr{L}_1 = b_2 , \tag{9}$$

where $r_1, r_3, \cdots, r_{2n+1}$ are local minima and $r_2, r_4, \cdots, r_{2n}$ are local maxima of $K^2(r)$, counted from infinity towards the origin ($r_1 > r_2 > r_3 \cdots$).

Both these first-order (in the "small" quantity $\mathscr{L}_1$) bounds have drawbacks: Thus (6) is evidently inapplicable to strong potentials (particularly at low energies) since g_M becomes > 1. Also the factor g_M in inconvenient to calculate. (9) on the other hand, in view of the restriction (8), is applicable in essence only to s wave scattering.

In Sec. 2 of this paper we derive yet another bound on R_1 which circumvents these drawbacks for many situations. We also present illustrative examples.

In the case of the second-order (in the "small" quantity $u_T - u$) remainder R_2 we have shown[3] that it is bounded by

$$|R_2| \leqslant \frac{|f|_M}{1 - g_M} \mathscr{L}_1 = B_0 , \tag{10}$$

where $f(r) = \int_0^\infty G_0(r,r') L u_T(r')\, dr'$, and the subscript M refers to the maximum value of quantity in question. This second-order bound, B_0, suffers from the same drawbacks as the first-order bound b_1. In Sec. 3 we discuss an alternate form of the second-order bound which, again, circumvents these difficulties for many situations. We also present generalizations of bounds to R_2 that were derived previously[3] using the Lippman–Schwinger equation, and obtain some new ones using a related approach.

In Sec. 4 we consider the generalization of the Kato identities to the case of multichannel scattering by many-body systems. A number of bounds developed previously for potential scattering are formally generalized to the many-body case, and we discuss the difficulties connected with their practical implementation. Finally, concluding remarks are given in Sec. 5.

2. USE OF THE PHASE-AMPLITUDE FORMALISM TO DETERMINE A FIRST ORDER BOUND

Calogero[6] has discussed in detail an alternate, though completely equivalent, formulation of the problem defined by Eq. (1). We shall term it the phase-amplitude formalism since u is written as

$$u(r) = \alpha(r) D_l(kr) \sin[\delta_l(kr) + \eta(r)] , \tag{11}$$

and (1) is replaced by the two first-order equations

$$\eta'(r) = -\frac{1}{k} U(r) D_l^2(kr) \sin 2[\delta_l(kr) + \eta(r)] \tag{12}$$

and

$$\alpha'(r) = -\frac{1}{2k} \alpha(r) U(r) D_l^2(kr) \sin 2[\delta_l(kr) + \eta(r)] . \tag{13}$$

$D_l(x)$ and $\delta_l(x)$ are the amplitude and phase functions for the modified spherical Bessel and Neumann (or Riccati–Bessel) functions.[6] $s_l(x) = xj(x)$ and $c_l(x) = -xn(x)$, since

$$s_l(x) = D_l(x) \sin\delta_l(x) \quad \text{and} \quad c_l(x) = D_l(x) \cos\delta_l(x) . \tag{14}$$

These functions, as well as the entire formalism, are discussed in detail in Calogero's book,[6] hence we shall not dwell on them here except to note that $D_l(x)$ is a positive, decreasing function of x for $l > 0$. (Indeed $D_0 = 1$, $D_1^2 = 1 + 1/x^2$, $D_2^2 = 1 + 3/x^2 + 9/x^4$ etc.) We are using a somewhat different notation from that of Calogero. The correspondence is $\alpha = \alpha_l$, $\eta = \delta_l$, $D_l = \hat{D}_l$, $\delta_l = \hat{\delta}_l$, $s_l = \hat{j}_l$, $c_l = -\hat{n}_l$, where Calogero's notation is given on the right in each case.

Thus $\alpha(r) D_l(kr)$ and $\eta(r)$ are, correspondingly, the amplitude and phase functions for any solution $u(r)$ of (1). If $\eta(r=0) = 0$ (so that $u(r=0) = 0$) then the solution, $\eta(r)$, of (12) approaches the physical phase shift η as $r \to \infty$.

Equation (13) implies that the amplitude function is

$$\alpha(r) = \alpha(\infty) \exp\left\{ -\frac{1}{2k} \int_r^\infty U(r') D_l^2(kr') \times \sin 2[\delta_l(kr') + \eta(r')]\, dr' \right\} , \tag{15}$$

where $\alpha(\infty) = A$. Also, any solution of (1) satisfies the inequality

$$|u(r)| \leqslant |\alpha(r)| D_l(kr) . \tag{16}$$

Now $\alpha(r)$ is unknown unless $\eta(r)$ is determined, which is equivalent to obtaining an exact solution of the problem. However, it follows from (15) that

$$\alpha(r) \leqslant |A|\, Z_l(r) = |A| \exp\left\{ \frac{1}{2k} \int_r^\infty |U(r')| D_l^2(kr')\, dr' \right\} . \tag{17}$$

$Z_l(r)$ is an easily calculable function, which decreases monotonically with r to its asymptotic value of unity. For example, if $U = -2e^{-2r}$ a.u.($\hbar = m = e = 1$), which is the example considered in earlier work,[3,4,7] we obtain $Z_0(r) = \exp[(1/2k) e^{-2r}]$. Thus, $Z_l(r)$ can be used to bound $|u(r)|$ and hence $|R_1|$. However, for $l \neq 0$, $Z_l(r)$ is usually highly singular at the origin, so that it cannot be used directly in the expression for R_1. We recall, however, that $|u(r)|$ increased from $|u(r=0)| = 0$ until at least the point r_0, which is the first zero of $K^2(r)$. (See, for example, the discussion in Mott and Massey,[8] p. 25). Thus

$$|u(r)| \leqslant |u(r_0)| \leqslant |A|\, Y_l(r_0), \quad \text{if } 0 \leqslant r \leqslant r_0 , \tag{18}$$

where

$$Y_l(r) = Z_l(r) D_l(kr) . \tag{19}$$

Defining

$$\bar{Y}(r) = Y_l(r_0) \quad \text{if } r \leqslant r_0$$
$$= Y_l(r) \quad \text{if } r > r_0 \tag{20}$$

we have $|u(r)| \leqslant |A|\, \bar{Y}_l(r)$, so that

$$|R_1| \leqslant |A| \int_0^\infty \bar{Y}_l(r) |L u_T(r)|\, dr = b_3 . \tag{21}$$

In comparing this new bound b_3, to the previous ones (6) and (9), we note that it has the advantage that no restriction on the strength of the potential (like $g_M < 1$) are necessary, and, unlike b_2, it can be used irrespectively of the sign of $K^2(r)$ (i.e. $l > 0$). Nevertheless for those cases where b_1 and/or b_2 are applicable, b_3 may not give the best value. Thus, for the exponential potential example already quoted, if $l = 0$ (whence $r_0 = 0$ would be taken), $b_2 = |A|\, \mathscr{L}_1 < b_3$ since

$\bar{Y}_l(r) \geqslant 1$ for all r. We can estimate the difference between b_2 and b_3 easily, since $b_3 < |A|\,\bar{Y}_0(0)\mathscr{L}_1$, so that $b_3/b_2 \lesssim e^{1/2k}$ for the example being considered. This ratio is thus not very large, except at low energies in which case b_2 has a substantial advantage over b_3. (The low energy case ($k \to 0$) must be treated differently, because the approximation $|\sin 2[\delta_l(kr) + \eta(r)]| = 1$ which has been made in deriving (17) is too crude. This, however, will not be discussed here.) However, if $l \neq 0$, b_2 is inapplicable, nor is b_1 if the potential is strong, whereas b_3 is easily calculated. For example if $U = -U_0 e^{-2r}$ then since

$$Y_1(r) = D_1(kr)\exp\left[\frac{U_0}{2k}\int_r^\infty e^{-2r'} D_1^2(kr')\,dr'\right] < \left[1 + \frac{1}{(kr)^2}\right]^{1/2} \exp\left[\frac{U_0}{4k}\left(1 + \frac{1}{k^2 r^2}\right)e^{-2r}\right],$$

$\bar{Y}_1(r)$ decreases with r extremely rapidly to unity from the maximum value $Y_1(r_0)$. This maximum value is itself not much larger than unity. Thus if U_0 is not too large $k^2 - U(r_0) - 2/r_0^2 = 0$ has the solution $r_0 \simeq \sqrt{2}/k$, so that $Y_1(r_0 = \sqrt{2}/k) \simeq \sqrt{\tfrac{1}{2}}\exp[(3U_0/8k)e^{-2\sqrt{2}/k}]$ which is largest at $k = 2\sqrt{2}$ when it takes on the value $Y_1 \simeq \sqrt{\tfrac{1}{2}}\exp(3U_0/e^{-2\sqrt{2}/k}/8k)$. This is about 1.35 if $U_0 = 2$ and 1.56 if $U_0 = 5$. The situation only improves for larger l values since r_0 increases with l. In short for $l > 0$ the first-order bound b_3 will be very nearly equal to $|A|\,\mathscr{L}_1$ for all k, and $\mathscr{L}_1$ can be made small by appropriate choice of the trial function u_T as we have demonstrated previously.[3,4] Thus we now have a calculable first-order bound suitable for any l and any strength of the potential.

3. SECOND-ORDER BOUNDS

If $w = u_T - u$, then w must satisfy the inhomogeneous equation

$$Lw = Lu_T \tag{22}$$

for which the solution can be formally written (see, for example Mott and Massey,[8] Ch. IV, p. 70).

$$w(r) = \int_0^\infty G(r,r')Lu_T(r')\,dr', \tag{23}$$

where

$$G(r,r') = -\frac{c}{k}u(r_<)\bar{u}(r_>). \tag{24}$$

$u,\bar{u}$ are linearly independent solutions of $Lw = 0$ which are regular and irregular at the origin (i.e. $u(r=0) = 0$ and $\bar{u}(r \to 0) \neq 0$, indeed usually $\bar{u}$ is singular at $r = 0$). $r_<$ and $r_>$ are, respectively, the smaller and larger of r,r' while $k/c =$ Wronskian of $\{u,\bar{u}\} = kA\bar{A}$, $A, \bar{A}$ being the asymptotic amplitudes of $u,\bar{u}$ respectively (these may be chosen arbitrarily, in particular, they may be chosen to be unity). If the potential U can be divided up into a "solvable" part U_1 and the remainder, i.e., $U = U_1 + U_2$, then rewriting (22) formally as $(K - U_1)w = U_2 w + Lu_T$, where $L = K - U$, it follows that

$$w(r) = \int G_1(r,r')U_2(r')w(r')\,dr' + \int G_1(r,r')Lu_T(r')\,dr', \tag{25}$$

where G_1 is the Green's function (24) but constructed from the (known) solutions of $(K - U_1)w = 0$.

It follows, from (25) that

$$|w(r)|_M \leqslant \max\text{ of}\left|\int G_1(r,r')Lu_T(r')\,dr'\right| \Big/ \left[1 - \max\text{ of}\int |G_1U_2|\,dr\right], \tag{26}$$

which is a generalization of the bound B_0 (Eq. (10)), and indeed reduces to it if $U_1 = 0$ and $U_2 = U$. It is useful in cases where

$$g_M = \max\text{ of}\int |G_0U|\,dr \geqslant 1$$

but

$$g_{1M} = \max\text{ of}\int |G_1U_2|\,dr < 1.$$

The Lippman–Schwinger equation can be used to generate yet another class of second-order bounds, somewhat similar to (10) (or its generalization (26)). This is because

$$|R_2|^2 = \left|\int wLu_T\,dr\right|^2 \leqslant \int |w|^2\rho\,dr \times \int \rho^{-1}||Lu_T|^2\,dr \tag{27}$$

by Schwartz's inequality, where ρ is some suitable, positive weight function (and such that all the indicated integrals exist). The integral $\int |w|^2\rho\,dr$ can, in turn, be bounded by successive applications of Schwartz's inequality to the integral equation that it satisfies, like in the approach used recently by Martin,[9] in his paper on the boundedness of the total cross section in potential scattering.

For the sake of simplicity we first consider a particular case where, if

$$Lu_T = (U_T - U)u_T, \tag{28}$$

the weight function ρ is taken to be $|U_T - U|$. Note that (28) can be viewed either as a definition of U_T or, alternatively, u_T can be thought of as the exact solution of the model problem (1) where U has been replaced by a solvable potential U_T. Then if G_T is the Green's function constructed from the known solutions of $(K - U_T)v = 0$, it follows from (25) (with $G_1 = G_T$, $U_2 = U_T - U$, $Lu_T = (U_T - U)u_T$) that

$$\int |w|^2|U_T - U| = \int\int w^*|U_T - U|G_T(\acute{U}_T - \acute{U})\acute{u}_T + \int\int w^*|U_T - U|G_T(\acute{U}_T - \acute{U})\acute{w}, \tag{29}$$

where, in order to simplify the notation, we have suppressed the differentials dr, dr' and have written $\acute{w}$ for $w(r')$, etc. Following Martin,[9] we apply Schwartz's inequality to each of the terms on the right, and obtain

$$\int |w|^2|U_T - U|\,dr \leqslant \frac{g_T{}^2}{(1-g_T)^2}\int |u_T|^2|U_T - U|\,dr \quad (30)$$

provided

$$g_T{}^2 = \int\int |U_T - U||G_T|^2|\acute{U}_T - \acute{U}| < 1\,. \quad (31)$$

Thus it follows from Eq. (27) that

$$|R_2| \leqslant \frac{g_T}{1-g_T}\int |u_T|^2|U_T - U|\,dr = k\bar{B}_4\,. \quad (32)$$

For example, if $u_T = s_l(kr)$, the solution with $U_T \equiv 0$ (so that $G_T = G_0$, the usual free particle Green's function), (32) becomes

$$|\tan\eta - \tan\eta_B| \leqslant \frac{g_B}{1-g_B}\tan\bar{\eta}_B = \bar{B}_4\,, \quad (33)$$

where we have used the usual normalization, $A = \sec\eta$, $A_T = 1$ and where we have written $k\tan\eta_B$ for $\int u_T L u_T\,dr = -\int s_l^2(kr)U(r)\,dr$, $k\tan\bar{\eta}_B$ for $\int s_l^2(kr)|U(r)|\,dr$, and assumed that $g_B^2 = \int\int |U||G_0|^2|\acute{U}| < 1$. Note that $|\tan\eta_B| \leqslant \tan\bar{\eta}_B$. Equation (33) implies that the Born approximation to a particular phase shift will approach the exact value as k and/or l become large. Indeed, $g_B/(1-g_B)$ is a generally decreasing function of k and l, and depends on the potential strength. Though readily calculable, $g_B/(1-g_B)$ cannot in general be evaluated in closed form, but simple, explicit (though crude) bounds to it are readily written down, which make the above observation obvious: Thus $g_B < n_l v_0/k$, where $v_0 = \int|U|\,dr$ is the potential strength, and $n_l = \max|s_l(kr_<)c_l(kr_>)|$ is a number close to unity. (34) then reduces to $|\tan\eta - \tan\eta_B|/\tan\bar{\eta}_B < nv_0/(k - nv_0)$. Alternatively, Martin[9] has pointed out that $g_B^2 < I/(2l+1)$, where

$$I = \frac{1}{(4\pi)^2}\int\int \frac{|U(r)||U(r')|}{|\mathbf{r}-\mathbf{r}'|}\,d^3r d^3r'\,,$$

which demonstrates explicitly the improvement of the Born approximation with increasing l (for potentials where I exists and is $< 2l+1$). More particularly, for the square well example considered previously,[3] we obtain for the s-wave ($l = 0$):

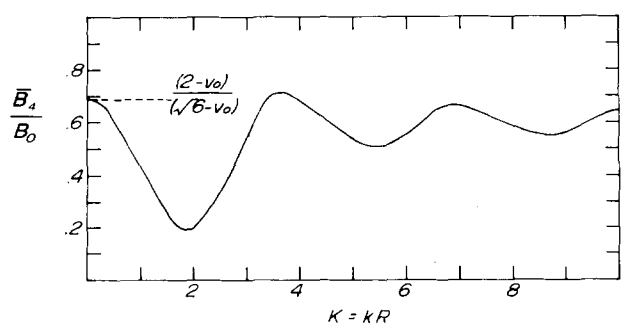


FIG. 1. Comparison of the bounds $\bar{B}_4$ and B_0 (see text) for a square well potential of strength $v_0 = 2mR^2V_0/\hbar^2 = 1$ and $l = 0$. B_0 is here the expression $f_B\mathscr{S}_l/k(1-g_M)$ given in Table 1 of Ref. 1, whereas $\bar{B}_4$ is given in (33) and (34), and $\tan\eta_B = \tan\bar{\eta}_B = (v_0/2K)[1 - (\sin 2K)/2K]$ for this example.

$$g_B{}^2 = \frac{v_0^2}{4K^2}\left[1 + \frac{\sin 2K}{K} + \frac{\cos 2K - 1}{K^2} - \frac{\sin^2 2K}{4K^2}\right]. \quad (34)$$

Comparing the bound $\bar{B}_4$, equation (33) to B_0 given in (10) (see Ref. 3) for this example we see that the results are comparable though $\bar{B}_4$ is somewhat better at all energies. The ratio $\bar{B}_4/B_0$ is plotted in Fig. 1 for a well of unit strength ($v_0 = 1$).

Of course analogous results may be obtained by using an arbitrary positive weight function ρ (but such that all indicated integrals exist), since (25) implies that

$$\int |w|^2\rho = \int\int w^*\rho G_1\acute{U}_2\acute{w} + \int\int w^*\rho G_1 L\acute{u}_T\,. \quad (35)$$

Performing similar manipulations to those preceeding Eq. (30) we find that

$$|R_2|^2 \leqslant kB_4 = \left(g_{\rho T} + \frac{g_{\rho 2}g_{2T}}{1-g_{22}}\right)^2\int |Lu_T|\,dr \times \int \rho^{-1}|Lu_T|^2\,dr\,, \quad (36)$$

where

$$g^2{}_{\rho T} = \int\int \rho|G_1|^2|L\acute{u}_T|\,,$$

$$g^2{}_{\rho 2} = \int\int \rho|G_1|^2|\acute{U}_2|\,,$$

$$g^2{}_{2T} = \int\int |U_2||G_1|^2|L\acute{u}_T|\,,$$

and

$$g^2{}_{22} = \int\int |U_2||G_1|^2|\acute{U}_2|\,, \quad \text{provided } g_{22} < 1.$$

Clearly variations on this theme are possible.

The phase-amplitude formalism can also be used to establish a new second-order bound, somewhat analogously to that which was done for the first order bound in Sec. 2. Thus, substituting (23) into the expression (5) for R_2 we obtain

$$R_2 = -\int\int G(r,r')Lu_T(r)Lu_T(r')\,drdr'\,. \quad (37)$$

Since $G(r,r')$ is constructed from the exact solutions $u(r)$ and $\bar{u}(r)$ [see Eq. (24)], each of which is bounded, in absolute value, by $|A|Y_l(r)$ [Eq. (19)], it is clear that a bound to $|G|$, hence to $|R_2|$ can be established. There is one difficulty, however: as has already been pointed out $Y_l(r)$ is, in general, highly singular at the origin, hence some means must be found of avoiding this region. This can be done because both $w(r) = u_T(r) - u(r)$ and $G(r,r')$ vanish at the origin, hence there is no purpose in using a singular function to bound them near the origin. Indeed, since u and u_T both vanish at the origin, $|w| = |u_T - u|$ will, of necessity, initially increase until it reaches a (locally) maximum value at some point $r > 0$. Also, for the type of potentials considered here, $u \propto r^{l+1}$ and $\bar{u} \propto r^{-l}$ for small r, so that $G \propto (r_</r_>)^l r_<$ which is an increasing function of r,r' for small r,r' (indeed it increases most quickly along $r = r'$). Thus there exists an $r_1' > 0$ such that $|G(r,r')| \leqslant |G(r_1',r_1')|$ for $r,r' < r_1'$. In either case, defining

$$\bar{\bar{Y}}(r) = Y_l(r_1) \quad \text{if } r \leqslant r_1$$
$$= Y_l(r) \quad \text{if } r > r_1$$

it then follows from (27) that

$$|R_2| \leqslant \frac{1}{k}\left\{\int_0^\infty |\bar{\bar{Y}}(r) L u_T(r)|\, dr\right\}^2 = B_3 . \tag{38}$$

Comparing this to the first-order bound (21) we see that $B_3 \sim b_3^2/k$. The bound B_3, unlike B_1 (Eq. (26)), has no restriction like max of $\int |G_1 U_2|\, dr < 1$. It requires, however, the choice of r_1. This number must lie within the region where $|u - u_T|$ or $|G(r,r')$ (Eq. 24) is an increasing function. Ideally, one would like to choose r_1 as large as possible since $Y_l(r)$ increases monotonically as $r \to 0$. We have not found, to date, a simple way of choosing a maximum r_1 on the basis of the properties of $U(r)$ only. Thus at the moment the choice of r_1 would have to be made on the basis of an approximate (power series) representation of $w(r)$ or $G(r,r')$ valid near the origin. This is clearly an inconvenient and inefficient way of choosing r_1 and the resulting bound would be unnecessarily large. An alternative way of overcoming the problem that $Y_l(r)$ (in general) singular at the origin, might be to choose u_T in such a way that $u_T \equiv 0$ for $r \leqslant$ some $r_1 > 0$, (but such that u_T', u_T'' exist and are continuous everywhere and, of course, u_T has the proper asymptotic form). The lower limits on the integrals in (37) and (38) are then not zero but $r_1 > 0$ and the difficulty associated with the possible singularity of $Y_l(r)$ at the origin would be avoided.

Lastly, we might remark that though we have considered individual partial waves thus far, there is no difficulty in writing down generalizations of most of the results to the case where the entire scattering amplitude and wave function are considered (see Sec. 4 below). Such an approach might be particularly useful for the case of scattering by nonspherical potentials (such as occur in the scattering by molecules), since in that case the determination of "exact" results by numerical integration is not at all trivial.

4. GENERALIZATION TO THE CASE OF MULTICHANNEL SCATTERING INVOLVING MANY-BODY TARGETS

We have previously stressed that the principal purpose of studying the potential scattering case is that such study may suggest generalizations which are applicable to the multichannel scattering processes that involve many-body targets. Indeed many of the results developed above for potential scattering generalize formally in a straightforward way, though their rigorous implementation is quite another matter.

To achieve the generalizations we first consider the extension of the Kato identities to the many-body problem. We consider any collision problem describable by the Schrödinger equation

$$\acute{H}\Psi_i = 0 , \tag{39}$$

where $\acute{H} = H - E$, H being the exact Hamiltonian of the system and E the total energy, which is assumed to be known. The subscript i stands for the collection of indices needed to specify the possible degenerate solutions at energy E. If A_{ij} is the exact scattering amplitude, obtained from the exact solutions Ψ_i, and A_{ij}^T is an approximation obtained from some approximate or trial solutions Ψ_i^T, then the following identity is obtained by using Green's theorem:

$$A_{ji} = A_{ji}^T - \frac{1}{4\pi}\frac{2m_j}{\hbar^2}\int \tilde{\Psi}_j^* \acute{H} \Psi_i^T\, d\tau , \tag{40}$$

where m_j is the reduced mass of the system in the jth channel, $d\tau$ is an element of the configuration space of the colliding system and Ψ_j, Ψ_i^T are, respectively, exact and approximate solutions having the appropriate asymptotic form. We stress that the identity holds for rearrangement as well as direct collisions, provided the wave functions have the correct asymptotic form, that is, that they satisfy (39) in the region where the interaction between the colliding system (in both incident and outgoing channels) can be neglected. For example for a particle (**x**) scattering from a many-body target (coordinates **r**) the total Hamiltonian is

$$H = -\frac{\hbar^2}{2m}\nabla_x^2 + V(\mathbf{x}) + H_A + V(\mathbf{x},\mathbf{r}) , \tag{41}$$

where H_A is the Hamiltonian of the isolated target. If $H_A \Psi_j(\mathbf{r}) = E_j \Psi_j(r)$, then at a total energy $E = \hbar^2 k_j^2/2m + E_j$ the exact solution of (39) that describes the collision process must have the asymptotic form

$$\Psi_i(\mathbf{x},\mathbf{r}) \underset{\mathbf{x}\to\infty}{\to} e^{i\mathbf{k}_i\cdot\mathbf{x}}\Psi_i(\mathbf{r}) + \sum_f A_{fi}\frac{e^{ik_f x}}{x}\Psi_f(\mathbf{r}) , \tag{42}$$

where in (42) the sum is taken over all states f that are accessible at the energy E. Similarly the identity (40) requires that

$$\Psi_i^T(\mathbf{x},\mathbf{r}) \underset{\mathbf{x}\to\infty}{\to} e^{i\mathbf{k}_i\cdot\mathbf{x}}\Psi_i(\mathbf{r}) + \sum_l A_{li}^T\frac{e^{ik_l x}}{x}\Psi_l(\mathbf{r}) , \tag{43}$$

where A_{li}^T are some approximate scattering amplitude elements (in the simplest case some or all of these may be zero). $\tilde{\Psi}_j$ in (40) refers to a solution of (39) in which all $\mathbf{k}_i$ are replaced by $-\mathbf{k}_i$, i.e., incoming waves (see, for example, Ref. 10, p. 116). The derivation of (40) assumes that the isolated target eigenfunctions are exactly known. This, of course, will not generally be the case, but we shall not consider this additional complication at present.

For rearrangement collisions the asymptotic forms will be more complicated than the example given in (42), and the derivation of (40) is somewhat more involved due to the nonorthogonality of the initial and final state eigenfunctions (Ref. 8, pp. 436–437), but this does not change the validity of the identity (40).

Though we have written (40) for elements of the scattering amplitude, analogous identities can be written in terms of related quantities, such as the transition matrix $T_{ji} = -(2\pi\hbar^2/m_j)A_{ji}$, S–matrix ($\mathbf{S} = I - i\mathbf{T}$) etc. Indeed, in practical applications the wavefunctions and amplitudes would be partially analyzed, and (for greatest convenience) the identity (40) would be written for the elements of the real, symmetric reactance or K–matrix (see, for example, Ref. 11, pp. 137–145).

We note that the remainder term

$$R_1 = c_j \int \tilde{\Psi}_j^* \acute{H} \Psi_i^T\, d\tau , \tag{44}$$

where

$$c_j = \frac{1}{4\pi}\frac{2m_j}{\hbar^2},$$

is first order in the "small" quantity $\Psi_i^T - \Psi_i$, since $\acute{H}(\Psi_i^T - \Psi_i) = \acute{H}\Psi_i^T$. Furthermore $\acute{H}\Psi_i^T$ vanishes outside the interaction region, hence the main contribution to R_1 comes from the interaction region. Indeed, if Ψ_i^T is a solution of a model problem $\acute{H}_T\Psi_i^T = 0$, where $\acute{H}_T$ is obtained from $\acute{H}$ by replacing the interaction potential V by a solvable model V_T, then $\acute{H}\Psi_i^T = \acute{H}_T\Psi_i^T + (V - V_T)\Psi_i^T$ (alternatively, this may be viewed as a definition of V_T, if Ψ_i^T is given), and (44) becomes

$$R_1 = \int \Psi_j^*(V - V_T)\Psi_i^T\, d\tau\,. \tag{45}$$

An identity in which the remainder is second order in "small" quantities follows immediately from (40):

$$A_{ji} = A_{ji}^T - c_j \int \Phi_j^{T*}\acute{H}\Psi_i^T\, dt + R_2\,, \tag{46}$$

where

$$R_2 = c_j \int (\Phi_j^T - \tilde{\Psi}_j)^*\acute{H}\Psi_i^T\, d\tau\,. \tag{47}$$

Strictly speaking, because $\acute{H}\Psi_i^T$ vanishes outside the interaction region, Φ_j^T need be a good approximation to $\tilde{\Psi}_j$ only in the interaction region for R_2 to be second-order "small" quantity (i.e., Φ_j^T need not be the same as Ψ_j^T and indeed need not have the correct asymptotic form). In standard applications, though (say the Kohn variational method), it is usual to take Φ_i^T to be the same as Ψ_i^T. This ensures, also, that the stationary approximation K_{ij} is also real and symmetric, though the trial values by themselves may not be. For a detailed discussion of this point see Ref. 12, Chap. 2. We might note, in passing, that approximations like the polarized-orbital approximation for (say) elastic $e^+ - H$ scattering can be viewed as being variational in the sense of (46), but with $\Phi^T \neq \Psi^T$. Indeed if Ψ^T is taken to be $F(\mathbf{x})\Psi_{\rm po}(\mathbf{r};\mathbf{x})$, the usual polarized orbital trial function (see, for example, Ref. 13), while $\Phi^T = G(\mathbf{x})\Psi_{1s}(\mathbf{r})$, where $\Psi_{1s}(\mathbf{r})$ is the target ground state, then the variational ansatz

$$\langle \delta\Phi_T|\acute{H}|\Psi_T\rangle = \langle \delta G\Psi_{1s}|\acute{H}|F\Psi_{\rm po}\rangle = 0 \tag{48}$$

yields the usual polarized-orbital equation for the coefficient F, viz.,

$$\langle \Psi_{1s}(\mathbf{r})|\acute{H}|\Psi_{\rm po}(\mathbf{r};\mathbf{x})\rangle F(x) = 0\,, \tag{49}$$

and (46) implies that the phase shifts obtained form (49) are equal to the exact ones to within the second-order remainder term $\propto \int (G\Psi_{1s} - \Psi_{\rm EXACT})\acute{H}(F\Psi_{\rm po} - \Psi_{\rm EXACT})d^3x d^3r$.

The identities (40) and (46) imply that if bounds could be determined to the remainder terms R_1 and R_2 [Eq. (44) and (47)], then bounds to the individual transition matrix elements (or, equivalently, the K-matrix elements) follow immediately, since if $|R_2| \leqslant B_{ji}$, say, then $|A_{ji} - \bar{A}_{ji}| \leqslant B_{ji}$. Various formal bounds of the type considered previously for potential scattering can be written down. Thus

$$|R_1| \leqslant |\Psi_j|_M c_j \int |\acute{H}\Psi_i^T|\, d\tau\,, \tag{50}$$

where M refers to the maximum value of the expression in question, and

$$|R_1|^2 \leqslant \int |\Psi_j|^2|W|\, d\tau \int |W|^{-1}|\acute{H}\Psi_i^T|^2\, d\tau\,, \tag{51}$$

where W is a suitable weight function. The problem is thus "reduced" to the determination of bounds to $|\Psi_j|_M$ or $\int|\Psi_i|^2|W\, d\tau$. Similar inequalities hold for R_2 except that Ψ_j in (50) and (51) must be replaced by $\delta\Psi_j = \Phi_j^T - \Psi_j$.

All the bounds that were derived previously for potential scattering by use of the Lippman–Schwinger equation, generalize in a straightforward way to the many-body case, since the generalized Lippman–Schwinger equation is readily written down. For example, if Φ_j^T is taken to have the correct asymptotic form, then $\delta\Psi_j$ satisfies

$$\delta\Psi_j(\tau) = \int G(\tau,\acute{\tau})[V(\acute{\tau})\delta\Psi_j(\acute{\tau}) + \acute{H}\Psi_j^T(\acute{\tau})]\, d\acute{\tau}\,, \tag{52}$$

where G is the Green's function constructed from the (known) solutions of (39) with the interaction part of the Hamiltonian, $V = 0$. It follows immediately from (52) that

$$|\delta\Psi_j(\tau)|_M \leqslant \frac{\left|\int G(\tau,\acute{\tau})\acute{H}\Psi_j^T(\acute{\tau})\, d\acute{\tau}\right|_M}{1 - \mathrm{MAX}\int |G(\tau,\acute{\tau})V(\acute{\tau})|\, d\acute{\tau}}, \tag{53}$$

provided the denominator is positive. Bounds of the type (30) are similarly obtained from (52), though they are clearly much more complicated because the Green's function appears quadratically there.

It is, unfortunately, impossible to evaluate such bounds without making approximations (and thus destroying the rigor of the bounds), because the Green's function in this case is very complicated. Thus, for a particle being scattered by a many-body target, $G(\tau,\acute{\tau})$ can be written as

$$G(\mathbf{x},\acute{\mathbf{x}},\mathbf{r},\acute{\mathbf{r}}) = \sum_\alpha G_\alpha(\mathbf{x},\acute{\mathbf{x}})\Psi_\alpha^*(\acute{\mathbf{r}})\Psi_\alpha(r)\,, \tag{54}$$

where the sum is taken over a complete set of eigenstates Ψ_α of the target, and

$$G_\alpha(\mathbf{x},\acute{\mathbf{x}}) = -\frac{1}{4\pi}\frac{e^{ik_\alpha|\mathbf{x}-\acute{\mathbf{x}}|}}{|\mathbf{x}-\acute{\mathbf{x}}|} \tag{55}$$

is the usual free-particle Green's function with $\hbar^2 k_\alpha{}^2/2m = E - E_\alpha$, E_α being the target eigenenergy, E the total energy of the colliding system, and m the reduced mass.

An obvious approximation, which would reduce the problem to tractable proportions is to replace all the G_α in (54) by some average value $G_{\bar{\alpha}}$, corresponding to some average $k_{\bar{\alpha}}$ (which thus enters as an adjustable parameter), and then to evaluate the remaining sum by closure to obtain

$$G(\mathbf{x},\acute{\mathbf{x}},\mathbf{r},\acute{\mathbf{r}}) \simeq G_{\bar{\alpha}}(x,\acute{\mathbf{x}})\delta(\mathbf{r} - \acute{\mathbf{r}})\,. \tag{56}$$

Indeed it is possible to generalize this somewhat by keeping a finite number of terms in (54) explicitly and summing the remainder using closure, as has been done in connection with the second Born approximation.[14,15] Still simpler estimates could be made by replacing the exact Ψ_j in (50) and

(51) by an (optimized) trial from Ψ_j^T. In the case of such nonrigorous approximations to the bounds more credence to the result would be lent if two different nonrigorous estimates of the bound (but both obtained by using the same trial function) gave similar results.

The actual utility of such approximations in conjunction with the estimation of bounds has to be tested on such relatively well studied systems as scattering by H and He targets. Even then the problem is by no means simple because the evaluation of integrals like those in (50) and (51) is not easy. Indeed one would like to avoid expressions that are quadratic in the Hamiltonian, if at all possible since, except in some special cases, such expressions become intractable. In practice, it would probably be necessary to determine approximations to such integrals, preferably by expressions which are upper bounds to them, as has been suggested previously.[3] It should be pointed out that even if bounds to $|\Psi_j|_M$ are not known, (50) can be used as a criterion for determining "optimal" values of any adjustable parameters in Ψ_i^T, since $|\Psi_j|_M$ is known to be finite, hence the bound to $|R_1|$ (and, indeed, to $|R_2|$) is minimized when $\int|\acute{H}\Psi_i^T|\,d\tau$ is minimized. Similarly, minimization of the variance (or "least squares") integral in (51) also minimizes a bound to the remainder (though, in general, a different bound). Such "least squares" criteria have been used with success for single-channel elastic scattering by hydrogen atom targets in the past, most recently by Abdel-Raouf.[16]

5. CONCLUDING REMARKS

We have considered further the problem of determining simultaneous upper and lower bounds to scattering parameters, by bounding the remainder terms that arise in the Kato identities which relate the exact to the approximate results.

For the case of (central) potential scattering, where both exact results and rigorous bounds are readily calculable, we have established a number of new bounds, some based on the use of the Lippman–Schwinger equation and others on the phase-amplitude formalism of the scattering problem. The latter have the advantage over the former in that they are applicable to potentials of arbitrary strength. Partial-wave scattering by an exponential potential, as well as by a square well potential have been used to illustrate some of the results, and to compare them to others that have been obtained previously.

It is probably fair to say that with our present and previous studies,[3,4] as well as those of other researchers (Refs. 5, 7, 9, 17, and citations therein), the problem of bounds in the case of scattering by central potentials is reasonably well understood, though a direct comparison of the different approaches would be worthwhile.

However, the purpose of studying bounds in the case of central potential scattering is in that they may lead to useful (tractable) generalizations to the case of multichannel scattering, involving many-body systems. We have thus considered the Kato identities for such multi-channel scattering, and have shown that at least those bounds which are based on the use of the Lippman–Schwinger equation generalize in a straight-forward way to the many-body case. The resulting expressions are very complicated, however, and cannot be evaluated without the introduction of nonrigorous approximations. In addition, the Kato identities in this many-body case (at least as presently formulated) presuppose that the separated colliding system can be exactly solved which, except in the special case of hydrogen-like targets (and/or projectiles) is not the case.

Thus at present, in the case of multichannel scattering by many-body targets, we are only in the position of determining approximate bounds to the individual K (or related)—matrix elements. How useful such nonrigorous estimates will be will have to await their detailed study and comparison with experiment as well as with those few multichannel systems (such as $e^{\pm}$–H,He scattering) where nearly exact calculations are, at least in principle, possible. We must stress that it is multi-channel processes that are of interest to us here. It is well known that "exact" single-channel (elastic) $e^{\pm}$–H results are available, but comparable results are not yet generally available when many channels are open (including, in particular, rearrangement channels such as occur in e^+–H scattering above 7 eV). Clearly such calculations would be very desirable.

It is important to continue seeking other expressions for the bounds, particularly expressions that avoid quadratic Hamiltonians and many-body Green's functions, as these obviously become extremely complicated even for few-body systems. In addition, the computation of bounds to scattering parameters in the many-body case requires the development of efficient methods of calculating multidimensional integrals or, at least, bounds to them. This in itself is a nontrivial problem.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft, through the program of the Sonderforschungsbereich 91 at the University of Kaiserslautern, and it is gratefully acknowledged. One of us (R. P.) would also like to thank the National Science and Engineering Research Council of Canada for financial support in the form of a graduate fellowship. We would also like to thank R. K. Nesbet for many stimulating conversations.

[1]D. G. Truhlar, J. Abdallah Jr. and R. L. Smith, "Algebraic Variational Methods in Scattering Theory," *Advances in Chemical Physics*, Vol. XXV (Wiley, New York, 1974).
[2]K. Takatsuka and T. Kueno, Phys. Rev. A. **19**, 1011 (1979).
[3]J. W. Darewych and R. Pooran, Can. J. Phys. **56**, 1358 (1978).
[4]J. W. Darewych and J. Sokoloff, J. Math. Phys. **20**, 736 (1979).
[5]T. Kato, Prog. Theoret. Phys. (Japan) **6**, 394 (1951).
[6]F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967).
[7]J. N. Bardsley, E. Gerjuoy and C. V. Sukumar, Phys. Rev. A **6**, 1813 (1972).
[8]N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collision* 3rd Ed. (Claredon, Oxford, 1965).
[9]A. Martin, Commun. Math. Phys. **69**, 89 (1979).
[10]S. Geltman, *Topics in Atomic Collision Theory* (Academic, New York, 1969).
[11]B. H. Bransden, *Atomic Collision Theory* (Benjamin, New York, 1970).
[12]R. K. Nesbet, *Variational Methods in Electron-Atom Scattering Theory*,

IBM Research Report RJ2533 (33073) 5/9/79, Yorktown Heights, N.Y., 1979 (also Plenum, New York, to be published).
[13]R. J. Drachman and A. Temkin, "The Polarized Orbital Approximation," *Case Studies in Atomic Collision Physics* II, edited by E. W. McDaniel and M. R. C. McDowell (North-Holland, Amsterdam, 1972), Chap. 6.
[14]H. S. W. Massey and C. B. O. Mohr, Proc. Roy. Soc. (London) A **146**, 880 (1934).
[15]A. R. Holt and B. L. Moiseiwitsch, J. Phys. B **1**, 36 (1968).
[16]M. A. Abdel-Raouf, J. Phys. B **12**, 3349 (1979).
[17]B. L. Burrows and A. J. Perks, J. Phys. B **12**, 2687 (1979).

The existence of wave operators for oscillating potentials [a)]

Allen Devinatz
Northwestern University, Evanston, Illinois 60201

(Received 6 December 1979; accepted for publication 14 March 1980)

The existence of the Møller wave operators is proved for Hamiltonians of the form $H = -\Delta + a\sin br^{\alpha}/r^{\beta} + V(x)$, where V is a short range potential, generally noncentral, and α and β take on suitable values including the case $\alpha = 1, \beta > \frac{1}{2}$.

1. INTRODUCTION AND RESULTS

All of the main features of quantum mechanical potential scattering with short-range potentials are now well understood.[1,2] However, there are long-range potentials of an oscillatory nature for which the situation is not yet so clear. Our purpose here is to make a contribution to this latter theory.

We shall be working in R^n with Schrödinger operators of the form

$$H = -\Delta + a\sin br^{\alpha}/r^{\beta} + V(x), \tag{1.1}$$

where V is a short-range potential which is not necessarily central. We shall prove the existence of the Møller wave operators for these potentials. For suitable α and β the existence and completeness of wave operators for Hamiltonians of the form (1.1) with V a central short-range potential have been recently obtained.[3,4]

Long-range oscillatory potentials of the form $\nabla\cdot W$, where W is a short-range R^n valued function, have been considered in the literature by various authors.[5,6] However, Hamiltonians of the form (1.1) are included in this setup only for $\alpha + \beta > 2$, and hence do not include what may be considered to be the benchmark case, namely $r^{-1}\sin r$. We note that a modification of the methods of this paper can be used to handle highly oscillatory potentials such as $e^r \sin e^r/(1+r)^{\beta}$ in place of $ar^{-\beta}\sin br^{\alpha}$.

In taking conditions on our short-range potential we shall follow the prescription given by V. Enss.[7] Let H_0 be the free Hamiltonian and let $F(\cdot)$ denote the projection operator from $L^2(R^n)$ onto the subspace of elements in L^2 with support in the set specified within the brackets. We shall suppose that

(i) V is a real-valued potential which is H_0-bounded, with H_0-bound less than 1, and

(ii) $\|V(H_0+i)^{-1}F(|x|\geqslant r)\| \equiv h(r) \in L^1(R^+, dr)$.

Theorem: Under the above specifications for V, the Møller wave operators for the Hamiltonian (1.1) exist for $2\beta + \alpha > 2$, or $\beta > \frac{1}{2}$ and $\alpha + \beta > 1$, and $\beta - \alpha < 2$ for $n > 3$, $\beta - \alpha < n/2$ for $n = 1,2,3$.

We should remark that the condition $\beta - \alpha < 2$ or $\beta - \alpha < n/2$ is needed only to eliminate a strong singularity at the origin. If we had used the potential $a\sin br^{\alpha}/(1+r)^{\beta}$, this condition would not be needed. We should also note that, for $V = 0$, B. Bourgeois[8] has shown the existence of the wave operators for the Hamiltonian (1.1) for $\beta > \frac{1}{2}$ and $\alpha + \beta > 1$.

2. PROOF OF THE THEOREM

Our proof will be constructed with the aid of several lemmas. Let

$$H' = -\Delta + a\sin br^{\alpha}/r^{\beta} \tag{2.1}$$

and let $\Omega_0^{\pm}$ be the Møller wave operators for H'. Our first lemma is of an elementary nature and is intended to indicate those facts about $\Omega_0^{\pm}$ which are sufficient to prove the existence of the wave operators for H.

Lemma 2.1: Suppose that the potential $ar^{-\beta}\sin br^{\alpha}$ has H_0-bound zero and that $\Omega_0^{\pm}$ exists. Suppose, also, that there is a dense collection $\mathcal{D} \subseteq D(H_0) \subseteq L^2(R^n)$ so that $f \in \mathcal{D}$ implies there is an $a > 0$ such that

$$\|F(|x| < a|t|)\Omega_0^{\pm} e^{-itH_0}(H_0+i)f\| \in L^1(R, dt). \tag{2.2}$$

Then the wave operators for H exist.

Proof: Let us first note that by the hypotheses on V and $ar^{-\beta}\sin br^{\alpha}$, the potential $ar^{-\beta}\sin br^{\alpha} + V(x)$ has H_0-bound less than 1, and thus H has a unique self-adjoint realization and $D(H) = D(H') = D(H_0)$. Thus $\exp(-itH')\Omega_0^{\pm} f \in D(H) \cap D(H')$.

Next we note that

$$\Omega^{\pm}(H, H_0) = \underset{t \to \pm\infty}{s\text{-lim}}\, e^{itH} e^{-itH'} \Omega_0^{\pm}, \tag{2.3}$$

provided this limit exists. By the usual technique used to prove Cook's theorem, i.e., integration of the derivative of the right-hand side of (2.3) evaluated at $f \in \mathcal{D}$, we have

$$\begin{aligned}\|(e^{itH}e^{-itH'} - e^{isH}e^{-isH'})\Omega_0^{\pm} f\| &\leqslant \int_s^t \|Ve^{-i\tau H'}\Omega_0^{\pm} f\|\, d\tau \\ &\leqslant \int_s^t \|V\Omega_0^{\pm} e^{-i\tau H_0} f\|\, d\tau. \end{aligned} \tag{2.4}$$

We have, of course, used the intertwining relation to get the last integral.

We may now write

$$\|V\Omega_0^{\pm} e^{-itH_0} f\| = \|V(H'+i)^{-1}\Omega_0^{\pm} e^{-itH_0}(H_0+i)f\|. \tag{2.5}$$

Further,

$$\begin{aligned}\|V(H'+i)^{-1}\Omega_0^{\pm} e^{-itH_0}(H_0+i)f\| &\leqslant \|V(H'+i)^{-1}F(|x| < a|t|)\Omega_0^{\pm} e^{-itH_0}(H_0+i)f\| \\ &\quad + \|V(H'+i)^{-1}F(|x| \geqslant a|t|)\Omega_0^{\pm} e^{-itH_0}(H_0+i)f\|. \end{aligned} \tag{2.6}$$

[a)]Research partially supported by NSF Grant MCS79-02538.

 0022-2488/80/092406-06$1.00

Since $\|V(H'+i)^{-1}\|$ is finite, by the hypothesis (2.2) the first term on the right of (2.6) is integrable.

Since V is a real-valued potential, it follows that $\|V(H_0+i)^{-1}F(|x|\geqslant a|t|)\|$ is integrable if and only if $\|F(|x|\geqslant a|t|)V(H_0+i)^{-1}\|$ is integrable. Also, using the fact that the potential $ar^{-\beta}\sin br^{\alpha}$ is H_0 bounded, it also follows that $\|V(H'+i)^{-1}F(|x|\geqslant a|t|)\|$ is integrable if and only if $\|F(|x|\geqslant a|t|)V(H'+i)^{-1}\|$ is integrable. Now

$$\begin{aligned}&\|F(|x|\geqslant a|t|)V(H'+i)^{-1}\|\\&\quad=\|F(|x|\geqslant a|t|)V(H_0+i)^{-1}(H_0+i)(H'+i)^{-1}\|\\&\leqslant\|(H_0+i)(H'+i)^{-1}\|\,\|F(|x|\geqslant a|t|)V(H_0+i)^{-1}\|.\end{aligned}\tag{2.7}$$

By our hypothesis on V we see that the term on the right is integrable. Thus the second term on the right of (2.6) is integrable. Hence, the lemma is established.

We see from Lemma 2.1 that in order to prove the theorem of Sec. 1, it is sufficient to establish (2.2) for a suitable dense set of functions. The remaining part of the paper is devoted to establishing this fact and constitutes the bulk of the work in this paper.

The hypothesis (2.2) involves only the operators H_0 and H'. If $\beta-\alpha<2$ for $n>3$ or $\beta-\alpha<n/2$ for $n=1,2,3$, then H' is H_0-compact and thus H_0-bounded with zero H_0-bound. Consequently, it has a unique self-adjoint realization. Further, since the potential of H' is radially symmetric we may reduce the study of the self-adjoint realization of H' to the study of the self-adjoint realizations of certain ordinary differential operators on R^+. As is well known, the self-adjoint realization of H' is unitarily equivalent to a direct sum of certain self-adjoint realizations of the ordinary differential operators in $L^2(R^+)$ given by

$$H_l'=-\frac{d^2}{dr^2}+\frac{\nu(l,n)}{r^2}+V_L(r),\tag{2.8}$$

where $\nu(l,n)=l(l+n-2)+(n-1)(n-3)/4$ and $V_L(r)=ar^{-\beta}\sin br^{\alpha}$. A detailed discussion of the exact boundary conditions needed at $r=0$ will be found in the appendix to Ref. 4.

In order to get the estimate (2.2) it is necessary to study certain solutions to the equation

$$H_l'\phi=\lambda^2\phi,\quad\lambda>0.\tag{2.9}$$

Lemma 2.2: Assume $\alpha,\beta>0$ and either $2\beta+\alpha>2$, or $\beta>\frac{1}{2}$ and $\alpha+\beta>1$. For every positive integer m there exists a finite number of points in R^+ so that if I is any open interval with compact closure in R^+ which is disjoint from this finite set of points, then for $\lambda\in I$ there exist solutions $\phi^{\pm}(r,\lambda)$ to (2.9) with the following properties:

(i) $\phi^{\pm}$ and $d\phi^{\pm}/dr$ are m-times differentiable in λ, and

$$\text{(ii)}\ \phi^{\pm}(r,\lambda)=e^{\pm i\lambda r}+\left[\sum_{j=1}^{k}c_j^{\pm}(\lambda)g_j^{\pm}(r)+c_{k+1}^{\pm}(r,\lambda)\right]e^{\pm i\lambda r}+\rho^{\pm}(r,\lambda),$$

where $c_j^{\pm}(\lambda)$ are infinitely differentiable, $g_j^{\pm}(r)=o(1)$ as $r\to\infty$, and $c_{k+1}^{\pm}$ and $\rho^{\pm}$ and their first m-derivatives with respect to λ are $O(r^{-1/2-\epsilon})$ as $r\to\infty$.

Let us delay the proof of this until later and use it now to complete the proof of our theorem. First we state one further lemma about solutions to (2.9).

Lemma 2.3: Suppose $\beta-\alpha<2$; then (2.9) has a solution $\eta(r,\lambda)$ such that, for $\lambda\in R^+$,

(i) $\eta(r,\lambda)$ and $d\eta(r,\lambda)/dr$ are infinitely differentiable in λ,

(iia) $\eta(r,\lambda)\to0$ as $r\to0$ for $l\neq0$ or $n\neq2$,

(iib) $r^{1/2}(\log r)^{-1}d\eta(r,\lambda)/d\lambda\to0$ as $r\to0$ for $l=0$, $n=2$.

This is Lemma 3.3 of Ref. 4. The conditions (iia) and (iib) are the boundary conditions needed to get the proper self-adjoint realizations of H_l'. See the appendix to Ref. 4 for details.

We now use the previous two lemmas to complete the proof of the theorem of Sec. 1. Since $\phi^{\pm}$ are linearly independent solutions to (2.9), we may write

$$\eta(r,\lambda)=c_1(\lambda)\phi^{+}(r,\lambda)+c_2(\lambda)\phi^{-}(r,\lambda),$$

$$\frac{d\eta(r,\lambda)}{dr}=c_1(\lambda)\frac{d\phi^{+}(r,\lambda)}{dr}+c_2(\lambda)\frac{d\phi^{-}(r,\lambda)}{dr}.$$

It follows that $c_1(\lambda)$ and $c_2(\lambda)$ are m-times differentiable functions of λ in I. Using the expansion (ii) for $\phi^{\pm}$ given in Lemma 2.2, we may write

$$\begin{aligned}\eta(r,\lambda)&=d_1(\lambda)\sin\lambda r+d_2(\lambda)\cos\lambda r\\&\quad+\left[\sum_{j=1}^{k}c_j^{+}(\lambda)g_j^{+}(r)+c_{k+1}^{+}(r,\lambda)\right]e^{i\lambda r}\\&\quad+\left[\sum_{j=1}^{k}c_j^{-}(\lambda)g_j^{-}(r)+c_{k+1}^{-}(r,\lambda)\right]e^{-i\lambda r}+\rho(r,\lambda),\end{aligned}\tag{2.10}$$

where now $c_j^{\pm}(\lambda)$ are m-times differentiable, and $c_{k+1}^{\pm}$ and ρ satisfy the same conditions as $c_{k+1}^{\pm}$ and $\rho^{\pm}$ of Lemma 2.2. Since $d_1^2+d_2^2\neq0$, we may divide both sides of (2.10) by this quantity and get another solution for (2.9) which satisfies the same properties as the old solution, except it may now be only m-times differentiable. Thus we may suppose that we have a solution of the form (2.10) with d_1 and d_2 m-times differentiable on I and $d_1^2+d_2^2=1$. The Monodromy theorem tells us that there is an m-times differentiable function $\delta(\lambda)$ on I such that

$$d_1(\lambda)+id_2(\lambda)=e^{i\delta(\lambda)}.\tag{2.11}$$

Thus in (2.10) we may write

$$d_1(\lambda)\sin\lambda r+d_2(\lambda)\cos\lambda r=\sin[\lambda r+\delta(\lambda)].\tag{2.12}$$

If we work with the free Hamiltonian, then the expansion (ii) of Lemma 2.2 is valid and there is an m-differentiable function $\delta_0(\lambda)$ and a solution $\eta_0(r,\lambda)$ to

$$-\frac{d^2\phi}{dr^2}+\frac{\nu(l,n)}{r^2}\phi=\lambda^2\phi,\tag{2.13}$$

which has the same properties as $\eta(r,\lambda)$ and

$$\begin{aligned}\eta_0(r,\lambda)&=\sin[r\lambda+\delta_0(\lambda)]+c^{+}(r,\lambda)e^{i\lambda r}\\&\quad+c^{-}(r,\lambda)e^{-i\lambda r}+\rho(r,\lambda).\end{aligned}\tag{2.14}$$

In this case the functions $c^{\pm}$ and ρ and their first m derivatives with respect to λ are $O(r^{-1/2-\epsilon})$ as $r\to\infty$.

Lemma 2.4: If $2\beta + \alpha > 2$, or if $\beta > \frac{1}{2}$ and $\alpha + \beta > 1$, and if $\beta - \alpha < 2$ for $n > 3$ and $\beta - \alpha < n/2$ for $1 \leqslant n \leqslant 3$, then the wave operators for H' exist.

Proof: We first note that the condition $\beta - \alpha < 2$ for $n > 3$ or $\beta - \alpha < n/2$ for $1 \leqslant n \leqslant 3$ implies the H' is H_0-compact and hence H_0-bounded with zero H_0-bound. Hence H' has a unique self-adjoint realization.

The proof of the existence of the wave operators for H' now follows standard techniques by using the expansions (2.10), (2.12), and (2.14). Let us indicate some of the details since there are some small points which may need elucidation.

Let us set

$$\delta_1(\lambda) = \delta(\lambda) - \delta_0(\lambda). \tag{2.15}$$

If $\tilde{f} \in C_0^\infty(I)$ let us set

$$f(r) = \int_0^\infty \eta_0(r,\lambda)\tilde{f}(\lambda)\, d\lambda, \tag{2.16}$$

$$\Omega_0^+ f(r) = \int_0^\infty \eta(r,\lambda) e^{-i\delta_1(\lambda)} \tilde{f}(\lambda)\, d\lambda. \tag{2.17}$$

The operator Ω_0^+ is actually unitarily equivalent to the wave operator for H', but we shall not change notation. We shall also not change our notation for H' or H_0 under this unitary equivalence. We note that we have suppressed the subscript l so that we are working generically. We also note that, as $I, \tilde{f}(\lambda)$, and l vary, we get a linear manifold of functions in an infinite direct sum of the spaces $L^2(R^+)$ which is unitarily equivalent to a dense set of functions in $L^2(R^n)$.

Using (2.16) and (2.17), we get

$$(\Omega_0^+ - I)e^{-itH_0} f = \int_0^\infty (\eta e^{-i\delta_1} - \eta_0) e^{-i\lambda^2 t} \tilde{f}\, d\lambda. \tag{2.18}$$

Now

$$\eta e^{-i\delta_1} - \eta_0 = \frac{1}{2i} e^{-i\lambda r} e^{-i\delta_0}(1 - e^{-2i\delta_1}) + \left[\sum_{j=1}^{k} c_j^+(\lambda) g_j^+(r) + c_{k+1}^+(r,\lambda)\right] e^{i\lambda r} + \left[\sum_{j=1}^{k} c_j^-(\lambda) g_j^-(r) + c_{k+1}^-(r,\lambda)\right] e^{-i\lambda r} + \rho. \tag{2.19}$$

Noting that $c_{k+1}^\pm$ and ρ are $O(r^{-1/2-\epsilon})$ as $r \to \infty$, if we put these terms into (2.18) we see the resulting function goes to zero in L^2 as $t \to \infty$. If we put the first term on the right of (2.19) and the terms involving $c_j^- g_j^-$ into (2.18) and integrate by parts we see that this also goes to zero in the L^2 norm as $t \to \infty$. It remains to deal with the terms $c_j^+ g_j^+$.

Let

$$h(s,t) = g_j^+(s) \int_0^\infty e^{i\lambda s} e^{-i\lambda^2 t} c_j^+(\lambda)\tilde{f}(\lambda)\, d\lambda.$$

Then

$$|h(s,t)|^2 = o(1)\left| \int_0^\infty e^{i\lambda s} e^{-i\lambda^2 t} c_j^+(\lambda)\tilde{f}(\lambda)\, d\lambda \right|^2,$$

so that

$$\int_r^\infty |h(s,t)|^2\, ds = o(1) \int_0^\infty \left| \int_0^\infty e^{i\lambda s} e^{-i\lambda^2 t} c_j^+(\lambda)\tilde{f}(\lambda)\, d\lambda \right|^2 ds = o(1) \int_0^\infty |c_j^+(\lambda)\tilde{f}(\lambda)|^2\, d\lambda.$$

Hence, independent of $t > 0$,

$$\int_r^\infty |h(s,t)|^2\, ds \to 0 \quad \text{as} \quad r \to \infty.$$

Using the Riemann–Lebesque lemma, this implies

$$\int_0^\infty |h(s,t)|^2\, ds \to 0 \quad \text{as} \quad t \to \infty.$$

We are now in a position to get the estimate (2.2) of Lemma 2.1. As we noted before, we will be working with operators which are unitarily equivalent to $\Omega_0^\pm$, H_0 and H', but we shall not change our notation.

