Analytic SU(3) states in a finite subgroup basis[a)]

P. E. Desmier and R. T. Sharp
Department of Physics, McGill University, Montreal, Quebec, Canada

J. Patera
Centre de Recherches de Mathématiques Appliquées, Université de Montréal, Montréal, Québec, Canada

(Received 10 February 1982; accepted for publication 26 March 1982)

A method is given for deriving branching rules, in the form of generating functions, for the decomposition of representations of SU(3) into representations of its finite subgroups. Interpreted in terms of an integrity basis, the generating functions define analytic polynomial basis states for SU(3), which are adapted to the finite subgroup.

PACS numbers: 02.20. + b, 02.30. + g

1. INTRODUCTION

Apart from the (double) point groups and permutation groups, finite groups have found relatively little application in physics. Some years ago, Fairbairn, Fulton, and Klink[1] considered the finite subgroups of SU(3)/C (of the eightfold way) as possible particle symmetry groups, with largely negative conclusions.

Recently there has been a growth in the interest taken by physicists in finite groups. If one restricts one's attention to representations of a space group G whose k vectors are rational with highest common denominator q, then one is dealing[2] with a finite group G/G_q, where G_q is the space group whose elementary displacements are q times larger than those of G. Finite groups have recently been used in conjunction with local gauge symmetry in the flavor sector of electroweak interactions.[3] A class of finite groups has been shown to be symmetries of general spin systems.[4] It has been suggested than an approximate integration over the manifold of color SU(3) transformations can be effected by summing over the transformations of a finite subgroup of SU(3). Some of these developments are discussed more fully by Abresch *et al.*[5]

In this paper it is shown how branching rules, in the form of generating functions, may be derived for SU(3)$\supset G$, where G is a finite subgroup of SU(3). The generating functions define an integrity basis, a finite number of subgroup multiplets, in terms of which general analytic SU(3)$\supset G$ basis states may be expressed. Generating functions for finite group tensors which are polynomials in the components of a given tensor are determined without the use of partial Molien functions; a formula is given for the sum of the numerator coefficients in such a generating function.

2. GENERALITIES

Reference 6 contains a prescription for the derivation of generating functions for branching rules from a compact Lie group to a finite group, starting with the character generator of the Lie group. Here we follow an alternative method which, at least for SU(3), is simpler to implement.

The SU(3) character generator,[6,7] with its two terms put over a common denominator, is

$$X(P,Q;\eta,\xi) = \frac{1-PQ}{(1-P\eta\xi)(1-P\eta^{-1}\xi)(1-P\xi^{-2})(1-Q\eta^{-1}\xi^{-1})(1-Q\eta\xi^{-1})(1-Q\xi^{2})}. \tag{2.1}$$

Now the coefficient of P^p in the expansion of $[(1-P\eta\xi)(1-P\eta^{-1}\xi)(1-P\xi^{-2})]^{-1}$ is the character of the SU(3) representation $(p,0)$; decomposed into characters of a finite subgroup G, it yields those subgroup multiplets which are polynomials of degree p in the multiplet 3 which spans (1,0) of SU(3). We are led to consider the generating function $B_{i3}(P)$, whose expansion

$$B_{i3}(P) = \sum_p C_p P^p \tag{2.2}$$

provides, as the coefficient C_p, the number of linearly independent G tensors transforming by the irreducible representation i, which are polynomials of degree p in the components of the tensor 3 contained in the (1,0) representations of SU(3). It is known[8–10] tht the function $B_{i3}(P)$ has the form

$$B_{i3}(P) = \sum_k n_k^{(i)} P^k \left[\prod_{h=1}^{3}(1-P^{d_h})\right]^{-1}; \tag{2.3}$$

the denominator factors correspond to three functionally independent scalars of degrees d_h, while the terms in the finite sum in the numerator correspond to tensors transforming by i, in number $n_k^{(i)}$ of degree k, which are linearly independent when their coefficients belong to the ring of denominator scalars. Some general properties of the $n_k^{(i)}$ are pointed out at the end of this section. Since (0,1) is conjugate to (1,0), it follows that

$$B_{i\bar{3}}(Q) = B_{\bar{i}3}(Q), \tag{2.4}$$

where $\bar{3}$ is the subgroup representation contained in (0,1) and $\bar{i}$ is the representation conjugate to i. Combining the representations found in $(p,0)$ with those contained in $(0,q)$, we

[a)] Supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

obtain

$$F_m(P,Q) = (1-PQ)\sum_{ii'} C_{ii'}^m B_{i3}(P) B_{i'\bar{3}}(Q)$$

$$= (1-PQ)\sum_{ii'} C_{ii'}^m B_{i3}(P) B_{\bar{i}'3}(Q), \qquad (2.5)$$

where $C_{ii'}^m$ is the multiplicity of the irreducible representation m in the direct product of representations i and i'. Then $F_m(P,Q)$ is the desired generating function. When expanded,

$$F_m(P,Q) = \sum_{p,q} r_{pq}^m P^p Q^q, \qquad (2.6)$$

it provides, as the coefficient r_{pq}^m, the multiplicity of the finite group representation m in the SU(3) representation (p,q).

To render $F_m(P,Q)$ into a "positive" form, the numerator must be written as a sum of terms, each containing as a factor one of the denominator factors $(1-P^{d_h})$, $(1-Q^{d_h})$. Then $F_m(P,Q)$ is a sum of terms each with five denominator factors [$5 = \frac{1}{2}(l+r)$, in agreement with Racah's[11] counting of labels] and a numerator which is a polynomial in P, Q with positive coefficients. In this form $F_m(P,Q)$ can be interpreted directly in terms of an integrity basis. Let η_1, η_2, η_3 be the (1,0) states, $\bar{\eta}_1, \bar{\eta}_2, \bar{\eta}_3$ the (0,1) states. Then the integrity basis consists of six denominator scalars, three of degrees d_1, d_2, d_3 in η_1, η_2, η_3 and three of degrees d_1, d_2, d_3 in $\bar{\eta}_1, \bar{\eta}_2, \bar{\eta}_3$. A numerator term P^pQ^q in $F_m(P,Q)$ represents an m tensor of degree p in η_1, η_2, η_3 and degree q in $\bar{\eta}_1, \bar{\eta}_2, \bar{\eta}_3$. With the knowledge of their degrees and transformation properties it is straightforward to determine the algebraic form of the elements of the integrity basis. An SU(3)$\supset G$ basis state corresponds to a numerator tensor, multiplied by a product of powers of the denominator scalars from the same term of $F_m(P,Q)$. To ensure that the states have the correct SU(3) transformation properties, i.e., belong to the representation (p,q) where p and q are the degrees in the unbarred and barred variables, one has two options. The first is to replace η_i by $\eta_i^{(1)}$ and $\bar{\eta}_i$ by $\epsilon_{ijk}\eta_j^{(1)}\eta_k^{(2)}$; this is called the two-particle scheme since it is based on two (1,0) representations. The second is to make the replacements

$$\eta_i \rightarrow a_i = \eta_i - B(N+3)^{-1}\partial_{\bar{\eta}_i},$$
$$\bar{\eta}_i \rightarrow a_i = \bar{\eta}_i - B(N+3)^{-1}\partial_{\eta_i}. \qquad (2.7)$$

Here $B = \Sigma_i \eta_i \bar{\eta}_i$ is an SU(3) scalar and $N = \Sigma_i[\eta_i\partial_{\eta_i} + \bar{\eta}_i\partial_{\bar{\eta}_i}]$ is the total degree; this is called the particle–antiparticle scheme since it is based on a (1,0) and a (0,1) representation. The proof that polynomials in a_i and $\bar{a}_j$ are traceless, i.e., orthogonal to any state containing B as a factor, is similar to that for the "traceless boson operators" of Lohe and Hurst.[12]

We now outline a procedure for determining the generating functions $B_{i3}(P)$, without the use of partial Molien functions. The polynomial G tensors of degree p for those contained in the symmetric plethysm $[p]$ (the Young tableau for $[p]$ is a single row of p boxes; a single box refers to the 3-dimensional representation 3 of G which spans (1,0) of SU(3)). Using the direct product formulas

$$[1]\otimes[p-1] = [p] + [p-1,1],$$
$$[1^2]\otimes[p-2] = [p-1,1] + [p-3],$$

we arrive at a recursion formula

$$[p] = [1]\otimes[p-1] - [1^2]\otimes[p-2] + [p-3], \qquad (2.8)$$

which determines the plethysm $[p]$ in terms of $[p-1]$, $[p-2]$, $[p-3]$. To carry out the iteration, all one needs are the Clebsch–Gordan series which involve $[1]$ or $[1^2]$, i.e., 3 or $\bar{3}$. After computing a number of $[p]$ [the interation of (2.8) is easily done with a computer] one notices that for some integer d_1, the representations contained in $[p]$ are also contained in $[p+d_1]$ for all $[p]$. This suggests a denominator scalar of degree d_1; to eliminate states containing it as a factor, replace $[p]$ by $[p]-[p-d_1]$. When this subtraction has been repeated twice more with the integers d_2 and d_3, the thrice-subtracted plethysm,

$$[p]-[p-d_1]-[p-d_2]-[p-d_3]+[p-d_1-d_2]$$
$$+[p-d_1-d_3]+[p-d_2-d_3] \qquad (2.9)$$
$$-[p-d_1-d_2-d_3],$$

will be found to vanish for $p > d_1+d_2+d_3-3$. The multiplicity of an i tensor in the subtracted plethysm (2.9) is the coefficient of P^p in the numerator of the generating function $B_{i3}(P)$.

We complete this section by noting two properties of the generating functions $B_{ij}(\lambda)$; they are straightforward generalizations of results given by Stanley[13] for Molien functions (i the scalar representation). The dimension t_j of the representation j is not now restricted to the value 3. In terms of the partial Molien functions $B_{ij}(\lambda)$ may be written[8–10,14]

$$B_{ij}(\lambda) = \frac{1}{N}\sum_s \left[N_s \chi_{i,s}^* \Big/ \prod_{h=1}^{t_j}(1-\lambda\alpha_{j,s}^{(h)})\right], \qquad (2.10)$$

where N is the order of the group, N_s the order of the class s, $\chi_{i,s}$ the character of the class s for the representation i, and $\alpha_{j,s}^{(h)}$ the h th eigenvalue of the matrix which represents an element of the class s for the representation j.

The first property has to do with the sum of the numerator coefficients $n_k^{(i)}$ when $B_{ij}(\lambda)$ is written in the form (2.3). We assume j is a faithful representation of G; that is the case for all the generating functions in this paper. Otherwise we deal with $G_j' = G/G_j$, where G_j is the subgroup of G consisting of elements represented by the identity in the representation j. The representations of G_j' form a subset of those of G. In (2.10) rewrite the term corresponding to the class s as

$$\frac{N_s\chi_{i,s}^*\Pi_h(1-\lambda^{d_h})/N\,\Pi_h(1-\lambda\alpha_{j,s}^{(h)})}{\Pi_h(1-\lambda^{d_h})}. \qquad (2.11)$$

The contribution of the class s to the sum $\Sigma_k n_k^{(i)}$ is the numerator of (2.11) evaluated at $\lambda = 1$. But this vanishes unless s is the identity class, for which $\alpha_{j,s}^{(h)} = 1$. We find

$$\sum_k n_k^{(i)} = t_i\left[\prod_h d_h\right]N^{-1}. \qquad (2.12)$$

The second property has to do with the symmetry of the coefficients $n_k^{(i)}$ about some $k = k_0$. Let us assume that the representation j is unimodular so that $\Pi_h\alpha_{j,s}^{(h)} = 1$; this is the case for all the generating functions in this paper. Replace λ

by λ^{-1} in $B_{ij}(\lambda)$. Then

$$\begin{aligned} B_{ij}(\lambda^{-1}) &= \frac{1}{N}\sum_s \frac{N_s\chi^*_{i,s}}{\Pi_h(1-\lambda^{-1}\alpha^{(h)}_{j,s})} \\ &= \frac{(-\lambda)^{t_j}}{N}\sum_s \frac{N_s\chi^*_{i,s}}{\Pi_h(1-\lambda\alpha^{(h)*}_{j,s})} \qquad (2.13) \\ &= (-\lambda)^{t_j} B_{\bar{i}j}(\lambda). \end{aligned}$$

The last step follows from the reality of $B_{ij}(\lambda)$. Imposing this symmetry on $B_{ij}(\lambda)$ in the form (2.3) leads to the identity

$$n_k^{(i)} = n_{k'}^{(\bar{i})} \qquad (2.14)$$

if

$$k + k' = \sum_h d_h - t_j. \qquad (2.15)$$

$\bar{i}$ is the representation conjugate to i.

3. SPECIFICS

In this section we consider three maximal finite subgroups of SU(3), namely $\Sigma(168)$, $\Sigma(648)$, $\Sigma(1080)$, as well as G(13,3,3), one of the "trihedral" subgroups [the order of G(m,n,3) is mn]. The three Σ groups are discussed and their generator matrices for the representation (1,0) of SU(3) are given in Ref. 1. The trihedral groups are discussed in Ref. 15. For each of these four subgroups of SU(3) we give the generating functions $B_{i3}(P)$, where 3 is the 3-dimensional representation contained in (1,0). The generating functions $F_m(P, Q)$ for general branching rules (and polynomial bases) is given in terms of $B_{i3}(P)$ by Eq. (2.5). We evaluate them in an explicit positive form only for $\Sigma(168)$.

An irreducible representation of $\Sigma(648)$ or of $\Sigma(1080)$ has definite triality, i.e., is found only in SU(3) representations of one triality [the triality of the SU(3) representation (p, q) is $(p-q)$ modulo 3]; a representation of $\Sigma(168)$ or $G(13,3,3)$ may be found in SU(3) representations of any triality.

TABLE I. Coefficients $n_k^{(i)}$ of P^k in the numerator of $B_{i3}(P)$ for $\Sigma(168)$ for $0 \leqslant k \leqslant 10$. The representation i is plotted vertically, the exponent k horizontally.

i \ k	0	1	2	3	4	5	6	7	8	9	10
8					1	1	1	2	1	1	1
7				1		1	1	1	1	1	1
6			1		1		1		1	1	1
$\bar{3}$				1		1					1
3		1							1	1	
1	1										

(a) The subgroup $\Sigma(168)$

The character table for $\Sigma(168)$ is found in Ref. 1, and is given by Littlewood.[16] The embedding is such that the representation denoted by $\bar{\Sigma}_3$ in Ref. 1 is contained in (1,0) of SU(3). We label the representations by their dimension: 1,3,$\bar{3}$,6,7,8. Our 3 is the $\bar{\Sigma}_3$ of Ref. 1. Representations 1,6,7,8 are self-conjugate; 3 and $\bar{3}$ are mutually conjugate. From the character table it is easy to determine Clebsch–Gordan series, in particular those involving 3 or $\bar{3}$. The iteration of (2.8) can then be carried out and the subtraction (2.9) implemented. The degrees d_1,d_2,d_3 of the denominator scalars turn out to be 4,6,14, so that the generating functions $B_{i3}(P)$ have the form $\Sigma_k n_k^{(i)} P^k/(1-P^4)(1-P^6)(1-P^{14})$. The numerator coefficients $n_k^{(i)}$ are given in Table I. Because of the symmetry (2.14), $n_k^{(i)} = n_{21-k}^{(\bar{i})}$, it is necessary to tabulate $n_k^{(i)}$ for $0 \leqslant k \leqslant 10$ only.

The generating function $F_m(P, Q)$ for general branching rules takes the form

$$\begin{aligned} F_m(P,Q) = \frac{1}{(1-P^{14})(1-Q^{14})}\Bigg[&\frac{N_m^{(1)}(P,Q)}{(1-P^4)(1-P^6)(1-Q^4)} + \frac{N_m^{(2)}(P,Q)}{(1-P^4)(1-P^6)(1-Q^6)} \\ &+ \frac{N_m^{(3)}(P,Q)}{(1-P^4)(1-Q^4)(1-Q^6)} + \frac{N_m^{(4)}(P,Q)}{(1-P^6)(1-Q^4)(1-Q^6)}\Bigg]. \end{aligned} \qquad (3.1)$$

We give a number of symmetry relations which serve to reduce the number and size of Tables II–VI in which are tabulated the numerator polynomials $N_m^{(k)}(P, Q)$ for $m = 1,3,6,7,8$ and for $k = 1,2$.

We have, first,

$$F_{\bar{3}}(P,Q) = F_3(Q,P). \qquad (3.2)$$

For $m = 1,6,7,8$, we have that

$$\begin{aligned} N_m^{(3)}(P,Q) &= N_m^{(1)}(Q,P), \\ N_m^{(4)}(P,Q) &= N_m^{(2)}(Q,P). \end{aligned} \qquad (3.3)$$

If we write

$$N_m^{(k)}(P,Q) = \sum_{p,q} n^{(k)}_{m;p,q} P^p Q^q \qquad (3.4)$$

then, for $m = 1,6,7,8$,

$$\begin{aligned} n^{(1)}_{m;p,q} &= n^{(1)}_{m;22-p,16-q}, \\ n^{(2)}_{m;p,q} &= n^{(2)}_{m;22-p,18-q}. \end{aligned} \qquad (3.5)$$

TABLE II. The coefficients of the polynomials $N_1^{(1)}(P, Q)$ and $N_1^{(2)}(P,Q)$ for $0 \leqslant p \leqslant 11$; p is plotted horizontally, q vertically. (a) The coefficients $n^{(1)}_{1;p,q}$; (b) The coefficients $n^{(2)}_{1;p,q}$.

(a)

q \ p	0	1	2	3	4	5	6	7	8	9	10
15					1		1				
14			1		1	1		1			
13			1	1							1
12			1		1	2	1				
11		1	1						1	1	
10			1		1	1	1				
9					1		1		1		
8		1	1	1	1						
7					1	1	1				
6			1	1		1					
5				1		½					
4			1			1					
3				½							1
2											
1											
0											

(b)

q \ p	0	1	2	3	4	5	6	7	8	9	10
18		1									
17		1				1					
16			1	1							
15				2	1	½					
14		1		1							
13		1		½	1	1	1				
12				1							
11				1		½		1			
10						1					
9		1		½							
8											
7								1			
6											
5											
4	½								½		
3											
2											
1											
0	½								½		

TABLE III. The coefficients of the polynomials $N_3^{(1)}(P,Q)$ and $N^{(2)}_{3;p,q}$; p is plotted horizontally, q vertically. (a) The coefficients $n^{(1)}_{3;p,q}$; (b) The coefficients $n^{(2)}_{3;p,q}$.

(a)

q \ p	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
16	1			1		1		1													
15		2	1		2		2														
14		1	3	2		1		2		2		1								1	
13	1		2	3	4		3														
12	1	1		3	4	3		3		1											
11	1	1			3	2	2		1		1		1		1						
10	1	1	2	3		3	2	3		1											
9	1	1	1	3	2		2	3	1		1		1								
8		1	1	2	3	4		2	3	2		1		1							
7		2	2	2	3	2	3		3	2	1										
6			1	2	3	2	3	2		2	1	2		2							
5	1	1	1	1	2	2	2	2	2		1	2	1		1		1				
4			1	2	2	2	2	2	1	1		2		2							
3	1		1	1	1	2	1	3	1	2	2		2	1	2						
2						2	1	2	2	1	2			1				1			
1				1	1	1	1		1		1	1	1			1					1
0									1	1		1					1				

(b)

q \ p	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
18	1		1																		
17		4													1						
16			6																		
15				8		1			1												
14		1			8		1														
13		2	1	1		8		2	2												
12			2				8		1			1									
11			2	4	1	2		3						1							
10					3		1		4												
9			2			4				4											
8							4				4										
7								1				4	1								
6									2				3		2						
5								1		2				3							
4		1								1	2		2		2		1				
3												1				1	1	1			
2													2		1	1	1				
1														2							
0		1																	1		

TABLE V. The coefficients of the polynomials $N_7^{(1)}(P,Q)$ and $N_7^{(2)}(P,Q)$ for $0 \leqslant p \leqslant 11$; p is plotted horizontally, q vertically. (a) The coefficients $n^{(1)}_{7;p,q}$; (b) The coefficients $n^{(2)}_{7;p,q}$.

(a)

q \ p	0	1	2	3	4	5	6	7	8	9	10
16	1	2		2		2					
15	2	4	4		4		2				
14		5	4½	6		6		3			
13	2		6	6	9		7		1		
12	2	4		8	7	9		2			
11	2	4	6		9	4½	3		1		
10	1	3	5	7		4	2	2			
9	1	3	5	5½	4		3	2			
8	1	3	3½	4	5	4		3	1	1	
7	1	2	3	4	5	5	4				
6	1	2	3	4	5	5	2½				
5	1	2	3	4	5	2½					
4		1	2	3	2						
3	1	1	2	1½							
2		1	½								
1											
0											

(b)

q \ p	0	1	2	3	4	5	6	7	8	9	10
18	5		1							1	
17		9	1	1	1						
16	1		13		2						
15		1		16½		2					
14	2		2		15		2	1		1	
13		4		2		13		2			
12			6		2		9½		1		
11				8		2		4			
10					6		1		1		
9	1					4					
8							4				
7								2			
6											
5											
4											
3											
2											
1											
0											

Finally, for $m = 3$, we have

$$n^{(3)}_{3;p,q} = n^{(1)}_{3;22-q,16-p},$$
$$n^{(4)}_{3;p,q} = n^{(2)}_{3;22-q,18-p}. \tag{3.6}$$

Because of (3.2) we do not give $N_5^{(k)}(P,Q)$. On account of (3.3) and (3.6) $N_m^{(3)}(P,Q)$ and $N_m^{(4)}(P,Q)$ are not tabulated. Finally, because of (3.5), for $m = 1,6,7,8$, we give $n^{(1)}_{m;\,p,q}$ and $n^{(2)}_{m;\,p,q}$ only for $0 \leqslant p \leqslant 11$.

The preservation of the symmetry (3.3) entails the appearance of some half-odd-integer coefficients; they may be eliminated by (i) increasing by $\frac{1}{2}$ each half-odd $n^{(k)}_{m;\,p,q}$ for which $k = 1$ or 2 and $q \geqslant p$, or for which $k = 3$ or 4 and $q > p$ while (ii) decreasing by $\frac{1}{2}$ all other half-odd $n^{(k)}_{m;\,p,q}$.

TABLE IV. The coefficients of the polynomials $N_6^{(1)}(P,Q)$ and $N_6^{(2)}(P,Q)$ for $0 \leqslant p \leqslant 11$; p is plotted horizontally, and q vertically. (a) The coefficients $n^{(1)}_{6;p,q}$; (b) The coefficients $n^{(2)}_{6;p,q}$.

(a)

q \ p	0	1	2	3	4	5	6	7	8	9	10	11
16	2	2		2		2						
15	2	3	4		4		1					
14		4	4	6		6		2				
13	1		5	5	8		4		1			
12	2	4		7	5	8		2		1		
11	1	3	4		8	4	3					
10	2	2	5	3		3	1	2		1		
9	1	2	4	5	6		2	2				
8	2	2	3	4	3	3		3	1			
7		2	3	4	5	3	4					
6	1	2	2	4	4	5	2					
5		2	2	3	4	2						
4	1		2		2							
3		1	1	1	3							
2	1		1									
1					1							
0												

(b)

q \ p	0	1	2	3	4	5	6	7	8	9	10	11
18	4											
17		8		1								
16			11		1							
15		1		14	1	2						
14	2		1	1	11		2	1				
13		4		2		10	2	2				
12			5		2	2	7		1			
11				7		2		3				
10					5		1		1			
9						4						
8						2	3					
7						1		1				
6												
5												1
4												
3												
2												
1												
0												

TABLE VI. The coefficients of the polynomials $N_8^{(1)}(P,Q)$ and $N_8^{(2)}(P,Q)$ for $0 \leqslant p \leqslant 11$; p is plotted horizontally, q vertically. (a) The coefficients $n^{(1)}_{8;p,q}$; (b) The coefficients $n^{(2)}_{8;p,q}$.

(a)

q \ p	0	1	2	3	4	5	6	7	8	9	10
16	2	2		3		3					
15	2	5	5		6		3				
14		4	5½	7		7		3			
13	3		7	7	9		8		1		1
12	2	4		8	8	10		2			
11	2	4	7		9	5½	3		2		
10	2	4	6	8		5	3	2			
9	1	4	5	5½	5		4	1	2		
8	1	3	4½	5	5	5		3	1		
7	2	2½	3	5	5	6	5				
6	1	2	4	5	5	6	2½				
5	1	1	3	5	5	3½					
4	1	2	3	4	2						
3		1	2	1½							
2		1	½								1
1		½									
0											

(b)

q \ p	0	1	2	3	4	5	6	7	8	9	10
18	5		1								
17		9½		2							
16	1	1	15		2	1					
15		1		18½		3					
14	3	1	2	1	15		2	2			
13		4½		3		15					
12			7	1	2	1	10½		1		
11				9		2		4½		1	
10					7		1		1		
9						4		1			
8							4				
7								2½			
6											
5											
4											
3											
2											
1											
0											

TABLE VII. Characters of triality 0 and triality 1 representations of $\Sigma(648)$. Triality 2 I.R.'s $3_2, 3_2', 3_2'', 6_2, 6_2', 6_2'', 9_2$ are the conjugates of $3_1, 3_1', 3_1'', 6_1, 6_1', 6_1'', 9_1$, respectively. Where three classes are listed in a single column, the entry is for the first of the three. The second and third are obtained by multiplying by 1,1 for triality 0, by ω,ω^2 for triality 1, and by ω^2,ω for triality 2.

	Order: 1 1 1	12 12 12	12 12 12	54 54 54	36 36 36	36 36 36	9 9 9	24	72	72
	Classes: 1 2 3	4 5 6	7 8 9	10 11 12	13 14 15	16 17 18	19 20 21	22	23	24
Triality 0										
1	1	1	1	1	1	1	1	1	1	1
$1'$	1	ω	ω^2	1	ω	ω^2	1	1	ω	ω^2
$\bar{1}'$	1	ω^2	ω	1	ω^2	ω	1	1	ω^2	ω
2	2	-1	-1	0	1	1	-2	2	-1	-1
$2'$	2	$-\omega$	$-\omega^2$	0	ω	ω^2	-2	2	$-\omega$	$-\omega^2$
$\bar{2}'$	2	$-\omega^2$	$-\omega$	0	ω^2	ω	-2	2	$-\omega^2$	$-\omega$
3	3	0	0	-1	0	0	3	3	0	0
8	8	2	2	0	0	0	0	-1	-1	-1
$8'$	8	2ω	$2\omega^2$	0	0	0	0	-1	$-\omega$	$-\omega^2$
$\bar{8}'$	8	$2\omega^2$	2ω	0	0	0	0	-1	$-\omega^2$	$-\omega$
Triality 1										
3_1	3	$\sqrt{3}e^{\pi i/18}$	$\sqrt{3}e^{35\pi i/18}$	1	$e^{17\pi i/9}$	$e^{\pi i/9}$	-1	0	0	0
$3_1'$	3	$\sqrt{3}e^{13\pi i/18}$	$\sqrt{3}e^{23\pi i/18}$	1	$e^{5\pi i/9}$	$e^{13\pi i/9}$	-1	0	0	0
$3_1''$	3	$\sqrt{3}e^{25\pi i/18}$	$\sqrt{3}e^{11\pi i/18}$	1	$e^{11\pi i/9}$	$e^{7\pi i/9}$	-1	0	0	0
6_1	6	$\sqrt{3}e^{31\pi i/18}$	$\sqrt{3}e^{5\pi i/18}$	0	$e^{5\pi i/9}$	$e^{13\pi i/9}$	2	0	0	0
$6_1'$	6	$\sqrt{3}e^{7\pi i/18}$	$\sqrt{3}e^{29\pi i/18}$	0	$e^{11\pi i/9}$	$e^{7\pi i/9}$	2	0	0	0
$6_1''$	6	$\sqrt{3}e^{19\pi i/18}$	$\sqrt{3}e^{17\pi i/18}$	0	$e^{17\pi i/9}$	$e^{\pi i/9}$	2	0	0	0
9_1	9	0	0	-1	0	0	-3	0	0	0

(b) The subgroup $\Sigma(648)$

Fairbairn, Fulton, and Klink[1] give a character table for those representations of $\Sigma(648)$ with triality 0. We worked out the rest of the table with the help of their generator matrices. There are three representations of dimension 3, three of dimension 6, and one of dimension 9 with triality 1; each has a conjugate representation with triality 2. The characters for representations with triality 0 and 1 are found in Table VII. The generating functions $B_{i3_1}(P)$ have the form $\Sigma_k n_k^{(i)} P^k/(1-P^9)(1-P^{12})(1-P^{18})$. The coefficients $n_k^{(i)}$ are given in Table VIII. Because of the relation $n_k^{(i)} = n_{36-k}^{(\bar{i})}$, it is necessary to tabulate them only for $0 \leqslant k \leqslant 18$. The representation 3_1 is contained in (1,0) of SU(3).

(c) The subgroup $\Sigma(1080)$

The characters of triality 0 representations are given in Ref. 1; McKay[17] provided us with the rest of the character table. It is given in Table IX. There are two representations of dimension 3, one each of dimensions 6,9,15 with triality 1; each has its conjugate representation with triality 2. The generating functions $B_{i3_1}(P)$ are of the form $\Sigma_k n_k^{(i)} P^k/(1-P^6)(1-P^{12})(1-P^{30})$. The coefficients $n_k^{(i)}$ are given in Table X; because of $n_k^{(i)} = n_{45-k}^{(\bar{i})}$, they are given only for $0 \leqslant k \leqslant 22$. The representation 3_1 is contained in (1,0) of SU(3).

(d) The subgroup $G(13,3,3)$

The order of this subgroup is 39. Its characters are given in Table XI; they are obtained from the analytic formulas given in Ref. 15. There are three representations of dimension 1, four of dimension 3. The generating functions $B_{i3}(P)$ are of the form $\Sigma_k n_k^{(i)} P^k/(1-P^3)(1-P^{13})(1-P^{16})$. The coefficients $n_k^{(i)}$, for $0 \leqslant k \leqslant 14$, are given in Table XII. For $15 \leqslant k \leqslant 29$ they are obtained from $n_k^{(i)} = n_{29-k}^{(\bar{i})}$. The representation 3 is contained in (1,0) of SU(3).

4. DISCUSSION

Since analytic basis states are now available, it is possible to develop the Racah algebra (generator and finite transformation matrix elements, Clebsch–Gordan or Wigner coefficients) of SU(3) in a finite subgroup basis.

TABLE VIII. Coefficients $n_k^{(i)}$ of P^k in the numerator of $B_{i3_1}(P)$ for $\Sigma(648)$ and $0 \leqslant k \leqslant 18$. The representation i is plotted vertically, the exponent k horizontally. (a) Triality 0; (b) Triality 1; (c) Triality 2.

(a)

i \ k	0	3	6	9	12	15	18
$8'$		1	1	1	4	3	3
$\bar{8}'$			2	2	2	5	3
8				3	3	3	6
3			1		2		3
$\bar{2}'$		1		1			
$2'$				2		2	
2						3	
$\bar{1}'$			1				
$1'$					2		
1	1						1

(b)

i \ k	1	4	7	10	13	16
9_1			2	2	5	3
$6_1''$		1	1	1	2	4
$6_1'$		1		3	2	2
6_1			1	3	1	5
$3_1''$			2		1	2
$3_1'$		1			3	
3_1	1			1		1

(c)

i \ k	2	5	8	11	14	17
9_2		1	1	4	3	6
$6_2''$		1	1	1	5	1
$6_2'$			3	1	3	3
6_2	1		1	2	2	2
$3_2''$		1		2	1	
$3_2'$		1	1		1	2
3_2			1	2		3

TABLE IX. Characters of triality 0 and triality 1 representations of $\Sigma(1080)$. $Z_1 = (1+\sqrt{5})/2$, $Z_2 = (1-\sqrt{5})/2$. Triality 2 I.R.'s $3_2, 3_2', 6_2, 9_2, 15_2$ are the conjugates of $3_1, 3_1', 6_1, 9_1, 15_1$, respectively. Where three classes are listed in a single column, the entry is for the first of the three. The second and third are obtained by multiplying by 1,1 for triality 0, by ω,ω^2 for triality 1, and by ω^2,ω for triality 2.

	Order: 1 1 1	120	45 45 45	72 72 72	72 72 72	90 90 90	120
	Classes: 1 1 1	2	3 3 3	4 4 4	5 5 5	6 6 6	7
Triality 0							
1	1	1	1	1	1	1	1
5	5	2	1	0	0	-1	-1
$5'$	5	-1	1	0	0	-1	2
8	8	-1	0	Z_1	Z_2	0	-1
$8'$	8	-1	0	Z_2	Z_1	0	-1
9	9	0	1	-1	-1	1	0
10	10	1	-2	0	0	0	1
Triality 1							
3_1	3	0	-1	Z_1	Z_2	1	0
$3_1'$	3	0	-1	Z_2	Z_1	1	0
6_1	6	0	2	1	1	0	0
9_1	9	0	1	-1	-1	1	0
15_1	15	0	-1	0	0	-1	0

TABLE X. Coefficients $n_k^{(i)}$ of P^k in the numerator of $B_{i3_t}(P)$ for $\Sigma(1080)$ and $0 \leqslant k \leqslant 22$. The representation i is plotted vertically, the exponent k horizontally. (a) Triality 0; (b) Triality 1; (c) Triality 2.

(a)

i \ k	0	3	6	9	12	15	18	21
10		1		2	1	2	2	2
9			1	1	2	1	2	2
8′				1	2	2	2	1
8			1	1	1	2	1	2
5′			1		1	1	1	1
5			1		1	1	1	1
1	1							

(b)

i \ k	1	4	7	10	13	16	19	22
15_1			2	2	3	3	3	3
9_1		1		1	2	2	2	2
6_1		1		2		1	1	1
$3'_1$			1		1			1
3_1	1					1	1	

(c)

i \ k	2	5	8	11	14	17	20
15_2		1	1	3	3	3	3
9_2			2	1	2	2	1
6_2	1		1		1	1	2
$3'_2$		1				1	1
3_2		1		1			1

The methods of this paper have already been applied to finite subgroups of SU(2), i.e., the double point groups.[10] The generating function for branching rules from SU(2) to its finite subgroup is just that for polynomial subgroup tensors based on the 2-dimensional representation contained in the spinor representation of SU(2).

The methods can also be applied to higher Lie groups and their finite subgroups.[18] We outline here how the calculation might proceed. The character generator of a Lie group can be written in the form of a fraction. The denominator is a product of factors of the form $1 - A_j \Pi_k \eta_k^{n_{jk}}$, one factor for each state of each fundamental irreducible representation; the variable A_j corresponds to the jth fundamental irreducible representation, η_k to the kth direction in weight space.

TABLE XII. Coefficients $n_k^{(i)}$ of P^k in the numerator of $B_{i3}(P)$ for $G(13,3,3)$ and $0 \leqslant k \leqslant 14$. The representation i is plotted vertically, the exponent k horizontally.

i \ k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\bar{3}'$				1	1	2	1	2	2	2	2	2	3	3	3
$3'$			1	1	1	1	1	2	1	2	2	3	3	3	3
$\bar{3}$			1		2		2	1	2	3	2	3	2	3	3
3		1		1		1	2	1	3	1	3	2	3	3	3
$\bar{1}'$						1		1		1	1	1	1	1	1
$1'$						1		1		1	1	1	1	1	1
1	1					1		1		1	1	1	1		1

The numerator is a polynomial in the A-variables; the coefficient of each product of powers is a sum of products of powers of the η-variables which decomposes into characters (some with negative signs) of irreducible representations of the Lie group. All the parts of the character generator can be interpreted in terms of polynomial subgroup tensors. The denominator factors involving A_j generate tensors based on those contained in the jth fundamental irreducible representation of the Lie group. The subgroup tensors arising from the l fundamental irreducible representations, and those corresponding to the numerator of the character generator, can be coupled consecutively using the Clebsch–Gordan series for the subgroup. In this way we obtain generating functions for branching rules from Lie group to finite subgroup. There remains the problem of separating the numerator to obtain the result in a "positive" form if it is to be interpreted in terms of an integrity basis.

The character generator for SU(3) in the form described above is given by Eq. (2.1). For $O(5)$ it is[6]

$$\frac{(1-B^2)(1+A^2B)-(1-B)AB(1,0)}{(1-A\eta)(1+A\eta^{-1})(1-A\xi)(1-A\xi^{-1})(1-B\eta\xi)(1-B\eta^{-1}\xi^{-1})(1-B\eta\xi^{-1})(1-B\eta^{-1}\xi)(1-B)}, \tag{4.1}$$

where $(1,0) \equiv \eta + \xi + \eta^{-1} + \xi^{-1}$ is the character of the representation (1,0).

For determining the generating functions $B_{ij}(A)$ based on the (perhaps) reducible subgroup representation j contained in one fundamental irreducible representation of the Lie group, one needs a generalization of Eq. (2.8). If the representation j has dimension t, the generalization is

$$[p] = \sum_{k=1}^{t} (-1)^{k-1}[1^k] \otimes [p-k]. \tag{4.2}$$

of course $[1^t]$ is the identity representation.

TABLE XI. Characters of representations of $G(13,3,3)$. $\epsilon = e^{2\pi i/3}$, $\omega = e^{2\pi i/13}$.

Order:	1	13	13	3	3	3	3
Classes:	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
$1'$	1	ε	ε^2	1	1	1	1
$\bar{1}'$	1	ε^2	ε	1	1	1	1
3	3	0	0	$\omega+\omega^3+\omega^9$	$\omega^2+\omega^5+\omega^6$	$\omega^4+\omega^{10}+\omega^{12}$	$\omega^7+\omega^8+\omega^{11}$
$\bar{3}$	3	0	0	$\omega^4+\omega^{10}+\omega^{12}$	$\omega^7+\omega^8+\omega^{11}$	$\omega+\omega^3+\omega^9$	$\omega^2+\omega^5+\omega^6$
$3'$	3	0	0	$\omega^2+\omega^5+\omega^6$	$\omega^4+\omega^{10}+\omega^{12}$	$\omega^7+\omega^8+\omega^{11}$	$\omega+\omega^3+\omega^9$
$\bar{3}'$	3	0	0	$\omega^7+\omega^8+\omega^{11}$	$\omega+\omega^3+\omega^9$	$\omega^2+\omega^5+\omega^6$	$\omega^4+\omega^{10}+\omega^{12}$

ACKNOWLEDGMENTS

We thank John McKay and Richard P. Stanley for helpful conversations.

[1] W. M. Fairbairn, T. Fulton, and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).

[2] D. Phaneuf and R. T. Sharp, in *Lecture Notes in Physics* (Springer, New York, 1980), Vol. 135, p. 517.

[4] M. Marcu and V. Rittenberg, J. Math. Phys. **22**, 2753 (1981).

[5] U. Abresch, A. Bovier, O. Lechtenfeld, M. Lüling, V. Rittenberg, and G. Weymans, J. Math. Phys. **21**, 2481 (1980).

[6] J. Patera and R. T. Sharp, "Generating functions for characters of group representations," in *Lecture Notes in Physics* (Springer, New York, 1979), Vol. 94.

[7] R. P. Stanley, J. Math. Phys. **21**, 2321 (1980).

[8] A. G. McLellan, J. Phys. C **7**, 3326 (1974).

[9] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **19**, 2362 (1978).

[10] P. E. Desmier and R. T. Sharp, J. Math. Phys. **20**, 74 (1979).

[11] G. Racah, Ergeb. Exacten. Naturwissen **37**, 28 (1965).

[12] M. A. Lohe and C. A. Hurst, J. Math. Phys. **12**, 1882 (1971).

[13] R. P. Stanley, Bull. Am. Math. Soc. **1**, 475 (1979).

[14] L. Michel, in *Group theoretical methods in Physics*, edited by R. T. Sharp and B. Kolman (Academic, New York, 1977).

[15] A. Bovier, M. Lüling, and D. Wyler, J. Math. Phys. **22**, 1543 (1981).

[16] D. E. Littlewood, *The Theory of Group Characters* (Clarendon, Oxford, England, 1950).

[17] J. McKay, private communication. The character table for $\Sigma(1080)$ was generated by the program CAYLEY, due to J. Cannon.

[18] P. E. Desmier, J. Patera, and R. T. Sharp, "Finite subgroup bases for compact Lie groups," to appear in Physica A.

Structure and matrix elements of the degenerate series representations of U($p+q$) and U(p,q) in a U(p)$\times$U(q) basis

Anatoli U. Klimyk
Institute for Theoretical Physics, Academy of Sciences of the Ukrainian SSR, Kiev 130, USSR

Bruno Gruber
Physics Department, Southern Illinois University, Carbondale, Illinois 62901

(Received 1 October 1981; accepted for publication 30 October 1981)

The representations of the most degenerate series of the group U(p,q) which are induced by the representations of the maximal parabolic subgroup are considered in this article. By making use of the infinitesimal operators of these representations in the U(p)$\times$U(q) basis the conditions are derived which are necessary and sufficient for *irreducibility*. For the *reducible* representations we describe their structure (composition series). We select from among the irreducible representations which are obtained in this article all representations of U(p,q) which admit unitarization. As a result we obtain the principal degenerate series, the supplimentary degenerate series, the discrete degenerate series, and the exceptional degenerate series of unitary representations of U(p,q). The U(p)$\times$U(q) spectrum of the representations of U($p+q$) with highest weights $(\lambda_1,0,\ldots,0,\lambda_2)$ is defined. We obtain the integral representation for the matrix elements of the degenerate representations of U(p,q) in the U(p)$\times$U(q) basis. The matrix elements of the irreducible representations of U($p+q$) with highest weights $(\lambda,0,\ldots,0)$, $(0,\ldots,0,\lambda)$ are evaluated in the U(p)$\times$U(q) basis.

PACS numbers: 02.20.Qs

I. INTRODUCTION

This paper is continuation of the work done in Refs. 1 and 2. In Ref. 1 the noncompact infinitesimal operators of the principal nonunitary series representations (and therefore, of the principal unitary series representations) of U(p,q) were found in explicit form. These representations are induced by finite dimensional representations of the minimal parabolic subgroup.[3,4] Using these infinitesimal operators, we have obtained in Ref. 2 the infinitesimal operators of the degenerate series representations of U(p,q) in the U(p)$\times$U(q) basis. Here we use these infinitesimal operators for an investigation of the most degenerate series representations. In particular, we obtain the *classification* of unitary representations of U(p,q) of the most degenerate series. Since we work with infinitesimal operators in the U(p)$\times$U(q) basis, we obtain the U(p)$\times$U(q) spectrum of representations [i.e., the decomposition of these representations into irreducible representations of U(p)$\times$U(q)]. The irreducible finite dimensional representations of U(p,q) with highest weights $(\lambda_1,0,\ldots,0,\lambda_2)$ are contained in certain representations of U(p,q) of the most degenerate series (see Ref. 2). Therefore we obtain the U(p)$\times$U(q) spectrum of the irreducible representations of U($p+q$) with highest weights $(\lambda_1,0,\ldots,0,\lambda_2)$ [and hence, with highest weights $(\lambda'_1,\lambda,\ldots,\lambda,\lambda'_2)$].