For simplicity, let us replace the coefficients of $\exp(\pm i\lambda r)$ in Formula (2.10) by $c^\pm(r,\lambda)$. Then from (2.16) and (2.17) we get

$$\begin{aligned}
&\Omega_0^+ e^{-itH_0}(H_0 + i) f(r) \\
&= \int_0^\infty \eta(r,\lambda) e^{-i\delta_1(\lambda)} e^{-it\lambda^2}(\lambda^2 + i)\tilde{f}(\lambda)\, d\lambda \\
&= \int_0^\infty e^{-i(\lambda r + \lambda^2 t)}[c^-(r,\lambda) - 1/2i] e^{-i\delta_1(\lambda)}(\lambda^2 + i)\tilde{f}\, d\lambda \\
&\quad + \int_0^\infty e^{i(\lambda r - \lambda^2 t)}[c^+(r,\lambda) + 1/2i] e^{-i\delta_1(\lambda)}(\lambda^2 + i)\tilde{f}\, d\lambda \\
&\quad + \int_0^\infty e^{-it\lambda^2} \rho(r,\lambda) e^{-i\delta_1(\lambda)}(\lambda^2 + i)\tilde{f}\, d\lambda.
\end{aligned}$$

The first integral may be integrated by parts twice and in so doing we find it is $O(|r + t|^{-2})$. If $\operatorname{supp}\tilde{f} \subseteq [a, \infty)$, then integrating the second integral by parts twice shows that it is $O(|r - 2at|^{-2})$. An integration by parts twice of the third integral shows that it is $O(r^{-1/2-\epsilon} t^{-2})$. Hence we see that

$$\|F(|x| < a|t|)\Omega_0^+ e^{-itH_0}(H_0 + i) f\| \in L^1(R, dt).$$

The remainder of the paper will be devoted to the following proof.

Proof of Lemma 2.2: Let us suppose, at first, that $2\beta + \alpha > 2$ and $\alpha \geqslant 1$. The differential equation (2.9) can be put in the matrix form

$$u' = \begin{bmatrix} 0 & 1 \\ V_L + \nu/r^2 - \lambda^2 & 0 \end{bmatrix} u, \tag{2.20}$$

where $V_L(r) = ar^{-\beta} \sin br^\alpha$. The term $\nu(l,n)/r^2$ and its derivatives decrease fast enough at infinity so that it plays no essential role in the computations which follow. Hence, for the sake of simplicity, we will drop it. Our computations will proceed very much like the computations in Ref. 4. In some aspects they are considerably easier since it is not necessary here to deal with complex values of λ. On the other hand, we are asking for more precise information so that we must pay a little closer attention to some of the estimates. We take this opportunity to note that the asymptotic estimates of Lemma 3.1 of Ref. 4 must be modified for those α and β where $\alpha < 1$, and $\alpha + \beta \leqslant 1$ or $\beta \leqslant \frac{1}{2}$. However, all of the conclusions of the paper remain valid except that we do not get the existence of the wave operator for all of the values of α and β that we claim, when $\alpha < 1$ (see Ref. 9).

Let us set

$$P=\begin{bmatrix} 1 & 1 \\ i\lambda & -i\lambda \end{bmatrix}$$

and let $u = Pw$. The Eq. (2.20) is transformed to

$$w' = (\Lambda + W)w, \tag{2.21}$$

where

$$W=\frac{iV_L}{2\lambda}\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \quad \Lambda = \operatorname{diag}[i\lambda, -i\lambda\,]. \tag{2.22}$$

Next, let us set

$$Q(r,\lambda)=\begin{bmatrix} 0 & q_{12}(r,\lambda) \\ q_{21}(r,\lambda) & 0 \end{bmatrix},$$

where Q is continuously differentiable in r and $Q(r,\lambda)\to 0$ as $r\to\infty$. If we make the transformation $w = (I + Q)y$, Eq. (2.21) is transformed (for large r) to

$$y' = (\Lambda + Y)y, \tag{2.23}$$

$$\text{where } (I + Q)Y = \Lambda Q - Q\Lambda + W + WQ - Q'. \tag{2.24}$$

We choose Q so that

$$Q' = \Lambda Q - Q\Lambda + W - \operatorname{diag} W. \tag{2.25}$$

This leads to the differential equations

$$q'_{12}(r,\lambda) = 2i\lambda q_{12}(r,\lambda) - iV_L/2\lambda,$$
$$q'_{21}(r,\lambda) = -2i\lambda q_{21}(r,\lambda) + iV_L/2\lambda.$$

Solutions to these equations are given by

$$q_{12}(r,\lambda) = (i/2\lambda)e^{2i\lambda r}\int_r^\infty e^{-2i\lambda s}V_L(s)\,ds,$$
$$q_{21}(r,\lambda) = -(i/2\lambda)e^{-2i\lambda r}\int_r^\infty e^{2i\lambda s}V_L(s)\,ds. \tag{2.26}$$

In the expression for V_L replace $\sin br^\alpha$ by sum of exponentials. For any real number γ, no matter how large, successive integration by parts of the integrals in (2.26) shows that q_{12} and q_{21} can be written as a sum of functions of the form

$$(e^{\pm ibr^\alpha}/r^{\alpha+\beta-1})c^{\pm}(\lambda)h^{\pm}(r) \quad \text{and} \quad h(r,\lambda), \tag{2.27}$$

where $c^{\pm}(\lambda)$ are infinitely differentiable, each $h^{\pm}(r)$ is of the form

$$h^{\pm}(r) = 1/r^\epsilon, \quad \text{some } \epsilon > 0,$$
$$h(r,\lambda) = O(1/r^\gamma), \tag{2.28}$$
$$\partial^k h(r,\lambda)/\partial\lambda^k = O(1/r^{\gamma-2k}).$$

If $\alpha = 1$ and $\lambda^2 = b^2/4$ we cannot integrate by parts to get the functions (2.27) so that in this case $|b|/2$ is one of the exceptional points in R^+. From (2.27) and (2.28) we see that as $r\to\infty$, and for $0\leqslant k\leqslant m$,

$$\partial^k Q/\partial\lambda^k = O(1/r^{\alpha+\beta-1}). \tag{2.29}$$

If in Eq. (2.23) we take $Y = \operatorname{diag} W + Y_0$, then Eq. (2.23) takes the form

$$y' = (\Lambda_0 + Y_0)y, \tag{2.30}$$

where $\Lambda_0 = \Lambda + \operatorname{diag} W$, and

$$(I + Q)Y_0 = -Q\operatorname{diag} W + WQ. \tag{2.31}$$

If we note that W and $\partial^k W/\partial\lambda^k$ are $O(r^{-\beta})$ as $r\to\infty$, then differentiations of (2.31) with respect to λ and the estimates (2.29) show that as $r\to\infty$, and for $0\leqslant k\leqslant m$,

$$\partial^k Y_0/\partial\lambda^k = O(1/r^{2\beta+\alpha-1}). \tag{2.32}$$

Moreover, if we operate on (2.31) by $(I + Q)^{-1}$, and expand this in a power series in Q, then from the form (2.27) of Q we see that the components of Y_0 are of the form

$$(e^{\pm ikbr^\alpha}/r^{2\beta+\alpha-1})c_k^{\pm}(\lambda)h_k^{\pm}(r) \quad \text{and} \quad h_0(r,\lambda), \tag{2.33}$$

where k runs over a finite set of nonnegative integers, the $c_k^{\pm}(\lambda)$ are infinitely differentiable, and the $h_k^{\pm}(r)$ and $h_0(r,\lambda)$ satisfy (2.28) for some sufficiently large γ.

We now transform the differential equation (2.30) by a transformation $y = (I + Q_1)y_1$, where we take the components $q_{11}^{(1)} = q_{22}^{(1)} = 0$, and, for $r\geqslant r_0 > 0$, r_0 sufficiently large,

$$q_{12}^{(1)}(r,\lambda) = -\exp\left(2\int_{r_0}^r \mu^+\right) \times \int_r^\infty \exp\left(-2\int_{r_0}^s \mu^+\right)[Y_0 - \operatorname{diag} Y_0]_{12}\,ds, \tag{2.34}$$

$$q_{21}^{(1)}(r,\lambda) = -\exp\left(2\int_{r_0}^r \mu^-\right) \times \int_r^\infty \exp\left(-2\int_{r_0}^s \mu^+\right)[Y_0 - \operatorname{diag} Y_0]_{21}\,ds,$$

where $\Lambda_0 = \operatorname{diag}[\mu^+, \mu^-]$. Note that the integrals in (2.34) converge since the integrands are $O(1/r^{2\beta+\alpha-1})$ and we are supposing $2\beta + \alpha > 2$.

Noting the form (2.33) of the components of Y_0, if we integrate the integrals in (2.34) successively by parts we find that $q_{12}^{(1)}$ and $q_{21}^{(1)}$ can be written as sums of functions of the form (2.33), where again (2.28) is valid. If $\alpha = 1$ we must exclude a finite number of points λ from R^+ in order to get functions of the form (2.33). We note that since the function $h_0(r,\lambda)$ of (2.33) has a very high rate of decrease at infinity, it is not necessary to integrate by parts any term involving h_0 as a multiple in order to get functions of the form (2.33).

From (2.33) and the estimates (2.28) we see that as $r\to\infty$ and for $0\leqslant k\leqslant m$

$$\partial^k Q_1/\partial\lambda^k = O(1/r^{2\beta+\alpha-1}). \tag{2.35}$$

The differential equation (2.30) transforms to

$$y'_1 = (\Lambda_0 + \operatorname{diag} Y_0 + Y_1)y_1 = (\Lambda_1 + Y_1)y_1, \tag{2.36}$$

where

$$(I + Q_1)Y_1 = -Q_1\operatorname{diag} Y_0 + Y_0Q_1. \tag{2.37}$$

If we proceed as before we find that for $0\leqslant k\leqslant m$

$$\partial^k Y_1/\partial\lambda^k = O(1/r^{2(2\beta+\alpha-1)}), \tag{2.38}$$

and the components of Y_1 are sums of functions of the form

$$(e^{\pm ikh^\alpha}/r^{2(2\beta+\alpha-1)})c_k^{\pm}(\lambda)h_k^{\pm}(r) \quad \text{and} \quad h_0(r,\lambda), \tag{2.39}$$

where the $c_k^{\pm}$ are infinitely differentiable and the $h_k^{\pm}(r)$ and $h_0(r,\lambda)$ satisfy (2.28).

We proceed by induction. At the mth stage we have a differential equation

$$y'_m = (\Lambda_m + Y_m)y_m, \tag{2.40}$$

where

$$\Lambda_m = \Lambda + \operatorname{diag} W + \sum_{k=0}^{m-1} \operatorname{diag} Y_k , \tag{2.41}$$

and

$$(I + Q_m)Y_m = -Q_m \operatorname{diag} Y_{m-1} + Y_{m-1} Q_m .$$

For $0 \leqslant k \leqslant m$, and for $r \to \infty$ we get

$$\partial^k Q_m / \partial \lambda^k = O(1/r^{2(m-1)(2\beta+\alpha-1)}), \tag{2.42}$$

and

$$\partial^k Y_m / \partial \lambda^k = O(1/r^{2m(2\beta+\alpha-1)}). \tag{2.43}$$

Further, the components of Y_m are of the form

$$(e^{\pm ikh^\alpha}/r^{2m(2\beta+\alpha-1)}) c_k^{\pm}(\lambda) h_k^{\pm}(r) \quad \text{and} \quad h_0(r,\lambda). \tag{2.44}$$

Let $\operatorname{diag} Y_j = \operatorname{diag}[\mu_j^+, \mu_j^-]$, $0 \leqslant j \leqslant m-1$. From (2.32) and (2.43) and the fact that $2\beta + \alpha - 1 > 1$, we see that

$$\int_{r_1}^{\infty} |\mu_j^{\pm}| < \infty ,$$

where r_1 is sufficiently large so that $\mu_j^{\pm}(r,\lambda)$ is defined for $r \geqslant r_1$. Let $\operatorname{diag} \Lambda_m = \operatorname{diag}[\nu^+, \nu^-]$. Since $\operatorname{diag}[\Lambda + W]$ is purely imaginary, we have

$$\max \sup_{r \geqslant r_1} \left| \exp \int_{r_1}^{r} \nu^{\pm} \right| \leqslant \max_j \exp \int_{r_1}^{\infty} |\mu_j^{\pm}| = M < \infty . \tag{2.45}$$

We now use Levinson's method to get certain solutions to the differential equation (2.40), together with asymptotic estimates. Choose $r_0 \geqslant r_1$ sufficiently large so that for all $\lambda \in I$

$$\int_{r_0}^{\infty} |Y_m(r,\lambda)| \, dr \leqslant 1/2M^2. \tag{2.46}$$

We are taking the norm of a matrix as the maximum of the absolute values of its entries. Let B be the space of all functions $f(r,\lambda)$, continuous on $[r_0,\infty) \times I$ so that $\partial^k f / \partial \lambda^k$ exist and are continuous on $[r_0,\infty) \times I$ for $1 \leqslant k \leqslant m$, and

$$\|f\| = \sum_{k=0}^{m} \sup_{[r_0,\infty)\times I} \left| \frac{1}{r^k} \frac{\partial^k f(r,\lambda)}{\partial \lambda^k} \right| < \infty .$$

Under this norm, B is a Banach space.

Let $\Gamma : B \to B$ be the bounded operator given by

$$(\Gamma f)(r,\lambda) = -\int_r^{\infty} \Psi(r,\lambda)\Psi(s,\lambda)^{-1} Y_m(s,\lambda) f(s,\lambda) \, d\lambda, \tag{2.47}$$

where

$$\Psi(r,\lambda) = \operatorname{diag}\left[\exp \int_{r_0}^{r} \nu^+ , \exp \int_{r_0}^{r} \nu^- \right].$$

From (2.45) it is clear that $|\Psi(r,\lambda)|$ and $|\Psi(s,\lambda)^{-1}|$ are both bounded by M on $[r_0,\infty) \times I$. Thus, from (2.46) and (2.47) it follows that $\|\Gamma\| \leqslant \frac{1}{2}$.

From the form of ν^+ and ν^-, and the estimates (2.32) and (2.43), it is clear that as $r \to \infty$

$$\frac{\partial^k \Psi}{\partial \lambda^k}, \frac{\partial^k \Psi^{-1}}{\partial \lambda^k} = O(r^k), \quad 1 \leqslant k \leqslant m.$$

Because of the estimate (2.43) on Y_m and its derivatives with respect to λ, Γf can be differentiated m times under the integral sign.

Let ψ^+ and ψ^- be the first and second columns of Ψ respectively. Clearly ψ^+ and ψ^- belong to B. Set

$$y_m^{\pm} = (I - \Gamma)^{-1} \psi^{\pm},$$

so that

$$y_m^{\pm} = \psi^{\pm} + \Gamma y_m^{\pm}. \tag{2.48}$$

The functions $y_m^{\pm}$ satisfy the differential equation (2.40). If we transform back to get solutions of the differential equation (2.21), we have

$$\begin{aligned} w^{\pm} &= \psi^{\pm} + [Q + (I+Q)Q_1 + (I+Q)(I+Q_1)Q_2 \\ &\quad + \cdots]\psi^{\pm} + (I+Q)(I+Q_1)\cdots(I+Q_m)\Gamma y_m^{\pm} \\ &= \psi^{\pm} + C^{\pm}\psi^{\pm} + R^{\pm}. \end{aligned} \tag{2.49}$$

From the estimates (2.29), (2.36), and (2.42) on $Q, Q_1, \ldots, Q_m$ it is clear that for $\alpha \geqslant 1$, $2\beta + \alpha > 2$, as $r \to \infty$, and for $0 \leqslant k \leqslant m$,

$$\partial^k C^{\pm} / \partial \lambda^k = O(1/r^{\alpha+\beta-1}) = O(1/r^{1/2+\epsilon}). \tag{2.50}$$

Further, since $y_m \in B$, $\partial^k y_m / \partial \lambda^k = O(r^k)$, $0 \leqslant k \leqslant m$, as $r \to \infty$. Hence from the integral expression for $\Gamma y_m^{\pm}$ and the rate of decrease of Y_m, it is clear that as $r \to \infty$, and for $0 \leqslant k \leqslant m$

$$\partial^k R^{\pm} / \partial \lambda^k = O(1/r^{1/2+\epsilon}). \tag{2.51}$$

Let $\operatorname{diag}[W + \Sigma_{r=0}^{m-1} Y_k] = \operatorname{diag}[\omega^+, \omega^-]$. Since the entries of $\operatorname{diag} W$ are conditionally integrable and the entries of $\operatorname{diag} \Sigma Y_k$ are absolutely integrable we may write

$$\exp \int_{r_0}^{r} \omega^{\pm} = c \exp\left(- \int_r^{\infty} \omega^{\pm} \right).$$

Thus we may write

$$\begin{aligned} \psi^{\pm} &= e^{\pm i\lambda r} \exp\left(\int_{r_0}^{r} \omega^{\pm} \right) p_{\pm} \\ &= c e^{\pm i\lambda r} \exp\left(- \int_r^{\infty} \omega^{\pm} \right) p_{\pm} \\ &= c e^{\pm i\lambda r} p_{\pm} + \left[\exp\left(- \int_r^{\infty} \omega^{\pm} \right) - 1 \right] e^{\pm i\lambda r} p_{\pm} , \end{aligned}$$

where $p_+ = (1,0)$, $p_- = (0,1)$.

From the form of the components of the Y_k [see (2.44)] it follows that $[\exp(-\int_r^{\infty} \omega^{\pm}) - 1]$ can be written as a sum of terms

$$\sum_1^k c_j^{\pm}(\lambda) g_j^{\pm}(r) + c_{k+1}^{\pm}(r,\lambda),$$

as in (ii) of Lemma 2.2. If we now transform by P and take the components of u, then we see we have solutions of (2.9) which satisfy the conditions of Lemma 2.2, at least for $r \geqslant r_0$. However, it is clear that these solutions can be extended to all of the positive axis. Thus Lemma 2.2 is proved for the case $\alpha \geqslant 1$, $2\beta + \alpha > 2$.

If $\alpha < 1$, $\alpha + \beta > 1$ and $\beta > \frac{1}{2}$, we proceed, at the initial stages, in a slightly different way. Let

$$\mu(r,\lambda) = i[\lambda^2 - V_L(r)]^{1/2},$$

and let us set

$$P(r,\lambda) = \begin{bmatrix} 1 & 1 \\ \mu(r,\lambda) & -\mu(r,\lambda) \end{bmatrix},$$

and $u = Pw$. Then the differential equation (2.20) is trans-

formed to

$$w' = (\Lambda + W)w,$$

where $\Lambda = \text{diag}[\mu, -\mu]$, and

$$W = \frac{1}{2}\frac{\mu'}{\mu}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} + O\left(\frac{1}{r^2}\right).$$

The proof now proceeds as before with the exception that μ'/μ replaces V_L in the computations. However, Levinson's technique now leads to solutions which are asymptotic to $\exp(\pm \int_{r_0}^{r} \mu)$, rather than $\exp(\pm i\lambda r)$. If we expand $(\lambda^2 - V_L)^{1/2}$ into a series we see we may write

$$\int_{r_0}^{r} \mu = i\lambda r + \gamma(r,\lambda),$$

where $\gamma(r,\lambda)$ converges to a finite number as $r \to \infty$ only if $\alpha + \beta > 1$ and $\beta > \frac{1}{2}$. This leads to the expansion (ii) in Lemma 2.2. Thus we shall consider Lemma 2.2 as proved.

[1]S. Agmon, Ann. Scuol. Norm. Sup. Pisa **2**, 151 (1975).
[2]V. Enss, Commun. Math. Phys. **61**, 285 (1978).
[3]J. Dollard and C. Friedman, Ann. Phys. **111**, 251 (1978).
[4]M. Ben-Artzi and A. Devinatz, J. Math. Phys. **20**, 594 (1979).
[5]M. Combescure and J. Ginibre, Ann. Inst. Henri Poincaré **24**, 17 (1976).
[6]M. Schechter, Lett. Math. Phys. **2**, 127 (1977).
[7]Reference 2.
[8]B. Bourgeois, thesis, University of Texas at Austin, 1979.
[9]M. Ben-Artzi and A. Devinatz, Errata, J. Math. Phys. **21**, 2472 (1980).

Properties of a covering space defined by Hawking

B. C. Haggman, G. W. Horndeski, and G. Mess
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 24 January 1980, accepted for publication 6 June 1980)

We investigate the topological properties of a covering space introduced by Hawking in 1967 to assist in his derivation of some singularity theorems.

1. INTRODUCTION

Let (M,g) be a Lorentzian manifold and let S be a slice in (M,g), i.e., a closed, connected, regular,[1] codimension one spacelike submanifold of M. Hawking[2] has defined a covering manifold M_H of M with projection π and claims that M_H is the largest covering manifold of M with the property that each component of the preimage of S is homeomorphic to S (see also p. 205 of Ref. 3). This claim is incorrect. In the next section we show that M_H is the largest covering manifold of M with the property that there exists at least one component of the preimage of S which is homeomorphic to S.

A slice N in a Lorentzian manifold $(\hat{M},\hat{g})$ is called a *partial Cauchy surface* in $(\hat{M},\hat{g})$ if every nonspacelike curve of class C^1 meets N in at most one point.

In the third section we show that if S is a slice in a time orientable spacetime (M,g), then each component of the preimage of S in the covering space M_H is a partial Cauchy surface in the spacetime (M_H,π^*g).

Due to the aforementioned results we shall be able to conclude that if S is a slice in a time orientable Lorentzian manifold (M,g), then there exists a covering space of (M,g) which admits a partial Cauchy surface which is homeomorphic to S. Consequently, Hawking's minor oversight concerning some of the properties of M_H will not affect his applications of M_H in the proofs of various singularity theorems.[2,3]

2. DEFINITION OF M_H AND SOME OF ITS TOPOLOGICAL PROPERTIES

In order to define M_H and establish some of its properties we do not require M to be a differentable manifold. Thus throughout this section we assume that M is a Hausdorff, connected, locally arcwise connected, and locally simply connected[4] topological space, and S is a connected, locally arcwise connected subspace of M.

Following Hawking,[2] we define $M_H := \{(p,[\lambda])|p\in M$ and $[\lambda]$ is an equivalence class of paths in M from S to p homotopic modulo S and $p\}$. If $(p,[\lambda])\in M_H$ and V is a simply connected neighborhood of p we define $V(p,[\lambda]) := \{(q,[\lambda\cdot\alpha])|q\in V$ and α is the unique (up to a homotopy) path class in V from p to $q\}$. The totality of sets $V(p,[\lambda])$ defines a basis for a topology on M_H in terms of which M_H is a Hausdorff, arcwise connected, and locally arcwise connected topological space. Let $\pi:M_H\to M$ be defined by $(p,[\lambda])\mapsto p$. It is easily seen that M_H is a covering space[5] of M with projection π. An important property of M_H is presented in

Proposition 2.1: *If $s\in S$ and $\tilde{s} := (s,[e_s])\in M_H$, where e_s is the constant path, then $\pi_*\Pi_1(M_H,\tilde{s}) = j_*\Pi_1(S,s)$, where*[6] *$j:S\to M$ is the natural injection and a lower star denotes the induced homomorphism between fundamental groups.*[7]

Proof: We first show that $j_*\Pi_1(S,s)\subset\pi_*\Pi_1(M_H,\tilde{s})$.

Let $\delta:I\to M$ ($I := [0,1]\subset\mathbb{R}$) be a loop starting at s and having its range contained in S. For every t with $0\leqslant t\leqslant 1$ we define $\delta_t: I\to M$ by $u\to\delta(tu)$. δ lifts to a map $\hat{\delta}: I\to M_H$ given by $t\to(\delta(t),[\delta_t])$. $\hat{\delta}$ will be a loop in M_H starting at $\tilde{s}$ provided $[\delta_0] = [\delta_1]$. But $\delta = \delta_1$ lies in S and so is homotopic modulo S and s to the constant loop δ_0. Thus $\hat{\delta}$ is a loop in M_H starting at $\tilde{s}$ and such that $\pi\circ\hat{\delta} = \delta$. Consequently, $j_*\Pi_1(S,s)\subset\pi_*\Pi_1(M_H,\tilde{s})$. Now for the converse.

If $\hat{\alpha}$ ($\hat{\alpha}(t) = (\alpha(t),[\lambda_t])$), is a loop in M_H starting at $\tilde{s}$ then $\pi\circ\hat{\alpha} = \alpha$ is a loop in M starting at s. Evidently $\hat{\alpha}$ provides one lift of α to M_H starting at $\tilde{s}$. A second lift of α to M_H starting at $\tilde{s}$ is given by $t\to(\alpha(t),[\alpha_t])$, where $\alpha_t(u) := \alpha(tu)$ for every t with $0\leqslant t\leqslant 1$. By the uniqueness of lifts starting at $\tilde{s}$ (cf., p. 151 of Ref. 5) we must have $[\lambda_t] = [\alpha_t]$ and so $[\alpha_0] = [\lambda_0] = [\lambda_1] = [\alpha_1]$ (where, recall that, the homotopy is modulo S and s). Thus $\alpha(=\alpha_1)$ must be homotopic modulo S and s to the constant path e_s, and hence α is loop homotopic to a loop whose range lies in S. (The loop and loop homotopy in question can be built using the homotopy from e_s to α modulo S and s.) As a result $\pi_*\Pi_1(M_H,\tilde{s})\subset j_*\Pi_1(S,s)$. □

Our next proposition concerns the relationship between S and its preimage in M_H.

Proposition 2.2: *Let $s\in S$ and set $\tilde{s} := (s,[e_s])$. The path component $\tilde{S}$ of $\Sigma := \pi^{-1}(S)$ which contains $\tilde{s}$ is homeomorphic to S under π. If λ is a path from S to s such that $[\lambda]\neq[e_s]$ then the path component of Σ passing through $(s,[\lambda])$ is the total space of a τ-sheeted covering space of S where τ is the number of distinct homotopy classes of the form $[\lambda\cdot\alpha]$ where α is a loop in M which determines an element of $j_*\Pi_1(S,s)$.*

Proof: Let Σ' denote a path component of Σ and let π' denote the unique map which makes the following diagram commutative:

$$\begin{array}{ccc} \Sigma' & \xrightarrow{\ j'\ } & M_H \\ {\scriptstyle\pi'}\downarrow & & \downarrow{\scriptstyle\pi} \\ S & \xrightarrow{\ j\ } & M. \end{array} \tag{2.1}$$

It is well known that Σ' is a covering space of S with projection π' (c.f. page 150 of Ref. 5). We shall now determine when π' is injective and hence a homeomorphism.

Suppose that $\hat{\alpha}$ is a path in M_H which goes from $(s,[\mu])$ to $(s,[\nu])$ and has range contained in Σ'. Upon projecting $\hat{\alpha}$ to M with π we obtain a loop α in M at s whose range lies in S. We lift α to the path $t\mapsto(\alpha(t),[\mu\cdot\alpha_t])$, where $\alpha_t(u) := \alpha(tu)$ for every t with $0\leqslant t\leqslant 1$. Due to the uniqueness of lifts of α

 0022-2488/80/092412-05$1.00

starting at $(s,[\mu])$ we must have $\hat{\alpha}(1) = (\alpha(1), [\mu\cdot\alpha_1])$ and hence $[\mu\cdot\alpha] = [\nu]$. Thus π' will be injective if and only if $[\mu\cdot\alpha] = [\mu]$ for all loops α which determine elements of $j_*\Pi_1(S,s)$. Consequently, the path component $\widetilde{S}$ of Σ which contains $(s,[e_s])$ is homeomorphic to S under π. However, if $[\mu]\neq[e_s]$, then $[\mu\cdot\alpha]$ need not equal $[\mu]$ for all loops α. Thus in this case the component Σ' of Σ passing through $(s[\mu])$ will be a τ-sheeted covering space where τ is the number of distinct homotopy classes of the form $[\mu\cdot\alpha]$, where α is a loop in M which determines an element of $j_*\Pi_1(S,s)$. □

At this point one should note that the definition of $\widetilde{S}$ given in Proposition 2.2 is actually independent of our choice of $s\in S$; i.e., if ζ is some other point in S then $(\zeta,[e_\zeta])\in\widetilde{S}$.

In general the covering space M_H will be irregular (see pages 163–164 of Ref. 5). However, in those instances for which M_H is regular we can use Proposition 2.2 to conclude that each component of the preimage of S is homeomorphic to S.

Propositions 2.1 and 2.2 essentially represent the essence of the covering space M_H, in the sense that all of the remaining results concerning M_H which we shall establish require only those propeties of M_H given in these propositions.

The relationship between the fundamental groups of $\widetilde{S}$ and M_H is provided by

Proposition 2.3: *If $s\in S$ and $\tilde{s} := (s,[e_s])\in M_H$ then $\tilde{j}_*\Pi_1(\widetilde{S},\tilde{s}) = \Pi_1(M_H,\tilde{s})$, where $\widetilde{S}$ is the component of the preimage of S which passes through $\tilde{s}$ and $\tilde{j}:\widetilde{S}\hookrightarrow M_H$.*

Proof: Due to Eq. (2.1) with Σ' replaced by $\widetilde{S}$ we obtain the following commutative diagram of groups and homomorphisms

$$\begin{array}{ccc} \Pi_1(\widetilde{S},\tilde{s}) & \xrightarrow{\tilde{j}_*} & \Pi_1(M_H,\tilde{s}) \\ \downarrow{\scriptstyle \pi'_*} & & \downarrow{\scriptstyle \pi_*} \\ \Pi_1(S,s) & \xrightarrow{j_*} & \Pi_1(M,s). \end{array}$$

It is apparent that $\tilde{j}_*\Pi_1(\widetilde{S},\tilde{s})\subset\Pi_1(M_H,\tilde{s})$, and so we must prove that $\Pi_1(M_H,\tilde{s})\subset\tilde{j}_*\Pi_1(\widetilde{S},\tilde{s})$. To that end let $\langle\hat{\alpha}\rangle\in\Pi_1(M_H,\tilde{s})$. Due to Proposition 2.1 we know that there exists $\langle\alpha\rangle\in\Pi_1(S,s)$ which is such that $j_*\langle\alpha\rangle = \pi_*\langle\hat{\alpha}\rangle$. Since $\widetilde{S}$ is homeomorphic to S, π'_* is an isomorphism, and hence there exists $\langle\beta\rangle\in\Pi_1(\widetilde{S},\tilde{s})$ for which $\pi'_*\langle\beta\rangle = \langle\alpha\rangle$. The commutativity of the above diagram and the fact that π_* is a monomorphism (see p. 154 of Ref. 5) imply that $\tilde{j}_*\langle\beta\rangle = \langle\hat{\alpha}\rangle$. □

The above corollary will be employed in the next section to show that if S is a slice in a time orientable spacetime then each component of the preimage of S separates M_H.

The covering space M_H is uniquely characterized (up to an isomorphism) by

Proposition 2.4: *M_H is the largest covering space of M which is such that there exists one component of the preimage of S which is homeomorphic to S under the covering map.*

Proof: Let $\hat{M}$ be a covering space of M with projection $\hat{\pi}$ which is such that there exists one component C of the preimage of S which is homeomorphic to S under $\hat{\pi}$. In order to prove the proposition we must show that M_H is a covering space of $\hat{M}$. To show this it suffices[8] to demonstrate that there exists a homomorphism from M_H into $\hat{M}$; i.e., a continuous map $p: M_H\to\hat{M}$ such that $\hat{\pi}\circ p = \pi$. This will be the case if $\pi_*\Pi_1(M_H,\tilde{s})\subset\hat{\pi}_*\Pi_1(\hat{M},\hat{s})$, where $s\in S$, $\tilde{s} := (s,[e_s])$ and $\hat{s}$ is the unique point in C for which $\hat{\pi}(\hat{s}) = s$. However, this is obvious since Proposition 2.1 tells us that $\pi_*\Pi_1(M_H,\tilde{s}) = j_*\Pi_1(S,s)$ and C is homeomorphic to S under $\hat{\pi}$. □

To illustrate the above theory we have the following

Example 2.1: Let $M := \mathbb{R}^2\setminus\{(0,0),(2,0)\}$ and let S denote the unit circle centered at (0,0) with subset topology. Evidently M and S satisfy our topological assumptions. We choose $s = (1,0)$ and let λ be the polygonal path from s to s whose image consists of the three line segments joining the points s, $(3,-1)$, $(3,1)$ and s. If we take α to be the loop in M based at s, which traverses S counterclockwise one time then the set $\{[\lambda\cdot\alpha^n]\}_{n\in\mathbb{Z}}$ ($\mathbb{Z}$: = the set of integers) consists of distinct elements. Thus the component Σ' of $\pi^{-1}(S)$ which contains $(s,[\lambda])$ as a countable infinity of sheets. It turns out that Σ' is homeomorphic to $\mathbb{R}$, and the only path component of $\pi^{-1}(S)$ which is homeomorphic to S is the one which contains $(s,[e_s])$.

In passing it should be noted that M admits many Lorentzian metrics in terms of which S is a slice—in fact, a partial Cauchy surface. One such metric is provided by $ds^2 = -dr^2 + r^2d\theta^2$ where r and θ are polar coordinates. □

This completes our brief study of some of the properties of M_H, and the preimage of S in M_H, which can be established under the mild topological restrictions imposed at the outset of this section. In the next section we shall establish a few more topological properties of M_H under the assumption that S is a slice in a time orientable Lorentzian manifold.

3. PARTIAL CAUCHY SURFACES

In this section we shall confine our attention to an n-dimensional Lorentzian manifold, $L_n = (M,g)$, where M is assumed to be a connected Hausdorff, n-dimensional, C^∞ manifold, and g is a class C^0 Lorentzian metric on M.

Let S be a slice in the Lorentzian manifold $L_n = (M,g)$. We make the covering space M_H determined by M and S into a connected, Hausdorff, n-dimensional, C^∞ manifold by demanding that the covering projection $\pi: M_H\to M$ be a local diffeomorphism. In term of this manifold structure we find that $\hat{L}_n := (M_H,\pi^*g)$ is a Lorentzian manifold and each component of the preimage of S is a slice in $\hat{L}_n$. If L_n is time orientable then so is $\hat{L}_n$.

The primary purpose of this section is to prove

Theorem 3.1: *If S is a slice in a time orientable Lorentzian manifold $L_n = (M,g)$ then each component of the preimage of S in M_H is a partial Cauchy surface in $\hat{L}_n = (M_H,\pi^*g)$.*

This result is an immediate consequence of the following two propositions which are of interest in their own right.

Proposition 3.1: *If S is a slice in a time orientable Lorentzian manifold $L_n = (M,g)$ which separates[9] M then S is a partial Cauchy surface in M.*

Proposition 3.2: *If S is a slice in a time orientable Lorent-*

zian manifold $L_n = (M,g)$, then each component of the preimage of S in M_H separates M_H.

Remark: Proposition 3.1 seems to be known by many relativists and is implicit in the work of Geroch[10]; however, we have not seen a published proof of this result. □

We shall now present the proofs of Propositions 3.1 and 3.2.

Proof[11] *of Proposition* 3.1: Since L_n is time orientable there exists a smooth, future pointing, timelike vector field Z on M. Let $\Phi: D \subset \mathbb{R}\times M \to M$ denote the flow of Z; i.e., $\Phi(t,p) := \gamma_p(t)$ where γ_p is the maximal integral curve of Z starting at p. Φ is a smooth function. We set $\theta := \Phi_0(1\times j)$, where $1:\mathbb{R}\to\mathbb{R}$ is the identity map and $j:S\hookrightarrow M$. Evidently θ is differentiable as a map from an open subset of $\mathbb{R}\times S$ into M, and it can be shown that since S is a regular submanifold of M there exists a connected neighborhood U of $\{0\}\times S$ in $\mathbb{R}\times S$ which is such that $U\subset \mathrm{dom}\Phi, \theta|_U$ is a diffeomorphism, and $\{0\}\times S$ is a deformation retract[12] of U.

Since S separates M, $M\setminus S$ consists of two (connected) components each of which has S as its topological boundary in M (see p. 107 of Ref. 13). Clearly $\theta(U)$ is a neighborhood of S in M and hence $\theta(U)$ meets each component of $M\setminus S$. Let $U^+ := \{(t,p)\in U \mid t>0\}$ and $U^- := (t,p)\in U \mid t<0\}$. U^+ and U^- are connected and $\theta(U^+)$ and $\theta(U^-)$ must lie in different components of $M\setminus S$. We let C^+ (C^-, resp.) denote the component of $M\setminus S$ which contains $\theta(U^+)$ ($\theta(U^-)$, resp.).

Employing the diffeomorphism θ we can show that if $s\in S$ then there exists a chart (W,x) of M at s with connected domain W which is such that:[14]

(i) $S\cap W = \{p\in W \mid x^0(p) = 0\}$;
(ii) if $x^0(p) > 0$, then $p\in W^+ := W\cap C^+$;
(iii) if $x^0(p) < 0$, then $p\in W^- := W\cap C^-$;
(iv) $Z = \dfrac{\partial}{\partial x^0}$ on W.

The chart (W,x) is said to be *adapted to Z at s.*

Let $c: J\subset\mathbb{R}\to M$ be a nonspacelike curve. c is either past pointing or future pointing. Without loss of generality we may assume that c is future pointing. Suppose that $c^{-1}(S)$ contains at least two points, a_1 and a_2, with $a_1 < a_2$. We set $k := c|_{[a_1,a_2]}$, and claim that $k^{-1}(S)$ is a finite set. This follows trivially from the fact that c is transverse to S; however, we proceed directly. Let $a\in k^{-1}(S)$ and set $s := k(a)$. If (W,x) is a chart of M at s which is adapted to Z then we have

$$\dot{c}(a) = \left.\frac{dc^i}{dt}\right|_a \left.\frac{\partial}{\partial x^i}\right|_s,$$

where $c^i := x^i\circ c$, and t is the standard chart on $\mathbb{R}$. Since $\dot{c}(a)$ is future pointing, we must have $dc^0/dt|_a > 0$, and hence c^0 is strictly increasing on a neighborhood of a. Thus there exists $\delta\in\mathbb{R}^+$ which is such that $c((-\delta+a,a))\subset W^-\subset C^-$ and $c((a,a+\delta))\subset W^+\subset C^+$. Consequently we can separate each element of $k^{-1}(S)$ by disjoint open sets. Since the interval $[a_1,a_2]$ is compact we can now conclude that the set $k^{-1}(S)$ is finite.

The above work shows that if $c^{-1}(S)$ contains at least two points then we can choose $a_1,a_2\in c^{-1}(S)$, with $a_1 < a_2$, so that $k := c|_{[a_1,a_2]}$ is such that $k^{-1}(S) = \{a_1,a_2\}$. Let (W,x) and (V,y) be charts of M at $s_1 := c(a_1)$ and $s_2 := c(a_2)$, respectively, which are adapted to Z. By the previous argument there exists $b_1,b_2\in\mathbb{R}$, such that $a_1 < b_1 < b_2 < a_2$, $c(b_1)\in W^+\subset C^+$, and $c(b_2)\in V^-\subset C^-$. We now have the contradiction that $c|_{[b_1,b_2]}$ is a continuous curve lying in $M\setminus S$ and joining points in different components of $M\setminus S$. This contradiction implies that if c is a nonspacelike curve then $c^{-1}(S)$ can contain at most one point and hence S is a partial Cauchy surface. □

Our proof of Proposition 3.2 will proceed by means of three lemmas. We begin with the following definition: a submanifold N of a manifold M will be said to be two-sided in M if N is a regular, connected, codimension one submanifold of M and there exists a smooth vector field defined on a neighborhood of N and nowhere tangent to N. E.g., if $L_n = (M,g)$ is time orientable then every slice in L_n is two-sided in M.

Employing an argument similar to the one presented in the proof of Proposition 3.1 to construct the map θ we can show that if N is a closed, two-sided submanifold of M then there exists a connected neighborhood V of S which is such that S is a deformation retract of V and S disconnects V into two connected components. We shall now exploit this fact to prove

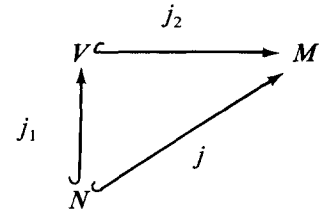
Lemma 3.1: *Let N be a closed, two-sided submanifold of a connected Hausdorff manifold M. If $j: N\hookrightarrow M$ and for some point $p\in N$, $j_*\Pi_1(N,p) = \Pi_1(M,p)$, then N separates M into two connected components.*

Proof: Let V be a connected neighborhood of N such that N is a deformation retract of V and $V\setminus N$ is disconnected. $W := M\setminus N$ is open and $V\cup W = M$. We want to prove that W is disconnected and has two components. This will be the case if we can show that $H_0(W)$ is isomorphic to $\mathbb{Z}\oplus\mathbb{Z}$, where for every $\alpha = 0,1,2,\ldots, H_\alpha(W)$ denotes the αth homology group[15] of W. This fact can easily be established through use of the Mayer–Vietoris theorem (see p. 74 of Ref. 15) which gives us the following exact sequence of abelian groups and homomorphisms:

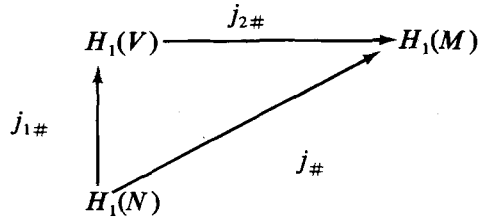
$$\cdots\to H_1(V)\oplus H_1(W) \xrightarrow{\Phi} H_1(M) \xrightarrow{\Delta} H_0(V\cap W) \xrightarrow{\psi}$$
$$H_0(V)\oplus H_0(W) \xrightarrow{\phi} H_0(M)\to 0. \tag{3.1}$$

The precise definition of the homomorphisms Δ,ψ, and ϕ is not required for what follows, and Φ will be defined momentarily.

Since $N\subset V\subset M$, we have the following commutative diagram of topological spaces and continuous maps:



This in turn gives rise to the commutative diagram of abelian groups and homomorphisms given below



where the # denotes the induced map on the homology groups. Since N is a deformation retract of V, $j_{1\#}$ is an isomorphism and hence $j_\#$ and $j_{2\#}$ have the same image.

The homomorphism $\Phi: H_1(V) \oplus H_1(W) \to H_1(M)$ is defined by $\Phi(\alpha \oplus \beta) := j_{2\#}(\alpha) + j_{3\#}(\beta)$, where $j_3: W \hookrightarrow M$. Since we are assuming that $j_* \Pi_1(N,p) = \Pi_1(M,p)$ we can use the Hurewitz homomorphism theorem (see p. 51 of Ref. 15) to conclude that $j_\# H_1(N) = H_1(M)$. Due to our previous work we know that $j_\#$ and $j_{2\#}$ have the same image and hence Φ is an epimorphism.

By the exactness of the sequence presented in Eq. (3.1) we see that Φ is an epimorphism iff $H_1(M) = \mathrm{Im}(\Phi) = \ker\Delta$ iff $0 = \mathrm{Im}\Delta = \ker\psi$ iff ψ is a monomorphism. Thus Eq. (3.1) gives rise to the following short exact sequence:

$$0 \to H_0(V \cap W) \xrightarrow{\psi} H_0(V) \oplus H_0(W) \xrightarrow{\phi} H_0(M) \to 0. \tag{3.2}$$

Since $H_0(V) \simeq \mathbb{Z}$, $H_0(V \cap W) \simeq \mathbb{Z} \oplus \mathbb{Z}$, and $H_0(M) \simeq \mathbb{Z}$, it is now an elementary matter to prove that $H_0(W) \simeq \mathbb{Z} \oplus \mathbb{Z}$ and hence N separates M into two (connected) components. □

If S is a two-sided slice in $L_n = (M,g)$ then it is clear that each component of the preimage of S in $\hat{L}_n = (M_H, \pi^* g)$ is a two-sided slice in $\hat{L}_n$. Thus we can now use Proposition 2.3 and Lemma 3.1 to deduce

Lemma 3.2: *Let $s \in S$ and set $\tilde{s} := (s, [e_s])$. If S is a two-sided slice in $L_n = (M,g)$ then the component $\tilde{S}$ of the preimage of S which passes through $\tilde{s}$ separates M_H into two connected components.*

Proposition 3.2 now follows from

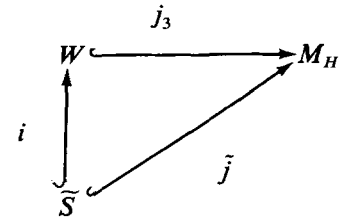
Lemma 3.3: *If S is a two-sided slice in $L_n = (M,g)$ then each component of the preimage of S separates M_H into two connected components.*

Proof: Let $\tilde{S}$ be the component of $\pi^{-1}(S)$ which passes through $\tilde{s} := (s, [e_s])$ where $s \in S$. Due to Lemma 3.2 and Proposition 2.3 we know that $\tilde{S}$ separates M_H into two connected components and $\tilde{j}_* \Pi_1(\tilde{S},\tilde{s}) = \Pi_1(M_H,\tilde{s})$, where $\tilde{j}: \tilde{S} \hookrightarrow M_H$. Due to the Hurewicz homomorphism we must also have $\tilde{j}_\# H_1(\tilde{S}) = H_1(M_H)$.

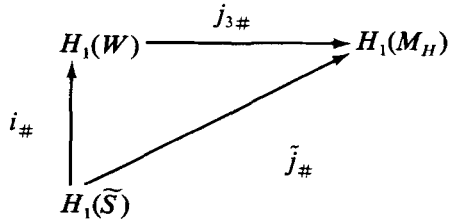
Let N be some other component of $\pi^{-1}(S)$. As mentioned previously N must be a two-sided slice in M_H and thus we can find a connected neighborhood V of N which is such that N is a deformation retract of V, $V \setminus N$ is disconnected with two components. We define $W := M_H / N$. W is open in M_H and $V \cup W = M_H$.

In order to complete the proof of this lemma we would like to apply the Mayer–Vietoris theorem argument used in the proof of Lemma 3.1 to the present decomposition of M_H. The key step in this argument involves showing that the map Φ appearing in Eq. (3.1) is an epimorphism. This can be proved as follows.

We have the following commutative diagram of topological spaces and continuous maps:



which in turn gives rise to the commutative diagram of abelian groups and homorphisms given below:



(3.3)

As mentioned above $\tilde{j}_\#$ is an epimorphism and hence diagram (3.3) implies that $j_{3\#}$ must be an epimorphism.

Now if $\alpha \oplus \beta \in H_1(V) \oplus H_1(W)$, then $\Phi(\alpha \oplus \beta) = j_{2\#}(\alpha) + j_{3\#}(\beta)$, where $j_2: V \hookrightarrow M_H$. Consequently, $\Phi: H_1(V) \oplus H_1(W) \to H_1(M_H)$ is an epimorphism.

Thus we can now use the Mayer–Vietoris sequence given in Eq. (3.1) to conclude that N separates M_H into two connected components as required. □

Lemma 3.3 tells us that if S is a two-sided slice in $L_n = (M,g)$ then each component of its preimage separates M_H. In view of Lemma 3.1 this leads one to suspect that if N is a component of the preimage of S then $j_* \Pi_1(N,p) = \Pi_1(M_H,p)$, where $p \in N$ and $j: N \hookrightarrow M_H$. However, in general this is not the case, as can be seen from Example 2.1.

We shall now conclude this section with a few remarks concerning a covering space constructed by Geroch.[10]

Let S be a two-sided slice in $L_n = (M,g)$, and let Z be a smooth vector field defined on a neighborhood of S and nowhere tangent to S. If $c: S^1 \to M$ is a smooth map transverse to S, θ is the "global" polar coordinate on S^1, and $c(\tau) \in S$, then we define

$$i(c,\tau) = \begin{cases} +1 \text{ if } dc\left(\dfrac{\partial}{\partial\theta}\Big|_\tau\right) \text{ and } Z[c(\tau)] \text{ point to the} \\ \text{same side of } dj(T_{c(\tau)}S) \text{ in } T_{c(\tau)}M, \text{ where} \\ j: S \hookrightarrow M; \text{and } -1 \text{ otherwise} \end{cases}$$

(where d denotes the derivative of a map). The *index of c with respect to the pair* (S,Z) is defined by

$$i(c,(S,Z)) := \sum_{\tau \in c^{-1}(S)} i(c,\tau).$$

If $s \in S$ and $\langle c \rangle \in \Pi_1(M,s)$ then we can regard c as a map of S^1 into M. We now define the index of $\langle c \rangle$ with respect to (S,Z) by $i(\langle c \rangle,(S,Z)) := i(k,(S,Z))$, where $k: S^1 \to M$ is any smooth map which is transverse to S and homotopic to c. $i(\langle c \rangle,(S,Z))$ is well defined due to standard results in differential topology (see, e.g., p. 132 of Ref. 13). Let $G = \{\langle c \rangle \in \Pi_1(M,s) \mid i(\langle c \rangle,(S,Z)) = 0\}$. G is a normal subgroup of $\Pi_1(M,s)$ whose definition is independent of our original choice of Z (so long as Z is nowhere tangent to S). Geroch's covering space M_G is defined (up to isomorphism)

to be the unique covering space of M with the property that $p_* \Pi_1(M_G,\tilde{s}) = G$, where $p:M_G \to M$ is the covering map. Since G is a normal subgroup of $\Pi_1(M,s)$, M_G is a regular covering space of M. It can be shown that M_G is such that each component of the preimage of S is homeomorphic to S and separates M_G. Thus we can now use Proposition 2.4 to conclude that M_H covers M_G, and Proposition 3.1 to conclude that if L_n is time orientable then each component of the preimage of S is a partial Cauchy surface in (M_G,p^*g). Lastly we would like to point out that Geroch has proved that M_G is the smallest covering space of M with the property that each component of the preimage of S separates the covering space.

ACKNOWLEDGMENTS

We wish to thank Professor P. N. Hoffman and Professor R. Geroch for their discussions on the topics treated in this paper. We also wish to thank the Natural Sciences and Engineering Research Council of Canada for financial support.

[1] A submanifold S of a manifold M is said to be a regular submanifold if the subset topology on S is equal to its manifold topology.

[2] S. W. Hawking, Proc. Roy. Soc. Lond. A **300**, 187 (1967).

[3] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (*Cambridge U. P., Cambridge*, 1973).

[4] By M being locally simply connected we mean that given any point $p \in M$ and a neighborhood V of p there exists a simply connected neighborhood U of p with $U \subset V$.

[5] For a definition of a covering space and some of its topological properties see W. S. Massey, *Algebraic Topology: An Introduction* (Harcourt, Brace and World, New York, 1967).

[6] If X is a subspace of the topological space Y, and $j:X \to Y$ is the natural injection, then we shall, in the future, convey this property of j by writing $j:X \hookrightarrow Y$.

[7] See Chapter 2 of Ref. 5 for the precise definitions of the fundamental group of a topological space and of the induced homomorphism between fundamental groups.

[8] The covering space theory used in this proof can be found on pages 156 and 160 of Ref. 5.

[9] S is said to separate M if $M \setminus S$ is disconnected.

[10] R. P. Geroch, J. Math. Phys. **8**, 782 (1967).

[11] For the details omitted here see B. C. Haggman, "Aspects of Fibre Bundles in General Relativity" M. Math. thesis (unpublished), University of Waterloo, Waterloo, Ontario, 1979.

[12] See p. 66 of Ref. 5 for the definition of "deformation retract."

[13] M. W. Hirsch, *Differential Topology* (Springer, New York, 1976).

[14] The coordinate functions of the chart x are $x^0,x^1,\ldots,x^{n-1}$.

[15] For the homology theory used in this proof see M. J. Greenberg, *Lectures on Algebraic Topology* (Benjamin, Reading, Mass., 1967).

Relationships between the group-theoretic and soliton-theoretic techniques for generating stationary axisymmetric gravitational solutions [a)]

Christopher M. Cosgrove [b)]
Department of Physics, Montana State University, Bozeman, Montana 59717

(Received 22 January 1980, accepted for publication 16 May 1980)

We investigate the precise interrelationships between several recently developed solution-generating techniques capable of generating asymptotically flat gravitational solutions with arbitrary multipole parameters. The transformations we study in detail here are the Lie groups **Q** and $\widetilde{\mathbf{Q}}$ of Cosgrove, the Hoenselaers–Kinnersley–Xanthopoulos (HKX) transformations and their SL(2) tensor generalizations, the Neugebauer–Kramer discrete mapping, the Neugebauer Bäcklund transformations I_1 and I_2, the Harrison Bäcklund transformation, and the Belinsky–Zakharov (BZ) one- and two-soliton transformations. Two particular results, among many reported here, are that the BZ soliton transformations are essentially equivalent to Harrison transformations and that the generalized HKX transformation may be deduced as a confluent double soliton transformation. Explicit algebraic expressions are given for the transforms of the Kinnersley–Chitre generating functions under all of the above transformations. In less detail, we also study the Kinnersley–Chitre β transformations, the non-null HKX transformations, and the Hilbert problems proposed independently by Belinsky and Zakharov, and Hauser and Ernst. In conclusion, we describe the nature of the exact solutions constructible in a finite number of steps with the available methods.

1. INTRODUCTION AND PRELIMINARY

Until recently, the number of exact solutions of the stationary axisymmetric vacuum gravitational field equations which could represent the gravitational field of a spinning mass formed a very short list indeed. This situation has now abruptly changed. In fact, several authors have recently proposed widely differing solution generating techniques which they claim are capable of generating, in principle, after infinitely many iterations, the entire class of asymptotically flat solutions from the static Weyl solutions, or even from just flat space.[1] It is a remarkable fact that we now have available many more transformations than we need to generate arbitrarily large classes of solutions. At this stage, the interrelationships between most of these various approaches are rather mysterious. This paper addresses the interesting problem of the interrelationships and attempts to catalog the exact solutions which can be constructed with the available methods.

In a series of papers,[2–8] Kinnersley and coworkers worked out an efficient formalism for the infinite dimensional Geroch group[9] and found two independent infinite dimensional subgroups which preserve asymptotic flatness. These will be called here the Kinnersley–Chitre $\beta^{(k)}$ groups[5] and the Hoenselaers–Kinnersley–Xanthopoulos (HKX) groups.[7] Cosgrove[10–12] sought groups outside the Geroch group and found one, **Q**, which preserves asymptotic flatness. This was used to construct the three-parameter Cosgrove–Tomimatsu–Sato (CTS) solution[13–16] with continuous deformation parameter δ and may also be used to enlarge this solution to an arbitrarily large number of parameters.[17] Independently of the above work on continuous transformation groups, Bäcklund transformations have been found by Harrison,[18] Neugebauer[19] and Belinsky and Zakharov,[20] and so a powerful theory, already well developed for the sine-Gordon and certain other nonlinear equations, is now applicable to the stationary axisymmetric problem in general relativity. In addition, Belinsky and Zakharov[20] and Hauser and Ernst[21] have reformulated the nonlinear problem in terms of a linear integral equation, the former working from the "inverse scattering" standpoint, the latter from the Kinnersley–Chitre formalism for the Geroch group.

The aim of this paper is to explore the relationships between these widely different solution generating techniques. The principal tool in this investigation will be the SL(2,R) tensor generating functions, $F_{AB}(t)$ and $G_{AB}(s,t)$, of Kinnersley and Chitre[22] as these will be shown to transform in a simple algebraic manner under most of the transformations we wish to consider. Reference 11 already contains a number of results about the interrelationships between Cosgrove's groups, **Q** and $\widetilde{\mathbf{Q}}$, Neugebauer's and Harrison's Bäcklund transformations and the HKX transformations. These results were obtained with relative ease using Neugebauer's commutation theorem[19] and did not need the generating functions. But to go further, we are forced to use the generating functions and some rather unexpected elegant relationships will emerge. Also, it is instructive to see Neugebauer's commutation theorem and some of the other results in Ref. 11 written out in terms of generating functions.

In Ref. 11, the following relationships between and properties of solution generating techniques were reported:

(i) Cosgrove's continuous groups, **Q** and $\widetilde{\mathbf{Q}}$, are essentially the same as Neugebauer's Bäcklund transformations, I_1 and I_2, respectively;

[a)]Research supported by National Science Foundation Grant PHY 79-19565.

[b)]Present address: Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125.

 0022-2488/80/092417-31$1.00

(ii) Harrison's Bäcklund transformation can be decomposed in the form of products, I_1I_2 and I_2I_1;

(iii) Using Neugebauer's commutation theorem, a composition theorem for Harrison transformations is given so that any number of successive Harrison transformations can be calculated by algebraic methods alone once the differential equation for the first transformation has been solved;

(iv) An even number of successive Harrison transformations (and I_1 alone) preserves asymptotic flatness (proofs appear here in Appendix B);

(v) With the aid of two examples, a qualitative similarity between the double Harrison transformation and the double HKX rank-zero transformation was exhibited. The double Harrison transform of any Weyl solution may always be obtained as the double HKX transform of a different Weyl solution, and the former transformation applied to the CTS solution gives the same result as the latter applied to a contiguous $\delta \to \delta \pm 1$ CTS solution.[17]

In this paper, the following further new results are reported:

(i) The transforms of the generating functions under both $\mathbf{Q}$ and $\widetilde{\mathbf{Q}}$ are calculated, the latter taking the form of a tensor equation;

(ii) Explicit formulas for the pseudopotentials α of Neugebauer and q of Harrison, which are defined by Riccati differential equations, and the eigenfunctions ψ and χ of Neugebauer and Kramer are given in terms of F_{AB};

(iii) The $\mu \to \mu^{(s)}$ transformation of Maison[23] is shown to be a subgroup of $\widetilde{\mathbf{Q}}$ or I_2;

(iv) Neugebauer's commutation theorem and results (i), (ii), and (iii) in the previous paragraph are written out in terms of generating functions;

(v) Explicit formulas for the transforms of F_{AB} and G_{AB} under the discrete Neugebauer–Kramer mapping[24] (I) are calculated (an unexpected result);

(vi) The Belinsky–Zakharov (BZ) one-soliton transformation[20] is shown to be equivalent to a product of (I) and a Harrison transformation in either order. The transform of the generating function F_{AB} is an SL(2) tensor equation. These transformations commute and are self-dual[25];

(vii) The transform of the generating function under the BZ two-soliton transformation, equivalent to a product of two one-solitons or of two Harrison transformations, is written out. The case where the two poles or "centers"[26] coincide is found to be equivalent to a generalization of the HKX rank-zero transformation[7] in which the γ_{22}'s are replaced by arbitrary null γ_{AB}'s.[27] A general commutation theorem is given for these generalized HKX transformations, which also preserve asymptotic flatness. The precise relationship between the original HKX rank-zero transformation and the double Harrison transformation is given;

(viii) Formulas for the transforms of the metric coefficient $e^{2\gamma}$ (see (1.1) below) under all of the above transformations are calculated;

(ix) The following transformations are in order of increasing generality: original HKX (two parameters), generalized HKX (three parameters), double Harrison or BZ two-soliton (four parameters), four-fold Neugebauer (eight parameters, restricted to six for asymptotic flatness preservation). A solution generated from the Weyl solutions or the CTS solution by any finite combination of the above transformations, including limits or confluences, can be obtained by iterating the original HKX transformations alone (including confluences such as the rank-one, rank-two, etc.) and even these are more general than needed. Also, all asymptotically flat solutions which have so far been constructed from the $\beta^{(k)}$ transformations of Ref. 5 and the linear integral equations[20,21] can be obtained with the above transformations acting on the Weyl solutions or special subclasses such as the Zipoy–Voorhees (ZV) solutions[28];

(x) The previous result shows that a catalog of all asymptotically flat solutions which can be generated by a finite number of iterations of the presently available methods would consist of two classes characterized by their "seed" or beginning solutions. The seed solutions are the Weyl solutions and the CTS solutions. The two classes overlap only when the deformation parameter δ in the latter class is an integer or when the CTS angular momentum parameter q is 0 or ± 1. Of course, the uniqueness conjectures imply that the latter class is necessarily contained in the former class in the limit of infinitely many iterations.

A number of other interesting problems will be treated in less detail. The non-null generalized HKX transformations are intimately related to the $\beta^{(k)}$ groups of Kinnersley and Chitre. These are both given in infinitesimal form and the finite exponentiated forms do not appear to be expressible by a simple algebraic formula involving the generating functions.[29,30] Nevertheless, a few interesting and potentially important commutation relations with other transformations can be deduced. An isolated case where Hoenselaers, Kinnersley and Xanthopoulos were successful in exponentiating the non-null HKX transformation is shown to be a fortunate instance where the formula for the null case works.

Another problem only briefly discussed here is the relationship between the integral equation of Hauser and Ernst,[21] deduced from the Kinnersley–Chitre formalism[3] for the Geroch group, and the integral equation of Belinsky and Zakharov[20] for the "non-soliton" part of the inverse scattering transform. By expressing Belinsky and Zakharov's variables in terms of the generating functions, the qualitative similarity between the two integral equation is immediately apparent, but the direct conversion of one to the other is rather awkward for several reasons. In a more recent paper,[21] Hauser and Ernst rederive their integral equation from the standpoint of a homogeneous Hilbert (Riemann) problem and this goes much of the way in relating the inverse scattering technique to the group-theoretic technique.

Our notation follows Kinnersley[2] except in one minor detail. The metric is

$$ds^2 = f(dt - \omega d\phi)^2 - f^{-1}\{e^{2\gamma}(d\rho^2 + dz^2) + \rho^2 d\phi^2\}, \tag{1.1}$$

where (ρ, z, ϕ, t) are canonical cylindrical coordinates and time, and f, ω and γ depend on ρ and z only. We define an Ernst potential[31] by

$$\mathscr{E} = f + i\psi, \quad \mathscr{E}^* = f - i\psi, \quad \nabla\psi = \rho^{-1} f^2 \tilde{\nabla}\omega, \tag{1.2a,b,c}$$

where $\nabla = (\partial/\partial\rho, \partial/\partial z)$, $\tilde{\nabla} = (\partial/\partial z, -\partial/\partial\rho)$. We next adopt the SL(2,R) tensor notation of Kinnersley[32] and define a symmetric tensor f_{AB} with components,

$$f_{11} = f, \quad f_{12} = f_{21} = -f\omega, \quad f_{22} = f\omega^2 - \rho^2 f^{-1}, \quad (1.3)$$

satisfying $f_{XA}f^{X}{}_{B} = -\rho^2\epsilon_{AB}$. Einstein's equations[2] imply the existence of a potential ψ_{AB} defined by

$$\nabla\psi_{AB} = -\rho^{-1}f_{A}{}^{X}\tilde{\nabla}f_{XB}, \quad (1.4)$$

with the property, $\psi_{AB} = \psi_{(AB)} + z\epsilon_{AB}$. From this, construct a tensor version of the Ernst potential,

$$H_{AB} = f_{AB} + i\psi_{AB}, \quad H^{*}_{AB} = f_{AB} - i\psi_{AB}, \quad (1.5)$$

which has the component, $H_{11} = f - i\psi = \mathscr{E}^{*}$.

Once H_{AB} has been constructed, a generating function $F_{AB}(t)$, which depends on the nonignorable coordinates, ρ and z, as well as a complex variable t, can be constructed by several methods, which are outlined in Appendix A. The coefficients in the power series expansion in t of $F_{AB}(t)$ are the Kinnersley–Chitre potentials $H^{(n)}_{AB}$, $n = 0,1,2,...$, but we will have little use for these potentials individually. Another generating function $G_{AB}(s,t)$ whose coefficients are the two-index potentials $N^{(m,n)}_{AB}$ of Kinnersley and Chitre is given in terms of $F_{AB}(t)$ by equation (A19). A number of useful algebraic and analytic properties of $F_{AB}(t)$ and $G_{AB}(s,t)$ are collected in Appendix A.

A point which must be emphasized is that H_{AB} and H^{*}_{AB} are only complex conjugates of each other when constructed from a real-valued metric and evaluated at real values of ρ and z. In this case, $F^{*}_{AB}(t)$ is the complex conjugate of $F_{AB}(t^{*})$. But since we wish to make use of the Neugebauer–Kramer mapping (I), which maps real-valued metrics to complex-valued, we must understand that $\mathscr{E}^{*}$ and H^{*}_{AB} are always defined as in (1.2) and (1.5) and so on for the higher potentials. Then $F^{*}_{AB}(t)$ is given in terms of $F_{AB}(t)$ by Eq. (A23).

Before launching into the theory of the recent solution generating techniques, we need a workable notation for some transformations which have been known for a long time. The SO(2,1) group of coordinate transformations, $t = \beta_1 t' + \beta_3\phi', \phi = \beta_2 t' + \beta_4\phi', \beta_1\beta_4 - \beta_2\beta_3 = 1$, transforms the metric according to

$$f'_{AB} = b_{A}{}^{C}b_{B}{}^{D}f_{CD}, \quad b_{A}{}^{B} = \begin{pmatrix}\beta_1 & \beta_2\\ \beta_3 & \beta_4\end{pmatrix} = \beta, \quad (1.6)$$

β denoting an unimodular matrix.[33] This is precisely the transformation law for SL(2) tensors and the same symmetry necessarily propagates through the hierachy of potentials. So the generating functions transform according to

$$F'_{AB}(t) = b_{A}{}^{C}b_{B}{}^{D}F_{CD}(t), \quad G'_{AB}(s,t) = b_{A}{}^{C}b_{B}{}^{D}G_{CD}(s,t). \quad (1.7)$$

This 3–parameter group will be denoted **L** and the element of **L** with matrix parameter β will be denoted $(L)_{\beta}$. We write $f'_{AB} = (L)_{\beta}f_{AB}$, $\{f',\omega',\psi'\} = (L)_{\beta}\{f,\omega,\psi\}$, etc. The three-parameter group **L** is generated by three one-parameter subgroups, to be called **A**, **U** and **L**, whose elements, $(A)_{\lambda}$, $(U)_{\mu}$ and $(L)_{\nu}$, respectively, are represented by β matrices,

$$\begin{pmatrix}1 & 0\\ -\lambda & 1\end{pmatrix}, \quad \begin{pmatrix}e^{\mu/2} & 0\\ 0 & e^{-\mu/2}\end{pmatrix}, \quad \begin{pmatrix}1 & \nu\\ 0 & 1\end{pmatrix}, \quad (1.8a,b,c)$$

respectively.[34] $(A)_{\lambda}$ and $(U)_{\mu}$ are trivial gauge changes: $(A)_{\lambda}\{f,\omega,\psi\} = \{f,\omega+\lambda,\psi\}$, $(U)_{\mu}\{f,\omega,\psi\} = \{e^{\mu}f, e^{-\mu}\omega, e^{\mu}\psi\}$. $(L)_{\nu}$ rotates the Killing vectors. Our notation ensures that $(L)_{\beta}(L)_{\alpha} = (L)_{\beta\alpha}$, $(A)_{\lambda}(A)_{\mu} = (A)_{\lambda+\mu}$, etc.

The dual[25] group **P** with elements $(P)_{\alpha}$, α an unimodular matrix, has the effect

$$\mathscr{E}' = \frac{\alpha_4\mathscr{E} - i\alpha_3}{\alpha_1 + i\alpha_2\mathscr{E}}, \quad (1.9a)$$

$$\mathscr{E}^{*\prime} = \frac{\alpha_4\mathscr{E}^{*} + i\alpha_3}{\alpha_1 - i\alpha_2\mathscr{E}^{*}}, \quad (1.9b)$$

$$\alpha = \begin{pmatrix}\alpha_1 & \alpha_2\\ \alpha_3 & \alpha_4\end{pmatrix}. \quad (1.9c)$$

This three-parameter group is generated by three one-parameter subgroups, **B**, **U** (same as above) and **P**, whose elements, $(B)_{\lambda}$, $(U)_{\mu}$ and $(P)_{\nu}$, respectively, are represented by α matrices

$$\begin{pmatrix}1 & 0\\ -\lambda & 1\end{pmatrix}, \quad \begin{pmatrix}e^{-\mu/2} & 0\\ 0 & e^{\mu/2}\end{pmatrix}, \quad \begin{pmatrix}1 & \nu\\ 0 & 1\end{pmatrix}, \quad (1.10a,b,c)$$

respectively. $(B)_{\lambda}$ is the trivial gauge change: $(B)_{\lambda}\{f,\omega,\psi\} = \{f,\omega,\psi+\lambda\}$. $(P)_{\lambda}$ is variously called the Ehlers transformation, NUT transformation or gravitational duality rotation and ν is often called the NUT parameter.[35] An alternative subgroup with $\alpha_1 = \alpha_4 = \cos\lambda/2$, $\alpha_2 = -\alpha_3 = \sin\lambda/2$ has the effect,

$$\xi' = e^{i\lambda}\xi, \quad \xi = \frac{1+\mathscr{E}}{1-\mathscr{E}}, \quad (1.11)$$

a formula due to Ernst.[31] The parameter λ here can also be called the NUT parameter. Now, Kinnersley and Chitre[4,36] have calculated the effect of $(P)_{\lambda}$ on the hierarchy of potentials. From their results, we can deduce

$$(P)_{\alpha}F_{11}(t) = \frac{\alpha_4F_{11}(t) + \alpha_3tF_{12}(t)}{\alpha_1 - i\alpha_2H_{11}}, \quad (1.12a)$$

$$(P)_{\alpha}F_{12}(t) = \frac{\alpha_1F_{12}(t) + (\alpha_2/t)F_{11}(t)}{\alpha_1 - i\alpha_2H_{11}} \quad (1.12b)$$

(recall $H_{11} = \mathscr{E}^{*}$). The $(P)_{\alpha}$ transform of $G_{AB}(s,t)$ may be deduced most easily from equation (A40).

The Neugebauer–Kramer mapping[24,25] (I) is defined by

$$(I)\{f,\omega,\psi\} = \{\rho f^{-1}, i\psi, -i\omega\}. \quad (1.13)$$

Also, we may define an (I^{*}) transformation[37] by replacing i by $-i$ in this formula. Further, we may cover both possibilities with a single symbol (I^{ϵ}) where $\epsilon = \pm 1$ and $(I^{+1}) = (I)$, $(I^{-1}) = (I^{*})$. The (I^{ϵ}) transform of the generating function $F_{AB}(t)$ is calculated in Sec. 3. A simple duality[25] theorem is: $(I^{\epsilon})(P)_{\alpha}(I^{\epsilon}) = (L)_{\beta}$ with $\beta_1 = \alpha_1$, $\beta_2 = -\epsilon i\alpha_2$, $\beta_3 = \epsilon i\alpha_3$, $\beta_4 = \alpha_4$.

The $\gamma^{(k)}_{AB}$ transformations of Kinnersley and Chitre with negative index k form an infinite dimensional group of gauge transformations,[4,36] earlier studied by Geroch.[9] The generating function $F_{AB}(t)$ transforms according to

$$F'_{AB}(t) = F_{A}{}^{X}(t)g_{XB}(t), \quad (1.14)$$

where the gauge functions $g_{AB}(t)$ depend on t only, not ρ and z, and are arbitrary except for the conditions,

$$g_{AB}(0)=\epsilon_{AB},$$
$$g_{AB}(t)=g^{*}{}_{AB}(t), \tag{1.15}$$
$$g_{XA}(t)g^{X}{}_{B}(t)=\epsilon_{AB}.$$

The $(B)_\lambda$ transformation is the special case where $g_{11}=-\lambda t, g_{12}=-g_{21}=1, g_{22}=0$. The **A**, **U** and **L** transformations are not included since $k=0$ in these cases. The dual of the gauge group is the same except that **A** is included and **B** is excluded.

Finally there is an affine group of transformations of the nonignorable coordinates. Define the **R** and **Z** transformations[38] by

$$(R)_\lambda\{\rho,z,H_{AB}\}=\{\Lambda\rho,\Lambda z,\Lambda H_{AB}\},\quad \Lambda=e^{\lambda}, \tag{1.16}$$
$$(Z)_\mu\{\rho,z,H_{AB}\}=\{\rho,z+\mu,H_{AB}+i\mu\epsilon_{AB}\}. \tag{1.17}$$

It is not difficult to construct the **R** and **Z** transforms of the higher potentials with the aid of the recurrence relations in Ref. 3. The results are collected in the formulas,

$$(R)_\lambda F_{AB}(t)=F_{AB}(\Lambda t),\quad (R)_\lambda G_{AB}(s,t)=G_{AB}(\Lambda s,\Lambda t), \tag{1.18}$$

$$(Z)_\mu F_{AB}(t)=(1-2\mu t)^{-1/2}F_{AB}\left(\frac{t}{1-2\mu t}\right). \tag{1.19}$$

There are also trivial reflections, $(\rho,z)\to(\rho,-z)$ and $(f,\omega,\psi)\to(f,-\omega,-\psi)$. The reflections, $(\rho,z)\to(-\rho,-z)$ and $(f,\omega,\psi)\to(-f,-\omega,-\psi)$ may be interpreted as $(R)_{i\pi}$ and $(U)_{i\pi}$, respectively.

The definition of asymptotic flatness we adopt is given by equation (B3) which allows a NUT parameter. This form is preserved by **P**, **R**, **Z** and the gauge group and can always be converted to the strict asymptotically Minkowskian form by an element of **P**. According to this definition, $(L)_\beta$ with $\beta_2\neq 0$ and (I) do not preserve asymptotic flatness.

It is easy to calculate the transforms of the metric coefficient $e^{2\gamma}$ under the above transformations.[39] The results are

$$(P)_\alpha e^{2\gamma}=e^{2\gamma},\quad (L)_\beta(f^{-1}e^{2\gamma})=f^{-1}e^{2\gamma}, \tag{1.20a,b}$$
$$(\text{gauge})e^{2\gamma}=e^{2\gamma},\quad (I)e^{2\gamma}=\rho^{1/2}f^{-1}e^{2\gamma}, \tag{1.20c,d}$$
$$(R)_\lambda=e^{2\gamma},\quad (Z)_\mu e^{2\gamma}=e^{2\gamma}. \tag{1.20e,f}$$

The latter two cases are to be interpreted as in Footnote 38. Since there is the extra gauge freedom to add a constant to γ, we could instead write $(R)_\lambda e^{2\gamma}=\Lambda e^{2\gamma}$ or $(R)_\lambda e^{2\gamma}=\Lambda^{1/2}e^{2\gamma}$, $\Lambda=e^{\lambda}$, both of which are natural choices in certain contexts.

2. EXPONENTIATION OF Q AND $\tilde{Q}$ USING THE GENERATING FUNCTIONS

In Refs. 10–12, we introduced a three-parameter transformation group **Q** which preserves asymptotic flatness (sketch of proof in Appendix B) and is not contained in the Geroch group. This is the free product of a one-parameter subgroup $\mathbf{Q}^0$ and the **R** and **Z** transformations and is locally isomorphic to SL(2,R). One form for the transformation, which is already exponentiated, involves the metric coefficient $e^{2\gamma}$.

$$e^{2\gamma}=(Q)_\delta e^{2\gamma}=e^{2\gamma},\quad \delta=\begin{pmatrix}\delta_1 & \delta_2\\ \delta_3 & \delta_4\end{pmatrix},\quad \det\delta=1, \tag{2.1}$$

$$\rho'=(Q)_\delta\rho=4\rho\,[\delta_3{}^2\rho^2+(\delta_3 z-2\delta_4)^2]^{-1}, \tag{2.2a}$$

$$z'=(Q)_\delta z=-2\,\frac{\delta_1\delta_3\rho^2+(\delta_1 z-2\delta_2)(\delta_3 z-2\delta_4)}{\delta_3{}^2\rho^2+(\delta_3 z-2\delta_4)^2}. \tag{2.2b}$$

As before,[38] (2.1) should be interpreted as $\gamma'(\rho',z')=\gamma(\rho,z)$ or $\gamma'(\rho,z)=\gamma(\rho'',z'')$, where $(\rho'',z'')=(Q)_\delta^{-1}(\rho,z)$. As is well known, γ is easily constructed by quadratures from f and ω or form f and ψ. In Ref. 15, the inverse problem is reduced to an ordinary differential equation and one has a choice of a Riccati equation, a second-order linear or a third-order linear equation. The Ernst potential is unique up to a **P** transformation (and a reflection $\psi\to-\psi$), which commutes with **Q**.

The one-parameter subgroup elements, $(Q^0)_{2t}$, $(Q^\kappa)_{2s}$, $(R)_{2\lambda}$ and $(Z)_{2\mu}$ are represented by δ matrices,

$$\begin{pmatrix}1 & 0\\ 2t & 1\end{pmatrix},\quad \begin{pmatrix}\cosh\kappa s & (\kappa/2)\sinh\kappa s\\ (2/\kappa)\sinh\kappa s & \cosh\kappa s\end{pmatrix}, \tag{2.3a,b}$$

$$\begin{pmatrix}e^{\lambda} & 0\\ 0 & e^{-\lambda}\end{pmatrix},\quad \begin{pmatrix}1 & -\mu\\ 0 & 1\end{pmatrix}, \tag{2.3c,d}$$

respectively. The CTS solutions[13] were originally derived[12] by seeking solutions which were invariant, up to a change of NUT parameter, under $\mathbf{Q}^\kappa$. In terms of coordinates,

$$\nu=y/x,\quad \eta=(x^2-1)/(1-y^2), \tag{2.4}$$

where (x,y) are prolate spheroidal coordinates defined by

$$\rho=\kappa(x^2-1)^{1/2}(1-y^2)^{1/2},\quad z=\kappa xy, \tag{2.5}$$

we have, under $(Q^\kappa)_{2s}$,

$$\gamma'(\nu',\eta')=\gamma(\nu,\eta),\quad \nu'=\frac{\nu-\tanh\kappa s}{1-\nu\tanh\kappa s},\quad \eta'=\eta. \tag{2.6}$$

The most general conditions under which the metric is invariant up to a change of NUT parameter occur when

$$\gamma(\nu,\eta)=f(\eta)+h\ln\frac{1+\nu}{1-\nu}, \tag{2.7}$$

h constant, f function of η only. In Ref. 16, this is shown to be equivalent to a slight generalization of "Rule (a)" of Tomimatsu and Sato[40] and, using the theory of the previous paper (Ref. 15), it uniquely determines a six-parameter class of solutions (NUT parameter included). The asymptotically flat subclass of the CTS solutions requires $h=0$ and a boundary condition on the differential equation for $H_4(\eta)$, one of the two transcendental functions which characterize the solutions.[13,16]

The above definition of **Q** only defines the transformed Ernst potential up to a **P** transformation. It is possible to give a precise formula for the infinitesimal **Q** transform of the Ernst potential and then exponentiate it using the generating functions. We use the script letter $\mathscr{Q}$ for the infinitesimal $\mathbf{Q}^0$ transformation so that

$$(Q^0)_t=1+t\mathscr{Q}+(t^2/2!)\mathscr{Q}^2+(t^3/3!)\mathscr{Q}^3+\cdots. \tag{2.8}$$

(In future, we shall drop the index zero from $(Q^0)_t$ and $(\tilde{Q}^0)_t$.) The infinitesimal transformation equations are

$$\mathscr{Q}\rho=\rho z,\quad \mathscr{Q}z=\tfrac{1}{2}(z^2-\rho^2), \tag{2.9}$$
$$\mathscr{Q}f=-\tfrac{1}{2}f\psi_{21},\quad \mathscr{Q}\omega=\tfrac{1}{2}(\psi_{22}+\omega\psi_{12}). \tag{2.10}$$

It is straightforward to show that $\mathscr{Q}\gamma=0$, which implies $e^{2\gamma}=e^{2\gamma}$, and that $(Q)_t(\rho,z)$ is given by (2.2) with δ given by

(2.3a). It turns out that the **Q** transform of $(I)f_{AB}$ can be written as a **P**–covariant tensor equation among the duals of the Kinnersley–Chitre potentials. Consequently, the original potentials and generating functions transform as tensors under the dual group $\tilde{\mathbf{Q}}$, defined by $(\tilde{Q})_\delta = (I)(Q)_\delta(I) = (I^*)(Q)_\delta(I^*)$, which commutes with **L**. Although $\tilde{\mathbf{Q}}$ alone does not preserve asymptotic flatness, its importance in combinations with **Q** has been amply demonstrated by Neugebauer.[19] We consider this case first. From (1.2), (1.13), (2.9), and (2.10),

$$\begin{aligned}\tilde{\mathscr{Q}} f &= \tfrac{1}{2} f(z - \psi_{21} + \omega\psi),\\ \tilde{\mathscr{Q}}\omega &= \tfrac{1}{4}[2\omega(\psi_{21}+z) + \psi_{22} - \psi(\omega^2 + \rho^2 f^{-2})].\end{aligned} \tag{2.11}$$

Using the defining relations for the potentials, we may successively build up the following formulas,

$$\tilde{\mathscr{Q}} f_{AB} = \tfrac{1}{4}(2z f_{AB} - f_{AX}\psi^X{}_B - f_{BX}\psi^X{}_A), \tag{2.12}$$

$$\tilde{\mathscr{Q}} H_{AB} = \tfrac{1}{4}[H^{(2)}_{AB} + iN^{(1,1)}_{AB}], \tag{2.13}$$

$$\tilde{\mathscr{Q}} N^{(m,n)}_{AB} = \tfrac{1}{4}[(m+1)N^{(m+1,n)}_{AB} + nN^{(m,n+1)}_{AB}]. \tag{2.14}$$

From (2.14) and (A15b), the infinitesimal transform of the generating function G_{AB} is

$$\tilde{\mathscr{Q}} G_{AB}(t_1,t_2) = \tfrac{1}{4}[(D_1 + D_2 - t_2^{-1})G_{AB}(t_1,t_2) + t_2^{-1}\epsilon_{AB}], \tag{2.15}$$

where D_1 and D_2 denote $\partial/\partial t_1$ and $\partial/\partial t_2$, respectively.[41] This expression is now very easily exponentiated, giving the explicit formulas,

$$(\tilde{Q})_{4s} G_{AB}(t_1,t_2) = \frac{t_2}{s+t_2} G_{AB}(s+t_1, s+t_2) + \frac{s}{s+t_2}\epsilon_{AB}, \tag{2.16}$$

$$(\tilde{Q})_{4s} F_{AB}(t) = -iS(s) F_{XA}(s) F^X{}_B(s+t), \tag{2.17}$$

$$\begin{aligned}(\tilde{Q})_{4s} H_{AB} &= -iS(s)F_{XA}(s)DF^X{}_B(s)\\ &= is^{-1}(G_{AB}(s,s) - \epsilon_{AB}),\end{aligned} \tag{2.18}$$

where $D = d/ds$ and $S(s)$ is defined by equation (A16a).[42] Equations (2.17) and (2.18) follow from (A19) and (A20a) which relate G_{AB} to F_{AB}.

The general $(\tilde{Q})_\delta$ transforms of F_{AB} and G_{AB} can be calculated by first expressing $\boldsymbol{\delta}$ as a product of matrices of the form (2.3a, c and d) in any order and using (1.18) and (1.19). The results are:

$$\begin{aligned}(\tilde{Q})_\delta F_{AB}(t) = &-i\left(1 + \frac{4\delta_2 t}{\delta_4}\right)^{-1/2} S\left(\frac{\delta_3}{4\delta_4}\right) F_{XA}\left(\frac{\delta_3}{4\delta_4}\right)\\ &\times F^X{}_B\left(\frac{\delta_3/4 + \delta_1 t}{\delta_4 + 4\delta_2 t}\right),\end{aligned} \tag{2.19}$$

$$\begin{aligned}&(\tilde{Q})_\delta G_{AB}(t_1,t_2)\\ &= \left[\frac{t_1}{t_1 - t_2} - \frac{t_2}{t_1 - t_2}\left(\frac{\delta_4 + 4\delta_2 t_2}{\delta_4 + 4\delta_2 t_1}\right)^{1/2}\frac{\delta_3 + 4\delta_1 t_1}{\delta_3 + 4\delta_1 t_2}\right]\epsilon_{AB}\\ &\quad + \frac{4t_2}{(\delta_4 + 4\delta_2 t_1)^{1/2}(\delta_4 + 4\delta_2 t_2)^{1/2}(\delta_3 + 4\delta_1 t_2)}\\ &\quad \times G_{AB}\left(\frac{\delta_3/4 + \delta_1 t_1}{\delta_4 + 4\delta_2 t_1}, \frac{\delta_3/4 + \delta_1 t_2}{\delta_4 + 4\delta_2 t_2}\right).\end{aligned} \tag{2.20}$$

Now the same formulas could serve for the **Q** transformation law if we replace the **L**-covariant tensors, F_{AB} and G_{AB}, by the **P**–covariant tensors, $(I)F_{AB}$ and $(I)G_{AB}$, respectively. The dual generating functions are not only better adapted for **Q**, but also for the $\beta^{(k)}$ transformations,[5] but these are isolated cases. Since most of the transformations to be discussed in this paper are either self-dual or involve products of both **Q** and $\tilde{\mathbf{Q}}$, it will make no difference whether we work with F_{AB} and G_{AB} or their duals. So next we turn to the more difficult problem of constructing a formula for $(Q)_{4s}F_{AB}(t)$, which cannot, of course, be a tensor equation.

Consider, first, the infinitesimal transformation $\mathscr{Q} - \tilde{\mathscr{Q}}$. From (2.10) and (2.11)

$$(\mathscr{Q} - \tilde{\mathscr{Q}})f = -\tfrac{1}{2} f(z + \omega\psi), \tag{2.21a}$$

$$(\mathscr{Q} - \tilde{\mathscr{Q}})\omega = \tfrac{1}{4}[\psi_{22} + 2z\omega + \psi(\omega^2 + \rho^2 f^{-2})]. \tag{2.21b}$$

Now, this is a familiar transformation in the Geroch group. It is precisely Geroch's $-(1/2)\mathscr{T}_2$ transformation,[9] $\mathscr{T}_2$ being the $\alpha = 2$ component of the SO(2,1) vector $\mathscr{T}_\alpha$. In the notation of Kinnersley and Chitre, it is the special $\gamma^{(1)}_{AB}$ transformation where the $\gamma^{(1)}_{AB}$ parameters are given by $\gamma_{11} = \gamma_{22} = 0$, $\gamma_{12} = \gamma_{21} = 1/8$. Kinnersley and Chitre use the symbol $\gamma^{(k)}_{AB}$ for both the infinitesimal operator (like $\mathscr{T}_\alpha$ and $\mathscr{Q}$) and the group parameter (a second-rank tensor to be contracted with $\gamma^{(k)}_{AB}$). Here, we wish to distinguish the two situations. We shall write Eq. (3.1) of Ref. 3 (restricting electrovac to vacuum) in the form,

$$\begin{aligned}\gamma^{(k)}_{XY} N^{(m,n)}_{AB} = &\ \epsilon_{A(X} N^{(m+k,n)}_{Y)B} - N^{(m,n+k)}_{A(X}\epsilon_{Y)B}\\ &+ \sum_{s=1}^{k} N^{(m,s)}_{A(X} N^{(k-s,n)}_{Y)B},\end{aligned} \tag{2.22}$$

which contains no parameters.[43] A linear combination of $\gamma^{(k)}_{11}$, $\gamma^{(k)}_{12}$ and $\gamma^{(k)}_{22}$ can be specified by contracting $\gamma^{(k)}_{XY}$ with a symmetric constant tensor, say q^{XY}. The infinitesimal $q^{XY}\gamma^{(k)}_{XY}$ transformation law reads,

$$\begin{aligned}q^{XY}\gamma^{(k)}_{XY} N^{(m,n)}_{AB} = &-q_A{}^X N^{(m+k,n)}_{XB} - q^X{}_B N^{(m,n+k)}_{AX}\\ &+ q^{XY}\sum_{s=1}^{k} N^{(m,s)}_{AX} N^{(k-s,n)}_{YB}.\end{aligned} \tag{2.23}$$

This expression holds also for negative k, except when $m + n = -k$, $m \geqslant 0$, $n \geqslant 1$, in which case it should read $\gamma^{(k)}_{XY} N^{(m,n)}_{AB} = -\epsilon_{A(X}\epsilon_{Y)B}$. Geroch's $K^\alpha\mathscr{T}_\alpha$ transformation is the same as $q^{XY}\gamma^{(1)}_{XY}$ with the identification,

$$\begin{aligned}K^1 &= K_3 = 2q^{11} = 2q_{22},\\ K^2 &= -2K_2 = 4q^{12} = -4q_{12},\\ K^3 &= K_1 = 2q^{22} = 2q_{11}.\end{aligned}$$

From the above remarks,

$$\mathscr{Q} - \tilde{\mathscr{Q}} = q^{XY}\gamma^{(1)}_{XY} \text{ with } q_{11} = q_{22} = 0, q_{12} = q_{21} = \tfrac{1}{8}. \tag{2.24}$$

Now, from (2.23) with $k = 1$ and Eqs. (A15) and (A22),

$$\begin{aligned}&q^{XY}\gamma^{(1)}_{XY} G_{AB}(t_1,t_2)\\ &\quad = t_1^{-1}[q_{AX}G^X{}_B(t_1,t_2) - S(t_1)q^{XY}F_{XA}(t_1)F_{YB}(t_2)]\\ &\qquad + t_2^{-1} q_{XB}(G_A{}^X(t_1,t_2) - \epsilon_A{}^X)\end{aligned} \tag{2.25}$$

(recall $\epsilon_A{}^X = -\epsilon^X{}_A = \delta^X_A$). Next, from equations (A20a) and (A21),

$$\begin{aligned}q^{XY}\gamma^{(1)}_{XY} F_{AB}(t) = &\ t^{-1}[q_{AX}F^X{}_B(t) + q_{XB}F_A{}^X(t)]\\ &+ iq_{AX}H^{YX}F_{YB}(t) - iq_{XY}H^X{}_A F^Y{}_B(t).\end{aligned} \tag{2.26}$$

Restrict attention to the components, $F_{11}(t)$ and $F_{12}(t)$, only, as the remaining components and all of $G_{AB}(t_1,t_2)$ may be obtained in terms of these two by equations (A39) and (A40). From (2.24) and (2.26),

$$(\mathscr{Q} - \bar{\mathscr{Q}})F_{11}(t) = \tfrac{1}{4}[-t^{-1}F_{11}(t) + iH_{11}F_{21}(t)], \tag{2.27a}$$

$$(\mathscr{Q} - \bar{\mathscr{Q}})F_{12}(t) = \tfrac{1}{4}iH_{11}F_{22}(t). \tag{2.27b}$$

But, from either (2.15) or (2.17),

$$\bar{\mathscr{Q}}F_{11}(t) = \tfrac{1}{4}[DF_{11}(t) - iH_{11}F_{21}(t) + iH_{12}F_{11}(t)], \tag{2.28a}$$

$$\bar{\mathscr{Q}}F_{12}(t) = \tfrac{1}{4}[DF_{12}(t) - iH_{11}F_{22}(t) + iH_{12}F_{12}(t)], \tag{2.28b}$$

$D = d/dt$. Finally, from (2.27) and (2.28),

$$\mathscr{Q}F_{11}(t) = \tfrac{1}{4}[D - t^{-1} + iH_{12}]F_{11}(t), \tag{2.29a}$$

$$\mathscr{Q}F_{12}(t) = \tfrac{1}{4}[D + iH_{12}]F_{12}(t). \tag{2.29b}$$

Now, the infinitesimal transformation equations (2.29) are not difficult to exponentiate. The method is to first write $(Q)_{4s}F_{12}(t) = F_{12}(t,s)$, so that (2.29b) reads

$$D_2F_{12}(t,s) = [D_1 + iD_1F_{12}(0,s)]F_{12}(t,s).$$

The general integral of this equation is

$$F_{12}(t,s) = \frac{if(s+t)}{f(s)+k},$$

$f(s)$ an arbitrary function of ρ, z and s, k an arbitrary function of ρ and z only. The particular integral satisfying $F_{12}(t,0) = F_{12}(t)$ and $F_{12}(0,s) = i$ is

$$F_{12}(t,s) = iF_{12}(s+t)/F_{12}(s).$$

Next, the integral of (2.29a) follows immediately by quadratures. The final results are:

$$(Q)_{4s}F_{11}(t) = \frac{it}{s+t}\frac{F_{11}(s+t)}{F_{12}(s)}, \tag{2.30a}$$

$$(Q)_{4s}F_{12}(t) = i\frac{F_{12}(s+t)}{F_{12}(s)}. \tag{2.30b}$$

These formulas could also be deduced from (2.28) and the Neugebauer commutation theorem by a circuitous method (see Sec. 4).

The transforms of $F_{21}(t)$, $F_{22}(t)$ and $G_{AB}(t_1,t_2)$ can now be obtained in several ways using the formulas in Appendix A. The results are:

$$(Q)_{4s}F_{21}(t) = -\frac{is}{s+t} + \frac{it}{s+t} \times \left[G_{21}(s,s+t) - G_{22}(s,s)\frac{F_{11}(s+t)}{F_{12}(s)}\right], \tag{2.31a}$$

$$(Q)_{4s}F_{22}(t) = i\left[G_{22}(s,s+t) - G_{22}(s,s)\frac{F_{12}(s+t)}{F_{12}(s)}\right], \tag{2.31b}$$

$$(Q)_{4s}G_{11}(t_1,t_2) = \frac{t_1t_2}{(s+t_1)(s+t_2)}G_{11}(s+t_1,s+t_2) - \frac{F_{11}(s+t_2)}{F_{12}(s)} \times \left[\frac{t_1t_2}{(s+t_1)(s+t_2)}G_{12}(s+t_1,s) - \frac{t_2}{s+t_2}\right], \tag{2.32a}$$

$$(Q)_{4s}G_{12}(t_1,t_2) = \frac{t_1}{s+t_1}G_{12}(s+t_1,s+t_2) - \frac{F_{12}(s+t_2)}{F_{12}(s)} \times \left[\frac{t_1}{s+t_1}G_{12}(s+t_1,s) - 1\right], \tag{2.32b}$$

$$(Q)_{4s}G_{21}(t_1,t_2) = \frac{t_2}{s+t_2}G_{21}(s+t_1,s+t_2) - \frac{s}{s+t_2} - \frac{t_2}{s+t_2}\frac{F_{11}(s+t_2)}{F_{12}(s)}G_{22}(s+t_1,s), \tag{2.32c}$$

$$(Q)_{4s}G_{22}(t_1,t_2) = G_{22}(s+t_1,s+t_2) - \frac{F_{12}(s+t_2)}{F_{12}(s)}G_{22}(s+t_1,s). \tag{2.32d}$$

3. FORMULAS FOR THE DUAL OF THE GENERATING FUNCTION $F_{AB}(t)$

Now that we know the transforms of $F_{AB}(t)$ under both $(Q)_{4s}$ and $(\tilde{Q})_{4s} = (I^\epsilon)(Q)_{4s}(I^\epsilon)$, $\epsilon = \pm 1$, we can attempt to calculate the transform of $F_{AB}(t)$ under the Neugebauer–Kramer mapping (I^ϵ). It is remarkable that such a formula exists at all. We first need explicit expressions for $F_{11}(s)$, $F_{12}(s)$, $F^*{}_{11}(s)$ and $F^*{}_{12}(s)$ in terms of either the $(Q)_{4s}$ or $(\tilde{Q})_{4s}$ transforms of f, ω and/or ψ.

From (2.30a) and (A41),

$$(Q)_{4s}\mathscr{E} = -\frac{i}{s}\frac{F^*{}_{11}(s)}{F^*{}_{12}(s)}, \quad (Q)_{4s}\mathscr{E}^* = \frac{i}{s}\frac{F_{11}(s)}{F_{12}(s)}, \tag{3.1a,b}$$

$$(Q)_{4s}f = (S^2(s)F_{12}(s)F^*{}_{12}(s))^{-1}f. \tag{3.1c}$$

An equation involving $F_{12}(s)$ alone may be obtained from $\nabla((Q)_{4s}\mathscr{E}^*)$. Using (A39), rewrite (A17) in terms of F_{11} and F_{12} to give

$$\nabla F_{1B} = (2f)^{-1}[(\nabla H_{11} - 2fS^{-1}\nabla S)F_{1B} + S^{-1}(\nabla_s H_{11})F^*{}_{1B}], \tag{3.2}$$

where $\nabla_s = (1-2sz)\nabla - 2s\rho\tilde{\nabla}$. Then, from (3.1b) and (A41),

$$\nabla((Q)_{4s}\mathscr{E}^*) = -\nabla_s\mathscr{E}^*/S^3(F_{12})^2. \tag{3.3a}$$

Alternatively, this could be written,

$$(Q)_{4s}(\nabla\mathscr{E}^*) = -\frac{(1-2sz)\nabla\mathscr{E}^* + 2s\rho\tilde{\nabla}\mathscr{E}^*}{S(F_{12})^2}. \tag{3.3b}$$

Let a single prime, a double prime, and a triple prime denote an (I^ϵ) transform, a $(Q)_{4s}(I^\epsilon)$ transform, and an (I^ϵ) $(Q)_{4s}(I^\epsilon) = (\tilde{Q})_{4s}$ transform, respectively. Then, from (1.13) and (3.3a),

$$\nabla(\rho'''f'''^{-1} - \epsilon\omega''') = \nabla(\mathscr{E}^{*\prime\prime}) = -\frac{\nabla_s\mathscr{E}^{*\prime}}{S^3(F'_{12})^2} = -\frac{\nabla_s(\rho f^{-1} - \epsilon\omega)}{S^3(F'_{12})^2}. \tag{3.4}$$

To calculate the left-hand side of (3.4), we first express $(\tilde{Q})_{4s}f_{AB}$ in terms of F_{11} and F_{12}. From (2.18) and (A28),

$$(\tilde{Q})_{4s}f_{AB} = -S^{-1}f_{XY}F^X{}_AF^Y{}_B. \tag{3.5}$$

Then, from (A39),

$$(\tilde{Q})_{4s} f = (4s^2 f)^{-1}[S((F_{11})^2 + (F^*_{11})^2) + 2(1-2sz)F_{11}F^*_{11}], \tag{3.6a}$$

$$(\tilde{Q})_{4s}(-f\omega) = (4s^2 f)^{-1}[S(F_{11}F_{12} + F^*_{11}F^*_{12}) + (1-2sz)(F_{11}F^*_{12} + F^*_{11}F_{12})]. \tag{3.6b}$$

Hence,

$$(\tilde{Q})_{4s}(\rho f^{-1} - \epsilon\omega) = \frac{\epsilon(F^*_{12} + \zeta^\epsilon F_{12})}{F^*_{11} + \zeta^\epsilon F_{11}}, \tag{3.7}$$

where $\zeta = \zeta(s)$ is the polar angular co-ordinate defined by (A.16b) and ζ^ϵ is to be read: ζ to the power of $\epsilon(=\pm 1)$. To obtain equation (3.7), we have used $(\tilde{Q})_{4s}\rho = S^{-2}\rho$ and (A.41) again and noticed that a common factor $F_{11} + \zeta^\epsilon F^*_{11}$ cancelled out in the numerator and denominator. Now take the gradient of both sides of (3.7) using (3.2). We obtain

$$\begin{aligned}\nabla((\tilde{Q})_{4s}(\rho f^{-1} - \epsilon\omega)) &= \frac{\epsilon i s}{S^2}[S^{-1}\nabla_s\mathscr{E} + \zeta^\epsilon\nabla\mathscr{E}^* - \zeta^\epsilon\nabla\mathscr{E} - \zeta^{2\epsilon}S^{-1}\nabla_s\mathscr{E}^* \\ &\quad + 2f\nabla\zeta^\epsilon]/(F^*_{11} + \zeta^\epsilon F_{11})^2 \\ &= \frac{4s^2\zeta^\epsilon f^2\nabla_s(\rho f^{-1} - \epsilon\omega)}{S^4(F^*_{11} + \zeta^\epsilon F_{11})^2}.\end{aligned} \tag{3.8}$$

Comparing this expression with the right-hand-side of (3.4), we find

$$(F'_{12})^2 = -\frac{1 - 2sz + 2\epsilon i s\rho}{4s^2 f^2}(F^*_{11} + \zeta^\epsilon F_{11})^2.$$

Finally, the sign of $F'_{12}(s)$ is determined from $F'_{12}(0) = i$. Hence, the explicit formula for $(I^\epsilon)F_{12}(s)$ is,

$$(I^\epsilon)F_{12}(s) = i(2sf)^{-1}[(1-2sz-2\epsilon i s\rho)^{1/2}F_{11}(s) + (1-2sz+2\epsilon i s\rho)^{1/2}F^*_{11}(s)], \tag{3.9a}$$

where both square roots take the value $+1$ when $s = 0$.

Now that $(I^\epsilon)F_{12}(s)$ has been found, the (I^ϵ) transforms of the other components, F^*_{12}, F_{11} and F^*_{11}, follow from similar but much easier calculations starting with (3.1a,b,c). The results are:

$$(I^\epsilon)F^*_{12}(s) = -i(2sf)^{-1}[(1-2sz+2\epsilon i s\rho)^{1/2}F_{11}(s) + (1-2sz-2\epsilon i s\rho)^{1/2}F^*_{11}(s)], \tag{3.9b}$$

$$(I^\epsilon)F_{11}(s) = \epsilon(2f)^{-1}[(1-2sz-2\epsilon i s\rho)^{1/2}F_{12}(s) + (1-2sz+2\epsilon i s\rho)^{1/2}F^*_{12}(s)], \tag{3.9c}$$

$$(I^\epsilon)F^*_{11}(s) = -\epsilon(2f)^{-1}[(1-2sz+2\epsilon i s\rho)^{1/2}F_{12}(s) + (1-2sz-2\epsilon i s\rho)^{1/2}F^*_{12}(s)]. \tag{3.9d}$$

Alternatively, (3.9b,d) could have been deduced from (3.9a,c) using the trivial symmetries, $(I^\epsilon)F^*_{12} = -(I^{-\epsilon})F_{12}$, $(I^\epsilon)F^*_{11} = (I^{-\epsilon})F_{11}$. A third approach is to derive (3.9c,d) by inverting (3.9a,b). The remaining two components of $(I^\epsilon)F_{AB}$ follow from equation (A39). The results are:

$$(I^\epsilon)F_{21}(s) = -(2f)^{-1}[(1-2sz-2\epsilon i s\rho)^{1/2}\mathscr{E}F_{12}(s) - (1-2sz+2\epsilon i s\rho)^{1/2}\mathscr{E}^*F^*_{12}(s)], \tag{3.10a}$$

$$(I^\epsilon)F_{22}(s) = -\epsilon i(2sf)^{-1}[(1-2sz-2\epsilon i s\rho)^{1/2}\mathscr{E}F_{11}(s) - (1-2sz+2\epsilon i s\rho)^{1/2}\mathscr{E}^*F^*_{11}(s)], \tag{3.10b}$$

$$(I^\epsilon)F^*_{21}(s) = -(I^{-\epsilon})F_{21}(s), \quad (I^\epsilon)F^*_{22}(s) = (I^{-\epsilon})F_{22}(s). \tag{3.10c,d}$$

In a similar manner, the transforms of G_{AB} and G^*_{AB} can be calculated from equation (A19) or (A40). We shall not write these out.

Formulas (3.9a-d) take their neatest form in terms of functions, P_{AB} and Q_{AB}, defined by

$$F_{AB}(s) = P_{AB}(s) + iQ_{AB}(s), \qquad F^*_{AB}(s) = P_{AB}(s) - iQ_{AB}(s). \tag{3.11}$$

$P_{AB}(s)$ and $Q_{AB}(s)$ would be the real and imaginary parts of $F_{AB}(s)$ if f_{AB}, ρ, z and s were all real. But since (I^ϵ) does not respect the reality of f_{AB}, $(I^\epsilon)P_{AB}$ and $(I^\epsilon)Q_{AB}$ will not always be real-valued functions. A few properties of P_{AB} and Q_{AB} are given in Appendix A. A simple transcription of (3.9a-d) gives

$$(I^\epsilon)P_{11}(s) = f^{-1}T_2(s)Q_{12}(s), \tag{3.12a}$$

$$(I^\epsilon)Q_{11}(s) = -\epsilon i f^{-1}T_1(s)P_{12}(s), \tag{3.12b}$$

$$(I^\epsilon)P_{12}(s) = \epsilon i(sf)^{-1}T_2(s)Q_{11}(s), \tag{3.12c}$$

$$(I^\epsilon)Q_{12}(s) = (sf)^{-1}T_1(s)P_{11}(s), \tag{3.12d}$$

where

$$T_1(s) = \{\tfrac{1}{2}[1 - 2sz + S(s)]\}^{1/2}, \qquad T_2(s) = \{\tfrac{1}{2}[-1 + 2sz + S(s)]\}^{1/2}, \tag{3.13}$$

with branches fixed (at least near $s = 0$) by

$$T_1(0) = 1, \quad T_1(s)T_2(s) = s\rho. \tag{3.14}$$

As an application of these formulas, let us check the self-duality of the **R** and **Z** transformations. From (1.18), (3.9) and (3.10),

$$(I^\epsilon)(R)_\lambda(I^\epsilon)\{F_{11}(t), F_{12}(t), F_{21}(t), F_{22}(t)\} = \{\Lambda^{-1}F_{11}(\Lambda t), F_{12}(\Lambda t), F_{21}(\Lambda t), \Lambda F_{22}(\Lambda t)\}. \tag{3.15}$$

This gives the relation,

$$(I^\epsilon)(R)_\lambda(I^\epsilon) = (U)_{-\lambda}(R)_\lambda, \tag{3.16}$$

where **U** is the trivial subgroup of **L** and **P** represented by (1.8b) and (1.10b), respectively. Similarly, from (1.19), (3.9) and (3.10),

$$(I^\epsilon)(Z)_\mu(I^\epsilon)\{F_{11}(t), F_{12}(t), F_{21}(t), F_{22}(t)\} = \left\{F_{11}\left(\frac{t}{1-2\mu t}\right), (1-2\mu t)^{-1}F_{12}\left(\frac{t}{1-2\mu t}\right), F_{21}\left(\frac{t}{1-2\mu t}\right), (1-2\mu t)^{-1}F_{22}\left(\frac{t}{1-2\mu t}\right)\right\}. \tag{3.17}$$

This transformation differs from $(Z)_\mu$ by a gauge transformation of the form (1.14). Thus we may write

$$(I^\epsilon)(Z)_\mu(I^\epsilon) = (\text{gauge})(Z)_\mu, \tag{3.18a}$$

with gauge functions given by

$$g_{11}(t) = g_{22}(t) = 0, \quad g_{12}(t) = (1-2\mu t)^{-1/2}, \quad g_{21}(t) = -(1-2\mu t)^{1/2}. \tag{3.18b}$$

Another application of our (I^c) transformation formulas is to check the following results,[43]

$$(I^c)\gamma_{11}^{(k)}(I^c) = \epsilon i\gamma_{22}^{(k-1)}, \tag{3.19a}$$

$$(I^c)\gamma_{12}^{(k)}(I^c) = -\gamma_{12}^{(k)}, \tag{3.19b}$$

$$(I^c)\gamma_{22}^{(k)}(I^c) = -\epsilon i\gamma_{11}^{(k+1)}, \tag{3.19c}$$

which appear in Ref. 12. There they were proved up to a gauge transformation for $k \geqslant 0$ by checking their truth directly for $k = 0$ and $k = 1$ when acting on f, ω and ψ and then employing the commutation relation,

$$[\gamma_{XY}^{(k)}, \gamma_{UV}^{(l)}] = -\epsilon_{X(U}\gamma_{V)Y}^{(k+l)} - \epsilon_{Y(U}\gamma_{V)X}^{(k+l)}, \tag{3.20}$$

which is our version of Eq. (3.4) of Ref. 3 (restricted to vacuum). In fact, Eqs. (3.19) are rigorously true for all positive and negative k, but a direct proof using Eqs. (2.22), (3.9) and (3.10) is very laborious. These equations will be proved in Sec. 6 as a corollary of the self-duality theorem for generalized HKX transformations.

4. BÄCKLUND TRANSFORMATIONS OF NEUGEBAUER AND HARRISON AND NEUGEBAUER'S COMMUTATION THEOREM

In Ref. 19, Neugebauer introduced two Bäcklund transformations, I_1 and I_2, which were shown in Ref. 11 to be equivalent to the 6–parameter transformation groups **PQ** and $\mathbf{L\tilde{Q}}$, respectively. However, Neugebauer's Bäcklund–theoretic approach has led to some remarkably powerful applications of these groups which were not likely to be foreseen in the purely group-theoretic approach. In particular, in direct analogy with the much simpler and better known sine–Gordon equation, he produced a commutation theorem which states that parameters can be chosen so that a four-fold product $I_2I_1I_2I_1$ reduces to the identity transformation and deduced from this a composition theorem which permits arbitrary products, $\cdots I_2I_1I_2I_1$, with arbitrary parameters to be calculated by purely algebraic manipulations once the first I_1 transformation has been determined. In Ref. 11, these theorems were used to factorize the Bäcklund transformation found earlier by Harrison[18] into the forms, I_2I_1 and I_1I_2, and to provide an analogous composition theorem for Harrison transformations. In this section, we translate these results into the very different language of generating functions.

The metric form implied in Neugebauer's work is the noncanonical form,

$$ds^2 = f(dt - \omega d\phi)^2 - f^{-1}\{e^{2\gamma}dx^1dx^2 + V^2d\phi^2\}, \tag{4.1}$$

where x^1 and x^2 are characteristic coordinates,

$$x^1 = \rho + iz, \quad x^2 = \rho - iz. \tag{4.2}$$

This differs from (1.1) in that the coefficient of $d\phi^2$ is now V^2 where V is an arbitrary harmonic function, i.e.,

$$\nabla^2 V = 4V_{,1,2} = 0, \tag{4.3}$$

$,_1$ and $,_2$ denoting $\partial/\partial x^1$ and $\partial/\partial x^2$, respectively. The role of the canonical coordinates (ρ,z) in (1.1) is now taken over by the pair (V,Z) where

$$\nabla V = \tilde{\nabla} Z \quad \text{or} \quad V_{,1} = iZ_{,1}, \quad V_{,2} = -iZ_{,2}, \tag{4.4}$$

and, provided V is not a constant or a function of x^1 only or x^2 only, then the coordinate change $(\rho',z') = (V,Z)$ is always available to return (4.1) to the form (1.1). (The transformation groups of the previous sections which change the coordinates can be rewritten so that (ρ,z) remains fixed while (V,Z) changes. For example, in equations (2.2) for the $(Q)_\delta$ and $(\tilde{Q})_\delta$ transformations, simply replace ρ by V and z by Z on both sides.) Neugebauer's field variables are[44]

$$M_1 = (2f)^{-1}\mathscr{E}_{,1}, \quad M_2 = (2f)^{-1}\mathscr{E}^*_{,1}, \quad M_3 = V^{-1}V_{,1}, \tag{4.5a,b,c}$$

$$N_1 = (2f)^{-1}\mathscr{E}^*_{,2}, \quad N_2 = (2f)^{-1}\mathscr{E}_{,2}, \quad N_3 = V^{-1}V_{,2}. \tag{4.5d,e,f}$$

The Bäcklund transformations, I_1 and I_2, are specified by a pseudopotential α and a coordinate γ [not to be confused with $e^{2\gamma}$ in (1.1) and (4.1)] satisfying the total differential equation,

$$d\gamma = \gamma(\gamma-1)M_3dx^1 + (\gamma-1)N_3dx^2, \tag{4.6}$$

whose solution is

$$\gamma = \zeta^2(s) = \frac{1 - 2sZ - 2isV}{1 - 2sZ + 2isV} \tag{4.7}$$

[cf. equation (A16b)], s the constant of integration. For the I_1 transformation, which we write $I_1(\alpha,\gamma)$ for the present, the transformation is given by

$$M_1' = \alpha M_1, \quad M_2' = (\gamma/\alpha)M_2, \quad M_3' = \gamma M_3, \tag{4.8a,b,c}$$

$$N_1' = (1/\alpha)N_1, \quad N_2' = (\alpha/\gamma)N_2, \quad N_3' = (1/\gamma)N_3, \tag{4.8d,e,f}$$

where α satisfies the total Riccati equation,

$$\begin{aligned} d\alpha = {} & [\alpha(\alpha-1)M_1 + (\alpha-\gamma)M_2 + \tfrac{1}{2}\alpha(\gamma-1)M_3]dx^1 \\ & + \left[(\alpha-1)N_1 + \frac{\alpha}{\gamma}(\alpha-\gamma)N_2 \right. \\ & \left. + \frac{\alpha}{2\gamma}(\gamma-1)N_3\right]dx^2. \end{aligned} \tag{4.9}$$

The $I_2(\alpha,\gamma)$ transformation is defined as $(I^*)I_1(\alpha,\gamma)(I^*)$ where now α satisfies (4.9) with M_i and N_i replaced by $(I^*)M_i$ and $(I^*)N_i$, respectively. Since

$$(I^*)M_1 = -M_2 + \tfrac{1}{2}M_3, \quad (I^*)M_2 = -M_1 + \tfrac{1}{2}M_3, \tag{4.10}$$

$$(I^*)N_1 = -N_1 + \tfrac{1}{2}N_3, \quad (I^*)N_2 = -N_2 + \tfrac{1}{2}N_3,$$

and $(I^*)M_3 = M_3$, $(I^*)N_3 = N_3$, the differential equation for α in $I_2(\alpha,\gamma)$ becomes

$$\begin{aligned} d\alpha = {} & [-(\alpha-\gamma)M_1 - \alpha(\alpha-1)M_2 \\ & + \tfrac{1}{2}(\alpha-1)(\alpha+\gamma)M_3]dx^1 \\ & + \left[-(\alpha-1)N_1 - \frac{\alpha}{\gamma}(\alpha-\gamma)N_2 \right. \\ & \left. + \frac{1}{2\gamma}(\alpha-1)(\alpha+\gamma)N_3\right]dx^2, \end{aligned} \tag{4.11}$$

and the $I_2(\alpha,\gamma)$ transformation law becomes

$$M_1' = \frac{\gamma}{\alpha}M_1 + \frac{\gamma}{2\alpha}(\alpha-1)M_3, \tag{4.12a}$$

$$M_2' = \alpha M_2 + \tfrac{1}{2}(\gamma-\alpha)M_3, \tag{4.12b}$$

$$N_1' = \frac{1}{\alpha}N_1 + \frac{\alpha-\gamma}{2\alpha\gamma}N_3, \quad N_2' = \frac{\alpha}{\gamma}N_2 + \frac{1-\alpha}{2\gamma}N_3, \tag{4.12c,d}$$

and $M_3' = \gamma M_3$, $N_3' = (1/\gamma)N_3$. It is important to specify which equation α satisfies if the context does not make it clear.

The group properties of I_1 and I_2 are expressed by

$$I_1(\alpha_2,\gamma_2)I_1(\alpha_1,\gamma_1) = I_1(\alpha_1\alpha_2,\gamma_1\gamma_2), \tag{4.13a}$$

$$I_1^{-1}(\alpha,\gamma) = I_1(\alpha^{-1},\gamma^{-1}), \tag{4.13b}$$

and similarly for I_2. (In (4.13a), α_2 and γ_2 satisfy (4.9) and (4.6) with M_1 replaced by $I_1(\alpha_1,\gamma_1)M_1$, etc.) Neugebauer's commutation theorem states that

$$I_2(\alpha^{(4)},\gamma^{(4)})I_1(\alpha^{(3)},\gamma^{(3)})I_2(\alpha^{(2)},\gamma^{(2)})I_1(\alpha,\gamma) = 1, \tag{4.14}$$

where

$$\alpha^{(2)} = \frac{\alpha-\gamma}{\gamma(\alpha-1)}, \quad \gamma^{(2)} = \gamma^{-1}, \tag{4.15a}$$

$$\alpha^{(3)} = \gamma/\alpha, \quad \gamma^{(3)} = \gamma, \tag{4.15b}$$

$$\alpha^{(4)} = \frac{\alpha-1}{\alpha-\gamma}, \quad \gamma^{(4)} = \gamma^{-1}. \tag{4.15c}$$

Neugebauer's composition theorem states that the variables $(\alpha^{(2)},\gamma^{(2)})$ in the second transformation in the product,

$$I_2(\alpha^{(2)},\gamma^{(2)})I_1(\alpha_1,\gamma_1), \tag{4.16}$$

$\alpha^{(2)}$ arbitrary and independent of α_1, etc., are given by

$$\alpha^{(2)} = \frac{\gamma_1\alpha_2 - \alpha_1\gamma_2}{\gamma_1(\alpha_2-\alpha_1)}, \quad \gamma^{(2)} = \gamma_2/\gamma_1, \tag{4.17}$$

where (α_1,γ_1) and (α_2,γ_2) are two solutions of the same Eqs. (4.9) and (4.6) with different integration constants. $(\alpha^{(2)},\gamma^{(2)})$ satisfies (4.11) and (4.6) with M_1 replaced by $I_1(\alpha_1,\gamma_1)M_1$, etc. The same formula serves equally well for the product, $I_1(\alpha^{(2)},\gamma^{(2)})\,I_2(\alpha_1,\gamma_1)$, with (α_1,γ_1) and (α_2,γ_2) satisfying (4.11) and (4.6). Consequently, the formula can be iterated so that all of the α and γ variables in an arbitrary product, $\cdots I_2I_1I_2I_1$, with arbitrary integration constants, can be expressed algebraically in terms of (α_1,γ_1), (α_2,γ_2), (α_3,γ_3), etc., all of which satisfy (4.9) and (4.6) with same M_i,N_i but with different integration constants. An interesting special case occurs when we put $\alpha_1 = \gamma_1 = 1$. We find

$$\alpha[\text{satisfying (4.11)}] = \frac{\alpha-\gamma}{\alpha-1}, \tag{4.18}$$

where the α on the right-hand side satisfies (4.9), and vice versa.