The representations $\pi_{\lambda_1\lambda_2}$ of Ref. 2 are representations of U(p,q) which are induced by one-dimensional representations of the maximal parabolic subgroup U(p,q). Using the action formula for the induced representations, we find the integral form for the matrix elements of the representations $\pi_{\lambda_1\lambda_2}$. We do not, however, evaluate these integrals due to their awkward nature. The integral form for the matrix elements of $\pi_{\lambda_1\lambda_2}$ leads to an integral form for the finite dimensional irreducible representations of U($p+q$) with highest weights $(\lambda,0,\ldots,0),(0,0,\ldots,\lambda)$. These integrals are evaluated in a trivial manner. Let us note here that we consider matrix elements of operators which correspond to "boosts".

The matrix elements of operators which correspond to arbitrary elements of U(p,q), or U($p+q$), can be reduced to a product of matrix elements for "boosts" and matrix elements of elements of U(p) and U(q).

II. STRUCTURE OF THE MOST DEGENERATE SERIES REPRESENTATIONS OF U(p,q)

In the following we use, without further explanation, the notation of Ref. 2. We shall investigate the representations $\pi_{\lambda_1\lambda_2}$ of Ref. 2. We use the formulas (3) and (4) of Ref. 2. It is convenient to introduce in these formulas the new parameters

$$M_p = m_{1p} + m_{pp}, \quad M_q = m'_{1q} + m'_{qq}, \tag{1}$$

$$J_p = m_{1p} - m_{pp}, \quad J_q = m'_{1q} - m'_{qq}, \tag{2}$$

$$\mu = \langle \Lambda,\alpha_1\rangle = (-\lambda_1+\lambda_2)/2, \tag{3}$$

$$\Lambda_1 = \lambda_1 + \lambda_2. \tag{4}$$

The representation $\pi_{\lambda_1\lambda_2}$ will then be denoted by $\pi_{\Lambda,\mu}$, where Λ_1 and μ are defined by (3) and (4). The basis elements $|m_p, \alpha,m_q,\beta\rangle$ will be denoted by

$$|M_p,M_q,J_p,J_q,\alpha,\beta\rangle \equiv |M_p,M_q,J_p,J_q\rangle.$$

We will omit the labels α and β, since in the formulas below these parameters do not change. The formulas (3) and (4) of Ref. 2 then take on the form

$$
\begin{aligned}
&d\pi_{\Lambda,\mu}(E_{p,p+1})|M_p,M_q,J_p,J_q\rangle \\
&\quad =\left(\mu+\frac{J_p-J_q}{2}-q+1\right)K^{+1}_{-1}K^{+1}_{-1}(\alpha,\beta)|M_p+1,M_q-1,J_p+1,J_q-1\rangle \\
&\qquad +\left(\mu+\frac{J_p+J_q}{2}\right)K^{+1}_{-q}K^{+1}_{-q}(\alpha,\beta)|M_p+1,M_q-1,J_p+1,J_q+1\rangle \\
&\qquad +\left(\mu-\frac{J_p+J_q}{2}-p-q+2\right)K^{+p}_{-1}K^{+p}_{-1}(\alpha,\beta)|M_p+1,M_q-1,J_p-1,J_q-1\rangle \\
&\qquad \left(\mu-\frac{J_p-J_q}{2}-p+1\right)K^{+p}_{-q}K^{+p}_{-q}(\alpha,\beta)|M_p+1,M_q-1,J_p-1,J_q+1\rangle ,
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
&d\pi_{\Lambda,\mu}(E_{p+1,p})|M_p,M_qJ_p,J_q\rangle \\
&\quad =\left(\mu-\frac{J_p-J_q}{2}-p+1\right)K^{-1}_{+1}K^{-1}_{+1}(\alpha,\beta)|M_p-1,M_q+1,J_p-1,J_q+1\rangle \\
&\qquad +\left(\mu-\frac{J_p+J_q}{2}-p-q+2\right)K^{-1}_{+q}K^{-1}_{+q}(\alpha,\beta)|M_p-1,M_q+1,J_p-1,J_q-1\rangle \\
&\qquad +\left(\mu+\frac{J_p+J_q}{2}\right)K^{-p}_{+1}K^{-p}_{+1}(\alpha,\beta)|M_p-1,M_q+1,J_p+1,J_q+1\rangle \\
&\qquad +\left(\mu+\frac{J_p-J_q}{2}-q+1\right)K^{-p}_{+q}K^{-p}_{+q}(\alpha,\beta)|M_p-1,M_q+1,J_p+1,J_q-1\rangle .
\end{aligned} \tag{6}
$$

It is evident that J_p and J_q are contained in the multiplicative factors of the right sides of (5) and (6) in the form $\pm(J_p+J_q)/2$, $\pm(J_p-J_q)/2$. This observation is important for the subsequent investigation of the representation $\pi_{\Lambda,\mu}$.

We now want to find the values of M_p,M_q,J_p,J_q which are admitted by the representations $\pi_{\Lambda,\mu}$. It was shown in Ref. 2 that in the representation $\pi_{\Lambda,\mu}$ the parameters $m_{1p},m_{pp},m'_{1q},m'_{qq}$ take on all possible integer values such that $m_{1p}\geqslant 0$, $m_{pp}\leqslant 0$, $m'_{1q}\geqslant 0$, $m'_{qq}\leqslant 0$, and

$$m_{1p}+m_{pp}+m'_{1q}+m'_{qq}=\Lambda_1. \tag{7}$$

Therefore, M_p and M_q take on all possible integer values such that

$$M_p+M_q=\Lambda_1. \tag{8}$$

It follows from (1), (2), (7), and (8) that J_p+J_q is even if Λ_1 is even and odd if Λ_1 is odd. The values of J_p and J_q, for fixed values M_p and M_q, are shown in Fig. 1 (small circles). Figure 1 shows the set of points (J_p,J_q) subject to the condition $0\leqslant|M_p|\leqslant|\Lambda_1|$. A change of M_p and M_q leads to a parallel shift of the lattice points (J_p,J_q) such that the apex of the lattice remains on $NABN'$. If $|M_p|\geqslant|\Lambda_1|$ then the apex is on AN. If $M_p\leqslant 0$ then the apex is on BN'. In each case the distance from the axis OJ_q is $|M_p|$.

It is clear that the set of points (J_p,J_q), for all M_p and M_q satisfying condition (8), covers the shaded domain of Fig. 2. Moreover, J_p,J_q are integers such that J_p+J_q has the same parity as Λ_1, i.e., J_p+J_q and Λ_1 are in the same congruence class mod 2, since $(-1)^{2\mu}=(-1)^{\Lambda_1}$.

Let us note that the set of points (M_p,M_q,J_p,J_q) for the representations $\pi_{\Lambda,\mu}$ depends on Λ_1, but does not depend on μ.

We return to formulas (5) and (6). With the help of these formulas the following lemma can be proven

Lemma 1: The representation $\pi_{\Lambda,\mu}$ is completely (and infinitesimally) irreducible if for any collection of numbers (M_p,M_q,J_p,J_q), which is admitted by the representation $\pi_{\Lambda,\mu}$, none of the numbers

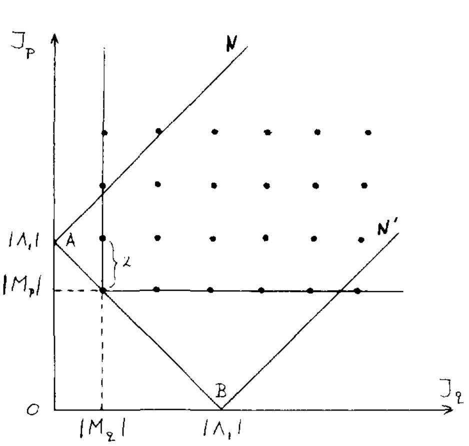


FIG. 1. The set of points (J_p,J_q) subject to the condition $0\leqslant|M_p|\leqslant|\Lambda_1|$.

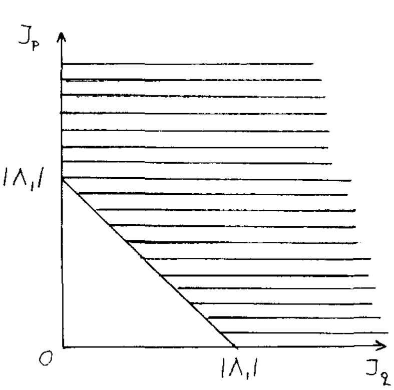


FIG. 2. The set of points (J_p,J_q) with M_p and M_q satisfying Eq. (8).

$$\mu + \frac{J_p - J_q}{2} - q + 1, \quad \mu + \frac{J_p + J_q}{2},$$

$$\mu - \frac{J_p + J_q}{2} - p - q + 2, \quad \mu - \frac{J_p - J_q}{2} - p + 1$$

[these are multiplicative factors in the coefficients of Eqs. (5) and (6)] is equal to 0.

This lemma is proven in the same way as is statement 7.1 in Ref. 5 and we omit the proof.

Theorem 1: The representation $\pi_{\Lambda,\mu}$ is irreducible if and only if 2μ is not an integer which has the same parity as Λ_1.

The irreducibility of these representations follows from Lemma 1. Indeed, $J_p + J_q$ (and therfore $J_p - J_q$) has the same parity as Λ_1. Hence, if 2μ is not an integer which has the same parity as Λ_1, then none of the numbers of Lemma 1 is equal to 0. Vice versa, every representation $\pi_{\Lambda,\mu}$ is reducible for which 2μ is an integer which has the same parity as Λ_1.

The reducibilty of these representations will be shown by considering separately each type (see cases 1–4 below).

Let $\pi_{\Lambda,\mu}$ be a representation for which 2μ is an integer with the same parity as Λ_1. We need the following lemma.

Lemma 2: Let π denote a subquotient representation of $\pi_{\Lambda,\mu}$ (i.e., a representation which is realized on a subspace of a quotient space) and let H denote the space of the representation π. The space H admits certain of the basis vectors $|M_p,M_q,J_p,J_q\rangle$. The representation π is *irreducible* if for every element $|M_p,M_q,J_p,J_q\rangle$ which is admitted by H (i.e., which is an element of H)

(a) the number

$$\mu + \frac{J_p - J_q}{2} - q + 1$$

can become equal to zero only if the elements $|M_p \pm 1, M_q \mp 1, J_p + 1, J_q - 1\rangle$ are not elements of H,

(b) the number

$$\mu + \frac{J_p + J_q}{2}$$

can become equal to zero only if the elements $|M_p \pm 1, M_q \mp 1, J_p + 1, J_q + 1\rangle$ are not elements of H,

(c) the number

$$\mu - \frac{J_p + J_q}{2} - p - q + 2$$

can become equal to zero if the elements $|M_p \pm 1, M_q \mp 1, J_p - 1, J_q - 1\rangle$ are not elements of H, and

(d) the number

$$\mu - \frac{J_p - J_q}{2} - p + 1$$

can become equal to zero only if the elements $|M_p \pm 1, M_q \mp 1, J_p - 1, J_q + 1\rangle$ are not elements of H,

The proof of this lemma is analogous to the proof of statement 7.2 of Ref. 5. We refer to Ref. 5 and omit here the proof.

In order to investigate the representations $\pi_{\Lambda,\mu}$ for which 2μ is an integer which has the same parity as Λ_1, we set the numbers of Lemma 2 equal to zero [i.e., the multipliers of the coefficients of the right-hand side of (5) and (6)],

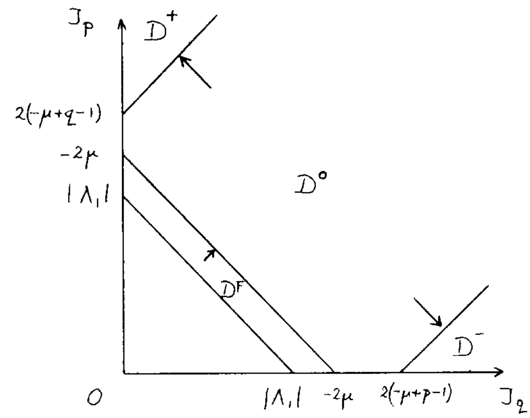


FIG. 3. Case 1: Graphical representation of Eqs. (9), (10), and (12).

$$\mu + \frac{J_p - J_q}{2} - q + 1 = 0, \tag{9}$$

$$\mu + \frac{J_p + J_q}{2} = 0, \tag{10}$$

$$\mu - \frac{J_p + J_q}{2} - p - q + 2 = 0, \tag{11}$$

$$\mu - \frac{J_p - J_q}{2} - p + 1 = 0. \tag{12}$$

We shall consider only those representations $\pi_{\Lambda,\mu}$ for which $\mu \leqslant (p + q - 1)/2$, because the representations $\pi_{\Lambda,\mu}$ and $\pi_{\Lambda_1, -\mu+p+q-1}$ contain the same irreducible constituents (see Sec. 4, Chap. 5 in Ref. 5). Since $\mu \leqslant (p + q - 1)/2$ and $J_p \geqslant 0$, $J_q \geqslant 0$, it follows that the relation (11) is not possible. Let us consider the relations (9), (10), and (12). We shall distinguish four cases.

Case 1: The representations $\pi_{\Lambda,\mu}$ for which 2μ is an integer with the same parity as Λ_1, and for which $-2\mu \geqslant |\Lambda_1|$. We represent the relations (9), (10), and (12) graphically and obtain Fig. 3. The set of points (J_p,J_q), admitted by the representation $\pi_{\Lambda,\mu}$, is divided into four parts denoted by D^F, D^0, D^+, D^-. The linear subspaces which correspond to the points (J_p,J_q) of the domains D^F, $D^F \cup D^0$, $D^F \cup D^0 \cup D^+$, $D^F \cup D^0 \cup D^-$ are invariant under the representation $\pi_{\Lambda,\mu}$. (Note that the points lying on the boundaries belong to the domain which contains the arrow pointing to the boundary.) The invariance is verified by using formulas

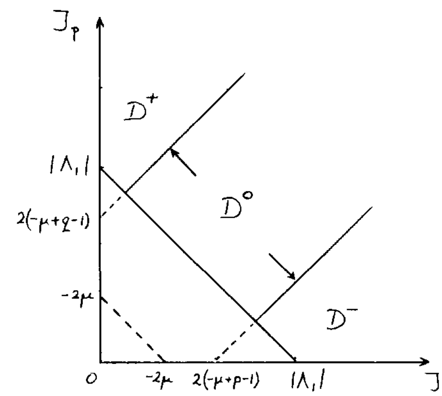


FIG. 4. Case 2: The set of points (J_p,J_q) and the three domains D^0, D^+, D^-.

(5) and (6). The infinitesimal operators $d\pi_{\Lambda,\mu}(E_{rs})$ of the other E_{rs} do not violate the invariance because $d\pi_{\Lambda,\mu}(E_{rs})$ differs from $d\pi_{\Lambda,\mu}(E_{p,p+1})$ and $d\pi_{\Lambda,\mu}(E_{p+1,p})$ only by the Clebsch–Gordan coefficients $K\,{}^{\pm 1(p)}_{\pm 1(q)}$ (see Ref. 2).

It follows from Lemma 2 that the restriction of the representation $\pi_{\Lambda,\mu}$ onto D^F, $(D^F \cup D^0)/D^F$, $(D^F \cup D^0 \cup D^+)/(D^F \cup D^0)$, $(D^F \cup D^0 \cup D^-)/(D^F \cup D^0)$ give irreducible representations of U(p,q). We denote them by $D^F_{\Lambda,\mu}$, $D^0_{\Lambda,\mu}$, $D^+_{\Lambda,\mu}$, $D^-_{\Lambda,\mu}$. The points (J_p,J_q) in Fig. 3 which are located in the domains D^F, D^0, D^+, D^-, correspond to these representations, respectively. The representations $\pi_{\Lambda,\mu}$ of this case have the following structure:

$$\begin{bmatrix} D^F_{\Lambda,\mu} & * & 0 & 0 \\ 0 & D^0_{\Lambda,\mu} & * & * \\ 0 & 0 & D^+_{\Lambda,\mu} & 0 \\ 0 & 0 & 0 & D^-_{\Lambda,\mu} \end{bmatrix}.$$

Here $*$ denotes a nonzero matrix.

The representation $D^F_{\Lambda,\mu}$ is finite dimensional. The highest weight of it is $(\lambda_1,0,...,0,\lambda_2)$, where λ_1, λ_2 are defined in terms of Λ_1,μ by means of (3) and (4) (see Ref. 2). The condition $J_p + J_q \leqslant -2\mu$ defines completely the set of points (M_p,M_q,J_p,J_q) which corresponds to $D^F_{\Lambda,\mu}$ and, in turn, this set defines completely by (1) and (2), the set of highest weights

$$(m_{1p},0,...,0,m_{pp})_{\mathrm{U}(p)}(m'_{1q},0,...,0,m'_{qq})_{\mathrm{U}(q)}$$

of the representations of U$(p)\times$U(q) which are contained in the representation $D^F_{\Lambda,\mu}$ of U(p,q) [and therefore, of U$(p+q)$], when restricted to U$(p)\times$U(q). This set of highest weights can be easily found for every particular case.

Case 2: The representations $\pi_{\Lambda,\mu}$ for which 2μ is an integer with the same parity as Λ_1, and for which $-2\mu < |\Lambda_1|$, $\mu \neq (p+q-1)/2$. Representing the relations (9), (10), and (12) graphically we divide the set of points (J_p,J_q) into three parts, D^0, D^+, D^- (see Fig. 4). The linear subspaces which correspond to D^0, $D^0 \cup D^+$, $D^0 \cup D^-$ are invariant under $\pi_{\Lambda,\mu}$. The representation $\pi_{\Lambda,\mu}$ realizes on D^0, $(D^0 \cup D^+)/D^0$. $(D^0 \cup D^-)/D^0$, irreducible representations of U(p,q). This is a consequence of Lemma 2. We denote these representations by $D^0_{\Lambda,\mu}$, $D^+_{\Lambda,\mu}$, $D^-_{\Lambda,\mu}$, respectively. The representations $\pi_{\Lambda,\mu}$ for this case have the structure

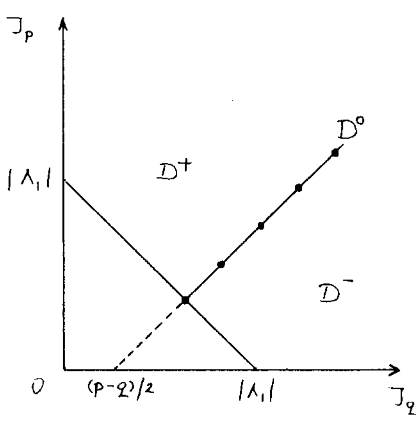


FIG. 5. Case 3: The domain D^0 as a line.

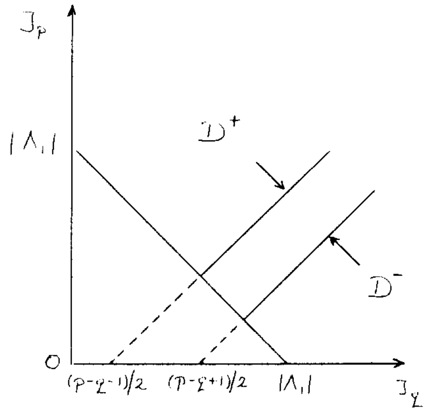


FIG. 6. Case 4: A graphical representation of Eqs. (9) and (12).

$$\begin{bmatrix} D^0_{\Lambda,\mu} & * & * \\ 0 & D^+_{\Lambda,\mu} & 0 \\ 0 & 0 & D^-_{\Lambda,\mu} \end{bmatrix}.$$

Case 3: The representations $\pi_{\Lambda,\mu}$ for which 2μ is an integer with the same parity as Λ_1, and for which $\mu = (p+q)/2 - 1$.

This case is really included in Case 2. We separate it in order to emphasize that in this case the domain D^0 turns into a line (see Fig. 5).

Case 4: The representations $\pi_{\Lambda,\mu}$ for which 2μ is an integer with the same parity as Λ_1, and for which $\mu = (p+q-1)/2$.

The relations (9) and (12) are represented graphically in Fig. 6. They divide the set of points (J_p,J_q) into two parts. Between the lines, according to relations (9) and (12), there are no points (J_p,J_q) admitted by $\pi_{\Lambda,\mu}$. The linear subspaces which correspond to the domains D^+ and D^- are invariant under $\pi_{\Lambda,\mu}$ [as a consequence of (5) and (6)] and the representation $\pi_{\Lambda,\mu}$ is decomposed into two irreducible representations of U(p,q). We denote them by $D^+_{\Lambda,\mu}$, $D^-_{\Lambda,\mu}$, respectively.

III. REPRESENTATIONS OF U(p,q) OF THE MOST DEGENERATE UNITARY SERIES

We have constructed the irreducible representations $\pi_{\Lambda,\mu}$, where 2μ is not an integer of the same parity as Λ_1, and the irreducible representations $D^+_{\Lambda,\mu}$, $D^0_{\Lambda,\mu}$, $D^-_{\Lambda,\mu}$, $D^F_{\Lambda,\mu}$. We now want to select those from among them which can be unitarized [i.e., those which are infinitesimally equivalent to unitary representations of U(p,q)].

Theorem 2: The following representations are unitarizable:

(1) the representations $\pi_{\Lambda,\mu}$, $\mu = i\rho + (p+q-1)/2$, where $\rho \in \mathbb{R}$, $\rho \neq 0$ (the principal most degnerate unitary series);

(2) the representations $\pi_{\Lambda,\mu}$ for which Λ_1 has the same parity as $p+q$, and for which μ is in the interval $(p+q)/2 - 1 < \mu < (p+q)/2$ (supplimentary most degenerate series);

(3) the representations $D^+_{\Lambda,\mu}, D^-_{\Lambda,\mu}$;

(4) the representations $D^0_{\Lambda,(p+q)/2-1}$.

Proof: Unitarizablilty of the class (1) representations follows from formulas (5) and (6). Indeed, unitarity of the representations of U(p,q) is equivalent to the condition that the noncompact infinitesimal operators satisfy, on the set of finite linear combinations of $|m_p,\alpha,m_q,\beta\rangle$, the relations

$$d\pi_{\Lambda,\mu}(E_{rs})^* = -d\pi_{\Lambda,\mu}(E_{sr}). \tag{13}$$

The formulas (5) and (6) show that this relation is satisfied for $E_{p,p+1}, E_{p+1,p}$. That Eq. (13) is satisfied for the other infinitesimal operators E_{rs} follows from the fact that the E_{rs} are obtained by means of commutations of $E_{p,p+1}, E_{p+1,p}$ with the compact infinitesimal operators. The commutations conserve the unitarity condition (13).

The representations of the classes (2), (3), (4) are not unitary in the basis $|m_p,\alpha,m_q,\beta\rangle$. Let us introduce a new basis:

$$|m_p,\alpha,m_q,\beta\rangle = a(m_{1p},m_{pp},m'_{1q},m'_{qq})^{1/2}|m_p,\alpha,m_q,\beta\rangle',$$

where the coefficients $a(m_{1p},m_{pp},m'_{1q},m'_{qq})$ are defined by formulas (5)–(8) of Ref. 2. Here m_1 is replaced by $\lambda_1 = (\Lambda_1/2) - \mu$, and m_2 by $\lambda_2 = (\Lambda_1/2) + \mu$. In the basis $|m_p,\alpha,m_q,\beta\rangle'$, the formulas (5) and (6) transform to the formulas

$$\begin{aligned}
&d\pi_{\Lambda,\mu}(E_{p,p+1})|M_p,M_q,J_p,J_q\rangle' \\
&= \left[\left(-\mu+\frac{J_p-J_q}{2}+p\right)\left(\mu+\frac{J_p-J_q}{2}-q+1\right)\right]^{1/2} K^{+1}_{-1}K^{+1}_{-1}(\alpha,\beta)|M_p+1,M_q-1,J_p+1,J_q-1\rangle' \\
&\quad+\left[\left(\mu+\frac{J_p+J_q}{2}\right)\left(-\mu+\frac{J_p+J_q}{2}+p+q-1\right)\right]^{1/2} K^{+1}_{-q}K^{+1}_{-q}(\alpha,\beta)|M_p+1,M_q-1,J_p+1,J_q+1\rangle' \\
&\quad+\left[\left(\mu-\frac{J_p+J_q}{2}-p-q+2\right)\left(-\mu-\frac{J_p+J_q}{2}+1\right)\right]^{1/2} K^{+p}_{-1}K^{+p}_{-1}(\alpha,\beta)|M_p+1,M_q-1,J_p-1,J_q-1,\rangle' \\
&\quad+\left[\left(\mu-\frac{J_p-J_q}{2}-p+1\right)\left(-\mu-\frac{J_p-J_q}{2}+q\right)\right]^{1/2} K^{+p}_{-q}K^{+p}_{-q}(\alpha,\beta)|M_p+1,M_q-1,J_p-1,J_q+1\rangle',
\end{aligned} \tag{14}$$

$$\begin{aligned}
&d\pi_{\Lambda,\mu}(E_{p+1,p})|M_p,M_q,J_p,J_q\rangle' \\
&= -\left[\left(-\mu+\frac{J_p-J_q}{2}+p-1\right)\left(\mu+\frac{J_p-J_q}{2}-q\right)\right]^{1/2} K^{-1}_{+1}K^{-1}_{+1}(\alpha,\beta)|M_p-1,M_q+1,J_p-1,J_q+1\rangle' \\
&\quad-\left[\left(-\mu+\frac{J_p+J_q}{2}+p+q-2\right)\left(\mu+\frac{J_p+J_q}{2}-1\right)\right]^{1/2} K^{-1}_{+q}K^{-1}_{+q}(\alpha,\beta)|M_p-1,M_q+1,J_p-1,J_q-1\rangle' \\
&\quad-\left[\left(-\mu-\frac{J_p+J_q}{2}\right)\left(\mu-\frac{J_p+J_q}{2}-p-q+1\right)\right]^{1/2} K^{-p}_{+1}K^{-p}_{+1}(\alpha,\beta)|M_p-1,M_q+1,J_p+1,J_q+1\rangle' \\
&\quad-\left[\left(-\mu-\frac{J_p-J_q}{2}+q-1\right)\left(\mu-\frac{J_p-J_q}{2}-p\right)\right]^{1/2} K^{-p}_{+q}K^{-p}_{+q}(\alpha,\beta)|M_p-1,M_q+1,J_p+1,J_q-1\rangle'.
\end{aligned} \tag{15}$$

The proof of unitarity of the representations of the classes (2), (3), and (4) is the same as for the case of the class (1) representations, but now we use the formulas (14) and (15) instead of the formulas (5) and (6). This proves the theorem.

It can be shown that the representations of Theorem 2 exhaust all unitarizable representations of U(p,q) among the irreducible representations $\pi_{\Lambda,\mu}$, $D^+_{\Lambda,\mu}$, $D^0_{\Lambda,\mu}$, $D^-_{\Lambda,\mu}$. The representations of class (1) and of class (3) for $\mu \neq (p+q-1)/2$ have been constructed in another manner in Ref. 6. It was shown in Ref. 6 that the representations $D^+_{\Lambda,\mu}$ and $D^-_{\Lambda,\mu}$ for $\mu \neq (p+q-1)/2$ belong to the discrete series, i.e., their matrix elements belong to $L^2(\mathrm{U}(p,q))$. Physicists call the representations of class (4) "ladder representations." Class (4) are also called representations of the exceptional series.

The representations of Theorem 2 for the case of the conformal group SU(2,2) were considered in Ref. 7. The structure of the representations $\pi_{\Lambda,\mu}$ for SU(2,2) were investigated in Ref. 8 (see also Ref. 9).

IV. THE REPRESENTATION $\pi_{\Lambda,\mu}$ IN GLOBAL FORM

We have obtained the representations $\pi_{\Lambda,\mu}$ in an infinitesimal form. Now we want to obtain them in global form. We consider in U(p,q) the minimal parabolic subgroup P_θ. This subgroup has a decomposition[3,5]

$$P_\theta = A_\theta N_\theta M_\theta = ANM_\theta(K),$$

where A and N are defined by the Langlands decomposition of the minimal parabolic subgroup $P = ANM$ (see Ref. 1). The subgroup A_θ consists of the matrices

$$a_\eta = \left[\begin{array}{ccc|ccc}
1 & & 0 & & & 0 \\
 & \ddots & & & & \\
0 & & 1 & & & \\
 & & \cosh\eta & \sinh\eta & & \\
\hline
 & & \sinh\eta & \cosh\eta & & 0 \\
 & & & & 1 & \\
 & 0 & & 0 & & \ddots \\
 & & & & & 1
\end{array}\right]\begin{array}{l} \}\,p \\ \\ \}\,q \end{array} \tag{16}$$

Its Lie algebra a_θ consists of the matrices

$$\left[\begin{array}{cc|cc}
0 & & & 0 \\
 & 0 & a & \\
\hline
 & a & 0 & \\
0 & & & 0
\end{array}\right], \quad a \in R.$$

The subgroup M_θ consists of the matrices

$$\left[\begin{array}{ccccc|ccccc} & & & & 0 & 0 & & & & \\ & & & & \cdot & \cdot & & & & \\ & * & & & \cdot & \cdot & & & * & \\ & & & & \cdot & \cdot & & & & \\ & & & & 0 & 0 & & & & \\ \hline 0 & \cdots & & 0 & e^{i\phi} & 0 & 0 & & \cdots & 0 \\ \hline 0 & \cdots & & 0 & 0 & e^{i\phi} & 0 & & \cdots & 0 \\ & & & & 0 & 0 & & & & \\ & & & & \cdot & \cdot & & & & \\ & & & & \cdot & \cdot & & & & \\ & * & & & \cdot & \cdot & & & * & \\ & & & & 0 & 0 & & & & \end{array}\right] \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \\ \\ \end{array}\right\} p-1 \\ \\ \\ \left.\begin{array}{l} \\ \\ \\ \\ \\ \end{array}\right\} q-1, \end{array} \quad (17)$$

where the $(p+q-2)\times(p+q-2)$ matrices, constructed from the four matrices denoted by $*$, constitute the subgroup $U(p-1,q-1)$. Thus, $M_\theta \sim U(p-1,q-1)\times U(1)$. The subgroup $M_\theta(K)$ coincides with the intersection of M_θ with $K = U(p)\times U(q)$. It is clear that $M_\theta(K) \sim U(p-1)\times U(q-1)\times U(1)$.

The subgroup N_θ can be obtained in the following manner. First one constructs the Lie algebra n_θ of N_θ. This algebra can be constructed with the help of the Lie algebra n of N. The Lie algebra n is generated by the root vectors

$$\omega_{\omega_i - \omega_j}, \quad e_{\omega_i + \omega_j}, \quad i<j, \quad e_{\omega_i}, e_{2\omega_i},$$

which were described in Ref. 1 The subalgebra n_θ is generated by the root vectors which belong to those restricted roots which are not identically equal to 0 on a_θ. It is easy to see that $e_{\omega_1 - \omega_j}, e_{\omega_1 + \omega_j}, j \neq 1, e_{\omega_1}, e_{2\omega_1}$ are the root vectors generating n_θ. In order to construct N_θ we have to construct exp n_θ. However, the structure of N_θ is complicated and thus we will not construct all elements of N_θ. For our further considerations it will be sufficient to deal with a subgroup of N_θ. This subgroup is generated by two root vectors $e_{\omega_1 - \omega_2}$, two root vectors $e_{\omega_1 + \omega_2}$, and the root vector $e_{2\omega_1}$. We exponentiate these root vectors. The two matrices $e_{\omega_1 - \omega_2}$ (as basis elements of a Lie algebra) generate the subgroup

$$N_{\omega_1 - \omega_2} = \begin{bmatrix} E_{p-2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\bar{Z} & \bar{Z} & 0 & 0 \\ 0 & Z & 1 & 0 & Z & 0 \\ 0 & Z & 0 & 1 & Z & 0 \\ 0 & 0 & \bar{Z} & -\bar{Z} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{q-2} \end{bmatrix}, \quad Z \in \mathbb{C},$$

where E_n is a unit $n\times n$ matrix. The two matrices $e_{\omega_1 + \omega_2}$ generate the subgroup

$$N_{\omega_1 + \omega_2} = \begin{bmatrix} E_{p-2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\bar{Z}_1 & \bar{Z}_1 & 0 & 0 \\ 0 & Z_1 & 1 & 0 & -Z_1 & 0 \\ 0 & Z_1 & 0 & 1 & -Z_1 & 0 \\ 0 & 0 & -\bar{Z}_1 & \bar{Z}_1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{q-2} \end{bmatrix}, \quad Z_1 \in \mathbb{C}$$

and the matrix $e_{2\omega_1}$ generates the subgroup

$$N_{2\omega_1} = \begin{bmatrix} E_{p-1} & 0 & 0 & \\ 0 & 1-it & -it & 0 \\ 0 & it & 1-it & 0 \\ 0 & 0 & 0 & E_{q-1} \end{bmatrix}, \quad t \in R.$$

Let us construct the subgroup N'_θ of N_θ generated by $N_{\omega_1 - \omega_2}, N_{\omega_1 + \omega_2}, N_{2\omega_1}$. To obtain this subgroup it is sufficient to multiply the matrices

$$N'_\theta = N_{\omega_1 - \omega_2} N_{\omega_1 + \omega_2} N_{2\omega_1} = \begin{bmatrix} E_{p-2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -(\bar{Z}_1 + \bar{Z}) & \bar{Z}_1 + \bar{Z} & 0 & 0 \\ 0 & Z_1 + Z & 1 + it - 2\bar{Z}_1 Z & -it + 2\bar{Z}_1 Z & -Z_1 + Z & 0 \\ 0 & Z_1 + Z & it - 2\bar{Z}_1 Z & 1 - it + 2\bar{Z}_1 Z & -Z_1 + Z & 0 \\ 0 & 0 & -\bar{Z}_1 + \bar{Z} & \bar{Z}_1 - \bar{Z} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{q-2} \end{bmatrix}. \quad (18)$$

Now we construct the representations $\pi^\theta_{\sigma,\lambda}$ of U(p,q) which are induced by the (irreducible) representations

$$hnm \to \exp[\lambda(\ln h)]\sigma(m), \quad h\in A_\theta, \quad n\in N_\theta, \quad m\in M_\theta, \tag{19}$$

of the subgroup $P_\theta = A_\theta N_\theta M_\theta$. The λ are linear forms on the Lie algebra a_θ and the σ are the one-dimensional representations of M_θ, which are identically equal to 1 on U($p-1,q-1$). The representation $\pi^\theta_{\sigma,\lambda}$ acts on the space $L^2_\sigma(K)$, $K = \mathrm{U}(p)\times\mathrm{U}(q)$ consisting of functions f of $L^2(K)$ such that

$$f(mk) = \sigma(m)f(k), \quad m\in M_\theta(K). \tag{20}$$

The operators $\pi^\theta_{\sigma,\lambda}(g)$ act on the functions $f\in L^2_\sigma(K)$ by the formula

$$\pi^\theta_{\sigma,\lambda}(g)f(k) = \exp[\lambda(\ln h)]f(k_g), \tag{21}$$

where k_g and h are defined by the decomposition

$$kg = hn(\exp X)k_g, \quad h\in A_\theta, \quad n\in N_\theta, X\in m_\theta\cap p, \quad k_g\in K$$

(see Ref. 3 and Theorem 1.3 in Ref. 5). For more details on the construction of the representations induced by the representations of parabolic subgroups see Chap. 5 of Ref. 5.

Since a_θ is a one-dimensional subalgebra the linear form λ is defined by a single number $\mu = \lambda(e)$, where e is a basis element of a_θ. We choose $e = (E_{p,p+1} + E_{p+1,p})/2$. The representation σ of Eq. (19) is defined by a single integer Λ_1, characterizing the representation $e^{i\phi}\to e^{i\Lambda_1\phi}$ of the subgroup U(1) of M_θ. Thus $\pi^\theta_{\sigma,\lambda}$ is defined by two numbers μ and Λ_1. For this reason we denote this representation by $\pi_{\Lambda,\mu}$ in what follows. The representation $\pi_{\Lambda,\mu}$ is in fact infinitesimally equivalent to the representation $\pi_{\Lambda,\mu}$ of U(p,q), which was considered in the previous sections. To verify this it is sufficient to construct for the representation (21) the noncompact infinitesimal operators by using Lemma 5.2 of Ref. 5. This construction is simple and we omit it. The noncompact infinitesimal operators for both representations are identical in appropriately chosen bases.

The representation σ is trivial (i.e., = 1) on U($p-1,q-1$) [and, therefore, on U($p-1$)$\times$U($q-1$)]. Hence, because of (20), the functions of the space $L^2_\sigma(K)$ can be considered as functions on the coset space

$$Y' = (\mathrm{U}(p-1)\times\mathrm{U}(q-1))\backslash(\mathrm{U}(p)\times\mathrm{U}(q)) \sim \mathrm{U}(p-1)\backslash\mathrm{U}(p)\times\mathrm{U}(q-1)\backslash\mathrm{U}(q).$$

We now introduce parameters on Y'. In order to do this we use the decomposition (see Ref. 10)

$$g = ha_n(\phi_n)\beta_n(\theta_n)h', \quad h,h'\in\mathrm{U}(n-1), \tag{22}$$

of the elements g of U(n). Here $a_n(\phi_n)$ differs from the unit $n\times n$ matrix by the last diagonal element, which is equal to $e^{i\phi}$; $\beta_n(\theta_n)$ has the form

$$\beta_n(\theta_n) = \begin{bmatrix} E_{n-2} & 0 & 0 \\ 0 & \cos\theta_n & -\sin\theta_n \\ 0 & \sin\theta_n & \cos\theta_n \end{bmatrix}, \quad 0\leqslant\theta\leqslant\pi/2. \tag{23}$$

In (22) one can take, instead of h and h', the elements $h\tilde h$ and $\tilde h^{-1}h'$, $\tilde h\in\mathrm{U}(n-2)$. Therefore, the unique decomposition can be chosen in the form

$$g = ha_n(\phi_n)\beta_n(\theta_n)a_{n-1}(\phi_{n-1})\beta_{n-1}(\theta_{n-1})\cdots a_2(\phi_2)\beta_2(\theta_2)a_1(\phi_1).$$

Hence, the coset space Y' is parametrized by

$$\phi_p,\theta_p,\phi_{p-1},\theta_{p-1},\dots,\phi_2,\theta_2,\phi_1; \quad \bar\phi_q,\psi_q,\dots,\bar\phi_2,\psi_2,\bar\phi_1. \tag{24}$$

Apart from U($p-1$)$\times$U($q-1$), the group $M_\theta(K)$ contains the subgroup U(1). Therefore, the functions of $L^2_\sigma(K)$ are, in fact, functions on the coset space

$$Y = (\mathrm{U}(p-1)\times\mathrm{U}(q-1)\times\mathrm{U}(1))\backslash(\mathrm{U}(p)\times\mathrm{U}(q)).$$

Since U(1) is realized in the form of the matrices given by Eq. (17), it becomes necessary to set $\bar\phi_q = -\phi_p$ in Eq. (24) in order to obtain a parametrization of the space Y.

The invariant measure on K leads to the following measure on Y (see, for example, Ref. 10):

$$dy = \frac{(p-1)!(q-1)!}{2\pi^{p+q-1}}\prod_{K=2}^{p}\sin^{2k-3}\theta_k\cos\theta_k d\theta_k\prod_{r=1}^{p}d\phi_r \times\prod_{K'=2}^{q}\sin^{2k'-3}\psi_{k'}\cos\psi_{k'}d\psi_{k'}\prod_{r'=1}^{q-1}\bar\phi_{r'}. \tag{25}$$

Thus, instead of the functions f of $L^2_\sigma(K)$ we can consider the corresponding functions of the parameters (24), in which $\bar\phi_q = -\phi_p$, or the functions depending on the elements of K of the form

$$k' = \prod_{K=2}^{p} a_k(\phi_k)\beta_k(\theta_k)\cdot a_1(\phi_1)a_q(-\phi_p)\beta_q(\psi_q) \times\prod_{k'=q-1}^{2} a_{k'}(\bar\phi_{k'})\beta_{k'}(\psi_{k'})a_1(\bar\phi_1). \tag{26}$$

V. THE INTEGRAL FORM FOR THE MATRIX ELEMENTS OF $\pi_{\Lambda,\mu}$

We shall consider the matrix elements of the representation operators which correspond to the elements (16) of U(p,q). We want to find an explicit form for the operators $\pi_{\Lambda,\mu}(a_\eta)$. In order to do this we use formula (21) and we find the parameters corresponding to the element k_g for $g = a_\eta$. If the element k' has the form (26) then

$$k'a_\eta = a_p(\phi_p)\beta_p(\theta_p)a_q(-\phi_p)\beta_q(\psi_q)a_\eta k'', \tag{27}$$

where

$$k'' = \prod_{k=p-1}^{2} a_k(\phi_k)\beta_k(\theta_k)a_1(\phi_1)\prod_{k'=q-1}^{2} a_{k'}(\bar\phi_{k'})\beta_{k'}(\psi_{k'})a_1(\bar\phi_1).$$

The element $a_p(\phi_p)\beta_p(\theta_p)a_q(-\phi_p)\beta_q(\psi_q)a_\eta$ is represented in the form

$$na_{\eta'}a_p(\phi'_p)\beta_p(\theta'_p)a_q(-\phi'_p)\beta_q(\psi_{q'}), \quad n\in N'_\theta. \tag{28}$$

Comparing these two elements we find $\eta',\phi'_p,\theta'_p,\psi'_q$ as functions of η, ϕ_p, θ_p, ψ_q. These functions are (for convenience we omit the indices p and q)

$$\sin\theta' = \sin\theta\,[(\cosh\eta\cos\theta - \sinh\eta\cos\psi e^{-2i\phi})^2 + \sin^2\theta\,]^{-1/2}, \tag{29}$$

$$\sin\psi' = \sin\psi[(\cosh\eta\cos\psi - \sinh\eta\cos\theta e^{2i\phi})^2 + \sin^2\psi]^{-1/2}, \tag{30}$$

$$\cos\theta' = \frac{\cosh\eta\cos\theta - \sinh\eta\cos\psi\, e^{-2i\phi}}{[(\cosh\eta\cos\theta - \sinh\eta\cos\psi\, e^{-2i\phi})^2 + \sin^2\theta\,]^{1/2}}, \tag{31}$$

$$\cos \psi' = \frac{\cosh \eta \cos \psi - \sinh \eta \cos \theta\, e^{2i\phi}}{[(\cosh \eta \cos \psi - \sinh \eta \cos \theta\, e^{2i\phi})^2 + \sin^2\psi]^{1/2}}, \tag{32}$$

$$e^{2i\phi'} = \left(\frac{(\cosh \eta \cos \theta - \sinh \eta \cos \psi\, e^{-2i\phi})^2 + \sin^2\theta}{(\cosh \eta \cos \psi - \sinh \eta \cos \theta\, e^{2i\phi})^2 + \sin^2\psi}\right)^{1/2} e^{2i\phi}, \tag{33}$$

$$e^{-2\eta'} = [(\cosh \eta \cos \theta - \sinh \eta \cos \psi\, e^{-2i\phi})^2 + \sin^2\theta]^{1/2} \times [(\cosh \eta \cos \psi - \sinh \eta \cos \theta\, e^{2i\phi})^2 + \sin^2\psi]^{1/2}. \tag{34}$$

According to (27)–(34) $\pi_{\Lambda,\mu}(a_\eta)$ acts upon the functions f by the formula

$$\pi_{\Lambda,\mu}(a_\eta) f(\phi_p,\theta_p,\phi_{p-1},\theta_{p-1},...,\phi_1,\psi_q,\bar{\phi}_{q-1},\psi_{q-1},...,\bar{\phi}_1) = e^{2\eta'\mu} f(\phi'_p,\theta'_p,\phi_{p-1},\theta_{p-1},...,\phi_1,\psi'_q,\bar{\phi}_{q-1},\psi_{q-1},...,\bar{\phi}_1). \tag{35}$$

To evaluate the matrix elements we have to choose an orthonormal basis in $L^2_\sigma(K)$. We do this by choosing the matrix elements of the representations of $U(p)\times U(q)$ as a basis. The irreducible representations of K are contained in $\pi_{\Lambda,\mu}$ not more than once (see Ref. 2). Moreover, $\pi_{\Lambda,\mu}$ contains, with unit multiplicity, all irreducible representations of $U(p)\times U(q)$ with the highest weights

$$(m_1,0,...,0m_2)_{U(p)}(n_1,0,...,0,n_2)_{U(q)}, \quad m_1\geqslant 0, \quad m_2\leqslant 0, \quad n_1\geqslant 0, \quad n_2\leqslant 0, \tag{36}$$

for which $m_1+m_2+n_1+n_2=\Lambda_1$. For convenience the representation with the highest weight (36) will be denoted by $[m]\times[n]$.

The matrix elements

$$(\dim[m]\times[n])^{1/2} D^{[m]\times[n]}_{\alpha_0\alpha}(k) \equiv |m,n,\alpha\rangle \tag{37}$$

can be taken for a basis for the space $L^2_\sigma(K)$, where α_0 denotes the double Gel'fand–Zetlin pattern (1) of Ref. 2, and corresponds to the invariance with respect to the subgroup $U(p-1)\times U(q-1)$. The symbol α in (37) represents the double Gel'fand–Zetlin pattern

$$\begin{bmatrix} m_1 & & 0 & \cdots & 0 & & m_2 & n_1 & & 0 & \cdots & 0 & & n_2 \\ & m'_1 & & 0 \cdots 0 & & m'_2 & & & n'_1 & & 0 \cdots 0 & & n'_2 & \\ & & & \cdots & & & & & & & \cdots & & & \end{bmatrix}. \tag{38}$$

The matrix elements of $\pi_{\Lambda,\mu}(a_\eta)$ are now evaluated between the basis functions (37). It is shown in a standard manner (see, for example, Ref. 11) that due to (35) these matrix elements depend only on Λ_1,μ and the first two rows of the patterns (38); moreover, the second row of both schemes α, $\tilde{\alpha}$ have to be the same, i.e.,

$$\langle m,n,\alpha|\pi_{\Lambda,\mu}(a_\eta)|\tilde{m},\tilde{n},\tilde{\alpha}\rangle = d^{\Lambda_1\mu}_{(mn)(\tilde{m}\tilde{n})(m'n')}(\eta), \tag{39}$$

where (m',n') is the second row for the schemes α and $\tilde{\alpha}$.

According to Ref. 10 (see also Ref. 12), the matrix element $D^{[m]}_{\beta_0\beta}(a_p(\phi)\beta_p(\theta))$ of the representation $[m]$ of $U(p)$ [here β_0 and β represent that part of the schemes α_0 and α of Eq. (37), which corresponds to the subgroup $U(p)$] is given by

$$D^{[m]}_{\beta_0\beta}(a_p(\phi)\beta_p(\theta)) \equiv D^m_{0,m'}(\phi,\theta) = e^{i\phi(-m_1-m_2)} d^m_{0,m'}(\theta), \quad m=(m_1,0,...,0,m_2), \quad m'=(m'_1,0,...,0,m'_2), \tag{40}$$

where

$$d^m_{0,m'}(\theta) = (\cos\theta)^{m'_1+m'_2-m_1-m_2} \sum_{k=m'_1}^{m_1} N(m_1,m_2,m'_1,m'_2,k)(\sin\theta)^{2k-m'_1-m'_2}. \tag{41}$$

The function $d^m_{0,m'}(\theta)$ can also be chosen in the form

$$d^m_{0,m'}(\theta) = (\cos\theta)^{m_1+m_2-m'_1-m'_2} \sum_{k=m_2}^{m'_2} N'(m_1,m_2,m'_1,m'_2,k)(\sin\theta)^{m'_1+m'_2-2k}. \tag{42}$$

The formula (41) is defined by (46) of Ref. 10, and (42) is defined by (47) of Ref. 10. The expressions for N and N' can be taken from (46) and (47) of Ref. 10. Let us note that in (40)–(42) the third row has to be taken to be equal to (0,...,0) since the matrix elements (39) depend on two rows of the scheme (38) only.

The matrix element (39) is defined by the formula

$$d^{\Lambda_1\mu}_{(mn)(\tilde{m}\tilde{n})(m'n')}(\eta) = \frac{(p-1)(q-1)}{\pi} \frac{[(\dim[m]\times[n])(\dim[\tilde{m}]\times[\tilde{n}])]^{1/2}}{\dim[m']\times[n']} \int_0^{\pi/2}\int_0^{\pi/2}\int_0^{\pi} d\theta\, d\psi\, d\phi \times \sin^{2p-3}\theta\cos\theta\sin^{2q-3}\psi\cos\psi\, \overline{D^{[m]}_{0,m'}(\phi,\theta) D^{[n]}_{0,n'}(-\phi,\psi)}\, e^{2\eta'\mu} D^{[\tilde{m}]}_{0,m'}(\phi',\theta') D^{[\tilde{n}]}_{0,n'}(-\phi',\psi'), \tag{43}$$

where $\eta',\phi',\theta',\psi'$ are defined by (29)–(34). Substituting the explicit expressions for $\eta',\phi',\theta',\psi'$ and using the formula (41), we obtain the integral form of the matrix elements (43) of the representation $\pi_{\Lambda,\mu}$,

$$d^{\Lambda_1\mu}_{(mn)(\tilde{m}\tilde{n})(m',n')}(\eta) = \frac{(p-1)(q-1)}{\pi} \frac{[(\dim[m]\times[n])(\dim[\tilde{m}]\times[\tilde{n}])]^{1/2}}{\dim[m']\times[n']} \times \sum_{K=m'_1}^{m_1} \sum_{k'=m'_1}^{\tilde{m}'} \sum_{s=n'_1}^{n_1} \sum_{s'=\tilde{n}'_1}^{\tilde{n}'} N(m_1,m_2,m'_1,m'_2,k) N(\tilde{m}_1,\tilde{m}_2,m'_1,m'_2,k') \times N(n_1,n_2,n'_1,n'_2,s) N(\tilde{n}_1,\tilde{n}_2,n'_1,n'_2,s') \int_0^{\pi/2}\int_0^{\pi/2}\int_0^{\pi} d\theta\, d\psi\, d\phi\, (\sin\theta)^{2(k+k'-m'_1-m'_2+p)-3}$$

$$\times(\cos\theta)^{m_1'+m_2'-m_1-m_2+1}(\sin\psi)^{2(s+s'-n_1'-n_2'+q)-3}(\cos\psi)^{n_1'+n_2'-n_1-n_2+1}$$
$$\times[(\cosh\eta\cos\theta-\sinh\eta\cos\psi\, e^{-2i\phi})^2+\sin^2\theta]^{(\Lambda_1-2\mu)/4-k'}[(\cosh\eta\cos\psi-\sinh\eta\cos\theta\, e^{2i\phi})^2$$
$$+\sin^2\psi]^{(\Lambda_1-2\mu)/4-s'}(\cosh\eta\cos\theta-\sinh\eta\cos\psi\, e^{-2i\phi})^{m_1'+m_2'-\tilde m_1-\tilde m_2}$$
$$\times(\cosh\eta\cos\psi-\sinh\eta\cos\theta e^{2i\phi})^{n_1'+n_2'-\tilde n_1-\tilde n_2}e^{2i\phi(m_1+m_2-\tilde m_1-\tilde m_2)}. \qquad (44)$$

Using the formula (42) for the matrix elements (40), we obtain a different expression for the matrix elements (43).

To obtain the matrix elements (44) in an explicit form, we have to evaluate the integrals. We do this for special values Λ_1 and μ in the following section.

VI. MATRIX ELEMENTS OF IRREDUCIBLE REPRESENTATIONS OF U($p+q$) WITH HIGHEST WEIGHTS (λ_1,0,...,0), (0,...,0, λ_2)

We use the formula (44) in order to evaluate elements for the irreducible representations of U($p+q$) with highest weights (0,...,0, λ), $\lambda \leqslant 0$, in the U(p)×U(q) basis. It is known that the representations of U($p+q$) with the highest weights ($-\lambda$,0,...,0) are contragradient to the representations with the highest weights (0,...,0, λ). Therefore, if the matrix elements for the irreducible representations of U($p+q$) with the highest weights (0,...,0, λ) are known then those of the representations with the highest weights ($-\lambda$,0,...,0) are known too.

In order to find the matrix elements for the irreducible unitary representations of U($p+q$) with highest weights (0,...,0,λ), we shall first evaluate the matrix elements for the finite dimensional irreducible representations of U(p,q) which have these highest weights. These representations are contained as subrepresentations in the representations $\pi_{\Lambda,\mu}$ for which $\Lambda_1=2\mu=\lambda$. In Sec. 2 we have explained how to find the U(p)×U(q) spectrum for the finite dimensional subrepresentations of $\pi_{\Lambda,\mu}$. In particular, it is easy to find that the finite dimensional representations of U(p,q) with the highest weight (0,...,0,λ), $\lambda\leqslant 0$, have a U(p)×U(q) spectrum

$$(0,...,0,\lambda)_{U(p)}(0,...,0)_{U(q)};(0,...,0,\lambda+1)_{U(p)}(0,...,0-1)_{U(q)};\quad (0,...,0,\lambda+2)_{U(p)}(0,...,0,-2)_{U(q)};...;\quad (0,...,0,)_{U(p)}(0,...,0,\lambda)_{U(q)}.$$

We consider the formula (44) for the matrix elements of the finite dimensional subrepresentations of $\pi_{\Lambda,\mu}$ for which $\Lambda_1=2\mu=\lambda$,

$$d^{\Lambda,\mu}_{(mn)(\tilde m\tilde n)(m'n')}(\eta)=\frac{(p-1)(q-1)}{\pi}\frac{[(\dim[m]\times[n])(\dim[\tilde m]\times[\tilde n])]^{1/2}}{\dim[m']\times[n']}$$
$$\times N(0,m_2,0,m_2',0)N(0,\tilde m_2,0,m_2',0)N(0,n_2,0,n_2',0)N(0,\tilde n_2,o,n_2',0)\int_0^{\pi/2}\int_0^{\pi/2}\int_0^{\pi}d\theta d\psi d\phi\,(\sin\theta)^{2(p-m_2')-3}$$
$$\times(\cos\theta)^{m_2'-m_2+1}(\sin\psi)^{2(q-n_2')-3}(\cos\psi)^{n_2'-n_2+1}(\cosh\eta\cos\theta-\sinh\eta\cos\psi\, e^{-2i\phi})^{m_2'-\tilde m_2}$$
$$\times(\cosh\eta\cos\psi-\sinh\eta\cos\theta\, e^{2i\phi})^{n_2'-\tilde n_2}e^{2i(m_2-\tilde m_2)\phi}. \qquad (45)$$

The powers of the trigonometrical functions in (45) are non-negative integers. Therefore the integral can be evaluated in a trivial manner. Since

$$(\cosh\eta\cos\theta-\sinh\eta\cos\psi\, e^{-2i\phi})^{m_2'-\tilde m_2}=\sum_{r=0}^{m_2'-\tilde m_2}\binom{m_2'-\tilde m_2}{r}(-1)^r\sinh^r\eta\cos^r\psi\, e^{-2ir\phi}\cosh^{m_2'-\tilde m_2-r}\eta\cos^{m_2'-\tilde m_2-r}\theta,$$

$$(\cosh\eta\cos\psi-\sinh\eta\cos\theta\, e^{2i\phi})^{n_2'-\tilde n_2}=\sum_{r'=0}^{n_2'-\tilde n_2}\binom{n_2'-\tilde n_2}{r'}(-1)^{r'}\sinh^{r'}\eta\cos^{r'}\theta e^{2ir'\phi}\cosh^{n_2'-\tilde n_2-r'}\eta\cos^{n_2'-\tilde n_2-r'}\psi,$$

the integral in the right-hand side of (45) is equal to

$$\sum_{r=0}^{m_2'-\tilde m_2}\sum_{r'=0}^{n_2'-\tilde n_2}\binom{m_2'-\tilde m_2}{r}\binom{n_2'-\tilde n_2}{r'}(-1)^{r+r'}(\sinh\eta)^{r+r'}(\cosh\eta)^{m_2'+n_2'-\tilde m_2-\tilde n_2-r-r'}\int_0^{\pi/2}\int_0^{\pi/2}\int_0^{\pi}d\theta\, d\psi\, d\phi\,(\sin\theta)^{2(p-m_2')-3}$$
$$\times(\cos\theta)^{2m_2'-m_2-\tilde m_2-r+r'+1}(\sin\psi)^{2(q-n_2')-3}(\cos\psi)^{2n_2'-n_2-\tilde n_2+r-r'+1}e^{2i(m_2-\tilde m_2+r'-r)\phi}. \qquad (46)$$

The relations (45) and (46) define completely the matrix elements for the finite dimensional irreducible representations of U(p,q) with highest weights (0,...,0,λ). To obtain the matrix elements of the corresponding representations of U(p,q) we have to make an analytic continuation in η:$\eta\to i\theta$. But this continuation does not transform the matrix a_η [see (16)] into a matrix of a rotation in the plane ($p,p+1$). These rotation matrices can be obtained in the following manner:

$$\hat s^{-1}\begin{pmatrix}\cosh\eta\sinh\eta\\ \sinh\eta\cosh\eta\end{pmatrix}\hat s=\begin{pmatrix}\cosh\eta\sinh\eta\\ -i\sinh\nu\cosh\nu\end{pmatrix}\xrightarrow{\eta\to i\theta}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sinh\theta & \cos\theta\end{pmatrix},$$

where

$$\hat s=\begin{pmatrix}1&0\\0&i\end{pmatrix}.$$

We can consider the action by $\hat{s}$ as an isomorphism pf $\mathrm{U}(p,q)$ onto a group $\mathrm{U}'(p,q)$. Hence, we can consider the following representations of $\mathrm{U}'(p,q)$:

$$\mathrm{U}'(p,q)\ni g' \equiv \hat{s}^{-1} g \hat{s} \rightarrow T_g, \quad g \in \mathrm{U}(p,q),$$

where T_g are the operators of finite dimensional representations of $\mathrm{U}(p,q)$. Thus, the formulas (45) and (46) define the matrix elements of the operators of the representation of $\mathrm{U}'(p,q)$ corresponding to the element

$$\begin{pmatrix} \cosh\eta & i\sinh\eta \\ -i\sinh\eta & \cosh\eta \end{pmatrix}.$$

Then an analytic continuation $\eta \rightarrow i\theta$ in (45) and (46) leads to the matrix elements of the representations of $\mathrm{U}(p+q)$ with highest weights $(0,\ldots,0,\lambda)$. Since $\cosh\eta \rightarrow \cos\theta$, $\sinh\eta \rightarrow i\sin\theta$ for $\eta \rightarrow i\theta$, we obtain for these matrix elements

$$\begin{aligned} & d_{(mn)(\tilde{m}\tilde{n})(m'n')}(\theta) \\ & = \frac{(p-1)(q-1)}{\pi} \frac{[(\dim[m]\times[n])(\dim[\tilde{m}]\times[\tilde{n}])]^{1/2}}{\dim[m']\times[n']} \\ & \times N(0,m_2,0,m_2',0) N(0,\tilde{m}_2,0,m_2',0) \\ & \times N(0,n_2,0,n_2',0) N(0,\tilde{n}_2,0,n_2',0) \\ & \times \sum_{r=o}^{m_2'-\tilde{m}_2} \sum_{r'=0}^{n_2'-\tilde{n}_2} \binom{m_2'-\tilde{m}_2}{r} \binom{n_2'-\tilde{n}_2}{r'} \\ & \times (-1)^{r+r'} B\left(p - m_2' - 1, m_2' - \frac{m_2+\tilde{m}_2+r-r'}{2} + 1\right) \\ & \times B\left(q - n_2' - 1, n_2' - \frac{n_2+\tilde{n}_2-r+r'}{2} + 1\right) \\ & \times i^{r+r'} (\sin\theta)^{r+r'} (\cos\theta)^{m_2'+n_2'-\tilde{m}_2-\tilde{n}_2-r-r'} \\ & \times \delta_{m_2+r',\tilde{m}_2+r}, \end{aligned} \tag{47}$$

where $B(.,.)$ is a beta function and δ is a Kronecker symbol. In this formula terms appear which are imaginary. In order to avoid their appearance, we have to introduce a new basis $|m,n\rangle' = i^{m_2}|m,n\rangle$, where $|m,n\rangle$ is the basis in which we have evaluated the matrix elements (47). The new basis leads to a multiplication of the matrix elements (47) by the factor $i^{m_2-\tilde{m}_2}$. Thus, on the right-hand side we have $i^{m_2-\tilde{m}_2+r+r'}$. Due to a Kronecker symbol it holds $m_2-\tilde{m}_2+r+r'=2r$. Therefore, in the new basis we have to replace the factor $i^{r+r'}$ in (47) by $(-1)^r$. Taking into account that $m_2+n_2 = \tilde{m}_2+\tilde{n}_2$, the matrix elements (47) in the basis $|m,n\rangle$ can be written as

$$\begin{aligned} & d_{(mn)(\tilde{m}\tilde{n})(m'n')}(\theta) = M(m,n,\tilde{m},\tilde{n},m',n') \\ & \times \sum_{r=\max(0,m_2-\tilde{m}_2)}^{\min(m_2'-\tilde{m}_2,n_2'-n_2)} \binom{m_2'-\tilde{m}_2}{r} \binom{n_2'-\tilde{n}_2}{\tilde{m}_2-m_2+r} \\ & \times (-1)^{\tilde{m}_2-m_2+r} B(p-m_2'-1, m_2'-m_2+1) \\ & \times B(q-n_2'-1, n_2'-n_2+1)(\sin\theta)^{2r+\tilde{m}_2-m_2} \\ & \times (\cos\theta)^{m_2'-\tilde{m}_2+n_2'-n_2-2r}, \end{aligned} \tag{47'}$$

where M denotes the numerical coefficient which preceeds the sum on the right-hand side of (47).

The matrix elements (47') for the representations of $\mathrm{U}(p+q)$ with the highest weight $(0,\ldots,0,\lambda)$ constitute a representation matrix which is not unitary. To obtain matrix elements in a unitary form it is necessary to introduce the new basis elements $|m,\alpha,n,\beta\rangle''$ by means of the formula

$$|m,\alpha,n,\beta\rangle'' = a(m,n)^{-1/2}|m,\alpha,n,\beta\rangle,$$

where the $a(m,n)$ are given by the formulas (5)–(8) of Ref. 2. Therefore, in unitary form the matrix elements for the representations of $\mathrm{U}(p+q)$ with highest weights $(0,\ldots,0,\lambda)$ have the form

$$D_{(mn)(\tilde{m}\tilde{n})(m'n')}(\theta) = \frac{a(\tilde{m},\tilde{n})^{1/2}}{a(m,n)^{1/2}} d_{(mn)(\tilde{m}\tilde{n})(m'n')}(\theta), \tag{48}$$

Since the operator $\hat{\pi}_g$ of the representation is contragradient to the representation π_g is defined by the relation $\hat{\pi}_g = \pi^T_{g^{-1}}$, where T is a transposition, it follows that the matrix elements $\tilde{D}_{(\mathbf{m},\mathbf{n})(\tilde{\mathbf{m}}\tilde{\mathbf{n}})(\mathbf{m}'\mathbf{n}')}(\theta)$ for the representations of $\mathrm{U}(p+q)$ with the highest weight $(-\lambda,0,\ldots,0)$, $\lambda \leqslant 0$ [this representation is contragradient to the representation with the highest weight $(0,\ldots,0,\lambda)$] are defined by the formula

$$\tilde{D}_{(\mathbf{mn})(\tilde{\mathbf{m}}\tilde{\mathbf{n}})(\mathbf{m}'\mathbf{n}')}(\theta) = D_{(mn)(\tilde{m}\tilde{n})(m'n')}(\theta).$$

Here $D\cdots(\theta)$ is defined by (48) [under the condition that $d\cdots(\theta)$ is real], and $m = (-m_2,0,\ldots,0)$ if $m = (0,\ldots,0,m_2)$. The same relations hold for $\tilde{\mathbf{m}},\mathbf{n},\tilde{\mathbf{n}},\mathbf{m}',\mathbf{n}'$.

The coefficient $(a(\tilde{m},\tilde{n})/a(m,n))^{1/2}$ in (48) is defined by the formula

$$(a(\tilde{m},\tilde{n})/a(m,n))^{1/2} = \prod_{j=o}^{\tilde{m}_2-m_2-1} \frac{(-\lambda+m_2+q+j)^{1/2}}{(m_2-p+j+1)^{1/2}},$$

if $\tilde{m}_2 > m_2$, and by the formula

$$(a(\tilde{m},\tilde{n})/a(m,n))^{1/2} = \prod_{j=0}^{m_2-\tilde{m}_2-1} \frac{(\tilde{m}_2-p+j+1)^{1/2}}{(-\lambda+\tilde{m}_2+q+j)^{1/2}}$$

if $\tilde{m}_2 < m_2$.

[1]A. U. Klimyk and B. Gruber, J. Math. Phys. **20**, 1995 (1979).

[2]A. U. Klimyk and B. Gruber, J. Math. Phys. **20**, 2011 (1979). Note that in Eq. (10) two signs need to be changed. In the line containing the equality sign, in the first factor beneath the square root, the plus sign following p must be changed into a minus sign (i.e., $p+1 \rightarrow p-1$). In the following line, in the second factor beneath the square root, the plus sign following q must be changed into a minus sign (i.e., $p+q+2 \rightarrow p+q-2$).

[3]G. Warner, *Harmonic Analysis on Semisimple Lie Groups* (Springer, Berlin, 1972), Vol. 1.

[4]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN, Warsaw, 1977).

[5]A. U. Klimyk, *Matrix Elements and Clebsch-Gordan Coefficients of Group Representations* (Naukova Dumka, Kiev, 1979) (in Russian).

[6]R. Raczka and J. Fisher, Commun. Math. Phys. **3**, 233 (1966); **4**, 8 (1967).

[7]A. Kilhberg, V. F. Müller, and F. Halbwachs, Commun. Math. Phys. **3**, 194 (1966).

[8]P. I. Holod, Dokl. Akad. Nauk Ukr. SSR, Ser. A **N12**, 1062 (1977).

[9]B. Speh, J. Funct. Anal. **33**, 95 (1979).

[10]A. U. Klimyk and A. M. Gavrilik, J. Math. Phys. **20**, 1624 (1979).

[11]N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, Rhode Island, 1968).

[12]N. Ja. Vilenkin, Sb. Nauchn. Tr. Mosk. Gos. Zaochn. Ped. Inst. **39**, 77 (1974) (in Russian).

Congruence classes of finite representations of simple Lie superalgebra[a)]

F. W. Lemire[b)]
Department of Mathematics, University of Windsor, Windsor, Ontario N9B 3P4, Canada

J. Patera[c)]
Centre de Récherches de Mathématiques Appliquées, Université de Montréal, Casepostale 6128, Succursale "A", Montréal, Québec, Canada

(Received 15 December 1981; accepted for publication 26 March 1982)

The concept of congruence of representations of Lie algebras is generalized and applied to the finite-dimensional representations of Lie superalgebras.

PACS numbers: 02.20.Sv, 02.20.Qs

I. INTRODUCTION

The SU(3) triality number introduced in Ref. 1 is a practical tool which simplifies the task of decomposing the tensor product of finite-dimensional, irreducible SU(3) representations. This concept has been generalized[2] to provide an equally useful equivalence relation (called congruence) on the finite-dimensional representations for each of the simple Lie groups.

The purpose of this paper is to point out that the congruence class concept is actually far more general than has been used to date. First, we provide a unified presentation of this concept which applies to a large class of representations (finite-dimensional or not) of both Lie algebras and Lie superalgebras. Secondly, we present explicit formulas for the labels of congruence classes of finite-dimensional representations of simple Lie superalgebras and illustrate their use.

The classification of finite-dimensional, irreducible representations of the simple Lie superalgebras has been completed by Kac.[3] In this classification Kac points out the existence of "typical" and "atypical" irreducible representations. The typical ones are direct summands in any representation in which they appear. The existence of atypical irreducible representations is related to the fact[4] that finite dimensional representations of simple Lie superalgebras are not completely reducible in general. This of course means that the decomposition of tensor products of such representations is much more difficult.[5] The congruence labels for these representations, having the same properties as those for Lie algebra representations with respect to tensor products, provide an easy-to-use tool to simplify this task. More precisely, the label of each summand of the decomposition must be equal to the sum of the labels of the two factors of the tensor product. Clearly the utility of this concept increases in proportion to the number of congruence classes which exist for a given algebra. In general we observe that there are far more congruence classes for finite-dimensional representation of Lie superalgebra than for the Lie algebra cases.

Section II of this paper is concerned with setting up the notation and general properties of the congruence relation. In Sec. III we provide explicit results on the congruence labels for finite-dimensional representations of simple Lie superalgebras and give some examples.

Let L denote a simple Lie (super) algebra over the complex number C with fixed Cartan superalgebra H.

II. NOTATIONS AND DEFINITIONS

A representation $\rho: L \to \mathrm{gl}(V)$ of L is said to be *H-integral* if and only if the representation space $V = \oplus_{\lambda\in H^*} V_\lambda$ where $V_\lambda = \{v\in V \mid \rho(h)v = \lambda(h)v \text{ for all } h\in H\}$ and if for any two weights λ_1 and λ_2 the difference $\lambda_1 - \lambda_2$ is an integral linear combination of simple roots of L. If Δ_1 and Δ_2 are two sets of simple roots of L, then each $\alpha\in\Delta_1$ can be expressed as an integral linear combination of the simple roots of Δ_2. Thus the definition of an H-integral representation is independent of the choice of simple roots for L. In particular, it is clear that every finite-dimensional, irreducible or more generally every indecomposable representation of L which admits a weight space decomposition with respect to H is H-integral. The class of H-integral representations is an extremely general one. Although there are indecomposable (even irreducible) representations of Lie (super) algebras which are not H-integral,[5,6] they have never appeared in any application so far.

Two H-integral representations $\rho_i: L \to \mathrm{gl}(V_i)$ for $i = 1,2$ of L are said to be *congruent* if and only in the representation $V_1 \oplus V_2$ is H-integral. Clearly a necessary and sufficient condition for this is that the difference of any two weights of $V_1 \oplus V_2$ is an integral linear combination of simple roots. The relation of congruence clearly provides an equivalence relation on the class of all H-integral representations of L. In order to make use of this equivalence relation to distinguish representations, we now indicate how one can label the equivalence classes of H-integral representations which can occur for a given simple Lie (super) algebra.

Let $\Delta = \{\alpha_1,...,\alpha_n\}$ be a fixed set of simple roots of L and denote by C the Cartan matrix of L associated with Δ. For the case of classical simple Lie superalgebras we use the conventions introduced by Kac.[3] Since the Killing form is nondegenerate on H^*, the matrix C is invertible. Set $C^{-1} = (g_{ij})$ and define $\lambda_i = \Sigma_{j=1}^n g_{ij}\alpha_j \in H^*$. Clearly $\{\lambda_1,...,\lambda_n\}$ forms a base of H^* which is "dual" to the base $\{\alpha_1,...,\alpha_n\}$, that is, $\langle\lambda_i, \alpha_j\rangle = \delta_{ij}$, where $\langle\cdot,\cdot\rangle$ is the scalar product on H^*. If V is an H-integral representation of L having $\Lambda = \Sigma_{i=1}^n a_i\lambda_i$ as a weight—i.e., $V_\Lambda \neq \{0\}$—then we define the congruence label of V to be the row vector $c(\rho,V)$ given by

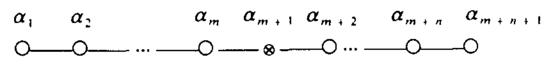
$$c(\rho,V) = (a_1 \cdots a_n)C^{-1} \quad \mathrm{mod}\ Z\times\cdots\times Z, \tag{1}$$

where Z denotes the integers. For certain classes of representations, in particular for finite-dimensional ones, the congruence classes are completely determined by a strict subset of the components of $c(\rho,V)$. For Lie algebras see Ref. 2 and

[a)] Work supported in part by the Ontario-Québec Exchange Programme, The Natural Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

 0022-2488/82/081409-06$02.50

TABLE I. Numbering of simple roots, Cartan matrix, and inverse of $A(m, n)$.



Cartan Matrix

$$\left(\begin{array}{cccccc|c|cccc}
2 & -1 & 0 &&&&&&&& \\
-1 & 2 & -1 &&&&&&&& \\
0 & -1 & 2 &&&&&&&& \\
&&& \ddots &&&&&&& \\
&&&& & 2 & -1 & 0 &&& \\ \hline
&&&& & -1 & 0 & 1 &&& \\ \hline
&&&& & 0 & -1 & 2 &&& \\
&&&&&&&& \ddots && \\
&&&&&&&&& 2 & -1 \\
&&&&&&&&& -1 & 2
\end{array}\right)$$

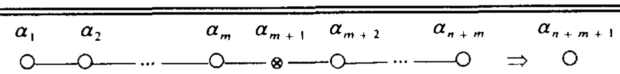
Inverse

$$\frac{1}{n-m}\left(\begin{array}{ccccc|c|cccc}
n-m+1 & n-m+2 & n-m+3 & \cdots & n & n+1 & -n & -n+1 & \cdots & -1 \\
n-m+2 & 2(n-m+2) & 2(n-m+3) & \cdots & 2n & 2(n+1) & 2(-n) & 2(-n+1) & \cdots & -2 \\
n-m+3 & 2(n-m+3) & 3(n-m+3) & \cdots & 3n & 3(n+1) & 3(-n) & 3(-n+1) & \cdots & -3 \\
\vdots & \vdots & \vdots & \vdots &&&&&& \\
n & 2n & 3n & \cdots & mn & m(n+1) & m(-n) & m(-n+1) & \cdots & -m \\ \hline
n+1 & 2(n+1) & 3(n+1) & \cdots & m(n+1) & (m+1)(n+1) & (m+1)(-n) & (m+1)(-n+1) & \cdots & -m-1 \\ \hline
n & 2n & 3n & \cdots & mn & (m+1)(n) & m(-n) & m(-n+1) & \cdots & -m \\
n-1 & 2(n-1) & 2(n-1) & \cdots & m(n-1) & (m+1)(n-1) & m(1-n) & (m-1)(-n+1) & \cdots & -m+1 \\
\vdots & \vdots & \vdots && \vdots & \vdots & \vdots &&& \\
1 & 2 & 3 & \cdots & m & m+1 & -m & -m+1 & \cdots & n-m-1
\end{array}\right)$$

for Lie superalgebras, see Table VIII. Since any other weight Λ' of the H-integral representation (ρ, V) is of the form $\Lambda' = \Lambda + \Sigma_{i=1}^{n} n_i\alpha_i$, where $n_i \in Z$, and the matrix C^{-1} transforms coordinates with respect to $\{\lambda_1,...,\lambda_n\}$ to coordinates with respect to $\{\alpha_1,...,\alpha_n\}$, it follows that the congruence vector of an H-integral representation is independent of the choice of the weight Λ. Finally it is immediate from the foregoing arguments that two H-integral representations of L are congruent if and only if their congruence labels coincide. The Cartan matrices and their inverses for simple Lie algebras are given in Ref. 7, for simple Lie superalgebras they are shown in Tables I–VI.

It is useful to point out an equivalent approach to the classification of H-integral representations of simple Lie (su-

TABLE II. Numbering of simple roots, Cartan matrix, and inverse for $B(n, m+1)$, $n > 0$.



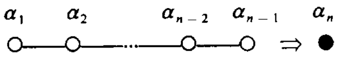
Cartan matrix

$$\left(\begin{array}{cccccc|c|ccccc}
2 & -1 & 0 &&&&&&&&& \\
-1 & 2 & -1 &&&&&&&&& \\
0 & -1 & 2 &&&&&&&&& \\
&&& \ddots &&&&&&&& \\
&&&& & 2 & -1 & 0 &&&& \\ \hline
&&&& & -1 & 0 & 1 &&&& \\ \hline
&&&& & 0 & -1 & 2 &&&& \\
&&&&&&&& \ddots &&& \\
&&&&&&&&& 2 & -1 & 0 \\
&&&&&&&&& -1 & 2 & -1 \\
&&&&&&&&& 0 & -2 & 2
\end{array}\right)$$

Inverse

$$\frac{1}{2}\left(\begin{array}{cccc|c|ccc}
2 & 2 & 2 & 2 & 2 & -2 & -2 & -1 \\
2 & 4 & 4 & 4 & 4 & -4 & -4 & -2 \\
2 & 4 & 6 & 6 & 6 & -6 & -6 & -3 \\
2 & 4 & 6 & 2m & 2m & -2m & -2m & -m \\ \hline
2 & 4 & 6 & 2m & 2(m+1) & -2(m+1) & -2(m+1) & -(m+1) \\ \hline
2 & 4 & 6 & 2m & 2(m+1) & -2m & -2m & -m \\
2 & 4 & 6 & 2m & 2(m+1) & -2m & 2(n-m-2) & n-m-2 \\
2 & 4 & 6 & 2m & 2(m+1) & -2m & 2(n-m-2) & (n-m-1)
\end{array}\right)$$

TABLE III. Numbering of simple roots, Cartan matrix, and inverse for $B(0,n)$.



Cartan matrix

$$\left[\begin{array}{cccccc|c} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & & & & \\ 0 & -1 & 2 & & & & \\ & & & \ddots & & & \\ & & & & & 2 & -1 \\ \hline & & & & & -2 & 2 \end{array}\right]$$

Inverse

$$\frac{1}{2}\left[\begin{array}{ccccc|c} 2 & 2 & 2 & & 2 & 1 \\ 2 & 4 & 4 & & 4 & 2 \\ 2 & 4 & 6 & & 6 & 3 \\ & & & \ddots & & \\ 2 & 4 & 6 & & 2(n-1) & n-1 \\ \hline 2 & 4 & 6 & & 2(n-1) & n \end{array}\right]$$

TABLE IV. Numbering of simple roots, Cartan matrix, and inverse for $C(n)$.

Cartan matrix

$$\left[\begin{array}{c|cccccc} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \hline -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & -2 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 \end{array}\right]$$

Inverse

$$\left[\begin{array}{c|cccccc} 2 & -2 & -2 & -2 & \cdots & -2 & -2 \\ \hline 2 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 2 & 0 & 2 & 2 & \cdots & 2 & 2 \\ 2 & 0 & 2 & 4 & \cdots & 4 & 4 \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ 2 & 0 & 2 & 4 & \cdots & 2(n-2) & 2(n-2) \\ 1 & 0 & 1 & 2 & \cdots & n-2 & n-1 \end{array}\right]$$

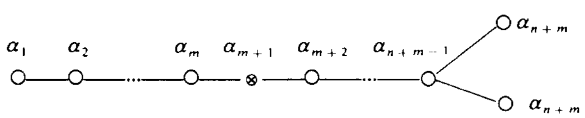
per) algebras suggested by I. Kaplansky. With the above notation we define $W = \{\Sigma_{i=1}^{n} a_i\lambda_i \,|\, a_i \in \mathbb{C}\}$ and consider W as an additive abelian group. Let R denote the subgroup of W generated by the simple roots $\Delta = \{\alpha_1,...,\alpha_n\}$ of L. Since, for any $\lambda \in H^*$, there exists an H-integral representation of L admitting λ as a weight, it is immediate that there is a one-one correspondence between the elements of the quotient group W/R and the congruence classes of H-integral representations of L.

The three operative properties of the congruence relation readily follow. First, if V_1 and V_2 are H-integral representations of L, then $V_1 \otimes V_2$ is again H-integral and its congruence vector is equal to the componentwise sum of the congruence vectors of the two factors

$$c(V_1 \otimes V_2) = c(V_1) + c(V_2), \quad \mathrm{mod} Z \times \cdots \times Z. \tag{2}$$

Secondly, if V is an H-integral representation of L and $V = V_1 \oplus \cdots \oplus V_k$, then the congruence vector of each summand is equal to the congruence vector of V. Finally, if L' is a semisimple subalgebra of L such that its adjoint representa-

TABLE V. Numbering of simple roots, Cartan matrix, and inverse for $(D(n, m+1)$.



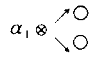
$$\left[\begin{array}{cccc|c|ccccc} 2 & -1 & 0 & & & & & & & \\ -1 & 2 & -1 & & & & & & & \\ 0 & -1 & 2 & & & & & & & \\ & & & 2 & -1 & 0 & & & & \\ \hline & & & -1 & 0 & 1 & & & & \\ \hline & & & 0 & -1 & 2 & \ddots & & & \\ & & & & & & & 2 & -1 & -1 \\ & & & & & & & -1 & 2 & 0 \\ & & & & & & & -1 & 0 & 2 \end{array}\right]$$

Inverse

$$\frac{1}{4}\left[\begin{array}{cccc|c|cccc} 4 & 4 & 4 & 4 & 4 & -4 & -4 & -2 & -2 \\ 4 & 8 & 8 & 8 & 8 & -8 & -8 & -4 & -4 \\ 4 & 8 & 12 & 12 & 12 & 12 & -12 & -6 & -6 \\ 4 & & & 4m & 4m & -4m & & & -2m \\ \hline 4 & & & 4m & 4(m+1) & -4(m+1) & & & -2(m+1) \\ \hline 4 & & & 4m & 4(m+1) & -4m & & & -2m \\ 4 & 8 & 12 & 4m & 4(m+1) & -4m & 4(n-m-3) & 2(n-m-3) & 2(n-m-3) \\ 2 & 4 & 6 & 2m & 2(m+1) & -2m & 2(n-m-3) & 2(n-m-2) & (n-m-3) \\ 2 & 4 & 6 & 2m & 2(m+1) & -2m & 2(n-m-3) & (n-m-3) & (n-m-1) \end{array}\right]$$

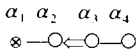
TABLE VI. Numbering of simple roots, Cartan matrices, and inverses for the exceptional simple lie superalgebras.

$D(2,1;\alpha)$



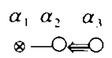
$$\left[\begin{array}{c|cc} 0 & 1 & \alpha \\ \hline -1 & 2 & 0 \\ -1 & 0 & 2 \end{array}\right], \qquad \frac{1}{2(1+\alpha)}\left[\begin{array}{c|cc} 4 & -2 & -2\alpha \\ \hline 2 & \alpha & -\alpha \\ 2 & -1 & 1 \end{array}\right]$$

F_4



$$\left[\begin{array}{c|ccc} 0 & 1 & 0 & 0 \\ \hline -1 & 2 & -2 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{array}\right], \qquad \frac{1}{3}\left[\begin{array}{c|ccc} 2 & -3 & -4 & -2 \\ \hline 3 & 0 & 0 & 0 \\ 2 & 0 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{array}\right]$$

G_3



$$\left[\begin{array}{c|cc} 0 & 1 & 0 \\ \hline -1 & 2 & -3 \\ 0 & -1 & 2 \end{array}\right], \qquad \frac{1}{2}\left[\begin{array}{c|cc} 1 & -2 & -3 \\ \hline 2 & 0 & 0 \\ 1 & 0 & 1 \end{array}\right]$$

tion on L reduces to H'-integral representations, then every H-integral representation of L is an H'-integral representation on L'.

When these properties of the congruence relation are combined with other simple invariants of representations such as dimension, index,[8] etc., it provides a useful and computationally practical tool in determining the decomposition of tensor products and certain branching rules.

Applications of congruence to the analysis of finite-dimensional representations of simple Lie algebras has been discussed in a previous paper.[2] We observe now that in its present generality the concept of congruence of representations is applicable to the analysis of H-integral infinite-dimensional representations of simple Lie algebras. The main difficulty in illustrating such applications is that the general classification of indecomposable infinite-dimensional representations of simple algebras is far from complete. For example, the tensor product of two infinite-dimensional irreducible representations of A_1 having double infinite strings of weight spaces contains no nonzero eigenvectors of the Casimir operator of A_1 and hence no irreducible subrepresentations. Nevertheless, each indecomposable representations in this decomposition must be H-integral, and their congruence label must be equal to the sum of the congruence labels of the two factors.

III. CONGUENCE CLASSES FOR SIMPLE LIE SUPERALGEBRAS

In Tables I–VI we list Cartan matrices and their inverses for the simple Lie superalgebras. It should be noted that, unlike the usual Lie algebra cases, the Cartan matrix for a simple Lie superalgebras depends on the choice of simple roots in the root system. Throughout this section we use the special sets of simple roots specified by Kac.[3] The particular choice has the properties of being similar to the simple roots of the corresponding Lie algebras and containing exactly one odd root.