Now, in Ref. 11, the special product,

$$I_2\left(\frac{\alpha-\gamma}{\gamma(\alpha-1)},\gamma^{-1}\right)I_1(\alpha,\gamma) = I_1(\alpha/\gamma,\gamma^{-1})I_2\left(\frac{\alpha-\gamma}{\alpha-1},\gamma\right), \tag{4.19}$$

was identified with Harrison's Bäcklund transformation.[18] According to one sign convention, the variables, t, u, v and w, of Harrison are related to Neugebauer's variables by

$$t = 4M_1 - 2M_3, \quad u = 4N_2 - 2N_3,$$
$$v = 4M_2 - 2M_3, \quad w = 4N_1 - 2N_3, \tag{4.20}$$

and the coordinate ζ and pseudopotential $q_{(H)}$ are related to Neugebauer's γ and α (satisfying (4.9)) by

$$\zeta = \gamma^{1/2}, \quad \alpha = \frac{\zeta(1+\zeta q_{(H)})}{\zeta + q_{(H)}}. \tag{4.21a,b}$$

According to a different sign convention,[45] M_1 would be replaced by M_2, N_1 by N_2 in Eq. (4.20) and $q_{(H)}$ would be replaced by $1/q_{(H)}$ in (4.21b). In either case, it is easy to check that Eq. (4.9) transforms into Harrison's total Riccati equation for $q_{(H)}$ and that the transformation equations for M_1, etc., agree. It is convenient to choose a different pseudopotential q defined by

$$q = -\frac{1+\zeta q_{(H)}}{\zeta + q_{(H)}} \tag{4.22}$$

and satisfying the total Riccati equation,

$$dq = [-M_1q(1+\zeta q) + M_2(q+\zeta)]dx^1 + [-N_2q(1+q\zeta^{-1}) + N_1(q+\zeta^{-1})]dx^2. \tag{4.23}$$

Then Eqs. (4.21a,b) become

$$\gamma = \zeta^2, \quad \alpha = -\zeta q. \tag{4.24a,b}$$

We shall write $H(\zeta,q)$ for the product transformation on either side of (4.19).

Now, a composition theorem for a product of two Harrison transformations, $H(\zeta',q')H(\zeta_1,q_1)$, may be derived as a special case of the composition theorem for four Neugebauer transformations. The product under consideration is

$$I_2\left(\frac{\alpha'-\gamma'}{\gamma'(\alpha'-1)},\gamma'^{-1}\right)I_1(\alpha',\gamma')I_2\left(\frac{\alpha-\gamma}{\gamma(\alpha-1)},\gamma^{-1}\right)I_1(\alpha,\gamma). \tag{4.25}$$

Here, $\alpha' = -\zeta_2q'$, $\gamma' = \zeta'^2 = \zeta_2^2$, $\alpha = -\zeta_1q_1$, $\gamma = \zeta_1^2$. We seek an algebraic formula for α' in terms of (α,γ) and (α_2,γ_2), $\gamma_2 = \zeta_2^2$, $\alpha_2 = -\zeta_2q_2$, which both satisfy (4.9) and (4.6) with same M_i, N_i but different integration constants.

Two iterations of (4.17) give

$$\alpha' = \frac{\gamma^{(2)}\alpha'^{(2)} - \alpha^{(2)}\gamma'^{(2)}}{\gamma^{(2)}(\alpha'^{(2)} - \alpha^{(2)})}, \quad \gamma' = \gamma'^{(2)}/\gamma^{(2)},$$

where $(\alpha^{(2)},\gamma^{(2)})$ is given by (4.15a) and $(\alpha'^{(2)},\gamma'^{(2)})$ is given by (4.17) with $(\alpha_1,\gamma_1) = (\alpha,\gamma)$. Thus,

$$\alpha' = \frac{\gamma_2(\gamma-1)\alpha + \gamma(1-\gamma_2)\alpha_2 + (\gamma_2-\gamma)\alpha\alpha_2}{\alpha[(\gamma-\gamma_2) + (\gamma_2-1)\alpha + (1-\gamma)\alpha_2]}.$$

Hence the composition theorem for Harrison transformations is

$$q' = \frac{\zeta_2(1-\zeta_1^2)q_1 - \zeta_1(1-\zeta_2^2)q_2 + (\zeta_2^2-\zeta_1^2)q_1q_2}{q_1[(\zeta_1^2-\zeta_2^2) + \zeta_1(1-\zeta_2^2)q_1 - \zeta_2(1-\zeta_1^2)q_2]}. \tag{4.26}$$

This q' satisfies (4.23) with $\zeta = \zeta_2$ and M_1 replaced by $H(\zeta_1,q_1)M_1$, etc. The same formula is obeyed by $q_{(H)}$. This formula can be iterated so that multiple Harrison transformations can be calculated by algebraic operations alone once (4.23) has been solved for the first transformation. In Ref. 11, explicit expressions for the transforms of f, ω, ψ and $e^{2\gamma}$ are given for the single and double Harrison transformations. These will not be written down here as the transforms of any potential in the hierarchy (excluding $e^{2\gamma}$) can be obtained directly from the transformation formulas for the generating function $F_{AB}(t)$ given below and in Sec. 5. The transforms of $e^{2\gamma}$ in terms of generating functions is given in Sec. 5.

In Appendix B, we show that the double Harrison transformation preserves asymptotic flatness. Hence an even number of Harrison transformations preserves asymptotic flatness. Since it is not difficult to solve (4.23) for the Weyl solutions, we can write down closed-form expressions for stationary solutions with all mass multipoles and any finite number of angular momentum multipoles chosen arbitrarily.

Let us now turn to the description of I_1, I_2 and H in terms of the groups, **PQ** and $\mathbf{L\tilde{Q}}$, and then bring in the generating functions. First, if Eqs. (4.5c,f) and (4.8c,f) are integrated in order to express V' and Z' in terms of V and Z, we obtain two more constants of integration (corresponding to trivial **R** and **Z** transformations) and the resulting expressions are the same as (2.2) with (ρ,z) replaced by (V,Z) and $s=(4\delta_4)^{-1}\delta_3$. Now, to identify I_1 with **PQ**, it is necessary and sufficient[15] to show that I_1 leaves invariant the metric coefficient $e^{2\gamma}$ (not to be confused with Neugebauer's coordinate). Einstein's equations involving $e^{2\gamma}$ can be written,

$$\gamma_{,1}=M_1M_2/M_3,\quad \gamma_{,2}=N_1N_2/N_3,$$
$$\gamma_{,1,2}=-\tfrac{1}{2}(M_1N_1+M_2N_2), \tag{4.27}$$

and, clearly, $e^{2\gamma'}=e^{2\gamma}$ on account of (4.8). Similary, $I_2=\mathbf{L\tilde{Q}}$ leaves invariant $V^{1/2}f^{-1}e^{2\gamma}$.

In future, we shall restrict the I_1 transformation to the form $(P)_\alpha(Q)_{4s}$ and I_2 to the form $(L)_\beta(\tilde{Q})_{4s}$, the use of s here agreeing with (4.7). Now, $(P)_\alpha$ also contains two trivial parameters. The non-trivial NUT parameter is the constant of integration arising from (4.9). The trivial parameters in the **U** and **B** (or **A**) subgroups arise from the integrations needed to recover $\mathscr{E}$ and $\mathscr{E}^*$ from M_1, etc. We could further restrict I_1 and I_2 to the forms, $(P)_\lambda(Q)_{4s}$ and $(L)_\lambda(\tilde{Q})_{4s}$, respectively, but sometimes other choices for the α and β matrices need to be considered.

Now, the results in Sec. 2 allow us to calculate explicit formulas for the pseudopotentials α [satisfying (4.9) or (4.11)] and q in terms of $F_{AB}(s)$. From (2.30a) and (1.9b),

$$(P)_\alpha(Q)_{4s}H_{11}=\frac{i\alpha_4F_{11}(s)+is\alpha_3F_{12}(s)}{\alpha_2F_{11}(s)+s\alpha_1F_{12}(s)}. \tag{4.28}$$

Then, from (3.2) and (A41),[46]

$$\nabla(P)_\alpha(Q)_{4s}H_{11}=-\frac{s^2S^{-3}(s)\nabla_sH_{11}}{[\alpha_2F_{11}(s)+s\alpha_1F_{12}(s)]^2}. \tag{4.29}$$

Then, from (4.28), (4.29), (4.5b,d) and (4.8b,d), the formula for α satisfying (4.9) is found to be

$$\alpha=-\zeta(s)\frac{\alpha_2F_{11}(s)+s\alpha_1F_{12}(s)}{\alpha_2F^*{}_{11}(s)+s\alpha_1F^*{}_{12}(s)}. \tag{4.30}$$

Similarly, from (2.18), (3.5), (1.7), and Appendix A,

$$(L)_\beta(\tilde{Q})_{4s}H_{AB}=is^{-1}(b_A{}^Cb_B{}^DG_{CD}(s,s)-\epsilon_{AB}), \tag{4.31}$$

$$\begin{aligned}(L)_\beta(\tilde{Q})_{4s}f_{AB}&=-S^{-1}(s)b_A{}^Cb_B{}^Df_{XY}F^X{}_C(s)F^Y{}_D(s)\\&=(4s^2f)^{-1}Sb_{(A}{}^Cb_{B)}{}^D(\zeta F^*{}_{1C}+F_{1C})\\&\quad\times(\zeta^{-1}F^*{}_{1D}+F_{1D}),\end{aligned} \tag{4.32}$$

$$\begin{aligned}&(\partial/\partial x^1)((L)_\beta(\tilde{Q})_{4s}H_{AB})\\&\quad=(2s^2f)^{-1}b_A{}^Cb_B{}^D(\zeta F^*{}_{1C}+F_{1C})\\&\quad\times[SM_2(\zeta F^*{}_{1D}+F_{1D})-2is\zeta V_{,1}F_{1D}].\end{aligned} \tag{4.33}$$

The $AB=11$ components of (4.32) and (4.33), together with (4.12b), give

$$\begin{aligned}\alpha&=\frac{b_1{}^D(\zeta(s)F^*{}_{1D}(s)+F_{1D}(s))}{b_1{}^D(\zeta^{-1}(s)F^*{}_{1D}(s)+F_{1D}(s))}\\&=\frac{\beta_1F_{11}(s)+\beta_2F_{12}(s)+\zeta(s)(\beta_1F^*{}_{11}(s)+\beta_2F^*{}_{12}(s))}{\beta_1F_{11}(s)+\beta_2F_{12}(s)+\zeta^{-1}(s)(\beta_1F^*{}_{11}(s)+\beta_2F^*{}_{12}(s))},\end{aligned} \tag{4.34}$$

which satisfies (4.11). These two α pseudopotentials are related by Eq. (4.18) when the parameters are related by $\beta_2/\beta_1=s\alpha_1/\alpha_2$. Compare this with the fact that these two α pseudopotentials are (I^*) duals of each other when $\beta_1=\alpha_1$, $\beta_2=i\alpha_2$, $\beta_3=-i\alpha_3$, $\beta_4=\alpha_4$. [As mentioned in Sec. 2, Neugebauer's commutation theorem can be used to deduce Eqs. (2.30a,b) from (2.17). The steps are: deduce (4.34) from (2.17) and (1.7) as we have just done; use (4.18) to deduce Eq. (4.30), obtaining the β matrix elements appropriate to the $(Q)_{4s}$ transformation from the infinitesimal limit; then Eqs. (3.1) and (3.3) follow algebraically from Eq. (4.5) here and (5.4) of Ref. 11; finally, the group property $(Q)_{4t}(Q)_{4s}=(Q)_{4(s+t)}$ applied to (3.1) and (3.3) gives (2.30a,b).]

From (4.24b) and (4.30), the formula for the pseudopotential q in $H(\zeta,q)$ is

$$q=\frac{F_{12}(s)+cF_{11}(s)}{F^*{}_{12}(s)+cF^*{}_{11}(s)}, \tag{4.35}$$

where c is the constant of integration arising from (4.23). If $H(\zeta,q)=(L)_\beta(\tilde{Q})_{-4s}(P)_\alpha(Q)_{4s}$, then

$$\alpha_2\beta_2=s\beta_1\alpha_4,\quad c=\alpha_2/s\alpha_1. \tag{4.36}$$

If $H(\zeta,q)=(P)_{\bar{\alpha}}(Q)_{-4s}(L)_{\bar{\beta}}(\tilde{Q})_{4s}$, then

$$\bar{\alpha}_2\bar{\beta}_2=s\bar{\alpha}_1\bar{\beta}_4,\quad c=\bar{\beta}_1/\bar{\beta}_2. \tag{4.37}$$

The cases, $c=0$ and $c=\infty$, are important special cases where the Harrison transformation maps Weyl solutions to Weyl solutions.

Equations (4.28) and (4.31) are special cases of general formulas given in Secs. 1 and 2. We can write down the general transforms of $F_{AB}(t)$ under I_1, I_2 and H as algebraic expressions. These can be iterated easily and eliminate the need for composition theorems such as (4.17) and (4.26). From (2.30) and (1.12),

$$\begin{aligned}&(P)_\alpha(Q)_{4s}F_{11}(t)\\&\quad=i\frac{\alpha_4t(s+t)^{-1}F_{11}(s+t)+\alpha_3tF_{12}(s+t)}{\alpha_2s^{-1}F_{11}(s)+\alpha_1F_{12}(s)},\end{aligned} \tag{4.38a}$$

$$\begin{aligned}&(P)_\alpha(Q)_{4s}F_{12}(t)\\&\quad=i\frac{\alpha_2(s+t)^{-1}F_{11}(s+t)+\alpha_1F_{12}(s+t)}{\alpha_2s^{-1}F_{11}(s)+\alpha_1F_{12}(s)}.\end{aligned} \tag{4.38b}$$

From (2.17) and (1.7),

$$(L)_\beta(\tilde{Q})_{4s}F_{AB}(t)=-iS(s)b_A{}^Cb_B{}^DF_{XC}(s)F^X{}_D(s+t). \tag{4.39}$$

We need to eliminate F_{21} and F_{22} from (4.39) using (A39). The result is

$$(L)_\beta(\tilde{Q})_{4s}F_{11}(t)$$

$$= \frac{tS(s)}{2s(s+t)f} A(s)A(s+t) - \frac{S(s)S(s+t)}{2(s+t)f} \times A(s)A^*(s+t) + \frac{S^2(s)}{2sf} A^*(s)A(s+t), \quad (4.40a)$$

$$(L)_\beta(\tilde{Q})_{4s}F_{12}(t) = \frac{tS(s)}{2s(s+t)f} A(s)B(s+t) - \frac{S(s)S(s+t)}{2(s+t)f} \times A(s)B^*(s+t) + \frac{S^2(s)}{2sf} A^*(s)B(s+t), \quad (4.40b)$$

where

$$A(t) = \beta_1 F_{11}(t) + \beta_2 F_{12}(t), \quad B(t) = \beta_3 F_{11}(t) + \beta_4 F_{12}(t).$$

As an application of Eqs. (4.38) and (4.40), let us determine the precise constraints on the parameters in the following statement of Neugebauer's commutation theorem:

$$H(\zeta,q) = (L)_\beta(\tilde{Q})_{-4s}(P)_\alpha(Q)_{4s} = (P)_{\bar{\alpha}}(Q)_{-4s}(L)_{\bar{\beta}}(\tilde{Q})_{4s}. \quad (4.41)$$

It is a straightforward calculation, so we just give the results. Apart from the four unit-determinant conditions and the three constraints implied by (4.36) and (4.37), there are also the following eight constraints:

$$\alpha_3\beta_3 = \bar{\alpha}_3\bar{\beta}_3 = 0, \quad (4.42a)$$

$$\frac{\bar{\alpha}_4\bar{\beta}_1}{\beta_1\alpha_4} = \frac{\bar{\alpha}_3\bar{\beta}_4}{\beta_1\alpha_3} = \frac{\bar{\alpha}_4\bar{\beta}_2}{\beta_2\alpha_1} = \frac{\bar{\alpha}_1\bar{\beta}_3}{\beta_3\alpha_4} = \frac{\bar{\alpha}_2\bar{\beta}_1}{\beta_4\alpha_2} = \frac{\bar{\alpha}_1\bar{\beta}_4}{\beta_4\alpha_1} = \pm 1. \quad (4.42b)$$

There are plenty of ways to solve these equations, but no particular solution (depending only on the nontrivial parameters s and c) seems to serve best in all cases. The simplest choice is

$$\alpha = \bar{\alpha} = \begin{pmatrix} 1 & cs \\ 0 & 1 \end{pmatrix}, \quad \beta = \bar{\beta} = \begin{pmatrix} 1 & c^{-1} \\ 0 & 1 \end{pmatrix} \quad (4.43)$$

[cf. (1.8c) and (1.10c)], but is not suitable for the special limits, $c = 0$ and $c = \infty$. With this choice, the Harrison transforms of $F_{11}(t)$ and $F_{12}(t)$ are given by

$$F'_{11}(t) = \left[-\frac{itS(s)}{2cs(s-t)fq} + \frac{i}{2csf} \right](F_{11}(t) + c^{-1}F_{12}(t)) + \frac{iS(t)}{2c(s-t)f}(F^*_{11}(t) + c^{-1}F^*_{12}(t)), \quad (4.44a)$$

$$F'_{12}(t) = \left[-\frac{itS(s)}{2cs(s-t)fq} + \frac{i}{2csf} \right](cst^{-1}F_{11}(t) + F_{12}(t)) + \frac{iS(t)}{2c(s-t)f}(cst^{-1}F^*_{11}(t) + F^*_{12}(t)), \quad (4.44b)$$

with q given by (4.35). These formulas can be made to serve also for the $c = 0$ and $c = \infty$ limits (not both at once) by applying trivial **A** and **U** transformations.

We defer further discussion of the Harrison transformation until Sec. 5 where we will have elegant tensor equations for products of the form $(I^c)H$.

We close this section by considering recent eigenvalue problems proposed by Neugebauer and Kramer[19] and Maison.[23] Neugebauer and Kramer have rewritten Eq. (4.9) for α as a pair of coupled linear equations for variables ψ and χ satisfying $\alpha = -\zeta\psi/\chi$, $\zeta = \zeta(s)$. Their equations are

$$d\psi = M_2(\psi + \zeta\chi)dx^1 + N_1(\psi + \zeta^{-1}\chi)dx^2, \quad (4.45a)$$

$$d\chi = M_1(\chi + \zeta\psi)dx^1 + N_2(\chi + \zeta^{-1}\psi)dx^2. \quad (4.45b)$$

Explicit formulas for ψ and χ are

$$\psi = S(s)[\alpha_2 F_{11}(s) + s\alpha_1 F_{12}(s)], \quad \chi = S(s)[\alpha_2 F^*_{11}(s) + s\alpha_1 F^*_{12}(s)]. \quad (4.46)$$

These are very easy to prove directly from Eq. (3.2), and provide a quick proof of Eq. (4.30). Maison[23] discovered a one-parameter transformation group and deduced from it a linear eigenvalue problem. His coordinates and matrix variables are easily transcribed into our notation [see (3.11), (3.13) with (ρ,z) replaced by (V,Z), and (A36b)]:

$$\tau = V, \quad \sigma = Z, \quad f = -s/T_1^2(s), \quad \gamma = \zeta(s),$$

$$\mu = f_A{}^B, \quad \omega = -\psi_A{}^B, \quad \mathbf{V}(s) = P_A{}^B(s), \quad \mathbf{U}(s) = -Q_A{}^B(s),$$

$$\bar{\mu} = (1/V)f_A{}^B, \quad \bar{\mathbf{U}}(s) = -\frac{S(s)}{T_1(s)} Q_A{}^B(s),$$

$$\bar{\mathbf{U}}(s)^{-1} = \frac{S(s)}{T_1(s)} Q^B{}_A(s).$$

Equations (3.5a,b) of Ref. 23,

$$\bar{\mu}^{(s)} = \bar{\mathbf{U}}(s)^{-1}\bar{\mu}\bar{\mathbf{U}}(s), \quad \tau^{(s)} = \left(\frac{\gamma(s) - \gamma(s)^{-1}}{4s\tau}\right)^2 \tau,$$

become

$$V'^{-1}f'_{AB} = \frac{S^2(s)}{VT_1^2(s)} f_{XY}Q^X{}_A(s)Q^Y{}_B(s), \quad (4.47a)$$

$$V' = -\frac{V}{S^2(s)}. \quad (4.47b)$$

From Eqs. (A38) and (4.47b), Eq. (4.47a) can be rewritten

$$f'_{AB} = +S^{-1}(s)f_{XY}F^X{}_A(s)F^Y{}_B(s).$$

This is precisely Eq. (3.5) above for the $(\tilde{Q})_{4s}$ transformation except for the plus sign (a trivial reflection). Equation (4.47b) also agrees with (2.2a) and (2.3a) for the $(\tilde{Q})_{4s}$ transformation except for sign.

5. RELATIONSHIP BETWEEN HARRISON TRANSFORMATIONS AND BELINSKY–ZAKHAROV SOLITON TRANSFORMATIONS

Belinsky and Zakharov[20] (BZ) have rewritten Einstein's equations for the cylindrical wave field (equivalent to the stationary axisymmetric field) in terms of a linear eigenvalue problem and then solved the latter by the inverse scattering technique. According to the analytic properties of certain functions which arise with respect to the spectral parameter, the transformations are classified as "soliton type" or "nonsoliton type." The former are Bäcklund transformations which we will show to be closely related to the Harrison and HKX transformations, and the latter are expressed implicitly by a linear integral equation very similar in form to the integral equation of Hauser and Ernst, and

therefore corresponding to (almost) the full Geroch group of transformations.

In order to start, we need to transcribe Belinsky and Zakharov's coordinates and field variables, denoted here with the superscript (BZ), into our own. Let us make the identifications,[47]

$$(t,x,y,z)^{(BZ)} = (z,t,\phi,i\rho), \tag{5.1a}$$

$$(\zeta,\eta)^{(BZ)} = (\tfrac{1}{2}(i\rho+z),\tfrac{1}{2}(i\rho-z)) = (\tfrac{1}{2}ix^2,\tfrac{1}{2}ix^1), \tag{5.1b}$$

$$(\alpha,\beta)^{(BZ)} = (i\rho,z), \tag{5.1c}$$

$$\mathbf{g}^{(BZ)} = f_{AB}, \tag{5.1d}$$

$$\{\mathbf{A},\mathbf{B}\}^{(BZ)} = \left\{2\rho^{-1}\frac{\partial f_{AX}}{\partial x^2}f^{XB}, -2\rho^{-1}\frac{\partial f_{AX}}{\partial x^1}f^{XB}\right\}, \tag{5.1e}$$

$$\{\mathscr{D}_1,\mathscr{D}_2\}^{(BZ)} = \left\{-2i\frac{\partial}{\partial x^2} - \frac{2\lambda}{\lambda-i\rho}\frac{\partial}{\partial\lambda}, \; -2i\frac{\partial}{\partial x^1} + \frac{2\lambda}{\lambda+i\rho}\frac{\partial}{\partial\lambda}\right\}. \tag{5.1f}$$

The extra variable λ in (5.1f) is the spectral parameter and will be seen to correspond to t in $F_{AB}(t)$. The actual relationship can be determined simply from the observation that differentiations with respect to t do not occur in the eigenvalue problems of Neugebauer and Kramer[19] and Maison.[23] Thus we can state

$$\lambda^{(BZ)} = (2t)^{-1}[1-2tz-S(t)], \tag{5.2a}$$

$$t = \lambda(\alpha^2+2\lambda\beta+\lambda^2)^{-1}. \tag{5.2b}$$

The BZ inverse scattering technique may be summarized as follows. Given a seed metric $\mathbf{g} = \mathbf{g}(\zeta,\eta)$ satisfying $\det\mathbf{g} = \alpha^2$, solve the compatible equations,

$$\mathscr{D}_1\psi = (\lambda-\alpha)^{-1}\mathbf{A}\psi, \quad \mathscr{D}_2\psi = (\lambda+\alpha)^{-1}\mathbf{B}\psi, \tag{5.3}$$

for the matrix function $\psi = \psi(\lambda)$, with initial data $\psi(0) = \mathbf{g}$. Next, attempt to construct a new solution $\mathbf{g}'$ by the ansatz:

$$\psi'(\lambda) = \chi(\lambda)\psi(\lambda). \tag{5.4}$$

The matrix function χ must satisfy

$$\mathscr{D}_1\chi = (\lambda-\alpha)^{-1}(\mathbf{A}'\chi - \chi\mathbf{A}), \tag{5.5}$$

$$\mathscr{D}_2\chi = (\lambda+\alpha)^{-1}(\mathbf{B}'\chi - \chi\mathbf{B}),$$

$$\chi(\infty) = \mathbf{I}, \quad \mathbf{g}' = \chi(\alpha^2/\lambda)\mathbf{g}\chi^T(\lambda). \tag{5.6}$$

The method of solution of (5.5) depends on the choice of analytic behavior of $\chi(\lambda)$ in the complex λ plane. If χ is taken to be analytic inside the circle $|\lambda|^2 = \alpha^2$, then (5.5) can be reformulated as a Hilbert problem leading to an integral equation whose kernel contains an arbitrary symmetric matrix function. This so-called "nonsoliton" part of the transformation accounts for most of the Geroch group (generalized HKX transformations are excluded since χ has double poles—see Sec. 6 below). If χ is taken to be a rational function of λ, then (5.5) can be solved in closed form. These are the pure soliton tranformations. The authors also show briefly how the two types can be combined.

First, a technical point must be clarified. The methods of BZ lead to new metric functions $\mathbf{g}'$ which do not obey $\det\mathbf{g}' = \alpha^2$ and therefore do not satisfy one of Einstein's equations. They get around this problem very neatly by defining a physical $\mathbf{g}'$ matrix by

$$\mathbf{g}'_{ph} = \alpha(\det\mathbf{g}')^{-1/2}\mathbf{g}'. \tag{5.7}$$

This satisfies all of Einstein's equation including $\det\mathbf{g}'_{ph} = \alpha^2$. However, as we wish to iterate the BZ transformations, we need one extra piece of information—a prescription for ψ'_{ph} constructed from (5.3) with $\mathbf{g}$ replaced by $\mathbf{g}'_{ph}$. A first integral of (5.3) is

$$\det\psi(\lambda) = (\lambda/t)f^2(t), \tag{5.8}$$

where $f(t)$ is an arbitrary function of t only, subject to $f(0) = 1$, and may be taken to be identically unity. The formulas

$$\psi'_{ph} = (\lambda/t)^{1/2}(\det\psi')^{-1/2}\psi' \tag{5.9a}$$

$$= (\det\chi)^{-1/2}\chi\psi \tag{5.9b}$$

may be proved in the same manner as (5.7).

The key step in this discussion is to derive a formula for $\psi(\lambda)$ in terms of the generating function $P_{AB}(t) = \frac{1}{2}(F_{AB}+F^*{}_{AB})$. First, write[47]

$$\psi(\lambda) = \theta_{AB}(t), \quad \chi(\lambda) = \chi_A{}^B(t). \tag{5.10}$$

In our notation, Eqs. (5.3) become

$$\nabla\theta_{AB} = -(2\rho^2S)^{-1}[\nabla_t f_{AX} + S\nabla f_{AX}]f^{XY}\theta_{YB}, \tag{5.11}$$

where $\nabla_t = (1-2tz)\nabla - 2t\rho\tilde{\nabla}$. Now, the qualitatively similar Eq. (A17) for F_{AB} can be rewritten in terms of P_{AB} using (A35b). Using also (A8a,b) and (A9a), we obtain

$$\begin{aligned}\nabla P_{AB} &= (\rho S)^{-2}[-T_1{}^2f^{XY}\nabla_t f_{AX} + \rho t f_{AX}\tilde{\nabla}_t f^{XY}]P_{YB} \\ &= -(2\rho^2S)^{-1}[\nabla_t f_{AX} + S\nabla f_{AX}]f^{XY}P_{YB} \\ &\quad - 2tS^{-2}(\tilde{\nabla}_t\rho)P_{AB}.\end{aligned} \tag{5.12}$$

A comparison of (5.11) with (5.12) reveals the relationship

$$\theta_{AB}(t) = t^{-1}f(t)S(t)P_{AB}(t), \tag{5.13}$$

where $f(t)$ is an arbitrary function of t only, subject to $f(0) = 1$. This is the same $f(t)$ which appears in (5.8) and will be taken to be identically unity in future.

In the case of the BZ n-soliton transformation, $\chi_A{}^B$ assumes both the following forms:

$$\chi_A{}^B = \epsilon_A{}^B + \sum_{k=1}^{n}\frac{R_{kA}{}^B}{\lambda-\mu(s_k)}, \tag{5.14a}$$

$$-(\det\chi)^{-1}\chi^B{}_A = \epsilon_A{}^B + \sum_{k=1}^{n}\frac{S_{kA}{}^B}{\lambda-\nu(s_k)}, \tag{5.14b}$$

where the R_k and S_k matrices (tensors) do not depend on λ (or t) and

$$\mu(s) = (2s)^{-1}[1-2sz-S(s)] = -s^{-1}T_2{}^2(s), \tag{5.15a}$$

$$\nu(s) = (2s)^{-1}[1-2sz+S(s)] = s^{-1}T_1{}^2(s) = -\rho^2/\mu(s). \tag{5.15b}$$

[Note that $\lambda = \mu(t)$.] Equation (5.14b) follows from (5.14a) or vice versa on account of (5.6). A simple but important consequence of the compatibility of (5.14a,b) is

$$\det\chi = \frac{[\lambda-\nu(s_1)][\lambda-\nu(s_2)]\cdots[\lambda-\nu(s_n)]}{[\lambda-\mu(s_1)][\lambda-\mu(s_2)]\cdots[\lambda-\mu(s_n)]}. \tag{5.16}$$

The determination of the explicit form of the R_k and S_k matrices depends only on the limiting form of Eqs. (5.5) and

(5.6) at the points $\lambda=\mu(s_k)$ and $\lambda=\nu(s_k)$. The reader is referred to Ref. 20 for the details.

With the aid of the three results (5.9b), (5.13) and (5.16) above, it is a straightforward matter to transcribe the BZ one-soliton and two-soliton transformation formulas into the language of generating functions. Their matrix equations automatically go over into SL(2) tensor equations. The one-soliton tranformation law becomes

$$P'_{AB}(t)=\left(\frac{\lambda-\mu}{\lambda-\nu}\right)^{1/2}\left[\epsilon_A{}^X-\left(\frac{\nu-\mu}{\lambda-\mu}\right)\times\left(\frac{h^E h^F P^C{}_E(s)P^X{}_F(s)f_{CA}}{h^E h^F f^{CD}P_{CE}(s)P_{DF}(s)}\right)\right]P_{XB}(t), \tag{5.17}$$

with $\lambda=\mu(t)$, $\mu=\mu(s)$, $\nu=\nu(s)$. Here, h^E is an arbitrary constant vector, but the right-hand side of (5.17) only depends on the ratio h^2/h^1. Choose the branch of the square root according to

$$(\lambda-\nu)\left(\frac{\lambda-\mu}{\lambda-\nu}\right)^{1/2}=-\frac{\epsilon i\rho}{T_1(t)}(1-t/s)^{1/2},\quad \epsilon=\pm 1. \tag{5.18}$$

Since $f_{AB}=DP_{AB}(0)$, we have

$$f'_{AB}=-\epsilon i\rho^{-1}\mu\left[f_{AB}+\frac{\nu-\mu}{\mu}\frac{h^E h^F P^C{}_E(s)P^D{}_F(s)f_{CA}f_{DB}}{h^E h^F f^{CD}P_{CE}(s)P_{DF}(s)}\right]. \tag{5.19}$$

Then from (A35b),

$$Q'_{AB}(t)=-\epsilon i\rho^{-1}\left(\frac{\lambda-\mu}{\lambda-\nu}\right)^{1/2}\left[\frac{\mu(\lambda-\nu)}{\lambda-\mu}\epsilon_A{}^X+\frac{\lambda(\nu-\mu)}{\lambda-\mu}\frac{h^E h^F P^C{}_E(s)P^X{}_F(s)f_{CA}}{h^E h^F f^{CD}P_{CE}(s)P_{DF}(s)}\right]Q_{XB}(t). \tag{5.20}$$

To achieve this form, we have used the identities (A4) and (A5). These identities, and corollaries such as (A6), will be used extensively throughout this section and Sec. 6 and we will not refer to them every time. Notice that P_{AB} and Q_{AB} are the natural variables for this analysis—the corresponding formulas written in terms of F_{AB} are rather inelegant. Notice also that the transformation does not respect the reality of f_{AB}.

Let us use the notation,

$$(\mathrm{BZ}{:}s,h^A), \tag{5.21}$$

for the transformation given by (5.17), (5.19) and (5.20). The formula for two successive one-soliton transformations,

$$(\mathrm{BZ}{:}s_2,k^A)(\mathrm{BZ}{:}s_1,h^A),$$

may be obtained by iterating (5.17). After a short calculation, we find

$$P''_{AB}(t)=\left(\frac{\lambda-\mu_2}{\lambda-\nu_2}\right)^{1/2}\left[\epsilon_A{}^X-\left(\frac{\nu_2-\mu_2}{\lambda-\mu_2}\right)\times\left(\frac{k^E k^E P'^C{}_E(s_2)P'^X{}_F(s_2)f'_{CA}}{k^E k^F f'^{CD}P'_{CE}(s_2)P'_{DF}(s_2)}\right)\right]P'_{XB}(t)$$
$$=\left(\frac{\lambda-\mu_2}{\lambda-\nu_2}\right)^{1/2}\left(\frac{\lambda-\mu_1}{\lambda-\nu_1}\right)^{1/2}\times\left[\epsilon_A{}^X+\frac{-(\nu_1-\mu_1)CA_A{}^X+(\nu_2-\mu_1)B\bar{B}_A{}^X}{(\lambda-\mu_1)(AC-B\bar{B})}+\frac{(\nu_1-\mu_2)\bar{B}B_A{}^X-(\nu_2-\mu_2)AC_A{}^X}{(\lambda-\mu_2)(AC-B\bar{B})}\right]P_{XB}(t), \tag{5.22}$$

where $\mu_1=\mu(s_1)$, $\mu_2=\mu(s_2)$, $\nu_1=\nu(s_1)$, $\nu_2=\nu(s_2)$ and

$$A_A{}^X=(\nu_1-\mu_1)^{-1}h_E h_F f_{CA}P^{CE}(s_1)P^{XF}(s_1), \tag{5.23a}$$
$$B_A{}^X=(\nu_1-\mu_2)^{-1}h_E k_F f_{CA}P^{CE}(s_1)P^{XF}(s_2), \tag{5.23b}$$
$$\bar{B}_A{}^X=(\nu_2-\mu_1)^{-1}k_E h_F f_{CA}P^{CE}(s_2)P^{XF}(s_1), \tag{5.23c}$$
$$C_A{}^X=(\nu_2-\mu_2)^{-1}k_E k_F f_{CA}P^{CE}(s_2)P^{XF}(s_2), \tag{5.23d}$$

$$A=A_X{}^X,\quad B=B_X{}^X,\quad \bar{B}=\bar{B}_X{}^X=(\mu_2/\mu_1)B,\quad C=C_X{}^X. \tag{5.24}$$

This is precisely the BZ two-soliton transformation. Similarly, iterating the one-soliton n times gives the n-soliton. The symmetry of (5.22)-(5.24) shows that the BZ transformations all commute. This two-soliton transformation will now be shown to be equivalent (up to a gauge change) to the double Harrison transformation, and therefore it has the important property of preserving asymptotic flatness. (It also respects the reality of the metric.) In Sec. 6, we deduce the HKX and generalized HKX transformations as limiting forms of Eq. (5.22).

Despite the obvious advantages of having the BZ transformations in the form of tensor equations, it is also equally desirable to abandon the SL(2) covariance and write the equations in terms of P_{11}, P_{12}, Q_{11} and Q_{12} only. From (A35a,b) and (A37),

$$P_{2B}(s)=-\omega P_{1B}(s)+\mu(s)f^{-1}Q_{1B}(s), \tag{5.25a}$$
$$Q_{2B}(s)=-\omega Q_{1B}(s)-\nu(s)f^{-1}P_{1B}(s), \tag{5.25b}$$
$$P_{11}(s)Q_{12}(s)-P_{12}(s)Q_{11}(s)=sS^{-2}(s)f. \tag{5.26}$$

Hence, under (BZ: s, h^A),

$$P'_{1B}(t)=\left(\frac{\lambda-\mu}{\lambda-\nu}\right)^{1/2}\times\left[P_{1B}(t)-\left(\frac{\nu-\mu}{\lambda-\mu}\right)\left(\frac{\mu Q^2P_{1B}(t)-\lambda PQQ_{1B}(t)}{\mu Q^2+\nu P^2}\right)\right], \tag{5.27a}$$

$$Q'_{1B}(t)=\epsilon i\rho^{-1}\mu\left(\frac{\lambda-\mu}{\lambda-\nu}\right)^{1/2}\left[\left(\frac{\lambda-\nu}{\lambda-\mu}\right)Q_{1B}(t)+\left(\frac{\nu-\mu}{\lambda-\mu}\right)\left(\frac{\lambda Q^2Q_{1B}(t)+\nu PQP_{1B}(t)}{\mu Q^2+\nu P^2}\right)\right], \tag{5.27b}$$

where

$$P=h^E P_{1E}(s),\quad Q=h^E Q_{1E}(s).$$

Our aim now is to establish the relationship

$$(\mathrm{BZ}{:}s,h^A)=(\mathrm{gauge})(U)_\alpha(A)_\beta(I^\epsilon)H(\zeta,q), \tag{5.28}$$

for some α, β and $g_{AB}(t)$, where ϵ is the same as in (5.18)–(5.20) and (5.27b) and

$$\zeta=\zeta(s),\quad q=\frac{h^E F_{1E}(s)}{h^E F^*{}_{1E}(s)}=\frac{P+iQ}{P-iQ},\quad h^1/h^2=c. \tag{5.29}$$

Start with formulas (4.44a,b) for the H transformation. Use a single prime for H, a double prime for $(I^c)H$ and a triple prime for the left-hand side of (5.28). In terms of P_{1B} and Q_{1B}, (4.44) reads

$$P'_{11}(t) = -(csf)^{-1}(P^2+Q^2)^{-1} \times \left[\frac{\lambda(\nu-\mu)}{(\lambda-\mu)(\lambda-\nu)} PQ\,[P_{11}(t)+c^{-1}P_{12}(t)] + \left(\frac{\lambda P^2}{\lambda-\mu}+\frac{\lambda Q^2}{\lambda-\nu}\right)[Q_{11}(t)+c^{-1}Q_{12}(t)]\right], \tag{5.30a}$$

$$Q'_{11}(t) = -(csf)^{-1}(P^2+Q^2)^{-1} \times \left[\frac{\lambda(\nu-\mu)}{(\lambda-\mu)(\lambda-\nu)} PQ\,[Q_{11}(t)+c^{-1}Q_{12}(t)] + \left(\frac{\nu P^2}{\lambda-\nu}+\frac{\mu Q^2}{\lambda-\mu}\right)[P_{11}(t)+c^{-1}P_{12}(t)]\right], \tag{5.30b}$$

where we have used (5.15) and (5.18). Corresponding formulas for $P'_{12}(t)$ and $Q'_{12}(t)$ are obtained from (5.30a,b) by replacing $P_{11}(t)+c^{-1}P_{12}(t)$ and $Q_{11}(t)+c^{-1}Q_{12}(t)$ by $cst^{-1}P_{11}(t)+P_{12}(t)$ and $cst^{-1}Q_{11}(t)+Q_{12}(t)$, respectively. Next, $P''_{1B}(t)$ and $Q''_{1B}(t)$ are constructed from Eqs. (3.12a–d). These formulas may be directly compared with Eqs. (5.27a,b). After a little algebra, we find

$$P'''_{11}(t) = (cs)^{-1}(1-t/s)^{-1/2}[\epsilon ic^{-1}P''_{11}(t) - tP''_{12}(t)], \tag{5.31a}$$

$$P'''_{12}(t) = (cs)^{-1}(1-t/s)^{-1/2}[-\epsilon iP''_{11}(t) + csP''_{12}(t)]. \tag{5.31b}$$

$Q'''_{1B}(t)$ and $Q''_{1B}(t)$ are related in exactly the same way. Hence, so also are $F'''_{1B}(t)$ and $F''_{1B}(t)$. Now, if (5.31a,b) is to represent a (gauge)$(U)_\alpha(A)_\beta$ transformation, then, from (1.7), (1.8a,b) and (1.14),

$$g_{11}(t) = -(cs)^{-1}t(1-t/s)^{-1/2}, \quad g_{12}(t) = (1-t/s)^{-1/2}, \tag{5.32a,b}$$

$$e^\alpha g_{21}(t) + \beta g_{11}(t) = -\epsilon ic^{-2}s^{-1}(1-t/s)^{-1/2},$$
$$e^\alpha g_{22}(t) + \beta g_{12}(t) = \epsilon i(cs)^{-1}(1-t/s)^{-1/2}.$$

Because of the supplementary conditions (1.15), all quantities are determined uniquely, and we find

$$g_{21}(t) = -(1-t/s)^{1/2}, \quad g_{22}(t) = 0, \tag{5.32c,d}$$

$$e^\alpha = \epsilon ic^{-2}s^{-1}, \quad \beta = \epsilon i(cs)^{-1}. \tag{5.32e,f}$$

This completes the proof that the BZ one-soliton is equivalent to a Harrison transformation followed by (I^c). The next result shows that we can reverse the order of (I^c) and H.

A remarkable property of the BZ transformations is their self-duality. This means that Eq. (5.17) could also be written as a P–covariant tensor equation among the dual generating functions $(I^c)P_{AB}(t)$. This is the first time a nontrivial transformation has been reported to have this property.[48] The precise statement of the theorem is

$$(I^c)(\mathrm{BZ}{:}s,h^A)(I^c) = (\mathrm{BZ}{:}s,k^A), \tag{5.33}$$

where

$$k^2/k^1 = -\epsilon' ish^1/h^2. \tag{5.34}$$

It may be proved without difficulty using Eqs. (5.27) and (3.12). An immediate corollary of the previous two results is that the BZ two-soliton transformation is identical to the double Harrison transformation. Thus, although the single Harrison transformation is not expressible as a tensor equation and there is no natural choice of gauge, the double Harrison transformation does not suffer from these disadvantages.

There is a third useful form for the BZ transformation involving $F_{11}(t)$, $F_{12}(t)$ and q. Let us take $\epsilon = +1$ from now on. From (5.27) and (5.29), under (BZ: s,h^A),

$$F'_{1B}(t) = -\frac{2T_1(t)(1-t/s)^{-1/2}}{[1+\zeta(t)][1+\zeta(s)q]} \times [-\zeta(t)F_{1B}(t) + \zeta(s)qF^*_{1B}(t)], \tag{5.35a}$$

$$F^{*\prime}_{1B}(t) = -\frac{2T_1(t)(1-t/s)^{-1/2}}{[1+\zeta(t)][\zeta(s)+q]} \times [\zeta(s)F_{1B}(t) - \zeta(t)qF^*_{1B}(t)]. \tag{5.35b}$$

These formulas are well suited to iteration. A formula for q' appropriate to a second (BZ: s_2, k^A) transformation is easily found to be

$$q' = \frac{k^E F'_{1E}(s_2)}{k^E F^{*\prime}_{1E}(s_2)} = \frac{\zeta_1+q_1}{1+\zeta_1 q_1}\,\frac{\zeta_1 q_1 - \zeta_2 q_2}{\zeta_1 q_2 - \zeta_2 q_1}, \tag{5.36}$$

where

$$\zeta_1 = \zeta(s_1), \quad \zeta_2 = \zeta(s_2), \qquad q_1 = \frac{h^E F_{1E}(s_1)}{h^E F^*_{1E}(s_1)}, \quad q_2 = \frac{k^E F_{1E}(s_2)}{k^E F^*_{1E}(s_2)}. \tag{5.37}$$

This is the composition theorem for BZ transformations corresponding to Eq. (4.26) for Harrison transformations. It can be deduced from the latter using (4.18) and (4.24). (If $\epsilon = -1$, F'_{1B} and $F^{*\prime}_{1B}$ would change places in (5.35a,b) and q' would be replaced by $1/q'$.) Now, the form of the double transformation (BZ: s_2, k^A)(BZ: s_1, h^A) is

$$F''_{1B}(t) = \frac{4T_1^2(t)(1-t/s_1)^{-1/2}(1-t/s_2)^{-1/2}}{(1+\zeta(t))^2} \times \frac{\zeta(t)(\zeta_1^2-\zeta_2^2)q_1q_2F^*_{1B}(t) - [\zeta^2(t)(\zeta_1q_2-\zeta_2q_1)+\zeta_1\zeta_2(\zeta_1q_1-\zeta_2q_2)]F_{1B}(t)}{\zeta_2(1-\zeta_1^2)q_1 - \zeta_1(1-\zeta_2^2)q_2 + (\zeta_2^2-\zeta_1^2)q_1q_2}. \tag{5.38}$$

The transform of $F^*{}_{1B}(t)$ is obtained by using $\zeta^* = \zeta^{-1}$, $q^* = q^{-1}$. The relation $\mathscr{E} = DF^*{}_{11}(0)$ gives

$$\mathscr{E}'' = \mathscr{E} + \frac{2(\zeta_2^2 - \zeta_1^2)f}{\zeta_1^2 - \zeta_2^2 + \zeta_1(1-\zeta_2^2)q_1 - \zeta_2(1-\zeta_1^2)q_2}, \tag{5.39}$$

which appears in Ref. 11. This formula is the basis for the proof of asymptotic flatness given in Appendix B.

Finally, we give an explicit formula for the BZ transform of the metric coefficient $e^{2\gamma}$. This can be achieved by purely algebraic methods as follows. Factorize the BZ transformation into a product of (I) and H according to (5.28). Then factorize the Harrison transformation into a product of four factors of the form (4.41). Since $\mathbf{P}$ and $\mathbf{Q}$ leave $e^{2\gamma}$ invariant and $\mathbf{L}$ and $\tilde{\mathbf{Q}}$ leave $\rho^{1/2}e^{2\gamma-2u}$ invariant, where $e^{2u} = f$, and the (I) transform of $e^{2\gamma}$ is given by (1.20d), we can construct the BZ transform of $e^{2\gamma}$. Further, since $e^{2\gamma-2u}$ is an SL(2) scalar, and the BZ transformation law preserves tensor character, it should be possible to assemble the transform of $e^{2\gamma-2u}$ in the form of a tensor equation. Indeed, this is the case, and the final result of the lengthy calculations is the following remarkably simple formula:

$$e^{2\gamma'-2u'} = \rho^{-5/2}s^{-2}S(s)T_1^2(s)h^E h^F f^{CD}P_{CE}(s)P_{DF}(s)e^{2\gamma-2u}. \tag{5.40}$$

This can be proved by substitution into the equations,

$$\begin{aligned}(\gamma-u)_{,1} &= (4\rho)^{-1}(f_{XY})_{,1}(f^{XY})_{,1},\\ (\gamma-u)_{,2} &= (4\rho)^{-1}(f_{XY})_{,2}(f^{XY})_{,2},\end{aligned} \tag{5.41}$$

but this is also a lengthy calculation requiring many applications of (A4)-(A6). The simplest method is to abandon the manifest covariance, and instead substitute Eq. (4.7d) of Ref. 11 for the Harrison transform of $e^{2\gamma}$ into Eqs. (4.27) here (a much easier calculation) and then show that this formula, together with (5.29) and (1.20d), imply (5.40). The details will be omitted. Using the notation, (5.23a–d) and (5.24), the formula for the BZ two-soliton or double Harrison transform of $e^{2\gamma-2u}$ may be written,

$$e^{2\gamma''-2u''} = \rho^{-6}S^2(s_1)S^2(s_2)(\rho^2+\nu_1\nu_2)^2(AC-B\bar{B})e^{2\gamma-2u}. \tag{5.42}$$

6. GENERALIZED HKX TRANSFORMATION AS A CONFLUENT DOUBLE SOLITON

In Ref. 7, Hoenselaers, Kinnersley, and Xanthopoulos exponentiated the infinite linear combination of γ_{11}'s,[43]

$$\sum_{k=0}^{\infty} \alpha s^k \gamma_{11}^{(k)}, \tag{6.1}$$

and found the neat formula,

$$G'_{11}(t_1,t_2) = G_{11}(t_1,t_2) + \frac{\alpha G_{11}(t_1,s)G_{11}(s,t_2)}{1-\alpha G_{11}(s,s)}. \tag{6.2}$$

This transformation, the HKX rank-zero transformation, preserves asymptotic flatness and may be iterated to produce stationary solutions with arbitrary multipoles. The transformations commute, and the subclasses with same s parameters from one-parameter groups, α being the canonical parameter. Iteration requires either changing the s parameter at each step, or constructing higher rank confluent transformations where the s parameters coalesce. The authors have also investigated the SL(2) generalization of the HKX transformation,[49]

$$\sum_{k=0}^{\infty} \alpha s^k q^{XY}\gamma_{XY}^{(k)}, \tag{6.3}$$

where q^{XY} is an arbitrary null symmetric tensor ($q^{XY} = h^X h^Y$). In this section, we find the form of the BZ two-soliton transformation or double Harrison transformation in the limit when the poles coalesce (a double-pole soliton) and identify the result with the generalized HKX transformation. The final formulas will be written as tensor equations among the generating functions, $F_{AB}(t)$ and $G_{AB}(t_1,t_2)$.

Start with Eqs. (5.22)–(5.24) which define the (BZ: s_2, k^A) (BZ:s_1, h^A) product transformation. If $s_2 = s_1 = s$, then from (A4)–(A6),

$$AC - B\bar{B} = (\nu-\mu)^{-2}(\det P)^2(\det f)(h_Y k^Y)^2. \tag{6.4a}$$

$$\begin{aligned}CA_A{}^X - B\bar{B}_A{}^X - \bar{B}B_A{}^X + AC_A{}^X\\ = (\nu-\mu)^{-2}(\det P)^2(\det f)(h_Y k^Y)^2\epsilon_A{}^X.\end{aligned} \tag{6.4b}$$

Hence, provided $h_Y k^Y \neq 0$, we obtain the identity transformation,

$$P'_{AB}(t) = P_{AB}(t).$$

The way to obtain a non-trivial confluent transformation is to write

$$s_1 = s, \quad s_2 = s+\epsilon, \quad k^A = h^A + \epsilon l^A, \tag{6.5}$$

and take the limit as $\epsilon \to 0$. (Note ϵ does not mean ± 1 here.) The calculation will be complicated by the fact that the denominator $AC - B\bar{B}$ and the numerator of the right-hand side of (5.22) are both of order ϵ^2. Write

$$B_A{}^X = A_A{}^X + \epsilon B_{1A}{}^X + \epsilon^2 B_{2A}{}^X + O(\epsilon^3), \tag{6.6a}$$

$$\bar{B}_A{}^X = A_A{}^X + \epsilon \bar{B}_{1A}{}^X + \epsilon^2 \bar{B}_{2A}{}^X + O(\epsilon^3), \tag{6.6b}$$

$$\begin{aligned}C_A{}^X = {} & A_A{}^X + \epsilon(B_{1A}{}^X + \bar{B}_{1A}{}^X)\\ & + \epsilon^2(B_{2A}{}^X + \bar{B}_{2A}{}^X + C_{2A}{}^X) + O(\epsilon^3),\end{aligned} \tag{6.6c}$$

$$A = A_X{}^X, \quad B_1 = B_{1X}{}^X, \quad \bar{B}_1 = \bar{B}_{1X}{}^X, \quad C_2 = C_{2X}{}^X. \tag{6.7}$$

Then, from (5.23a-d),

$$A_A{}^X = (\nu-\mu)^{-1}h_E h_F f_{CA}P^{CE}P^{XF}, \tag{6.8a}$$

$$\begin{aligned}B_{1A}{}^X = {} & (\nu-\mu)^{-1}h_E f_{CA}P^{CE}(h_F DP^{XF} + l_F P^{XF})\\ & + \frac{D\mu}{\nu-\mu}A_A{}^X,\end{aligned} \tag{6.8b}$$

$$\begin{aligned}\bar{B}_{1A}{}^X = {} & (\nu-\mu)^{-1}h_F f_{CA}P^{XF}(h_E DP^{CE} + l_E P^{CE})\\ & - \frac{D\nu}{\nu-\mu}A_A{}^X,\end{aligned} \tag{6.8c}$$

$$\begin{aligned}C_{2A}{}^X = {} & (\nu-\mu)^{-1}f_{CA}(h_E DP^{CE} + l_E P^{CE})(h_F DP^{XF} + l_F P^{XF})\\ & - \frac{D\nu}{\nu-\mu}B_{1A}{}^X + \frac{D\mu}{\nu-\mu}\bar{B}_{1A}{}^X,\end{aligned} \tag{6.8d}$$

where $D = d/ds$, $P_{AB} = P_{AB}(s)$. Expressions for $B_{2A}{}^X$ and $\bar{B}_{2A}{}^X$ will not be needed. The denominator and numerator in (5.22) take the forms,

$$AC - B\bar{B} = \epsilon^2(AC_2 - B_1\bar{B}_1) + O(\epsilon^3), \tag{6.9}$$

$$\frac{-(\nu_1-\mu_1)CA_A{}^X+(\nu_2-\mu_1)B\bar{B}_A{}^X}{\lambda-\mu_1}$$
$$+\frac{(\nu_1-\mu_2)\bar{B}B_A{}^X-(\nu_2-\mu_2)AC_A{}^X}{\lambda-\mu_2}$$
$$=\epsilon^2\left\{\frac{\nu-\mu}{\lambda-\mu}(B_1\bar{B}_{1A}{}^X+\bar{B}_1B_{1A}{}^X-AC_{2A}{}^X-C_2A_A{}^X)\right.$$
$$+\frac{D\nu}{\lambda-\mu}(B_1A_A{}^X-AB_{1A}{}^X)+\frac{(\lambda-\nu)D\mu}{(\lambda-\mu)^2}(A\bar{B}_{1A}{}^X$$
$$\left.-\bar{B}_1A_A{}^X)-\frac{D\mu D\nu}{(\lambda-\mu)^2}AA_A{}^X\right\}+O(\epsilon^3). \quad (6.10)$$

Some spurious terms in (6.8b,c,d) can be removed by choosing new variables,

$$B'_A{}^X=B_{1A}{}^X-\frac{D\mu}{\nu-\mu}A_A{}^X,\quad \bar{B}'_A{}^X=\bar{B}_{1A}{}^X+\frac{D\nu}{\nu-\mu}A_A{}^X, \quad (6.11a,b)$$

$$C'_A{}^X=C_{2A}{}^X+\frac{D\nu}{\nu-\mu}B_{1A}{}^X-\frac{D\mu}{\nu-\mu}\bar{B}_{1A}{}^X, \quad (6.11c)$$

$$B'=B'_X{}^X=\bar{B}'_X{}^X,\quad C'=C'_X{}^X. \quad (6.12)$$

Now, from (5.22), (6.9) and (6.10), the limiting form of the transformation is

$$P'_{AB}(t)=\frac{\lambda-\mu}{\lambda-\nu}P_{AB}(t)+\left[(AC'-B'^2)-\frac{D\mu D\nu}{(\nu-\mu)^2}A^2\right]^{-1}$$
$$\times\left\{\frac{(2\lambda-\mu-\nu)D\mu D\nu}{(\nu-\mu)(\lambda-\mu)(\lambda-\nu)}AA_A{}^X\right.$$
$$+\frac{D\nu}{\lambda-\nu}(B'A_A{}^X-AB'_A{}^X)$$
$$-\frac{D\mu}{\lambda-\mu}(B'A_A{}^X-A\bar{B}'_A{}^X)$$
$$-\frac{\nu-\mu}{\lambda-\nu}(C'A_A{}^X+AC'_A{}^X$$
$$\left.-B'B'_A{}^X-B'\bar{B}'_A{}^X)\right\}P_{XB}(t). \quad (6.13)$$

Equation (6.13) is rather cumbersome, but can be reduced to a very elegant expression involving the complex generating functions F_{AB} and G_{AB} by a lengthy series of algebraic steps. Let us treat the denominator first.

$$AC'-B'^2=(\nu-\mu)^{-2}[h_Eh_Ff_{CD}P^{CE}P^{DF}f_{XY}(h_UDP^{XU}$$
$$+l_UP^{XU})(h_VDP^{YV}+l_VP^{YV})-h_Ef_{CD}P^{CE}$$
$$\times(h_FDP^{DF}+l_FP^{DF})h_Vf_{XY}P^{YV}$$
$$\times(h_UDP^{XU}+l_UP^{XU}]$$
$$=(\nu-\mu)^{-2}h_Ef_{CD}f_{XY}P^{CE}(h_UDP^{XU}+l_UP^{XU})$$
$$\times[h_FP^{DF}(h_VDP^{YV}+l_VP^{YV})$$
$$-h_VP^{YV}(h_FDP^{DF}+l_FP^{DF})]$$
$$=(\nu-\mu)^{-2}h_Ef_{CD}f_X{}^DP^{CE}(h_UDP^{XU}+l_UP^{XU})$$
$$\times h_VP_Z{}^V(h_FDP^{ZF}+l_FP^{ZF})$$
$$=-\rho^2(\nu-\mu)^{-2}[h_EP_X{}^E(h_UDP^{XU}+l_UP^{XU})]^2$$
$$=-\rho^2(\nu-\mu)^{-2}[h^Eh^FP_{XE}DP^X{}_F$$
$$+s\mu S^{-2}h_Xl^X]^2, \quad (6.14)$$

where we have used (A4), (A5), (A9a), (A36a), and (5.15a). (In future, we refer only to identities which specifically involve the generating functions.) The qualitative similarity between $P_{XE}DP^X{}_F$ in (6.14) and $F_{XA}DF^X{}_B$ which appears in (A25) suggests that we should look for an expression involving $G_{AB}(s,s)$. From (A25), (A34), (A35b), and (A36a),

$$G_{AB}(s,s)=(1+s^2S^{-1}D\nu)\epsilon_{AB}-\rho^{-2}S^2\nu P_{XA}DP^X{}_B$$
$$+i\rho^{-2}\nu f^{XY}P_{XA}P_{YB}. \quad (6.15)$$

Now define

$$\alpha=-(sh_Xl^X)^{-1},$$
$$M=\mathrm{Re}(1-\alpha h^Eh^FG_{EF}(s,s)), \quad (6.16)$$
$$N=\mathrm{Im}(1-\alpha h^Eh^FG_{EF}(s,s)).$$