For all $\Lambda \in H^*$ which satisfy certain conditions listed below, there exist finite-dimensional irreducible representations having highest weight Λ. Conversely, every finite-dimensional irreducible representation has a highest weight which satisfies these conditions. For convenience we recall these conditions. If $\Lambda \in H^*$, then there exists a finite-dimensional irreducible representation having the highest weight $\Lambda = \Sigma a_i \lambda_i$ if and only if the a_i's satisfy the following conditions[3]:

(1) $a_i \in Z_+$ for all *even* simple roots α_i;

(2) the linear combination k of the a_i's given in Table VIIA is a nonnegative integer;

(3) if the value of k is strictly less than b given in Table VIIA, then the a_i's must satisfy the additional conditions of Table VIIB.

Analyzing the range of coordinates of the congruence vectors (1) for finite-dimensional irreducible representations of simple Lie superalgebras, we find that the congruence class is completely determined by a subset of the coordinates of $c(\rho, V)$. Therefore, we introduce a new congruence vector $C(\rho, V)$ whose components are given in Table VIII in terms of the a_i's for all simple Lie superalgebras.

One should observe from this list that there are infinitely many congruence classes for the finite-dimensional irreducible representations of the Lie superalgebras $A(m,n)$, $C(n)$, and $D(2,1;\alpha)$ (for some α). The other algebras admit only finitely many congruence classes.

Let us illustrate the use of the congruence concept with some examples. First, we observe that the even subalgebra of any simple Lie superalgebras is an integral subalgebra. This follows since the Cartan subalgebra of the Lie superalgebra

TABLE VIIA. Linear combinations k and constants b for simple Lie superalgebras G.

G	k	b
$B(0,n)$	$a_n/2$	0
$B(m,n)$, $m>0$	$a_n - a_{n+1} - \cdots - a_{m+n-1} - \frac{1}{2}a_{m+n}$	m
$D(m,n)$	$a_n - a_{n+1} - \cdots - a_{m+n-2} - \frac{1}{2}(a_{m+n-1} + a_{m+n})$	m
$D(2,1;\alpha)$ $\alpha \in C, \alpha \neq 0, -1$	$(1+\alpha)^{-1}(2a_1 - a_2 - \alpha a_3)$	2
$F(4)$	$\frac{1}{3}(2a_1 - 2a_2 - 4a_3 - 2a_4)$	4
$G(3)$	$\frac{1}{2}(a_1 - 2a_2 - 3a_3)$	4

TABLE VIIB. Supplementary conditions applicable when $k < b$.

G	Conditions on a_i's
$B(m,n)$	$a_{n+k+1} = \cdots = a_{m+n} = 0$
$D(m,n)$	$a_{n+k+1} = \cdots = a_{m+n} = 0$ for $k \leqslant m-2$
	$a_{m+n-1} = a_{m+1}$ for $k = m-1$
$D(2,1;\alpha)$, $\alpha \in C$	all $a_i = 0$ for $k = 0$
$\alpha \neq 0,1$	$(a_3+1)\alpha = \pm(a_2+1)$ for $k = 1$
F_4	all $a_i = 0$ for $k = 0$
	$k \neq 1$
	$a_2 = a_4 = 0$ for $k = 2$
	$a_2 = 2a_4 + 1$ for $k = 3$
$G(3)$	all $a_i = 0$ for $k = 0$
	$k \neq 1$
	$a_2 = 0$ for $k = 2$

is, by convention, taken to be the Cartan subalgebra of its even part. Thus, if we take any finite-dimensional indecomposable representation of the simple Lie superalgebra L and reduce it as a representation of the even subalgebra L_0, all of its direct summands must be congruent as L_0 representations. This property of congruence has been used implicity by Hurni and Morel[9] to provide an explicit description of the finite-dimensional irreducible representations of SU(M/N) and also by Marcu[10] in his classification of all finite-dimensional indecomposable representations of spl(2,1).

The concept of congruence also provides an aid in determining the summands in the decomposition of a tensor product. For example, using the notation previously introduced, the finite-dimensional irreducible representation of $A(1,0)$ are uniquely determined by the value of their highest weight $\Lambda = a_1\lambda_1 + a_2\lambda_2$. In fact, for $\Lambda \in H^*$ with $a_1 \in Z_+$, there exists a finite-dimensional irreducible representation V_Λ having Λ as its highest weight—we label this representa-

TABLE VIII. Congruence vectors $C(\rho, V) = (c_1, c_2, \cdots)$ for finite-dimensional irreducible representations of simple Lie superalgebras given as linear combinations of the coordinates of the highest weight $\Lambda = \Sigma a_i\lambda_i$.

G	Congrence class determined by
$A(m,n)$	$c_1 = a_1 + 2a_2 + \cdots + ma_m + na_{m+2} + \cdots + a_{m+n+1}$, mod $m-n$ $c_2 = [(n+1)/(m-n)]a_{m+1}$, mod Z $c_3 = [(m+1)/(m-n)]a_{m+1}$, mod Z
$B(0,n)$	all finite representations are congruent
$B(n,m+1)$ $n>0$	$c_1 = a_{m+n+1}$, mod 2 $c_2 = \sum_{i=0,1,\cdots} a_{2i+1} - \frac{1}{2}(1-(-1)^{m+1})a_{m+1}$, mod 2
$C(n)$	$c_1 = a_1$, mod Z $c_2 = a_n$, mod 2
$D(n,m+1)$	$c_1 = a_{m+n} + a_{m+n+1}$, mod 2 $c_2 = 2a_1 + 4a_2 + \cdots + 2(m+1)a_{m+1} + 2ma_{m+2} + \cdots + 2(m-n+2)a_{m+n} + (m-n+3)a_{m+n+1}$, mod 4 $c_3 = 2a_1 + 4a_2 + \cdots + 2(m+1)a_{m+1} + 2ma_{m+2} + \cdots + (m-n+3)a_{m+n} + (m-n+1)a_{m+n+1}$, mod 4
$D(2,1;\alpha)$, $\alpha \in C$, $\alpha \neq 0, -1$	$c_1 = [1/(1+\alpha)](2a_1 + a_2 + a_3)$ $c_2 = [1/2(1+\alpha)](-2a_1 + \alpha a_2 - a_3)$ $c_3 = [1/2(1+\alpha)](-2\alpha a_1 - \alpha a_2 + a_3)$
$F(4)$	$c_1 = 2a_1$, mod 2
$G(3)$	all finite representations are congruent

tion by the pair (a_1,a_2). With this notation we have that

$$(0,2i+\tfrac{1}{2})\otimes(1,i+1)$$
$$=(1,3i+\tfrac{3}{2})\oplus(2,3i+\tfrac{3}{2})\oplus(0,3i+\tfrac{5}{2})\oplus(1,3i+\tfrac{5}{2}),$$
$$4\times 8 = \quad 8 \quad + \quad 12 \quad + \quad 4 \quad + \quad 8,$$
$$(2i+\tfrac{1}{2})+(i+1)$$
$$=3i+\tfrac{3}{2}=3i+\tfrac{3}{2}=3i+\tfrac{5}{2}=3i+\tfrac{5}{2}, \quad \mathrm{mod}\ Z..$$

The first equation describes the tensor product and its decomposition, the second equation gives the corresponding dimension of the representations, and the final equation provides the appropriate label for the congruence class of each representation. We observe the congruence label of each summand in the decomposition is equal to the sum of the congruence labels of the two factors. This property along with others has in fact been used by Marcu[5] to provide a graphical scheme for decomposing the tensor product of any two indecomposable representations of $A(1,0)$.

ACKNOWLEDGMENT

The authors are grateful for the hospitality of the Aspen Center for Physics. The main part of this work was completed during our stay at the Center (July 1981).

[1]G. E. Baird and L. C. Biedenharn, *Symmetry Principles at High Energy*, edited by B. Korsunoglu and A. Perlmutter (Freeman, San Francisco, 1964).
[2]F. Lemire and J. Patera, J. Math. Phys. **21**, 2026 (1980).
[3]V. Kac, Adv. Math. **26**, 8 (1977).
[4]D. Djoković and G. Hochschild, Ill. J. Math. **20**, 134 (1976).
[5]M. Marcu, J. Math. Phys. **21**, 1284 (1980).
[6]D. Arnal and G. Pinczon, C. R. Acad. Sci. **272**, 1369 (1971).
[7]E. Dynkin, A. M. S. Transl., Ser. 2, **6**, 111 (1957), Table 2.
[8]J. McKay, J. Patera, and R. T. Sharp, J. Math. Phys. **22**, 2770 (1981).
[9]J. P. Hurni and B. Morel, " Irreducible Representations of SU(M/N)," J. Math. Phys. (to be published).
[10]M. Marcu,J. Math. Phys. **21**, 1277 (1980).

Supersymmetry and Lie algebras

R. Y. Levine and Y. Tomozawa
Department of Physics, University of Michigan, Ann Arbor, Michigan 48109

(Received 1 December 1981; accepted for publication 5 February 1982)

Starting from the standard supersymmetry algebra, an infinite Lie algebra is constructed by introducing commutators of fermionic generators as members of the algebra. From this algebra a finite Lie algebra results for fixed momentum analogous to the Wigner analysis of the Poincaré algebra. It is shown that anticommutation of the fermionic charges plays the role of a constraint on the representation. Also, it is suggested that anticommuting parameters can be avoided by using this infinite Lie algebra with fermionic generators modified by a Klein transformation.

PACS numbers: 02.20.Sv, 11.30.Pb

I. INTRODUCTION

Supersymmetry is unique as a symmetry of nature in that bosons and fermions are grouped together in the same multiplet.[1,2] This feature is essential for the construction of a sensible supergravity[3] theory, but also means that any low energy theory has to incorporate supersymmetry breaking. A deeper understanding of how supersymmetry may arise could certainly shed light on its breaking. The existence of a fermionic charge in supersymmetry requires that the algebra be defined by anticommutation as well as commutation relations. While this allows the evasion of the Coleman–Mandula no go theorem,[4] the resulting algebra is not a Lie algebra and the parameters for infinitesimal transformations are anticommuting numbers (Grassmann variables). This leads naturally to an extension of Minkowski space, known as superspace, in which spinors are attached to each space-time point.[2] With anticommuting parameters, one has the conceptual problem of nilpotent translation parameters. Further, all continuous symmetries in nature have been represented by a Lie algebra. It is therefore natural to ask whether supersymmetry can be represented by a Lie algebra.

In this paper fermionic anticommutation relations are used to construct the commutation relations of an infinite Lie algebra. In this algebra successive multiplication by the momentum operator defines new generators. The algebra thus obtained has both the Wigner representation of the Poincaré algebra and the standard supersymmetry representations. The latter arises when anticommutation of the fermionic charges is used as a constraint. The standard superalgebra requires the parameters of infinitesimal supersymmetry transformations to be anticommuting c-numbers in order to have a finite closed algebra. With our formalism commuting c-numbers close the algebra due to the added generators. However, in order to preserve the spin statistics relationship it is necessary to modify the fermionic generators with a Klein transformation.

In Sec. II, an explicit construction of the infinite Lie algebra is presented. Section III contains the resulting finite Lie algebra for fixed momentum, which is analogous to the Wigner analysis of the Poincaré algebra. This section also contains constructions of massive and massless representations of the finite algebra with the anticommutation constraint. A discussion of the modification of fermionic generators by the Klein transformation is in Sec. IV followed by general discussion in Sec. V.

II. THE INFINITE LIE ALGEBRA

The superalgebra is defined by the following commutation and anticommutation relations[1,2]:

$$\begin{aligned}
&[J_{\mu\nu},J_{\lambda\rho}] = i(\delta_{\mu\lambda}J_{\nu\rho} - \delta_{\mu\rho}J_{\nu\lambda} + \delta_{\nu\rho}J_{\mu\lambda} - \delta_{\nu\lambda}J_{\mu\rho}),\\
&[J_{\mu\lambda},P_\nu] = i(\delta_{\mu\nu}P_\lambda - \delta_{\lambda\nu}P_\mu),\\
&[P_\mu,P_\nu] = 0,\\
&[S^\alpha,P_\mu] = 0,\\
&[S^\alpha,J_{\lambda\nu}] = \tfrac{1}{2}(\sigma_{\lambda\nu})_{\alpha\beta}S^\beta,\\
&\{S^\alpha,S^\beta\} = i(\gamma_\mu C)_{\alpha\beta}P^\mu.
\end{aligned}\tag{2.1}$$

Here $J_{\mu\nu}$ and P_λ are the generators of the Poincaré algebra, $S^\alpha(\alpha = 1,2,3,4)$ are the fermionic Majorana generators of supersymmetry, and C is the charge conjugation matrix $(C^+C = 1,\ C^T = -C,\ C^{-1}\gamma_\mu C = -\gamma_\mu^T)$. In this paper only this simplest algebra is considered although the extension to the case of a fermionic Dirac generator or multiple Majorana generators is trivial. In order to form the Lie algebra we consider the commutator

$$[S^\alpha,S^\beta] \equiv T^{\alpha\beta}.\tag{2.2}$$

If $T^{\alpha\beta}$ was expressible as a linear combination of the generator of the superalgebra we would have

$$T^{\alpha\beta} = a(\gamma_5\gamma_\mu C)^{\alpha\beta}P^\mu,\tag{2.3}$$

where a is a dimensionless number. Note that C, $(\gamma_5 C)$, and $(\gamma_5\gamma_\mu C)$ are antisymmetric and $(\gamma_\mu C)$ and $(\sigma_{\mu\nu}C)$ are symmetric. Using the Jacobi identity (2.4), where

$$[[A,B],C] = \{\{B,C\},A\} - \{\{C,A\},B\}\tag{2.4}$$

$$= -[[B,C],A] - [[C,A],B],\tag{2.5}$$

with $A = S^\alpha$, $B = S^\beta$, and $C = S^\delta$, it is easy to show that (2.3) is inconsistent. Therefore, we conclude that $T^{\alpha\beta}$ is a new generator. Furthermore, there is no consistent second-order operator in the Poincaré algebra that $T^{\alpha\beta}$ could equal. For example,

$$T^{\alpha\beta} = a(\gamma_5\gamma_\mu C)^{\alpha\beta}W_\mu,\tag{2.6}$$

 0022-2488/82/081415-07$02.50

where $W^\mu = -i/2\,\epsilon^{\mu\lambda\nu\rho} J_{\lambda\nu} P_\rho$ is the Pauli–Lubanski vector,[5] is contradicted by the Jacobi identity (2.4) with $A = S^\alpha$, $B = S^\beta$, and $C = S^\delta$. (This is proven with an appropriate Fierz transformation of the γ-matrices.)

Because $T^{\alpha\beta}$ is an independent generator we consider the commutators with $J_{\mu\nu}$, P_μ, S^α, and itself:

$$\begin{aligned}
[T^{\alpha\beta},S^\delta] &= 2iS^\alpha(\gamma_\mu C)_{\beta\delta}P^\mu - 2i(\gamma_\mu C)_{\alpha\delta}P^\mu S^\beta,\\
[T^{\alpha\beta},T^{\eta\delta}] &= 2iT^{\eta\alpha}(\gamma_\mu C)_{\delta\beta}P^\mu - 2iT^{\delta\alpha}(\gamma_\mu C)_{\eta\beta}P^\mu\\
&\quad - 2iT^{\eta\beta}(\gamma_\mu C)_{\delta\alpha}P^\mu + 2iT^{\delta\beta}(\gamma_\mu C)_{\eta\alpha}P^\mu,\\
[T^{\alpha\beta},P_\mu] &= 0,\\
[T^{\alpha\beta},J_{\mu\nu}] &= \tfrac12(\sigma_{\mu\nu})_{\alpha\delta}T_{\delta\beta} - \tfrac12(\sigma_{\mu\nu})_{\beta\delta}T_{\delta\alpha}.
\end{aligned} \tag{2.7}$$

The first of these is constructed using the anticommutation relation in (2.1) and the identity of (2.4). The rest follow from the commutation relations in (2.1) and the identity (2.5). From (2.7) it can be seen that there are new generators $S^\alpha P_\mu$ and $T^{\alpha\beta}P_\mu$. Again the commutators of these new members with all the previous generators and with themselves must be considered. We exhibit these in Appendix A. From these relations one must include as new members of the algebra the operators on the right-hand side of (A1):

$$S^\alpha P_\mu P_\nu, \quad T^{\alpha\beta}P_\mu P_\nu,$$
$$S^\alpha P_\mu P_\nu P_\lambda, \quad T^{\alpha\beta}P_\mu P_\nu P_\lambda.$$

Obviously, there are a finite number of generators of the form $S^\alpha P_\mu \cdots P_\omega$ and $T^{\alpha\beta}P_\mu\cdots P_\omega$ added for each order of commutation. Thus, we obtain an infinite Lie algebra with generators $J_{\mu\nu}$, P_μ, S_α, $T^{\alpha\beta}$, $S^\alpha P_\mu$, $T^{\alpha\beta}P_\mu$,..., $S^\alpha P_\mu\cdots P_\omega$, $T^{\alpha\beta}P_\mu\cdots P_\omega$,... . The added generators are of a geometrical series type and the resulting algebra is called an affine Lie algebra.[6] Note that the Casimir operators of this algebra are identical to the Casimir operators of the superalgebra. Also, it will be shown that the representations of the superalgebra are those representations of the infinite algebra satisfying the anticommutation relation in (2.1) as a constraint. Instead of studying the infinite Lie algebra directly, we will consider in the next section the finite Lie algebra which results from a fixed momentum condition. For this to be consistent we must use the operator W_μ instead of $J_{\mu\lambda}$, since W_μ commutes with P_λ.

Incidently, note with W_μ alone we have the commutators

$$\begin{aligned}
[W^\mu,W^\nu] &= \epsilon^{\mu\nu\lambda\rho}W_\lambda P_\rho,\\
[W^\mu P_\lambda,W^\nu] &= \epsilon^{\mu\nu\kappa\rho}W_\kappa P_\rho P_\lambda,
\end{aligned} \tag{2.8}$$

and so on. This forms an infinite Lie algebra with generators W_μ, $W_\mu P_\lambda$, $W_\mu P_\lambda P_\nu$,... similar to the structure above. A finite SU(2) algebra follows for fixed timelike momentum.It is important to realize that the analysis of the finite algebra for fixed momentum is equivalent to an analysis of the Poincaré algebra.

III. THE FINITE ALGEBRA FOR FIXED MOMENTUM

The algebra generated by W_μ, $T^{\alpha\beta}$, and S^α for fixed momentum is defined by the commutation relations

$$\begin{aligned}
[W^\mu,W^\nu] &= \epsilon^{\mu\nu\lambda\rho}W_\lambda P_\rho,\\
[W^\mu,S^\alpha] &= (i/4)\epsilon^{\mu\lambda\nu\rho}(\sigma_{\lambda\nu})_{\alpha\beta}S^\beta P_\rho,\\
[T^{\alpha\beta},W^\mu] &= (i/4)\epsilon^{\mu\lambda\nu\rho}(\sigma_{\lambda\nu})_{\beta\delta}T^{\delta\alpha}P_\rho\\
&\quad - (i/4)\epsilon^{\mu\lambda\nu\rho}(\sigma_{\lambda\nu})_{\alpha\delta}T^{\delta\beta}P_\rho,\\
[S^\alpha,S^\beta] &= T^{\alpha\beta},\\
[T^{\alpha\beta},T^{\eta\delta}] &= 2iT^{\eta\alpha}(\gamma_\mu C)_{\delta\beta}P^\mu - 2iT^{\delta\alpha}(\gamma_\mu C)_{\eta\beta}P^\mu\\
&\quad - 2iT^{\eta\beta}(\gamma_\mu C)_{\delta\alpha}P^\mu + 2iT^{\delta\beta}(\gamma_\mu C)_{\eta\alpha}P^\mu,\\
[T^{\alpha\beta},S^\delta] &= 2iS^\alpha(\gamma_\mu C)_{\beta\delta}P^\mu - 2i(\gamma_\mu C)_{\alpha\delta}P^\mu S^\beta.
\end{aligned} \tag{3.1}$$

For fixed momentum we have the finite algebra generated by $\{W_\mu, T^{\alpha\beta}, S^\delta\}$ (denoted A_{WTS}) and a subalgebra generated by $\{W_\mu, T^{\alpha\beta}\}$ (denoted A_{WT}). A Casimir operator of the superalgebra (and therefore the infinite algebra) is also a Casimir operator of A_{WTS}. Denoting A_{JPS} to be the superalgebra and $C(A)$ to be the set of Casimir operators of an algebra A we have the inclusion relations

$$C(A_{JPS}) \subset C(A_{WTS}) \subset C(A_{WT}). \tag{3.2}$$

Also, the irreducible representations of A_{JPS} correspond to irreducible representations of A_{WTS} in the same way that the irreducible representations of A_W for fixed momentum correspond to irreducible representations of the Poincaré algebra. For both statements the converse is not true. In fact the representations of A_{JPS} are obtained by enforcing the anticommutation relation

$$\{S^\alpha,S^\beta\} = i(\gamma_\mu C)_{\alpha\beta}P^\mu \tag{3.3}$$

as a constraint on the representations of A_{WTS} (as well as the infinite algebra). Similar to Wigner,[5] we give an explicit construction of A_{WTS} in the rest frame and for mass zero.

A. Timelike momentum (massive particle representation)

Choosing the rest frame $P_\mu = (0,0,0,iP_0)$, we first rearrange the generators so that the group structure of A_{WT} is transparent.[7] Define

$$\begin{aligned}
L_3 &= \frac{(T^{23}-T^{14})}{4P_0}, \quad M_3 = \frac{(T^{23}+T^{14})}{4P_0},\\
N_3 &= \frac{W_3}{P_0} - \frac{(T^{23}+T^{14})}{4P_0},\\
L_+ &= \frac{T^{12}}{2P_0}, \quad M_+ = \frac{T^{24}}{2P_0},\\
N_+ &= \left(\frac{W_+}{P_0} - \frac{T^{24}}{2P_0}\right),\\
L_- &= \frac{-T^{34}}{2P_0}, \quad M_- = -\frac{T^{13}}{2P_0},\\
N_- &= \left(\frac{W_-}{P_0} + \frac{T^{13}}{2P_0}\right)
\end{aligned} \tag{3.4}$$

(note that that $W_j = \frac12\epsilon_{jkl}J^{kl}P_0$, $W_0 = 0$). From (3.1) with $P_\mu = (\mathbf{0},iP_0)$ one has that the operators **L, M, N** all commute with each other and each generate an SU(2) algebra;

$$[N_3,N_\pm] = \pm N_\pm, \quad [N_+,N_-] = 2N_3, \text{ etc.} \tag{3.5}$$

In other words, $A_{WT} = \mathrm{SU(2)}\times\mathrm{SU(2)}\times\mathrm{SU(2)}$. Trivially then, the Casimirs of A_{WT} are given by $\mathbf{N}^2$, $\mathbf{L}^2$, $\mathbf{M}^2$, where $N_\pm = N_1 \pm iN_2$, etc. We note that N_i ($i = 1,2,3$) commutes with S^α ($\alpha = 1,2,3,4$).

To construct A_{WTS} we normalize $\tilde{S}^\alpha = S^\alpha/(2P_0)^{1/2}$ and have the commutation relations

$$
\begin{aligned}
&[L_3,\tilde{S}^\alpha] = \pm\tfrac{1}{2}\tilde{S}^\alpha\begin{cases} +, & \alpha = 1,2 \\ -, & \alpha = 3,4, \end{cases}\\
&[M_3,\tilde{S}^\alpha] = \pm\tfrac{1}{2}\tilde{S}^\alpha\begin{cases} +, & \alpha = 2,4 \\ -, & \alpha = 1,3, \end{cases}\\
&[L_+,\tilde{S}^3] = \tilde{S}^1, \quad [L_-,\tilde{S}^2] = \tilde{S}^4,\\
&[L_+,\tilde{S}^4] = \tilde{S}^2, \quad [L_-,\tilde{S}^1] = \tilde{S}^3,\\
&[M_+,\tilde{S}^1] = -\tilde{S}^2, \quad [M_-,\tilde{S}^2] = -\tilde{S}^1,\\
&[M_+,\tilde{S}^3] = -\tilde{S}^4, \quad [M_-,\tilde{S}^4] = -\tilde{S}^3,\\
&[\tilde{S}^1,\tilde{S}^2] = L_+, \quad [\tilde{S}^2,\tilde{S}^4] = M_+,\\
&[\tilde{S}^3,\tilde{S}^4] = -L_-, \quad [\tilde{S}^1,\tilde{S}^3] = -M_-,\\
&[\tilde{S}^1,\tilde{S}^4] = (M_3 - L_3), \quad [\tilde{S}^2,\tilde{S}^3] = (L_3 + M_3),
\end{aligned}
\tag{3.6}
$$

with the rest zero. From the root vector diagram exhibited in Fig. 1 for these relations, one has that the algebra generated by **L**, **M** and S^α is Sp(4)$\equiv C_2$ or, equivalently, SO(5)$\equiv B_2$. Therefore, the algebra A_{WTS} is Sp(4)$\times$SU(2), where SU(2) is generated by **N**.

The irreducible representations of A_{WTS} are determined by (λ_1,λ_2,N), where (λ_1,λ_2) are the highest weight values in the Sp(4) representation.[8] Members of the representation are designated by $(\lambda_1,\lambda_2,L,L_3,M,M_3;N,N_3)$. However, in the following it is shown that the anticommutation relation (3.3) restricts the values of λ_1 and λ_2 while leaving N unconstrained due to its commuting with S^α.

In the rest frame, (3.3) is given by

$$\{S^1,S^4\} = -P_0, \quad \{S^2,S^3\} = P_0, \quad (S^\alpha)^2 = 0, \tag{3.7}$$

and all other anticommutators zero. From the Majorana condition,

$$S = C\bar{S}^T \tag{3.8}$$

or

$$(S^4)\dagger = -S^1 \quad \text{and} \quad (S^3)\dagger = S^2. \tag{3.9}$$

We are led to identify the following operators[2]:

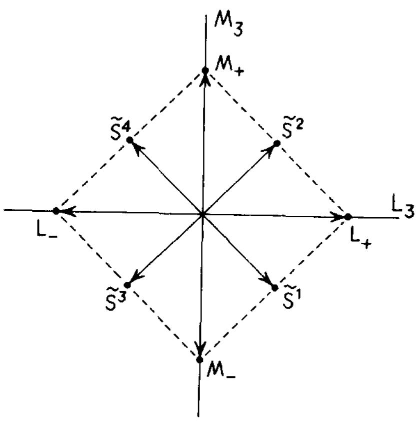


FIG. 1. Root vector diagrams for Sp(4) algebra. The generators are identified at the head of the corresponding root vector. The axes refer to the Cartan subalgebra with $H_1 = L_3$ and $H_2 = M_3$.

$$
\begin{aligned}
a_1 &= \frac{S^1}{\sqrt{P_0}}, \quad a_1^* = -\frac{S^4}{\sqrt{P_0}},\\
a_2 &= \frac{S^2}{\sqrt{P_0}}, \quad a_2^* = \frac{S^3}{\sqrt{P_0}}.
\end{aligned}
\tag{3.10}
$$

From the commutation relations between J_{12} and S^α in Eq. (2.1), we have

$$[J_{12},a_i] = \mp\tfrac{1}{2}a_i \quad \text{for} \quad i = \begin{cases}1\\2\end{cases}. \tag{3.11}$$

Therefore, we identify $a_1(a_2)$ as the annihilation operator for $J_z = \frac{1}{2}(-\frac{1}{2})$ and $a_1^*(a_2^*)$ as the creation operator for $J_z = \frac{1}{2}(-\frac{1}{2})$. Defining number operators $n_1 = a_1^*a_1$ and $n_2 = a_2^*a_2$ we have

$$
\begin{aligned}
N_3 &= J_{12} + \frac{n_2 - n_1}{2}, \quad N_+ = J_+ - a_1^*a_2,\\
N_- &= J_- + a_1a_2^*,\\
L_3 &= \frac{1 - n_1 - n_2}{2}, \quad M_3 = \frac{n_1 - n_2}{2},
\end{aligned}
\tag{3.12}
$$

and[9]

$$\mathbf{L}^2 + \mathbf{M}^2 = \tfrac{3}{4}. \tag{3.13}$$

Equation (3.13) implies that the irreducible representation is restricted to $(L,M) = (\frac{1}{2},0)$ or $(0,\frac{1}{2})$ or, equivalently, $(\lambda_1,\lambda_2) = (1,0)$ with arbitrary N. This gives the identification of n_1 and n_2 as follows:

$$
\begin{aligned}
&n_1 = 0, \quad n_2 = 0, \quad L_3 = \tfrac{1}{2}, \quad M_3 = 0,\\
&n_1 = 1, \quad n_2 = 1, \quad L_3 = -\tfrac{1}{2}, \quad M_3 = 0,\\
&n_1 = 1, \quad n_2 = 0, \quad L_3 = 0, \quad M_3 = \tfrac{1}{2},\\
&n_1 = 0, \quad n_2 = 1, \quad L_3 = 0, \quad M_3 = -\tfrac{1}{2}.
\end{aligned}
\tag{3.14}
$$

We can, therefore, replace L and M by n_1 and n_2 and arrive at the set $\{N,N_3,n_1,n_2\}$ or $\{N,J_{12},n_1,n_2\}$ as the commuting operators. These bases are used by Salam and Strathdee to construct the explicit representations of the superalgebra.[2] For completeness we reconstruct this representation. Using Eq. (3.12) one constructs the $N = 0$ representations containing $J = \{\frac{1}{2},0,0\}$ and the $N = \frac{1}{2}$ representation with $J = \{1,\frac{1}{2},\frac{1}{2},0\}$. For arbitrary $N > 0$ the representation contains $J = \{N + \frac{1}{2},N,N,N - \frac{1}{2}\}$ with a total of $4(2N + 1)$ states. The two states with $J = N$ correspond to $(n_1,n_2) = (0,0)$ and $(1,1)$. The parity operation

$$S \rightarrow S' = e^{i\eta}\gamma_4 S, \tag{3.15}$$

with the Majorana condition $S' = C\bar{S}'^T$, requires that $\eta = \pi/2$ (or $-\pi/2$) and

$$P\,|N,N_3,n_1,n_2\rangle = (-1)^{n_1 + n_2}|N,N_3,n_1,n_2\rangle, \tag{3.16}$$

where P is the parity operator. Thus the $J = N$ states are of opposite parity.

As described above, among the representations of A_{WTS} = Sp(4)$\times$SU(2) only the Sp(4) spinor representation is allowed by the constraint (3.3). Therefore, the only nonconstant Casimir is $\mathbf{N}^2$. Note from Eq. (3.4) this can be written

$$N_j = \frac{1}{P_0}\,[\,W_j - (i/4)SC^{-1}\gamma_j\gamma_5 S\,] \tag{3.17}$$

and can be generalized relativistically by defining

$$K_\mu = W_\mu - (i/4) S C^{-1} \gamma_\mu \gamma_5 S, \tag{3.18}$$

with

$$\mathbf{N}^2 = \left(K_\mu^2 - \frac{(K\cdot P)^2}{P^2} \right)\Bigg|_{P=(0,iP_0)}. \tag{3.19}$$

From the commutators

$$[K_\mu S^\alpha] = -\tfrac{1}{2}(\gamma_5)_{\alpha\beta} S^\beta P_\mu \tag{3.20}$$

and

$$[K_\mu, P_\lambda] = 0, \tag{3.21}$$

we have $(K_\mu P_\nu - K_\nu P_\mu)^2$ commuting with all operators S^α, $J_{\mu\nu}$, and P_μ. This is the relativistic expression of the Casimir operator $\mathbf{N}^2$. This operator, with P^2, forms the set of Casimir operators for A_{JPS}.

B. Lightlike momentum (massless particles)

Taking $P_\mu = (0,0,p,ip)$ in Appendix B, all commutation relations for $A_{W\hat{T}\hat{S}}$ are given, where

$$\hat{S}^1 = \frac{(S^1 + S^3)}{2}, \quad \hat{S}^2 = \frac{(S^2 + S^4)}{2},$$
$$\hat{S}^3 = \frac{(S^3 - S^1)}{2}, \quad \hat{S}^4 = \frac{(S^4 - S^2)}{2}, \tag{3.22}$$

and

$$\hat{T}^{\alpha\beta} = [\hat{S}^\alpha, \hat{S}^\beta].$$

Note from Eq. (B7) that the group structure of $A_{W\hat{T}\hat{S}}$ is $G \times \mathrm{U}(1)$, where U(1) is generated by $\hat{T}^{23}$, which commutes with all generators of $A_{W\hat{T}\hat{S}}$. From Eq. (B6) the Casimir operator is

$$K_{\mu\nu}^2 = -2(\hat{T}^{23})^2 p^2. \tag{3.23}$$

Instead of analyzing G and its representations, we consider the constraint condition (3.3),

$$\{\hat{S}^1, \hat{S}^4\} = -p, \tag{3.24}$$

and all other $\{\hat{S}^i, \hat{S}^j\} = 0$. Define creation and annihilation operators

$$\frac{\hat{S}^1}{\sqrt{p}} = a, \quad -\frac{\hat{S}^4}{\sqrt{p}} = a^*,$$
$$\frac{\hat{S}^2}{\sqrt{p}} = b, \quad -\frac{\hat{S}^3}{\sqrt{p}} = b^*, \tag{3.25}$$

with

$$\{a, a^*\} = 1, \tag{3.26}$$

$$\{b, b^*\} = 0, \tag{3.27}$$

and all others vanish. Equation (3.26) leads us to identify $a(a^*)$ as the annihilation (creation) operator of a spin up state. (Note $[J_{12}, \hat{S}^1] = -\tfrac{1}{2}\hat{S}^1$ and $[J_{12}, \hat{S}^4] = \tfrac{1}{2}\hat{S}^4$). Equation (3.27) implies the operation of b on any state $|\psi\rangle$ is zero:

$$b\,|\psi\rangle = b^*|\psi\rangle = 0. \tag{3.28}$$

It immediately follows that

$$\hat{T}^{23} = p[b, b^*] = 0. \tag{3.29}$$

In fact, all $\hat{T}^{ij}$ except $\hat{T}^{14}$ vanish. With this result it follows that K_3 and $K\pm$, defined in Eq. (B5), form the Euclidean algebra E_2. As seen from Eq. (B9),

$$[K_+, K_-] = 0,$$
$$[K_3, K_\pm] = \pm K_\pm p. \tag{3.30}$$

As is shown in Appendix C, this implies that $K_+ = K_- = 0$ for a finite dimensional representation. This is analogous to the $W^2 = 0$ condition in the Wigner analysis of lightlike momentum. In this case the representation is characterized by generalized helicity Λ,

$$\Lambda = \frac{\mathbf{K}\cdot\mathbf{P}}{p^2} = \frac{K_3}{p} = \frac{(W_3 - \tfrac{1}{2}\hat{T}^{14})}{p}$$
$$= (J_{12} - n + \tfrac{1}{2}), \tag{3.31}$$

where $n = a^*a$. As is shown by Salam and Strathdee,[2] the representation is characterized by two states, $|j_3\rangle$ and $|j_3 + \tfrac{1}{2}, n = 1\rangle$, where $a|j_3\rangle = 0$ and $|j_3 + \tfrac{1}{2}, n = 1\rangle = a^*|j_3\rangle$. Both have $\Lambda = (j_3 + \tfrac{1}{4})$. It is obvious that the parity operator acting on these states gives a basis set of the opposite helicity. It should be emphasized that, as for the massive case, among all possible representations of $A_{W\hat{T}\hat{S}}$ the supersymmetry representation obtained above is selected by the constraint (3.3).

IV. SPIN STATISTICS IN THE INFINITE LIE ALGEBRA

In the previous sections it has been shown that the representations of superalgebra are equivalent to the representations of the infinite Lie algebra with the anticommutation relations among the fermionic charges as a constraint. From states in the irreducible representations of A_{WTS}, field operators can be constructed by a standard method.[10] The infinitesimal tranformation of the field operator $\Phi(x)$ by fermionic generators is given by

$$\delta\Phi = i\,\bar{\epsilon}^\alpha [S^\alpha, \Phi], \tag{4.1}$$

where ϵ^α is a constant spinor. The choice of ϵ^α to be a commuting parameter contradicts the spin statistics relation in the case of fermionic fields $\Phi \equiv \psi$. The standard procedure is to use anticommuting c-numbers as parameters for supersymmetry transformations. These parameters also anticommute with fermionic fields.

For the infinite Lie algebra we require commuting parameters because the commutator of two supersymmetry transformations should be a generator of the algebra,

$$[\bar{\epsilon}_1 S, \bar{\epsilon}_2 S] = \bar{\epsilon}_1^\alpha \bar{\epsilon}_2^\beta T^{\alpha\beta}. \tag{4.2}$$

In order to resolve the spin statistics problem in Eq. (4.1), we use the Klein transformation[11] to define a new fermionic operator,

$$S'^\alpha = i(-1)^{N_F} S^\alpha, \tag{4.3}$$

where N_F is the fermionic number operator. Note that the fermionic content of S^α is not defined; that is, being a Majorana spinor, S^α is a mixture of the ± 1 eigenvectors of N_F. However, the Klein operator $(-1)^{N_F}$ has definite anticommutation relations with all fermionic operators regardless of Dirac or Majorana properties:

$$\{(-1)^{N_F},\psi\} = 0 \tag{4.4}$$

and

$$\{(-1)^{N_F},S^\alpha\} = 0, \tag{4.5}$$

where ψ and S^α are Dirac or Majorana. This is seen by the following argument. Letting ψ be a Dirac field satisfying

$$[N_F,\psi] = \pm\psi, \tag{4.6}$$

we have

$$e^{i\pi N_F}\psi e^{-i\pi N_F} = e^{\pm i\pi}\psi = -\psi. \tag{4.7}$$

This equation implies (4.4) for Dirac fields with definite fermion number. A Majorana field ψ_1 or ψ_2 can be expressed in terms of a Dirac field ψ by

$$\psi_1 = \frac{(\psi + \psi^C)}{\sqrt{2}} \tag{4.8}$$

or

$$\psi_2 = \frac{(\psi - \psi^C)}{\sqrt{2}\,i}. \tag{4.9}$$

It is easy to see that Eq. (4.7) and therefore (4.4) is valid for Majorana fields without definite fermion number.

With these new operators, Eq. (4.3), we have

$$\delta\psi = [\bar{\epsilon}^\alpha S'^\alpha,\psi] = -\bar{\epsilon}^\alpha\{S^\alpha,\psi\}(-1)^{N_F} \tag{4.10}$$

for fermionic fields ψ. Therefore, we have

$$[\bar{\epsilon}^\alpha S'^\alpha,\bar{\epsilon}^\beta S'^\beta] = \bar{\epsilon}^\alpha\bar{\epsilon}^\beta T^{\alpha\beta} \tag{4.11}$$

because $(-1)^{2N_F} = 1$. In terms of operators we have the equation

$$[S'^\alpha,S'^\beta] = T^{\alpha\beta} \equiv T'^{\alpha\beta}. \tag{4.12}$$

The constraint is given by

$$\{S'^\alpha,S'^\beta\} = i(\gamma_\mu C)_{\alpha\beta}P^\mu. \tag{4.13}$$

The infinite Lie algebra should be modified using S'. Equations (2.1), (2.7), (3.3), and those in Appendix A are altered by the replacement

$$S^\alpha \to S'^\alpha,$$
$$T^{\alpha\beta} \to T'^{\alpha\beta}, \tag{4.14}$$

This modification alters neither the structure nor the physical content of A_{WTS}. Note that the Majorana condition for S' implies

$$S' = i(-1)^{N_F} C\bar{S}^T = C\bar{S}'^T. \tag{4.15}$$

Therefore, the sign change in the constraint Eq. (4.13) does not alter the definition of creation and annihilation operators in Eq. (3.10).

V. CONCLUSION

The standard supersymmetry transformation is exponentiated[12] using infinitesimal anticommuting parameters. We have avoided this by constructing an infinite Lie algebra and using the Klein transformation. The anticommutation relations of fermionic generators becomes a constraint on the representations of the infinite Lie algebra which yields the standard supersymmetry representations.

While anticommuting parameters are natural in superspace and the superspace formalism is convenient for constructing field theories, it is extremely difficult to comprehend a physical reality in such a space. Our formalism replaces this difficult concept with an infinite Lie algebra which uses commuting parameters. This new viewpoint of supersymmetry may help in understanding the nature of the symmetry, its breaking, and supergravity.

ACKNOWLEDGMENTS

We would like to express our gratitude for useful discussions with R. Akhoury, I. Bars, F. Gürsey, K. T. Hecht, J. Leveille, D. Peterson, M. Veltman, D. Williams, and A. C. T. Wu. This work is supported in part by the U.S. Department of Energy.