Hence, from (6.15), we have

$$h^Eh^FP_{XE}DP^X{}_F+s\mu S^{-2}h_Xl^X=-\alpha^{-1}\mu S^{-2}M, \quad (6.17)$$

$$A=s\mu(\alpha S)^{-1}N. \quad (6.18)$$

In terms of M and N, the denominator in (6.13) is given by

$$AC'-B'^2-\frac{D\mu D\nu}{(\nu-\mu)^2}A^2=-\frac{\rho^2s^2\mu^2}{\alpha^2S^6}(M^2+N^2). \quad (6.19)$$

In the same manner that (6.14) and (6.19) were derived, we can also derive the following formulas:

$$C'A_A{}^X+AC'_A{}^X-B'B'_A{}^X-B'\bar{B}'_A{}^X$$
$$=-\rho^2(\nu-\mu)^{-2}[h^Eh^FP_{YE}DP^Y{}_F$$
$$+s\mu S^{-2}h_Yl^Y]^2\epsilon_A{}^X$$
$$=-\frac{\rho^2s^2\mu^2}{\alpha^2S^6}M^2\epsilon_A{}^X, \quad (6.20)$$

$$B'A_A{}^X-AB'_A{}^X=(\nu-\mu)^{-2}h^Eh^Ff_A{}^Cf^{XD}P_{CE}P_{DF}$$
$$\times[h^Uh^VP_{YU}DP^Y{}_V+s\mu S^{-2}h_Yl^Y]$$
$$=-\left(\frac{s^2\mu}{\alpha S^4}\right)Mh^Eh^Ff_A{}^Cf^{XD}P_{CE}P_{DF}, \quad (6.21)$$

$$B'A_A{}^X-A\bar{B}'_A{}^X=(\nu-\mu)^{-2}\rho^2h^Eh^FP_{AE}P^X{}_F$$
$$\times[h^Uh^VP_{YU}DP^Y{}_V+s\mu S^{-2}h_Yl^Y]$$
$$=-\frac{\rho^2s^2\mu}{\alpha S^4}Mh^Eh^FP_{AE}P^X{}_F. \quad (6.22)$$

Since these quantities are going to be contracted with $P_{XB}(t)$, we should next look for expressions involving $G_{AB}(s,t)$. Write

$$U_{AB}=\mathrm{Re}\left(G_{AB}(s,t)-\frac{s}{s-t}\epsilon_{AB}\right),$$
$$V_{AB}=\mathrm{Im}\left(G_{AB}(s,t)-\frac{s}{s-t}\epsilon_{AB}\right). \quad (6.23)$$

Then, from (A19), (A34), (A35b), and (5.18),

$$U_{AB}=\frac{\nu-\mu}{\lambda-\mu}P_{XA}P^X{}_B(t),$$
$$V_{AB}=\frac{\nu-\mu}{\mu(\lambda-\nu)}f^{XY}P_{YA}P_{XB}(t). \quad (6.24)$$

The results of the contractions are now

$$AA_A{}^X P_{XB}(t) = -\frac{s^3\mu^2(\lambda-\mu)}{\alpha S^3} Nh^E h^F Q_{AE} U_{FB}, \tag{6.25}$$

$$(C'A_A{}^X + AC'_A{}^X - B'B'_A{}^X - B'\bar{B}'_A{}^X)P_{XB}(t) = -\frac{\rho^2 s^2\mu^2}{\alpha^2 S^6} M^2 P_{AB}(t), \tag{6.26}$$

$$(B'A_A{}^X - AB'_A{}^X)P_{XB}(t) = \frac{s^3\mu^3(\lambda-\nu)}{\alpha S^5} Mh^E h^F Q_{AE} V_{FB}, \tag{6.27}$$

$$(B'A_A{}^X - A\bar{B}'_A{}^X)P_{XB}(t) = \frac{\rho^2 s^3\mu(\lambda-\mu)}{\alpha S^5} Mh^E h^F P_{AE} U_{FB}. \tag{6.28}$$

With the substitutions, (6.19) and (6.25)–(6.28), Eq. (6.13) simplifies to

$$P'_{AB}(t) = \frac{\lambda-\mu}{\lambda-\nu} P_{AB}(t) + (M^2+N^2)^{-1} \times \left[\frac{\alpha(2\lambda-\mu-\nu)}{\lambda-\nu} Nh^E h^F Q_{AE} U_{FB} - \alpha Mh^E h^F Q_{AE} V_{FB} + \alpha Mh^E h^F P_{AE} U_{FB} - \frac{\nu-\mu}{\lambda-\nu} M^2 P_{AB}(t) \right]. \tag{6.29}$$

This can be further simplified using the identity

$$NP_{AB}(t) = \alpha h^E h^F \left(\frac{\lambda-\nu}{\nu-\mu} P_{AE} V_{FB} - \frac{\lambda-\mu}{\nu-\mu} Q_{AE} U_{FB} \right), \tag{6.30}$$

which is proved using (6.18), (6.24), (A5) and (A35b). The final result is

$$P'_{AB}(t) = P_{AB}(t) + \frac{\alpha h^E h^F}{M^2+N^2} \times (MP_{AE}U_{FB} - MQ_{AE}V_{FB} + NP_{AE}V_{FB} + NQ_{AE}U_{FB}). \tag{6.31}$$

To find $F'_{AB}(t)$, we first need $Q'_{AB}(t)$, which itself needs $f'_{AB} = DP'_{AB}(0)$. From (6.23) and (A22), we have $D_2U_{AB}(s,0) = s^{-1}SQ_{BA}$, $D_2V_{AB}(s,0) = -s^{-1}SP_{BA}$. Hence

$$f'_{AB} = f_{AB} + \frac{\alpha(\nu-\mu)h^E h^F}{M^2+N^2} \times (MP_{AE}Q_{BF} + MQ_{AE}P_{BF} - NP_{AE}P_{BF} + NQ_{AE}Q_{BF}). \tag{6.32}$$

A straightforward application of (A35a,b), (A36a,b), (A38), (6.18), (6.24), (6.31), and (6.32) gives

$$Q'_{AB}(t) = Q_{AB}(t) + \frac{\alpha h^E h^F}{M^2+N^2} \times (MQ_{AE}U_{FB} + MP_{AE}V_{FB} - NP_{AE}U_{FB} + NQ_{AE}V_{FB}). \tag{6.33}$$

Equations (6.31) and (6.33) may be collected in the single equation,

$$P'_{AB}(t) + iQ'_{AB}(t) = P_{AB}(t) + iQ_{AB}(t) + \frac{\alpha h^E h^F (P_{AE}+iQ_{AE})(U_{FB}+iV_{FB})}{M+iN}.$$

Hence, using the definitions (A34), (6.16) and (6.23) and writing $q^{AB} = h^A h^B$, we finally arrive at the following much shorter and neater formula for the confluent two-soliton transformation:

$$F'_{AB}(t) = F_{AB}(t) + \frac{\alpha q^{CD} F_{AC}(s)\{G_{DB}(s,t) - [s/(s-t)]\epsilon_{DB}\}}{1-\alpha q^{EF} G_{EF}(s,s)}. \tag{6.34}$$

Then, from (A19), with help from (A18) and (A32), we get

$$G'_{AB}(t_1,t_2) = G_{AB}(t_1,t_2) + \frac{\alpha q^{CD}\{G_{AC}(t_1,s) - [t_1/(t_1-s)]\epsilon_{AC}\}\{G_{DB}(s,t_2) - [s/(s-t_2)]\epsilon_{DB}\}}{1-\alpha q^{EF} G_{EF}(s,s)}. \tag{6.35}$$

The right-hand side of (6.34) is singular at $t = s$. This pole can be removed by changing gauge by means of the gauge functions

$$g_{AB}(t) = \epsilon_{AB} - \frac{\alpha t}{s-t} q_{AB}. \tag{6.36}$$

The transformation now takes the form

$$F'_{AB}(t) = F_{AB}(t) - \frac{\alpha t}{s-t} q_{XB} F_A{}^X(t) + \frac{\alpha q^{CD} F_{AC}(s)\{G_D{}^X(s,t) - [s/(s-t)]\epsilon_D{}^X\}\{\epsilon_{XB} - [\alpha t/(s-t)]q_{XB}\}}{1-\alpha q^{EF} G_{EF}(s,s)}. \tag{6.37}$$

Let us use the notation

$$(\mathrm{HKX}{:}s,\alpha q^{AB}), \tag{6.38}$$

for the transformation (6.37), in anticipation of the identification with the HKX transformations. [It is easy to check that (6.37) reduces to (6.2) when $AB = 11$ and $q^{11} = 1$, other $q^{XY} = 0$.] The pole at $t = s$ is absent in (6.37), so we can multiply two transformations with same s parameters. In fact,

$$q^{CD}F'_{AC}(s) = \frac{q^{CD}F_{AC}(s)}{1-\alpha q^{EF}G_{EF}(s,s)}, \tag{6.39a}$$

$$q^{CD}\left(G'_{DB}(s,t) - \frac{s}{s-t}\epsilon_{DB}\right) = \left[q^{CD}G_D{}^X(s,t)\{\epsilon_{XB} - [\alpha t/(s-t)]q_{XB}\} - [s/(s-t)]q^C{}_B\right]\{1-\alpha q^{EF}G_{EF}(s,s)\}^{-1}, \tag{6.39b}$$

$$q^{EF}G'_{EF}(s,s) = \frac{q^{EF}G_{EF}(s,s)}{1-\alpha q^{EF}G_{EF}(s,s)}. \tag{6.39c}$$

A straightforward calculation now shows that

$$(\mathrm{HKX}{:}s,\beta q^{AB})(\mathrm{HKX}{:}s,\alpha q^{AB}) = (\mathrm{HKX}{:}s,(\alpha+\beta)q^{AB}), \tag{6.40}$$

which means that these transformations form a one-parameter group with α as canonical parameter. It is therefore sufficient to identify the infinitesimal limit of (6.37) with the infinitesimal HKX transformation (6.3). The infinitesimal form of (6.37) is

$$\begin{aligned} F'_{AB}(t) = {} & F_{AB}(t) - \frac{\alpha t}{s-t} q_{XB}F_A{}^X(t) \\ & + \alpha q^{CD}F_{AC}(s)\left(G_{DB}(s,t) - \frac{s}{s-t}\epsilon_{DB}\right) \\ & + O(\alpha^2). \end{aligned} \tag{6.41}$$

Then, from (A19) and (A32),

$$\begin{aligned} G'_{AB}(t_1,t_2) = {} & G_{AB}(t_1,t_2) \\ & + \frac{\alpha}{s-t_2} q_{YB}(-t_2G_A{}^Y(t_1,t_2) + sG_A{}^Y(t_1,s)) \\ & + \frac{\alpha t_1}{t_1-s} q_{AY}(G^Y{}_B(t_1,t_2) - G^Y{}_B(s,t_2)) \\ & + \alpha q^{CD}G_{AC}(t_1,s)G_{DB}(s,t_2) + O(\alpha^2). \end{aligned} \tag{6.42}$$

It is not difficult to expand the right-hand side of (6.42) as a triple power series in t_1, t_2, and s. The coefficient $t_1^m t_2^n$ is the transform of $N_{AB}^{(m,n)}$ and the coefficient of s^k in the latter may be compared directly with Eq. (2.23). We find that the infinitesimal transformation (6.42) is precisely the generalized HKX transformation (6.3), thus completing the proof.

Now, return to the original confluent BZ transformation given by (634) and (6.35). Although $F'_{AB}(t)$ and $G'_{AB}(t_1,t_2)$ have poles whenever t, t_1 or $t_2 = s$, product transformations with same s parameters are well defined by taking an obvious limit. Formulas (6.34) and (6.35) may be obtained from the HKX transformation equations by the gauge change inverse to (6.36) and which is expressed by the gauge functions,

$$g_{AB}(t) = \epsilon_{AB} + \frac{\alpha t}{s-t} q_{AB}. \tag{6.43}$$

When this is expressed in terms of the $\gamma_{XY}^{(k)}$ groups with negative index k, we find that it has the form,

$$\sum_{k=1}^{\infty} \alpha s^{-k} q^{XY}\gamma_{XY}^{(-k)}. \tag{6.44}$$

Since this commutes with (6.3), we can identify the transformation (6.34) with the open-ended HKX sum

$$\sum_{k=-\infty}^{\infty} \alpha s^k q^{XY}\gamma_{XY}^{(k)}. \tag{6.45}$$

Let this be called the extended HKX transformation. With an obvious notation, we have the theorem

$$(\mathrm{ext.HKX}{:}s,\alpha q^{AB}) = \lim_{\epsilon\to 0}(\mathrm{BZ}{:}s+\epsilon,h^A+\epsilon l^A)(\mathrm{BZ}{:}s,h^A), \tag{6.46}$$

with $\alpha = -(sh_X l^X)^{-1}$, $q^{AB} = h^A h^B$.

Now, we would expect from the commutation relation (3.20) that HKX transformations commute when they have the same q^{XY} parameter but different α and s parameters. But, since BZ transformations always commute, it follows that the extended HKX transformations commute also for different q^{XY} parameters. The corresponding commutation theorem for HKX transformations of the form (6.3) is

$$\begin{aligned} & (\mathrm{HKX}{:}s_2,\beta k'^A k'^B)(\mathrm{HKX}{:}s_1,\alpha h^A h^B) \\ & \quad = (\mathrm{gauge})(\mathrm{HKX}{:}s_1,\alpha h'^A h'^B)(\mathrm{HKX}{:}s_2,\beta k^A k^B), \end{aligned} \tag{6.47a}$$

$$h'^A = h^A - \frac{\beta s_1}{s_1-s_2} h_X k^X k^A, \tag{6.47b}$$

$$k'^A = k^A - \frac{\alpha s_2}{s_1-s_2} h_X k^X h^A.$$

A direct proof is straightforward but rather tedious.

Another property of the extended HKX transformations which follows from (6.46) is their self-duality. An immediate consequence of (6.46) and (5.33) is the following theorem:

$$(I^\epsilon)(\mathrm{ext.HKX}{:}s,\alpha q^{AB})(I^\epsilon) = (\mathrm{ext.HKX}{:}s,\alpha q'^{AB}), \tag{6.48}$$

where

$$q'^{11} = -\epsilon' i s^{-1} q^{22}, \quad q'^{12} = -q^{12}, \quad q'^{22} = \epsilon' i s q^{11}. \tag{6.49}$$

Take the infinitesimal limits (α small) of both sides of (6.48), expand as a power series in s, and look at the coefficient of s^k. Then, by comparing the coefficients of q^{11}, q^{12} and q^{22}, we arrive at a proof of the duality theorem (3.19).

The generalized HKX transform of the metric coefficient $e^{2\gamma}$ can be obtained without difficulty from the confluent form of Eq. (5.42), using (6.9), (6.19) and (6.16). The result is the very compact formula,

$$\begin{aligned} e^{2\gamma'-2u'} = {} & (M^2+N^2)e^{2\gamma-2u} \\ = {} & (1-\alpha q^{EF}G_{EF}(s,s)) \\ & \times(1-\alpha q^{EF}G^*{}_{EF}(s,s))e^{2\gamma-2u}, \end{aligned} \tag{6.50}$$

$e^{2u} = f$, $e^{2u'} = f'$. This result can be generalized to the case of a product of any number of generalized HKX transformations with any parameters, including higher rank transformations. All these transformations involve α parameters and s parameters. Under product transformations, the transforms of H_{AB}, $F_{AB}(t)$ and $G_{AB}(t_1,t_2)$ can all be written as rational functions of the α parameters, e.g.,

$$H'_{AB} = N_{AB}/D.$$

The denominator D is the same for H'_{AB}, F'_{AB} and G'_{AB}. Let the numerator and denominator be normalized by $N_{AB} = H_{AB}$, $D = 1$ when all of the α parameters are zero, and we assume that common factors have been cancelled out. Then the transform of $e^{2\gamma-2u}$ is given by the very simple formula,

$$e^{2\gamma'-2u'} = DD^*e^{2\gamma-2u}. \tag{6.51}$$

It is not difficult to prove this very general result by induction on the number of iterations.

The right-hand side of the formula,

$$H'_{AB} = H_{AB} - i\alpha s^{-1} S(s) \frac{q^{CD} F_{AC}(s) F_{BD}(s)}{1 - \alpha q^{EF} G_{EF}(s,s)}, \quad (6.52)$$

for the extended HKX transformation[50] contains a quadratic surd, the function $S(s)$. If the sign of $S(s)$ is changed, the new H'_{AB} must still satisfy Einstein's equations. Let us denote by $\tilde{F}_{AB}(s)$ the result of changing the sign of $S(s)$ in $F_{AB}(s)$. Then, since S occurs only in even powers in (A17), we must have

$$\tilde{F}_{AB}(s) = F_A{}^X(s) k_{XB}(s), \quad (6.53)$$

as for a gauge change, where $k_{AB}(s)$ does not depend on ρ and z. The inverse of (6.53) must have the same form so that $k_A{}^X k_{XB} = \epsilon_{AB}$. Changing the sign of S in Eqs. (A18) and (A23) gives the conditions

$$k_{AB} = k_{BA}, \quad k_{XA} k^X{}_B = -\epsilon_{AB}, \quad k^*{}_{AB} = -k_{AB}, \quad (6.54)$$

which are different from (1.15) for g_{AB}. We cannot make any statement about $k_{AB}(0)$. When the sign of S is changed in (6.52), the result is another HKX transformation of the same type with α replaced by $-\alpha$ and q^{AB} replaced by $q_{CD} k^{AC} k^{BD}$.

To illustrate this last result, consider the Weyl solutions. The generating function components and HKX transform ($q^{11} = 1$, other $q^{AB} = 0$) for this case are calculated in Ref. 7. It is a trivial matter to change the sign of S in their formulas. We find

$$\tilde{F}_{11}(s) = isF_{12}(s), \quad \tilde{F}_{12}(s) = -is^{-1}F_{11}(s),$$
$$\tilde{F}_{21}(s) = isF_{22}(s), \quad \tilde{F}_{22}(s) = -is^{-1}F_{21}(s).$$

Under this reflection, the HKX transformation with $q^{11} \neq 0$ becomes one with $q^{22} \neq 0$ (other $q^{AB} = 0$). This special symmetry occurs only because of the preferred gauge chosen for F_{AB} in the static case. If an arbitrary gauge were chosen for F_{AB}, then changing the sign of S would convert the special HKX transformation to a general one with arbitrary null q^{AB}.

Equation (6.37) can also be derived by iterating the nontensor Eqs. (4.44a,b) for the Harrison transformation. The poles at $t = s$ in the right-hand sides of (4.44a,b) are absent, so that $F'_{11}(s)$ and $F'_{12}(s)$ are well defined. The limit is very similar to the $t \to s$ limit of Eq. (A40). Consider a Harrison transformation with $c = c_1$ (denoted by a prime) followed by one with $c = c_2$ (double prime) with same s parameters. We find

$$q = \frac{c_1 F_{11} + F_{12}}{c_1 F^*{}_{11} + F^*{}_{12}}, \quad (6.55a)$$

$$q'^{-1} = \frac{c_2 F^{*\prime}{}_{11} + F^{*\prime}{}_{12}}{c_2 F'_{11} + F'_{12}}$$
$$= q + 2is\, q(c_1 + c_2) f'(c_1 F_{11} + F_{12})(c_1 F^*{}_{11} + F^*{}_{12})$$
$$\times \{1 + (c_1 + c_2)[G_{11} + (1/c_1)(G_{12} + G_{21}) + (1/c_1^2) G_{22}]\}^{-1}, \quad (6.55b)$$

$$f' = (4c_1{}^2 s^2 f)^{-1} [S(s)(q + q^{-1}) + 2(1 - 2sz)], \quad (6.55c)$$

where $F_{AB} = F_{AB}(s)$, $F'_{AB} = F'_{AB}(s)$, $G_{AB} = G_{AB}(s,s)$. A glance at the denominators of (6.55b) and (6.37) shows that the parameters must be related by

$$\alpha q^{11} = -(c_1 + c_2),$$
$$\alpha q^{12} = -(c_1 + c_2)/c_1, \quad (6.56)$$
$$\alpha q^{22} = -(c_1 + c_2)/c_1{}^2.$$

With this identification, and using Eq. (A40), we find, after a lengthy calculation,

$$F''_{11}(t) = (c_1/c_2)^2 \tilde{F}_{11}(t),$$
$$F''_{12}(t) = \tilde{F}_{12}(t) + (c_1/c_2)(c_1 + c_2)\tilde{F}_{11}(t), \quad (6.57)$$

where the tilde ($\sim$) denotes the generalized HKX transform given by (6.37) (not as in the preceding two paragraphs). This shows that

$$\tilde{F}_{AB}(t) = (U)_\lambda (A)_\mu F''_{AB}(t),$$
$$e^\lambda = (c_2/c_1)^2, \quad (6.58)$$
$$\mu = (c_2/c_1)(c_1 + c_2).$$

In Ref. 11, we pointed out a qualitative similarity between double HKX rank-0 and double Harrison transformations. The former applied to a Weyl or CTS solution gives the same result as the latter applied to a different Weyl or CTS solution, respectively. This can now be explained. The Harrison transformations with $c = 0$ or $c = \infty$ and the BZ transformations with $h^1 = 0$ or $h^2 = 0$ map Weyl solutions to Weyl solutions. [According to (5.28) and (5.29), $c = h^1/h^2$.] From (6.46) and (5.28), the original HKX transformation is the product of a special Harrison transformation with $c = \infty$ followed by a Harrison transformation with same s, arbitrary c, and then a gauge change. The double HKX (different s parameters) is therefore a fourfold Harrison transformation and the factors can be rearranged so that the two Weyl-to-Weyl transformations are done first. The double Harrison transformation with $c = 0$ in the first factor gives the HKX with the sign of $S(s)$ changed or with the γ_{11}'s replaced by γ_{22}'s (recall Footnote 43). The same arguments apply to the CTS solutions and their generalizations.

A remarkable consequence of the previous result is that the generalized HKX transform of any Weyl or CTS solution is the same as a combined HKX rank-0 and rank-1 transform (of the original type in Ref. 7) of a different Weyl or CTS solution, respectively. We shall demonstrate this explicitly for the Weyl solutions. The general solution is given by

$$\mathscr{E} = \mathscr{E}^* = e^{2\chi}, \quad \nabla_3{}^2 \chi = 0, \quad (6.59)$$

$\nabla_3{}^2 = \partial^2/\partial\rho^2 + \partial^2/\partial z^2 + \rho^{-1}\partial/\partial\rho$. The generating function components for this case were calculated in Ref. 7. From their results, we find

$$G_{11}(s,t) = -\frac{it}{2S(t)} + e^{\beta(s)+\beta(t)}\left[1 + \frac{s + t - 4stz}{sS(t) + tS(s)}\right], \quad (6.60a)$$

$$G_{12}(s,t) = \frac{s}{s-t} + e^{\beta(s)-\beta(t)} \frac{s - t - sS(t) - tS(s)}{2(s-t)S(t)}, \quad (6.60b)$$

$$G_{21}(s,t) = -\frac{s}{s-t} + te^{-\beta(s)+\beta(t)} \frac{s - t + sS(t) + tS(s)}{2s(s-t)S(t)}, \quad (6.60c)$$

$$G_{22}(s,t)=\frac{i}{2sS(t)}e^{-\beta(s)-\beta(t)}\left[1-\frac{s+t-4stz}{sS(t)+tS(s)}\right], \tag{6.60d}$$

where $\beta(t)$ is defined up to an additive integration constant (or function of t) by

$$\nabla\beta(t)=S^{-1}(t)[(1-2tz)\nabla\chi-2t\rho\,\tilde{\nabla}\chi]. \tag{6.61}$$

The double HKX rank-zero transformation is expressed by

$$\mathscr{E}^{*\prime}=\mathscr{E}^{*}+\frac{\alpha_1F_1G'_1(1-\alpha_2G(s_2,s_2))+\alpha_2F_2G'_2(1-\alpha_1G(s_1,s_1))+\alpha_1\alpha_2[F_1G'_2G(s_1,s_2)+F_2G'_1G(s_2,s_1)]}{1-\alpha_1G(s_1,s_1)-\alpha_2G(s_2,s_2)+\alpha_1\alpha_2[G(s_1,s_1)G(s_2,s_2)-G(s_1,s_2)G(s_2,s_1)]}, \tag{6.62}$$

where $F_1=F_{11}(s_1), F_2=F_{11}(s_2), G'_1=D_2G_{11}(s_1,0), G'_2=D_2G_{11}(s_2,0)$ and $G(s,t)=G_{11}(s,t)$. The effect on the Weyl solution $\mathscr{E}^{*}=e^{2\chi}$ is given by

$$\mathscr{E}^{*\prime}=e^{2\chi}\frac{1+i\alpha_1A_-e^{2\beta_1}+i\alpha_2B_-e^{2\beta_2}-\alpha_1\alpha_2[(\zeta_2-\zeta_1)/(\zeta_2+\zeta_1)]^2A_-B_-e^{2\beta_1+2\beta_2}}{1+i\alpha_1A_+e^{2\beta_1}+i\alpha_2B_+e^{2\beta_2}-\alpha_1\alpha_2[(\zeta_2-\zeta_1)/(\zeta_2+\zeta_1)]^2A_+B_+e^{2\beta_1+2\beta_2}}, \tag{6.63}$$

where $\beta_1=\beta(s_1), \beta_2=\beta(s_2)$ and

$$A_\pm=\frac{1}{4}\frac{s_1s_2}{s_1-s_2}\frac{(1\pm\zeta_1)^2(\zeta_2^2-\zeta_1^2)}{\zeta_1^2(1-\zeta_2^2)},$$

$$B_\pm=\frac{1}{4}\frac{s_1s_2}{s_1-s_2}\frac{(1\pm\zeta_2)^2(\zeta_2^2-\zeta_1^2)}{\zeta_2^2(1-\zeta_1^2)}.$$

Here, we have used $\zeta_1=\zeta(s_1)$ and $\zeta_2=\zeta(s_2)$ as coordinates. (We could instead use prolate spheroidal coordinates as in Ref. 11, but there are six ways of doing this depending on the signs of s_1 and s_2.) The double BZ transform of $\mathscr{E}^{*}=e^{2\chi}$ can be obtained directly from (5.39) with

$$q_1=\frac{h^{1}s_1e^{2\beta_1}+ih^{2}}{h^{1}s_1e^{2\beta_1}-ih^{2}},\quad q_2=\frac{k^{1}s_2e^{2\beta_2}+ik^{2}}{k^{1}s_2e^{2\beta_2}-ik^{2}}.$$

The result is

$$\mathscr{E}^{*\prime}=\frac{(1+\zeta_1)(1+\zeta_2)}{(1-\zeta_1)(1-\zeta_2)}e^{2\chi}\left(\frac{1-i[(\zeta_2+\zeta_1)/(\zeta_2-\zeta_1)](C_-e^{2\beta_1}-D_-e^{2\beta_2})+C_-D_-e^{2\beta_1+2\beta_2}}{1-i[(\zeta_2+\zeta_1)/(\zeta_2-\zeta_1)](C_+e^{2\beta_1}-D_+e^{2\beta_2})+C_+D_+e^{2\beta_1+2\beta_2}}\right), \tag{6.64}$$

$$C_\pm=\frac{h^{1}s_1}{h^{2}}\left(\frac{1+\zeta_1}{1-\zeta_1}\right)^{\pm1},\quad D_\pm=\frac{k^{1}s_2}{k^{2}}\left(\frac{1+\zeta_2}{1-\zeta_2}\right)^{\pm1}.$$

Now define

$$e^{2\bar{\chi}}=\frac{(1+\zeta_1)(1+\zeta_2)}{(1-\zeta_1)(1-\zeta_2)}e^{2\chi}, \tag{6.65a}$$

$$e^{2\bar{\beta}_1}=\left(\frac{s_1-s_2}{s_1s_2}\right)^2\frac{\zeta_1^2(1-\zeta_2^2)}{(1-\zeta_1^2)(\zeta_2-\zeta_1)^2}e^{2\beta_1}, \tag{6.65b}$$

$$e^{2\bar{\beta}_2}=\left(\frac{s_1-s_2}{s_1s_2}\right)^2\frac{\zeta_2^2(1-\zeta_1^2)}{(1-\zeta_2^2)(\zeta_2-\zeta_1)^2}e^{2\beta_2}, \tag{6.65c}$$

satisfying (6.61). A comparison of (6.63) with (6.64) reveals the theorem

$$(\mathrm{HKX}{:}s_2,\alpha_2q^{AB})(\mathrm{HKX}{:}s_1,\alpha_1q^{AB})e^{2\bar{\chi}}=(\mathrm{BZ}{:}s_2,k^{A})(\mathrm{BZ}{:}s_1,h^{A})e^{2\chi}, \tag{6.66}$$

where $q^{11}=1$, other $q^{AB}=0$, and the parameters are related by

$$\alpha_1=-4\frac{h^{1}s_1}{h^{2}}\frac{s_1s_2}{s_1-s_2},\quad \alpha_2=4\frac{k^{1}s_2}{k^{2}}\frac{s_1s_2}{s_1-s_2}. \tag{6.67}$$

These relations would change if we were to choose different integration constants for $\bar{\beta}_1$ and $\bar{\beta}_2$ in (6.65b,c).

Now, the generalized HKX transform of the Weyl solution $\mathscr{E}^{*}=e^{2\chi}$ is given by

$$\mathscr{E}^{*\prime}=e^{2\chi}\frac{1+\alpha q^{12}[1+S^{-1}+2sD\beta(s)]+\frac{1}{2}i\alpha S^{-2}[(1-2sz-S)sq^{11}e^{2\beta(s)}+(1-2sz+S)s^{-1}q^{22}e^{-2\beta(s)}]}{1+\alpha q^{12}[1-S^{-1}+2sD\beta(s)]+\frac{1}{2}i\alpha S^{-2}[(1-2sz+S)sq^{11}e^{2\beta(s)}+(1-2sz-S)s^{-1}q^{22}e^{-2\beta(s)}]}. \tag{6.68}$$

We claim that this can be attained by a combined HKX rank-0 and rank-1 transformation acting on the Weyl solution,

$$\mathscr{E}^* = e^{2\bar{\chi}} = \left(\frac{1+\zeta(s)}{1-\zeta(s)}\right)^2 e^{2\chi}. \tag{6.69}$$

The combined HKX rank-0, rank-1, ..., rank-N transform of $\mathscr{E}^*$ is given by the matrix equation (3.19) of Ref. 7. Putting $N=1$, $u=s$, $\alpha^{(0)}=\lambda$, $u\alpha^{(1)}=\mu$ in their formula, we get

$$\mathscr{E}^{*\prime} = \mathscr{E}^* + \frac{\lambda FG' + \mu(DFG' + FDG') + \mu^2(FG'G_{12} + DFDG'G - FDG'G_2 - DFG'G_1)}{1 - \lambda G - \mu(G_1+G_2) + \mu^2(G_1G_2 - GG_{12})}, \tag{6.70}$$

where we have used the abbreviations,

$$F = F_{11}(s), \quad DF = DF_{11}(s),$$
$$G' = D_2G_{11}(s,0), \quad DG' = D_1D_2G_{11}(s,0),$$
$$G = G_{11}(s,s), \quad G_1 = D_1G_{11}(s,s),$$
$$G_2 = D_2G_{11}(s,s), \quad G_{12} = D_1D_2G_{11}(s,s).$$

Substitution of (6.69) and (6.60) into (6.70) will yield (6.68) after a very tedious calculation. We can avoid this by observing that (6.70) can be obtained as a confluent double rank-zero transformation by putting

$$s_1 = s, \quad s_2 = s+\epsilon, \quad \alpha_1 = \lambda - \mu\epsilon^{-1}, \quad \alpha_2 = \mu\epsilon^{-1} \tag{6.71}$$

in (6.62) and taking the limit as $\epsilon \to 0$. Now, as $s_2 \to s_1 = s$, the right-hand side of (6.66) tends to the generalized (extended) HKX transformation according to theorem (6.46). The same limit applied to the left-hand side of (6.66) gives the combined HKX rank-0 and rank-1 according to (6.71). The parameters are related by

$$\lambda = -4(h^1/h^2)s^2 - 4\alpha^{-1}(s/h^2)^2,$$
$$\mu = -4(h^1/h^2)s^3, \quad q^{AB} = h^A h^B, \tag{6.72}$$

provided we define $\bar{\beta}_1$ and $\bar{\beta}_2$ by the $s_2 \to s_1$ limit of Eqs. (6.65b,c).

7. MISCELLANEOUS QUESTIONS

In this section, we discuss a few properties of the non-null HKX transformations, the Kinnersley–Chitre $\beta^{(k)}$ transformations[5] and the linear integral equations of Refs. 20 and 21.

The infinitesimal form of the non-null HKX transformation is defined by (6.3) where now q^{XY} is allowed to be an arbitrary nonnull symmetric tensor. The infinitesimal transforms of $F_{AB}(t)$ and $G_{AB}(t_1,t_2)$ are given by (6.41) and (6.42) without modification. However, formula (6.37) for the finite transformation does not hold when q^{XY} is non-null. In fact, it does not appear to be possible to write down a simple algebraic expression such as (6.37) for the exponential non-null HKX transformation, although there are large classes of special solutions whose transforms can be written down explicitly.

Equation (6.42) for the four components of $G'_{AB}(t_1,t_2)$ can be decoupled. First, using (A4)–(A6) and (A33), we get the contractions,

$$q^{AB}G'_{AB}(t_1,t_2) = q^{AB}G_{AB}(t_1,t_2) + \alpha q^{AB}G_{AB}(t_1,s)q^{CD}G_{CD}(s,t_2) + \alpha q^2 G_A{}^A(t_1,s)G_B{}^B(s,t_2) + 2\alpha q^2 \left[\frac{s(t_1-t_2)}{(t_1-s)(s-t_2)}G_A{}^A(t_1,t_2) - \frac{s}{s-t_2}G_A{}^A(t_1,s) - \frac{t_1}{t_1-s}G_A{}^A(s,t_2)\right] + O(\alpha^2), \tag{7.1}$$

$$G'_A{}^A(t_1,t_2) = G_A{}^A(t_1,t_2) + \alpha G_A{}^A(t_1,s)q^{CD}G_{CD}(s,t_2) + \alpha G_A{}^A(s,t_2)q^{CD}G_{CD}(t_1,s) + 2\alpha q^{AB} \times \left[\frac{s(t_1-t_2)}{(t_1-s)(s-t_2)}G_{AB}(t_1,t_2) - \frac{s}{s-t_2}G_{AB}(t_1,s) - \frac{t_1}{t_1-s}G_{AB}(s,t_2)\right] + O(\alpha^2), \tag{7.2}$$

where the real or pure imaginary constant q is defined by

$$q_{XY}q^{XY} = -2q^2, \quad q_{XA}q^X{}_B = -q^2\epsilon_{AB}. \tag{7.3}$$

Now define

$$K_\epsilon(t_1,t_2) = q^{AB}G_{AB}(t_1,t_2) + \epsilon q G_A{}^A(t_1,t_2), \quad \epsilon = \pm 1. \tag{7.4}$$

Then (7.1) and (7.2) may be combined into a single equation for K_ϵ:

$$K'_\epsilon(t_1,t_2) = K_\epsilon(t_1,t_2) + \alpha K_\epsilon(t_1,s)K_\epsilon(s,t_2) + 2\epsilon\alpha q \times \left[\frac{s(t_1-t_2)}{(t_1-s)(s-t_2)}K_\epsilon(t_1,t_2) - \frac{s}{s-t_2}K_\epsilon(t_1,s) - \frac{t_1}{t_1-s}K_\epsilon(s,t_2)\right] + O(\alpha^2). \tag{7.5}$$

(In the special case, $q^{12}=q^{21}=1$, $q^{11}=q^{22}=0$, we have $K_{+1}=2G_{12}$, $K_{-1}=2G_{21}$.) When $q=0$, Eq. (7.5) is very easy to exponentiate. The result is that $K_\epsilon (= q^{AB}G_{AB})$ transforms according to Eq. (6.2). In the general case, we can rewrite (7.5) as a functional differential equation. Let $u=(t_1-s)/s$, $v=(t_2-s)/s$ and write $K'_\epsilon(t_1,t_2) = L_\epsilon(u,v,\alpha)$, where α is finite. Then

$$D_3L_\epsilon(u,v,\alpha) = L_\epsilon(u,0,\alpha)L_\epsilon(0,v,\alpha) + 2\epsilon q \times \left[\left(\frac{1}{u}-\frac{1}{v}\right)L_\epsilon(u,v,\alpha) + \frac{1}{v}L_\epsilon(u,0,\alpha) - \left(1+\frac{1}{u}\right)L_\epsilon(0,v,\alpha)\right]. \tag{7.6}$$

Notice that s does not appear in (7.6). Unfortunately, this does not mean that the nonnull HKX can be exponentiated for general s if it can be done for the limiting cases, $s=0$ or $s=\infty$.

It turns out that the nonnull HKX transformation can be exponentiated in the limit as $s \to \infty$. Hoenselaers, Kin-

nersley and Xanthopoulos have also succeeded with this case using different methods.[49] First, it is easy to check from Eq. (A17) that gauge functions can be chosen so that $F_{AB}(t)$ is analytic at $t=\infty$. [Since $F_{AB}(t)$ has quadratic branch points at the two complex zeros of $S(t)$, this statement assumes that the branch cut is a finite segment from one point to the other, crossing the real axis. If we wish to keep $F_{AB}(t)$ analytic, or having poles at worst, along the real axis, then $F_{AB}(-\infty) = \tilde{F}_{AB}(+\infty)$ where $\tilde{F}_{AB}$ is described in the paragraph containing Eqs. (6.53) and (6.54).] Equation (A18) shows that $F_{AB}(\infty)$ is a null tensor. Next, (A18) and (A19) show that $G_{AB}(s,\infty)$, $G_{AB}(\infty,t)$ and $G_{AB}(\infty,\infty)$ are all well defined provided the branch of $S(s)$ is chosen consistently near $s=\infty$. From (A32) and (A33), it follows that $G_{AB}(s,\infty)$ and $G_{AB}(\infty,\infty)$ are null tensors. All of the formulas (A17)–(A43) have well defined limiting forms as s and/or t tends to infinity.

Put $s=\infty$ in Eq. (7.5) to give

$$K'_\epsilon(t_1,t_2) = K_\epsilon(t_1,t_2) + \alpha K_\epsilon(t_1,\infty) \times (K_\epsilon(\infty,t_2) - 2\epsilon q) + O(\alpha^2), \tag{7.7}$$

which can be exponentiated by four elementary integrations. (Solve for $K'_\epsilon(\infty,\infty)$ first, then $K'_\epsilon(t_1,\infty)$, $K'_\epsilon(\infty,t_2)$ and finally $K'_\epsilon(t_1,t_2)$.) The result is

$$K'_\epsilon(t_1,t_2) = K_\epsilon(t_1,t_2) - \frac{(\sinh q\alpha)K_\epsilon(t_1,\infty)(K_\epsilon(\infty,t_2) - 2\epsilon q)}{(\sinh q\alpha)K_\epsilon(\infty,\infty) - qe^{\epsilon q\alpha}}. \tag{7.8}$$

This gives us two linear combinations of the four components of $G'_{AB}(t_1,t_2)$. The remaining components can be found in a fully covariant manner by first observing that $q^{AB} + \epsilon q\epsilon^{AB}$ is null and must therefore factorize into the form,

$$q^{AB} + \epsilon q\epsilon^{AB} = g^A_\epsilon h^B_\epsilon, \tag{7.9}$$

so that $h^B_\epsilon q_{AB} = \epsilon q h_{\epsilon A}$. ($\epsilon = \pm 1$ is not a tensor index and does not take part in the summation convention.) Define a vector $H_{\epsilon A}$ by

$$H_{\epsilon A}(t_1,t_2) = h^B_\epsilon G_{AB}(t_1,t_2), \tag{7.10}$$

so that $K_\epsilon = g^A_\epsilon H_{\epsilon A}$. When $s=\infty$, Eq. (6.42) reduces to

$$G'_{AB}(t_1,t_2) = G_{AB}(t_1,t_2) + \alpha q_{YB}G_A{}^Y(t_1,\infty) + \alpha q^{CD}G_{AC}(t_1,\infty)G_{DB}(\infty,t_2) + O(\alpha^2). \tag{7.11}$$

Contract with h^B_ϵ and use (A5) and (A33) to obtain

$$H'_{\epsilon A}(t_1,t_2) = H_{\epsilon A}(t_1,t_2) + \alpha H_{\epsilon A}(t_1,\infty) \times (K_\epsilon(\infty,t_2) - 2\epsilon q) + O(\alpha^2). \tag{7.12}$$

Since we know $K'_\epsilon(\infty,t_2)$, this can be exponentiated with two elementary integrations. The result is

$$H'_{\epsilon A}(t_1,t_2) = H_{\epsilon A}(t_1,t_2) - \frac{(\sinh q\alpha)H_{\epsilon A}(t_1,\infty)(K_\epsilon(\infty,t_2) - 2\epsilon q)}{(\sinh q\alpha)K_\epsilon(\infty,\infty) - qe^{\epsilon q\alpha}}. \tag{7.13}$$

The individual components of $G'_{AB}(t_1,t_2)$ may now be recovered by means of the identities,

$$G_{AB} = (h_{+X}h^X_-)^{-1}(h_{+B}H_{-A} - h_{-B}H_{+A}), \tag{7.14}$$

$$q(h_{+A}h_{-B} + h_{-A}h_{+B}) = -h_{+X}h^X_- q_{AB}. \tag{7.15}$$

We thus arrive at the following formulas for the $s=\infty$ limit of the nonnull HKX transforms of the generating functions:

$$+ \frac{q^{-1}(\tanh q\alpha)q^{CD}G_{AC}(t_1,\infty)(G_{DB}(\infty,t_2) - \epsilon_{DB})}{1 - q^{-1}(\tanh q\alpha)q^{EF}G_{EF}(\infty,\infty)}, \tag{7.16}$$

$$F'_{AB}(t) = F_{AB}(t) + \frac{q^{-1}(\tanh q\alpha)q^{CD}F_{AC}(\infty)(G_{DB}(\infty,t) - \epsilon_{DB})}{1 - q^{-1}(\tanh q\alpha)q^{EF}G_{EF}(\infty,\infty)}. \tag{7.17}$$

When $q=0$, these formulas reduce to (6.34), (6.35) and (6.37) for the null case with $s=\infty$. But the similarity goes deeper than this. Formulas (6.34), (6.35) and (6.37) actually hold when q^{AB} is non-null in the limiting case $s=\infty$. They can be proved directly by showing that they have the correct infinitesimal limits (given by (6.41) and (6.42)) and satisfy the group property (6.40) except that $\alpha+\beta$ on the right-hand side must be replaced by $(\alpha+\beta)/(1+q^2\alpha\beta)$. When we replace α by $q^{-1}(\tanh q\alpha)$, we get (7.16) and (7.17) above which satisfy (6.40) without modification.

Let us determine the effect of the nonnull HKX transformation with $s=\infty$ on the Weyl solutions. The transform of $\mathscr{E}^* = e^{2\chi}$ can be obtained from Eq. (6.68) by putting $s=\infty$, replacing α by $q^{-1}(\tanh q\alpha)$ and allowing q^{AB} to be nonnull. The result is

$$\mathscr{E}^{*\prime} = e^{2\chi}\frac{1 + q^{-1}(\tanh q\alpha)(q^{12} - \frac{1}{4}ir^{-2}(z+r)q^{11}e^{2\beta(\infty)})}{1 + q^{-1}(\tanh q\alpha)(q^{12} - \frac{1}{4}ir^{-2}(z-r)q^{11}e^{2\beta(\infty)})}, \tag{7.18}$$

where $r=(\rho^2+z^2)^{1/2}$. Compare this with the $s=\infty$ limit of the original null HKX transform with $q^{11}=1$, other $q^{AB}=0$, given in Ref. 7, which is

$$\mathscr{E}^{*\prime} = e^{2\chi}\frac{1 - \frac{1}{4}i\alpha r^{-2}(z+r)e^{2\beta(\infty)}}{1 - \frac{1}{4}i\alpha r^{-2}(z-r)e^{2\beta(\infty)}}. \tag{7.19}$$

Clearly, (7.18) can be converted into (7.19) by a trivial reparametrization, so the nonnull HKX gives us nothing new in this case.

A number of interesting commutation and conjugation theorems can be proved for the nonnull HKX transformations using only the infinitesimal transformation equation (6.42) or its extended version. The extended nonnull HKX transformation is defined by exponentiating (6.45) and is related to the original by the gauge functions,

$$g_{AB}(t) = \left(\cosh\frac{\alpha q t}{s-t}\right)\epsilon_{AB} + q^{-1}\left(\sinh\frac{\alpha q t}{s-t}\right)q_{AB}, \tag{7.20}$$

reducing to (6.43) when $q=0$. Thus the infinitesimal extended nonnull HKX transform of G_{AB} is given by

$$G'_{AB}(t_1,t_2) = G_{AB}(t_1,t_2) + \alpha q^{CD}\left(G_{AC}(t_1,s) - \frac{t_1}{t_1-s}\epsilon_{AC}\right)$$

$$\times\left(G_{DB}(s,t_2)-\frac{s}{s-t_2}\epsilon_{DB}\right)+O(\alpha^2). \quad (7.21)$$

It is not difficult to show that extended nonnull HKX transformations commute for all values of their α, s and q^{AB} parameters. The corresponding theorem for the nonextended transformations is

$$(\mathrm{HKX}{:}s_2,\alpha_2 q'^{AB}_2)(\mathrm{HKX}{:}s_1,\alpha_1 q_1^{AB})$$
$$=(\text{gauge})(\mathrm{HKX}{:}s_1,\alpha_1 q'^{AB}_1)(\mathrm{HKX}{:}s_2,\alpha_2 q_2^{AB}), \quad (7.22a)$$
$$q'^{AB}_1=q_1^{CD}g_2{}^A{}_C(s_1)g_2{}^B{}_D(s_1),$$
$$q'^{AB}_2=q_2^{CD}g_1{}^A{}_C(s_2)g_1{}^B{}_D(s_2), \quad (7.22b)$$

where $g_{1AB}(t)$ and $g_{2AB}(t)$ are given by (7.20) with (α,s,q,q^{AB}) replaced by $(\alpha_1,s_1,q_1,q_1^{AB})$ and $(\alpha_2,s_2,q_2,q_2^{AB})$, respectively. This reduces to theorem (6.47) when $q_1=q_2=0$.

We already know that BZ transformations always commute with extended null HKX transformations. They also commute with extended nonnull HKX transformations. This can be proved directly by writing (7.19) in terms of F_{11} and F_{12} using (A40) and using Eqs. (5.29) and (5.35a,b) for the BZ transformation (with different s parameters). However, at an early stage in the calculation, the equation to be checked is of the form $q^{XY}[\ldots]_{XY}=0$, where none of the terms in the bracket depend on q^{AB}. Since the coefficients of q^{11}, q^{12} and q^{22} vanish identically in the null case, they must also vanish identically in the nonnull case.

The null and nonnull HKX transformations also obey the following conjugation theorems:

$$(Z)_{\mu}(\mathrm{HKX}{:}s,\alpha q^{AB})(Z)_{-\mu}$$
$$=\left(\mathrm{HKX}{:}\frac{s}{1+2\mu s},\frac{\alpha}{1+2\mu s}q^{AB}\right); \quad (7.23)$$
$$(\tilde{Q})_{4u}(\mathrm{HKX}{:}s,\alpha q^{AB})(\tilde{Q})_{-4u}$$
$$=\exp\left(-\frac{\alpha u}{s-u}q^{XY}\gamma^{(0)}_{XY}\right)\left(\mathrm{HKX}{:}s-u,\frac{\alpha s}{s-u}q^{AB}\right); \quad (7.24)$$
$$(\tilde{Q})_{4s}(\mathrm{HKX}{:}s,\alpha q^{AB})(\tilde{Q})_{-4s}$$
$$=\exp(\alpha q^{XY}\gamma^{(0)}_{XY})\exp(\alpha s q^{XY}\gamma^{(1)}_{XY}); \quad (7.25)$$
$$(\tilde{Q})_{4u}(\text{ext.HKX}{:}s,\alpha q^{AB})(\tilde{Q})_{-4u}$$
$$=\left(\text{ext.HKX}{:}s-u,\frac{\alpha s}{s-u}q^{AB}\right); \quad (7.26)$$
$$(I^c)(\text{ext.HKX}{:}s,\alpha q^{AB})(I^c)=(\text{ext.HKX}{:}s,\alpha q'^{AB})$$

with

$$(q'^{11},q'^{12},q'^{22})=(-\epsilon i s^{-1}q^{22},-q^{12},\epsilon i s q^{11}); \quad (7.27)$$
$$(Q)_{4u}(\text{ext.HKX}{:}s,\alpha q^{AB})(Q)_{-4u}$$
$$=\left(\text{ext.HKX}{:}s-u,\frac{\alpha s}{s-u}q'^{AB}\right)$$

with

$$(q'^{11},q'^{12},q'^{22})=\left(\frac{s}{s-u}q^{11},q^{12},\frac{s-u}{s}q^{22}\right); \quad (7.28)$$
$$(\text{gauge})(\text{ext.HKX}{:}s,\alpha q^{AB})(\text{gauge})^{-1}$$
$$=(\text{ext.HKX}{:}s,\alpha q'^{AB})$$

with

$$q'^{AB}=q^{CD}g_C{}^A(s)g_D{}^B(s). \quad (7.29)$$

Theorems (7.23), (7.24), (7.26) and (7.29) are easy to prove directly from Eqs. (6.42), (7.21), (1.19), (2.16) and (1.14). Theorem (7.25) is a simple limiting form of (7.24). Theorem (7.27) follows automatically from theorem (3.19) which was deduced from theorem (6.48) where q^{AB} was null. Theorem (7.28) follows from (7.26) and (7.27). Theorem (7.25) shows that the nonnull HKX can be exponentiated if $q^{XY}\gamma^{(1)}_{XY}$ can be exponentiated (q^{XY} nonnull) and vice versa. Theorem (7.23) gives the impression that the general nonnull HKX transformation law can be deduced from the $s=\infty$ limit which is known. Unfortunately, the $O(s^{-1})$ term in the asymptotic expansion as $s\to\infty$ is also needed.

Kinnersley and Chitre[5,51] discovered that the commuting infinitesimal transformations,[43]

$$\beta^{(k)}=\gamma^{(k+2)}_{11}+\gamma^{(k)}_{22}, \quad (7.30)$$

individually preserve asymptotic flatness. The HKX-type sums,

$$(\mathrm{KC}{:}s,\alpha)=\exp\left\{\sum_{k=0}^{\infty}\alpha s^k\beta^{(k-1)}\right\}, \quad (7.31a)$$

may be regarded as generating functions for the $\beta^{(k)}$ and are special cases of the (extended) nonnull HKX transformation. In fact, it is easy to prove the following results:

$$(\mathrm{KC}{:}s,\alpha)=(A)_{-\alpha/s}(\text{gauge})(\mathrm{KHX}{:}s,\alpha q^{AB}), \quad (7.32a)$$
$$(\text{ext.KC}{:}s,\alpha)=(\text{ext.HKX}{:}s,\alpha q^{AB}), \quad (7.32b)$$
$$q^{11}=s^{-1},\quad q^{12}=0,\quad q^{22}=s;$$
$$(I^c)(\mathrm{KC}{:}s,\alpha)(I^c)=(\mathrm{HKX}{:}s,-\epsilon i\alpha q'^{AB}), \quad (7.33a)$$
$$(I^c)(\text{ext.KC}{:}s,\alpha)(I^c)=(\text{ext.HKX}{:}s,-\epsilon i\alpha q'^{AB}), \quad (7.33b)$$
$$q'^{11}=1,\quad q'^{12}=0,\quad q'^{22}=-1,$$

The gauge functions in (7.32a) may be constructed by using the relationship of the extended nonnull HKX to the nonnull HKX given by (7.20), together with the facts that (7.33a) is free of gauge and the dual of a gauge transformation is readily obtainable from (1.14), (1.15), (3.12a–d), and (1.8a). They are found to be

$$g_{11}(t)=(s-t)\sin\frac{\alpha t}{s-t}\cos\frac{\alpha t}{s-t},$$
$$g_{12}(t)=\cos^2\frac{\alpha t}{s-t}+\frac{s}{t}\sin^2\frac{\alpha t}{s-t}-\frac{\alpha}{s}(s-t)\sin\frac{\alpha t}{s-t}\cos\frac{\alpha t}{s-t},$$
$$g_{21}(t)=-\cos^2\frac{\alpha t}{s-t}-\frac{t}{s}\sin^2\frac{\alpha t}{s-t},$$
$$g_{22}(t)=\left(\frac{1}{s}-\frac{1}{t}\right)\sin\frac{\alpha t}{s-t}\cos\frac{\alpha t}{s-t}+\frac{\alpha}{s}\left(\cos^2\frac{\alpha t}{s-t}+\frac{t}{s}\sin^2\frac{\alpha t}{s-t}\right).$$

The full infinite-dimensional group generated by the $\beta^{(k)}$, $k\geqslant 0$, can be constructed by iterating $\beta^{(0)}$ and $\mathbf{Q}^0$ alone. This follows from (7.25) and (7.33a) which give the result:

$$\exp(-\alpha\beta^{(-1)})(\mathrm{KC}:s,\alpha) = (Q)_{-4s}\exp(\alpha s\beta^{(0)})(Q)_{4s}. \tag{7.34}$$

The $\beta^{(-1)}$ transformation is precisely the gravitational duality rotation given by (1.11) and therefore generates Taub-NUT space from Schwarzschild. Kinnersley and Chitre realized that $\beta^{(0)}$ has the important property of generating Kerr from Schwarzschild. $\beta^{(0)}$ and $\beta^{(1)}$ combined give Kerr–NUT. All the remaining $\beta^{(k)}$ transformations merely reparametrize the Kerr-NUT solution. When applied to the static Zipoy–Voorhees solutions,[28]

$$e^{2\chi} = \left(\frac{x-1}{x+1}\right)^{\delta}, \tag{7.35}$$

with δ an integer ($\delta = 1$ is Schwarzschild) and coordinates (x,y) defined by (2.5), the $\beta^{(k)}$ transformations can be exponentiated by the methods of Ref. 5. The resulting stationary solutions, the KC solutions, have 2δ parameters (including κ, excluding NUT) and are generated by $\beta^{(-1)}$, $\beta^{(0)},\ldots,\beta^{(2\delta-2)}$ only, being reparametrized by higher $\beta^{(k)}$. These techniques will also succeed with the more general static solution,

$$e^{2\chi} = \left(\frac{x_1-1}{x_1+1}\right)\left(\frac{x_2-1}{x_2+1}\right)\cdots\left(\frac{x_n-1}{x_n+1}\right), \tag{7.36}$$

which is a superposition of n Schwarzschild particles of different masses on the z axis. The coordinates (x_i,y_i) are defined as in (2.5) except that the foci are to be moved from $z = \pm\kappa$ to arbitrary positions on the z axis. [We can also include factors of the form $(1+y_i)/(1-y_i)$ in (7.36) which are not asymptotic to unity but the transforms will nevertheless be asymptotically flat in general, and we can change signs of individual x_i's and y_i's.] The resulting stationary solution is the corresponding non-linear superposition of n Kerr–NUT particles with arbitrary parameters. This interpretation is due to Neugebauer and Kramer[19] who calculated the $n = 2$ case explicitly by applying a four-fold Harrison transformation to flat space.[52] A remarkable feature discovered by these authors is that algebraic conditions among the parameters can be written down such that the z axis is regular between the particles, thus balancing gravitational attraction with rotational repulsion.

At this stage the solution (7.36), with factors $(1+y_i)/(1-y_i)$ included, is the most general seed solution for which the Kinnersley–Chitre $\beta^{(k)}$ transforms can be calculated. (Of course, transforms of transforms are a trivial further step.) The transforms of the general Weyl solution and the CTS solution with δ not an integer have not yet been found. But the superposition of n Kerr particles can be calculated with much less labor with Harrison or BZ transformations, which can be applied to any stationary solution. In fact, the superposition of n Kerr particles is precisely the general $2n$-fold Harrison or $2n$-fold BZ transform of flat space. The Kinnersley–Chitre $\delta = n$ solution is the limiting case where n of the s parameters are equal to $1/2\kappa$ and the remaining n are equal to $-1/2\kappa$. For example, the KC $\delta = 2$ solution is the product of two generalized (null) HKX transformations applied to flat space. Indeed, also, the generation of Kerr–NUT from Schwarzschild can be done with HKX transformations. In Eq. (4.22) of Ref. 7, which gives the double HKX transform of the Zipoy–Voorhees solution (7.35) with $s_1 = 1/2\kappa$, $s_2 = -1/2\kappa$, simply put $\delta = -1$ and change the sign of x. We defer further discussion of the generation of exact solutions to Sec. 8.

Now, it is quite possible that the class of seed solutions amenable to treatment with KC and nonnull HKX transformations could be enlarged using the integral equation formalism of Hauser and Ernst (HE).[21] Their method of generating new solutions from old is to solve the linear singular integral equation,

$$f_A{}^B(t) + \frac{1}{2\pi i}\int_C \frac{(f_A{}^X(s) + s^{-1}\epsilon_A{}^X)K_X{}^B(s)}{s-t}\,ds = 0, \tag{7.37}$$

for $f_A{}^B(t)$. Then

$$F'_{AB}(t) = [\epsilon_A{}^X + tf_A{}^X(t)]F_{XB}(t) \tag{7.38}$$

is the generating function for the new solution. The kernel $K_A{}^B(s)$ is defined by[53]

$$K_A{}^B(s) = (\det F)^{-1}F_{AX}(s)\tau^X{}_Y(s)F^{BZ}(s), \tag{7.39}$$

$$\begin{aligned}\tau^X{}_Y(s) &= \exp[\gamma^X{}_Y(s)] + \epsilon^X{}_Y \\ &= (1-\cos\delta)\epsilon^X{}_Y + \delta^{-1}(\sin\delta)\gamma^X{}_Y,\end{aligned} \tag{7.40}$$

where $\delta = (\det\gamma)^{1/2}$ and $\gamma_{AB}(s)$ depends on s only and satisfies $(\gamma_{AB}(s))^* = \gamma_{BA}(s^*)$. The contour C is a closed positively-oriented contour surrounding the origin in the complex s plane lying in the region of analyticity of $\gamma_{AB}(s)$ (annulus around $s = 0$) which is itself contained in the region of analyticity of $F_{AB}(s)$ and $f_A{}^B(s)$ (open neighborhood of $s = 0$).

In the first paper of Ref. 21, HE give an explicit construction of the tensor $\gamma_{AB}(s)$ in terms of the $\gamma^{(k)}$ and $\tau^{(k)}$ transformations of Kinnersley and Chitre.[3,4] They have already demonstrated, with particular examples, that the integral equation (7.37) can be solved by standard techniques in the known cases: (i) where the element of the Geroch group specified by $\gamma_{AB}(s)$ can be exponentiated for all stationary solutions (e.g., HKX) and (ii) where a general method of exponentiation is not available but the explicit transforms of particular solutions can be constructed (e.g., KC generation of Kerr from Schwarzschild). Since the kernel $K_A{}^B(s)$ of (7.37) depends on both the transformation and the seed solution, it would be of great interest to have a catalog of all kernels for which Eq. (7.37) can be solved by known methods.

More recently, HE have rederived their integral equation by posing an homogeneous Hilbert problem, and this derivation closely parallels the derivation of the "nonsoliton" part of the inverse scattering transform of Belinsky and Zakharov.[20] In a sense, HE have solved equation (A17) [equivalent to the eigenvalue problem (4.45)] by a corresponding inverse scatterng technique. The BZ integral equation is

$$\begin{aligned}\rho(\lambda) = {}& [\mathbf{I} - \mathbf{G}(\lambda)][\mathbf{I} + \mathbf{G}(\lambda)]^{-1} \\ & \times \left[\mathbf{I} - \frac{1}{\pi i}(P)\int_\Gamma \frac{\rho(\mu)}{\mu-\lambda}\,d\mu\right],\end{aligned} \tag{7.41}$$

where

$$\mathbf{G}(\lambda) = \psi(\lambda)\mathbf{G}_0(\lambda)\psi^{-1}(\lambda) \tag{7.42}$$

and $\mathbf{G}_0(\lambda)$ satisfies $\mathscr{D}_1\mathbf{G}_0 = \mathscr{D}_2\mathbf{G}_0 = 0$. (Refer to Sec. 5 for notation.) New solutions are generated from old according to Eq. (5.9b) where

$$\chi^{-1}(\lambda) = \mathbf{I} - \frac{1}{\pi i}\int_{\Gamma^+} \frac{\rho(\mu)}{\mu - \lambda}\, d\mu. \tag{7.43}$$

Two distinct but equivalent integral equations can be written down for χ instead of ρ. Here, Γ is the contour $|\mu| = \alpha$ in the complex μ plane[54] and, in (7.41) where λ lies on Γ, the integral takes its Cauchy principal value. In (7.43), Γ^+ is the same as Γ if λ is inside Γ, but is to be deformed just to the outside of $\mu = \lambda$ if λ is on Γ.

The key to the comparison of the methods of BZ and HE is equations (5.2), (5.9) and (5.10) which relate BZ's variables to the generating functions. According to (5.1f) and (5.2), $\mathbf{G}_0(\lambda)$ is a function of t only, not ρ and z, and so it corresponds to HE's $\tau^A{}_B(s)$. Then, comparing (7.42) with (7.39), we see that $\mathbf{G}(\lambda)$ takes the place of the kernel $K_A{}^B(s)$ except that the role of $F_{AB}(t)$ is taken over by $P_{AB}(t)$. In fact, equations (5.9b) and (7.38) have essentially the same form. Using (5.6), (5.7), (5.9b), (5.10), (5.13), (5.15a,b), (A23), (A34), and (A35b), we obtain the identification

$$\begin{aligned}\epsilon_A{}^X + tf_A{}^X(t) &= tS^{-1}(t)\Big\{ -[\det\chi(\lambda)]^{-1/2}\chi_A{}^Y(t)(\mu(t)\epsilon_Y{}^X + if_Y{}^X) \\ &\quad + \left(\frac{\det\chi(\lambda)}{\det\chi(0)}\right)^{1/2}\tilde{\chi}_A{}^Y(t)(\nu(t)\epsilon_Y{}^X + if_Y{}^X)\Big\}, \end{aligned}\tag{7.44}$$

where $\tilde{\chi}_A{}^B(t) = \chi(\alpha^2/\lambda)$ and is simply $\chi_A{}^B(t) = \chi(\lambda)$ with the sign of $S(t)$ changed, and $f_Y{}^X$ on the right-hand side is the SL(2) metric tensor defined by (A7). When the change of variable (5.2) is employed, the complex λ plane becomes a two-sheeted Riemann t surface. The two integral equations are not exactly equivalent since the contour Γ will cross over singularities of the integrand when it is deformed to coincide with C. This is manifest in the fact that (7.37) contains all of the generalized HKX transformations while (7.41) does not contain those for which $\chi(\mu)$ has double poles inside the contour Γ. When the soliton transformations are incorporated into the BZ scheme, it is then more general than the formalism of HE, but the latter is somewhat more elegant and easier to manage when transformations in the Geroch group only are under consideration. Although we have remarked that P_{AB} is a better field variable than F_{AB} for the pure soliton transformations (except for the HKX confluent cases), it would be interesting to see the HE formalism expanded to include mixed soliton and nonsoliton transformations.

8. CONCLUSION: GENERATION OF EXACT SOLUTIONS

Now that we have analyzed in detail the interrelationships between the various solution generating techniques for the stationary axisymmetric gravitational field equations, we can reasonably attempt to catalog all of the exact asymptotically flat solutions which are constructible with the present methods.

According to the conjectures, any solution generating algorithm which adds one or more new multipole parameters at each iteration will give the general asymptotically flat stationary solution in the limit of infinitely many iterations, if the general asymptotically flat Weyl solution is taken as the seed or beginning solution. For example, exactly one additional parameter at each step may be achieved by applying a HKX rank-0, followed by a rank-1, rank-2, and so on, with same s parameters. However, there are relatively simple exact solutions (e.g., Tomimatsu–Sato,[40] Kinnersley–Chitre[5]) which cannot be generated from any Weyl solution with this algorithm in finitely many iterations.

The Cosgrove–Tomimatsu–Sato (CTS) solutions[13–16] and their enlargements[11] with δ not an integer and angular momentum parameter q different from 0 or ± 1 cannot be obtained from any Weyl solution in a finite number of iterations of any of the algebraic transformations discussed in this paper for the simple reason that the transformations are algebraic and the CTS solutions involve transcendental functions, $H_4(\eta)$ and $K(\nu,\eta)$. The cases where δ is an integer are the soliton solutions representing a nonlinear superposition of δ Kerr–NUT particles or contractions thereof and can be generated from flat space by 2δ Harrison or BZ transformations or from Weyl solutions of the form (7.36) by HKX transformations.

Now, all of the exact solutions outside the CTS class which are either already known or can be calculated with the presently known methods may be generated from the Weyl solutions by a combination of the methods described in this paper. So the Weyl solutions and the CTS solutions are the only seed solutions we need consider. (Notwithstanding these statements, the conjectures in Hauser and Ernst's work imply that a class of kernels should exist for the integral equation (7.37) such that the ZV solutions[28] transform into the CTS solutions. Since none of the known algebraic transformations have this property, such an integral equation representation would be highly desirable.) Considering the variety of transformations available, however, the list of solutions generated from Weyl and CTS would be somewhat chaotic unless we take into account the relationships between the transformations and some of their specific properties.

We know that the following transformations preserve asymptotic flatness (subject to provisos in Appendix B) and are written in order of increasing generality:

original HKX (two parameters);

generalized HKX (three parameters);

{double Harrison (four parameters);
{Belinsky–Zakharov two-soliton (four parameters);

4-fold Neugebauer $I_2I_1I_2I_1$ (eight parameters, restricted to six for asymptotic flatness preservation).

We claim that all solutions which can be generated from the Weyl or CTS solutions by any finite combination of the above transformations can be generated with the original HKX transformations alone (including confluences such as the rank-one, etc.). The most elegant formulas, however, result when we use $2n$-fold Harrison or BZ $2n$-soliton transformations only. A determinant expression for the $2n$-fold Harrison or BZ transform of the general Weyl solutions is given in Ref. 19. It is interpreted there as a non-linear superposition of the original Weyl solution and n Kerr–NUT particles.

In Sec. 6, we demonstrated explicitly that the double Harrison or BZ two-soliton transform of any Weyl solution can always be obtained as the double HKX transform of a different Weyl solution. Also, the original HKX with $q^{11}\neq 0$ can be replaced by one with $q^{22}\neq 0$ or the sign of $S(s)$ can be changed. The same property holds for the CTS solutions (including the asymptotically nonflat $h\neq 0$ cases in Refs. 12, 13, and 16). The double Harrison transform of the CTS solution is calculated in Ref. 11. The generating function $F_{AB}(t)$ and hence the HKX transform can also be obtained by a similar line of reasoning, with the aid of formulas in Secs. 2 and 3 here. These results arise because the double HKX transformation (either $q^{11}\neq 0$ or $q^{22}\neq 0$, other $q^{AB}=0$) can be factorized as a fourfold BZ transformation such that the first two factors transform Weyl to Weyl and CTS to CTS ($\delta\to\delta\pm 1$). Since BZ transformations commute, a $2n$-fold Harrison or BZ transform of any Weyl or CTS solution is again a $2n$-fold HKX transform of a different Weyl or CTS solution. Further, since the generalized HKX transformation is a confluent form of the BZ two-soliton transformation, the generalized HKX transform of a Weyl or CTS solution is the same as a combined HKX rank-0 and rank-1 transform of a different Weyl or CTS solution. This was also demonstrated explicitly in Sec. 6 for the Weyl solutions. Thus asmptotically flat solutions which can be generated from the Weyl or CTS solutions by any combination of HKX, generalized HKX, Harrison or BZ transformations can also be obtained by iterating any single one of these types of transformation (including special cases and limits).

Products of Neugebauer's Bäcklund transformations I_1 and I_2 (or, equivalently, Cosgrove's groups **PQ** and **LQ̃**, respectively) of the form, $\cdots I_2I_1I_2I_1$, are more general still but do not give any further asymptotically flat solutions. We need to know when such products preserve asymptotic flatness. Recall that Harrison's Bäcklund transformation can be factorized into the forms, I_2I_1 and I_1I_2, where the parameters in the second factor depend on those in the first factor, the latter being arbitrary. The general fourfold product $I_2I_1I_2I_1$ with eight parameters can be written as $I_2(I_2I_1)(I_2I_1)I_1$ where the factors in parentheses are Harrison transformations. Since I_1 does and I_2 does not, in general, preserve asymptotic flatness, this latter product will preserve asymptotic flatness only when the final I_2 is the identity. Thus $I_2I_1I_2I_1$ preserves asymptotic flatness when the second I_2I_1 is a Harrison transformation. But I_1 transforms Weyl to Weyl and leaves CTS invariant (up to a change of NUT parameter). Consequently, the $I_2I_1I_2I_1$ transform of a Weyl or CTS solution, if asymptotically flat, is the same as the double Harrison transform of a different Weyl or the same CTS solution, respectively.

An alternative argument is to write

$$I_2I_1I_2I_1=(I_2I_1)(I_1I_2)(I_2I_1)I_1, \tag{8.1}$$

where the factors in parentheses are Harrison or identity transformations. Thus $I_2I_1I_2I_1$ preserves asymptotic flatness when one of the three factors in parentheses on the right is the identity. Similarly, we can rewrite arbitrary products in the following manner:

$$I_2I_1\cdots I_2I_1I_2I_1=(I_2I_1)(I_1I_2)\cdots(I_1I_2)(I_2I_1)I_1, \tag{8.2}$$

$$I_1I_2\cdots I_2I_1I_2I_1=(I_1I_2)(I_2I_1)\cdots(I_1I_2)(I_2I_1)I_1, \tag{8.3}$$

where the factors in parentheses are Harrison or identity transformations. (If I_2 occurs first, write $I_2=I_2I_1I_1^{-1}$.) These preserve asymptotic flatness when the right-hand side contains an even number of non-trivial Harrison transformations.[55] Thus these products do not give any further new solutions when the seed solutions are the Weyl or CTS solutions.

With one possible exception, the Kinnersley–Chitre $\beta^{(k)}$ and nonnull HKX transformations and the integral equation methods have not yet yielded solutions distinct from those which can already be generated with the (generalized) null HKX, Harrison, BZ and Neugebauer transformations. However, the integral equation methods are still being explored and when a catalog of kernels is developed, we may find new solutions. The $s=\infty$ limit of the nonnull HKX transformation gives the same result as the $s=\infty$ limit of the original HKX when applied to the Weyl solutions, but appears to give a distinct new solution when applied to the CTS solutions.

The main success of the $\beta^{(k)}$ transformations was the generation of Kerr–NUT from Schwarzschild and the discovery of the Kinnersley–Chitre (KC) solutions for $\delta\geqslant 2$. We have already remarked that the KC $\delta=n$ solution is a limiting case of the superposition of n Kerr–NUT particles which is the $2n$-fold Harrison or BZ $2n$-soliton transform of flat space. But according to the above discussion, these solutions must also be constructible with HKX transformations. In fact, the generation of Kerr–NUT from Schwarzschild is contained in Eq. (4.22) of Ref. 7 which gives the double HKX transform of the Zipoy–Voorhees solution (7.35) in the limit when the two centers[26,42] coincide with the foci of spheroidal coordinates ($s_1=1/2\kappa$, $s_2=-1/2\kappa$). Putting $\delta=1$ gives a subclass of the KC $\delta=3$ solution, while putting $\delta=-1$ gives Kerr–NUT. The replacement of (x,y,δ) by $(-x,-y,-\delta)$ changes the right-hand side only of Eq. (4.22) of Ref. 7. This symmetry corresponds to changing the signs of $S(s_1)$ and $S(s_2)$ or replacing the HKX transformations with $q^{11}\neq 0$ by ones with $q^{22}\neq 0$. The authors also remarked that the case where $\delta=0$ gives a subclass of the KC $\delta=2$ solution, where the Tomimatsu–Sato parameter q is unity. The full KC $\delta=2$ solution can be constructed by putting $\delta=-2$ and then applying two further HKX rank-1 transformations with $s=\pm 1/2\kappa$. The same result can also be obtained by applying two generalized HKX transformations to flat space ($\delta=0$). More generally, the superposition of n Kerr–NUT particles can be obtained from the superposition of n Schwarzschild particles given by (7.36) by first changing the sign of the spheroidal coordinates x_i in (7.36) and then applying $2n$ HKX transformations whose centers coincide with the $2n$ foci. These statements are straightforward corollaries of theorem (6.66) and Eq. (6.65a).

ACKNOWLEDGMENTS

I wish to thank F. J. Ernst, I. Hauser, D. Kramer, G. Neugebauer, and many other authors for sending me preprints of their recent work. Also, special thanks are due to

my wife, Bronwynne, for her invaluable assistance in the typing of a difficult manuscript.

APPENDIX A: PROPERTIES OF THE GENERATING FUNCTIONS

In Ref. 2, Kinnersley compares different covariant formulations for the stationary axisymmetric vacuum and electrovac fields and decides in favor of an **L**–covariant SL(2,R) tensor formulation. In the vacuum case, the groups **P** and **L** are isomorphic and so the **P**–covariant and **L**–covariant formulations are isomorphic, except that the complex potential notation in the latter is more elegant. In addition, there is the further choice between SO(2,1) vector potentials (as used by Geroch[9]) and SL(2) tensor potentials,[2-8] and a comparison of the recurrence relations in Ref. 9 and Ref. 3 shows that the latter has proved to be much more successful.

Vectors h_A and second-rank tensors K_{AB} transform under $(L)_\beta$ according to

$$h'_A = b_A{}^C h_C, \quad K'_{AB} = b_A{}^C b_B{}^D K_{CD}, \tag{A1}$$

and so on for higher-rank tensors, where $b_1{}^1 = \beta_1$, $b_1{}^2 = \beta_2$, $b_2{}^1 = \beta_3$, $b_2{}^2 = \beta_4$, $\beta_1\beta_4 - \beta_2\beta_3 = 1$. Indices are raised and lowered with $\epsilon_{AB} = \epsilon^{AB}$ ($\epsilon_{11} = \epsilon_{22} = 0$, $\epsilon_{12} = -\epsilon_{21} = 1$) according to the rule for spinors:

$$h^A = \epsilon^{AX} h_X, \quad h_A = \epsilon_{XA} h^X. \tag{A2}$$

Observe that

$$\epsilon_A{}^B = -\epsilon^B{}_A = \delta^B_A. \tag{A3}$$

Two very useful general properties of tensors are

$$K_{XA} K^X{}_B = K_{AX} K_B{}^X = (\det K)\epsilon_{AB}, \tag{A4}$$

$$L_{AB} - L_{BA} = L_X{}^X \epsilon_{AB}. \tag{A5}$$

In Eq. (A5), L may be a tensor or product of tensors of any rank with just two indices shown. For example, (A4) and (A5) imply

$$K_{AB} K_{CD} - K_{AD} K_{CB} = (\det K)\epsilon_{AC}\epsilon_{BD}. \tag{A6}$$

Equation (A4) shows that the inverse of the matrix K_{AB} is $K^{BA}/(\det K)$ and the inverse of $K_A{}^B$ is $-K^B{}_A/(\det K)$. A second-rank tensor q_{AB} is said to be null if $q_{XY} q^{XY} = 0$. This means that $q_{XA} q^X{}_B = 0$. All vectors are identically null ($h_X h^X = 0$). Under **L**, a symmetric null tensor q_{AB} (not identically zero) can be transformed so that either q_{11} or q_{22} is its only nonvanishing component. A general nonnull tensor q_{AB} can be transformed so that q_{12} and q_{21} are its only nonvanishing components.

All the tensor potentials in the hierarchy are derived from the 2×2 block f_{AB} of the metric tensor:

$$f_{11} = f, \quad f_{12} = f_{21} = -f\omega, \quad f_{22} = f\omega^2 - \rho^2 f^{-1}. \tag{A7}$$

The nonsymmetric tensor version of the Ernst potential is defined by

$$H_{AB} = f_{AB} + i\psi_{AB}, \quad \nabla\psi_{AB} = -\rho^{-1} f_A{}^X \tilde{\nabla} f_{XB}, \tag{A8a,b}$$

where $\nabla = (\partial/\partial\rho, \partial/\partial z)$, $\tilde{\nabla} = (\partial/\partial z, -\partial/\partial\rho)$. Corollaries of these formulas are

$$f_{XA} f^X{}_B = -\rho^2 \epsilon_{AB}, \quad H_{AB} - H_{BA} = 2iz\epsilon_{AB}, \tag{A9a,b}$$

$$\nabla f_{AB} = \rho^{-1} f_A{}^X \tilde{\nabla}\psi_{XB}, \quad \nabla H_{AB} = -i\rho^{-1} f_A{}^X \tilde{\nabla} H_{XB}. \tag{A10a,b}$$

The hierarchy of potentials, $H^{(n)}_{AB}$ and $N^{(m,n)}_{AB}$, is defined by the consistent set of recurrence relations,

$$H^{(0)}_{AB} = i\epsilon_{AB}, \quad H^{(1)}_{AB} = H_{AB}, \quad H^{(n)}_{AB} = iN^{(0,n)}_{AB}, \tag{A11a,b,c}$$

$$N^{(m,n+1)}_{AB} - N^{(m+1,n)}_{AB} = iN^{(m,1)}_{AX} H^{(n)X}{}_B, \quad (m,n) \neq (-1,0) \text{ or } (0,-1), \tag{A12}$$

$$N^{(m,n)}_{AB} - N^{*(n,m)}_{BA} = H^{*(m)}_{XA} H^{(n)X}{}_B, \quad (m,n) \neq (0,0), \tag{A13}$$

$$\nabla N^{(m,n)}_{AB} = H^{*(m)}_{XA} \nabla H^{(n)X}{}_B, \tag{A14}$$

* denoting complex conjugate (assuming f_{AB} real; if f_{AB} complex, * does not change f_{AB}, ψ_{AB}, etc., but only changes the sign of i in the above definitions). We define also $N^{(m,0)}_{AB} = 0$ for $m \geqslant 1$, $N^{(0,0)}_{AB} = \epsilon_{AB}$, and all potentials with a negative index are taken to be zero.

The most efficient method for calculating the potentials is to calculate the generating functions,

$$F_{AB}(t) = \sum_{n=0}^{\infty} H^{(n)}_{AB} t^n, \tag{A15a}$$

$$G_{AB}(s,t) = \sum_{m,n=0}^{\infty} N^{(m,n)}_{AB} s^m t^n, \tag{A15b}$$

for which methods are given in Refs. 4, 5, and 7 and here. In this analysis, the spherical radial and angular coordinates,[42]

$$S(t) = [(1-2tz)^2 + 4t^2\rho^2]^{1/2}, \tag{A16a}$$

$$\zeta(t) = \left(\frac{1-2tz-2it\rho}{1-2tz+2it\rho}\right)^{1/2}, \tag{A16b}$$

frequently arise. $S(t)$ was introduced by Kinnersley and Chitre[4] and $\zeta(t)$ by Harrison[18] in quite different contexts. Elimination of $N^{(1,n)}_{AB}$ from (A12) with $m = 0$ and (A14) with $m = 1$ gives

$$[-it^{-1}(1-2tz)\epsilon_{XA} - 2f_{XA}]\nabla F^X{}_B(t) = (\nabla H_{AX}) F^X{}_B(t).$$

Then (A4) and (A10b) give the partial differential equation,

$$\nabla F_{AB}(t) = itS^{-2}(t)[(1-2tz)\nabla H_{AX} - 2t\rho\tilde{\nabla} H_{AX}] F^X{}_B(t), \tag{A17}$$

which is the most important defining relation for $F_{AB}(t)$, but there are other methods better suited for calculation in particular cases. A first integral[4] which is now easily checked is

$$F_{XA}(s) F^X{}_B(s) = F_{XA}(s) F_B{}^X(s) = -S^{-1}(s)\epsilon_{AB}. \tag{A18}$$

With the aid of this integral, (A11)–(A13) imply the following algebraic relations:

$$G_{AB}(s,t) = \frac{s}{s-t}\epsilon_{AB} + \frac{tS(s)}{s-t} F_{XA}(s) F^X{}_B(t), \tag{A19}$$

$$G_{AB}(0,t) = -iF_{AB}(t), \quad G_{AB}(s,0) = \epsilon_{AB}, \tag{A20a,b}$$

$$D_1 G_{AB}(0,t) = -t^{-1}(\epsilon_{AB} + iF_{AB}(t)) - H_{AX} F^X{}_B(t), \tag{A21}$$

$$D_2 G_{AB}(s,0) = s^{-1}[\epsilon_{AB} - iS(s) F_{BA}(s)], \tag{A22}$$

$$F^*{}_{AB}(t) = S^{-1}(t)[2it f_{AX} F^X{}_B(t) - (1-2tz) F_{AB}(t)], \tag{A23}$$

$$G^*{}_{AB}(s,t) = G_{BA}(t,s) + \epsilon_{AB} + F_{XA}(s) F^{*X}{}_B(t). \tag{A24}$$

[*Proof*: (A20) and (A24) follow immediately from (A11) and (A13), respectively. Multiply (A12) by $s^m t^n$ and sum over m, n to obtain a relation between $G_{AB}(s,t)$, $D_2 G_{AB}(s,0)$ and $F_{AB}(t)$; put $s=0$ to get (A21) and put $t=s$ and use (A18) to get (A22). Then (A19) follows automatically. Finally, take $\partial/\partial t$ of (A24) at $t=0$ and use (A21) and (A22) to get (A23).] It is easy to show that equations (A17), (A18) and (A23) are respected by gauge transformations of the form (1.14).

By combining (A19)–(A24), a number of other useful relations can be deduced:

$$G_{AB}(s,s) = \epsilon_{AB} - sS(s)F_{XA}(s)DF^{X}{}_{B}(s), \tag{A25}$$

$$G^{*}{}_{AB}(s,s) - G_{BA}(s,s) = \epsilon_{AB} + F_{XA}(s)F^{*X}{}_{B}(s), \tag{A26}$$

$$G_{AB}(s,s) - G_{BA}(s,s) = \left(1 + \frac{1-2sz}{S^2(s)}\right)\epsilon_{AB}, \tag{A27}$$

$$G_{AB}(s,s) - G^{*}{}_{AB}(s,s) = 2isS^{-1}(s)f_{XY}F^{X}{}_{A}(s)F^{Y}{}_{B}(s), \tag{A28}$$

$$F^{*}{}_{AX}(s)F_{B}{}^{X}(s) = S^{-2}(s)[2isf_{AB} + (1-2sz)\epsilon_{AB}], \tag{A29}$$

$$F^{*}{}_{XA}(s)F^{X}{}_{B}(s) = (1-2sz)S^{-2}(s)\epsilon_{AB} + 2isS^{-1}(s)f_{XY}F^{X}{}_{A}(s)F^{Y}{}_{B}(s), \tag{A30}$$

$$f_{XY}F^{*X}{}_{A}(s)F^{*Y}{}_{B}(s) = f_{XY}F^{X}{}_{A}(s)F^{Y}{}_{B}(s), \tag{A31}$$

$$sS(t)G_{AB}(s,t) - tS(s)G_{BA}(t,s) = \frac{s^2S(t) - t^2S(s)}{s-t}\epsilon_{AB}, \tag{A32}$$

$$G_{AX}(t_1,s)G^{X}{}_{B}(s,t_2) = \frac{s(t_1-t_2)}{(t_1-s)(s-t_2)}G_{AB}(t_1,t_2) - \frac{t_1}{t_1-s}G_{AB}(s,t_2) - \frac{s}{s-t_2}G_{AB}(t_1,s). \tag{A33}$$

In some cases, it is desirable to split F_{AB} into its real and imaginary parts. Define P_{AB} and Q_{AB} by

$$F_{AB} = P_{AB} + iQ_{AB}, \quad F^{*}{}_{AB} = P_{AB} - iQ_{AB}. \tag{A34}$$

Then, from (A18) and (A23),

$$P_{AB}(t) = -\frac{T_2^2(t)}{t\rho^2}f_{AX}Q^{X}{}_{B}(t), \tag{A35a}$$

$$Q_{AB}(t) = -\frac{T_1^2(t)}{t\rho^2}f_{AX}P^{X}{}_{B}(t), \tag{A35b}$$

$$P_{XA}(t)P^{X}{}_{B}(t) = -\frac{T_2^2(t)}{S^2(t)}\epsilon_{AB}, \tag{A36a}$$

$$Q_{XA}(t)Q^{X}{}_{B}(t) = \frac{T_1^2(t)}{S^2(t)}\epsilon_{AB}, \tag{A36b}$$

$$P_{AX}(t)Q_{B}{}^{X}(t) = -Q_{AX}(t)P_{B}{}^{X}(t) = tS^{-2}(t)f_{AB}, \tag{A37}$$

$$P_{XA}(t)Q^{X}{}_{B}(t) = P_{XB}(t)Q^{X}{}_{A}(t) = \frac{T_1^2(t)}{t\rho^2}f_{XY}P^{X}{}_{A}(t)P^{Y}{}_{B}(t) = -\frac{T_2^2(t)}{t\rho^2}f_{XY}Q^{X}{}_{A}(t)Q^{Y}{}_{B}(t) = tS^{-1}(t)f_{XY}F^{X}{}_{A}(t)F^{Y}{}_{B}(t), \tag{A38}$$

where $T_1(t)$ and $T_2(t)$ are defined by (3.13) and (3.14).

Frequently in this paper, it was found necessary to express all of the generating function components in terms of just F_{11} and F_{12}. Equation (A.23) with $A=1$ gives

$$F_{2B}(t) = (2itf)^{-1}[(1-2tz-2itf\omega)F_{1B}(t) + S(t)F^{*}_{1B}(t)]. \tag{A39}$$

When F_{2B} is eliminated from the right-hand side of (A19), we obtain

$$2s(s-t)G_{AB}(s,t) = 2s^2\epsilon_{AB} - iS(s)f^{-1}[(s-t)F_{1A}(s)F_{1B}(t) + sS(t)F_{1A}(s)F^{*}_{1B}(t) - tS(s)F^{*}_{1A}(s)F_{1B}(t)]. \tag{A40}$$

The pole at $t=s$ is absent since

$$F^{*}{}_{11}(s)F_{12}(s) - F_{11}(s)F^{*}{}_{12}(s) = 2isS^{-2}(s)f, \tag{A41}$$

which is the $AB=11$ component of Eq. (A29). When F_{2B} is eliminated from (A17), the result is Eq. (3.2).

In Ref. 5, Kinnersley and Chitre give differential equations which involve F_{11} or F_{12} alone (not together with $F^{*}{}_{11}$ or $F^{*}{}_{12}$). Equations (2.8) and (2.11) of that reference are

$$f_{11}\nabla_3{}^2F_{1B}(t) = \nabla H_{11}\cdot\nabla F_{1B}(t), \tag{A42}$$

$$(1-2tz)\tilde{\nabla}H_{11}\cdot\nabla F_{1B}(t) + 2t\rho\nabla H_{11}\cdot\nabla F_{1B}(t) = it(\nabla H_{11}\cdot\tilde{\nabla}H_{12})F_{1B}(t), \tag{A43}$$

where $\nabla_3{}^2 = \partial^2/\partial\rho^2 + \partial^2/\partial z^2 + \rho^{-1}\partial/\partial\rho$. Equation (A43) may be solved by the method of characteristics and the general integral involving one arbitrary function of one variable should then be substituted into (A42) to give a second-order linear ordinary differential equation for the unknown function. Any two linearly independent solutions may be taken to be F_{11} and F_{12}, subject to (A11a,b), because of gauge freedom [see (1.14) and (1.15)]. A qualitatively similar pair of equations can also be written down for F_{2B}, but are not needed as F_{2B} can always be obtained algebraically from the $A=1$ component of Eq. (A17), or from (A39) if $F^{*}{}_{11}$ can be obtained unambiguously from F_{11} (no problem when f_{AB} real).

An alternative method is to exploit the fact that the $(Q)_{4s}$ transform of the Ernst potential can be obtained by algebraic methods once a total Riccati equation has been solved for a pseudopotential $\alpha = -\zeta q$. A Riccati equation is equivalent to a second-order linear ordinary differential equation and is therefore a much simpler type than Eq. (A43). The component F_{12} can then be obtained immediately from Eq. (3.3), and the other components from (3.1) and (A39).

APPENDIX B: ASYMPTOTIC FLATNESS

If a stationary axisymmetric vacuum solution is to represent the exterior gravitational field of an isolated rotating collection of matter, then the metric of space–time should be asymptotic to Minkowski space–time as $r\to\infty$, uniformly with respect to θ, where (r,θ) are spherical coordinates defined by

$$\rho = r\sin\theta, \quad z = r\cos\theta. \tag{B1}$$

Thus the Ernst potential assumes a Taylor (Laurent) expansion,

$$\mathscr{E} = 1 - 2mr^{-1} + A_1(\theta)r^{-2} + A_2(\theta)r^{-3} + \ldots, \tag{B2}$$

where m is the (real) Schwarzschild mass and $A_n(\theta)$, $n \geqslant 1$, is a complex-valued polynomial in $\cos\theta$. Substitution of (B2) into Ernst's equation [a special case of (A42)] gives an imhomogeneous Legendre equation of degree n for $A_n(\theta)$. Thus $A_n(\theta)$ is the sum of a particular integral depending on $A_0 = -2m, A_1, \ldots, A_{n-1}$ and a complimentary function, $(2a_n + ib_n)P_n(\cos\theta)$, where a_n and b_n are arbitrary real constants representing the multipole parameters. ($a_0 = -m$ is the mass, $b_1 = -2J$ is angular momentum.) The linearly independent solution $Q_n(\cos\theta)$ is to be rejected since it has logarithmic singularities on the z axis ($\theta = 0, \pi$).

In our work, it is not convenient to insist that all solutions adopt the strict asymptotically Minkowskian form (B2). The definition of asymptotic flatness we adopt is the slightly more general form

$$\mathscr{E} = C + iD + (2a_0 + ib_0)r^{-1} + A_1(\theta)r^{-2} + A_2(\theta)r^{-3} + \ldots, \tag{B3}$$

where C, D, a_0 and b_0 are real constants ($C \neq 0$) and $A_n(\theta)$, $n \geqslant 1$, is a polynomial in $\cos\theta$ of degree n, being a linear combination of $(2a_n + ib_n)P_n(\cos\theta)$ and a particular integral satisfying an inhomogeneous Legendre equation as before. The angular momentum monopole b_0 is the NUT parameter. The Ernst potential (B2) transforms into the form (B3) under a **P** transformation. Conversely, an element of **P** can always be found such that any Ernst potential of the form (B3) transforms back to the form (B2).

There are many exact solutions not possessing the asymptotic form (B3) which have asymptotically vanishing Riemann tensors. These are excluded from the present definition but some of these solutions are potentially physically meaningful and deserve further study. Examples are certain subclasses of the $h \neq 0$ CTS solutions,[13,16] and contractions such as the Ernst solutions.[16,56] Recently, Harrison[57] has shown that large classes of such solutions can be generated from the Weyl solutions with a single Harrison transformation.

The series (B3) can be extended to the left through positive powers of r. If all the coefficients are polynomials in $\cos\theta$, such solutions will be called "multipole type." If positive powers only of r occur, the solution could describe the vacuum interior of a rotating hollow body. The general case could describe the hollow region between two rotating shells (provided the NUT singularity is absent). It is widely believed that the Geroch group acts transitively on these solutions. Multipole character is also preserved by **L**, **Q**, $\tilde{\mathbf{Q}}$ and Harrison transformations, but not by (I) or single BZ transformations.

The asymptotic expansions of higher potentials, generating functions and pseudopotentials can be determined in a straightforward manner from their defining relations and (B3). The generating function components $F_{11}(t)$ and $F_{12}(t)$ are best calculated from equations (A42) and (A43). We find the asymptotic behavior,

$$F_{1B}(t) = F_{1B}(r,\theta,t) = F^{(0)}_{1B}(t)r^{-1} + F^{(1)}_{1B}(\theta,t)r^{-2} + O(r^{-3}), \tag{B4}$$

where $F^{(0)}_{1B}$ is independent of θ, $F^{(1)}_{1B}$ is linear in $\cos\theta$, and higher coefficients are polynomials in $\cos\theta$. $F^{(0)}_{1B}$ is an arbitrary function of t subject to (A41) and the remaining coefficients are then determined in terms of $F^{(0)}_{1B}$ and the multipole parameters. [The coefficient of $\cos\theta$ in $F^{(1)}_{1B}$ is determined separately by eliminating F_{2B} from (A17) and (A39).] From (A39), we find

$$F_{2B}(t) = F^{(0)}_{2B}(t) + O(r^{-1}), \tag{B5}$$

where $F^{(0)}_{2B}$ is independent of θ, the $O(r^{-1})$ term is linear in $\cos\theta$, etc. These formulas hold equally well for $t = \infty$, but break down for $t = 0$. From (A16a,b),

$$S(t) = \epsilon[2tr - \cos\theta + O(r^{-1})], \tag{B6a}$$

$$\zeta(t) = -\epsilon e^{i\theta}\left[1 + \frac{i\sin\theta}{2tr} + O(r^{-2})\right], \tag{B6b}$$

where $\epsilon = \operatorname{sgn} t$, $t \neq 0$. The pseudopotential q can be found either from (4.23) or (4.35). The result is

$$q = q^{(0)}(s)[1 + iq^{(1)}(\theta,s)\, r^{-1} + O(r^{-2})], \tag{B7}$$

where $q^{(0)}(s)$ is an arbitrary phase, $q^{(0)}q^{(0)*} = 1$, independent of θ, and $q^{(1)}$ is real and linear in $\cos\theta$, etc.

It is now easy to test the Harrison, BZ and generalized HKX transformations for asymptotic flatness preservation. First, the transform of the Ernst potential under the double Harrison or BZ two-soliton transformation is given by Eq. (5.39). From (B3), (B6b) and (B7), we get

$$\mathscr{E}'' = C + iD + \left\{(2a_0 + ib_0) + \frac{(s_1 - s_2)C}{s_1 s_2[\epsilon_1 q_1^{(0)}(s_1) - \epsilon_2 q_2^{(0)}(s_2)]}\right\} r^{-1} + O(r^{-2}), \tag{B8}$$

where $\epsilon_1 = \operatorname{sgn} s_1$, $\epsilon_2 = \operatorname{sgn} s_2$. It is easy to see that the coefficients of higher powers of r^{-1} are polynomials in $\cos\theta$. Therefore, the double Harrison or double BZ transformation preserves asymptotic flatness, provided $\epsilon_1 q_1^{(0)}(s_1) \neq \epsilon_2 q_2^{(0)}(s_2)$. This supplementary condition means that it is always possible to choose one particular integration constant in the second Harrison or BZ transformation such that asymptotic flatness is not preserved. For example, the transform of flat space is the full Kerr–NUT class including an asymptotically nonflat limiting case, $\xi = -ie^{i\lambda}y$ [see (1.11) and (2.5)]. From (5.38), $F''_{1B}(t)$ has the correct asymptotic form (B4). Under a single Harrison transformation (4.44),

$$\mathscr{E}' = r(a + b\cos\theta) + O(1), \quad F'_{1B}(t) = O(1), \tag{B9a,b}$$

a, b complex constants, which is of multipole type but not asymptotically flat according to our definition.

It is a simple corollary of theorem (6.46) and (B8) that the generalized HKX transformation preserves asymptotic flatness. This can also be proved directly from either (6.34) or (6.37). The nonnull HKX transformation also preserves asymptotic flatness. This follows from the group property (6.40) and the infinitesimal transformation law (6.41).

The transformations **Q** and $\tilde{\mathbf{Q}}$ need to be treated differently as they change the canonical coordinates. Let a prime denote the $(Q)_{4s}$ or $(\tilde{Q})_{4s}$ transform. Then, from (2.2), (2.3a), and (B1),

$$r = \frac{r'}{S'(-s)}, \quad \cos\theta = \frac{\cos\theta' + 2sr'}{S'(-s)}, \tag{B10a,b}$$

$$S(t) = \frac{S'(t-s)}{S'(-s)}, \quad S(s) = \frac{1}{S'(-s)}. \tag{B11a,b}$$

The asymptotic behavior of the transformed solution ($r' \to \infty$) depends on the behavior of the original solution in the neighborhood of the point $(\rho,z) = (0,1/2s)$. The latter behavior is basically the same for all multipole type solutions, irrespective of their asymptotic form, provided $(0,1/2s)$ is not some special singular point (NUT singularities are allowed). To calculate asymptotic expansions for $\mathscr{E}$, $F_{1B}(t)$, and $F_{1B}(s)$ as $r' \to \infty$, we must rewrite Eqs. (A42) and (A43) in terms of coordinates (r',θ') and solve for the coefficients once again. The results are:

$$\mathscr{E} = \mathscr{E}^{(0)}(s) + \mathscr{E}^{(1)}(s)(\cos\theta')r'^{-1} + O(r'^{-2}), \tag{B12}$$

$$F_{1B}(t) = F^{(0)}_{1B}(s,t) + O(r'^{-1}), \quad t \neq 0,s, \tag{B13}$$

$$F_{1B}(s) = \tilde{F}^{(0)}_{1B}(s)r' + O(1), \tag{B14}$$

$$F_{2B}(t) = O(1), \quad F_{2B}(s) = O(1). \tag{B15a,b}$$

In each case, the leading term is independent of θ', the next is linear in $\cos\theta'$ and further coefficients are polynomials in $\cos\theta'$.

It is now easy to see from Eqs. (3.1a,b,c) that $(Q)_{4s}$ transforms multipole type solutions to asymptotically flat solutions, provided $(\rho,z) = (0,1/2s)$ is not a singular point of the original solution. Also, (2.30a,b) shows that $F'_{1B}(t) = O(r'^{-1})$. The asymptotic regions of the original solution map onto the neighborhood of the point $(\rho',z') = (0,-1/2s)$ of the transformed solution. If the original solution were asymptotically non-flat, then this point would be a singularity of the latter solution and the inverse transformation $(Q)_{-4s}$ would then not be asymptotic flatness preserving. Similarly, from Eqs. (2.17), (2.18) and (3.6a), multipole type solutions transform under $(\tilde{Q})_{4s}$ into solutions whose asymptotic behavior is given by (B9) (with primes on r and θ).

From the preceding discussion, I_2 preserves multipole character while I_1 maps multipole type solutions to asymptotically flat solutions. This means that I_1I_2 preserves asymptotic flatness provided it is not a Harrison transformation. The same holds for $I_1I_2I_1$. According to (8.3), both these transformations can be rewritten as a product of an I_1 and two Harrison transformations. Similarly, the right-hand side of (8.3) preserves asymptotic flatness because it contains an even number of Harrison transformations,[55] and the left-hand side does so because the final I_1 maps a multipole type solution to an asymptotically flat solution.

[1]These claims are still conjecture, but there is an overwhelming body of evidence. Hauser and Ernst, in particular, seem very close to providing a proof in their context (see Ref. 21). It should be emphasized that the conjectures can only apply to "multipole type" solutions which are defined in Appendix B. In Refs. 4 and 5, Kinnersley and Chitre identified two infinite-dimensional subgroups of the Geroch group (Ref. 9), the former acting transitively on the class of multipole type Weyl solutions which includes flat space, the latter leaving flat space invariant. These results show that flat space is sufficient as a beginning solution and that the Geroch group is infinitely multiply transitive.

[2]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977) (Paper I).
[3]W. Kinnersley and D. M. Chitre, J. Math. Phys. **18**, 1538 (1977) (Paper II).
[4]W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 1926 (1978) (Paper III).
[5]W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 2037 (1978) (Paper IV).
[6]C. Hoenselaers, J. Math. Phys. **20**, 2526 (1979) (Paper V).
[7]C. Hoenselaers, W. Kinnersley, and B. C. Xanthopoulos, J. Math. Phys. **20**, 2530 (1979) (Paper VI).
[8]Related unnumbered papers are: W. Kinnersley, J. Math. Phys. **14**, 651 (1973); C. Hoenselaers, J. Math. Phys. **17**, 1264 (1976); D. M. Chitre, J. Math. Phys. **19**, 1625 (1978); W. Kinnersley and D. M. Chitre, Phys. Rev. Lett. **40**, 1608 (1978); C. Hoenselaers, W. Kinnersley and B. C. Xanthopoulos, Phys. Rev. Lett. **42**, 481 (1979); T. C. Jones, J. Math. Phys. (in press); C. Hoenselaers, (preprint); B. C. Xanthopoulos, (preprint).
[9]R. Geroch, J. Math. Phys. **13**, 394 (1972). The basic idea is in the conclusion of the earlier paper: J. Math. Phys. **12**, 918 (1971).
[10]C. M. Cosgrove, Proceedings of the Einstein Centenary Summer School on Gravitational Radiation and Collapsed Objects, Perth, Western Australia, 22–31 January 1979.
[11]C. M. Cosgrove, Proceedings of the The Second Marcel Grossmann Meeting on the Recent Developments of General Relativity, Trieste, Italy, 5–11 July 1979.
[12]C. M. Cosgrove, Ph.D. Thesis, University of Sydney (Feb. 1979) (unpublished).
[13]C. M. Cosgrove, J. Phys. A: Math. Gen. **10**, 1481 (1977).
[14]C. M. Cosgrove, J. Phys. A: Math. Gen. **10**, 2093 (1977).
[15]C. M. Cosgrove, J. Phys. A: Math. Gen. **11**, 2389 (1978).
[16]C. M. Cosgrove, J. Phys. A: Math. Gen. **11**, 2405 (1978).
[17]A seven-parameter enlargement of the CTS solution appears in Ref. 11. This is an excellent example of how a knowledge of the interrelationships between transformations discovered independently (Refs. 10–12, 18, and 19) was necessary to achieve success.
[18]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978).
[19]G. Neugebauer, J. Phys. A: Math. Gen. (letters) **12**, L 67 (1979). Also, recent preprints by G. Neugebauer and D. Kramer give recursion relations, repeated transforms of the Weyl solutions and the generation of a non-linear superposition of Kerr–NUT particles (submitted to J. Phys. A: Math. Gen. and Phys. Lett.).
[20]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1953 (1978) [Sov. Phys. JETP **48**, 985 (1978)]; Zh. Eksp. Teor. Fiz. **77**, 3 (1979) [Sov. Phys. JETP **50**, 1 (1979)].
[21]I. Hauser and F. J. Ernst, Phys. Rev. D **20**, 362 (1979); Phys. Rev. D **20**, 1783 (1979). The derivation from a Hilbert problem is given in J. Math. Phys. **21**, 1126 (1980).
[22]The generating functions were introduced in Refs. 4 and 5. Several properties are collected here in Appendix A.
[23]D. Maison, Phys. Rev. Lett. **41**, 521 (1978); J. Math. Phys. **20**, 871 (1979).
[24]G. Neugebauer and D. Kramer, Ann. Phys. (Leipzig) **24**, 62 (1969); A. Sackfield, J. Phys. A: Math. Gen. **8**, 506 (1975). This is the S transformation of Ref. 19. We use the symbol (I^*) for S and (I) for the complex conjugate. Note $(I)^{-1} = (I)$.
[25]The dual of a field variable (e.g., metric component, potential, generating function, etc.) is defined as the (I) or (I^*) transform, perhaps with a complex factor $\pm i$ removed. The dual of a transformation, say $\mathbf{Q}$, is defined by conjugation with (I) or (I^*): $\tilde{\mathbf{Q}} = (I)\mathbf{Q}(I)$. A field variable or transformation is said to be self-dual if the dual is the same as the original, or differs in only a trivial manner.
[26]These centers are preferred points on the z axis which act as poles of preferred spherical polar radial and angular coordinates. [See Eqs. (A16a,b) and Ref. 42 below.]
[27]More precisely, the infinitesimal operators $\gamma^{(k)}_{AB}$ are to be contracted with an arbitrary null symmetric tensor q^{AB}, then multiplied by αs^k and summed over k. The original HKX transformation is the case where $q^{11} = q_{22} = 1$ is the only non-zero component. In Sec. 2 [paragraph containing Eqs. (2.22) and (2.23)] and Footnote 43, we introduce a slightly different meaning for the γ_{AB} symbols in order to distinguish operators from parameters.
[28]D. Zipoy, J. Math. Phys. **7**, 1137 (1966); B. Voorhees, Phys. Rev. D **2**, 2119 (1970). These solutions are the static limits of the CTS solutions. Special cases are Schwarzschild ($\delta = 1$) and Curzon ($\delta = \infty$).
[29]D. M. Chitre has privately communicated that he has a method of exponentiating the $\beta^{(k)}$ using the methods of Ref. 5, but I have no details.

Hauser and Ernst (Ref. 21) have reduced the exponentiation of the $\beta^{(k)}$ or any other element of the Geroch group to the solution of a linear integral equation.

[30]In Secs. 7 and 8, we show that the Kinnersley–Chitre solutions of Ref. 5 can be constructed most easily using Harrison or Belinsky–Zakharov transformations. They can also be constructed with HKX transformations.

[31]F. J. Ernst, Phys. Rev. **167**, 1175 (1968). In Ref. 2, Kinnersley defines ψ by Eq. (1.2c) with the sign of the right-hand side changed. Our definition follows the conventions of Refs. 10–16 and 40. Consequently, Kinnersley's $\mathscr{E}$ corresponds to our $\mathscr{E}^*$ and we write $H_{11} = \mathscr{E}^*$.

[32]Some operational rules are given in Appendix A, but the reader is also advised to inspect Ref. 2.

[33]SO(2,1) rather than SL(2) since β and $-\beta$ give the same transformation.

[34]The use of the symbol **L** in both these situations should cause no confusion in the sequel.

[35]This is in acknowledgment of the considerable surge of interest in the global effects of the NUT parameter in Taub–NUT space which began with the two papers: E. T. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963); C. W. Misner, J. Math. Phys. **4**, 924 (1963). The solution itself appeared previously in the references: A. H. Taub, Ann. Math. **53**, 472 (1951); A. Papapetrou, Ann. Phys. (Leipzig) **12**, 309 (1953); J. Ehlers, "Les théories relativistes de la gravitation," Paris, CNRS (1959).

[36]Our one-parameter groups, **A**, **B**, **U**, **L**, and **P**, correspond to Kinnersley and Chitre's $\gamma^{(0)}_{22}$, $\gamma^{(-1)}_{11}$, $\gamma^{(0)}_{12}$, $\gamma^{(0)}_{11}$, and $\gamma^{(1)}_{22}$, respectively.

[37]Our sign conventions require us to identify the Neugebauer–Kramer mappings in Refs. 2 and 19 with (I^*).

[38]Equation (1.16) is to be interpreted as $H'_{AB}(\rho',z') = \Lambda H_{AB}(\rho,z)$, where $\rho' = \Lambda\rho$, $z' = \Lambda z$. If primes are dropped on the coordinates, this would read: $H'_{AB}(\rho,z) = \Lambda H_{AB}(\Lambda^{-1}\rho,\Lambda^{-1}z)$.

[39]The defining equations for γ are written out in detail in Ref. 15. See also Eqs. (4.27) and (5.41).

[40]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972); Prog. Theor. Phys. **50**, 95 (1973).

[41]The advantage of this notation is evident in expressions such as $D_2G_{AB}(t_1,0)$, etc. Similarly, the symbol D will be used for ordinary differentiation of a function of one variable, as in $H_{AB} = DF_{AB}(0)$.

[42]$S(t)$ and $\zeta(t)$, defined by Eqs. (A16a,b), are spherical polar radial and angular coordinates, respectively, with the pole at $(\rho,z) = (0,1/2t)$. Under $(Q)_{4s}$ and $(\tilde{Q})_{4s}$, they transform according to: $S'(t) = S(s+t)/S(s)$, $\zeta'(t) = \zeta(s+t)/\zeta(s)$.

[43]The individual component γ_{XY} defined by (2.22) is called γ^{XY} by Kinnersley and Chitre, since γ^{XY} is its nonvanishing coefficient, e.g., their γ_{22} is our γ_{11}.

[44]Here, we have identified the symbol f in Ref. 19 with $\mathscr{E}$. It could equally well be identified with $\mathscr{E}^*$.

[45]The different sign conventions arise from the reflections, $\omega \to -\omega$, $\mathscr{E} \to \mathscr{E}^*$, etc. as in Ref. 44. Harrison's work contains no sign ambiguities, so the identification (4.5) should correspond precisely to our latter sign convention. The former convention holds if we change the sign of the g_{12} term in Harrison's metric.

[46]Here, $S(s) = [(1-2sZ)^2 + 4s^2V^2]^{1/2}$ and $\zeta(s)$ is given by (4.7).

[47]The position of the indices in (5.1d,e) and (5.10) is chosen so that matrix equations become tensor equations.

[48]In Ref. 11, we claimed that the Harrison transformations commuted and were self-dual. These statements were based, respectively, on the symmetric appearance of (ζ_1,q_1) and (ζ_2,q_2) in Eqs. (4.8) of that reference and (5.38) and (5.39) here, and on the interchangeability of I_1 and I_2 in Neugebauer's commutation theorem. They are valid, however, only up to a change of integration constant in the pseudopotentials and a change of gauge. In this paper, however, the integration constants and gauge are specified unambiguously in the formulas involving generating functions.

[49]Private communications. The results have not been published. They did not carry the analysis far enough to reach our Eq. (6.37), but they certainly had all the mathematical apparatus in place and knew how it could be done. The same comment applies to Eqs. (7.16) and (7.17) for the $s = \infty$ limit of the nonnull HKX transformation.

[50]In the case of the generalized HKX transformation given by (6.37), a pure imaginary constant tensor $-i\alpha s^{-1}q_{AB}$ would be added to the right-hand side of (6.52).

[51]The principal results appeared first in the fourth paper of Ref. 8.

[52]The double Kerr–NUT solution of Neugebauer and Kramer is contained in the seven-parameter enlargement of the CTS solutions derived in Ref. 11. In fact, the special case $\delta = 1$ of the latter is the former.

[53]We write $(\det F)^{-1}$ rather than $-S(s)$ because Hauser and Ernst allow gauge transformations of the form (1.14) more general than those which obey (1.15). In particular, the $\tau^{(k)}$ transformations of Ref. 4 are included. This necessitates minor changes to many of the formulas in Appendix A. For example, (A18) would take the form, $F_{XA}F^X{}_B = -h(s)S^{-1}(s)\epsilon_{AB}$, where $h(s)$ is a complex-valued function of s only.

[54]α is real for cylindrical wave fields. To apply Eqs. (7.41)–(7.43) to stationary axisymmetric fields where $\alpha = i\rho$, we must analytically continue f_{AB}, $F_{AB}(s)$, etc. to complex values of (ρ,z) and then restrict ρ to be pure imaginary. Afterwards, this process can be reversed.

[55]In (8.3), where I_1 occurs last, the generic case involves an even number of Harrison transformations. A different explanation of why this product transformation preserves asymptotic flatness is given in Appendix B.

[56]F. J. Ernst, J. Math. Phys. **18**, 233 (1977).

[57]B. K. Harrison, Phys. Rev. D **21**, 1695 (1980).

On the singular eigenfunctions for linear transport in an exponential atmosphere [a)]

Edward W. Larsen
Theoretical Division, University of California, Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545

G. C. Pomraning and V. C. Badham
School of Engineering and Applied Science, University of California, Los Angeles, California 90024

(Received 11 December 1979; accepted for publication 9 May 1980)

We prove that under an explicit condition on the parameters in the isotropic-scattering linear transport equation for an exponential atmosphere, the continuum eigensolutions developed by Millikin and Siewert are complete on the half range $0<\mu\leqslant 1$. We also treat numerically the equation for the outgoing flux, which can be derived using these eigenfunctions, and we show that excellent numerical results are obtained if the above condition is satisfied, while poor results are obtained if the condition is sufficiently violated. Finally, we describe a method for constructing elementary solutions of the anisotropic-scattering transport equation for an exponential atmosphere.

I. INTRODUCTION

In this article we consider the construction of the solution of the following half-space linear transport problem, with isotropic scattering and an exponentially varying scattering ratio:

$$\mu\frac{\partial}{\partial z}\psi(z,\mu)+\psi(z,\mu)=\tfrac{1}{2}ce^{-z/s}\int_{-1}^{1}\psi(z,\mu')d\mu',$$
$$0<z<\infty,\quad -1\leqslant\mu\leqslant 1,\quad 0\leqslant c,\quad 0\leqslant s, \tag{1.1}$$

$$\psi(0,\mu)=f(\mu),\quad 0<\mu\leqslant 1, \tag{1.2}$$

$$\lim_{z\to\infty}\psi(z,\mu)=0. \tag{1.3}$$

Here $\psi(z,\mu)$ is the to-be-determined flux of particles at the point z, traveling in the direction $\theta=\cos^{-1}\mu$, and $f(\mu)$ is the prescribed incident flux on the half-space.

The above problem has been considered by Martin,[1] Siewert,[2] and Mullikin and Siewert.[3] However, these authors do not aim to construct the entire flux $\psi(z,\mu)$ for all z and μ, but rather to derive and analyze a singular integral equation for the unknown outgoing flux, $\psi(0,\mu)$ for $-1\leqslant\mu<0$, in terms of the known incident flux. In Ref. 1, Martin derives this equation by directly manipulating Eq. (1.1). (Martin's analysis is generalized in Ref. 3 for a more complicated problem.) In Refs. 2 and 3, Siewert and Mullikin derive the equation in a different way by: (i) constructing a family of continuum elementary solutions, $\psi_\nu(z,\mu)$ for $0\leqslant\nu\leqslant 1$, of Eq. (1.1), (ii) assuming that the solution $\psi(z,\mu)$ can be expanded as

$$\psi(z,\mu)=\int_0^1 a(\nu)\psi_\nu(z,\mu)d\nu, \tag{1.4}$$

and (iii) setting $z=0$ in Eq. (1.4) and using orthogonality properties of the elementary solutions to eliminate the unknown expansion coefficients $a(\nu)$.

If $c\leqslant 1$, the half-space $0<z<\infty$ is subcritical and a unique solution of problem (1.1)–(1.3) exists.[1] For this case, the outgoing and incident fluxes must satisfy the singular integral equation derived in Refs. 1 and 3, and thus this equation is guaranteed to have at least one solution. It is not known in general whether this equation has more than one solution, and correspondingly, whether there exist extra (undetermined) conditions which the outgoing and incident fluxes must satisfy. However, in Ref. 1, it is shown that if the condition

$$\frac{c}{2}\left[\left(\frac{s}{s+1}\right)\ln(1+2s)+\frac{\pi s}{s+1}+\left(\frac{s}{s+1}\right)^{1/2}\right]<1 \tag{1.5}$$

is satisfied, then the singular integral equation for the outgoing flux has a unique solution. In Ref. 3, results of numerical solutions of this equation are presented. In particular, the albedo is given for an isotropic incident flux, $c=1$, and $s=0.5$, 1, 1.5, and 2. The albedos agree to three significant figures with those computed from direct numerical solutions of the transport problem (1.1)–(1.3). However, Millikin and Siewert also reported unsatisfactory results in attempting to compute finite slab geometry albedos.[3]

The overall purpose of this article is to present new analytical and numerical results concerning the elementary solutions developed by Mullikin and Siewert. First, we prove that if the condition [which is slightly sharper than Eq. (1.5)]

$$\frac{c}{2}\left\{\left(\frac{s}{s+1}\right)\ln(1+2s)+\min\left[\frac{\pi s}{s+1}+\left(\frac{s}{s+1}\right)^{1/2},\pi\right]\right\}<1 \tag{1.6}$$

is met, then the singular integral equation for the expansion coefficient $a(\nu)$ [cf. Eq. (1.4)] possesses a unique solution. This implies that the elementary solutions ψ_ν are complete

[a)]Work by the first author (E.W.L.) was performed under the auspices of the U.S. Department of Energy, and work by the second and third authors (G.C.P. and V.C.B.) was partly supported by the National Science Foundation under grant ENG-77-24992.

 0022-2488/80/092448-07$1.00

on the half-range $0<\mu\leqslant 1$, and that the expansion (1.4) is legitimate. Second, we present the results of numerical experiments which indicate the following: if the condition (1.6) is satisfied, then numerical solutions of the singular integral equation for the half-space outgoing flux agree very well with independent calculations based on standard numerical solutions of the transport equation. However, if the condition (1.6) is sufficiently violated, then numerical solutions of this singular integral equation break down and become unphysical. (Our results agree with those of Mullikin and Siewert in the sense that for some values of c and s, numerical solutions of the singular integral equation for the outgoing flux are shown to be very accurate. However, our calculations also show that for other values of c and s this is not true. We do not know whether this occurs for the same reason as that which caused Mullikin and Siewert to obtain unsatisfactory results for finite slab problems.) The third basic purpose of this article is to show how to construct elementary solutions for anisotropic scattering problems.