APPENDIX A

Consider the commutation relations for $\{S^\alpha P_\mu, T^{\alpha\beta}P_\mu\}$ which appear on the right-hand side of Eq. (2.7). These operators are generators for the infinite Lie algebra.

$$
\begin{aligned}
&[S^\alpha P_\mu, S^\beta P_\lambda] = T^{\alpha\beta}P_\mu P_\lambda,\\
&[S^\alpha P_\mu, T^{\beta\eta}P_\lambda] = -2iS^\beta(\gamma_\rho C)_{\eta\alpha}P^\rho P_\mu P_\lambda\\
&\qquad + 2i(\gamma_\rho C)_{\beta\alpha}S^\eta P^\rho P_\mu P_\lambda,\\
&[T^{\alpha\beta}P_\lambda, T^{\eta\delta}P_\nu] = [2iT^{\eta\alpha}(\gamma_\mu C)_{\delta\beta}P^\mu\\
&\qquad - 2iT^{\delta\alpha}(\gamma_\mu C)_{\eta\beta}P^\mu\\
&\qquad - 2iT^{\eta\beta}(\gamma_\mu C)_{\delta\alpha}P^\mu\\
&\qquad + 2iT^{\delta\beta}(\gamma_\mu C)_{\eta\alpha}P^\mu]P_\lambda P_\nu,\\
&[S^\alpha P_\mu, S^\beta] = T^{\alpha\beta}P_\mu,\\
&[T^{\alpha\beta}P_\mu, S^\eta] = 2iS^\alpha(\gamma_\rho C)_{\beta\eta}P^\rho P_\mu\\
&\qquad - 2i(\gamma_\rho C)_{\alpha\eta}S^\beta P^\rho P_\mu,\\
&[S^\alpha P_\mu, P_\lambda] = 0,\\
&[T^{\alpha\beta}P_\mu, P_\lambda] = 0,\\
&[S^\alpha P_\mu, T^{\beta\eta}] = -2iS^\beta(\gamma_\rho C)_{\eta\alpha}P^\rho P_\mu\\
&\qquad + 2i(\gamma_\rho C)_{\beta\alpha}S^\eta P^\rho P_\mu,\\
&[T^{\alpha\beta}P_\lambda, T^{\eta\delta}] = [2iT^{\eta\alpha}(\gamma_\mu C)_{\delta\beta}P^\mu - 2iT^{\delta\alpha}(\gamma_\mu C)_{\eta\beta}P^\mu,\\
&\qquad - 2iT^{\eta\beta}(\gamma_\mu C)_{\delta\alpha}P^\mu\\
&\qquad + 2iT^{\delta\beta}(\gamma_\mu C)_{\eta\alpha}P^\mu]P_\lambda,\\
&[S^\alpha P_\mu, J_{\lambda\nu}] = iS^\alpha(-\delta_{\lambda\mu}P_\nu + \delta_{\mu\nu}P_\lambda)\\
&\qquad + \tfrac{1}{2}(\sigma_{\lambda\nu})_{\alpha\beta}S^\beta P_\mu,
\end{aligned}
\tag{A1}
$$

$$[T^{\alpha\beta}P_\mu, J_{\lambda\nu}] = iT^{\alpha\beta}(-\delta_{\lambda\mu}P_\nu + \delta_{\mu\nu}P_\lambda) + \tfrac{1}{2}(\sigma_{\lambda\nu})_{\alpha\delta}T_{\delta\beta}P_\mu - \tfrac{1}{2}(\sigma_{\lambda\nu})_{\beta\delta}T_{\delta\alpha}P_\mu. \tag{A2}$$

Note that Eqs. (A1) are obtained from Eq. (2.7) by multiplying by the appropriate momentum operators. Equation (A2) has no new generators on the right-hand side.

APPENDIX B: MASSLESS CASE $P_\mu = (0,0,p,ip)$

Introduce the variables

$$\hat{S}^1 = \left(\frac{S^1+S^3}{2}\right), \quad \hat{S}^2 = \left(\frac{S^2+S^4}{2}\right),$$
$$\hat{S}^3 = \left(\frac{S^3-S^1}{2}\right), \quad \hat{S}^4 = \left(\frac{S^4-S^2}{2}\right). \tag{B1}$$

With the constraint equation (3.3) we have

$$\{\hat{S}^1,\hat{S}^4\} = -p,$$
$$\{\hat{S}^1,\hat{S}^2\} = \{\hat{S}^1,\hat{S}^3\} = \{\hat{S}^2,\hat{S}^4\} = \{\hat{S}^2,\hat{S}^3\} = \{\hat{S}^3,\hat{S}^4\} = 0. \tag{B2}$$

In terms of the $\hat{S}$ variables the commutators $\hat{T}^{\alpha\beta}$ are given by

$$\hat{T}^{12} = [\hat{S}^1,\hat{S}^2] = \tfrac{1}{4}(T^{12}+T^{14}-T^{23}+T^{34}),$$
$$\hat{T}^{13} = \tfrac{1}{2}T^{13},$$
$$\hat{T}^{14} = \tfrac{1}{4}(-T^{12}+T^{14}+T^{23}+T^{34}),$$
$$\hat{T}^{23} = \tfrac{1}{4}(T^{12}+T^{14}+T^{23}-T^{34}),$$
$$\hat{T}^{24} = \tfrac{1}{2}T^{24},$$
$$\hat{T}^{34} = \tfrac{1}{4}(T^{12}-T^{14}+T^{23}+T^{34}). \tag{B3}$$

Also,

$$W_1 = (J_{23}+iJ_{24})p, \quad W_3 = (J_{12})p,$$
$$W_2 = -(J_{13}+iT_{14})p, \quad W_4 = i(J_{12})p. \tag{B4}$$

Expressing the components of K_μ [Eq. (3.18)] in terms of $\hat{T}$ amd W,

$$K_+ = (W_+ - \tfrac{1}{2}T^{24}) = (W_+ - \hat{T}^{24}),$$
$$K_- = (W_- + \tfrac{1}{2}T^{13}) = (W_- + \hat{T}^{13}),$$
$$K_3 = W_3 - \tfrac{1}{4}(T^{14}+T^{23}) = W_3 - \tfrac{1}{2}(\hat{T}^{14}+\hat{T}^{23}),$$
$$K_4 = i(W_3 - \tfrac{1}{4}(-T^{12}+T^{34})) = i(W_3 - \tfrac{1}{2}(\hat{T}^{14}-\hat{T}^{23})), \tag{B5}$$

and the Casimir operator is given by

$$K^2_{\mu\nu} = -2(K_\mu P^\mu)^2 = -2(\hat{T}^{23})^2p^2. \tag{B6}$$

The commutation relations for $A_{W\hat{T}\hat{S}}$ are given by the following:

$\hat{T}^{23}$ commutes with all $\hat{S}^\alpha$ and W^i as seen from Eq. (B6), (B7)

K_+,K_- commute with all $\hat{S}^\alpha$, $\hat{T}^{\alpha\beta}$, (B8)

$$[K_+,K_-] = -2p\hat{T}^{23}, \tag{B9}$$
$$[K_3,K_\pm] = \pm K_\pm p, \tag{B10}$$
$$[K_3/p,\hat{T}^{12}] = \hat{T}^{12},$$
$$[K_3/p,\hat{T}^{34}] = -\hat{T}^{34}, \tag{B11}$$
$$[K_3/p,\hat{S}^1] = \tfrac{1}{2}\hat{S}^1,$$
$$[K_3/p,\hat{S}^2] = \tfrac{1}{2}\hat{S}^2,$$
$$[K_3/p,\hat{S}^3] = -\tfrac{1}{2}\hat{S}^3,$$
$$[K_3/p,\hat{S}^4] = -\tfrac{1}{2}\hat{S}^4, \tag{B12}$$
$$[\hat{T}^{j4},\hat{S}^1] = -2p\hat{S}^j, \quad j=1,2,3$$
$$[\hat{T}^{1j},\hat{S}^4] = 2p\hat{S}^j, \quad j=2,3,4 \tag{B13}$$
$$[\hat{T}^{14},\hat{T}^{1k}] = -2p\hat{T}^{1k}, \quad k=2,3$$
$$[\hat{T}^{14},\hat{T}^{j4}] = 2p\hat{T}^{j4}, \quad j=2,3$$
$$[\hat{T}^{12},\hat{T}^{34}] = -2p\hat{T}^{23},$$
$$[\hat{T}^{13},\hat{T}^{24}] = 2p\hat{T}^{23}, \tag{B14}$$

the rest is zero.

APPENDIX C: PROOF THAT $K_\pm = 0$ FOR FINITE REPRESENTATION OF E_2

Consider the Euclidean algebra E_2 commutation relations for $\{K_+,K_-,K_3\}$:

$$[K_+,K_-] = 0, \tag{C1}$$
$$[K_3,K_\pm] = \pm K_\pm, \tag{C2}$$

where $K_+{}^\dagger, = K_-$. The Casimir operator is K_-K_+ and Eq. (C2) implies $K_+(K_-)$ is the raising (lowering) operator for the K_3 eigenvalues. Consider the minimum K_3 eigenstates, $|k_{\min}\rangle$, defined by $K_-|k_{\min}\rangle = 0$ and $K_3|k_{\min}\rangle = k_{\min}|k_{\min}\rangle$. From (C1),

$$K_-K_+|k_{\min}\rangle = 0, \tag{C3}$$

and therefore $K_-K_+ = 0$ for the entire representation because it is a Casimir operator. This implies $K_+ = K_- = 0$ for any finite representation.

[1] D. V. Volkov and V. P. Akulov, Phys. Lett. B **46**, 109 (1973); J. Wess and B. Zumino, *ibid.* **49**, 52 (1974); Nucl. Phys. B **70**, 39 (1974); **78**, 1 (1974); for further references see review articles. P. Fayet and S. Ferrara, Phys. Rep. **32C**, 249 (1977); A. Salam and J. Strathdee, Fortschr. Phys. **26**, 57 (1978).

[2] A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974); **80**, 499 (1974); Phys. Rev. D **4**, 1521 (1975); P. Goddard, Nucl. Phys. B **88**, 429 (1975).

[3] P. Van Nieuwenhuizen, Phys. Rep. **68C**, 189 (1981).

[4] S. Coleman and J. Mandula, Phys. Rev. **159**, 1251 (1967).

[5] E. P. Wigner, Ann. Math. **40**, 149 (1939).

[6] V. G. Kac and D. Peterson, Bull. Am. Math. Soc. **3**, 1057 (1980); V. G. Kac, Izv. Akad. Nauk SSSR Ser. Mat. **32** (1968) [Math. USSR-Izv. **2**, 1271 (1968)]; R. V. Moody, J. Algebra **10**, 211 (1968).

[7] We take $C = (\gamma_2\gamma_4)$ and all γ-matrices are in the Pauli–Dirac representation.

[8] B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New York, 1970); K. T. Hecht, Nucl. Phys. **63**, 177 (1965); **A102**, 11 (1967).

[9] Note that the quadratic Casimir for Sp(4) is given by
$\mathbf{L}^2 + \mathbf{M}^2 - (1/4P_0)\{S^1,S^4\} + (1/4P_0)\{S^2,S^3\}$.

[10] H. Joos, Fortschr. Phys. **10**, 65 (1962); S. Weinberg, Phys. Rev. **133**,

B1318 (1964); **134**, B882 (1964); E. H. Wichmann (unpublished); S. Weinberg, *Lectures on Particles and Field Theory, Brandeis Summer Institute in Theoretical Physics 1964* (Prentice-Hall, Englewood Cliffs, N. J., 1965); S. Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1966).
[11]O. Klein, J. Phys. Radium **9**, 1 (1938); K. Nishijima, Prog. Theor. Phys. **5**, 187 (1950); S. Oneda and H. Umezawa, *ibid.* **9**, 685 (1953); T. Kinoshita, Phys. Rev. **96**, 199 (1954); **110**, 978 (1958); G. Lüders, Z. Naturforsch **13A**, 251 (1958); H. Araki, J. Math. Phys. **2**, 267 (1961).
[12]F. A. Berezin and G. I. Kac, Mat. Sb. **82**, 124 (1970) [Math. USSR **11**, 311 (1970)].

General superposition of solitons and various ripplons of a two-dimensional nonlinear Schrödinger equation

Akira Nakamura
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676 [a)]

(Received 12 January 1982; accepted for publication 19 March 1982)

We consider the two-dimensional nonlinear Schrödinger equation of Benney–Roskes. It is shown that the equation admits superposition solutions of solitons and various ripplons.

PACS numbers: 02.30. + g

1. INTRODUCTION

Recently, it has been found that certain nonlinear evolution equations simultaneously admit two different types of solutions, solitons and ripplons (simple similarity type explode-decay mode solutions).[1–8] Soliton solutions represent propagating waves with constant speed and constant profile, whereas ripplons represent more dynamical waves whose profiles grow and then decay with time. Ripplon solutions can be expected to play important roles in explaining dynamical phenomena such as explosions.

In two-dimensional (2D) systems, we know three equations whose ripplon solutions are known explicitly. They are the 2D-KdV (or the Kadomtsev-Petviashvili) equations,[1–3,6] 2D cubic nonlinear Schrödinger (*NLS*) equation,[4] and the 2D-NLS equation of Benney–Roskes.[5,8] The last equation is written as[9–14]

$$iu_t - \beta u_{xx} + \gamma u_{yy} + \delta u^* uu - 2wu = 0, \tag{1.1a}$$

$$\beta w_{xx} + \gamma w_{yy} - \beta\delta(u^*u)_{xx} = 0, \tag{1.1b}$$

where β, γ, δ are real constants. Throughout the paper subscripts x, y, t denote partial derivatives.

By using the technique of Hirota bilinear theory, Nakamura recently derived 1D-like ripplons and the superpositions among themselves.[5] It was also shown that there exists simple transformations which transform arbitrary propagating-wave solutions with permanent profile to the explode–decay mode (ripplon) solution, which implies that if we know multiple solitons, multiple periodic waves, multiple lump solitons, and so on, we can derive explode–decay mode multiple solitons, multiple periodic waves, multiple lump solitons and so on.[8] Then the question arises whether superposition between ordinary waves and ripplons is possible. The purpose of this paper is to show that this superposition is actually possible. For this purpose, we use the technique of Zakharov–Shabat inverse spectral transform (IST) method. Anker and Freeman have already applied the scheme to this equation and derived soliton solutions.[13] We will show that the same IST scheme can generate ripplons. Then from the linearity of the IST scheme itself, it can be shown that the properly superposed state of solitons and ripplons is also the solution.

In Sec. 2, we briefly review the IST formalism and in Sec. 3, we derive solitons, lump solitons, ripplons, and lump ripplons and show that all of them can be superposed.

[a)] Present address: Physics Laboratory, Osaka University of Foreign Studies, Aoh Madani, Minoo City, Osaka 562 Japan.

2. IST FORMALISM

For the convenience of the later calculations and notation, we review very briefly the Zakharov–Shabat IST scheme.[15] Following them, we introduce the Volterra operator $\widehat{K}$ and integral operator $\widehat{F}$ by

$$\widehat{K}\psi(x) \equiv \int_x^\infty dz\, K(x,z)\psi(z), \quad \widehat{F}\psi(x) \equiv \int_{-\infty}^\infty dz\, F(x,z)\psi(z), \tag{2.1}$$

where $K(x,z)$, $F(x,z)$ are (n,n) matrices with their (i,j) element denoted by $K_{ij}(x,z)$, $F_{ij}(x,z)$ respectively, $\psi(x)$ is an $(n,1)$ matrix, and it is assumed that $(1+\widehat{K})^{-1}$ exists. We consider a pair of commuting operators M_1 and M_2

$$[M_1,M_2] \equiv M_1M_2 - M_2M_1 = 0, \tag{2.2}$$

and the transformation from M_i to $\widetilde{M}_i$ by

$$\widetilde{M}_i(1+\widehat{K}) - (1+\widehat{K})M_i = 0, \quad (i = 1,2), \tag{2.3}$$

leading to

$$[\widetilde{M}_1,\widetilde{M}_2] = 0, \tag{2.4}$$

which provides nonlinear evolution equations with respect to the variables K_{ij}. The explicit solutions for the variables K_{ij} in turn can be given by solving the integral equation

$$F(x,z) + K(x,z) + \int_x^\infty ds\, K(x,s)F(s,z) = 0, \tag{2.5}$$

where $F(x,z)$ is determined from the relation

$$[\widehat{F},M_i] = 0, \quad (i = 1,2). \tag{2.6}$$

We consider the case where Eq. (2.6) has solutions in the form of separation of variables

$$F(x,z) = \sum_{i=1}^N f_i(x)\bar{f}_i(z), \tag{2.7}$$

where f_i, $\bar{f}_i$ are appropriate (n,n) matrices. By assuming the form of $K(x,z)$ as

$$K(x,z) = \sum_{i=1}^N K_i(x)\bar{f}_i(z), \tag{2.8}$$

and inserting Eqs. (2.7) and (2.8) into Eq. (2.5), $K_i(x)$ and thus $K(x,z)$ can be solved as

$$(K_1 \ldots K_N) = -(f_1 \ldots f_N)L^{-1}, \tag{2.9a}$$

$$K(x,x) = \sum_{i=1}^N K_i(x)\bar{f}_i(x) = -(f_1 \ldots f_N)L^{-1}\begin{pmatrix}\bar{f}_1\\ \vdots \\ \bar{f}_N\end{pmatrix}, \tag{2.9b}$$

$$[(i,j)\text{block of } L\,] \equiv \delta_{ij}1 + \int_x^\infty ds\,\bar{f}_i(s)f_j(s), \quad (1 \leqslant i,j \leqslant N). \tag{2.9c}$$

Here L is an (nN, nN) matrix consisting of N^2 blocks of (n,n) matrices. In the case of $N=1$ and simple F of the form

$$F(x,z) = f(x)\bar{f}(z) = \begin{pmatrix} 0, & f_{12}(x) \\ f_{21}(x), & 0 \end{pmatrix}\begin{pmatrix} \bar{f}_{11}(z), & 0 \\ 0, & \bar{f}_{22}(z) \end{pmatrix}, \tag{2.10}$$

we have

$$K(x,x) = \left[1 - \int_x^\infty ds\, f_{12}(s)\bar{f}_{11}(s) \int_x^\infty ds\, f_{21}(s)\bar{f}_{22}(s)\right]^{-1} \times \begin{pmatrix} f_{12}(x)\bar{f}_{11}(x)\int_x^\infty ds\, f_{21}(s)\bar{f}_{22}(s), & -f_{12}(x)\bar{f}_{22}(x) \\ -f_{21}(x)\bar{f}_{11}(x), & f_{21}(x)\bar{f}_{22}(x)\int_x^\infty ds\, f_{12}(s)\bar{f}_{11}(s) \end{pmatrix}. \tag{2.11}$$

Expression (2.9) shows that so far as the solution F of Eq. (2.6) is obtained in the form of separation of variables, Eq. (2.7), the superposed solution exists and is given by Eqs. (2.9(b)) and (2.9(c)). In the next section we will show that all of the solitons, lump solitons, ripplons, and lump ripplons can be generated in the form of Eq. (2.7), thus superposition of these are also a solution.

3. SOLITONS, LUMP SOLITONS, RIPPLONS, LUMP RIPPLONS, AND THEIR SUPERPOSITIONS

First we follow the procedures of Anker and Freeman and derive ordinary solitons.[13] However, for the sake of simplicity, we take the simplified boundary condition $u(x = \pm\infty) = 0$ instead of their nonvanishing boundary condition $u(x=\pm\infty)\neq 0$. We choose a pair of commuting operators M_1, M_2, as

$$M_1 = \beta_0\partial_y + \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}\partial_x, \quad M_2 = \alpha_0\partial_t + \partial_x^2, \tag{3.1}$$

where $\alpha_0, \beta_0, \gamma_1, \gamma_2$ are scalar constants (β,γ and $\beta_0, \gamma_1, \gamma_2$ are completely unrelated, different constants). Equation (2.3) then determines $\widetilde{M}_1$, $\widetilde{M}_2$ as

$$\widetilde{M}_1 = M_1 + (\gamma_1 - \gamma_2)\begin{pmatrix} 0 & \xi_{12} \\ -\xi_{21} & 0 \end{pmatrix},$$

$$\widetilde{M}_2 = M_2 + 2\begin{pmatrix} \xi_{11,x} \xi_{12,x} \\ \xi_{21,x} \xi_{22,x} \end{pmatrix}, \tag{3.2}$$

$$\xi_{ij} \equiv K_{ij}(x,z)|_{z=x}. \tag{3.3}$$

Then Eq. (2.4) is written explicitly for each matrix element as

$$-\alpha_0(\gamma_1-\gamma_2)\xi_{12,t} + (\gamma_1+\gamma_2)\xi_{12,xx} + 2\beta_0\xi_{12,xy} - 2(\gamma_1-\gamma_2)\xi_{12}(\xi_{11}-\xi_{22})_x = 0, \tag{3.4a}$$

$$\alpha_0(\gamma_1-\gamma_2)\xi_{21,t} + (\gamma_1+\gamma_2)\xi_{21,xx} + 2\beta_0\xi_{21,xy} - 2(\gamma_1-\gamma_2)\xi_{21}(\xi_{11}-\xi_{22})_x = 0, \tag{3.4b}$$

$$\beta_0(\xi_{11}+\xi_{22})_y + \tfrac{1}{2}(\gamma_1+\gamma_2)(\xi_{11}+\xi_{22})_x + \tfrac{1}{2}(\gamma_1-\gamma_2)(\xi_{11}-\xi_{22})_x = 0, \tag{3.4c}$$

$$\beta_0(\xi_{11}-\xi_{22})_y + \tfrac{1}{2}(\gamma_1+\gamma_2)(\xi_{11}-\xi_{22})_x + \tfrac{1}{2}(\gamma_1-\gamma_2)(\xi_{11}+\xi_{22})_x + 2(\gamma_1-\gamma_2)\xi_{12}\xi_{21} = 0. \tag{3.4d}$$

We set the value of parameters as $\gamma_1 = -\gamma_2 (\equiv\gamma_0)$. Then Eq. (3.4c) gives $\xi_{11}+\xi_{22} = \gamma_0\phi_x$, $\xi_{11}-\xi_{22} = -\beta_0\phi_y$, where we have newly introduced the quantity $\phi = \phi(x,t)$. Further, by considering the coordinate transformation[13]

$$X = \beta^{1/2}(\beta_0 x - i\gamma_0 y), \quad Y = \gamma^{1/2}(\beta_0 x + i\gamma_0 y), \tag{3.5a}$$

$$x = \frac{1}{2\beta_0}\left(\frac{X}{\beta^{1/2}} + \frac{Y}{\gamma^{1/2}}\right), \quad y = \frac{i}{2\gamma_0}\left(\frac{X}{\beta^{1/2}} - \frac{Y}{\gamma^{1/2}}\right), \tag{3.5b}$$

and choosing the parameter values as

$$\alpha_0 = \beta_0^2 = -4/\delta, \tag{3.6}$$

and under the condition $\gamma_0\phi = (\gamma_0\phi)^*$, by denoting $2\beta\gamma_0\phi_{xx} \equiv w$, $\xi_{12}\equiv u$, $(\xi_{21}\equiv\xi_{12}^*)$, we see that Eqs. (3.4a–3.4d) reduce to Eq. (1.1) with subscripts x,y replaced by X,Y. For simplicity, we consider $\delta>0$, which gives pure imaginary β_0 or $\beta_0 = i|\beta_0|$. We also take real γ_0. In our present calculation X,Y are physical variables and should be real. This and the above parametrization for β_0, γ_0 imply both x and y to be pure imaginary, then each integral path in the previous section should be interpreted as being on the imaginary axis instead of usual real axis $\int_x^\infty ds \to \int_x^{+i\infty} ds$. We note that this formal modification of upper bound of integration $+\infty \to +i\infty$ does not affect any of the remaining arguments and the whole scheme holds in the same manner as the usual real x,y case.

Now we consider explicit solutions. From Eqs. (2.6) and (3.1) we have equations for F as

$$\beta_0 F_y + \begin{pmatrix} \gamma_0 & 0 \\ 0 & -\gamma_0 \end{pmatrix} F_x + F_z \begin{pmatrix} \gamma_0 & 0 \\ 0 & -\gamma_0 \end{pmatrix} = 0, \tag{3.7a}$$

$$\alpha_0 F_t + F_{xx} - F_{zz} = 0. \tag{3.7b}$$

The simple exponential-type solution to Eqs. (3.7) is given by

$$F(x,z) = \begin{pmatrix} 0, & \exp[m_i x + n_i z + \beta_0^{-1}\gamma_0(n_i - m_i)y + \alpha_0^{-1}(n_i^2 - m_i^2)t + \eta_{0i}] \\ \exp[m_i' x + n_i' z + \beta_0^{-1}\gamma_0(m_i' - n_i')y + \alpha_0^{-1}(n_i'^2 - m_i'^2)t + \eta_{0i}'], & 0 \end{pmatrix} \tag{3.8a}$$

$$= \begin{pmatrix} 0, & \exp[m_i(x - \beta_0^{-1}\gamma_0 y - \alpha_0^{-1}m_i t) + \eta_{0i}] \\ \exp[m_i'(x + \beta_0^{-1}\gamma_0 y - \alpha_0^{-1}m_i' t) + \eta_{0i}'], & 0 \end{pmatrix}\begin{pmatrix} \exp[n_i'(z - \beta_0^{-1}\gamma_0 y + \alpha_0^{-1}n_i' t)], & 0 \\ 0, & \exp[n_i(z + \beta_0^{-1}\gamma_0 y + \alpha_0^{-1}n_i t)] \end{pmatrix} \tag{3.8b}$$

$$\equiv f_i^s(x)\bar{f}_i^s(z). \tag{3.8c}$$

Here $m_i, n_i, \eta_{0i}, m_i', \eta_i', \eta_{0i}'$ are arbitrary complex constants. Via Eq. (2.11), this gives the 1-soliton solution

$$u = K_{12}(x,x) = -\exp(N_i^s)/\{1 - [(m_i + n_i')(m_i' + n_i)]^{-1}\exp(D_i^s)\},$$

$$N_i^s \equiv (m_i + n_i)x + \beta_0^{-1}\gamma_0(n_i - m_i)y + \alpha_0^{-1}(n_i^2 - m_i^2)t + \eta_{0i},$$
$$D_i^s \equiv N_i^s + (m_i' + n_i')x - \beta_0^{-1}\gamma_0(n_i' - m_i')y + \alpha_0^{-1}(n_i'^2 - m_i'^2)t + \eta_{0i}'. \tag{3.9}$$

Here for the integral $\int_x^{+i\infty} ds$ in Eq. (2.11) to converge, we take the parameters as $\mathrm{Im}(m_i + n_i') > 0$, $\mathrm{Im}(m_i' + n_i) > 0$. On the other hand, Eqs. (3.5) give the relation

$$(x - \beta_0^{-1}\gamma_0 y + c_1)(x + \beta_0^{-1}\gamma_0 y + c_2) = (2\beta_0^2\beta)^{-1}(X + \bar{c}_1)^2 + (2\beta_0^2\gamma)^{-1}(Y + \bar{c}_2)^2,$$
$$\bar{c}_1 \equiv \tfrac{1}{2}\beta_0\beta^{1/2}[(1+i)c_1 + (1-i)c_2], \quad \bar{c}_2 \equiv \tfrac{1}{2}\beta_0\gamma^{1/2}[(1-i)c_1 + (1+i)c_2]. \tag{3.10}$$

By using Eqs. (3.5) and (3.10) and choosing the parameters as

$$m_i' = -m_i^*, \quad n_i' = -n_i^*, \quad \eta_{0i}' = \eta_{0i}^*, \tag{3.11}$$

Eq. (3.9) can be rewritten in the variables X, Y as

$$u = -\exp(\eta_i)/[1 + \exp(\eta_i + \eta_i^* + \tau_{ii*}^s)],$$
$$\eta_i \equiv k_i X + l_i Y + \omega_i t + \eta_{0i},$$
$$k_i \equiv [(1-i)m_i + (1+i)n_i]/(2\beta_0\beta^{1/2}),$$
$$l_i \equiv [(1+i)m_i + (1-i)n_i]/(2\beta_0\gamma^{1/2}),$$
$$\omega_i \equiv i(-\beta)k_i^2 + i\gamma l_i^2,$$
$$\exp(-\tau_{ii*}^s) \equiv (k_i + k_i^*)^2 2\beta/\delta + (l_i + l_i^*)^2 2\gamma/\delta. \tag{3.12}$$

This is the 1-soliton solution of Eq. (1.1).

Next we consider generation of lump solitons. Certain lump solitons to the present equation were derived by Satsuma and Ablowitz by taking limiting procedures of the known soliton solutions.[14] Here we derive more general types of lump solitons. Since Eqs. (3.7a) and (3.7b) are linear, arbitrary derivatives of F with respect to various parameters [such as m_i, m_i', n_i, n_i' in Eqs. (3.8)] and their linear combinations are again the solution to Eqs. (3.7a) and (3.7b).[1,6] For our example, we consider only the simplest of such series of arbitrary higher-order derivatives. We take the solution

$$F = (\partial_{m_i} + \partial_{m_i'})\begin{pmatrix} 0, & (m_i + n_i')\exp[m_i x + n_i z + \beta_0^{-1}\gamma_0(n_i - m_i)y + \alpha_0^{-1}(n_i^2 - m_i^2)t + \eta_{0i}] \\ (m_i' + n_i)\exp[m_i' x + n_i' z + \beta_0^{-1}\gamma_0(m_i' - n_i')y + \alpha_0^{-1}(n_i'^2 - m_i'^2)t + \eta_{0i}'], & 0 \end{pmatrix},$$
$$= \begin{pmatrix} 0, & \partial_{m_i}(m_i + n_i')\exp[m_i x - \beta_0^{-1}\gamma_0 m_i y - \alpha_0^{-1}m_i^2 t + \eta_{0i}] \\ \partial_{m_i'}(m_i' + n_i)\exp[m_i' x + \beta_0^{-1}\gamma_0 m_i' y - \alpha_0^{-1}m_i'^2 t + \eta_{0i}'], & 0 \end{pmatrix}$$
$$\times\begin{pmatrix} \exp[n_i'(z - \beta_0^{-1}\gamma_0 y + \alpha_0^{-1}n_i' t)], & 0 \\ 0, & \exp[n_i(z + \beta_0^{-1}\gamma_0 y + \alpha_0^{-1}n_i t)] \end{pmatrix} \equiv f_i^{ls}(x)\bar{f}_i^{ls}(z). \tag{3.13}$$

Via Eq. (2.11), this gives the lump 1-soliton solution

$$u = -[1 + (m_i + n_i')(x - \beta_0^{-1}\gamma_0 y - 2\alpha_0^{-1}m_i t)]\exp(N_i^s)/\{1 - (x - \beta_0^{-1}\gamma_0 y - 2\alpha_0^{-1}m_i t)$$
$$\times(x + \beta_0^{-1}\gamma_0 y - 2\alpha_0^{-1}m_i' t)\exp(D_i^s)\}, \tag{3.14}$$

where N_i^s and D_i^s are the same as in Eq. (3.9). By using Eqs. (3.5), (3.10), and (3.11), solution (3.14) can be written in the variables X, Y as

$$u = -\left\{1 + [\tfrac{1}{2}(1+i)\beta^{1/2}(k_i + k_i^*) + \tfrac{1}{2}(1-i)\gamma^{1/2}(l_i + l_i^*)]\left[\frac{(1-i)X}{2\beta^{1/2}} + \frac{(1+i)Y}{2\gamma^{1/2}} - ((1+i)\beta^{1/2}k_i + (1-i)\gamma^{1/2}l_i)t\right]\right\}$$
$$\times\exp(\eta_i)/\left\{1 + \left[\frac{\delta(X + X_i^{ls}t)^2}{8\beta} + \frac{\delta(Y + Y_i^{ls}t)^2}{8\gamma}\right]\exp(\eta_i + \eta_i^*)\right\},$$
$$X_i^{ls} \equiv -\beta^{1/2}[i\beta^{1/2}(k_i - k_i^*) + \gamma^{1/2}(l_i + l_i^*)],$$
$$Y_i^{ls} \equiv -\gamma^{1/2}[\beta^{1/2}(k_i + k_i^*) - i\gamma^{1/2}(l_i - l_i^*)], \tag{3.15}$$

where k_i, l_i, η_i are the same as in Eq. (3.12). In the special limit of k_i, l_i being pure imaginary $k_i \equiv ik_{R_i}$, $l_i \equiv il_{Ri}$ (k_{Ri}, l_{Ri} = real), by denoting $\exp(\eta_{0i}) \equiv -\rho_i^s$, lump 1-soliton solution (3.15) becomes

$$u = \rho_i^s \exp[i(k_{Ri}X + l_{Ri}Y + \omega_{Ri}t)]/\{1 + 8^{-1}\delta\rho_i^{s*}\rho_i^s[\beta^{-1}(X + 2\beta k_{Ri}t)^2 + \gamma^{-1}(Y - 2\gamma l_{Ri}t)^2]\},$$
$$\omega_{Ri} \equiv \beta k_{Ri}^2 - \gamma l_{Ri}^2. \tag{3.16}$$

Equation (3.16) is the lump 1-soliton solution previously obtained by Nakamura.[8] The lump solitons considered by Satsuma and Ablowitz are this type of solutions, whose denominators are always finite polynominals of x, y, t and constants.

Now we consider generation of ripplons. Equation (3.7) have the following solutions:

$$F(x,z)=\begin{pmatrix}0, & (t+t_i)^{-1}\exp\{\alpha_0[(x-\beta_0^{-1}\gamma_0 y+x_i)^2-(z+\beta_0^{-1}\gamma_0 y+z_i)^2]/[4(t+t_i)]+\theta_{0i}\}\\ (t+t_i)^{-1}\exp\{\alpha_0[(x+\beta_0^{-1}\gamma_0 y+x_i')^2-(z-\beta_0^{-1}\gamma_0 y+z_i')^2]/[4(t+t_i)]+\theta_{0i}'\}, & 0\end{pmatrix},$$

$$=\begin{pmatrix}0, & (t+t_i)^{-1}\exp\{\alpha_0(x-\beta_0^{-1}\gamma_0 y+x_i)^2/[4(t+t_i)]+\theta_{0i}\}\\ (t+t_i)^{-1}\exp\{\alpha_0(x+\beta_0^{-1}\gamma_0 y+x_i')^2/[4(t+t_i)]+\theta_{0i}'\}, & 0\end{pmatrix}$$
$$\times\begin{pmatrix}\exp\{-\alpha_0(z-\beta_0^{-1}\gamma_0 y+z_i')^2/[4(t+t_i)]\}, & 0\\ 0, & \exp\{-\alpha_0(z+\beta_0^{-1}\gamma_0 y+z_i)^2/[4(t+t_i)]\}\end{pmatrix}\equiv f_i^r(x)\bar{f}_i^r(z). \tag{3.17}$$

Via Eq. (2.11), this gives the 1-ripplon solution

$$u=(t+t_i)^{-1}\exp(N_i^r)/\{1-4\alpha_0^{-1}\exp(D_i^r)\},$$
$$N_i^r\equiv\alpha_0[(x-\beta_0^{-1}\gamma_0 y+x_i)^2-(x+\beta_0^{-1}\gamma_0 y+z_i)^2]/[4(t+t_i)]+\theta_{0i},$$
$$D_i^r\equiv N_i^r+\alpha_0[(x+\beta_0^{-1}\gamma_0 y+x_i')^2-(x-\beta_0^{-1}\gamma_0 y+z_i')^2]/[4(t+t_i)]+\theta_{0i}'. \tag{3.18}$$

Here for the integral $\int_x^{+i\infty} ds$ in Eq. (2.11) to converge, we take the parameters as $\mathrm{Im}(x_i-z_i')>0$, $\mathrm{Im}(x_i'-z_i)>0$ for $\alpha_0/(t+t_i)>0$ and $\mathrm{Im}(x_i-z_i')<0$, $\mathrm{Im}(x_i'-z_i)<0$ for $\alpha_0/(t+t_i)<0$.

By using Eqs. (3.5) and (3.10) and choosing the parameters as

$$x_i'=-x_i^*,\ z_i'=-z_i^*,\ \theta_{0i}'=\theta_{0i}^*, \tag{3.19}$$

Eq. (3.18) can be rewritten in the variables X,Y as

$$u=-(t+t_i)^{-1}\exp(\theta_i)[1+\exp(\theta_i+\theta_i^*+\tau_{ii*})]^{-1},$$
$$\theta_i\equiv\frac{i(X+X_i)^2}{4(t+t_i)(-\beta)}+\frac{i(Y+Y_i)^2}{4(t+t_i)\gamma}+\theta_{0i},$$
$$X_i\equiv\tfrac{1}{2}\beta_0\beta^{1/2}[(1+i)x_i+(1-i)z_i],$$
$$Y_i\equiv\tfrac{1}{2}\beta_0\gamma^{1/2}[(1-i)x_i+(1+i)z_i],$$
$$\exp(-\tau_{ii*})\equiv\frac{[(iX_i)+(iX_i)^*]^2}{2\beta\delta}+\frac{[(iY_i)+(iY_i)^*]^2}{2\gamma\delta}. \tag{3.20}$$

This is the 1-ripplon solution first derived by Nakamura through the Hirota bilinear method.[5]

In a manner similar to the derivation of lump solitons from the ordinary solitons, we can generate lump ripplons from the ordinary ripplons. We take the solution for F as

$$F=(\partial_{x_i}+\partial_{x_i'})\begin{pmatrix}0, & (t+t_i)^{-1}(x_i-z_i')\exp\{\alpha_0[(x-\beta_0^{-1}\gamma_0 y+x_i)^2-(z+\beta_0^{-1}\gamma_0 y+z_i)^2]/[4(t+t_i)]+\theta_{0i}\}\\ (t+t_i)^{-1}(x_i'-z_i)\exp\{\alpha_0[(x+\beta_0^{-1}\gamma_0 y+x_i')^2-(z-\beta_0^{-1}\gamma_0 y+z_i')^2]/[4(t+t_i)]+\theta_{0i}'\}, & 0\end{pmatrix}$$
$$\equiv\begin{pmatrix}0, & \partial_{x_i}(t+t_i)^{-1}(x_i-z_i')\exp\{\alpha_0(x-\beta_0^{-1}\gamma_0 y+x_i)^2/[4(t+t_i)]+\alpha_0 a_i x_i/2+\bar{\theta}_{0i}\}\\ \partial_{x_i'}(t+t_i)^{-1}(x_i'-z_i)\exp\{\alpha_0(x+\beta_0^{-1}\gamma_0 y+x_i')^2/[4(t+t_i)]+\alpha_0 a_i' x_i'/2+\bar{\theta}_{0i}'\}, & 0\end{pmatrix}$$
$$\times\begin{pmatrix}\exp\{-\alpha_0(z-\beta_0^{-1}\gamma_0 y+z_i')^2/[4(t+t_i)]\}, & 0\\ 0, & \exp\{-\alpha_0(z+\beta_0^{-1}\gamma_0 y+z_i)^2/[4(t+t_i)]\}\end{pmatrix}\equiv f_i^{lr}(x)\bar{f}_i^{lr}(z). \tag{3.21}$$

Here we set $\theta_{0i}=\alpha_0 a_i x_i/2+\bar{\theta}_{0i}$, $\theta_{0i}'=\alpha_0 a_i' x_i'/2+\bar{\theta}_{0i}'$ with $a_i, a_i', \bar{\theta}_{0i}, \bar{\theta}_{0i}'$ representing arbitrary constants whose derivatives with respect to x_i, x_i' vanish. Via Eq. (2.11), this gives lump 1-ripplon solutions as

$$u=(t+t_i)^{-1}\{1+(x_i-z_i')[x-\beta_0^{-1}\gamma_0 y+x_i+a_i(t+t_i)]/[2(t+t_i)]\}\exp(N_i^r)/$$
$$\{1-(t+t_i)^{-2}[x-\beta_0^{-1}\gamma_0 y+x_i+a_i(t+t_i)][x+\beta_0^{-1}\gamma_0 y+x_i'+a_i'(t+t_i)]\exp(D_i^r)\}, \tag{3.22}$$

where N_i^r, D_i^r are the same as in Eq. (3.18). By using Eqs. (3.5), (3.10), and (3.19), the solution (3.22) can be rewritten in the variables X,Y as

$$u=-(t+t_i)^{-1}\{1+\frac{1}{8(t+t_i)}$$
$$\times[(1-i)(X_i-X_i^*)/\beta^{1/2}+(1+i)(Y_i-Y_i^*)/\gamma^{1/2}]$$
$$\times[(1-i)(X+X_i+\bar{A}_i t)/\beta^{1/2}$$
$$+(1+i)(Y+Y_i+\bar{B}_i t)/\gamma^{1/2}]\}\exp(\theta_i)/$$
$$\Big\{1+\frac{\delta}{8(t+t_i)^2}[(X+\bar{X}_i+\bar{A}_i(t+t_i))^2/\beta$$
$$+(Y+\bar{Y}_i+\bar{B}_i(t+t_i))^2/\gamma]\exp(\theta_i+\theta_i^*)\Big\},$$
$$\bar{X}_i\equiv\tfrac{1}{2}\beta^{1/2}[(X_i+X_i^*)/\beta^{1/2}+i(Y_i-Y_i^*)/\gamma^{1/2}],$$
$$\bar{Y}_i\equiv\tfrac{1}{2}\gamma^{1/2}[-i(X_i-X_i^*)/\beta^{1/2}+(Y_i+Y_i^*)/\gamma^{1/2}],$$
$$\bar{A}_i\equiv\tfrac{1}{2}\beta_0\beta^{1/2}[(1+i)a_i+(1-i)a_i'],$$
$$\bar{B}_i\equiv\tfrac{1}{2}\beta_0\gamma^{1/2}[(1-i)a_i+(1+i)a_i'], \tag{3.23}$$

where X_i, Y_i, θ_i are the same as in Eq. (3.20). This is the lump 1-ripplon solution of Eqs. (1.1).

We choose the parameter $a_i'=-a_i^*$ so that $\bar{A}_i$ and $\bar{B}_i$ become real arbitrary constants. In the special limit of X_i, Y_i

being real or $X_i \equiv X_{Ri}$, $Y_i \equiv Y_{Ri}$, the present lump 1-ripplon becomes

$$u = -(t+t_i)^{-1}\exp\left[\frac{i(X+X_{Ri})^2}{4(t+t_i)(-\beta)} + \frac{i(Y+Y_{Ri})^2}{4(t+t_i)\gamma} + \theta_{0i}\right] \Big/ \left\{1 + \frac{\delta}{8(t+t_i)^2}[(X+X_{Ri}+\overline{A}_i(t+t_i))^2/\beta + (Y+Y_{Ri}+\overline{B}_i(t+t_i))^2/\gamma]\exp(\theta_{0i}+\theta^*_{0i})\right\}, \quad (3.24)$$

where X_{Ri}, Y_{Ri}, $\overline{A}_i$, $\overline{B}_i$ (θ_{0i}) are real (complex) arbitrary constants. This is the simplified lump 1 − ripplon solution first obtained by Nakamura.[8]

So far we have obtain 1-soliton, lump 1-soliton, 1-ripplon, and lump 1 -ripplon in the framework of IST formalism. As mentioned earlier, this implies that all of these can be superposed. It is seen as follows. Since Eqs. (3.7) is linear for F, obviously arbitrary linear combinations of the solutions are again the solutions for F. Thus we can take the general superposed solution of F as

$$F(x,z) = \sum_{i=1}^{N_s} f_i^s(x)\overline{f}_i^s(z) + \sum_{i=N_s+1}^{N_s+N_{ls}} f_i^{ls}(x)\overline{f}_i^{ls}(z) + \sum_{i=N_s+N_{ls}+1}^{N_s+N_{ls}+N_r} f_i^r(x)\overline{f}_i^r(z) + \sum_{i=N_s+N_{ls}+N_r+1}^{N_s+N_{ls}+N_r+N_{lr}} f_i^{lr}(x)\overline{f}_i^{lr}(z). \quad (3.25)$$

Here $f_i^s, \overline{f}_i^s; f_i^{ls}, \overline{f}_i^{ls}; f_i^r, \overline{f}_i^r$ and $f_i^{lr}, \overline{f}_i^{lr}$ are the same as those given, respectively, in Eqs., (3.8), (3.13), (3.17), and (3.21). The values of various constants can be taken arbitrarily different for each i.

Then, at least in principle, $K(x,x)$ and thus $u(u(x) = K_{12}(x,x))$ can be solved by Eqs. (2.9b) and (2.9c). This solution corresponds to the superposition solution of N_s-soliton, N_{ls}-lump soliton, N_r-ripplon, and N_{lr}-lump ripplon.

ACKNOWLEDGMENTS

This work was partially supported by AFOSR and ONR.

[1] R. S. Johnson and S. Thompson, Phys. Lett. A **66**, 279 (1978).
[2] L. G. Redekopp, Stud. Appl. Math. **63**, 185 (1980).
[3] A. Nakamura, Phys. Rev. Lett. **46**, 751 (1981).
[4] A. Nakamura, J. Phys. Soc. Jpn. **50**, 2467 (1981).
[5] A. Nakamura, J. Math. Phys. **23**, 417, (1982).
[6] A. Nakamura, J. Phys. Soc. Jpn. **51**, 19 (1982).
[7] A. Nakamura, "A one-space-dimensional nonlinear evolution equation which admits simple similarity-type explode-decay mode solutions, periodic waves and solitons," (1981).
[8] A. Nakamura, "Explode-decay mode lump solitons of a two-dimensional nonlinear Schrodinger equation," Phys. Lett. A **88**, 55 (1982).
[9] D. J. Benney and G. J. Roskes, Stud. Appl. Math. **48**, 377 (1969).
[10] A. Davey and K. Stewartson, Proc. R. Soc. London Ser. A **338**, 101 (1974).
[11] N. C. Freeman and A. Davey, Proc. R. Soc. London Ser. A **344**, 427 (1975).
[12] M. J. Ablowitz and R. Haberman, Phys. Rev. Lett. **35**, 1185 (1975).
[13] D. Anker and N. C. Freeman, Proc. R. Soc. London Ser. A **360**, 529 (1978).
[14] J. Satsuma and M. J. Ablowitz, J. Math. Phys. **20**, 1496 (1979).
[15] V. E. Zakharov and A. B. Shabat, Func. Anal. Appl. **8**, 226 (1974).

Korteweg-de Vries surfaces and Bäcklund curves[a)]

B. A. Kupershmidt
Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545[b)]

(Received 8 January 1982; accepted for publication 19 February 1982)

It is shown that every point $w(\epsilon)$ on the curve $\gamma_r(\epsilon)$ representing a 1-parameter family of integrable equations containing a given rth Korteweg–de Vries (KdV) equation $w(0)$, also belongs to a different integrable curve $\Gamma_r(\epsilon,\nu)$. Symmetries of the resulting surface make it possible to construct a curve of Bäcklund transformations, that is, infinitesimal automorphisms, of points on $\gamma_r(\epsilon)$ starting with the usual infinitesimal automorphisms of $w(0)$. In addition, we obtain four new Bäcklund transformations of the second order for all higher KdV equations.

PACS numbers: 02.30.Jr, 02.30.Qy

1. INTRODUCTION

The main results of this paper concern the existence of a second parameter for higher KdV equations (see Ref. 1) and infinitesimal automorphisms for equations associated with the first parameter (see Ref. 2).

We begin by explaining this in the simplest case of the KdV equation itself,

$$u_t = 6uu_x - u_{xxx} = X_2(u). \tag{1.1}$$

Here the subscripts t and x denote partial derivatives with respect to t and x, respectively, and $X_r(u)$ corresponds to the rth flow in the hierarchy of KdV equations which can be given by the Lax representation

$$L_t = [P, L\,], \tag{1.2}$$

where

$$L = -\xi^2 + u, \quad \xi \equiv \partial/\partial x, \tag{1.3}$$

and P is an isobaric differential operator of order $2r-1$, $r \geqslant 1$.

Now, Gardner observed (see Ref. 3) that is w satisfies the Gardner equation

$$w_t = 6ww_x - w_{xxx} + 6\epsilon^2 w^2 w_x, \tag{1.4}$$

then $u = (g(\epsilon))\,(w)$, given by the map

$$g(\epsilon): w \rightarrow u = w + \epsilon^2 w^2 + \epsilon w_x, \tag{1.5}$$

satisfies the KdV equation (1.1). Originally, this fact had provided the shortest proof that the KdV equation has an infinite number of conserved densities $H_q(u)$, $q = 1,2,...$, that is, equalities of the form

$$\frac{\partial H_q(u)}{\partial t} = \partial J_q(u), \quad \partial \equiv \partial/\partial x,$$

which follow formally from (1.1). Here $H_q(u)$ and $J_q(u)$ are differential polynomials in u, i.e., polynomials in u and its x derivatives $u^{(j)} = \partial^j u/\partial x^j$. This one-line proof consists of inverting (1.5):

$$w = \sum_{q=0}^{\infty} h_q(u)\epsilon^q,$$

then substituting this expresssion in (1.4) rewritten in the conservation form

$$\frac{\partial}{\partial t} w = \partial(3w^2 - w_{xx} + 2\epsilon^2 w^3),$$

and finally equating powers of ϵ in both sides of the derived equality.

Let us look more closely at the Gardner equation (1.4). First, notice that all conserved densities $H_q(u)$ of the KdV equation generate, via the map $g(\epsilon)$ (1.5), conserved densities $H_q((g(\epsilon))(w))$ of the Gardner equation. Thus the Gardner equation is *integrable,* meaning: has an infinite number of conserved densities. Secondly, the Gardner equation depends polynomially upon ϵ and therefore can be considered as a *curve* in the space of all (evolution) equations, and an *integrable curve* at that. For the points on this curve we can use the suggestive notation $\gamma_2(\epsilon)$. Thus the KdV equation is just the point $\gamma_2(0)$, and we can consider it as the natural base point of the curve $\gamma_2(\epsilon)$. In general, when an integrable system is included in an integrable family, we call such family a *deformation* of the original system. Moreover, we can and shall take one further step: When every member σ of an integrable family Σ is included in a one-parameter integrable curve $\sigma(\epsilon)$ which intersects Σ only at $\sigma = \sigma(0)$, we call the new family $\cup_{\sigma\in\Sigma}\sigma(\epsilon) = \Sigma(\epsilon)$ a *deformation* of Σ. Thirdly, the Gardner deformation (1.4) has one additional property: it is supplemented by the map $g(\epsilon)$ (1.5) which is regular in ϵ and sends all points on the curve $\gamma_2(\epsilon)$ into one point $\gamma_2(0)$, if we agree not to distinguish between equations and their solutions. Such an agreement is certainly legitimate if we keep in mind that our evolution equations represent just a traditional way of writing evolution fields (see, e.g., Chap. I of Ref. 4). In general, we call a map regular in ϵ, $f(\epsilon)$: $\Sigma(\epsilon) \rightarrow \Sigma(0) = \Sigma$, such that $f(0) = \mathrm{Id}$, a *reduction*. It should be immediately noted that reductions do not have to and often cannot accompany deformations.

The alert reader might have remarked that the above definition of deformations must be augmented by a device to exclude curves with systems which are mutually "equivalent" in one sense or another, say, under changes of variables. Another source of triviality may occur when, for a given evolution field $\sigma(0)$, the increment $\sigma(\epsilon) - \sigma(0)$, considered also as an evolution field, commutes with $\sigma(0)$ or has an infinity of conserved densities in common with it. Fortunately, for all systems considered in this paper, the nontriviality of deformations follows from a few general statements. This will be done in the last Sec. 5, in order not to interrupt the presentation.

[a)] Supported in part by NSF Grant MCS 800 3104 and the US Department of Energy.

[b)] Permanent address: Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109.

To sum up: The KdV equation allows the deformation (1.4) together with the reduction (1.5). In reality, even a more general phenomenon occurs (see Ref. 2): one can *deform further* the curve (1.4). Namely, if

$$p_t = 6p_x C(1+\epsilon^2 C) - p_{xxx} + 2\epsilon^2 v^2 p_x^3, \tag{1.6}$$

where

$$C = C(\epsilon,v,p) = \frac{\sinh(2\epsilon vp)}{2\epsilon v} + \frac{\sinh^2(\epsilon vp)}{\epsilon^2}, \tag{1.7}$$

then $w = (G(\epsilon,v))(p)$, where the map $G(\epsilon,v)$ is defined by

$$G(\epsilon,v): p \to w = C(\epsilon,v,p) + vp_x, \tag{1.8}$$

satisfies the Gardner equation (1.4). Notice that on the resulting surface $\Gamma_2(\epsilon,v)$, consisting of the points (1.6), those points with $v = 0$ form the curve $\gamma_2(\epsilon)$ and the map (1.8) provides a reduction of the surface $\Gamma_2(\epsilon,v)$ onto the curve $\gamma_2(\epsilon)$.

Now, it was proved in Ref. 2 that there exist curves $\gamma_r(\epsilon)$, with the same reduction (1.5), which deform all the higher KdV equations. Our first problem is, then: Are there any analogous surfaces $\Gamma_r(\epsilon,v)$ for $r > 2$, preferably with the same reduction (1.8), as for $r = 2$? The construction of $\gamma_r(\epsilon)$ in Ref. 2 was based on the fact that for every r, the rth KdV equation is a bi-Hamiltonian system and the reduction (1.5) can be interpreted as a "canonical transformation" (see, e.g., Ref. 5). Since the Gardner equation is not bi-Hamiltonian anymore and (1.6) is not even Hamiltonian, the reasoning based on the Hamiltonian formalism is no longer applicable. On the other hand, note that $G(0,v) = g(v)$. This implies that any construction of the surfaces $\Gamma_r(\epsilon,v)$ will provide, as a bonus, a construction of the curve $\gamma_r(\epsilon)$ as $\Gamma_r(0,\epsilon)$. Conversely, if we are to build $\Gamma_r(\epsilon,v)$ starting with $\gamma_r(\epsilon)$, we should first of all find a non-Hamiltonian construction of $\gamma_r(\epsilon)$ such that the method can be deformed into construction of $\Gamma_r(\epsilon,v)$.

This is exactly what we shall do. Here is the clue of how to proceed. Recall that, parallel to the hierarchy of the KdV equations (1.2) and (1.3), there is another hierarchy of Modified KdV (MKdV) equations (see, e.g., Ref. 5), given by (1.2), with

$$L = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\xi + \begin{pmatrix} 0 & v \\ -v & 0 \end{pmatrix}; \tag{1.9}$$

for example, MKdV_2 has the form

$$v_t = 6v^2 v_x - v_{xxx}. \tag{1.10}$$

The main relation between the two hierarchies is provided by the map M ("Miura transformation")

$$M: v \to u = v^2 + v_x, \tag{1.11}$$

which maps solutions of MKdV_r into those of KdV_r.

If we now put (Ref. 2, Sec. 5)

$$v = \epsilon(w + \epsilon^{-2}/2), \tag{1.12}$$

$$\tilde{u} = u - \epsilon^{-2}/4, \tag{1.13}$$

then (1.11) turns into

$$\tilde{u} = w + \epsilon^2 w^2 + \epsilon w_x, \tag{1.14}$$

which is almost (1.5); a few more remarks will drop the tilde from $\tilde{u}$. We shall do this in Sec. 2. In Sec. 3 we construct: (a) curves $\bar{\gamma}_r(\epsilon)$ which deform the MKdV equations, (b) their reductions $\bar{g}(\epsilon): \bar{\gamma}_r(\epsilon) \to \bar{\gamma}_r(0)$, and (c) the deformation of the map M, $M(\epsilon): \gamma_r(\epsilon) \to \gamma_r(\epsilon)$. Starting with the map $M(\epsilon)$, in Sec. 4 we adapt our arguments of Sec. 2 to depend upon the new parameter v. This enables us to construct the desired surfaces $\Gamma_r(\epsilon,v)$ together with their reduction $G(\epsilon,v)$.

Let us now turn to our second problem, that of existence of Bäcklund curves. Note that the Gardner equation (1.4) depends upon ϵ^2 while the reduction (1.5) depends upon ϵ. Therefore, we obtain a "Bäcklund transformation," that is, an infinitesimal automorphism, of the KdV equation, simply as

$$b(\epsilon) = g(-\epsilon) \circ g(\epsilon)^{-1}. \tag{1.15}$$

It was laboriously proved in Ref. 6, Chap. III, that (1.15) provides the Bäcklund transformation for all higher KdV equations as well. We give a very short proof of this fact in Sec. 2.

Now observe that (1.6) depends on v^2 whereas the reduction $G(\epsilon,v)$ depends on v. As before, one gets a Bäcklund transformation of the Gardner equation:

$$B(\epsilon,v) = G(\epsilon,-v) \circ G(\epsilon,v)^{-1}, \tag{1.16}$$

which deforms $b(\epsilon)$ since $b(\epsilon) = B(0,\epsilon)$. Combining (1.5) and (1.16), we obtain four new Bäcklund transformations of the KdV equation:

$$g(\pm\epsilon) \circ G(\epsilon,-v) \circ G(\epsilon,v)^{-1} \circ g(\pm\epsilon)^{-1}. \tag{1.17}$$

As might be expected, equations from $\Gamma_r(\epsilon,v)$ also depend upon v^2, for all r. We prove this in Sec. 4. This will guarantee that Bäcklund transformations (1.16) and (1.17) are valid also for all r's.

2. CONSTRUCTION OF THE CURVES $\gamma_r(\epsilon)$

Let X_r and Y_r denote the rth MKdV and MKdV field respectively. Their "trajectories" are solutions of corresponding evolution equations

$$u_t = X_r(u),$$

$$v_t = Y_r(v).$$

More generally, let $\alpha_1,\ldots,\alpha_n$ be arbitrary constants (say, from $\mathbb{R}$ or $\mathbb{C}$). Consider linear combinations

$$X^\alpha = \Sigma \alpha_i X_i, \tag{2.1}$$

$$Y^\alpha = \Sigma \alpha_i Y_i \tag{2.2}$$

and their solutions

$$u_t = X^\alpha(u), \tag{2.3}$$

$$v_t = Y^\alpha(v). \tag{2.4}$$

We shall use the following well-known facts.

Proposition 2.1: (i) If v satisfies (2.4), then $u = M(v) = v^2 + v$ satisfies (2.3) (see, e.g., Ref. 7); (ii) if v satisfies (2.4), then $(-v)$ satisfies it also (see, e.g., Ref. 7). Equivalently,

$$Y^\alpha(-v) = -Y^\alpha(v). \tag{2.5}$$

(iii) For any constant c, there exists a lower-triangular matrix Ω^c which has ones on the diagonal, depends polynomially upon c, and such that

$$X^{\alpha}(\bar{u}+c)=X^{\bar{\alpha}}(\bar{u}), \tag{2.6}$$

where the vectors α and $\bar{\alpha}$ are related by

$$\bar{\alpha}=\Omega^{c}\alpha. \tag{2.7}$$

In other words, if u satisfies (2.3) then $\bar{u}=u-c$ satisfies

$$\bar{u}_t=X^{\bar{\alpha}}(\bar{u}). \tag{2.8}$$

Proposition 2.1(iii) follows from Ref. 1, Chap. II, formula (17).

Now let us rescale the modified variable v by

$$w=\epsilon^{-1}v-\tfrac{1}{2}\epsilon^{-2}. \tag{2.9}$$

If v satisfies (2.4), then w will satisfy

$$\dot{w}=Z^{\alpha}(w), \tag{2.10}$$

where

$$Z^{\alpha}(w)=\epsilon^{-1}Y^{\alpha}\big[\epsilon(w+\tfrac{1}{2}\epsilon^{-2})\big]. \tag{2.11}$$

Simultaneously, the map M (1.11) will send solutions of (2.10) into solutions of (2.3) via

$$\begin{aligned} u &= v^2+v_x=\epsilon^2(w+\tfrac{1}{2}\epsilon^{-2})^2+\epsilon w_x\\ &=\tfrac{1}{4}\epsilon^{-2}+w+\epsilon^2w^2+\epsilon w_x. \end{aligned}$$

Therefore, applying Proposition 2.5 (iii) with

$$c=\tfrac{1}{4}\epsilon^{-2}, \tag{2.12}$$

we obtain the map

$$g(\epsilon):w\rightarrow\bar{u}=w+\epsilon^2w^2+\epsilon w_x, \tag{2.13}$$

which sends solutions of (2.10) into solutions of (2.8), with $\bar{\alpha}=\Omega^{1/4\epsilon^2}\alpha$. Since $(\Omega^{c})^{-1}=\Omega^{-c}$, we get finally

Theorem 2.2: For any $\alpha=(\alpha_1,\ldots,\alpha_n)$, if w satisfies

$$w_t=Z^{\beta}(w),\quad \beta=\Omega^{-1/4\epsilon^2}\alpha, \tag{2.14}$$

then $u=(g(\epsilon))(w)$ satisfies (2.3).

Remark 2.3: Thus, we can deform not only individual KdV equations but also their linear combinations.

Remark 2.4: We do not need to worry about whether a deformed equation [such as (2.14)] is regular in a deformation parameter. It is always regular so long as a reduction is regular [such as (2.13)] (see, e.g., Ref. 2).

Now it is clear how the deformed equations (2.14) depend upon ϵ.

Theorem 2.5: The right-hand side of (2.14) depends upon ϵ^2.

Proof: $\Omega^{-1/4\epsilon^2}$ is polynomial in $(-1/4\epsilon^2)$, by Proposition 2.1 (iii); therefore $\beta=\epsilon^{-1/4\epsilon^2}\alpha$ is even in ϵ. Since $Z^{\beta}(w)=\Sigma\beta_iZ_i(w)$, it is enough to look only at the $Z_i(w)$. By (2.11),

$$Z_i(w,\epsilon)=\epsilon^{-1}Y_i\big[\epsilon(w+\tfrac{1}{2}\epsilon^{-2})\big].$$

Hence, by (2.5),

$$\begin{aligned} Z_i(w,-\epsilon) &= (-\epsilon)^{-1}Y_i\big[-\epsilon(w+\tfrac{1}{2}\epsilon^{-2})\big]\\ &= -\epsilon^{-1}(-1)Y_i\big[\epsilon(w+\tfrac{1}{2}\epsilon^{-2})\big]\\ &= Z_i(w,\epsilon). \end{aligned}$$

3. THE CURVES $\bar{\gamma}_r(\epsilon)$ AND THE MAP $M(\epsilon)$

In this section we prepare the ground for the construction of the surfaces $\Gamma_r(\epsilon,\nu)$ in Sec. 4. As was mentioned in the Introduction, what we are aiming at is a curve of propositions which goes through Proposition (2.1).

Let us consider two maps

$$M(\epsilon):q\rightarrow w=\frac{\sinh^2\epsilon q}{\epsilon^2}+q_x, \tag{3.1}$$

$$Mg(\epsilon):q\rightarrow v=\frac{\sinh(2\epsilon q)}{2\epsilon}+\epsilon q_x. \tag{3.2}$$

Lemma 3.1: We have the commutative diagram

$$M\circ Mg(\epsilon)=g(\epsilon)\circ M(\epsilon). \tag{3.3}$$

Proof: We have

$$\begin{aligned} (M\circ Mg(\epsilon))(q) &= M\Big(\frac{\sinh2\epsilon q}{2\epsilon}+\epsilon q_x\Big)\\ &= \Big(\frac{\sinh2\epsilon q}{2\epsilon}+\epsilon q_x\Big)^2+\Big(\frac{\sinh2\epsilon q}{2\epsilon}+\epsilon q_x\Big)_x\\ &= \frac{\sinh^2 2\epsilon q}{4\epsilon^2}+\sinh2\epsilon q\cdot q_x+\epsilon^2q_x^2\\ &\quad+\cosh2\epsilon q\cdot q_x+\epsilon q_{xx}, \end{aligned}$$

$$\begin{aligned} (g(\epsilon)\circ M(\epsilon))(q) &= (g(\epsilon))\Big(\frac{\sinh^2\epsilon q}{\epsilon^2}+q_x\Big)\\ &= \frac{\cosh2\epsilon q-1}{2\epsilon^2}+q_x+\epsilon^2\Big(\frac{\cosh2\epsilon q-1}{2\epsilon^2}+q_x\Big)^2\\ &\quad+\epsilon\Big(\frac{\cosh2\epsilon q-1}{2\epsilon^2}+q_x\Big)_x\\ &= \frac{\cosh2\epsilon q-1}{2\epsilon^2}+q_x+\frac{\cosh^2 2\epsilon q-2\cosh2\epsilon q+1}{4\epsilon^2}\\ &\quad+(\cosh2\epsilon q-1)q_x+\epsilon^2q_x^2+\sinh2\epsilon q\cdot q_x+\epsilon q_{xx}. \end{aligned}$$

Remark: $M(\epsilon)$ can be considered as a deformation of $M=M(0)$.

Definition 3.2: $\bar{\gamma}_r(\epsilon)=(Mg(\epsilon)^{-1})(Y_r)$.

Corollary 3.3: $\bar{\gamma}_r(\epsilon)$ depends upon ϵ^2.

Proof: By Lemma 3.1, $\bar{\gamma}_r(\epsilon)=(M(\epsilon)^{-1})(\gamma_r(\epsilon))$. By (3.1), $M(\epsilon)$ depends upon ϵ^2; by Theorem 2.5, $\gamma_r(\epsilon)$ depends upon ϵ^2.

Remarks: (i) As in Sec. 2, we could defined deformations not only of fields Y_r but also of their linear combinations Y^{α} by

$$\bar{\gamma}^{\alpha}(\epsilon)=(Mg(\epsilon)^{-1})(Y^{\alpha}). \tag{3.4}$$

Corollary 3.3 would still remain true. (ii) The map (3.2) is the reduction of $\bar{\gamma}^{\alpha}(\epsilon)$ onto Y^{α}. (iii) Of course, one would like to know that points on the curve $\bar{\gamma}_r(\epsilon)$ or [or $\bar{\gamma}^{\alpha}(\epsilon)$] represent finite fields, that is, evolution equations with only a finite number of derivatives. That this is true, was essentially proven many times in the physical literature (see, e.g., Ref. 8). For the sake of completeness, the proof is reproduced here.

The Lax representations (1.2) for the MKdV equations are the integrability conditions for

$$L\psi=\lambda\psi, \tag{3.5}$$

$$\psi_t=P\psi, \tag{3.6}$$

where L is given by (1.9), $\psi=(\psi_1,\psi_2)^t$ is a column vector, and P runs over special differential operators which make sense of (1.2) (see, e.g., Ref. 5). Equivalently, we can think of P as a multiplication operator, if we express x derivatives of ψ's using (3.5). In any case, what we need is the possibility to rewrite (3.6) as an evolution equation for $\Gamma = \psi_1/\psi_2$:

$$\Gamma_t = a + b\Gamma + c\Gamma^2, \tag{3.7}$$

where a, b, and c are some polynomials in λ and differential polynomials in v. On the other hand, (3.5) is equivalent to

$$\Gamma_x = 2\lambda\Gamma + v(\Gamma^2 - 1), \tag{3.8}$$

which can be transformed into (3.2) by identifications

$$\lambda = \epsilon^{-1}/2, \quad \Gamma = \tanh(\epsilon q).$$

Substituting this into (3.7) and eliminating v in favor of q by (3.2), we arrive at the desired equations.

4. CONSTRUCTION OF THE SURFACES $\Gamma_r(\epsilon,\nu)$

We shall follow the path of Sec. 2. For this we need an analog of Proposition (2.1) for the pair $\bar{\gamma}^\alpha(\epsilon)$, $\gamma^\alpha(\epsilon)$. We already have Theorem (2.2), which is an analog of Proposition 2.1 (i). The generalization of the property 2.1 (ii) for $\bar{\gamma}^\alpha(\epsilon)$ is given by

Proposition 4.1: Let us write evolution equations for points on the curve $\bar{\gamma}^\alpha(\epsilon)$ as

$$q_t = V^\alpha(q,\epsilon). \tag{4.1}$$

If q satisfies (4.1) then $(-q)$ satisfies it also. In other words,

$$V^\alpha(-q,\epsilon) = -V^\alpha(q,\epsilon). \tag{4.2}$$

Proof: Let $v = (Mg(\epsilon))(q)$. By (3.2),

$$-q = -(Mg(\epsilon))^{-1}(v) = (Mg(\epsilon)^{-1})(-v),$$

and Proposition 2.1 (ii) does the job.

The property 2.1 (iii) can be generalized as follows.

Proposition 4.2: Let w be a point on the curve $\gamma^\alpha(\epsilon)$ satisfying (2.14), which we shall write as

$$w_t = E^\alpha(w,\epsilon^2). \tag{4.3}$$

Then, for any constant $\bar{c}$,

$$\hat{w} = (1 + 2\epsilon^2\bar{c})(w - \bar{c}) \tag{4.4}$$

is a point on the curve $\gamma^{\hat{\alpha}}(\hat{\epsilon})$, where

$$\hat{\epsilon} = \frac{\epsilon}{1 + 2\epsilon^2\bar{c}}, \tag{4.5}$$

$$\hat{\alpha} = \Omega^{\bar{c} + \epsilon^2\bar{c}^2}\alpha. \tag{4.6}$$

Proof: We begin with w from (4.3) and u from (2.3), connected via Theorem 2.2. Then we have

$$\begin{aligned}\bar{u} &= u - \bar{c} - \epsilon^2\bar{c}^2 = w + \epsilon^2 w^2 + \epsilon w_x - \bar{c} - \epsilon^2\bar{c}^2\\ &= (w - \bar{c})(1 + 2\epsilon^2\bar{c}) + \epsilon^2(w - \bar{c})^2 + \epsilon(w - \bar{c})_x\\ &= \hat{w} + \hat{\epsilon}^2\hat{w}^2 + \hat{\epsilon}\hat{w}_x,\end{aligned}$$

by (4.5). Thus $g(\hat{\epsilon})$ sends $\hat{w}$ to $\bar{u} = u - c$, where $c = \bar{c} + \epsilon^2\bar{c}^2$. Now apply Proposition (2.1) (iii) and get (4.6).

Remark: For $\epsilon = 0$, Proposition 4.2 is just Proposition 2.1 (iii).

Now we can easily repeat the reasoning of Sec. 2 and get

Theorem 4.3: Let q be a point on the curve $\bar{\gamma}^\alpha(\epsilon)$ and let $w = (M(\epsilon))(q)$ be a corresponding point on $\gamma^\alpha(\epsilon)$. Then

$$p = \nu^{-1}\,(1 + \epsilon^2\nu^{-2})^{1/2}(q - \tfrac{1}{2}\epsilon^{-1}\sinh^{-1}\epsilon\nu^{-1}) \tag{4.7}$$

lies on the surface $\Gamma^{\hat{\alpha}}(\hat{\epsilon},\nu)$, where

$$\hat{\epsilon} = \epsilon(1 + \epsilon^2\nu^{-2})^{-1/2}, \tag{4.8}$$

and $\hat{\alpha}$ is given by (4.6) with

$$\bar{c} = \tfrac{1}{2}\nu^{-2}[1 + (1 + \epsilon^2\nu^{-2})^{1/2}]^{-1}. \tag{4.9}$$

The reduction $G(\hat{\epsilon},\nu)$ sends p to $\hat{w}$, which is related to $w = (M(\epsilon))\,(q)$ by (4.4).

Proof: We have to show that

$$(G(\hat{\epsilon},\nu))(p) = \hat{w}. \tag{4.10}$$

First substitute (4.9) into (4.5) and get

$$1 + 2\epsilon^2\bar{c} = (1 + \epsilon^2\nu^{-2})^{1/2},$$

$$\hat{\epsilon} = \epsilon(1 + 2\epsilon^2\bar{c})^{-1} = \epsilon(1 + \epsilon^2\nu^{-2})^{-1/2}, \tag{4.11}$$

which is (4.8). Now

$$\begin{aligned}&(G(\hat{\epsilon},\nu))(p)\\ &= \frac{\sinh 2\hat{\epsilon}\nu p}{2\hat{\epsilon}\nu} + \frac{\cosh 2\hat{\epsilon}\nu p - 1}{2\hat{\epsilon}^2} + \nu p_x\\ &= (1 + 2\epsilon^2\bar{c})\left[\frac{\sinh 2\hat{\epsilon}\nu p}{2\epsilon\nu} + \frac{\cosh 2\hat{\epsilon}\nu p - 1}{2\epsilon\hat{\epsilon}}\right.\\ &\quad \left. + \frac{\nu}{1 + 2\epsilon^2\bar{c}}p_x\right].\end{aligned}$$

On the other hand,

$$\begin{aligned}\frac{\hat{w}}{1 + 2\epsilon^2\bar{c}} &= (w - \bar{c}) = \frac{\cosh 2\epsilon q - 1}{2\epsilon^2} + q_x - \bar{c}\\ &= \frac{1}{2\epsilon^2}\cosh 2\epsilon\left(\frac{\nu p}{(1 + \epsilon^2\nu^{-2})^{1/2}} + \frac{1}{2\epsilon}\sinh^{-1}\frac{\epsilon}{\nu}\right)\\ &\quad + \frac{\nu}{(1 + \epsilon^2\nu^{-2})^{1/2}}p_x - \bar{c} - \frac{1}{2\epsilon^2}\\ &= \frac{1}{2\epsilon^2}\left[\cosh 2\hat{\epsilon}\nu p(1 + \epsilon^2/\nu^2)^{1/2} + \sinh\hat{\epsilon}\nu p\cdot\frac{\epsilon}{\nu}\right]\\ &\quad + \frac{\nu}{1 + 2\epsilon^2\bar{c}}p_x - \frac{1}{2\epsilon\hat{\epsilon}},\end{aligned}$$

which is just what is needed.

Corollary 4.4: Points on $\Gamma^\alpha(\epsilon,\nu)$ depend upon ν^2 (notice that we have dropped hats from α and ϵ).

Proof: By (4.9), $\bar{c}$ depends upon ν^2 and by (4.5) and (4.6), the same is true for $\hat{\alpha}$ and $\hat{\epsilon}$. If q is a point on $\bar{\gamma}^\alpha(\epsilon)$ satisfying (4.1), then (4.7) yields

$$\begin{aligned}p_t &= \nu^{-1}(1 + \epsilon^2\nu^{-2})^{1/2}q_t\\ &= \nu^{-1}(1 + \epsilon^2\nu^{-2})^{1/2}V^\alpha[\nu(1 + \epsilon^2\nu^{-2})^{-1/2}p\\ &\quad + \tfrac{1}{2}\epsilon^{-1}\sinh^{-1}\epsilon\nu^{-1}],\end{aligned}$$

which is an even function of ν, by (4.2).

Remark: As was mentioned in the Introduction, our construction of $\Gamma_r(\epsilon,\nu)$ can be considered as a deformation of the construction of $\Gamma_r(\epsilon)$. To see this, put $\epsilon = 0$ in (4.7) and recover (2.9), with ν renamed ϵ.

5. NONTRIVIALITY OF DEFORMATIONS

As was mentioned in the Introduction, we have to investigate three possible types of trouble.

(A) First we check that evolution equations $\sigma = \sigma(0)$ and $\sigma(\epsilon)$ are not equivalent under any changes of variables. Here $\sigma(\epsilon)$ stands for any of the deformations constructed in the preceding sections.

To do this we use the existence of reductions for all deformations constructed above. Let us write them as

$$f(\epsilon): u_i \mapsto u_i + \epsilon\phi_i, \quad i = 1,...,n, \tag{5.1}$$

where n is the number of dependent variables (so far we have had $n = 1$), ϕ_i are functions which are analytic in ϵ and $u_i^{(\mu)}$, where $\mu = (\mu_1,...,\mu_m)$, is a multi-index denoting $\partial^{|\mu|}u_i/\partial x_1^{\mu_1}\cdots\partial x_m^{\mu_m}$, and $x_1,...,x_m$ are independent variables (we have had $x_1 = x$ up to now).

We want to show that there does not exist any finite inversion of the map (5.1). The word "finite" emphasizes the difference with the formal inversion of (5.1) in the formal power series in ϵ—such an inversion always exists. Of course, we can take advantage of the simplicity of the case $n = 1$. Then by the classification theorem (Theorem 4 in Ref. 9) of jet symmetries, (5.1) must be a contact transformation of the corresponding 1-jet manifold, and if there are any x derivatives of u involved in ϕ_1 of (5.1) this is impossible. The argument can be made more explicit in the case in which we are primarily interested: $m = 1$. Then if, say,

$$f: u \to \phi(x,u,...,u^{(k)}), \quad f^{-1}: u \to \bar{\phi}(x,u,...,u^{(l)}),$$
$$k > 0, \quad l \geqslant 0$$

then

$$0 = \frac{\partial u}{\partial u^{(k+l)}} = \frac{\partial \phi}{\partial u^{(k)}}\frac{\partial \bar{\phi}}{\partial u^{(l)}} \neq 0,$$

a contradiction.

On the other hand, for $n > 1$, e.g., if we want to treat the general scalar Lax equations (1.2) with

$$L = \xi^{n+1} + \sum_{i=1}^{n} u_i \xi^{i-1}, \tag{5.2}$$

the above argument cannot be applied because in the nonscalar case $n > 1$ there *are* invertible differential operators, like, say,

$$f: u \to u, \quad v \to v + \epsilon\phi(x,u,u_x,...). \tag{5.3}$$

To study such maps, we derive a simple sufficient criterion of nonivertibility. First

Definition 5.1: Suppose we have a map $\bar{f}: u \to y$ given by functions

$$y_j = y_j(x,u_i,...,u_i^{(\mu)}), \quad j = 1,...,N, \ i = 1,..., n.$$

The Frechet derivative of $\bar{f}$ is the following matrix differential operator $D(\bar{f})$:

$$D(\bar{f})_{ji} = \Sigma_\mu \frac{\partial y_j}{\partial u_i^{(\mu)}}\partial^\mu, \quad \partial^\mu := \partial_1^{\mu_1}\cdots\partial_1^{\mu_m},$$

$$\partial_k = \partial/\partial x_k.$$

Theorem 5.2: For a map f given by (5.1), let us denote by $l(f)$ its linearization with respect to u near $u = 0$. If $\det D(l(f))$ is a differential operator (of positive order), then the map (5.1) is not finitely invertible.

Proof: If f were invertible, then $l(f)$ would be also, and

$$l(f)l(f^{-1}) = \text{Id},$$

therefore

$$D(l(f))\circ D(l(f^{-1})) = \text{Id},$$

thus

$$\det D(l(f^{-1})) = [\det D(l(f))]^{-1},$$

a contradiction, since the rhs does not exist as a finite differential operator.

Remark 5.3: The above criterion allows one to easily analyze deformations for the case (5.2) (see Chap. III, Ref. 6).

Remark 5.4: In looking for a possible equivalence of equations $\sigma(0)$ and $\sigma(\epsilon)$, one might wish to allow also x and t to mix with $u_i^{(\mu)}$'s, and instead of requiring the full inversion of (5.1), to ask merely for a mapping which takes solutions of $\sigma(0)$ into those of $\sigma(\epsilon)$. But Vinogradov's theorem guarantees that nothing will change (Theorem 4, Ref. 10; Theorem 7.6, Ref. 11).

(B) Secondly, to prevent the appearance of fields of the sort $X_r + \epsilon X_{\bar{r}}$, we have to check that $\sigma(\epsilon) - \sigma(0)$ does not commute with $\sigma(0)$. To do this, we consider both $\sigma(0)$ and $\sigma(\epsilon)$ living on the same "jet bundle" (i.e., having the same coordinates) so that we can commute them. Thus, for example, we write $g(\epsilon)$ as

$$f: u \to u + \epsilon^2 u^2 + \epsilon u_x.$$

Let us write down the fact that the fields $\sigma(\epsilon)$ and $\sigma = \sigma(0)$ are f-related:

$$\sigma(\epsilon)f = f\sigma,$$

so

$$\sigma(\epsilon) = f\sigma f^{-1}.$$

Then we have to show that

$$[f\sigma f^{-1},\sigma] \neq 0. \tag{5.4}$$

To do this, note that in each of the three cases (1.5), (1.8), and (3.2) our map f has the form

$$f: u \to u + \epsilon u_x + O(\epsilon^2),$$

so

$$f^{-1}: u \to u - \epsilon u_x + O(\epsilon^2).$$

On the other hand, all evolution equations we have met have the form

$$u_t = \text{const} \times u^{(N)} + O(<N), \tag{5.5}$$

where $O(<N)$ stands for terms which depend upon $u^{(j)}$ with $j < N$. In particular, anything which commutes with σ in the form (5.5), must obviously have the same form as well. So it is enough to show that $(f\sigma f^{-1} - \sigma)$ has no terms *linear* in u and its x derivatives. Let us write then

$$\sigma = \sigma_1 + \sigma_2 + \cdots,$$

where σ_i is a homogeneous differential polynomial in u of degree i. Since $f\sigma_i f^{-1}$ has no components of degree less than i, and

$$f\sigma_1 f^{-1} \equiv \sigma_1 + O(\epsilon^2),$$

we get

$$\deg(f\sigma f^{-1}-\sigma)>1,$$

which is just what we needed.

(C) Finally, we do not want to have families of the form $X+\epsilon\bar{X}$, where X and $\bar{X}$ have the same conserved densities. Again, we shall use the existence of the reduction to derive a contradiction to the assumption that conserved densities are the same for σ and $\sigma(0)$.

We have three cases to consider: $\mathrm{MKdV}_r(\epsilon)$, $\mathrm{KdV}_r(\epsilon)$, and $\mathrm{KdV}_r(\epsilon,\nu)$. First, note that since $\mathrm{KdV}_r(0,\nu)=\mathrm{KdV}_r(\nu)$, it is enough to consider only MKdV_r and KdV_r fields. In both cases the description of all conserved densities is well known: for each positive integer n, one has conserved densities

$$H_n=(-1)^n\frac{u^{(n)2}}{2}+O(<n), \tag{5.6}$$

where $O(<n)$ are differential polynomials of degree *more* than 2. Since

$$f{:}u\to u+\epsilon u_x+(\text{terms of degree}\geqslant 2),$$

then conserved densities of $\sigma(\epsilon)$ via the pullback of H_n's of (5.6) have the form, modulo exact derivatives,

$$(-1)^n\frac{u^{(n)2}+\epsilon^2u^{(n+1)2}}{2}+(\text{terms of degree}\geqslant 2).$$

Thus, assuming the sameness of conserved densities, we must have

$$H_n(f(u))\approx H_n(u)-\epsilon^2H_{n+1}(u), \tag{5.7}$$

where $\approx$ means "equal modulo Im ∂".

Let us compute functional derivatives with respect to u of both sides of (5.7). For the left-hand side we have

$$\begin{aligned}
&\frac{\delta}{\delta u}[H_n(f(u))]\\
&=D(f)^+\frac{\delta H_n(\bar{u})}{\delta(\bar{u})}\Big|_{\bar{u}=f(u)}\\
&=[1-\epsilon\partial+O(\epsilon^2)][\bar{u}^{2n}+O(\leqslant 2n-2)]_{\bar{u}=\epsilon u_x+O(<1)}\\
&=[1-\epsilon\partial+O(\epsilon^2)][\epsilon u^{(2n+1)}+O(\leqslant 2n-1)]\\
&=-\epsilon^2u^{(2n+2)}+[1+O(\epsilon^2)]\epsilon u^{(2n+1)}+O(\leqslant 2n).
\end{aligned}$$

For the right-hand side we have

$$\begin{aligned}\frac{\delta H}{\delta u}-\epsilon^2\frac{\delta H_{n+1}}{\delta u}&=u^{(2n)}+O(\leqslant 2n-2)\\ &\quad-\epsilon^2[u^{(2n+2)}+O(\leqslant 2n)],\end{aligned}$$

a discrepancy in the $u^{(2n+1)}$ terms.

ACKNOWLEDGMENTS

The author appreciates valuable comments made by D. Burns, I. Dolgachev, Yu. I. Manin, and A. Greenspoon. The hospitality of F. Hirzebruch at the University of Bonn, where part of this paper was written, is gratefully acknowledged.

[1] I. M. Gel'fand and L. A. Dikii, "The asymptotics of the resolvent of Sturm–Liouville equations and the algebra of the Korteweg–de Vries equations," Usp. Math. Nauk **30**, 67–100 (1975) [Russ. Math. Surv. **30** (5), 77–113 (1975)].

[2] B. A. Kupershmidt, "On the nature of the Gardner transformation," J. Math. Phys. **22**, 449–451 (1981).

[3] R. M. Miura, C. S. Gardner, M. D. Kruskal, "Korteweg–de Vries equation and generalizations. II. Existence of conservation laws and constants of motion," J. Math. Phys. **9**, 1204–1209 (1968).

[4] Yu. I. Manin, "Algebraic aspects of non-linear differential equations," Itogi Nauki i Tekhniki Ser. Sovremennye Prob. Mat. **11**, 5–152 (1978) [J. Sov. Math. **11**, 1–122 (1979)].

[5] B. A. Kupershmidt and G. Wilson, "Modifying Lax equations and the second Hamiltonian structure," Invent. Math. **62**, 403–436 (1981).

[6] B. A. Kupershmidt, "Deformations of integrable systems," Proc. R. Irish Acad. (to appear) (1982).

[7] M. Adler and J. Moser, "On a class of polynomials connected with the Korteweg-de Vries equation," Commun. Math. Phys. **61**, 1–30 (1978).

[8] M. Wadati, H. Sanuki, and K. Konno, "Relationships among inverse method, Bäcklund transformations, and an infinite number of conservation laws," Progr. Theor. Phys. **53**, 419–436 (1975).

[9] A. M. Vinogradov, "Multivalued solutions and principle of classification of nonlinear differential equations," Dokl. Akad. Nauk SSSR **210**, 11–14 (1973) [Soviet Math. Dokl. **14**, 661–665 (1973)].

[10] A. M. Vinogradov, "The theory of higher symmetries of non-linear partial differential equations," Dokl. Akad. Nauk, SSSR **248**, 274–278 (1979) [Soviet Math. Dokl. **20**, 985–990 (1979).

[11] A. M. Vinogradov, "Geometry of nonlinear differential equations," in Problems in geometry **11**, 89–134 (1980).