The singular integral equation for $a(\nu)$, which we analyze in this article, has a similar structure to the singular integral equation for the outgoing flux derived in Refs. 1 and 3. [This is the reason for the similarity between the conditions (1.5) and (1.6).] However, these singular integral equations appear not to be equivalent; the latter is derived from the former in Ref. 3, but we have not been able to reverse the procedure. Moreover, in Ref. 1 the inequality (1.5) is derived in the context of the Hilbert space $\mathscr{L}_2[0,1]$, whereas our derivation of Eq. (1.6) is in the context of $X_2[0,1]$, the (larger) space of functions $g(\mu)$ such that $\mu g(\mu)\in\mathscr{L}_2[0,1]$. The various inequalities used in Ref. 1 to derive the condition (1.5) are, with appropriate modifications, applicable here, and we have made full use of them in our derivation of the condition (1.6).

The present note is an extension of Ref. 3, but is also a continuation of other recent work on the construction of elementary solutions of transport equations in continuously variable media.[4] In all of the media considered in Ref. 4, continuum and discrete solutions are extant, and for one class of media half-range completeness of the elementary solutions is proved, under a certain condition, using an approach similar to the method used here. The unusual feature of the present analysis is that under condition (1.6), there actually is no discrete solution. This is undoubtedly attributable to the fact that in Ref. 4, scattering ratios $c(z)$ are considered which satisfy

$$\lim_{z\to\infty} c(z) = k > 0, \tag{1.7}$$

whereas

$$\lim_{z\to\infty} c(z) = \lim_{z\to\infty} ce^{-z/s} = 0. \tag{1.8}$$

Physically, the half-range completeness of the continuum eigenfunctions under condition (1.6) must be due to the fact that the medium becomes essentially a pure absorber so rapidly, with increasing distance from the boundary, that a discrete elementary solution with lesser decay rate than the continuum solutions is not established.

To summarize, Refs. 1–3 and the present article [a] have presented the following information regarding the singular eigensolutions of Eq. (1):

(i) They are complete on the half-range for c and s satisfying Eq. (1.6). [a)]

(ii) Their definition can be extended to apply to anisotropic scattering transport equations. [a)]

(iii) They can be used to derive a singular integral equation for the outgoing flux from a half-space.[2,3] This equation has a unique solution for c and s satisfying Eq. (1.5).[1] For a given c, the F_N method gives numerical solutions of this equation for s sufficiently small,[2,3] but not for s sufficiently large. [a)]

(iv) They can be used to derive a singular integral equation for the outgoing fluxes from a finite slab, but the F_N method does not give satisfactory numerical solutions of this equation.[3]

An outline of this paper follows. In Sec. II we establish notation by deriving the elementary solutions of Ref. 3 and writing the singular integral equation for $a(\nu)$ in a suitable form. In Sec. III we derive an estimate which is used to obtain the condition (1.6). Then, in Sec. IV, we present and give a detailed discussion of our numerical results. Finally, in Sec. V, we outline a method for the construction of elementary continuum solutions of anistropic-scattering problems.

II. FORMULATION OF THE INTEGRAL EQUATION

Following Mullikin and Siewert,[3] we seek elementary solutions of Eq. (1.1) of the form

$$\psi_\nu(z,\mu) = f_\nu(\mu)e^{-z/\nu} + g_\omega(\mu)e^{-z/\omega}, \tag{2.1}$$

where ω is defined in terms of ν by

$$\frac{1}{\omega} = \frac{1}{\nu} + \frac{1}{s}. \tag{2.2}$$

Introducing this ansatz into Eq. (1.1) and equating the coefficients of $\exp[-1/\nu - (n/s)]z$ for $n = 0$, 1, and 2 gives

$$[1-(\mu/\nu)]\,f_\nu(\mu) = 0, \tag{2.3}$$

$$[1-(\mu/\omega)]g_\omega(\mu) = \frac{c}{2}\int_{-1}^{1} f_\nu(\mu')\,d\mu', \tag{2.4}$$

$$0 = \int_{-1}^{1} g_\omega(\mu')\,d\mu'. \tag{2.5}$$

These equations have the solutions

$$f_\nu(\mu) = \delta(\nu-\mu), \tag{2.6}$$

$$g_\omega(\mu) = \frac{c\omega}{2}\left[\frac{1}{\omega-\mu} - \delta(\omega-\mu)\ln\left(\frac{1+\omega}{1-\omega}\right)\right], \tag{2.7}$$

provided ν and ω both lie in the interval $[-1,1]$; for any $s>0$ this is guaranteed for

$$0\leqslant\nu\leqslant 1. \tag{2.8}$$

The general solution of Eq. (1.1) which is constructable from these elementary solutions and which satisfies Eq. (1.3) is

$$\psi(z,\mu) = \int_0^1 a(\nu)\psi_\nu(z,\mu)\,d\nu. \tag{2.9}$$

It remains to satisfy Eq. (1.2). Using the above results, this equation is

$$f(\mu) = \int_0^1 a(\nu)\left\{\delta(\nu-\mu) + \frac{c\omega}{2} \times\left[\frac{1}{\omega-\mu} - \delta(\omega-\mu)\ln\left(\frac{1+\omega}{1-\omega}\right)\right]\right\} d\nu, \quad (2.10)$$

where ω is defined in terms of ν by Eq. (2.2). By making the change of variables

$$\nu = \frac{\omega s}{s-\omega}, \quad (2.11)$$

this equation can be written as

$$f(\mu) = (I + \tfrac{1}{2}cL)a(\mu), \quad 0<\mu\leqslant 1, \quad (2.12)$$

where I is the identity operator and L is defined by

$$La(\mu) = \int_0^{s/(s+1)} \frac{\omega s}{s-\omega} a\left(\frac{\omega s}{s-\omega}\right)\frac{s}{s-\omega} \times\left[\frac{1}{\omega-\mu} - \delta(\omega-\mu)\ln\left(\frac{1+\omega}{1-\omega}\right)\right] d\omega. \quad (2.13)$$

To provide a setting for proving conditions under which Eq. (2.12) has a unique solution, we define the Banach space $X = X_2[0,1]$ as

$$X = \left\{g(\mu)\,\middle|\, \|g\|^2 = \int_0^1 |\mu g(\mu)|^2\, d\mu < \infty\right\}. \quad (2.14)$$

We shall require $f\in X$, and we seek a solution $a\in X$ of Eq. (2.12). A unique solution exists, and in principle is constructable from the Neumann series, if $L : X\to X$ is bounded and

$$\tfrac{1}{2}c\|L\| < 1. \quad (2.15)$$

In the next section, we derive an explicit estimate on $\|L\|$, and then the inequality (2.15) becomes an explicit condition on c and s which guarantees a unique solution of Eq. (2.12). This condition, of course, will be the inequality (1.6). If this is satisfied, then the resulting unique solution of Eq. (2.12) is admissible in Eq. (2.9) provided $\psi(z,\mu)$ is interpreted as an element of the Banach space

$$Y = \left\{h(\mu)\,\middle|\, \|h\|^2 = \int_{-1}^1 |\mu h(\mu)|^2\, d\mu < \infty\right\} \quad (2.16)$$

for every z; this easily follows from results in Ref. 5.

III. ESTIMATE ON $\|L\|$

In this section, we let $g\in X$ be arbitrary. By Eq. (2.13), we may write

$$L = L_1 + L_2, \quad (3.1)$$

where

$$L_1 g(\mu) = -\int_0^{s/(s+1)} \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right)\frac{s}{s-\omega}\delta(\omega-\mu)\ln\left(\frac{1+\omega}{1-\omega}\right) d\omega = \begin{cases} -\dfrac{\mu s}{s-\mu} g\left(\dfrac{\mu s}{s-\omega}\right)\dfrac{s}{s-\mu}\ln\left(\dfrac{1+\mu}{1-\mu}\right), & 0\leqslant\mu<\dfrac{s}{s+1} \\ 0, & \dfrac{s}{s+1}\leqslant\mu\leqslant 1, \end{cases} \quad (3.2)$$

and

$$L_2 g(\mu) = \int_0^{s/(s+1)} \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right)\frac{s}{s-\omega}\frac{1}{\omega-\mu} d\omega. \quad (3.3)$$

Following Martin's approach in Ref. 1, we shall derive estimates on $\|L_1 g\|$ and $\|L_2 g\|$. First,

$$\|L_1 g\|^2 = \int_0^{s/(s+1)} \left|\mu\frac{\mu s}{s-\mu} g\left(\frac{\mu s}{s-\mu}\right)\frac{s}{s-\mu}\ln\left(\frac{1+\mu}{1-\mu}\right)\right|^2 d\mu \leqslant \left[\left(\frac{s}{s+1}\right)\ln\left(\frac{1+s/(s+1)}{1-s/(s+1)}\right)\right]^2 \times\int_0^{s/(s+1)} \left|\frac{\mu s}{s-\mu} g\left(\frac{\mu s}{s-\mu}\right)\right|^2\left(\frac{s}{s-\mu}\right)^2 d\mu. \quad (3.4)$$

Introducing the change of variables

$$t = \frac{\mu s}{s-\mu}, \quad dt = \left(\frac{s}{s-\mu}\right)^2 d\mu, \quad (3.5)$$

immediately gives the bound

$$\|L_1 g\| \leqslant \left[\left(\frac{s}{s+1}\right)\ln(1+2s)\right]\|g\|. \quad (3.6)$$

To derive a useful estimate for $\|L_2 g\|$, we need the following result[6]; if $u(x)\in\mathscr{L}_2(-\infty,+\infty)$, then the function $v(x)$, defined by

$$v(x) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{u(x')}{x'-x} dx', \quad (3.7)$$

is also in $\mathscr{L}_2(-\infty,+\infty)$, and

$$\int_{-\infty}^{+\infty} |v(x)|^2\, dx = \int_{-\infty}^{+\infty} |u(x)|^2\, dx. \quad (3.8)$$

Using this result, we can obtain two bounds on $L_2 g$. First, if we interpret Eq. (3.3) as defining $L_2 g(\mu)$ for all $-\infty<\mu<+\infty$, then

$$\|L_2 g\|^2 = \int_0^1 |\mu L_2 g(\mu)|^2\, d\mu \leqslant \int_0^1 |L_2 g(\mu)|^2\, d\mu \leqslant \int_{-\infty}^{+\infty} |L_2 g(\mu)|^2\, d\mu$$

$$= \pi^2 \int_0^{s(s+1)} \left| \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right) \right|^2 \left(\frac{s}{s-\omega}\right)^2 d\omega$$
$$= \pi^2 \|g\|^2. \tag{3.9}$$

Next, we may write

$$\mu L_2 g(\mu) = g_1 + g_2(\mu), \tag{3.10}$$

where

$$g_1 = \int_0^{s/(s+1)} \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right) \frac{s}{s-\omega} d\omega, \tag{3.11}$$

$$g_2(\mu) = -\int_0^{s/(s+1)} \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right) \frac{s}{s-\omega} \frac{\omega}{\omega-\mu} d\omega. \tag{3.12}$$

Using the Cauchy–Schwartz inequality, we get

$$|g_1|^2 \leq \int_0^{s/(s+1)} 1^2 \, d\omega \int_0^{s/(s+1)} \left| \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right) \right|^2 \left(\frac{s}{s-\omega}\right)^2 d\omega$$
$$= \frac{s}{s+1} \|g\|^2, \tag{3.13}$$

and using Eqs. (3.7) and (3.8), we get

$$\int_0^1 |g_2(\mu)|^2 \, d\mu$$
$$\leq \int_{-\infty}^{+\infty} |g_2(\mu)|^2 \, d\mu$$
$$= \pi^2 \int_0^{s/(s+1)} \left| \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right) \right|^2 \left(\frac{s}{s-\omega}\right)^2 \omega^2 \, d\omega$$
$$\leq \left(\frac{\pi s}{s+1}\right)^2 \int_0^{s/(s+1)} \left| \frac{\omega s}{s-\omega} g\left(\frac{\omega s}{s-\omega}\right) \right|^2 \left(\frac{s}{s-\omega}\right)^2 d\omega$$
$$= \left(\frac{\pi s}{s+1}\right)^2 \|g\|^2. \tag{3.14}$$

Eqs. (3.10), (3.13), (3.14), and the triangle inequality combine to give

$$\|L_2 g\| = |g_1| + \left\{ \int_0^1 |g_2(\mu)|^2 \, d\mu \right\}^{1/2}$$
$$\leq \left[\left(\frac{s}{s+1}\right)^{1/2} + \frac{\pi s}{s+1} \right] \|g\|. \tag{3.15}$$

Hence, we obtain from the inequalities (3.9) and (3.15)

$$\|L_2 g\| \leq \min\left[\frac{\pi s}{s+1} + \left(\frac{s}{s+1}\right)^{1/2}, \pi \right] \|g\|. \tag{3.16}$$

Finally, by Eqs. (3.1), (3.6), and (3.16), we get

$$\|L\| \leq \|L_1\| + \|L_2\| \leq \left\{ \left(\frac{s}{s+1}\right) \ln(1+2s) + \min\left[\frac{\pi s}{s+1} + \left(\frac{s}{s+1}\right)^{1/2}, \pi \right] \right\}. \tag{3.17}$$

Introducing this bound into the inequality (2.15), we obtain the desired result, Eq. (1.6).

IV. NUMERICAL RESULTS

In this section we shall investigate numerically, and in more detail than in Ref. 3, the singular integral equation for the outgoing flux. That is, we follow Ref. 3 and: (i) derive from Eq. (1.4) this singular integral equation, (ii) apply the F_N method[3] to obtain a numerical solution, (iii) compute the albedo, and (iv) compare this albedo with the albedo generated by a standard, independent numerical method.

Our purpose is partly to show that for a given value of c, and s sufficiently small, the F_N method gives excellent numerical results. In this sense, we verify the half-space results reported by Mullikin and Siewert.[2,3] However, we also show that for s sufficiently large, the F_N results break down and become unphysical. We do not know whether the cause for this is the same as that which led to the complications reported by Mullikin and Siewert in their attempt to solve finite slab problems. However we shall, later in this section, discuss these numerical difficulties for large s.

The singular integral equation for the outgoing flux is obtained from Eq. (1.4) by using the full-range orthogonality of the Mullikin-Siewert (MS) eigenfunctions, and is given in Ref. 3 as

$$\int_0^1 \mu \psi(0,-\mu) \psi_\nu \, d\mu = \int_0^1 \mu f(\mu) \psi_\nu(0,-\mu) \, d\mu, \quad 0 < \nu \leq 1. \tag{4.1}$$

For a given incident flux $f(\mu)$, the right side of Eq. (4.1) can be evaluated, and the result is a singular integral equation for the outgoing flux $\psi(0,-\mu)$. This is a nonstandard integral equation because the integrand is singular at two points, namely $\mu = \nu$ and $\mu = \omega$. We have numerically solved Eq. (4.1) by the F_N method,[3] which consists of representing the outgoing flux by an Nth order polynomial

$$\psi(0,-\mu) \simeq \sum_{n=0}^{N} a_n \mu^n, \quad 0 < \mu \leq 1. \tag{4.2}$$

With this representation, the integration on the left side of Eq. (4.2) can be explicitly performed. Subsequent evaluation of Eq. (4.2) at $N+1$ discrete values of ν in the range $0 < \nu < 1$ leads to an $(N+1)\times(N+1)$ matrix equation for the expansion coefficients a_n. We have solved this matrix equation numerically by using a standard linear equation solver. Once the coefficients a_n are found, the albedo (reflection probability) is computed from

$$R = \int_0^1 \mu \psi(0,-\mu) \, d\mu \Big/ \int_0^1 \mu f(\mu) \, d\mu$$
$$\simeq \sum_{n=0}^{N} \frac{a_n}{n+2} \Big/ \int_0^1 \mu f(\mu) \, d\mu. \tag{4.3}$$

In our work, the $N+1$ discrete values of ν were chosen in various ways, including the positive roots of the $2(N+1)$-order Legendre polynomials used by Mullikin and Siewert,[3] and the simple equal spacing

$$\nu_m = (2m+1)/(2N+2), \quad m = 0,1,\ldots,N. \tag{4.4}$$

Our numerical results were insensitive to this choice, and the results reported here utilized Eq. (4.4). Also, we repeated our calculations using Legendre polynomials in μ rather than the power series (4.2), but again our results were unchanged.

We summarize in Fig. 1 our numerically computed albedo for various values of s, with $c = 0.99$ and $f(\mu) = 1$. Figure 1 requires a good deal of discussion. The solid curve represents the "exact" albedos calculated using the ANISN computer code.[7] This corresponds to a direct numerical so-

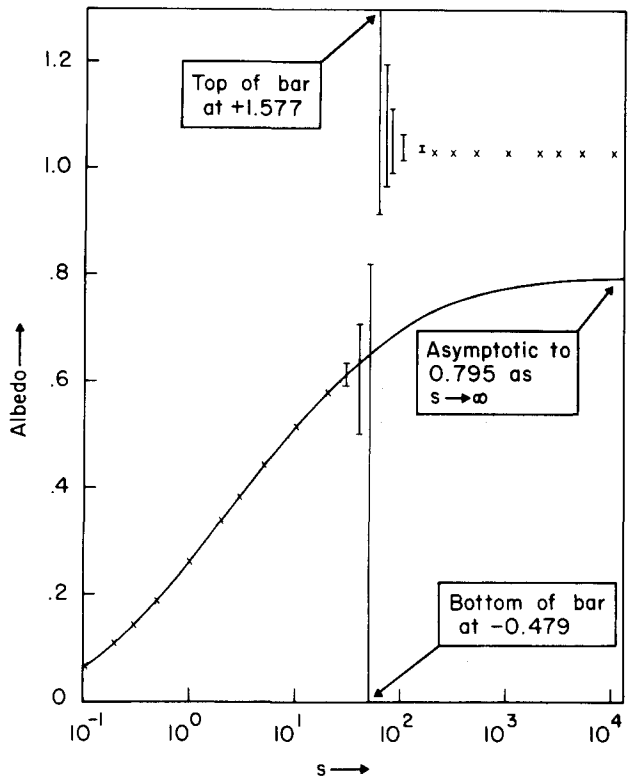


FIG. 1. The albedo, versus s, for an isotropic incident flux and $c = 0.99$.

lution of the transport equation by the standard, well-tested, method of discrete ordinates. The size of the spatial mesh ($\Delta z \simeq 0.1$) and the order of the angular quadrature ($M \simeq 32$) were chosen to insure results for the albedo accurate to at least three significant figures. The solid curve in Fig. 1 thus represents the exact albedo as a function of s, against which we compare the Mullikin-Siewert (MS) value.

We first focus our attention upon the crosses (somewhat obscured by the solid curve) in Fig. 1 at $s = 0.1, 0.2, 0.3$ and 0.5 which represent converged MS albedos. For $c = 0.99$, the value of s for which Eq. (1.6) becomes an equality is $s \simeq 0.570$. Thus, for these s, the MS eigenfunctions are complete *and* the MS albedos agree very well with the discrete ordinates albedos. In addition, we have compared the outgoing angular flux generated by ANISN to the MS outgoing flux, given by Eq. (4.2), and found that these more detailed quantities also agreed very well, for the above values of s.

In particular, for $s = 0.5$ the expression given by Eq. (4.3) was well-converged (to at least three significant figures) for $N = 7$. Larger values of s correspond to a more highly scattering half space, and one would accordingly expect that the outgoing angular flux would be a smoother function of μ. This expectation is borne out by our ANISN calculations. Hence, in the absence of numerical instabilities or eigenfunction incompleteness, one ought to find a converged (to at least three significant figures) result for $N = 7$ and $s > 0.5$. We found this to be the case for s as large as 20, and these converged values, which agree with ANISN calaculations, are also indicated by crosses on Fig. 1. [This suggests that the inequality (1.6) is quite conservative.] For values of s between 20 and 150, however, convergence to three significant figures was not observed at $N = 7$. For these values of s, the tips of the vertical lines on Fig. 1 show the $N = 7$ and $N = 8$ results. For $s > 150$, the MS treatment again apparently converged to three significant figures at $N = 7$, but gave $R = 1.03$ as the albedo (which we again indicated by crosses on Fig. 1). This is obviously an incorrect result, since the albedo is bounded from above by unity. In fact, for any value of s, it must be bounded from above by the $s = \infty$ result, which is $R = 0.795$ for $c = 0.99$.

Obviously, the plotting of the $N = 7$ and $N = 8$ results in Fig. 1 is somewhat arbitrary, but this figure does give a visual overview of the difficulties of the MS treatment for large values of s. Table I gives more detail, for representative

TABLE I. The albedo for isotropic incidence with $c = 0.99$.

N \ s	0.1	0.5	5	20	30	50	100	1000
2	0.0703	0.1942	0.4454	0.6168	0.6627	0.7182	0.7913	0.9627
3	0.0674	0.1884	0.4423	0.5651	0.5772	0.5355	−0.7505	1.0681
4	0.0691	0.1897	0.4416	0.5898	0.6468	0.7486	0.9014	1.0301
5	0.0678	0.1892	0.4417	0.5737	0.5748	0.2633	1.1678	1.0354
6	0.0685	0.1895	0.4418	0.5830	0.6425	0.8000	0.9871	1.0332
7	0.0679	0.1894	0.4417	0.5798	0.5923	−0.4788	1.0614	1.0328
8	0.0682	0.1895	0.4418	0.5797	0.6261	0.8221	1.0165	1.0322
9	0.0679	0.1894	0.4417	0.5808	0.6096	−1.0235	1.0407	1.0318
10	0.0681	0.1895	0.4418	0.5799	0.6148	0.8154	1.0258	1.0314
11	0.0680	↓	↓	0.5803	0.6160	−0.0002	1.0348	1.0311
12	↓	↓	↓	0.5803	0.6127	0.7746	1.0289	1.0309
13	↓	↓	↓	0.5802	0.6156	0.4872	1.0329	1.0307
14	↓	↓	↓	0.5803	0.6139	0.7027	1.0301	1.0306
15	↓	↓	↓	0.5802	0.6145	0.6384	1.0321	1.0305
16	↓	↓	↓	↓	0.6145	0.6397	1.0305	1.0303
17	↓	↓	↓	↓	0.6143	0.6747	1.0317	1.0302
18	↓	↓	↓	↓	0.6145	0.6284	1.0307	1.0302

values of s, of the MS treatment as N, the order of the polynomial representation, is increased up to $N = 18$. There appear to be two likely explanations for the above numerical breakdown at large s. The first possibility is a numerical instability which becomes manifest for large s; the second possibility is that the MS eigenfunctions become incomplete for large s and thereby render Eqs. (1.4) and (4.1) invalid. We shall very briefly discuss these here.

First let us discuss the possibility of a numerical instability, and write Eq. (4.1) as

$$M(s)\psi(0,-\mu) = P(s)f(\mu). \tag{4.5}$$

Since the numerical breakdown occurs for large s, it may be instructive to examine Eq. (4.5) for $s = \infty$. We have

$$M(\infty)\psi(0,-\mu) = P(\infty)f(\mu). \tag{4.6}$$

This is an equation corresponding to a homogeneous medium, for which an explicit analysis is possible. In fact, the following results easily follow from Ref. 5 and Ref. 8, p. 86, Eq. (346):

(i) For any $f \in L_2(0,1)$, there exists a solution $\psi(0,-\mu) \in L_2(0,1)$.

(ii) $M(\infty)\left[\dfrac{\chi^+(\mu)}{\Lambda^+(\mu)}\right] = 0.$

(The functions χ and Λ are defined in Ref. 8, pp. 79,80.) If we now think of $M(s)$ as approximately $M(\infty)$ for large s, and the F_N (or any other numerical) method approximating $M(s)$ by a finite-dimensional operation (a matrix), we see that no finite-dimensional operator can duplicate the property of simultaneously having its range be the entire (finite-dimensional) space *and* having a null space. This inconsistency for $s = \infty$ could be a source of numerical difficulties for $s \gg 1$.

The other possibility, that of missing elementary solutions in Eq. (1.4), seems very unlikely because of some new results communicated to us privately by T. W. Mullikin.[9] Mullikin has shown that the Fourier transform of the scalar flux $\hat{\phi}(\lambda)$, is analytic for all finite points λ in the complex plane off the cut $[-i\infty, -i]$. The contributions to the transform can only come from the cut and from $\lambda = \infty$. In terms of singular eigenfunction representation ($\nu = 1/iz$), this means that all contributions must come from $0 \leqslant \nu \leqslant 1$. (The point at ∞ maps into 0.) Thus, if the representation (1.4) is invalid, it is by virtue of missing elementary solutions corresponding to ν *on* the cut $0 \leqslant \nu \leqslant 1$, not off it. In our view this seems implausible.

However, the notion of a missing discrete solution makes considerable sense if one thinks physically. As $s \to \infty$, it is easy to show that the MS eigenfunctions reduce to the standard continuum eigenfunction.[8] Moreover, in this limit, one can easily show that the transport equation solution converges pointwise to the solution of the $s = \infty$ (homogeneous medium) problem, which *does* have a discrete solution. Thus, one might expect that as s increases beyond the bound for half-range completeness given by the inequality (1.6), there is a value, say s^*, above which a new elementary solution of the transport equation, corresponding to a point out of the continuum $0 \leqslant \nu \leqslant 1$ and independent of the continuum eigensolutions, becomes manifest. However, Mullikin's result indicates that this is not the case.

Thus, the precise analytic behavior of the singular integral equation for the outgoing flux for large s, and the cause of the numerical difficulties, remain open questions.

V. ANISOTROPIC SCATTERING

Here we shall very briefly describe a method for constructing elementary continuum solutions for "exponential" atmospheres with anisotropic scattering. For simplicity, we shall only explicitly treat the case of linearly anisotropic scattering; the extension to higher-order anisotropic scattering problems is straightforward.

The transport equation with linearly anisotropic scattering is[8]

$$\mu\frac{\partial}{\partial z}\psi(z,\mu) + \psi(z,\mu) = \frac{c}{2}e^{-z/s}\int_{-1}^{1}(1 + a\mu\mu')\psi(z,\mu')\,d\mu', \tag{5.1}$$

and we take as the solution form for the elementary solutions

$$\psi_\nu(z,\mu) = f_0(\mu)e^{-z/\nu} + \sum_{n=1}^{2} f_n(\mu)e^{-z/\omega_n}, \tag{5.2}$$

where

$$\frac{1}{\omega_n} = \frac{1}{\nu} + \frac{n}{s}, \quad n = 1,2. \tag{5.3}$$

Combining Eqs. (5.1)–(5.3), we obtain the following equations for $f_n(\mu)$:

$$f_0(\mu)\left(1 - \frac{\mu}{\nu}\right) = 0, \tag{5.4}$$

$$f_1(\mu)\left(1 - \frac{\mu}{\omega_1}\right) = \frac{c}{2}\int_{-1}^{1}(1 + a\mu\mu')f_0(\mu')\,d\mu', \tag{5.5}$$

$$f_2(\mu)\left(1 - \frac{\mu}{\omega_2}\right) = \frac{c}{2}\int_{-1}^{1}(1 + a\mu\mu')f_1(\mu')\,d\mu', \tag{5.6}$$

$$0 = \frac{c}{2}\int_{-1}^{1}(1 + a\mu\mu')f_2(\mu')d\mu'. \tag{5.7}$$

The first three of these equations can be solved recursively, yielding the general solutions

$$f_0(\mu) = \delta(\nu - \mu), \tag{5.8}$$

$$f_1(\mu) = \alpha_1\delta(\omega_1 - \mu) + \frac{c\omega_1}{2}\frac{1 + a\mu\nu}{\omega_1 - \mu}, \tag{5.9}$$

$$f_2(\mu) = \alpha_2\delta(\omega_2 - \mu) + \alpha_1\left(\frac{c\omega_2}{2}\frac{1 + a\mu\omega_1}{\omega_2 - \mu}\right) + \left(\frac{c\omega_1}{2}\right)\left(\frac{c\omega_2}{2}\right)\frac{\beta_0 + \beta_1\mu}{\omega_2 - \mu}, \tag{5.10}$$

where

$$\beta_n = \int_{-1}^{1}(a\mu')^n\frac{1 + a\mu'\nu}{\omega_1 - \mu'}\,d\mu', \quad n = 0,1, \tag{5.11}$$

and where α_1 and α_2 are arbitrary constants. Equation (5.7) however is equivalent to the two equations

$$0 = \int_{-1}^{1}\mu^n f_2(\mu)\,d\mu, \quad n = 0,1, \tag{5.12}$$

which uniquely determine α_1 and α_2 as

$$\alpha_1 = -\frac{c\omega_1}{2}\beta_0, \tag{5.13}$$

$$\alpha_2 = ac\omega_1 \frac{c\omega_2}{2} \int_{-1}^{1} \frac{\mu}{\omega_2 - \mu} d\mu. \tag{5.14}$$

The expression (5.10) for $f_2(\mu)$ now simplifies to

$$f_2(\mu) = \alpha_2 \delta(\omega_2 - \mu) - (ac\omega_1) \frac{c\omega_2}{2} \frac{\mu}{\omega_2 - \mu}. \tag{5.15}$$

The solutions (5.2) are completely specified, and are valid provided $-1 \leqslant \nu \leqslant 1$ and $-1 < \omega_1, \omega_2 < 1$. For $s > 0$, this is guaranteed for $0 \leqslant \nu \leqslant 1$.

It is clear that the above procedure can be applied to problems with any definite order of anisotropic scattering; if the scattering kernel contains N Legendre polynomials, then the ansatz (5.2) must be expanded to sum from 1 to N. The functions $f_n(\mu)$ will always consist of a delta and a principal value function, and it is certain that, as in the isotropic-scattering case considered above in detail, the solutions (5.2) will be complete on the half range $0 < \mu \leqslant 1$ if an appropriate inequality is satisfied.

ACKNOWLEDGMENT

The authors would like to acknowledge many provocative and helpful discussions with Professors T. W. Mullikin and C. E. Siewert.

[1]B. J. Martin, SIAM J. Appl. Math., **20**, 703 (1971).
[2]C. E. Siewert, Sixth Conference on Transport Theory, Tucson, Arizona, April 2–6, 1979.
[3]T. W. Mullikin and C. E. Siewert, "Radiative Transfer in Inhomogeneous Atmospheres," Ann. Nucl. Energy (in press). This article extends the analysis in Ref. 1 and contains details of the results presented in Ref. 2.
[4]G. C. Pomraning and E. W. Larsen, J. Math. Phys. **21**, 1603 (1980).
[5]E. W. Larsen, S. Sancaktar and P. F. Zweifel, J. Math. Phys. **16**, 1117 (1975).
[6]E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford U.P., London, 1950).
[7]W. W. Engle, Jr., *The User's Manual for ANISN: A One-Dimensional Discrete Ordiantes Transport Code with Anisotropic Scattering*, K-1693, Oak Ridge Gaseous Diffusion Plant (1967). Available from Radiation Shielding Information Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, as CCC-82/ANISN (March 1967).
[8]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison–Wesley, Reading, Mass., 1967).
[9]T. W. Mullikin (private communication).

Interaction of isovector scalar mesons with simple sources. II. Isobars and meson scattering

M. Bolsterli
T Division, Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545

(Received 1 February 1980; accepted for publication 6 May 1980)

For the case of a static source, appropriate special coherent states are defined so as to give a relatively simple treatment of states of any isospin in the no-meson and one-meson approximations. The analogous localized coherent states for nonstatic sources are described and shown to be suitable for use in translated-localized-state calculations.

I. INTRODUCTION

Noncovariant models for meson-nucleon interactions have become more interesting in recent years as it has become more evident that nucleons and mesons are not elementary fields but composites of elementary fields. In particular, meson-nucleon field theories without divergences can be used as models for describing properties of systems of mesons and nucleons at excitation energies that are not too high; that is, excitations below the antinucleon threshold. A fundamental problem in these models is the state corresponding to a single nucleon. One of the approaches to this problem is through the use of static models, that is, systems in which the meson field interacts with a static source.

For neutral mesons interacting with a static source, the solution is simple.[1] Interaction of an isovector meson field with an isospinor source is not so simple. The basic theoretical tool has been the intermediate-coupling method introduced by Tomonoga[2] and utilized in other work.[3-6] The basic state vector in this method is one in which the mesons are in a single mode; that is, only the meson operator

$$A_i^\dagger = \int b(k) a_i^\dagger(k)\, dk \tag{1}$$

is used to produce the basic state from the meson vacuum.

Recently, the connection between the Tomonoga method and coherent states in general was considered,[7] and it was shown that a special coherent state has some advantages over the usual basic state vector. The usual basic state $|\rangle_T$ is such that meson hole states are possible; a proper one-meson approximation would involve both the states $a_i^\dagger(k)|\rangle_T$ and $a_i(k)|\rangle_T$; these latter have been ignored in the past. The special coherent state $|\rangle_S$ of Ref. 7 is an eigenstate of all the isoscalar one-meson annihilation operators

$$\tau\cdot a(k)|\rangle_S = b(p)|\rangle_S; \tag{2}$$

Then, within a particular isospin subspace, meson hole states need not be considered, and the one-meson approximation involves only the states $\tau\cdot a^\dagger(k)|\rangle_S$. Moreover, the techniques for going from static coherent states to nonstatic ones can be applied to this system to give approximation methods for the case in which the nucleon is allowed to recoil.

In Ref. 7 only the case of isospin $\frac{1}{2}$ was treated and meson-nucleon scattering was not considered. This paper extends the results of Ref. 7 to higher isospins (isobars) and to the calculation of meson-nucleon scattering.

II. ENERGY SPECTRUM IN THE NO-MESON APPROXIMATION

The Hamiltonian, as in Ref. 7, is

$$H = T - V - V^\dagger,$$

$$T = \int_0^\infty \omega(k) a^\dagger(k)\cdot a(k)\, dk, \tag{3}$$

$$V = \tau\cdot \int_0^\infty v^*(k) a(k)\, dk,$$

where a, $a^\dagger$, and τ are isovectors, $v^*(k)$ is a form factor, and $\omega(k)$ is the meson energy, usually $(k^2 + m^2)^{1/2}$; only s-wave mesons interact with the source, so k is scalar, with spherical factors all incorporated in $v(k)$. When it is convenient, spherical tensor components of τ and $a(k)$ can be used. The nonzero commutator of the a operators is

$$[a_i(p), a_j^\dagger(q)] = \delta_{ij}\delta(p-q), \tag{4}$$

and only the anticommutator of the τ_i will be used explicitly

$$\{\tau_i, \tau_j\} = 2\delta_{ij}. \tag{5}$$

The nucleon state $|0m)$ has isospin $T = \frac{1}{2}$ and $T_3 = m$ with $m = \pm\frac{1}{2}$; the quantum number m will usually be omitted.

As in Tomonoga's work, consider a single mode for the mesons; that is, let

$$A_i^\dagger = \int b(k) a_i^\dagger(k)\, dk. \tag{6}$$

Within this mode, define the fundamental state $|nb)$ with $T = n + \frac{1}{2}$ as

$$|nb) = (n!)^{-1/2}\{A_{i_1}^\dagger A_{i_2}^\dagger \cdots A_{i_n}^\dagger |0)\}^{n+1/2}, \tag{7}$$

where the curly bracket indicates vector coupling of the n $A^\dagger$ operators with the $T = \frac{1}{2}$ state $|0)$ to give total $T = n + \frac{1}{2}$. Since

$$\{\tau|0)\}^{3/2} = 0, \tag{8}$$

it follows that if τ_j is substituted for $A_j^\dagger$ in Eq. (6), then the result is zero, that is,

$$\tau\cdot a(k)|nb) = 0. \tag{9}$$

 0022-2488/80/092455-06$1.00

The special one-mode coherent state $|nb\rangle$ satisfies

$$\tau\cdot a(k)|nb\rangle = b(k)|nb\rangle \tag{10}$$

and it must be of the form of a general one-mode state with $T = n + \frac{1}{2}$; hence

$$|nb\rangle = \sum_{\mu=0}^{\infty} c_{n,\mu}(\tau\cdot A^{\dagger})^{\mu}|nb\rangle/\sqrt{\Gamma_n}$$

$$= \sum_{\mu=0}^{\infty} [c_{n,2\mu} + c_{n,2\mu+1}\tau\cdot A^{\dagger}](A^{\dagger}\cdot A^{\dagger})^{\mu}|nb\rangle/\sqrt{\Gamma_n}, \tag{11}$$

where Γ_n will be chosen to normalize the state $|nb\rangle$. The commutation relations (4) and (5) give

$$[\tau\cdot a(k), A^{\dagger}\cdot A^{\dagger}] = 2b(k)\tau\cdot A^{\dagger},$$

$$\{\tau\cdot a(k), \tau\cdot A^{\dagger}\}|nb\rangle = b(k)\tau\cdot\tau|nb\rangle + 2A^{\dagger}\cdot a(k)|nb\rangle \tag{12}$$

$$= (2n+3)b(k)|nb\rangle;$$

it follows that Eq. (10) will be satisfied for

$$c_{n,2\mu} = \frac{(n+\frac{1}{2})!}{2^{2\mu}\mu!(n+\frac{1}{2}+\mu)!},$$

$$c_{n,2\mu+1} = \frac{(n+\frac{1}{2})!}{2^{2\mu+1}\mu!(n+\frac{3}{2}+\mu)!}. \tag{13}$$

Let $\gamma_n(x)$ be defined by

$$\gamma_n(x) = \sum_{0}^{\infty} c_{\mu}x^{\mu}; \tag{14}$$

then the special coherent state $|nb\rangle$ is

$$|nb\rangle = \gamma_n(\tau\cdot A^{\dagger})|nb\rangle/\sqrt{\Gamma_n}. \tag{15}$$

The normalization of the state is easily determined; first it is clear from (10) that

$$\tau\cdot A|nb\rangle = B|nb\rangle,$$

$$B \equiv \int |b(k)|^2\,dk, \tag{16}$$

so that

$$\langle nb|nb\rangle = \gamma_n(B)(nb|nb\rangle/\sqrt{\Gamma_n} = \gamma_n(B)(nb|nb\rangle/\Gamma_n$$

$$= \gamma_n(B)B^n/\Gamma_n, \tag{17}$$

$$\Gamma_n(B) = B^n\gamma_n(B).$$

Thus, for every L_2 function $b(k)$ there is a one-mode coherent state (15) with $T = n + \frac{1}{2}$ that satisfies Eq. (10). Any one of these states can be used as the basis for a set of N-meson approximations, $N = 0,1,2,\ldots$, where the N-meson approximation is obtained by minimizing the expectation of the Hamiltonian within the subspace of states of the form

$$|nb\rangle, \int f(q_1\cdots q_k)\prod_{i=1}^{k}\tau\cdot a^{\dagger}(q_i)\,dq_i|nb\rangle, \tag{18}$$

with $k \leqslant N$. Again, the point of the coherent state is that no $\tau\cdot a(q)$ operators need be used in generating the states. Within the N-meson approximation, the energy of the lowest $T = n + \frac{1}{2}$ state is a functional $F_{N,n}\{b\}$ of $b(k)$, so that the best $b(k)$ in the N-meson approximation is the one that minimizes the energy functional $F_{N,n}\{b\}$.

In particular, in the no-meson approximation, $F_{0,n}\{b\}$ is just the expectation of H in the state $|nb\rangle$

$$F_{0,n}\{b\} = \langle nb|H|nb\rangle. \tag{19}$$

To evaluate this, the following matrix element is needed:

$$\langle nb|a^{\dagger}(p)\cdot a(q)|nb\rangle$$

$$= \sum_{\mu} c_{\mu}(nb|(\tau\cdot A)^{\mu}a^{\dagger}(p)\cdot a(q)|nb\rangle/\sqrt{\Gamma_n}$$

$$= \sum_{\mu} c_{\mu}(nb|\Big\{b^*(p)\sum_{j=0}^{\mu-1}(\tau\cdot A)^{\mu-j-1}\tau$$

$$\cdot a(q)(\tau\cdot A)^{j} + a^{\dagger}(p)\cdot a(q)B^{\mu}\Big\}|nb\rangle/\sqrt{\Gamma_n} \tag{20}$$

$$= b^*(p)b(q)\sum_{\mu} c_{\mu}\{\mu B^{n+\mu-1} + nB^{n+\mu-1}\}/\Gamma_n$$

$$= b^*(p)b(q)\Gamma_n'(b)/\Gamma_n(B).$$

The general matrix element of this type is

$$\langle nb|:\prod_{k=1}^{K} a^{\dagger}(p_k)\cdot a(q_k):|nb\rangle$$

$$= \Big[\prod_{k=1}^{K} b(p_k)b^*(q_k)\Big]\Gamma_n^{(K)}(B)/\Gamma_n(B). \tag{21}$$

Thus, the number of mesons in the state $|nb\rangle$ is

$$N_{0,n}\{b\} = \int \langle nb|a^{\dagger}(k)\cdot a(k)|nb\rangle\,dk = B\frac{\Gamma_n'(B)}{\Gamma_n(B)}$$

$$\equiv BG_n(B), \tag{22}$$

and it follows also that

$$F_{0,n}\{b\} = WG_n(B) - \int [v^*(k)b(k) + b^*(k)v(k)]\,dk, \tag{23}$$

$$W = \int \omega(k)|b(k)|^2\,dk = \bar{\omega}B,$$

where $\bar{\omega}$ is the average energy per meson in the field. Variation of $F_{0,n}$ gives

$$b_{0,n}(k) = \frac{v(k)}{\omega(k)G_n(B) + WG_n'(B)}, \tag{24}$$

and the equations that must be solved take the form

$$B = \int \frac{|v(k)|^2\,dk}{[\omega(k)G_n(B) + WG_n'(B)]^2},$$

$$W = \int \frac{\omega(k)|v(k)|^2\,dk}{[\omega(k)G_n(B) + WG_n'(B)]^2}, \tag{25}$$

where the functions $\omega(k)$, $v(k)$, and $G_n(B)$ are known; the two equations are to be solved for $B_{0,n}$ and $W_{0,n}$. Then the energy functional becomes

$$E_{0,n} \equiv F_{0,n}\{b_{0,n}\} = -W_{0,n}[G_n(B_{0,n}) + 2B_{0,n}G_n'(B_{0,n})]. \tag{26}$$

Note that $E_{0,n}$ is greater than or equal to the usual intermediate-coupling value of the energy in the no-meson approximation, since the latter minimizes the expectation of H over all states of the form of Eq. (11) without the restriction of Eq. (10). However, for both strong and weak coupling, the

two energies are the same, and, as was noted above, the one-meson states are much simpler when Eq. (10) is used.

In order to discuss the limits of weak and strong coupling, the following expressions for $\Gamma_n(x)$ are needed:

$$\Gamma_n(x) \underset{x\to 0}{\to} x^n + \frac{x^{n+1}}{(2n+3)} + \frac{x^{n+2}}{2(2n+3)} + \cdots \tag{27}$$

$$\Gamma_n(x) \underset{x\to\infty}{\to} (2n+1)!!\, e^x \times\left(\frac{1}{x} - \frac{(n+1)^2}{2x^2} + \frac{n(n+1)^2(n+2)}{8x^3} - \cdots\right).$$

The first of these follows from the definition of Γ_n, Eqs. (13), (14), and (17); for the second, see the Appendix. These expressions give for $G_n(x)$

$$\begin{aligned} G_n(x) &\underset{x\to 0}{\to} \frac{n}{x} + \frac{1}{2n+3} + \frac{2n+2}{(2n+3)^2}x + \cdots, \\ G_n'(x) &\underset{x\to 0}{\to} -\frac{n}{x^2} + \frac{2n+2}{(2n+3)^2}\cdots, \\ G_n(x) &\underset{x\to\infty}{\to} 1 - \frac{1}{x} + \frac{(n+1)^2}{2x^2} + \cdots, \\ G_n'(x) &\underset{x\to\infty}{\to} \frac{1}{x^2} - \frac{(n+1)^2}{x^3} + \cdots, \end{aligned} \tag{28}$$

and therefore

$$\begin{aligned} G_n + 2xG_n' &\underset{x\to 0}{\to} -\frac{n}{x} + \frac{1}{2n+3}\cdots, \\ &\underset{x\to\infty}{\to} 1 + \frac{1}{x} + \cdots. \end{aligned} \tag{29}$$

From Eq. (25), W is positive, so that Eqs. (26) and (29) show that for $n>0$ and weak coupling ($B\to 0$), $E_{0,n}$ is positive. Thus, the state with $b(k)=0$ has the lowest energy in this case. That is, for a given $n>0$, the lowest solution has $b(k)=0$ until B reaches a value such that $G_n(B)+2BG_n'(B)$ is positive. At that point the solution Eq. (24) becomes lower in energy. Eq. (29) shows that in strong coupling $b(k)$ is always given by Eq. (24).

For the $T=\frac{1}{2}$ state in weak coupling, $n=0$ and $G_0=\frac{1}{3}+\cdots$, so that

$$E_{0,0,\mathrm{WC}} = -\tfrac{1}{3}W_{0,0} = -3\int \frac{|v(k)|^2}{\omega(k)}\,dk, \tag{30}$$

which is also the result in perturbation theory.

In strong coupling, B is large, and Eqs. (25) give

$$W_{0,n,\mathrm{SC}} = \int \frac{|v(k)|^2\,dk}{\omega(k)} \tag{31}$$

and, hence, with Eq. (26),

$$E_{0,n,\mathrm{SC}} = -\int \frac{|v(k)|^2\,dk}{\omega(k)}. \tag{32}$$

Thus, the g^2 term in $E_{0,n}$ is the same for all n. Moreover, from Eqs. (28) it is easy to see that the term of order one in $E_{0,n}$ is also independent of n; the first n-dependent term is the one of order g^{-2}. Hence, in strong coupling all isobars are stable.

As was noted previously,[7] Eq. (32) for $n=0$ is a considerable improvement on the work of Ref. 8. Here it is evident that for $n>0$, Eq. (32) is even more of an improvement on the selfconsistent state method of Ref. 8.

III. ONE-MESON APPROXIMATION

Consider a coherent state $|nb\rangle$ of the form given by Eq. (15); the state satisfies Eq. (10). Then for isospin $T=n+\frac{1}{2}$, the one-meson subspace based on the coherent state $|nb\rangle$ is defined to be the subspace consisting of the state $|nb\rangle$ and all the states

$$\tau\cdot a^\dagger(p)|nb\rangle, \tag{33}$$

or, equivalently, the subspace consisting of $|nb\rangle$ and all the states

$$\begin{aligned} |nb\,p\rangle &= c^\dagger(p)|nb\rangle, \\ c^\dagger(p) &= \tau\cdot a^\dagger(p) - b^*(p). \end{aligned} \tag{34}$$

These latter are orthogonal to $|nb\rangle$

$$\langle nb\,|nbp\rangle = 0. \tag{35}$$

The anticommutator of the c's follows from Eqs. (4) and (5)

$$\begin{aligned} \{c(p),c^\dagger(q)\} &= 3\delta(p-q) + 2a^\dagger(q)\cdot a(p) \\ &\quad - 2b(p)c^\dagger(q) - 2c(p)b^*(q) - 2b(p)b^*(q) \end{aligned} \tag{36}$$

and, of course,

$$c(p)|nb\rangle = 0. \tag{37}$$

From Eqs. (36) and (20), it follows immediately that

$$\langle nbp|nbq\rangle = 3\delta(p-q) + 2b(p)b^*(q)(G_n - 1). \tag{38}$$

From the previous section

$$\begin{aligned} \langle nb\,|H\,|nb\rangle &= F_{0,n}\{b\} = WG_n - U - U^* \\ U &= \int v^*(k)b(k)\,dk, \end{aligned} \tag{39}$$

and Eqs. (4), (5), (20), (21), and (36) give

$$\begin{aligned} \langle nbp|H\,|nb\rangle &= w(p) \\ &= [\omega(p) + 2U^*(1-G_n)]b(p) - 3v(p), \\ \langle nbp|H\,|nbq\rangle &= 3\delta(p-q)[F_{T,n}\{b\} + \omega(p)] \\ &\quad + 2b(p)b^*(q)(G_n - 1)[\omega(p) + \omega(q) + \beta_n], \end{aligned} \tag{40}$$

$$F_{T,n}\{b\} = WG_n + \tfrac{1}{3}(U+U^*) = F_{0,n}\{b\} + \tfrac{4}{3}(U+U^*),$$

$$\beta_n = WG_n + U + U^* + W\frac{G_n'}{G_n - 1}.$$

The general state in the nb one-meson subspace can be written

$$|nbZf\rangle = Z\,|nb\rangle + \frac{1}{\sqrt{3}}\int f(p)|nbp\rangle\,dp \tag{41}$$

and the equations for Z and f are obtained by the variational principle

$$\frac{\partial}{\partial Z^*}\langle nbZf|H-\lambda\,|nbZf\rangle=0,$$
$$\frac{\delta}{\delta f^*(p)}\langle nbZf|H-\lambda\,|nbZf\rangle=0, \tag{42}$$

where λ is the usual Lagrange multiplier that turns out to be the energy. The matrix element can be written

$$\begin{aligned}\langle nbZf|H-\lambda\,|nbZf\rangle &= Z^*Z(F_{0,n}-\lambda)+Z^*\int y_n^* f+Z\int f^* y_n\\ &\quad+\int f^*(F_{T,n}+\omega-\lambda)f\\ &\quad+\alpha_n(\lambda)\int f^* b\int b^* f\\ &\quad+\eta_n\left[\int f^*\omega b\int b^* f+\int f^* b\int \omega b^* f\right]\end{aligned} \tag{43}$$

$$y_n(p)=(1/\sqrt{3})\omega(p),$$
$$\alpha_n(\lambda)=\tfrac{2}{3}(\beta_n-\lambda(G_n-1)),$$
$$\eta_n=\tfrac{2}{3}(G_n-1).$$

The right-hand side of Eq. (43) is a sum of separable terms, so that Eqs. (42) can be solved by simple algebra and quadrature. However, the algebra is complicated, so that it is useful to consider first an approximation to Eq. (43) in order to obtain some insight into the nature of the solution.

Consider first just neglecting the terms with products of integrals; suppose the variational matrix were

$$\begin{aligned}M &= Z^*Z(F_{0,n}-\lambda)+Z^*\int y_n^* f\\ &\quad+Z\int f^* y_n+\int f^*(F_{T,n}+\omega-\lambda)f,\end{aligned} \tag{44}$$

with corresponding equations for Z and f

$$(\lambda-\omega(k)-F_{T,n})f(k)=Zy_n(k),$$
$$(\lambda-F_{0,n})Z=\int y_n^* f. \tag{45}$$

This is a standard separable-potential scattering problem with the following well-known method of solution. Let the function $D_{n0}(z)$ of the complex variable z be given by

$$D_{n0}(z)=F_{0,n}-z+\int\frac{|y_n(k)|^2}{z-F_{T,n}-\omega(k)}\,dk. \tag{46}$$

Then the bound states of the system are the roots of

$$D_{n0}(z)=0 \tag{47}$$

and the scattering phase shift at energy E is the phase of $D_{n0}(E-i0)$. Note that the cut in $D_{n0}(z)$ starts at $F_{T,n}+m$ and

$$F_{T,n}=F_{0,n}+\tfrac{4}{3}(U+U^*). \tag{48}$$

Since $b_{1,n}(k)$ can be expected to be like $b_{0,n}(k)$ of Eq. (24), it follows that v is positive. Hence, $F_{T,n}>F_{0,n}$ and the unique root $F_{1,n}$ of Eq. (53) satisfies

$$F_{1,n}<F_{0,n}. \tag{49}$$

Thus, in this approximation to the one-meson approximation, there is a single bound state at $F_{1,n}$ and a nonzero phase shift starting at $F_{T,n}+m$. In the region between $F_{1,n}+m$ and $F_{T,n}+m$ the phase shift is zero.

The full Eqs. (42) can also be solved by elementary methods; the result can again be stated in terms of a D function, with

$$D_n(z)=D_{n0}(z)+D_{n1}(z),$$
$$D_{n1}(z)=\frac{I_{by}I_{yb}Y+\eta_n(I_{by}I_{\omega by}+I_{yb}I_{\omega yb})+n_n^2 I_{bb}I_{\omega by}I_{\omega yb}}{X^2-I_{bb}Y},$$
$$I_{fg}(z)\equiv\int\frac{f^*(k)g(k)}{z-F_{T,n}-\omega(k)}\,dk,$$
$$I_{\omega fg}(z)=\int\frac{\omega(k)f^*(k)g(k)}{z-F_{T,n}-\omega(k)}\,dk,\ \text{etc.}, \tag{50}$$
$$X=1-\eta I_{\omega bb},$$
$$Y=\alpha+\eta^2 I_{\omega\omega bb}.$$

The cut in D_n still runs from $F_{T,n}+m$ to $+\infty$; that is, the phase shift is zero in the energy region below $F_{T,n}+m$. The effect of D_{n1} is to possibly produce extra zeros of $D_n(z)$ below $F_{T,n}+m$; that is, in the one-meson approximation there can be more than one bound state. Let $F_{1,n}\{b\}$ be the lowest one of these; $F_{1,n}$ is a functional of $b(k)$.

In the one-meson approximation, the best $b(k)$ is the one that minimizes $F_{1,n}\{b\}$. Thus, $b(k)$ is variationally determined. Once $b(k)$ is determined, then the phase-shift for meson scattering in the one-meson approximation is also determined. In effect, this is a variational method for computing meson scattering by a static source. In the N-meson approximation, $b(k)$ and the meson scattering parameters are variationally determined by minimization of $F_{N,n}\{b\}$.

IV. NONSTATIC SOURCE

As in Ref. 7, the Hamiltonian is

$$\begin{aligned}H &= T_F+T_B+V+V^{\dagger},\\ T_F &= \int\tilde\psi^\dagger(p)t(p)\tilde\psi(p)\,dp,\\ T_B &= \int\omega(k)a^\dagger(k)\cdot a(k)\,dk,\\ V^\dagger &= -\int a^\dagger(k)\cdot\tilde\psi^\dagger(p)W(p,q)\tau\tilde\psi(q)\delta(p+k-q)\\ &\quad\times dp\,dq\,dk,\end{aligned} \tag{51}$$

where the current operator $W(p,q)$ does not involve the isospin. Let the operators $\tilde\rho(k)$ and $\tilde\tau$ be defined by

$$\tilde\rho(k)=\int\delta(p+k-q)\tilde\psi^\dagger(p)W(p,q)\tilde\psi(q)\,dp\,dq,$$
$$\tilde\tau=\int\tilde\psi^\dagger(p)\tau\tilde\psi(p)\,dp. \tag{52}$$

Then

$$\tilde\tau\tilde\rho(k)=\int\delta(p+k-q)\tilde\psi^\dagger(p)W(p,q)\tilde\psi(q)\,dp\,dq$$

$$+\int \delta(p+k-q)\tilde{\psi}^{\dagger}(s)\tilde{\psi}^{\dagger}(p)W(p,q)\tilde{\psi}(q)\tau\tilde{\psi}(s)$$
$$\times dp\,dq\,ds. \tag{53}$$

In the one-Fermion sector the last term in Eq. (53) is zero and

$$\tilde{\tau}\tilde{\rho}(k) \doteq \int \delta(p+k-q)\tilde{\psi}^{\dagger}(p)W(p,q)\tau\tilde{\psi}(q)\,dp\,dq, \tag{54}$$

where the symbol $\doteq$ will be used to mean "equals in the one-Fermion sector." Thus, in the one-Fermion sector the interaction Hamiltonian is

$$V^{\dagger} \doteq -\int \tilde{\tau}\cdot a^{\dagger}(k)\tilde{\rho}(k)\,dk. \tag{55}$$

The field operators $\tilde{\tau}$ in the one-Fermion sector obey the same commutation and anticommutation relations as the τ matrices. Thus, the appropriate localized special coherent state in the no-meson approximation is

$$|x;n,b,f\rangle = \gamma_n(\tilde{\tau}\cdot A^{\dagger}(x))|x;n,b,f\rangle,$$
$$A^{\dagger}(x) = \int b(k)a^{\dagger}(k)e^{-ik\cdot x}\,dk, \tag{56}$$
$$|x;n,b,f\rangle = (n!)^{-1/2}\{A_{i_1}{}^{\dagger}(x)\cdots A_{i_n}{}^{\dagger}(x)|x;f\rangle\}^{n+1/2},$$
$$|x;f\rangle = \int e^{-ip\cdot x}\tilde{\psi}^{\dagger}(p)\tilde{f}(p)dp|\Omega\rangle.$$

The factors $e^{-ik\cdot x}$ and $e^{-ip\cdot x}$ are useful for going from localized states to translated localized states.[9] Again, the coherent states of (56) satisfy

$$\tilde{\tau}\cdot a(k)|x;n,b,f\rangle = b(k)e^{-ik\cdot x}|x;n,b,f\rangle; \tag{57}$$

generally, the algebra of the preceding sections applies with appropriate modifications. Thus

$$D_n(x) \equiv \left\langle \frac{x}{2};nbf\right| - \left.\frac{x}{2};nbf\right\rangle = D_{nB}(x)D_F(x),$$
$$D_F(x) = \int e^{ip\cdot x}\tilde{f}^{\dagger}(p)\tilde{f}(p)\,dp,$$
$$D_{nB}(x) = \Gamma_n(B(x)),$$
$$B(x) = \int e^{ik\cdot x}|b(k)|^2\,dk,$$
$$A_n(x) \equiv \langle x/2;nbf|H|-x/2;nbf\rangle \tag{58}$$
$$= \Gamma_n(B)\left\{t(x) - \int [b^*(k)\tilde{\rho}_f(k,x) + \tilde{\rho}_f^*(k;-x)b(k)]\,dk\right\}$$
$$+ \Gamma_n'(B(x))D_F(x)\int \omega(k)|b(k)|^2 e^{ik\cdot x}\,dk,$$
$$t(x) = \int e^{ip\cdot x}\tilde{f}^{\dagger}(p)t(p)\tilde{f}(p)\,dp,$$
$$\tilde{\rho}_f(k,x) = \int \delta(p+k-q)\tilde{f}^{\dagger}(p)W(p,q)\tilde{f}(q)e^{iq\cdot x}\,dp\,dq.$$

As in Refs. 7 and 10, the weak-coupling approximation is obtained by taking $f(p)$ to be constant in the translated-localized-state (TLS) energy functional

$$F_{\mathrm{TLS}} = \frac{\int A(x)\,dx}{\int D(x)\,dx}; \tag{59}$$

the result is the same as second order peturbation theory, namely, for $n>0$ the minimum occurs for $b(k)=0$ and for $n=0$ the weak-coupling energy is

$$E_{WC,n=0} = t(0) - 3\int \frac{|W(k,0)|^2}{\omega(k)+k^2/2M}\,dk, \tag{60}$$

where $t(p)$ has been taken to be equal to $t(0)+p^2/2M$. In strong coupling, the localized-state energy functional F_{LS},

$$F_{\mathrm{LS}} = A(0)/D(0) \tag{61}$$

becomes independent of n

$$F_{\mathrm{LS;SC}}\{f\} = \int \tilde{f}^{\dagger}(p)t(p)\tilde{f}(p) - \int \frac{|\tilde{\rho}_f(k,0)|^2}{\omega(k)}\,dk, \tag{62}$$

and is identical with the strong-coupling energy functional for isoscalar field. It is evident that the static model is related to a corresponding nonstatic source only in the case of strong coupling.

ACKNOWLEDGMENT

This work was performed under the auspices of the United States Department of Energy.

APPENDIX

Let

$$\gamma_n(x) = \sum_{\mu\ \mathrm{even}} \frac{(n+\frac{1}{2})!}{2^{\mu}(\mu/2)!(n+\frac{1}{2}+\mu/2)!}x^{\mu}$$
$$+ \sum_{\mu\ \mathrm{odd}} \frac{(n+\frac{1}{2})!}{2^{\mu}((\mu-1)/2)!(n+1+(\mu/2))!}x^{\mu}$$
$$= \alpha_n(x) + \beta_n(x).$$

Then

$$\alpha_n'(x) = \sum_{\mu\ \mathrm{odd}} \frac{(n+\frac{1}{2})!x^{\mu}}{2^{\mu}((\mu-1)/2)!(n+1+\mu/2)!} = \beta_n(x),$$
$$\gamma_n(x) = \left(1+\frac{d}{dx}\right)\alpha_n(x),$$
$$\frac{1}{x}\alpha_n'(x) = \sum_{\mu\ \mathrm{even}} \frac{(n+\frac{1}{2})!}{2^{\mu+1}(\mu/2)!(n+\frac{3}{2}+\mu/2)!}x^{\mu}$$
$$= \frac{1}{2n+3}\alpha_{n+1}(x),$$
$$\alpha_n(x) = \frac{2n+1}{x}\frac{d}{dx}\alpha_{n-1}(x)$$
$$= (2n+1)!!\left(\frac{1}{x}\frac{d}{dx}\right)^n\alpha_0(x),$$
$$\alpha_0(x) = 1 + \frac{x^2}{2.3} + \frac{x^4}{2.3.4.5} + \cdots = \frac{\sinh x}{x}.$$

Thus,

$$\gamma_n(x) = (2n+1)!!\left(1+\frac{d}{dx}\right)\left(\frac{1}{x}\frac{d}{dx}\right)^n \frac{\sinh x}{x}$$

$$\underset{\infty}{\to} (2n+1)!!\left(1+\frac{d}{dx}\right)\left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{e^x}{2x}\right),$$

$$= (2n+1)!!1 + \frac{d}{dx}\left[\frac{e^x}{2}\left(\frac{1}{x^{n+1}} - \frac{n(n+1)}{2x^{n+2}}\right.\right.$$

$$\left.\left. + \frac{(n-1)n(n+1)(n+2)}{8x^{n+3}} - \cdots\right)\right],$$

$$= (2n+1)!!e^x\left(\frac{1}{x^{n+1}} - \frac{(n+1)^2}{2x^{n+2}}\right.$$

$$\left. + \frac{n(n+1)^2(n+2)}{8x^{n+3}} - \cdots\right).$$

[1]See, e.g., E.M. Henley and W. Thirring, *Elementary Quantum Field Theory* (McGraw-Hill, New York, 1962).
[2]S. Tomonoga, Progr. Theor. Phys. (Japan) **2**, 6 (1947).
[3]F. Harlow and B. Jacobsohn, Phys. Rev. **93**, 333 (1954).
[4]T. D. Lee and R. Christian, Phys. Rev. **94**, 1760 (1954).
[5]M. H. Friedman, T. D. Lee, and R. Christian, Phys. Rev. **100**, 1494 (1955).
[6]E. M. Henley and T. D. Lee, Phys. Rev. **101**, 1536 (1956).
[7]M. Bolsterli, Phys. Rev. D **16**, 1749 (1977).
[8]M. Bolsterli, Phys. Rev. D **6**, 3501 (1972).
[9]M. Bolsterli, Phys. Rev. D **13**, 1727 (1976).

Quantum field theory of particles of indefinite mass. I

Levere Hostler
Physics Department, Wilkes College, Wilkes-Barre, Pennsylvania 18766

(Received 4 March 1980; accepted for publication 16 May 1980)

A quantum field theory of particles of indefinite mass is derived using rigorous correspondence arguments starting with a classical theory of particles of indefinite mass. The classical theory can be recovered in a suitable limiting case as $\hbar \to 0$ by means of Ehrenfest's theorem. Our deviation leads to a quantum mechanical wave equation which turns out to be basically the same equation investigated earlier by Fock, Nambu, and others, but differs from this earlier equation in our use of a new evolution parameter—herein called "evolution-time"—defined as proper time divided by the classical mass. Owing to our use of evolution time as the development parameter of the system our Fock equation is without any reference to a mass parameter, in contrast to the older Fock equation. The indefiniteness of the particle mass frees the ordinary time, t—herein called "observer's time"—of any fixed relationship to the evolution time, and the two times becomes quite independent parameters. The Hamiltonian of our system turns out to be (minus half) the total mass squared of the system and is a constant of the motion. In order to guarantee negative definiteness of the Hamiltonian of the second quantized system, it is necessary to quantize our Fock equation using Fermi–Dirac statistics. A real scalar field is described by an equation which is second-order in the evolution time, obtained by iterating our first-order Fock equation. The second-order Fock equation must be second quantized using Bose–Einstein statistics, in order to preserve the interpretation of the Fourier amplitudes as creation and annihilation operators. The propagator for the second-order Fock equation has a Pauli–Villars type sum of terms, in which one term describe the propagation of timelike states, the other term describes space-like states.

I. INTRODUCTION

Efforts to cast the laws of quantum field theory in different forms have been spurred by the desire to overcome the divergence difficulties of canonical field theory and to overcome the limitations of perturbation theory. Also there is the possibility that a particular form of quantum field theory may provide a convenient language for treating particular problems. That a formalism like the one presented here may have some utility in elementary particle physics is suggested by a recent work of Feynman, Kislinger, and Ravndal.[1] Although Feynman *et al.* state that they are just "curve fitting", their use of mass Hamiltonians[2] could be fit nicely into the first quantized form of the theory presented here. The theory presented here is a form of "proper time theory." Such theories have been investigated earlier by a number of authors including Fock,[3] Nambu,[4] Katayama *et al.*,[5] and Fanchi.[6] Also, the indefinite mass concept has been investigated earlier by Greenberger.[7]

In Sec. II we construct a classical theory of indefinite mass particles by use of an ensemble concept. Our ensemble is described in Hamiltonian langauge with the Hamiltonian being (minus half) the mass squared of the system. The evolution parameter, herein called "evolution time," of our Hamiltonian system differs from the proper time by a factor equal to the reciprocal mass of the particle. The indefiniteness of the mass of the particle described by the ensemble eliminates any possibility of a definite relation between evolution time and ordinary time t, herein called "observer's time," so that evolution time and observer's time become quite independent parameters. First quantization of this classical theory is carried out in Sec. III and is found to lead to essentially the same equation investigated earlier by Fock,[3] Nambu,[4] and others. The principal difference between our Fock equation, (3.16), and the equation originally investigated by Fock is the total absence of a mass parameter in the equation.[8] Our classical ensemble picture can be recovered from our Fock equation in a suitable limiting case as $\hbar \to 0$ by use of Ehrenfest's theorem. We interpret the wavefunction of our Fock equation as a spacetime probability amplitude, so that at any particular evolution time the quantity $|\phi(\mathbf{x},\tau)|^2 d^4x$ gives the probability to find the particle in d^3r during the observer's time dx^0. When we proceed in Sec. IV. A to second quantize, we find that our Fock equation must be second quantized using Fermi–Dirac statistics, in order to guarantee the negative definiteness of the Hamiltonian. The real scalar field is taken up in Sec. IV.B. We describe this field by a second-order (in evolution time) equation. The second quantization of this second-order Fock equation must be carried out using Bose–Einstein statistics, in order to preserve the interpretation of the Fourier amplitudes as creation and annihilation operators. The propagator (4.41) of the second-order Fock equation has the form of a Pauli–Villars[9] type of sum, there being one term describing timelike states and one term describing spacelike states. Fully interacting fields will be investigated in a companion paper.

II. CLASSICAL THEORY

We begin with the known covariant equations of motion[10]

$$m \frac{d^2x^\mu}{ds^2} = qF^{\mu\nu} \frac{dx_\nu}{ds} \tag{2.1}$$

 0022-2488/80/092461-07$1.00

of a charged particle in an electromagnetic field $F^{\mu\nu}$. As is known, equations (2.1) admit the integral of the motion

$$\frac{dx^\mu}{ds}\frac{dx_\mu}{ds} = \text{constant}. \tag{2.2}$$

In order to assure the proper-time interpretation of the evolution parameter s only solutions for which the integral of the motion (2.2) is equal to unity are allowed. The constraint equation (2.2), which is a nonholonomic constraint, can be lifted by the following device: we introduce a new evolution parameter τ related to the proper time s by a scale factor involving the mass of the particle

$$s \equiv m\tau. \tag{2.3}$$

The parameter τ will be referred to as the *evolution-time* of the system. To distinguish between this evolution-time τ and the time x^0 the latter will be called the *observer's time*. In terms of evolution-time the equations (2.1) take the form

$$\frac{d^2x^\mu}{d\tau^2} = qF^{\mu\nu}\frac{dx_\nu}{d\tau}, \tag{2.4}$$

from which the mass parameter has disappeared. The mass parameter reappears through the integral of the motion [analogous to (2.2)]

$$\frac{dx^\mu}{d\tau}\frac{dx_\mu}{d\tau} = \text{constant} = m^2. \tag{2.5}$$

Equation (2.4) is quite suitable for describing a classical ensemble of particles, the various members of the ensemble being like particles aside from their mass. The same differential equations (2.4) would correctly describe the motion of any member of the ensemble, the particular member that we are looking at being evidenced by the value of the integral of the motion (2.5). From this ensemble point of view there are now no "useless" solutions of the relativistic equations of motion.

As is known, a description of a physical system from an ensemble point of view is best carried out within the framework of Hamiltonian dynamics. We therefore transform Eq. (2.4) into a Hamiltonian form:

$$\frac{dx^\mu}{d\tau} = \frac{\partial H}{\partial p^\mu} = PB(x^\mu;H) \tag{2.6a}$$

$$\frac{dp_\mu}{d\tau} = -\frac{\partial H}{\partial x^\mu} = PB(p_\mu;H), \tag{2.6b}$$

$$H = -\tfrac{1}{2}(p^\mu + q\Omega^\mu)(p_\mu + q\Omega_\mu); \tag{2.7}$$

where Ω_μ is the 4–potential of the electromagnetic field, defined through the condition $F_{\mu\nu} = \partial_\mu\Omega_\nu - \partial_\nu\Omega_\mu$. The symbols $PB(;)$ here signify the classical Poisson bracket[11]

$$PB(A;B) \equiv \sum_{\mu=0}^{3}\left(\frac{\partial A}{\partial x^\mu}\frac{\partial B}{\partial p_\mu} - \frac{\partial A}{\partial p_\mu}\frac{\partial B}{\partial x^\mu}\right). \tag{2.8}$$

Our classical statistical ensemble is now described by a phase space density $\rho(\mathbf{x},\mathbf{p},\tau)$ obeying a Liouville's equation

$$\frac{\partial\rho}{\partial\tau} + PB(\rho;H) = 0. \tag{2.9}$$

Our Hamiltonian (2.7) is readily shown to be a constant of the motion and to have the simple physical interpreation of minus half the mass-squared of the particle.

Note that to a particular evoluton-time τ there will in general correspond a statistical distribution of x^0—values in the ensemble; owing to the indefiniteness in the particle mass there is now no fixed relationship between evolution time and observer's time. Indeed, in a theory with indefinite mass the observer's time x^0 and the evolution-time τ become quite independent parametrs; because of the statistical distribution of x^0—values in a theory with indefinite mass, the observer's time x^0 will enter the following formalism entirely on a par with the coordinates x^1, x^2, x^3.

III. QUANTUM THEORY

The Hamiltonian equations of motion (2.6) suggest a covariant quantization procedure which parallels the canonical formalism except for the use of our new evolution parameter τ. We thus promote the variables x^μ, p_μ to the status of self-adjoint operators on Hilbert space. The basic commutators of our theory must be obtained from the classical relations

$$PB(x^\mu;p_\nu) = \delta^\mu{}_\nu \tag{3.1}$$

by setting up a suitable correspondence

$$PB(A;B) \propto i[A;B] \tag{3.2}$$

between the classical Poisson bracket and quantum mechanism commutator. In setting up this correspondence we appeal to the usual nonrelativistic commutation relations as a guide. These nonrelativistic commutation relations

$$\left[x^k; m\frac{dx^j}{dx^0} + q\Omega^j\right] = i\delta_{kj}, \quad k,j = 1,2,3, \tag{3.3}$$

can be written in the form

$$\left[x^k; \frac{dx^j}{d(x^0/m)} + q\Omega^j\right] = i\delta_{kj}. \tag{3.4}$$

Since in the nonrelativistic limit the evolution-time τ is approximately equal to x^0/m, we can adopt the commutation relation

$$\left[x^k; \frac{dx^j}{d\tau} + q\Omega^j\right] = -ig^{kj}, \quad k,j = 1,2,3, \tag{3.5}$$

in our relativistic theory without altering the classical limit. Generalizing, Eqs. (3.5) to incorporate time components; we have $[x^\mu;(dx^\nu/d\tau) + q\Omega^\nu] = -ig^{\mu\nu}$. Incorporating the relation $p^\mu = -[(dx^\mu/d\tau) + q\Omega^\mu]$ (from the equation of motion (2.6a)), we have

$$[x^\mu;p_\nu] = i\delta^\mu{}_\nu. \tag{3.6}$$

These are the basic commutators in our relativistic theory. Equations (3.6) together with Eq. (3.1) give the relation $[x^\mu;p_\nu] = iPB(x^\mu;p_\nu)$ indicating that the proportionality constant in the general formula (3.2) is equal to minus one:

$$PB(A;B) = (1/i)[A;B]. \tag{3.7}$$

Accordingly, the Heisenberg equations of motion (2.6a) and (2.6b) go over into

$$\frac{dx^\mu}{d\tau} = \frac{1}{i}[x^\mu;H] \tag{3.8a}$$

and

$$\frac{dp_\mu}{d\tau} = \frac{1}{i}[p_\mu;H] \tag{3.8b}$$

in the quantum mechanical case. Equations (3.8a) and (3.8b) and Eq. (3.6) are the basic equations of our new relativistic quantum theory. It can be verified by direct calculation that Eq. (3.8a) reproduces the correct relation $p^\mu = -[(dx^\mu/d\tau) + q\Omega^\mu]$ between the 4–velocity $\dot{x}^\mu \equiv dx^\mu/d\tau$ and the 4–momentum p^μ; and Eq. (3.8b) gives essentially the equation of motion (2.4) with which we started. For this calculation the commutator [which follows from Eq. (3.6)]

$$[F(\mathbf{x});p_\nu] = i\frac{\partial F(\mathbf{x})}{\partial x^\nu} \tag{3.9}$$

is required.

Let the evolution of operators in our theory be effected by means of a unitary operator $U(\tau)$ such that

$$x^\mu(\tau) = U^{-1}x^\mu(0)U \tag{3.10a}$$

and

$$p_\nu(\tau) = U^{-1}p_\nu(0)U. \tag{3.10b}$$

Substituting in the Heisenberg equations of motion (3.8a) and (3.8b) leads to the following differential equation for $U(\tau)$

$$H[\mathbf{x}(0),\mathbf{p}(0)]U + \frac{1}{i}\frac{dU}{d\tau} = 0. \tag{3.11}$$

The key to the contact between experimental observations and our theoretical structure must lie in expectation values such as

$$\langle F\rangle = \langle\chi|F(\mathbf{x}(\tau),\mathbf{p}(\tau),\tau)|\chi\rangle, \tag{3.12}$$

in which $|\chi\rangle$ is a Heisenberg state vector independent of evolution-time τ. The expectation value (3.12) can be reduced to an expression containing only the initial values $\mathbf{x}(0)$, $\mathbf{p}(0)$ of the canonical variables

$$\begin{aligned}\langle F\rangle &= \langle\chi|F(U^{-1}\mathbf{x}(0)U,U^{-1}\mathbf{p}(0)U,\tau)|\chi\rangle \\ &= \langle\chi|U^{-1}F(\mathbf{x}(0),\mathbf{p}(0),\tau)U|\chi\rangle,\end{aligned}$$

or

$$\langle F\rangle = \langle\phi|F(\mathbf{x}(0),\mathbf{p}(0),\tau)|\phi\rangle, \tag{3.13}$$

in which the τ–dependent state vector $|\phi\rangle$ is defined as

$$|\phi\rangle \equiv U(\tau)|\chi\rangle. \tag{3.14}$$

This transformation in which the evolution-time dependence has been moved from the operators to the state vectors corresponds to the familiar transition from the Heisenberg to the Schrödinger picture. In fact, the physical consequences of the theory can be considered as contained in Eq. (3.13), where the Schrödinger picture state vector is given by Eq. (3.14). A differential equation for the direct calculation of the Schrödinger wave function $|\phi\rangle$ can be obtained by combining the defining Eq. (3.14) with the equation of motion (3.11) of the evolution operator U

$$H[\mathbf{x}(0),\mathbf{p}(0)]|\phi\rangle + (1/i)(\partial/\partial\tau)|\phi\rangle = 0. \tag{3.15}$$

If we substitute the explicit expression (2.7) for the Hamiltonian function and go over to a coordinate representation in the usual way, Eq. (3.15) becomes

$$\left\{-\tfrac{1}{2}\left(i\frac{\partial}{\partial x^\mu} - q\Omega_\mu\right)\left(i\frac{\partial}{\partial x_\mu} - q\Omega^\mu\right) + \frac{1}{i}\frac{\partial}{\partial\tau}\right\}\phi(\mathbf{x},\tau) = 0, \tag{3.16}$$

in which the coordinate space form

$$p_\nu \equiv \frac{1}{i}\frac{\partial}{\partial x^\nu} \tag{3.17}$$

of the momentum operator, which can be derived from Eq. (3.6), appears. Equation (3.16) is the Schrödinger equation of our new relativistic quantum theory. It is basically the same equation investigated earlier by Fock,[3] and Nambu,[4] and differs from their equation only in our elimination of any explicit reference to a mass parameter.[8] We have here arrived at Eq. (3.16), which we shall refer to as the first-order Fock equation, through rigorous correspondence arguments which bring out a particular physical interpretation. Taking our cue from the classical ensemble picture of the system under investigation, we postulate that our Fock equation must describe a particle whose mass has no sharp value. Indeed, if we take $|\phi(\mathbf{x},\tau)|^2\, d^4x$ to be the probability to observe the particle in d^3r during the time interval dx^0 of observer's time, then we shall have an Ehrenfest's theorem for Eq. (3.16) which will guarantee our recovering the classical ensemble picture of Sec. II in a suitable limiting case as $\hbar \to 0$. This interpretation of the wavefunction of Eq. (3.16) as a spacetime probability amplitude is corroborated by the existence of a "5–current" convervation law

$$\frac{\partial}{\partial\tau}[\phi^\dagger\phi] + \partial_\mu[\tfrac{1}{2}\phi^\dagger(i\overleftrightarrow{\partial}^\mu - 2q\Omega^\mu)\phi] = 0, \tag{3.18}$$

according to which the integral over all spacetime $\int d^4x\,|\phi|^2$ is a constant of the motion. Accordingly, if the integral $\int d^4x\,|\phi|^2$ is normalized to unity at one evolution time, ϕ will remain so normalized for all evolution time, as befits a proper probability amplitude. Note that the object $(\tfrac{1}{2})\phi^\dagger(i\overleftrightarrow{\partial}^\mu - 2q\Omega^\mu)\phi$ that we are inclined to identify with the usual probability 4–current is here no longer a conserved 4–current. On the other hand, if we integrate Eq. (3.18) over all evolution time, a conserved 4–current does emerge

$$\begin{aligned}&\partial_\mu j^\mu = 0 \\ &j^\mu \equiv \int_{-\infty}^{\infty} d\tau\, \tfrac{1}{2}\phi^\dagger(i\overleftrightarrow{\partial}^\mu - 2q\Omega^\mu)\phi.\end{aligned} \tag{3.19}$$

It is this j^μ which is to be identified with the conserved 4–current that we are accustomed to.

IV. SECOND QUANTIZATION, FREE FIELDS

A. The first-order Fock equation: Particles obeying Fermi–Dirac statistics

To second quantize the first-order Fock equation we can simply imitate the procedure of the usual canonical formalism except for the use of our new evolution parameter τ. A Lagrangian density for the free field form of Eq. (3.16) is

$$\mathcal{L} = (1/2i)(\dot{\phi}^\dagger\phi - \phi^\dagger\dot{\phi}) + \tfrac{1}{2}(\partial_\mu\phi^\dagger)\partial^\mu\phi. \tag{4.1}$$

To discover the operator structure of the quantized theory we appeal to a Schwinger action principle,[12] writing

$$\delta\phi(2) = (1/i)[\phi(2);\delta W]$$

and

$$\delta\phi(2)^\dagger = (1/i)[\phi(2)^\dagger;\delta W], \tag{4.2}$$

in which

$$\delta W = \int d^4x \left\{\delta\phi \frac{\partial \mathcal{L}}{\partial \dot{\phi}} - \phi\delta\left(\frac{\partial \mathcal{L}}{\partial \dot{\phi}}\right) + \delta\phi^\dagger \frac{\partial \mathcal{L}}{\partial \dot{\phi}^\dagger} - \phi^\dagger\delta\left(\frac{\partial \mathcal{L}}{\partial \dot{\phi}^\dagger}\right)\right\}, \tag{4.3}$$

is the change in the action $W \equiv \int d\tau \int d^4x\, \mathcal{L}$ due to a variation of the generalized coordinates and canonical momenta at evolution time $\tau = \tau_2$. We would obtain either Bose–Einstein or Fermi–Dirac statistics according as we treat the $\delta\phi(2)$ and $\delta\phi^\dagger(2)$ as commuting or anticommuting c–numbers. Now in the following it will be seen that the use of Bose–Einstein statistics in connection with the first-order Fock equation leads to a non negative definite Hamiltonian. To avoid this problem we must quantize the first-order Fock equation using Fermi–Dirac statistics. The action principle then leads to the one (evolution) time anticommutation relations

$$\{\phi(2);\phi(1)^\dagger\} = \delta^4(21). \tag{4.4}$$

To proceed, we require the Fourier decomposition of the free field. This is

$$\phi(\mathbf{x},\tau) = \int \frac{d^4k}{(2\pi)^4} a(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x} + (ik^2/2)\tau}. \tag{4.5}$$

Note that completeness of the continuum states $\exp(-i\mathbf{k}\cdot\mathbf{x})$ requires that the integral (4.5) include spacelike states $k^2 < 0$ as well as timelike states $k^2 > 0$. These spacelike states will eventually be given a special treatment, but for the moment it is convenient to proceed using (4.5) in which they appear on the same footing as the timelike states. By inverting the Fourier transform in Eq. (4.5) and using Eq. (4.4), we find the momentum space anticommutation relations

$$\{a(\mathbf{k}_2);a(\mathbf{k}_1)^\dagger\} = (2\pi)^4\delta^4(\mathbf{k}_2 - \mathbf{k}_1). \tag{4.6}$$

The Hamiltonian H of the system is identified through the change in the action functional due to a variation of the evolution time associated with the various points on the orbit of our dynamical system:

$$\delta W = -\delta\tau H. \tag{4.7}$$

We find $H = \int d^4x\, \mathcal{H}$ where the Hamiltonian density $\mathcal{H}$ has the form:

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial \dot{\phi}}\dot{\phi} + \frac{\partial \mathcal{L}}{\partial \dot{\phi}^\dagger}\dot{\phi}^\dagger - \mathcal{L}. \tag{4.8}$$

The expression (4.8) evaluates to

$$\mathcal{H} = -\tfrac{1}{2}(\partial_\mu\phi^\dagger)\partial^\mu\phi. \tag{4.9}$$

Next we substitute our Fourier representation (4.5) into Eq. (4.9) and integrate over all spacetime, in order to express the total Hamiltonian in the occupation number representation. The result is

$$H = \int \frac{d^4k}{(2\pi)^4} a(\mathbf{k})^\dagger a(\mathbf{k})(-\tfrac{1}{2}k^2). \tag{4.10}$$

It is at this point that we encounter the above mentioned difficulty with negative definiteness of the Hamiltonian: timelike states make a negative contribution in Eq. (4.10), spacelike states make a positive contribution. If we had been quantizing using Bose–Einstein statistics, we would here have an impasse. With the use of Fermi–Dirac statistics, however, we can readily extricate ourselves from difficulty. We simply redefine creation and annihilation operators for the spacelike states, in effect interchanging $a(\mathbf{k})$ and $a(\mathbf{k})^\dagger$. Thus, we define $\mathcal{A}(\mathbf{k}) \equiv a(\mathbf{k})^\dagger$ for spacelike states, and interpret $\mathcal{A}(\mathbf{k})$ and not $a(\mathbf{k})$ as the annihilation operator of our theory. We note that the anticommutation relations (4.6) permit either interpretation. The following equations summarize our results in the new representation. The free field expansion (4.5) and anticommutation relations (4.6) go over into the forms

$$\phi = \int_{k^2>0} \frac{d^4k}{(2\pi)^4} a(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x} + ik^2\tau/2} + \int_{k^2<0} \frac{d^4k}{(2\pi)^4} \mathcal{A}(\mathbf{k})^\dagger e^{-i\mathbf{k}\cdot\mathbf{x} + ik^2\tau/2}, \tag{4.11}$$

and

$$\{a(\mathbf{k}_2);a(\mathbf{k}_1)^\dagger\} = (2\pi)^4\delta^4(\mathbf{k}_2 - \mathbf{k}_1), \quad k^2 > 0, \tag{4.12a}$$

$$\{\mathcal{A}(\mathbf{k}_2);\mathcal{A}(\mathbf{k}_1)^\dagger\} = (2\pi)^4\delta^4(\mathbf{k}_2 - \mathbf{k}_1), \quad k^2 < 0, \tag{4.12b}$$

respectively. Our new Hamiltonian is (we continue to use the same symbol H for the Hamiltonian, although in going over to the new representation we have dropped an infinite c–number term):

$$H = \int_{k^2>0} \frac{d^4k}{(2\pi)^4} a(\mathbf{k})^\dagger a(\mathbf{k})(-\tfrac{1}{2}k^2) + \int_{k^2<0} \frac{d^4k}{(2\pi)^4} \mathcal{A}(\mathbf{k})^\dagger \mathcal{A}(\mathbf{k})(-\tfrac{1}{2}|k^2|). \tag{4.13}$$

Next we look at the free particle propagator. This is defined in the present context as the vacuum expectation value of an ordered product of field operators, the ordering here being with respect to the evolution time τ

$$i\Delta_F(21) \equiv \langle 0|T(\phi(2)\phi(1)^\dagger)|0\rangle, \tag{4.14}$$

$$T(\phi(2)\phi(1)^\dagger) \equiv \theta(\tau_2 - \tau_1)\phi(2)\phi(1)^\dagger - \theta(\tau_1 - \tau_2)\phi(1)^\dagger\phi(2).$$

The details of the calculation of a vacuum expectation value such as (4.14) are quite standard.[13] The end result of the calculations is most simply expressed in momentum space. This result is

$$i\Delta_F(2,1) \equiv \int \frac{d^4\mathbf{k}}{(2\pi)^4} e^{-i\mathbf{k}\cdot(\mathbf{x}_2 - \mathbf{x}_1)} i\Delta_F(\mathbf{k},\tau_2 - \tau_1), \tag{4.15}$$

where

$$i\Delta_F(\mathbf{k},\tau_2 - \tau_1) = \theta(\tau_2 - \tau_1)e^{ik^2(\tau_2 - \tau_1)/2}\theta(k^2) - \theta(\tau_1 - \tau_2)e^{ik^2(\tau_2 - \tau_1)/2}\theta(-k^2). \tag{4.16}$$

Another form of Eq. (4.16), which eliminates the explicit appearance of theta functions, is

$$i\Delta_F(\mathbf{k},\tau_2-\tau_1)=\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{d\lambda e^{ik^2\lambda/2}}{\lambda(1-i\epsilon)-(\tau_2-\tau_1)}. \tag{4.17}$$

In the following we shall refer to Eq. (4.16) or (4.17) as a "mixed" representation of the propagator, since it is a momentum space representation in the coordinates x^μ, but can be considered a coordinate space representation in τ. A pure Fourier representation is obtained by transforming in the evolution time as well. Defining

$$i\Delta_F(\mathbf{k},\tau_2-\tau_1)\equiv\int_{-\infty}^{\infty}\frac{dN}{2\pi}i\Delta_F(\mathbf{k},N)e^{iN(\tau_2-\tau_1)/2}, \tag{4.18}$$

we find

$$i\Delta_F(\mathbf{k},N)=\frac{i}{k^2-N(1-i\epsilon)},\quad \epsilon>0. \tag{4.19}$$

We note here an interesting correspondence between boundary conditions in evolution time and the usual Feynman boundary conditions. According to Eq. (4.16) all timelike states propagate strictly forward in evolution time. This strictly retarded propagation in evolution time for the timelike states implies exactly the forward and backward propagation in observer's time that we are accustomed to; because for $N>0$ the $i\epsilon$ prescription in Eq. (4.19) is precisely the Feynman prescription.[14]

B. The second-order Fock equation: Particles obeying Bose–Einstein statistics

We arrive at an equation that can be quantized using Bose–Einstein statistics by considering a real field. The first-order Fock equation

$$\left(-\tfrac{1}{2}i\partial_\mu i\partial^\mu+\frac{1}{i}\frac{\partial}{\partial\tau}\right)\phi=0 \tag{4.20}$$

is not suitable for describing a real field, since it does not remain invariant under complex conjugation. To rememdy this we eliminate imaginaries in Eq. (4.20) by iteration, obtaining

$$\left\{(\tfrac{1}{2}\Box^2)^2+\frac{\partial^2}{\partial\tau^2}\right\}\phi=0,$$
$$-\Box^2\equiv i\partial_\mu i\partial^\mu. \tag{4.21}$$

We shall refer to Eq. (4.21) as the "second-order Fock equation." In eliminating imaginaries by iteration we have paid a rather high price viz., our wave equation now involves the D'Alembertian operator ($\Box^2$) raised to the second power. The unpleasantness of this situation originates from our previous experience in which the observer's time is the evolution parameter. As regards the evolution parameter in the present case, it enters our wave equation (4.21) only in second order. This appearance of the evolution parameter in second order is quite ordinary and ameliorates our situation somewhat. On the other hand, the wave equation (4.21) still has the unusual feature that particle and antiparticle remain distinct even though the field is real. For this reason the use of Eq. (4.21) to describe a real field may turn out to be provisional. Perhaps we should think of the second-order Fock equation as describing more or less phenomenologically the overall dynamics of a particle which is really a composite system made up of two Fermions.

With the above reservations, we proceed to take the second-order Fock equation seriously and to explore its consequences. Due to the high-order derivatives in the wave equation, we must now allow the Lagrangian density to include a dependence on $\partial_\alpha\partial_\beta\phi$. The action principle then leads to the Lagrangian equations of motion

$$\frac{d}{d\tau}\left(\frac{\partial\mathscr{L}}{\partial\dot\phi}\right)+\partial_\alpha\frac{\partial\mathscr{L}}{\partial\partial_\alpha\phi}-\partial_\alpha\partial_\beta\frac{\partial\mathscr{L}}{\partial\partial_\alpha\partial_\beta\phi}-\frac{\partial\mathscr{L}}{\partial\phi}=0. \tag{4.22}$$

The change in the action due to a variation of the generalized coordinates and canonical momenta at the end points is

$$\delta W=\int d^4x\left\{\delta\phi\frac{\partial\mathscr{L}}{\partial\dot\phi}-\phi\delta\left(\frac{\partial\mathscr{L}}{\partial\dot\phi}\right)\right\}, \tag{4.23}$$

and the Hamiltonian is

$$H=\int d^4x\left(\frac{\partial\mathscr{L}}{\partial\dot\phi}\dot\phi-\mathscr{L}\right). \tag{4.24}$$

For the particular case of the second-order Fock equation, the Lagrangian density is

$$\mathscr{L}=-\tfrac{1}{2}\dot\phi^2+\tfrac{1}{8}(\Box^2\phi)^2. \tag{4.25}$$

The Hamiltonian of our particular system is [from Eqs. (4.24) and (4.25)]

$$H=-\int d^4x\left|\tfrac{1}{2}\dot\phi^2+\tfrac{1}{8}(\Box^2\phi)^2\right|. \tag{4.26}$$

To obtain the quantum conditions we require

$$\delta\phi(2)=(1/i)[\phi(2);\delta W],$$
$$\delta\dot\phi(2)=(1/i)[\dot\phi(2);\delta W], \tag{4.27}$$

in which [from Eqs. (4.23) and (4.25)]

$$\delta W=\int d^4x_1\{-\delta\phi(1)\dot\phi(1)+\phi(1)\delta\dot\phi(1)\}. \tag{4.28}$$

The one (evolution) time commutation relations

$$[\phi(2);\dot\phi(1)]=(1/i)\delta^4(21) \tag{4.29}$$

now follow, assuming the variations $\delta\phi(2)$ and $\delta\dot\phi(2)$ to behave like commuting c–numbers. The option of treating the $\delta\phi(2)$ and $\delta\dot\phi(2)$ like anticommuting c–numbers will be ruled out in anticipation of the fact that we cannot then have a consistent interpretation of the Fourier amplitudes as creation and annihilation operators.

To proceed, we go into momentum space, writing

$$\phi\equiv\int\frac{d^4k}{(2\pi)^4(|k^2|)^{1/2}}\{a(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}+ik^2\tau/2}+a(\mathbf{k})^\dagger e^{i\mathbf{k}\cdot\mathbf{x}-ik^2\tau/2}\}. \tag{4.30}$$

As in our earlier treatment of the first-order Fock equation, spacelike states will eventually be given special treatment; but for the moment we continue using the representation (4.30) in which they appear on the same footing as the timelike states. In order to obtain the commutators of the Fourier amplitudes, we first project out a particular amplitude by forming[15]

$$\int d^4x\ (e^{-i\mathbf{k}\cdot\mathbf{x}+ik^2\tau/2})^*(1/i)(\partial/\partial\tau)\phi$$
$$= \epsilon(k^2)(|k^2|)^{1/2}a(\mathbf{k}),$$
$$\epsilon(k^2)\equiv +1, \quad k^2>0; \quad \epsilon(k^2)\equiv -1, \quad k^2<0. \tag{4.31}$$

To calculate the commutator $[a(\mathbf{k}_2); a(\mathbf{k}_1)^\dagger]$ we substitute the representation (4.31) and use the coordinate space commutation relations (4.29). The result of this calculation is

$$[a(\mathbf{k}_2); a(\mathbf{k}_1)^\dagger] = \epsilon(k_2^2)(2\pi)^4\delta^4(\mathbf{k}_2-\mathbf{k}_1). \tag{4.32}$$

Because of the appearance of the factor $\epsilon(k_2^2)$ in Eq. (4.32) it is clear that it is $a(\mathbf{k})$ and not $a(\mathbf{k})^\dagger$ which creates spacelike states. Accordingly we change our notation to reflect this, defining $\mathscr{b}(\mathbf{k})\equiv a(\mathbf{k})^\dagger$ for $k^2<0$. Then the above equations read

$$\phi = \int_{k^2>0}\frac{d^4k}{(2\pi)^4(|k^2|)^{1/2}}\{a(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}+ik^2\tau/2} + a(\mathbf{k})^\dagger e^{i\mathbf{k}\cdot\mathbf{x}-ik^2\tau/2}\} + \int_{k^2<0}\frac{d^4k}{(2\pi)^4(|k^2|)^{1/2}}\{\mathscr{b}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}-ik^2\tau/2} + \mathscr{b}(\mathbf{k})^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}+ik^2\tau/2}\}, \tag{4.33}$$

and

$$[a(\mathbf{k}_2);a(\mathbf{k}_1)^\dagger] = (2\pi)^4\delta^4(\mathbf{k}_2-\mathbf{k}_1), \quad k_2^2>0,$$
$$[\mathscr{b}(\mathbf{k}_2);\mathscr{b}(\mathbf{k}_1)^\dagger] = (2\pi)^4\delta^4(\mathbf{k}_2-\mathbf{k}_1), \quad k_2^2<0. \tag{4.34}$$

By using these representations we can transform the Hamiltonian (4.26) into momentum space

$$H = \int_{k^2>0}\frac{d^4k}{(2\pi)^4}a(\mathbf{k})^\dagger a(\mathbf{k})(-k^2/2) + \int_{k^2<0}\frac{d^4k}{(2\pi)^4}\mathscr{b}(\mathbf{k})^\dagger\mathscr{b}(\mathbf{k})(-|k^2|/2), \tag{4.35}$$

Note the negative definiteness of the Hamiltonian. If we had arrived at Eq. (4.32) with an anticommutator instead of a commutator on the left-hand side of the equation, there would be no way of maintaining the interpretation of the Fourier amplitudes of spacelike states as creation and annihilation operators; there is thus no consistent way of quantizing the second-order Fock equation with Fermi–Dirac statistics. This justifies the above decision to quantize using Bose–Einstein statistics.

The propagation function for our second-order Fock equation is defined in analogy to Eq. (4.14). Since the calculation is quite standard, we omit the details. The end result of the calculation is again most simply expressed in momentum space. Defining $\langle 0|T(\phi(2)\phi(1))|0\rangle \equiv iD_F(21)$, and

$$iD_F(21)\equiv\int\frac{d^4k}{(2\pi)^4}e^{-i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}iD_F(\mathbf{k},\tau_2-\tau_1), \tag{4.36}$$

we find

$$iD_F(\mathbf{k},\tau_2-\tau_1) = \frac{e^{i|k^2|(\tau_2-\tau_1)/2}}{|k^2|}. \tag{4.37}$$

A representation free of absolute value signs is

$$iD_F(\mathbf{k},\tau_2-\tau_1) = (1/2\pi i)\int_{-\infty}^{\infty}(i/2)\,d\lambda\ e^{i\lambda k^2/2} \times\ln([\lambda(1-i\epsilon)]^2-[\tau_2-\tau_1]^2), \tag{4.38}$$

and a pure Fourier representation is

$$iD_F(\mathbf{k},N) = \frac{2i}{(k^2)^2-N^2+i\epsilon}, \tag{4.39}$$

where $iD_F(\mathbf{k},N)$ is defined through the relation

$$iD_F(\mathbf{k},\tau_2-\tau_1)\equiv\int_{-\infty}^{\infty}\frac{dN}{2\pi}e^{iN(\tau_2-\tau_1)/2}\,iD_F(\mathbf{k},N). \tag{4.40}$$

In Equation (4.38) the branch of the multiple valued function is specified by requiring $\arg(\lambda(1-i\epsilon)\pm(\tau_2-\tau_1))\to 0$ as $\lambda\to+\infty$, where the arc signifies the phase.

Equation (4.39) can be rearranged as follows ($i\epsilon$ terms have been dropped except as required to define the singularities)

$$iD_F(\mathbf{k},N) = \frac{i}{|N|}\left\{\frac{1}{k^2-|N|(1-i\epsilon)} - \frac{1}{k^2+|N|(1-i\epsilon)}\right\}. \tag{4.41}$$

The first and second terms here describe the propagation of timelike and spacelike states, respectively. The fact that the term in Eq. (4.41) for timelike states is proportional to the simple propagator (4.19) for the first-order Fock equation indicates that perhaps we have not strayed very far after all from customary conceptions in our decision to take the second-order Fock equation seriously. The second term in Eq. (4.41) has the form of a Pauli–Villars regulator. The propagator for the timelike states is thus regularized through the effect of the space-like states.

[1] R. P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971).
[2] A still earlier application of mass Hamiltonian is due to K. Fujimura, T. Kobayashi, and M. Namiki, Prog. Theor. Phys. (Kyoto) **44**, 193 (1970).
[3] V. A. Fock, Phys. Z. Sowjetunion **12**, 404 (1937).
[4] Yoichiro Nambu, Prog. Theor. Phys. **5**, 82 (1950).
[5] Yasuhisa Katayama, Katuro Sawada, and Shuji Takagi, Prog. Theor. Phys. **5**, 14 (1950).
[6] John R. Fanchi, Phys. Rev. D **20**, 3108 (1979). See also L. P. Horwitz and C. Piron, Helv. Phys. Acta **46**, 317 (1973).
[7] Daniel M. Greenberger, J. Math. Phys. **15**, 395 (1974); **15**, 406 (1974).
[8] Feynman discussed the Fock equations without a mass parameter in an appendix to his article which appeared in Phys. Rev. **80**, 440 (1950).
[9] W. Pauli and F. Villars, Rev. Mod. Phys. **21**, 434 (1949).
[10] We use Heaviside–Lorentz (rationalized Gaussian) units with $\hbar = c = 1$. Our spacetime metric is such that we distinguish between contravariant and covariant quantities and use the signature $g^{00} = -g^{11} = -g^{22} = -g^{33} = +1$. We use the notation $\partial_\mu \equiv \partial/\partial x^\mu$, and $\partial^\mu \equiv g^{\mu\nu}\partial_\nu$.
[11] Herbert Goldstein, *Classical Mechanics* (Addision–Wesley, Reading, Massachusetts, 1965).
[12] Julian Schwinger, Phys. Rev. **82**, 914 (1951); and Phys. Rev. **91**, 713 (1953).
[13] See for example, James D. Bjorken and Sidney D. Drell, *Relativistic Quantum Fields* (McGraw–Hill, New York, 1965).
[14] This connection between the two types of boundary conditions was first noticed by Nambu (Ref. 4) who was working with the original Fock equation containing a mass parameter.
[15] There is a 5–current conservation law analogous to Eq. (3.18) for the second-order Fock equation. Assuming a complex wave function and an external electromagnetic field, this conservation law is:

$$\frac{\partial\rho}{\partial\tau} + \partial_\mu J^\mu = 0, \text{ where}$$
$$\rho = \phi^*\frac{1}{i}\dot{\phi} - \phi\frac{1}{i}\dot{\phi}^*$$

$$\mathbf{J} = \tfrac{1}{4}\phi^{+}(\mathbf{\Pi} + \overleftarrow{\mathbf{\Pi}}^{*})\Pi^{2}\phi + \tfrac{1}{4}\phi^{+}(\overleftarrow{\Pi}^{*})^{2}(\mathbf{\Pi} + \overleftarrow{\mathbf{\Pi}}^{*})\phi,$$

$\mathbf{\Pi} \equiv i\Box - q\mathbf{\Omega}$. (The form of the wave equation is now $\left\{\left(\frac{\Pi^{2}}{2}\right)^{2} + \frac{\partial^{2}}{\partial\tau^{2}}\right\}\phi = 0$.)

Because of this conservation law the integral $\int d^{4}x\,\rho$ is independent of evolution time and provides a kind of inner product for complex solutions of the second-order Fock equation. The form of the integral in Eq. (4.31) was chosen with this fact in mind.

On computing eigenvalues in radiative transfer

C. E. Siewert
Nuclear Engineering Department, North Carolina State University, Raleigh, North Carolina 27650

(Received 18 July 1979; accepted for publication 15 November 1979)

The Wiener–Hopf factorization of the dispersion function is used to deduce explicit expressions for the discrete eigenvalues in the theory of radiative transfer.

I. INTRODUCTION

One of the first tasks encountered in exact[1,2] or, in some cases,[3,4] approximate analysis of the basic equation of radiative transfer[5]

$$\mu \frac{\partial}{\partial \tau} I(\tau,\mu) + I(\tau,\mu) = \frac{\omega}{2} \sum_{l=0}^{L} (2l+1) f_l P_l(\mu) \int_{-1}^{1} P_l(\mu') I(\tau,\mu') \, d\mu', \quad (1)$$

is that of computing the discrete eigenvalues. Thus, if we substitute

$$I_\nu(\tau,\mu) = \phi(\nu,\mu) e^{-\tau/\nu}, \quad \nu \notin [-1,1], \quad (2)$$

into Eq. (1), then the required eigenvalues $\pm \nu_{\alpha-1}$, $\alpha = 1,2,3,\ldots,\kappa$, can readily be seen to be the zeros of the dispersion function

$$\Lambda(z) = 1 + z \int_{-1}^{1} \psi(x) \frac{dx}{x-z}, \quad (3)$$

where the characteristic function[5] is

$$\psi(\mu) = \frac{\omega}{2} \sum_{l=0}^{L} (2l+1) f_l g_l(\mu) P_l(\mu). \quad (4)$$

Here ω is the albedo for single scattering, and the f_l (with $f_0 = 1$) are the coefficients in a Legendre expansion of the phase function. In addition, the polynomials $g_l(\mu)$, of order l, are those introduced by Chandrasekhar,[5] i.e.,

$$g_0(\nu) = 1, \quad (5a)$$

$$g_1(\nu) = h_0 \nu, \quad (5b)$$

$$g_2(\nu) = \tfrac{1}{2}(h_0 h_1 \nu^2 - 1), \quad (5c)$$

and, in general,

$$(l+1) g_{l+1}(\nu) = \nu h_l g_l(\nu) - l g_{l-1}(\nu), \quad (6)$$

with

$$h_l = (2l+1)(1 - \omega f_l). \quad (7)$$

It is apparent that $\Lambda(z)$ is analytic in the complex plane cut from -1 to 1 along the real axis and that we can use the argument principle[6] to compute the number κ of $\pm$ pairs of zeros of $\Lambda(z)$ in the cut plane. We note, however, that although the use of the argument principle is conceptually straightforward, the calculation can pose problems for specific cases.

Before proceeding to establish some explicit expressions for the zeros of $\Lambda(z)$, we can integrate Eq. (3) to find

$$\Lambda(z) = 1 + z\psi(z) \log\left(\frac{z-1}{z+1}\right) + \omega z \sum_{l=1}^{L} (2l+1) f_l g_l(z) \Gamma_l(z), \quad (8)$$

where the polynomials $\Gamma_l(z)$ can be generated from

$$(2l+1) z\Gamma_l(z) = -\delta_{l,0} + (l+1) \Gamma_{l+1}(z) + l\Gamma_{l-1}(z), \quad (9)$$

with

$$\Gamma_0(z) = 0, \quad (10a)$$

$$\Gamma_1(z) = 1, \quad (10b)$$

and

$$\Gamma_2(z) = \tfrac{3}{2} z. \quad (10c)$$

II. ANALYSIS

In order to establish some useful expressions concerning $\Lambda(z)$ we first let z tend to infinity in Eq. (3) to find

$$\Lambda(\infty) = 1 - \omega \sum_{l=0}^{L} f_l W_l, \quad (11)$$

where

$$W_l = \left(\frac{2l+1}{2}\right) \int_{-1}^{1} g_l(\mu) P_l(\mu) \, d\mu. \quad (12)$$

We can use the recursive relation

$$(2l+1) \mu P_l(\mu) = (l+1) P_{l+1}(\mu) + l P_{l-1}(\mu), \quad (13)$$

and Eq. (6) to deduce from Eq. (12) that

$$(2l+1) W_{l+1} = h_l W_l, \quad (14a)$$

with

$$W_0 = 1, \quad (14b)$$

and therefore we can write

$$\Lambda(\infty) = \prod_{l=0}^{L} (1 - \omega f_l). \quad (15)$$

We can also deduce from Eq. (3) that

$$\Lambda(z) \to \Lambda(\infty) + \frac{a_2}{z^2} + \frac{a_4}{z^4} + \cdots, \quad \text{as } z \to \infty, \quad (16)$$

where

$$a_2 = -\omega \sum_{l=0}^{L} f_l \left(\frac{2l+1}{2}\right) \int_{-1}^{1} \mu^2 g_l(\mu) P_l(\mu) \, d\mu = -\omega \sum_{l=0}^{L} f_l B_l, \quad (17a)$$

and

$$a_4 = -\omega \sum_{l=0}^{L} f_l \left(\frac{2l+1}{2}\right) \int_{-1}^{1} \mu^4 g_l(\mu) P_l(\mu) \, d\mu = -\omega \sum_{l=0}^{L} f_l C_l. \quad (17b)$$

If we now use Eqs. (6) and (13), we find from Eq. (17a) a convenient way to compute the coefficients B_l, viz.,

 0022-2488/80/092468-03$1.00

$$(2l+1)\ B_{l+1} = h_l B_l + \frac{(l+2)^2}{(2l+5)(2l+3)} h_l W_l - \frac{l^2}{2l-1} W_{l-1}, \quad l \geqslant 0, \tag{18}$$

with

$$B_0 = \tfrac{1}{3}, \tag{19a}$$

$$B_1 = \tfrac{3}{5} h_0, \tag{19b}$$

and

$$B_2 = \tfrac{1}{3}\left(\tfrac{6}{7} h_0 h_1 - 1\right). \tag{19c}$$

In a similar manner we find we can express the coefficients C_l as

$$(2l+1)\ C_{l+1} = h_l C_l + \frac{(l+2)^2}{(2l+5)(2l+3)} h_l T_l - \frac{l^2}{2l-1} T_{l-1}, \quad l \geqslant 0, \tag{20a}$$

where

$$T_l = B_l + \frac{1}{2l+5}\left[\frac{(l+3)^2}{2l+7} + \frac{(l+2)^2}{2l+3}\right] W_l, \tag{20b}$$

with

$$C_0 = \tfrac{1}{5}, \tag{21a}$$

$$C_1 = \tfrac{3}{7} h_0, \tag{21b}$$

and

$$C_2 = \tfrac{2}{7}\left(\tfrac{5}{6} h_0 h_1 - 1\right). \tag{21c}$$

The work of Muskhelishvili[7] can now be used to establish a Wiener–Hopf factorization of $\Lambda(z)$, i.e.,

$$\Lambda(z) = \Lambda(\infty) X(z) X(-z) \prod_{\alpha=1}^{\kappa} (\nu_{\alpha-1}^2 - z^2), \tag{22}$$

where

$$X(z) = \frac{1}{(1-z)^{\kappa}} \exp\left[\frac{1}{\pi}\int_0^1 \Theta(t)\frac{dt}{t-z}\right]. \tag{23}$$

Here

$$\Theta(t) = \arg[\lambda(t) + i\pi t\psi(t)], \tag{24a}$$

or

$$\Theta(t) = \tan^{-1}\left[\frac{\pi t\psi(t)}{\lambda(t)}\right], \tag{24b}$$

with

$$\lambda(t) = 1 + t\psi(t)\ln\left(\frac{1-t}{1+t}\right) + \omega t \sum_{l=1}^{L} (2l+1)\, f_l\, g_l(t)\, \Gamma_l(t), \tag{25}$$

is a continuous function of t, with $\Theta(0) = 0$ and $\Theta(1) = \kappa\pi$.

Considering first the case $\kappa = 1$, we can solve Eq. (22) to find

$$\nu_0^2 = z^2 + \Lambda(z)[\Lambda(\infty) X(z) X(-z)]^{-1}, \quad \kappa = 1, \tag{26}$$

which clearly is an explicit expression for ν_0^2 for any value of z. We can find two particularly concise expressions for ν_0^2 by setting $z = 0$ or by letting $z \to \infty$ in Eq. (26):

$$\nu_0^2 = \frac{1}{\Lambda(\infty)} \exp\left[-\frac{2}{\pi}\int_0^1 \Theta(t)\frac{dt}{t}\right], \quad \kappa = 1, \tag{27}$$

and

$$\nu_0^2 = 1 - \frac{2}{\pi}\int_0^1 t\Theta(t)\, dt + \frac{\omega}{\Lambda(\infty)} \sum_{l=0}^{L} f_l B_l, \quad \kappa = 1. \tag{28}$$

For the case $\kappa = 2$, two different values of z can be used in

$$(\nu_0^2 - z^2)(\nu_1^2 - z^2) = \Lambda(z)[\Lambda(\infty) X(z) X(-z)]^{-1}, \quad \kappa = 2, \tag{29}$$

to yield two equations that can be solved simultaneously to yield ν_0^2 and ν_1^2. We use $z = 0$ and let $z \to \infty$ in Eq. (29) to find

$$\nu_0^2 = A + (A^2 - B)^{1/2}, \quad \kappa = 2, \tag{30a}$$

and

$$\nu_1^2 = A - (A^2 - B)^{1/2}, \quad \kappa = 2, \tag{30b}$$

where

$$A = 1 - \frac{1}{\pi}\int_0^1 t\Theta(t)\, dt + \frac{1}{2}\frac{\omega}{\Lambda(\infty)} \sum_{l=0}^{L} f_l B_l, \tag{31a}$$

and

$$B = \frac{1}{\Lambda(\infty)} \exp\left[-\frac{2}{\pi}\int_0^1 \Theta(t)\frac{dt}{t}\right]. \tag{31b}$$

Finally, for the case $\kappa = 3$, we deduce from Eq. (22) the three equations

$$\nu_0^2\nu_1^2\nu_2^2 = \frac{1}{\Lambda(\infty)} \exp\left[-\frac{2}{\pi}\int_0^1 \Theta(t)\frac{dt}{t}\right], \quad \kappa = 3, \tag{32a}$$

$$\nu_0^2 + \nu_1^2 + \nu_2^2 = 3 - \Theta_1 + \frac{\omega}{\Lambda(\infty)} \sum_{l=0}^{L} f_l B_l, \quad \kappa = 3, \tag{32b}$$

and

$$\nu_0^2\nu_1^2 + \nu_0^2\nu_2^2 + \nu_1^2\nu_2^2 = 3(1-\Theta_1) + \Theta_3 + \frac{1}{2}\Theta_1^2 + (3-\Theta_1)\frac{\omega}{\Lambda(\infty)} \times \sum_{l=0}^{L} f_l B_l - \frac{\omega}{\Lambda(\infty)} \sum_{l=0}^{L} f_l C_l, \quad \kappa = 3, \tag{32c}$$

TABLE I. Numerical results.

ω	ν_0		ν_1		ν_2	
	Explicit	Refined	Explicit	Refined	Explicit	Refined
0.1	1.030046	1.030042				
0.5	1.536814	1.536814	1.054992	1.054987		
0.95	7.480699	7.480699	1.666787	1.666787	1.019564	1.019586

where

$$\Theta_\alpha = \frac{2}{\pi}\int_0^1 t^\alpha \Theta(t)\,dt. \tag{33}$$

Equations (32) clearly can be solved simultaneously to yield ν_0^2, ν_1^2, and ν_2^2.

It is apparent that the foregoing procedure can be used to reduce the task of finding the zeros of $\Lambda(z)$ to one of solving a polynomial equation of order κ in the square of the desired eigenvalues. This procedure has the additional merit that it provides results for each of the desired solutions. Such results can, if necessary , be refined by using them as first approximations in an iterative solution. In order to demonstrate the accuracy of the explicit solutions given by Eqs. (27), (30), and (32) for the cases $\kappa = 1$, 2, and 3, we quote in Table I the results of Maiorino[8] who used the 80 point quadrature scheme of Gauss to compute the integrals required to evaluate the explicit solutions and a Newton–Raphson scheme to compute the refined results. The results in Table I are based on the $L = 20$ scattering law defined by[9]

$$(2l+1)\,f_l^L = \left(\frac{L+1}{2L}\right)[lf_{l-1}^{L-1} + (2l+1)\,f_l^{L-1} + (l+1)\,f_{l+1}^{L-1}], \tag{34}$$

with $f_0^L = 1$ and $f_l^L = 0$, if $l > L$.

ACKNOWLEDGMENTS

The author is grateful to J.R. Maiorino for communicating the numerical results shown in Table I. This work was supported in part by National Science Foundation grant ENG. 7709405.

[1]K.M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[2]J.R. Mika, Nucl. Sci. Eng. **11**, 415 (1961).
[3]C.E. Siewert and P. Benoist, Nucl. Sci. Eng. **69**, 156 (1979).
[4]C.E. Siewert, Astrophy. Space Sci. **58**, 131 (1978).
[5]S. Chandrasekhar, *Radiative Transfer* (Oxford University, London, 1950).
[6]L.V. Ahlfors, *Complex Analysis* (McGraw-Hill, New York, 1953).
[7]N.I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).
[8]J.R. Maiorino (private communication, 1979).
[9]H.G. Kaper, J.K. Shultis, and J.G. Veninga, J. Comput. Phys. **6**, 288 (1970).

ERRATA

Erratum: Spectral and scattering theory for the adiabatic oscillator and related potentials[J. Math. Phys. 20, 594 (1979)]

Matania Ben-Artzi[a] and Allen Devinatz
Department of Mathematics, Northwestern University, Evanston, Illinois 60201

(Received 4 March 1980; accepted for publication 14 March 1980)

In Lemma 3.1, when $0<\alpha<1$ the asymptotic formulas

(b) $\phi_+(r,z)\sim e^{i\sqrt{z}r}$, $\phi'_+(r,z)\sim i\sqrt{z}e^{i\sqrt{z}r}$, $r\to\infty$,

will hold only in the range $\alpha+\beta>1$ and $\beta>1/2$. In the region $\alpha<1$, $\alpha+\beta\leqslant 1$ or $\beta\leqslant 1/2$, $2\beta-\alpha>0$, the asymptotic formulas (b) must be replaced by

(b') $\phi_+(r,z)\sim e^{i\int_{r_0}^{r}\sqrt{z-V_L}}$,

$\phi'_+(r,z)\sim i\sqrt{z-V_L(r)}\,e^{i\int_{r_0}^{r}\sqrt{z-V_L}}$, $r\to\infty$.

The reason is that if we write

$$i\int_{r_0}^{r}\sqrt{z-V_L}=izr+R(r,z),$$

then $R(r,z)$ converges to a finite number as $r\to\infty$ only if $\alpha+\beta>1$ and $\beta>1/2$.

All of the results in Sec. 4 of the paper, where Lemma 3.1 is used, remain valid with the replacement (b'). The proofs require only minor adjustments. The existence of the wave operators of Theorem 5.1, cases (iii) and (iv), are no longer valid for $\alpha<1$, since the proof of existence requires the asymptotics given by (b). A corrected (and in some intances, improved) set of α,β values for which the wave operators exist is given in Ref. 1. The wave operators are asymptotically complete for the set of α,β values for which they exist.

[a]Present address: Institute of Mathematics, Hebrew University, Jerusalem, Israel.

[1]A. Devinatz, J. Math. Phys. **21**, 2406 (1980).

 0022-2488/80/092471-01$0.00