Interior and exterior solutions for boundary value problems in composite elastic media. I. Two-dimensional problems

D. L. Jain
Department of Mathematics, University of Delhi, Delhi-110007, India

R. P. Kanwal
Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802

(Received 29 January 1981; accepted for publication 10 April 1981)

We study two-dimensional problems of elasticity when a homogeneous and isotropic solid of an arbitrary shape is embedded in an infinite homogeneous isotropic medium of different properties. Solutions are obtained both inside the guest and host media. These solutions are derived by first transforming the boundary value problems to the equivalent integral equations. The interior displacement field is obtained by a simple method of truncation. By this method the integral equations are recast into an infinite number of algebraic equations and a systematic scheme of solutions is constructed by an appropriate truncation. The exterior solutions are obtained by substituting the interior solutions in the integral equations valid for the entire medium. The boundaries considered are rectangular cylinder, equilateral triangular prism, and elliptic cylinder and its limiting configurations. It emerges that the solutions for the elliptic cylinder and its limiting configurations are exact.

PACS numbers: 03.40.Dz

1. INTRODUCTION

Chen and Young[1] have recently studied the displacement fields inside some three-dimensional elastic solids embedded in an infinite, homogeneous, and isotropic medium. They have treated the problem by an integral equation technique due to Waterman[2] and Eyges[3]. By this technique the integral equations governing these boundary value problems are transformed into an infinite system of algebraic equations, which are then suitably truncated to yield exact or approximate inner solutions.

Our aim is threefold. First, we extend the method to be applicable in deriving the interior displacement fields in two-dimensional linearly elastic inclusions of arbitrary shape embedded in an infinite homogeneous and isotropic medium. Secondly, we present a technique which helps us in obtaining the displacement fields in the host medium also, both in two- as well as three-dimensional problems. We use this scheme to present the exterior solutions for the configurations which have already been studied in Ref. 1. Finally, we present the interior and exterior solutions for those configurations such as a triaxial ellipsoid and elliptic cylinder of finite height for which the solutions are not available. Solutions for the two-dimensional problems are presented in this part while those for the three-dimensional case will appear in part II of this paper.

It emerges that the very first approximation yields the *exact solutions* for an infinite elliptic cylinder and its limiting configurations. All the known results in this field, such as those given in Refs. 4–7, follow as small corollaries. A few results appear to be new even for some simple configurations.

In addition to the inclusions we also discuss the cavities of arbitrary shapes. In their case we obtain the expressions for the strain energy stored in the host medium per unit height.

2. GOVERNING INTEGRAL EQUATIONS

Let (x, y, z) be a Cartesian coordinate system. An elastic homogeneous isotropic cylinder with axis along the z axis is embedded in an infinite homogeneous isotropic medium with Lamé's constants λ_1 and μ_1 and density ρ_1. The material of the cylinder has the elastic constants λ_2 and μ_2 and the density ρ_2. The section of the cylinder by the x, y plane occupies the region S_2, the exterior domain is S_1, while the bound-

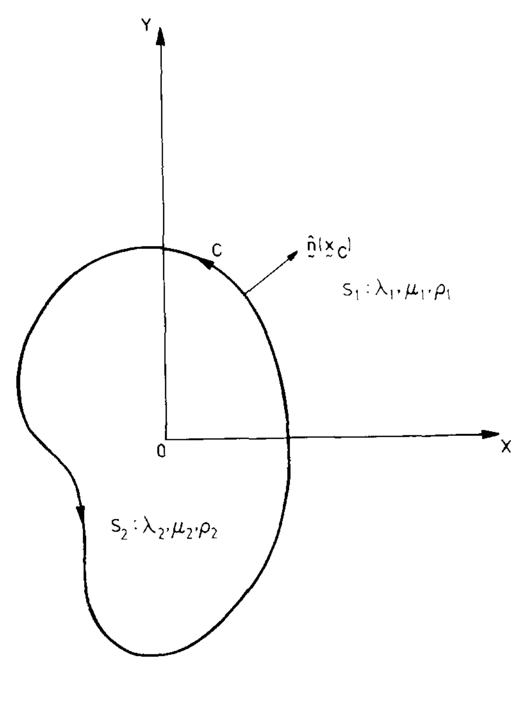


FIG. 1. Geometry of the section of the elastic cylinder by the $x - y$ plane.

 0022-2488/82/081433-11$02.50

ary is C. The origin of the coordinate system is situated at the centroid of S_2, which is assumed to be symmetrical with respect to the x or y axes. The configuration is explained in Fig. 1.

The constant stiffness tensors $C^{\alpha}_{ijkl}(\mathbf{x})$, $\mathbf{x}=(x,y)\in S_\alpha$, $\alpha=1,2$, are defined as

$$C^{\alpha}_{ijkl}(\mathbf{x}) = C^{\alpha}_{ijkl} = \lambda_\alpha \delta_{ij}\delta_{kl} + \mu_\alpha(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}), \quad \mathbf{x}\in S_\alpha, \tag{2.1}$$

where δ's are Kronecker deltas. In this analysis Latin indices have the range 1,2,3 and Greek indices have the range 1,2. Let $u^0(\mathbf{x})$ be the displacement field in the infinite host medium occupying the whole region $S=S_1+C+S_2$, due to prescribed stresses at infinity, so that it satisfies the equilibrium equations.

$$C^{1}_{i\alpha k\beta} u^0_{k,\beta\alpha}(\mathbf{x}) = 0, \quad \mathbf{x}=(x,y)\in S, \tag{2.2}$$

in the absence of body forces.

In this analysis we shall employ the distributional formula

$$\begin{aligned}\overline{\text{div}}\,(C_{i\alpha k\beta}(\mathbf{x})u_{k,\beta}(\mathbf{x})) &= \text{div}\,(C_{i\alpha k\beta}(\mathbf{x})u_{k,\beta}(\mathbf{x})) + [t_i]\delta(\mathbf{x}-\mathbf{x}_C)\\ &+ C_{k\alpha i\beta}(\mathbf{x})_c \hat{n}_\alpha(\mathbf{x}_C)[u_k]\frac{\bar{\partial}}{\partial x_\beta}\delta(\mathbf{x}-\mathbf{x}_C), \quad \mathbf{x}\in S,\end{aligned} \tag{2.3}$$

where the bar denotes the distributional derivative, $[F]=F(\mathbf{x}_C)|_- - F(\mathbf{x}_C)|_+$ is the jump of the function F from S_2 to S_1, the quantities $t_i = C_{i\alpha k\beta}u_{k,\beta}n_\alpha$ are the components of the traction vector, and $\hat{n}_\alpha$ are the components of the unit normal vector $\hat{\mathbf{n}}$ to the curve C. Also, the displacement field $u_k(\mathbf{x})$ is equal to $u^1_k(\mathbf{x})$ for $\mathbf{x}\in S_1$, $u^2_k(\mathbf{x})$ in S_2, and the same is true for the tensor $C_{i\alpha k\beta}$. The point $\mathbf{x}_C$ is on C and $\delta(\mathbf{x}-\mathbf{x}_C)$ is the Dirac delta function.

The first term in Eq. (2.3) vanishes both in S_1 and S_2. The other two terms vanish because the boundary conditions require that the displacements and tractions be continuous across the curve C. Thus the global equation

$$\overline{\text{div}}\,(C_{i\alpha k\beta}(\mathbf{x})u_{k,\beta}(\mathbf{x})) = 0, \quad \mathbf{x}\in S \tag{2.4}$$

incorporates all the conditions mentioned above. Putting

$$C_{i\alpha k\beta}(\mathbf{x}) = C^{1}_{i\alpha k\beta} + (\Delta C_{i\alpha k\beta})\theta(\mathbf{x}),$$

where

$$\Delta C_{i\alpha k\beta} = C^{2}_{i\alpha k\beta} - C^{1}_{i\alpha k\beta}$$

and $\theta(\mathbf{x})$ is the Heaviside function

$$\theta(\mathbf{x}) = \begin{cases} 0, & \mathbf{x}\in S_1,\\ \frac{1}{2}, & \mathbf{x}\in C,\\ 1, & \mathbf{x}\in S_2,\end{cases}$$

in (2.4), we obtain

$$\begin{aligned}\overline{\text{div}}\,(C^{1}_{i\alpha k\beta}u_{k,\beta}(\mathbf{x})) &= -\overline{\text{div}}\,(\Delta C_{i\alpha k\beta}u_{k,\beta}(\mathbf{x}))\theta(\mathbf{x})\\ &+ \Delta C_{i\alpha k\beta}u_{k,\beta}(\mathbf{x}_C)n_\alpha(\mathbf{x}_C)\delta(\mathbf{x}-\mathbf{x}_C), \quad \mathbf{x}\in S.\end{aligned} \tag{2.5}$$

In the above derivation we have also used the fact that

$$\overline{\text{grad}}\,\theta(\mathbf{x}) = -\hat{\mathbf{n}}(\mathbf{x}_C)\delta(\mathbf{x}-\mathbf{x}_C).$$

For the integral equation formulation of this boundary value problem, we define a Green's function $G_{km}(\mathbf{x},\mathbf{x}')$ when the whole space is filled with the elastic medium whose stiffness tensor is $C^{1}_{ijk\beta}$ and the tensor G_{km} satisfies the differential equation

$$\text{div}(C^{1}_{i\alpha k\beta}G_{km,l}(\mathbf{x},\mathbf{x}')) = -\delta_{im}\delta(\mathbf{x},\mathbf{x}'), \quad \mathbf{x},\mathbf{x}'\in S. \tag{2.6}$$

The values of these functions are

$$\begin{aligned}G_{\alpha\beta}(\mathbf{x},\mathbf{x}') &= G_{\beta\alpha}(\mathbf{x},\mathbf{x}')\\ &= -\frac{1}{8\pi}\left\{\frac{1}{\mu_1}\nabla^2\delta_{\alpha\beta} + \left(\frac{1}{\lambda_1+2\mu_1} - \frac{1}{\mu_1}\right)\right.\\ &\quad\left.\times\frac{\partial^2}{\partial\alpha\partial\beta}(R^2\ln R)\right\} \quad \alpha,\beta, = 1,2,\end{aligned} \tag{2.7}$$

$$G_{33}(\mathbf{x},\mathbf{x}') = -(1/2\pi\mu_1)(\ln R), \tag{2.8}$$

$$G_{\alpha 3}(\mathbf{x},\mathbf{x}') = G_{3\alpha}(\mathbf{x},\mathbf{x}') = 0, \tag{2.9}$$

where $R=|\mathbf{x},\mathbf{x}'|$.

Multiplying (2.5) by $G_{im}(\mathbf{x})$ and (2.6) by $u_i(\mathbf{x})$, subtracting and integrating over S, and using Green's theorem we obtain

$$\begin{aligned}u_m(\mathbf{x}') = u^0_m(\mathbf{x}') &- \int_{S_2}\text{div}(\Delta C_{i\alpha k\beta}u_{k,\beta}(\mathbf{x}))G_{im}(\mathbf{x},\mathbf{x}')\,dS_2\\ &+ \Delta C_{i\alpha k\beta}\oint_C u_{k,\beta}(\mathbf{x}_C)|_- G_{im}(\mathbf{x}_C,\mathbf{x}')n_\alpha(\mathbf{x}_C)dl_C,\end{aligned} \tag{2.10}$$

where dl_C is the arc length along C. Next, we use the relation

$$\begin{aligned}\int_{S_2}\text{div}(\Delta C_{i\alpha k\beta}u_{k,\beta}(\mathbf{x}))G_{im}(\mathbf{x},\mathbf{x}')\,dS_2 &= \Delta C_{i\alpha k\beta}\oint_C u_{k,\beta}(\mathbf{x}_C)|_- G_{im}(\mathbf{x}_C,\mathbf{x}')\hat{n}_\alpha(\mathbf{x}_C)dl_C\\ &- \Delta C_{i\alpha k\beta}\int_{S_2}G_{im,\alpha}(\mathbf{x},\mathbf{x}')u_{k,\beta}(\mathbf{x})\,dS_2\end{aligned} \tag{2.11}$$

in (2.10), interchange $\mathbf{x}$ and $\mathbf{x}'$, and get

$$u_m(\mathbf{x}) = u^0_m(\mathbf{x}) + \Delta C_{i\alpha k\beta}\int_{S_2}G_{im,\alpha}(\mathbf{x},\mathbf{x}')u_{k,\beta'}(\mathbf{x}')\,dS'_2, \quad \mathbf{x}\in S \tag{2.12}$$

where $u_{k,\beta'}(\mathbf{x}') = \partial u_k(\mathbf{x}')/\partial x'_\beta$.

In order to solve the integral equation (2.12) we decouple it and make some changes in the indices so that they take the forms

$$u_\alpha(\mathbf{x}) = u^0_\alpha(\mathbf{x}) + \Delta C_{\beta\gamma\nu\delta}\int_{S_2}G_{\alpha\beta,\gamma}(\mathbf{x},\mathbf{x}')u_{\nu,\delta'}(\mathbf{x}')\,dS'_2, \quad \alpha=1,2, \quad \mathbf{x}\in S, \tag{2.13}$$

$$u_3(\mathbf{x}) = u^0_3(\mathbf{x}) + \Delta\mu\int_{S_2}G_{33,\gamma}(\mathbf{x},\mathbf{x}')u_{3,\gamma'}(\mathbf{x}')\,dS'_2, \quad \mathbf{x}\in S, \tag{2.14}$$

where $\Delta\mu = \mu_2 - \mu_1$. Setting

$$u_i(\mathbf{x}) = u^0_i(\mathbf{x}) + u^s_i(\mathbf{x}), \tag{2.15}$$

we obtain the integral equations for the disturbed field $\mathbf{u}^s(\mathbf{x})$ as

$$u^s_\alpha(\mathbf{x}) = \Delta C_{\beta\gamma\nu\delta}\int_{S_2}G_{\alpha\beta,\gamma}(\mathbf{x},\mathbf{x}')u_{\nu,\delta'}(\mathbf{x}')\,dS'_2, \quad \alpha=1,2, \tag{2.16}$$

$$u_3^s(\mathbf{x}) = \Delta\mu \int_{S_2} G_{33,\gamma}(\mathbf{x},\mathbf{x}') u_{3,\gamma'}(\mathbf{x}')\, dS_2'. \tag{2.17}$$

Equation (2.14) embodies the antiplane strain problems in the two-dimensional elastostatic composite media and has been solved by us in Ref. 8, where we have solved various kinds of the potential problems in composite media such as electrostatics. We shall process here Eq. (2.13), which embodies the two-dimensional plane strain problems.

3. INTERIOR SOLUTION OF THE INTEGRAL EQUATION (2.13)

To present the systematic approximations we differentiate Eq. (2.13) n times and obtain

$$\begin{aligned} &u_{\alpha,p_1\cdots p_n}(\mathbf{x}) - u^0_{\alpha,p_1\cdots p_n}(\mathbf{x}) \\ &= \Delta C_{\beta\gamma\nu\delta} \int_{S_2} G_{\alpha\beta,\gamma p_1\cdots p_n}(\mathbf{x},\mathbf{x}') u_{\nu,\delta'}(\mathbf{x}')\, dS_2' \\ &= (-1)^{n+1} \Delta C_{\beta\gamma\nu\delta} \int_{S_2} G_{\alpha'\beta',\gamma p_1'\cdots p_n'}(\mathbf{x},\mathbf{x}') u_{\nu,\delta'}(\mathbf{x}')\, dS_2' \\ &\qquad \mathbf{x}\in S_2, \end{aligned} \tag{3.1}$$

where the p's have values 1,2. Now, we expand the quantities $u_{\nu,\delta'}(\mathbf{x}')$ in a Taylor series about the origin $\mathbf{0}$ when $\mathbf{x}'\in S_2$ so that

$$\begin{aligned} u_{\nu,\delta'}(\mathbf{x}') &= \sum_{s=0}^{\infty} \frac{1}{s!} \{u_{\nu,\delta' q_1'\cdots q_s'}(\mathbf{0})\} x'_{q_1} \cdots x'_{q_s} \\ &= \sum_{s=0}^{\infty} \frac{1}{s!} \{u_{\nu,\delta q_1\cdots q_s}(\mathbf{x})\}_{\mathbf{x}=0} x'_{q_1} \cdots x'_{q_s}, \end{aligned} \tag{3.2}$$

where the q's also have the values 1,2. Substituting this value in (3.1) and setting $\mathbf{x} = \mathbf{0}$ in both sides, we have

$$\begin{aligned} &u_{\alpha,p_1\cdots p_n}(\mathbf{0}) - u^0_{\alpha,p_1\cdots p_n}(\mathbf{0}) \\ &= (-1)^{n+1} \Delta C_{\beta\gamma\nu\delta} \sum_{s=0}^{\infty} \frac{1}{s!} T_{\alpha\beta,\gamma p_1\cdots p_n q_1\cdots q_s} u_{\nu,\delta q_1\cdots q_s}(\mathbf{0}), \end{aligned} \tag{3.3}$$

where

$$\begin{aligned} &T_{\alpha\beta,\gamma p_1\cdots p_n q_1\cdots q_s} \\ &= \int_{S_2} G_{\alpha'\beta',\gamma p_1'\cdots p_n'}(\mathbf{0},\mathbf{x}') x'_{q_1} \cdots x'_{q_s}\, dS_2' \\ &= \int_{S_2} G_{\alpha\beta,\gamma p_1\cdots p_n}(\mathbf{x},\mathbf{0}) x_{q_1} \cdots x_{q_s}\, dS_2, \end{aligned} \tag{3.4}$$

which depends only on the elastic constants of the host medium and the geometry of the inclusion.

The inner solution is derived by truncating the system of equations (3.3) to a finite number of equations involving an equal number of unknowns. Then the coefficients $u_{\alpha,p_1\cdots p_n}(\mathbf{0})$ can be solved in terms of the known coefficients $u^0_{\alpha,p_1\cdots p_n}(\mathbf{0})$ in the Taylor expansion of $u^0_\alpha(\mathbf{x})$. It emerges fortunately that the lowest order truncation gives a close approximation in most of the cases and *exact solutions* for the elliptical and circular cylinders. Accordingly, for $n = 0$, $u_\alpha(\mathbf{0}) = u^0_\alpha(\mathbf{0})$, while for $n = 1$, $s = 0$, we have

$$u_{\alpha,p}(\mathbf{0}) - u^0_{\alpha,p}(\mathbf{0}) = \Delta C_{\beta\gamma\nu\delta} T_{\alpha\beta,\gamma p} u_{\nu,\delta}(\mathbf{0}), \tag{3.5}$$

where

$$\begin{aligned} T_{\alpha\beta,\gamma p} &= \int_{S_2} G_{\alpha\beta,\gamma p}(\mathbf{x},\mathbf{0})\, dS_2 \\ &= \{\mu_1^{-1} \delta_{\alpha\beta} t_{\gamma pkk} + (M_1^{-1} - \mu_1^{-1}) t_{\alpha\beta\gamma p}\}, \\ M_1 &= \lambda_1 + 2\mu_1, \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} t_{\alpha\beta\gamma p} &= -\frac{1}{8\pi} \int_{S_2} \frac{\partial^4}{\partial x_\alpha \partial x_\beta \partial x_\gamma \partial x_p} [r^2 \ln r]\, dS_2, \\ r &= |\mathbf{x}|. \end{aligned} \tag{3.7}$$

The quantities $t_{\alpha\beta\gamma p}$ are the shape factors. They depend on the shape of the inclusion and are totally symmetric and dimensionless.

When we substitute the value $\Delta C_{\beta\gamma\nu\delta} = C^2_{\beta\gamma\nu\delta} - C^1_{\beta\gamma\nu\delta}$ from (2.1) in (3.5) we have

$$\begin{aligned} &u_{\alpha,p}(\mathbf{0}) - u^0_{\alpha,p}(\mathbf{0}) \\ &= \Delta\lambda T_{\alpha\gamma,\gamma p} u_{\nu,\nu}(\mathbf{0}) \\ &\quad + \Delta\mu (T_{\alpha\beta,\gamma p} u_{\beta,\gamma}(\mathbf{0}) + T_{\alpha\beta,\gamma p} u_{\gamma,\beta}(\mathbf{0})), \end{aligned} \tag{3.8}$$

where $\Delta\lambda = \lambda_2 - \lambda_1$.

In the next stage we decompose $u_{\alpha,p}$ into the symmetric and antisymmetric parts $u_{\alpha p}$ and $a_{\alpha p}$, respectively,

$$\begin{aligned} u_{\alpha p}(\mathbf{x}) &= \tfrac{1}{2}(u_{\alpha,p}(\mathbf{x}) + u_{p,\alpha}(\mathbf{x})), \\ a_{\alpha p}(\mathbf{x}), &\tfrac{1}{2}(u_{\alpha,p}(\mathbf{x}) - u_{p,\alpha}(\mathbf{x})), \end{aligned} \tag{3.9}$$

and define

$$T^{\pm}_{\alpha\beta\gamma p} = \tfrac{1}{2}(T_{\alpha\beta,\gamma p} \pm T_{p\beta,\gamma\alpha}). \tag{3.10}$$

Then relation (3.5) gives rise to two relations, namely,

$$u_{\alpha p}(\mathbf{0}) - u^0_{\alpha p}(\mathbf{0}) = \Delta C_{\beta\gamma\nu\delta} T^{+}_{\alpha\beta,\gamma p} u_{\nu\delta}(\mathbf{0}), \tag{3.11}$$

$$a_{\alpha p}(\mathbf{0}) - a^0_{\alpha p}(\mathbf{0}) = \Delta C_{\beta\gamma\nu\delta} T^{-}_{\alpha\beta,\gamma p} u_{\nu\delta}(\mathbf{0}). \tag{3.12}$$

Equation (3.11) is equivalent to

$$\begin{aligned} &u_{\alpha p}(\mathbf{0}) - (\Delta\lambda T^{+}_{\alpha\beta,\beta p} u_{\nu\nu}(\mathbf{0}) + 2\Delta u T^{+}_{\alpha\beta,\gamma p u_{\beta\gamma}}(\mathbf{0})) \\ &= u^0_{\alpha p}(\mathbf{0}). \end{aligned} \tag{3.13}$$

Setting $p = \alpha$ and summing on α, this equation reduces to

$$\begin{aligned} &u_{\alpha,\alpha}(\mathbf{0})[1 + (\Delta\lambda / M_1)] \\ &- (2\Delta\mu / M_1)[t_{11\beta\beta} u_{11}(\mathbf{0}) + t_{22\beta\beta}(\mathbf{0}) u_{22}(\mathbf{0})] = u^0_{\beta\beta}(\mathbf{0}), \end{aligned} \tag{3.14}$$

where we have used relations (3.6) and (3.10). Similarly,

$$\begin{aligned} u_{\epsilon\epsilon}(\mathbf{0}) &= \frac{1}{2} \left(\alpha_\epsilon + \frac{\beta_\epsilon \delta_\epsilon}{\gamma_\epsilon}\right)^{-1} \\ &\times \left\{ \left[1 + \frac{\alpha_\epsilon + (1 + \Delta\lambda / 2\Delta\mu)\delta_\epsilon}{\gamma_\epsilon}\right] u^0_{11}(\mathbf{0}) \right. \\ &\left. + \left[\frac{\alpha_\epsilon + (1 + \Delta\lambda / 2\Delta\mu)\delta_\epsilon}{\gamma_\epsilon} - 1\right] u^0_{22}(\mathbf{0}) \right\}, \quad \epsilon = 1,2, \end{aligned} \tag{3.15}$$

$$u_{12}(\mathbf{0}) = [1 + \Delta\mu\{\mu_1^{-1} - 4(M_1^{-1} - \mu_1^{-1}) t_{1122}\}]^{-1} u^0_{12}(\mathbf{0}), \tag{3.16}$$

and

$$\begin{aligned} \alpha_\epsilon &= 1 - 2\Delta\mu [t_{\epsilon\epsilon\alpha\alpha} / M_1 - 2(M_1^{-1} - \mu_1^{-1}) t_{\epsilon\epsilon\eta\eta}], \\ \beta_\epsilon &= -(\Delta\lambda / M_1) t_{\epsilon\epsilon\alpha\alpha} - 2\Delta\mu [M_1^{-1} - \mu_1^{-1}] t_{\epsilon\epsilon\eta\eta}, \\ \gamma_\epsilon &= 1 + \Delta\lambda / M_1 - (2\Delta\mu / M_1) t_{\eta\eta\alpha\alpha}, \\ \delta_\epsilon &= (2\Delta\mu / M_1)(t_{\epsilon\epsilon\alpha\alpha} - t_{\eta\eta\alpha\alpha}), \end{aligned} \tag{3.17}$$

$\epsilon \neq \eta$, ϵ, $\eta = 1,2$, and ϵ and η are not summed.

As regards the antisymmetric quantity $a_{\alpha p}(\mathbf{0})$ we observe that $a_{11}(0) = a_{22}(0) = 0$ and, therefore, $a_{\alpha\alpha}(0) = 0$. Also $a^0_{11}(0) = a^0_{22}(0) = 0$. This leaves only $a_{12}(0)$ and $a_{21}(0)$ to be determined. Indeed, by substituting the values of $\Delta C_{\beta\gamma\lambda\delta}$ in (3.12) and using relations (3.6) and (3.10) we find that

$$a_{12}(0) = -a_{12}(0) = a^0_{12}(0). \tag{3.18}$$

4. VALUES OF SHAPE FACTORS FOR VARIOUS CYLINDERS

(i). Rectangular cylinder. Let the rectangular cross section have the edges of length $2a$ and $2b$ parallel to the coordinate axes with its center at the origin. Their equations are $x = \pm b, y = \pm a$. Then there are the following nonvanishing shape factors:

$$t_{1111} = -\frac{1}{\pi}\left\{2 \arctan\frac{b}{a} - \frac{ab}{a^2+b^2}\right\}, \tag{4.1a}$$

$$t_{2222} = -\frac{1}{\pi}\left\{2 \arctan\frac{a}{b} - \frac{ab}{a^2+b^2}\right\}, \tag{4.1b}$$

$$t_{1122} = t_{2211} = -\frac{1}{\pi}\left[\frac{ab}{a^2+b^2}\right], \tag{4.1c}$$

$$t_{11\alpha\alpha} = -\frac{2}{\pi}\arctan\frac{b}{a}; \quad t_{22\alpha\alpha} = -\frac{2}{\pi}\arctan\frac{a}{b}. \tag{4.1d}$$

For a *square cylinder*, $b \to a$ so that the above values reduce to

$$t_{1111} = t_{2222} = \{(1/2\pi) - \tfrac{1}{2}\}, \tag{4.2a}$$

$$t_{1122} = t_{2211} = -1/2\pi, \tag{4.2b}$$

$$t_{11\alpha\alpha} = t_{22\alpha\alpha} = -\tfrac{1}{2}. \tag{4.2c}$$

(ii). Prism. Let the prism have the equilateral triangle whose cross section is described by the equations

$$y = \pm(1/\sqrt{3})(x + 2a/3); \quad x = a/\sqrt{3},$$

where $2a$ is the length of each side of the triangular cross section and the centroid of this section lies at the origin. Then the nonvanishing shape factors are

$$t_{1111} = t_{2222} = -\tfrac{3}{8}, \tag{4.3a}$$

$$t_{1122} = t_{2211} = -\tfrac{1}{8}, \tag{4.3b}$$

$$t_{11\alpha\alpha} = t_{22\alpha\alpha} = -\tfrac{1}{2}. \tag{4.3c}$$

(iii). Elliptic cylinder. It is convenient to introduce the elliptic coordinates (ξ,η) such that

$$x = c\cosh\xi\cos\eta, \quad y = c\sinh\xi\sin\eta,$$

where $2c$ is the focal length. The semimajor axis a and the minor axis b are $a = c\cosh\xi_0$, $b = c\sinh\xi_0$, and the bounding curve C of the elliptic cross section (Rather the elliptic section is given by $x^2/a^2 + y^2/b^2 \leqslant 1, \xi < \xi_0$) $x^2/a^2 + y^2/b^2 = 1$ is given by $\xi = \xi_0$, that is, $x = c\cosh\xi_0\cos\eta, y = c\sinh\xi_0\sin\eta$, where $(\pm c,0)$ are the coordinates of the two foci. In this case the nonvanishing shape factors are

$$t_{1111} = -\tfrac{1}{2}e^{-\xi_0}\sinh\xi_0(2 - \cosh\xi_0 e^{-\xi_0}), \tag{4.4a}$$

$$t_{2222} = -\tfrac{1}{2}e^{-\xi_0}\cosh\xi_0(2 - \sinh\xi_0 e^{-\xi_0}), \tag{4.4b}$$

$$t_{1122} = t_{2211} = -\tfrac{1}{4}e^{-2\xi_0}\sinh 2\xi_0, \tag{4.4c}$$

$$t_{11\alpha\alpha} = -e^{-\xi_0}\sinh\xi_0; \quad t_{22\alpha\alpha} = -e^{-\xi_0}\cosh\xi_0. \tag{4.4d}$$

For a circular cylinder these values reduce to

$$t_{1111} = t_{2222} = -\tfrac{3}{8}, \tag{4.5a}$$

$$t_{1122} = t_{2211} = -\tfrac{1}{8}, \tag{4.5b}$$

$$t_{11\alpha\alpha} = t_{22\alpha\alpha} = -\tfrac{1}{2}, \tag{4.5c}$$

which are the same as for the equilateral prism.

5. EVALUATION OF THE INTERIOR DISPLACEMENT FIELD

(i). Square cylinder, equilateral triangular prism, circular cylinder. When we substitute the values of the shape factors from the previous section into formulas (3.15)–(3.18) we find that the values of $u_{\alpha p}(0)$ and $a_{\alpha p}(0)$ are the same for a prism, and a circular cylinder to this approximation. These values for a square cylinder, an equilateral triangular prism, and a circular cylinder are

$$u_{11}(0) = \tfrac{1}{2}[A_1^{-1} + B^{-1}]u^0_{11}0 + \tfrac{1}{2}[B^{-1} - A_1^{-1}]u^0_{22}(0), \tag{5.1}$$

$$u_{22}(0) = \tfrac{1}{2}[B^{-1} - A_1^{-1}]u^0_{11}(0) + \tfrac{1}{2}[A_1^{-1} + B^{-1}]u^0_{22}(0), \tag{5.2}$$

$$u_{12}(0) = u_{21}(0) = A_2^{-1}u^0_{12}(0); \quad a_{12}(0) = -a_{21}(0) = a^0_{12}(0). \tag{5.3}$$

where

$$A_i = [1 + C_i(\Delta\mu/\mu_1)], \quad i = 1,2, \tag{5.4}$$

$$B = \left(1 + \frac{\Delta\mu + 3\Delta K}{3K_1 + 4\mu_1}\right), \quad C_i = \frac{3a_iK_1 + 4b_i\mu_1}{3K_1 + 4\mu_1}, \tag{5.5}$$

while $\Delta K = \Delta\lambda + \tfrac{2}{3}\Delta\mu$ and $K_1 = \lambda_1 + \tfrac{2}{3}\mu_1$. The values of the quantities a's and b's for a *square cylinder* are

$$a_1 = 2/\pi, \quad b_1 = 1/2\pi + \tfrac{3}{4},$$
$$a_2 = 1 - 2/\pi, \quad b_2 = 1 - 1/2\pi, \tag{5.6}$$

and for the equilateral triangular prism or a circular cylinder they are

$$a_1 = a_2 = \tfrac{1}{2}, \quad b_1 = b_2 = \tfrac{7}{8}. \tag{5.7}$$

From this analysis it follows that the *inner displacement field* in all these three cases is given as

$$u_1(\mathbf{x}) = u^0_1(0) + (u_{11}(0))x + (u_{12}(0) + a_{12}(0))y,$$

$$u_2(\mathbf{x}) = u^0_2(0) + (u_{12}(0) - a_{12}(0))x + u_{22}(0)y, \tag{5.8}$$

for $\mathbf{x} \in S_2$, where we have used relations (3.9) in the Taylor expansions of $u_1(\mathbf{x})$ and $u_2(\mathbf{x})$, and $u_{11}(0)$, $u_{12}(0)$, $u_{22}(0)$, and $a_{12}(0)$ are given by (5.1)–(5.7) in terms of the values of $u^0_{11}(0)$, $u^0_{12}(0)$, $u^0_{22}(0)$, and $a^0_{12}(0)$, which can be easily derived from the known displacement field $u^0_\alpha(\mathbf{x})$.

(ii). Rectangular cylinder. Now we substitute the values of the shape factor for the rectangular cylinder from the previous section into formulas (3.17) and obtain

$$\alpha_1 = 1 - 2\Delta\mu\left[-\frac{2}{\pi}\left(M_1^{-1}\arctan\frac{b}{a}\right) + \frac{2}{\pi}(M_1^{-1} - \mu_1^{-1})\frac{ab}{a^2+b^2}\right], \tag{5.9a}$$

$$\beta_1 = \frac{2\Delta\mu}{\pi}\left(M_1^{-1}\arctan\frac{b}{a}\right)$$

$$+\frac{2\Delta\mu}{\pi}(M_1^{-1}-\mu_1^{-1})\frac{ab}{a^2+b^2}, \quad (5.9b)$$

$$\gamma_1 = 1 + M_1^{-1}\Delta\lambda + (4/\pi)\Delta\mu M_1^{-1}\arctan(a/b), \quad (5.9c)$$

$$\delta_1 = \frac{4}{\pi}M_1^{-1}\Delta\mu\left(\arctan\frac{a}{b} - \arctan\frac{b}{a}\right), \quad (5.9d)$$

and the values of $\alpha_2, \beta_2, \gamma_2$, and δ_2 are derived from these formulas by interchanging a and b.

When we substitute these values in formulas (3.15) we get the values of the quantities $u_{11}(0)$ and $u_{22}(0)$ as

$$\begin{pmatrix} u_{11}(0) \\ u_{22}(0) \end{pmatrix} = \frac{1}{2D}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} u_{11}^0(0) \\ u_{22}^0(0) \end{pmatrix}, \quad (5.10)$$

where

$$A_{11} = 2 + M_1^{-1}(\Delta\lambda + 2\Delta\mu) - \frac{4}{\pi}\Delta\mu(M_1^{-1} - \mu_1^{-1}) \times \frac{ab}{a^2+b^2} + \frac{2}{\pi}M_1^{-1}(2\Delta\mu + \Delta\lambda) \times\left(\arctan\frac{a}{b} - \arctan\frac{b}{a}\right),$$

$$A_{12} = -\Delta\lambda M_1^{-1}\left[1 + \frac{2}{\pi}\left(\arctan\frac{b}{a} - \arctan\frac{a}{b}\right)\right] - \frac{4}{\pi}\Delta\mu(M_1^{-1}-\mu_1^{-1})\frac{ab}{a^2+b^2},$$

$$D = [1 + M_1^{-1}(\Delta\lambda + \Delta\mu)] \times\left\{1 - \frac{4}{\pi}\Delta\mu(M_1^{-1}-\mu_1^{-1})\frac{ab}{a^2+b^2}\right\} + \frac{16}{\pi^2}M_1^{-2}\Delta\mu(\Delta\mu + \Delta\lambda)\arctan\frac{a}{b}\arctan\frac{b}{a} + \frac{\Delta\mu}{M_1},$$

while the expressions for A_{22} and A_{21} follow by interchanging a and b in the above formulas for A_{11} and A_{12}, respectively.

Similarly, the value of $u_{12}(0)$ is

$$u_{12}(0) = \left[1 + \Delta\mu\left\{\mu_1^{-1} + \frac{4}{\pi}(M_1^{-1}-\mu_1^{-1})\frac{ab}{a^2+b^2}\right\}\right]^{-1} \times u_{12}^0(0), \quad (5.11a)$$

while

$$a_{12}(0) = a_{12}^0(0) + \frac{2\Delta\mu}{\pi\mu_1}\left(\arctan\frac{b}{a} - \arctan\frac{a}{b}\right)u_{12}(0). \quad (5.11b)$$

The inner displacement field for the rectangular cylinder is also given by Eq. (5.8) except that now the values of $u_{11}(0)$, $u_{12}(0)$, $u_{22}(0)$, and $a_{12}(0)$ are given by relations (5.10) and (5.11). When $b \to a$, these results reduce to those for the square cylinder, as derived above.

(iii). Elliptic cylinder. In this case the values of the coefficients $\alpha_1, \beta_1, \gamma_1$, and δ_1 are

$$\alpha_1 = 1 + 2(e^{-\xi_0}\sinh\xi_0)\Delta\mu[1/M_1 - (1/M_1 - 1/\mu_1)e^{-\xi_0}\cosh\xi_0], \quad (5.12a)$$

$$\beta_1 = e^{-\xi_0}\sinh\xi_0[\Delta\lambda/M_1 + e^{-\xi_0}\Delta\mu(1/M_1 - 1/\mu_1) \times\cosh\xi_0], \quad (5.12b)$$

$$\gamma_1 = 1 + \frac{\Delta\lambda}{M_1} + 2e^{-\xi_0}\frac{\Delta\mu}{M_1}\cosh\xi_0; \quad \delta_1 = \frac{2\Delta\mu}{M_1}e^{-2\xi_0}. \quad (5.12c)$$

The values of α_2, β_2, and γ_2 are obtained from the above formulas by interchanging $\sinh\xi_0$ and $\cosh\xi_0$, while $\delta_2 = -\delta_1$.

Next, we substitute these values in formulas (3.15) and get values of $u_{11}(0)$ and $u_{22}(0)$ from formulas (5.10), where the values of the A 's and D now are

$$A_{11} = 2 + 2\Delta\lambda M_1^{-1}e^{-\xi_0}\cosh\xi_0 + \Delta\mu\mu_1^{-1}e^{-2\xi_0}\sinh 2\xi_0 + 2\Delta\mu M_1^{-1}e^{-2\xi_0}\cosh\xi_0(2\cosh\xi_0 + \sinh\xi_0), \quad (5.13a)$$

$$A_{12} = 2\Delta\mu M_1^{-1}e^{-\xi_0}\cosh\xi_0 + \mu_1^{-1}\Delta\mu e^{-2\xi_0}\sinh 2\xi_0 - 2\Delta\lambda M_1^{-1}e^{-\xi_0}\sinh\xi_0 - 2\Delta\mu M_1^{-1}e^{-2\xi_0} \times\cosh\xi_0(\cosh\xi_0 + 2\sinh\xi_0), \quad (5.13b)$$

$$D = \{1 + (\Delta\lambda + \Delta\mu)M_1^{-1}\} + \{1 - \Delta\mu(M_1^{-1} - \mu_1^{-1})e^{-2\xi_0}\sinh 2\xi_0\} + \Delta\mu M_1^{-1}\{1 + 2M_1^{-1}(\Delta\lambda + \Delta\mu)e^{-2\xi_0}\sinh 2\xi_0\}, \quad (5.13c)$$

and the values of the quantities A_{21} and A_{22} are derived from the above formulas for A_{12} and A_{11}, respectively, by interchanging $\sinh\xi_0$ and $\cosh\xi_0$. The coefficient $u_{12}(0)$ has the value

$$u_{12}(0) = [1 + \Delta\mu\{\mu_1^{-1} + (M_1^{-1} - \mu_1^{-1})e^{-2\xi_0}\sinh 2\xi_0\}]^{-1} \times u_{12}^0(0). \quad (5.14)$$

Furthermore,

$$a_{12}(0) = a_{12}^0(0) - (\Delta\mu/\mu_1)e^{-2\xi_0}u_{12}(0). \quad (5.15)$$

Finally, by substituting the above values of $u_{11}(0)$, $u_{12}(0)$, $u_{22}(0)$, and $a_{12}(0)$ in formulas (5.8), we obtain the required displacement field inside the elastic elliptic cylinder. When $\xi_0 \to \infty$ we recover the corresponding values for the circular cylinder, as derived earlier.

(iv). Infinite strip. For an infinite strip $-c \leqslant x \leqslant c, y = 0$, $-\infty < z < \infty$, we can derive the corresponding formulas by taking the limit $\xi_0 \to 0$ in (5.12)–(5.15). This yields

$$u_{11}(0) = u_{11}^0(0), \quad (5.16a)$$

$$u_{22}(0) = \left[1 + \left(\frac{\Delta\lambda + 2\Delta\mu}{M_1}\right)\right]^{-1}\left\{-\frac{\Delta\lambda}{M_1}u_{11}^0(0) + u_{22}^0(0)\right\}, \quad (5.16b)$$

$$u_{12}(0) = [1 + \Delta\mu/\mu_1]^{-1}u_{12}^0(0), \quad (5.16c)$$

$$a_{12}(0) = a_{12}^0(0) - (\Delta\mu/\mu_1)u_{12}(0). \quad (5.16d)$$

Substituting relations (5.16a)–(5.16d) in (5.8), we obtain the required inner field inside the strip.

We shall now prove that the above inner displacement fields as derived by the first approximation for the elliptic cylinder and its limiting configurations of the circular cylinder and the infinite strip are *exact solutions* of the governing integral equations (2.13). It is assumed that the infinite host medium is subjected to a constant prescribed stress field in the absence of the body forces. Accordingly, the known displacement field $u_\alpha^0(\mathbf{x})$ is linear in x and y. For this purpose, it suffices to establish that the integrals

$$I_{\alpha\beta,\gamma}(\mathbf{x}) = \int_{S_2} G_{\alpha\beta,\gamma}(\mathbf{x},\mathbf{x}')\,dS_2', \quad \mathbf{x}\in S_2,$$

are linear in x and y when the section S_2 is an ellipse because, using this result, it follows from Eq. (2.13) that the inner solution $u_\alpha(\mathbf{x})$, $\mathbf{x}\in S_2$ is also linear in x and y for an elliptic cylinder. This is true since in this case the quantities $u_{\vartheta,\delta'}(\mathbf{x}')$ are constants and therefore all the terms on both sides of Eq. (2.13) become linear in x and y. Accordingly, the above approximate inner solution for an elliptic cylinder by the first truncation of Eq. (3.3) is the exact inner solution. Let us therefore prove that $I_{\alpha\beta,\gamma}(\mathbf{x})$, $\mathbf{x}\in S_2$ is linear in x and y. To establish this we shall prove that for an elliptic cylinder the following relation holds:

$$J_{\alpha\beta,\gamma pq}(\mathbf{x}) = \frac{\partial^2}{\partial x_p \partial x_q} I_{\alpha\beta,\gamma}(\mathbf{x}) = \int_{S_2} G_{\alpha\beta,\gamma pq}(\mathbf{x},\mathbf{x}')\, dS_2' = 0, \quad p,q = 1,2 \quad \mathbf{x}\in S_2, \tag{5.17}$$

It amounts to proving that for an elliptic cylinder

$$\left\{\frac{\partial^n}{\partial x_{\alpha_1}\partial x_{\alpha_2}\cdots\partial x_{\alpha_n}}(J_{\alpha\beta,\gamma pq}(\mathbf{x}))\right\}_{x=0} = \left\{\int_{S_2} G_{\alpha\beta,\gamma pq\alpha_1\alpha_2\cdots\alpha_n}(\mathbf{x},\mathbf{x}')dS_2'\right\}_{\mathbf{x}=0} = 0,$$
$$p,q,\alpha_1,\ldots\alpha_n = 1,2, \quad n = 0,1,2,3,\ldots. \tag{5.18}$$

To prove this we first observe that

$$\left(\int_{S_2} G_{\alpha\beta,\gamma pq\alpha_1\alpha_2\cdots\alpha_n}(\mathbf{x},\mathbf{x}')dS_2'\right)_{(\mathbf{x})=(\mathbf{0})} = (-1)^{n+1}\int_S G_{\alpha'\beta',\gamma' p' q'\alpha_1'\alpha_2'\cdots\alpha_n'}(\mathbf{0},\mathbf{x}')\, dS_2' = (-1)^{n+1}\int_{S_2} G_{\alpha\beta,\gamma pq\alpha_1\cdots\alpha_n}(\mathbf{x},\mathbf{0})\, dS_2. \tag{5.19}$$

Also, from relation (2.7) it follows that $G_{\alpha\beta}(\mathbf{x},\mathbf{0})$ involves second order partial derivatives of $|\mathbf{x}|^2 \ln|\mathbf{x}|$. Therefore, in order to establish (5.18) with the help of (5.19) we have to prove that for an elliptic cylinder

$$\int_{S_2} \frac{\partial^N}{\partial\alpha_1\cdots\partial\alpha_N}(r^2\ln r)dS_2 = 0, \quad \alpha_1\ldots\alpha_N = 1,2, \quad N\geqslant 5. \tag{5.20}$$

Due to the symmetry of the elliptic section S_2, (5.20) is obviously satisfied for all odd values of $N\geqslant 5$. Therefore, it remains to prove that

$$\int_{S_2} \frac{\partial^{2L}}{\partial\alpha_1\cdots\partial\alpha_{2L}}(r^2\ln r)dS_2 = 0 \quad \alpha_1,\ldots,\alpha_{2L} = 1,2, \; L\geqslant 3. \tag{5.21}$$

Now for the elliptic section $S_2{:}x^2/a^2 + y^2/b^2 \leqslant 1$, relation (5.21) reduces to

$$\int_{x^2/a^2+y^2/b^2\leqslant l} \frac{\partial^m}{\partial x_1^m}\left(\frac{\partial}{\partial x_2}\right)^{2L-m}(r^2\ln r)dS_2 = 0, \quad 0\leqslant m\leqslant 2L$$

which is again true for odd values of m because of the symmetry of the region of integration. It therefore remains to prove that

$$\int_{x^2/a^2+y^2/b^2\leqslant l} \left\{\frac{\partial^{2n}}{\partial x_1^{2n}}\left(\frac{\partial}{\partial x_2}\right)^{2(L-n)}(r^2\ln r)\right\}dS_2 = 0, \quad 0\leqslant n\leqslant L \text{ for } L\geqslant 3. \tag{5.22}$$

Now the integrand in (5.22) is of the form $(1/r^{2(L-1)})$ $\times\{a_n\cos 2L\phi + b_n\cos(2L-1)\phi\}$, where the coefficients a_n and b_n depend on the values of the integer n while $dS_2 = r\,dr\,d\phi$. Accordingly, this equation becomes

$$\lim_{\epsilon\to 0}\int_{\phi=0}^{2\pi}\left[\{a_n\cos 2L\phi + b_n\cos 2(L-1)\phi\} + \int_{r=\epsilon}^{(\cos^2\phi/a^2+\sin^2\phi/b^2)^{1/2}}\frac{dr}{r^{2L-3}}\right]d\phi = 0 \quad \text{if } L\geqslant 3, \tag{5.23}$$

where we have excluded an infinitesimal region lying inside a circle of radius ϵ with center (0) form the domain of integration and have used the polar form of the ellipse, namely, $r = (\cos^2\phi/a^2 + \sin^2\phi/b^2)^{-1/2}$. The value of the inner integral in (5.23) is

$$\frac{1}{(2L-4)}\left\{\frac{1}{\epsilon^{2L-4}} - \left(\frac{\cos^2\phi}{a^2} + \frac{\sin^2\phi}{b^2}\right)^{L-2}\right\},$$

which can be expressed in terms of the known coefficients c_m as

$$\frac{1}{(2L-4)}\left\{\frac{1}{\epsilon^{(2L-4)}} - \sum_{m=0}^{L-2} c_m\cos 2m\phi\right\},$$

which when substituted in (5.23) makes it an identity in view of the orthogonal properties of the trigonometric cosine functions, and we have completed the proof that *the very first approximation for the elliptic cylinder is an exact solution.*

6. HIGHER ORDER APPROXIMATIONS

Let us now display the power of the method by presenting the estimate of the accuracy of the first lowest order approximation for those configurations for which it does not yield an exact solution. Let us, for example, take the case of the square cylinder. For simplicity we consider a purely plane static strain

$$u^0_{\alpha,\beta}(\mathbf{x}) = \mathrm{U}\delta_{\alpha\beta}, \quad \mathbf{x}\in S \tag{6.1}$$

applied to the host medium before inserting this cylinder. Then

$$u^0_{1,1}(\mathbf{x}) = u^0_{2,2}(\mathbf{x}) = \mathrm{U}, \; u^0_{1,2}(\mathbf{x}) = u^0_{2,1}(\mathbf{x}) = 0, \tag{6.2}$$

so that

$$u^0_1(\mathbf{x}) = \mathrm{U}x, \quad u^0_2(\mathbf{x}) = \mathrm{U}y. \tag{6.3}$$

By symmetry, there are the following three independent nonzero coefficients in the Taylor expansion of the displacement field $u_\alpha(\mathbf{x})$ inside the square cylinder,

$$A = u_{1,1}(\mathbf{0}) = u_{2,2}(\mathbf{0}), \tag{6.4a}$$
$$B = u_{1,111}(\mathbf{0}) = u_{2,222}(\mathbf{0}), \tag{6.4b}$$
$$C = u_{1,222}(\mathbf{0}) = u_{2,111}(\mathbf{0}). \tag{6.4c}$$

From Eq. (3.3) if follows that for the second approximation we take $n = 1$, $s = 2$ so that

$$u_{\alpha,p}(\mathbf{0}) - u^0_{\alpha,p}(\mathbf{0}) = C_{\beta\gamma\nu\delta}\left[T_{\alpha\beta,\gamma p}u_{\nu,\delta}(\mathbf{0}) + (1/2!)T_{\alpha\beta,\gamma p,q_1q_2}u_{\nu,\delta q_1q_2}(\mathbf{0})\right]. \tag{6.5}$$

Identifying p with α because of the symmetry considerations and using (6.2), the above relation reduces to

$$[1 - (\Delta\lambda + \Delta\mu)T_{\alpha\beta,\beta\alpha}]A - (\Delta\mu/2)T_{\alpha\beta,\beta\alpha,\beta\beta}B - (\Delta\mu/2)[2T_{\alpha\beta,\gamma\alpha,\beta\gamma} - 3T_{\alpha\beta,\beta\alpha,\beta\beta}]C = \mathrm{U}, \tag{6.6}$$

where an index appearing four times is to be successively set equal to 1,2 and then summed.

Similarly, for $n = 3$, $s = 2$, relation (3.3) yields

$$u_{\alpha,p_1p_2p_3}(0) = \Delta C_{\beta\gamma\nu\delta}\left[T_{\alpha\beta,\gamma p_1p_2p_3}\mu_{\nu,\delta}(0) + \tfrac{1}{2}T_{\alpha\beta,\gamma p_1p_2p_3,q_1q_2}u_{\nu,\delta q_1q_2}(0)\right]. \tag{6.7}$$

Identifying p_1 with α in (6.7) and p_3 with p_2, we have

$$u_{\alpha,\alpha pp} = \Delta C_{\beta\gamma\nu\delta}\left[T_{\alpha\beta,\gamma\alpha pp}u_{\nu,\delta}(0) + \tfrac{1}{2}T_{\alpha\beta,\gamma\alpha pp,q_1q_2}u_{\nu,\delta q_1q_2}(0)\right],$$

which reduces to

$$B + C = 0. \tag{6.8}$$

To find the third relation for A,B, and C, we set $\alpha = p_1 = p_2 = p_3$ in (6.7) and obtain

$$C\left[2 + \Delta\mu\{T_{\alpha\beta,\beta\alpha\alpha\alpha,\gamma\gamma} + 2T_{\alpha\beta,\gamma\alpha\alpha\alpha,\beta\gamma} - 4T_{\alpha\beta,\beta\alpha\alpha\alpha,\beta\beta}\}\right] = -2A(\Delta\lambda + \Delta\mu)T_{\alpha\beta,\beta\alpha\alpha\alpha}. \tag{6.9}$$

To derive the values of A,B, and C we evaluate the shape factors of orders six and eight occurring in Eq. (6.6), (6.8), and (6.9). For the square cylinder the required values are

$$T_{\alpha\beta,\beta\alpha} = -M_1^{-1}, \quad T_{\alpha\beta,\beta\alpha,\beta\beta} = 2M_1^{-1}(2/\pi - \tfrac{1}{2}) = 0.28M_1^{-1},$$

$$T_{\alpha\beta,\beta\alpha\alpha\alpha,\gamma\gamma} = (12/\pi)M_1^{-1} = 3.6M_1^{-1}; \quad T_{\alpha\beta,\beta\alpha\alpha\alpha} = -(2/\pi)M_1^{-1} = -0.64M_1^{-1},$$

$$T_{\alpha\beta,\beta\alpha\alpha\alpha,\beta\beta} = 6\mu^{-1} + (2/\pi - 4)M_1^{-1} = 6\mu_1^{-1} - 3.36M_1^{-1},$$

$$T_{\alpha\beta,\gamma\alpha\alpha\alpha,\beta\gamma} = \mu_1^{-1}(3 - 6/\pi) + (M_1^{-1} - \mu_1^{-1})(-3 + 3/\pi) = 3.12\mu_1^{-1} - 2.04M_1^{-1},$$

Substituting these values in Eqs. (6.6), (6.8), and (6.9), which we solve for A,B, and C, and get

$$A = \left\{[1 + M_1^{-1}(\Delta\lambda + \Delta\mu)] - \frac{(0.1536)\Delta\mu(\Delta\lambda + \Delta\mu)M_1^{-2}}{[1 + \Delta\mu\{6.48M_1^{-1} - 8.88\mu_1^{-1}\}]}\right\}^{-1} U, \tag{6.10a}$$

$$C = -B = 0.64(\Delta\lambda + \Delta\mu)M_1^{-1} \times\{1 + \Delta\mu[6.48M_1^{-1} - 8.88\mu_1^{-1}]\}^{-1}A. \tag{6.10b}$$

Now, when $u_1^0(x) = Ux$, $u_2^0(\mathbf{x}) = Uy$, we have from the first approximate inner solution for a square cylinder as given by Eq. (5.1)–(5.6),

$$A = u_{1,1}(0) = u_{11}(0) = \left(1 + \frac{\Delta\mu + 3\Delta K}{3K_1 + 4\mu_1}\right)U = \left(1 + \frac{\Delta\mu + \Delta\lambda}{M_1}\right)U = u_{2,2}(0), \tag{6.11a}$$

$$u_{1,2}(0) = 0 = u_{2,1}(0). \tag{6.11b}$$

Note that we do not get any information about the values of the coefficients B and C. When $(\Delta\lambda)M_1^{-1}$, $(\Delta\mu)M_1^{-1}$, and $(\Delta\mu)\mu_1^{-1}$ are very small, the value of A in (6.11) obtained from the first approximation is almost equal to that obtained by the second approximation (6.10a).

7. CYLINDRICAL CAVITY AND STRAIN ENERGY

When the inclusion is a cylindrical cavity of an arbitrary shape we have

$$\lambda_2 = \mu_2 = 0; \quad \Delta\lambda = -\lambda_1, \Delta\mu = -\mu_1;$$
$$\Delta C_{\beta\gamma\nu\delta} = -C^1_{\beta\gamma\vartheta\delta}.$$

Let us assume that the whole infinite medium is subjected to the constant stress field $\tau^0_{\alpha\beta}(\mathbf{x})$, where

$$\tau^0_{11}(\mathbf{x}) = \tau^0_{22}(\mathbf{x}) = T; \quad \tau^0_{12}(\mathbf{x}) = 0 \quad \mathbf{x} \in S. \tag{7.1}$$

Hence,

$$u_1^0(\mathbf{x}) = \frac{Tx}{2(\lambda_1 + \mu_1)}; \quad u_2^0(\mathbf{x}) = \frac{Ty}{2(\lambda_1 + \mu_1)}, \quad \mathbf{x} \in S, \tag{7.2a}$$

$$u_1^0(0) = u_2^0(0) = 0;$$

$$u^0_{1,1}(0) = u^0_{2,2}(0) = \frac{T}{2(\lambda_1 + \mu_1)} = u^0_{11}(0) = u^0_{22}(0), \tag{7.2b}$$

$$a^0_{12}(0) = 0, \ u_{1,2}(0) = u^0_{2,1}(0) = u^0_{21}(0) = u^0_{12}(0) = (0). \tag{7.2c}$$

It follows from Eqs. (3.15) and (3.16) that the first approximation yields the inner displacement field solution

$$u_1(\mathbf{x}) = u_{1,1}(0)x, \quad u_2(\mathbf{x}) = (u_{2,2}(0))y, \ \mathbf{x} \in S_2, \tag{7.3a}$$

where

$$u_{\epsilon,\epsilon}(0) = u_{\epsilon\epsilon}(0) = \frac{T}{2}\left[\frac{1}{(\lambda_1 + \mu_1)(\alpha_\epsilon\gamma_\epsilon + \beta_\epsilon\delta_\epsilon)}\right] \times\left[\alpha_\epsilon + \left(1 + \frac{\lambda_1}{2\mu_1}\right)\delta_\epsilon\right]. \tag{7.3b}$$

$$\alpha_\epsilon = 1 + 2\mu_1\left[\frac{t_{\epsilon\epsilon\alpha\alpha}}{M_1} - 2(M_1^{-1} - \mu_1^{-1})t_{\epsilon\epsilon\eta\eta}\right], \tag{7.3c}$$

$$\beta_\epsilon = \frac{\lambda_1}{M_1}t_{\epsilon\epsilon\alpha\alpha} + 2(M_1^{-1} - \mu_1^{-1})\,t_{\epsilon\epsilon\eta\eta}, \tag{7.3d}$$

$$\gamma_\epsilon = 1 - \frac{\lambda_1}{M_1} + \frac{2\mu_1}{M_1}t_{\eta\eta\alpha\alpha}, \tag{7.3e}$$

$$\delta_\epsilon = \frac{-2\mu_1}{M_1}(t_{\eta\eta\alpha\alpha} - t_{\epsilon\epsilon\alpha\alpha}), \tag{7.3f}$$

while $\epsilon \neq \eta$ and ϵ,η are not summed. Consequently, the inner stress field vanishes, i.e.,

$$\tau_{11}(\mathbf{x}) = \tau_{22}(\mathbf{x}) = \tau_{12}(\mathbf{x}) = 0, \ x \in S_2, \tag{7.4}$$

as expected. Using boundary conditions across the curve C, we get

$$u_1(\mathbf{x}_C) = u_1(\mathbf{x}_C)|_+ = (u_{1,1}(0))x_C;$$
$$u_2(\mathbf{x}_C) = u_2(\mathbf{x}_C)|_+ = (u_{2,2}(0))y_C, \tag{7.5a}$$
$$\tau_{n\alpha}(\mathbf{x}_C) = \tau_{\alpha\beta}(\mathbf{x}_C)|_+\hat{n}_\beta(\mathbf{x}_C) = 0, \tag{7.5b}$$

where $\mathbf{x}_C = (x_C, y_C)$ and $u_{1,1}(0)$ and $u_{2,2}(0)$ are given by (7.3).

As in (2.15), the stress field $\tau_{\alpha\beta}(\mathbf{x})$ can be written as

$$\tau_{\alpha\beta}(\mathbf{x}) = \tau^0_{\alpha\beta}(\mathbf{x}) + \tau^s_{\alpha\beta}\mathbf{x}, \quad \mathbf{x} \in S, \tag{7.6}$$

where $\tau^s_{\alpha\beta}(\mathbf{x})$ represents the disturbance in the constant applied stress field $\tau^0_{\alpha\beta}(\mathbf{x})$ due to the presence of the cylindrical cavity in the host medium. Substituting relations (2.15) and (7.6) in (7.5) we have

$$u_1^s(\mathbf{x}_C) = (u_{1,1}(0))\mathbf{x}_C - \frac{Tx_C}{2(\lambda_1 + \mu_1)};$$

$$u_2^s(\mathbf{x}_C) = (u_{2,2}(\mathbf{0}))y_C - \frac{Ty_C}{2(\lambda_1 + y_C)}; \tag{7.7a}$$

$$\tau_{n\alpha}^s(\mathbf{x}_C) = \tau_{\alpha\beta}^s(\mathbf{x}_C)|_+ \hat{n}_\beta(\mathbf{x}) = -\tau_{\alpha\beta}^0(\mathbf{x}_C)|_+ \hat{n}_\beta(\mathbf{x}_C), \tag{7.7b}$$

$$\tau_{nl}^s(\mathbf{x}_C) = -T\hat{n}_1(\mathbf{x}_C), \quad \tau_{n2}^s(\mathbf{x}_C) = -T\hat{n}_2(\mathbf{x}_C), \tag{7.7c}$$

$$\tau_{nn}^s(\mathbf{x}_C) = \tau_{n\alpha}^s\, \hat{n}_\alpha(\mathbf{x}_C) = -T, \tag{7.7d}$$

where we have used the results (7.1) and (7.2).

The elastic energy E stored in the host medium per unit height due to the presence of the cylindrical cavity is given as

$$E = -\tfrac{1}{2}\oint_C (u_\alpha^s(\mathbf{x}_C)\hat{n}_\alpha(\mathbf{x}_C)\tau_{nn}^s(\mathbf{x}_C)dl_C + \tfrac{1}{2}\oint_{C_\infty} (u_\alpha^s(\mathbf{x})\hat{n}_\alpha(\mathbf{x}_{C_\infty}))\tau_{nn}^s(\mathbf{x}_{C_\infty})dl_{C_\infty}, \tag{7.8}$$

where the second integral vanishes when we appeal to the known far-field behavior of $u_\alpha^s(\mathbf{x})$ and $\tau_{\alpha\beta}^s(\mathbf{x})$. When we substitute the values of $u_\alpha^s(\mathbf{x}_c)$, $\tau_{nn}^s(\mathbf{x}_c)$ from (7.7a) and (7.7d) in (7.8), we get the required formula for E,

$$E = \frac{T}{2}\oint_C \left[\left\{u_{1,1}(\mathbf{0}) - \frac{T}{2(\lambda_1 + \mu_1)}\right\}x_C\hat{n}_1(\mathbf{x}_C) + \left\{u_{2,2}(\mathbf{0}) - \frac{T}{2(\lambda_1 + \mu_1)}\right\}y_C\hat{n}_2(\mathbf{x}_C)\right]dl_C, \tag{7.9}$$

where $u_{1,1}(\mathbf{0})$, $u_{2,2}(\mathbf{0})$ are given in (7.3) in terms of the shape factors of the inclusion. Since the inner solution was derived by using the first approximation, we find that formula (7.9) for E is exact for elliptic cylindrical and circular cylindrical cavities as well as for an infinite crack.

We now derive the expressions for the strain energy E per unit height for different configurations.

(i). Elliptic Cylindrical Cavity. The ellipse C is given by

$$x_C = c\cosh\xi_0\cos\eta, \quad y_C = c\sinh\xi_0\sin\eta,$$

$$\hat{n}_1(\mathbf{x}_C) = \sinh\xi_0\cos\eta/h_0, \quad \hat{n}_2(\mathbf{x}_C) = \cosh\xi_0\sin\eta/h_0,$$

$$h_0 = (\cosh^2\xi_0 - \cos^2\eta)^{1/2}, \quad dl_C = h_0 c\, d\eta.$$

Substituting these values in (7.9) we get

$$E = \frac{T}{2}c^2\sinh\xi_0\cosh\xi_0\int_0^{2\pi}\left[(u_{1,1}(\mathbf{0}))\cos^2\eta + (u_{2,2}(\mathbf{0}))\sin^2\eta - \frac{T}{2(\lambda_1 + \mu_1)}\right]d\eta,$$
$$= \frac{\pi T}{2}c^2\sinh\xi_0\cosh\xi_0\left[u_{1,1}(\mathbf{0}) + u_{2,2}(\mathbf{0}) - \frac{T}{(\lambda_1 + \mu_1)}\right], \tag{7.10}$$

where $u_{\epsilon,\epsilon}(\mathbf{0})$ are given by (7.3b). When we put the values of the shape factors for the elliptic cylinder from (4.4) in (7.3) we derive the values of $u_{\epsilon,\epsilon}(\mathbf{0})$, which when substituted in (7.10) yield

$$E = \frac{T\pi}{4}c^2\sinh 2\xi_0\left[\frac{TM_1\cosh 2\xi_0}{\mu_1(\lambda_1 + \mu_1)\sinh 2\xi_0} - \frac{T}{\lambda_1 + \mu_1}\right],$$
$$= \frac{T^2\pi}{4}\left[\frac{a^2 + b^2}{\mu_1} + \frac{(a-b)^2}{\lambda_1 + \mu_1}\right], \tag{7.11}$$

where a and b are the semiprincipal axes of the ellipse C.

When $b \to a$ in (7.11), we derive the corresponding result for a *circular cylindrical cavity* of radius a, namely,

$$E = \frac{T^2\pi}{2\mu_1}a^2 = \frac{T^2 A}{2\mu_1}, \tag{7.12}$$

where $A = \pi a^2$ is the area enclosed.

When $b \to 0$, in (7.11) we obtain the corresponding result for the *infinite crack* of width $2a$,

$$|x| \leqslant a, \quad y = 0, \quad -\infty < z < \infty.$$

This value is

$$E = \frac{T^2\pi a^2}{4}\left\{\frac{1}{\mu_1} + \frac{1}{\lambda_1 + \mu_1}\right\} = \frac{T^2\pi a^2 M_1}{4\mu_1(\lambda_1 + \mu_1)}. \tag{7.13}$$

Even this limiting formula appears to be new.

(ii). Square Cylindrical Cavity. In this case, C consists of four sides of the square of length $2a$ with center at 0, that is,

$$|x| \leqslant a, \ y = \pm a; \quad |y| \leqslant a, \ x = \pm a.$$

When we substitute the values of the shape factors from Eq. (4.2) in (7.3), we obtain

$$u_{1,1}(\mathbf{0}) = u_{2,2}(\mathbf{0}) = TM_1/2\mu_1(\lambda_1 + \mu_1),$$

so that formula (7.9) yields, in this case,

$$E = \frac{4T}{2}\left[\frac{TM_1}{2\mu_1(\lambda_1 + \mu_1)} - \frac{T}{2(\lambda_1 + \mu_1)}\right]\int_{-a}^{a} a\, dy,$$

or

$$E = T^2 A/2\mu_1, \tag{7.14}$$

where $A = 4a^2$ is the area enclosed by C. This formula agrees with (7.12) for the circular cylindrical cavity.

8. EXTERIOR SOLUTIONS

To derive the displacement field in the host medium we again appeal to the governing integral equation (2.13), valid in the entire plane S. In this equation we now substitute the value of the displacement field in the guest medium, which is valid up to the surface of the cylinder. To understand the method let us first discuss the case of the circular cylinder.

(i). Circular Cylinder. Let us assume that the prescribed stress field is such that we have the uniform tension T in the direction of the x axis before the cylinder is inserted in it. Accordingly, the displacement components are

$$u_1^0(\mathbf{x}) = \frac{T(\lambda_1 + 2\mu_1)x}{4\mu_1(\lambda_1 + \mu_1)}, \quad u_2^0(\mathbf{x}) = \frac{-\lambda_1 Ty}{4\mu_1(\lambda_1 + \mu_1)}. \tag{8.1}$$

In terms of the polar coordinates $x = r\cos\vartheta$, $y = r\sin\vartheta$, this assumption yields the following displacement and stress fields:

$$u_r^0(\mathbf{x}) = (Tr/2E_1)(1 + \sigma_1)\{(1 - 2\sigma_1) + \cos 2\vartheta\}, \tag{8.2a}$$

$$u_\vartheta^0(\mathbf{x}) = -(Tr/2E_1)(1 + \sigma_1)\sin 2\vartheta, \tag{8.2b}$$

$$\tau_{rr}^0(\mathbf{x}) = (T/2)(1 + \cos 2\vartheta), \tag{8.3a}$$

$$\tau_{r\vartheta}^0(\mathbf{x}) = -(T/2)\sin 2\vartheta, \tag{8.3b}$$

$$\tau_{\vartheta\vartheta}^0(\mathbf{x}) = (T/2)(1 - \cos\vartheta), \tag{8.3c}$$

where E is Young's modulus and σ is Poisson's ratio,

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}; \quad \sigma = \frac{\lambda}{2(\lambda + \mu)},$$

and the subscript 1 stands for the host medium as before.

Exact Interior Solution in the Region $r<a$. Recall that relation (5.8) gives the exact value of the displacement field inside the circular cylinder for a general prescribed uniform stress field at infinity. For the present case of the uniform tension T in the direction of the x axis, it reduces to

$$u_1(\mathbf{x})=\frac{T(1-\sigma_1)}{2}\left\{\frac{1-2\sigma_2}{[(1-2\sigma_2)\mu_1+\mu_2]}+\frac{2}{[\mu_1+(3-4\sigma_1)\mu_2]}\right\}x, \tag{8.4a}$$

$$u_2(\mathbf{x})=\frac{T(1-\sigma_1)}{2}\left\{\frac{1-2\sigma_2}{[(1-2\sigma_2)\mu_1+\mu 2]}-\frac{2}{[\mu_1+(3-4\sigma_1)\mu_2]}\right\}y, \tag{8.4b}$$

$$u_r(\mathbf{x})=\frac{Tr(1-\sigma_1)}{2}\left\{\frac{1-2\sigma_2}{[(1-2\sigma_2)\mu_1+\mu_2]}+\frac{2\cos 2\vartheta}{[\mu_1+(3-4\sigma_1)\mu_2]}\right\}, \tag{8.4c}$$

$$u_\vartheta(\mathbf{x})=-\frac{T(1-\sigma_1)r\sin 2\vartheta}{[\mu_1+(3-4\sigma_1)\mu_2]}, \tag{8.4d}$$

for $r<a$.

From these values we find that the interior stress field is

$$\tau_{rr}(\mathbf{x})=\mu_2T(1-\sigma_1)\left\{\frac{1}{(1-2\sigma_2)\mu_1+\mu_2}+\frac{2\cos 2\vartheta}{\mu_1+(3-4\sigma_1)\mu_2}\right\}, \tag{8.5a}$$

$$\tau_{r\vartheta}(\mathbf{x})=-\mu_2T(1-\sigma_1)\left\{\frac{2\sin 2\vartheta}{\mu_1+(3-4\sigma_1)\mu_2}\right\}, \tag{8.5b}$$

$$T_{\vartheta\vartheta}(\mathbf{x})=\mu_2T(1-\sigma_1)\left\{\frac{1}{[(1-2\sigma_1)\mu_1+\mu_2]}-\frac{2\cos 2\vartheta}{[\mu_1+(3-4\sigma_1)\mu_2]}\right\}. \tag{8.5c}$$

As far as we are aware, the exact solutions (8.4) and (8.5), even for such a simple boundary and such a simple prescribed field, are given here for the first time.

To derive the displacement field in the region $r>a$ we appeal to the integral equation (2.16), namely,

$$u_\alpha^s(\mathbf{x})=\Delta C_{\beta\gamma\nu\delta}\int_{S_2}G_{\alpha\beta,\gamma}(\mathbf{x},\mathbf{x}')u_{\nu,\delta'}(\mathbf{x}')dS_2', \quad \alpha=1,2. \tag{8.6}$$

When we substitute the values of constants $u_{\nu,\delta'}(\acute{\mathbf{x}})$ as calculated from the linear interior solution (8.4) and use (2.1) to write down $\Delta C_{\beta\gamma\nu\delta}$ explicitly, we obtain

$$\begin{aligned}u_1^s(\mathbf{x})=(\lambda_2-\lambda_1)&\frac{[T(1-\sigma_1)(1-2\sigma_2)]}{[(1-2\sigma_2)\mu_1+\mu_2]}\int_{S_2}G_{1\gamma,\gamma}(\mathbf{x},\mathbf{x}')dS_2'\\&+2(\mu_2-\mu_1)\int_{S_2}[G_{11,1}(\mathbf{x},\mathbf{x}')u_{1,1}(\mathbf{x}')\\&+G_{12,2}(\mathbf{x},\mathbf{x}')u_{2,2}(\mathbf{x}')]dS_2', \quad r>a.\end{aligned} \tag{8.7}$$

Next, we substitute the value of the Green's functions (2.7) in (8.7). The values of R^2 and $\ln R$ in (2.7) for the present case are

$$R^2=r^2+r'^2-2rr'\cos(\vartheta-\vartheta')$$

and

$$\ln R=\ln r-\sum_{n=1}^{\infty}\left(\frac{r'}{r}\right)^n\frac{\cos n(\vartheta-\vartheta')}{n}, \quad r>r'.$$

After carrying out the integration in (8.7) and simplifying, we obtain that formula for $u_1^s(\mathbf{x})$. Similarly, we derive the value of $u_2^s(\mathbf{x})$. They finally yield

$$u_r^s(\mathbf{x})=\frac{A}{r}+\left[-\frac{B}{r^3}+\frac{2C}{r}(1-\sigma_1)\right]\cos 2\vartheta, \quad r>a, \tag{8.8a}$$

$$u_\vartheta^s(\mathbf{x})=-\left[\frac{B}{r^3}+\frac{C}{r}(1-2\sigma_1)\right]\sin 2\vartheta, \quad r>a, \tag{8.8b}$$

$$\tau_{rr}^s(\mathbf{x})=2\mu_1\left[-\frac{A}{r^2}+\left(\frac{3B}{r^4}-\frac{2C}{r^2}\right)\cos 2\vartheta\right], \quad r>a, \tag{8.9a}$$

$$\tau_{r\vartheta}^s(\mathbf{x})=2\mu_1\left[\frac{3B}{r^4}-\frac{C}{r^2}\right]\sin 2\vartheta, \quad r>a, \tag{8.9b}$$

$$\tau_{\vartheta\vartheta}^s(\mathbf{x})=2\mu_1\left[\frac{A}{r^2}-\frac{3B}{r^4}\cos 2\vartheta\right], \quad r>a, \tag{8.9c}$$

where

$$A=\frac{Ta^2}{4\mu_1}\frac{(1-2\sigma_2)\mu_1-(1-2\sigma_1)\mu_2}{(1-2\sigma_2)\mu_1+\mu_2},$$

$$B=\frac{Ta^4}{4\mu_1}\frac{(\mu_1-\mu_2)}{\mu_1+(3-4\sigma_1)\mu_2}; \quad C=\frac{2B}{a^2}.$$

These formulas agree with the ones given by Goodier[9] and by us.[10]

(ii). Elliptic Cylinder. Let us assume that the whole host medium is subjected to a uniform tension T applied in the direction making angle α with the x axis before the guest elliptic cylinder is inserted. Therefore,

$$u_1^0(\mathbf{x})=\frac{T[2(\lambda_1+\mu_1)\cos^2\alpha-\lambda_1]}{4\mu_1(\lambda_1+\mu_1)}x+\frac{T\sin 2\alpha}{4\mu_1}y, \tag{8.10a}$$

$$u_2^0(\mathbf{x})=\frac{T\sin 2\alpha}{4\mu_1}x+\frac{T[2(\lambda_1+\mu_1)\sin^2\alpha-\lambda_1]}{4\mu_1(\lambda_1+\mu_1)}y, \tag{8.10b}$$

$$\tau_{11}^0(\mathbf{x})=T\cos^2\alpha, \tau_{22}^0(\mathbf{x})=T\sin^2\alpha,$$

$$\tau_{12}^0(\mathbf{x})=\frac{T\sin 2\alpha}{2}, \tag{8.10c}$$

$$u_{11}^0(\mathbf{0})=\frac{T[2(\lambda_1+\mu_1)\cos^2\alpha-\lambda_1]}{4\mu_1(\lambda_1+\mu_1)}, \tag{8.10d}$$

$$u_{22}^0(\mathbf{0})=\frac{T[2(\lambda_1+\mu_1)\sin^2\alpha-\lambda_1]}{4\mu_1(\lambda_1+\mu_1)}, \tag{8.10e}$$

$$u_{12}^0(\mathbf{0})=u_{21}^0(\mathbf{0})=\frac{T\sin 2\alpha}{4\mu_1}, \quad a_{12}^0(\mathbf{0})=0. \tag{8.10f}$$

Accordingly, the exact linear interior solution in this case follows from Sec. 5 to be

$$\begin{aligned}&u_1(\mathbf{x})=u_{11}(\mathbf{0})x+(u_{12}(\mathbf{0})+a_{12}(\mathbf{0}))y,\\&u_2(\mathbf{x})=(u_{12}(\mathbf{0})-a_{12}(\mathbf{0}))x+u_{22}(\mathbf{0})y,\\&\mathbf{x}\in S_2,\end{aligned} \tag{8.11}$$

where $u_{\alpha\beta}(\mathbf{0})$, $a_{12}(\mathbf{0})$ are given by Eqs. (5.10), (5.13)–(5.15) in terms of known constants $u_{\alpha\beta}^0(\mathbf{0})$, $a_{12}^0(\mathbf{0})$ defined in (8.10d), (8.10e), and (8.10f).

We again use the elliptic coordinates $\mathbf{x} = (\xi,\eta)$ as introduced in Sec. 5 and process the integral equation (8.6), which we write as

$$u^s_\alpha(\mathbf{x}) = -\Delta\lambda\int_{S_2} G_{\alpha\beta,\beta'}(\mathbf{x},\mathbf{x}')u_{\delta,\delta'}(\mathbf{x}')dS'_2 - \Delta\mu\int_{S_2} G_{\alpha\beta,\gamma'}(\mathbf{x},\mathbf{x}')[u_{\beta,\gamma'}(\mathbf{x}') + u_{\gamma,\beta'}(\mathbf{x}')]dS'_2, \quad \xi\geqslant\xi_0, \tag{8.12}$$

where S_2 is the region $\xi < \xi_0$ and we have used the property $G_{\alpha\beta,\gamma}(\mathbf{x},\mathbf{x}') = -G_{\alpha\beta,\gamma'}(\mathbf{x},\mathbf{x}')$. When we use the inner solution (8.11) and apply Green's theorem, the above equation reduces to

$$\begin{aligned} u^s_\alpha(\mathbf{x}) = & -\Delta\lambda\bar{u}_{\delta\delta}(\mathbf{0})\oint_C [G_{a1}(\mathbf{x},\mathbf{x}')\hat{\mathbf{i}} + G_{\alpha 2}(\mathbf{x},\mathbf{x}')\hat{\mathbf{i}}]\cdot\hat{\mathbf{n}}(\mathbf{x}'_C)dl'_C \\ & -\Delta\mu\oint_C [\bar{u}_{\beta 1}(\mathbf{0})G_{\alpha\beta}(\mathbf{x},\mathbf{x}')\hat{\mathbf{i}} + \bar{u}_{\beta 2}(\mathbf{0})G_{\alpha\beta}(\mathbf{x},\mathbf{x}')\hat{\mathbf{j}}]\cdot\hat{\mathbf{n}}(\mathbf{x}'_C)dl'_C \\ & -\Delta\mu\oint_C [\bar{u}_{1\beta}(\mathbf{0})G_{\alpha\beta}(\mathbf{x},\mathbf{x}')\hat{\mathbf{i}} + \bar{u}_{2\beta}(\mathbf{0})G_{\alpha\beta}(\mathbf{x},\mathbf{x}')\hat{\mathbf{j}}]\cdot\hat{\mathbf{n}}(\mathbf{x}'_C)dl'_C \\ & \xi > \xi_0, \end{aligned} \tag{8.13}$$

where $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are the unit vectors along the x and y axes and dl'_C is the element of length along the elliptic contour C, whose equation is $\xi = \xi_0$, and constants $\bar{u}_{\alpha\beta}(\mathbf{0}) = u_{\alpha,\beta'}(\mathbf{x}') = u_{\alpha\beta}(\mathbf{0}) + a_{\alpha\beta}(\mathbf{0})$, $\mathbf{x}' \in S_2$.

We we substitute the values of Green's functions from Eq. (2.7) in (8.13), we readily obtain

$$\begin{aligned} u^s_1(\mathbf{x}) \\ = & \frac{c}{8\pi}(\Delta\lambda u_{\delta\delta}(\mathbf{0}) + 2\Delta\mu u_{11}(\mathbf{0}))\left[\frac{1}{\mu_1}\nabla^2 + \left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial x^2}\right]I_1(\mathbf{x}) + \frac{c}{8\pi}(\Delta\lambda u_{\delta\delta}(\mathbf{0}) + 2\Delta\mu u_{22}(\mathbf{0}))\left[\left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial x\partial y}\right]I_2(\mathbf{x}) \\ & + \frac{c}{4\pi}\Delta\mu u_{12}(\mathbf{0})\left\{\left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial x\partial y}(I_1(\mathbf{x})) + \left[\frac{1}{\mu_1}\nabla^2 + \left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial x^2}\right]I_2(x)\right\}, \quad \xi > \xi_0, \end{aligned} \tag{8.14a}$$

$$\begin{aligned} u^s_2(\mathbf{x}) = & \frac{c}{8\pi}(\Delta\lambda u_{\delta\delta}(\mathbf{0}) + 2\Delta\mu u_{11}(\mathbf{0}))\left[\left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial x\partial y}\right]I_1(\mathbf{x}) + \frac{c}{8\pi}(\Delta\lambda u_{\delta\delta}(\mathbf{0}) + 2\Delta\mu u_{22}(\mathbf{0})) \\ & \times\left[\frac{1}{\mu_1}\nabla^2 + \left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial y^2}\right]I_2(\mathbf{x}) + \frac{c}{4\pi}\Delta\mu u_{12}(\mathbf{0})\left\{\left(\frac{1}{\mu_1}\nabla^2 + \left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial y^2}\right)I_1(\mathbf{x}) + \left(\frac{1}{M_1} - \frac{1}{\mu_1}\right)\frac{\partial^2}{\partial x\partial y}I_2(\mathbf{x})\right\}, \\ & \xi > \xi_0, \end{aligned} \tag{8.14b}$$

where

$$\begin{aligned} \mathrm{I}_1(\mathbf{x}) = & \int_{\eta'=0}^{2\pi}(R^2\ln R)_{\xi'=\xi_0}\sinh\xi_0\cos\eta' d\eta' = \frac{c^2\pi}{2}\sinh\xi_0\{[\tfrac{1}{6}e^{-3\xi}\cosh 3\xi_0 - \tfrac{1}{2}e^{-\xi}\cosh\xi_0]\cos 3\eta \\ & - [e^{\xi}\cosh\xi_0 + \tfrac{1}{2}e^{-3\xi}\cosh\xi_0 + e^{-\xi}(3\cosh\xi_0 + \tfrac{1}{2}\cosh 3\xi_0)]\cos\eta - 4\cosh\xi\cosh\xi_0\cos\eta\ln(ce^{\xi}/2)\}, \quad \xi > \xi_0, \end{aligned} \tag{8.15a}$$

$$\begin{aligned} \mathrm{I}_2(\mathbf{x}) = & \int_{\eta'=0}^{2\pi}(R^2\ln R)_{\xi'=\xi_0}\cosh\xi_0\sin\eta' d\eta' = \tfrac{1}{2}c^2\pi\cosh\xi_0\{[\tfrac{1}{6}e^{-3\xi}\sinh 3\xi_0 - \tfrac{1}{2}e^{-\xi}\sinh\xi_0]\sin 3\eta \\ & - [e^{\xi}\sinh\xi_0 + \tfrac{1}{2}e^{-3\xi}\sinh\xi_0 + e^{-\xi}(\tfrac{1}{2}\sinh 3\xi_0 - 3\sinh\xi_0)]\sin\eta \\ & - 4\sinh\xi\sinh\xi_0\sin\eta\ln(\mathrm{c}e^{\xi}/2)\}, \quad \xi > \xi_0, \end{aligned} \tag{8.15b}$$

and we have used the relations

$$R^2 = |\mathbf{x} - \mathbf{x}'|^2 = (c^2/2)\{\cosh 2\xi + \cosh 2\xi' + \cos 2\eta + \cos 2\eta' - 4\cosh\xi\cosh\xi'\cos\eta\cos\eta' - 4\sinh\xi\sinh\xi'\sin\eta\sin\eta'\}$$

and

$$\ln R = \xi - \ln\left(\frac{2}{c}\right) - 2\sum_{n1}^{\infty}\frac{e^{-n\xi}}{n}[\cosh(n\xi')\cos n\eta\cos n\eta' + \sinh(n\xi')\sin n\,\eta\,\sin n\,\eta'], \quad \xi > \xi'.$$

To obtain the expressions for $u^s_1(\mathbf{x})$ and $u^s_2(\mathbf{x})$ from formulas (8.14), we require the differential operators

$$\frac{\partial}{\partial x} = \frac{1}{c^2h}\left(\sinh\xi\cos\eta\frac{\partial}{\partial\xi} - \cosh\xi\sin\eta\frac{\partial}{\partial\eta}\right), \tag{8.16a}$$

$$\frac{\partial}{\partial y} = \frac{1}{c^2h}\left(\cosh\xi\sin\eta\frac{\partial}{\partial\xi} + \sinh\xi\cos\eta\frac{\partial}{\partial\eta}\right), \tag{8.16b}$$

and the relation $h^2 = \cosh^2\xi - \cos^2\eta$. Now, we substitute the values of the integrals $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ from (8.15) in (8.14) and use the above formulas to obtain the required expressions for the exterior displacements $u^s_1(\mathbf{x})$, $u^s_2(\mathbf{x})$ in the elliptic coordinates. From these solutions we can evaluate the outer stress field $\tau^s_{\alpha\beta}(\mathbf{x})$ or $\tau_{\alpha\beta}(\mathbf{x})$. All these expressions are quite cumbersome except the expression for $\tau_{\alpha\alpha}(\mathbf{x})$, which is of great physical significance to evaluate the stress intensity factor. When the guest elliptic cylinder is a cavity, $\tau_{\alpha\alpha}(\mathbf{x})$ is given as

$$\tau_{\alpha\alpha}(\mathbf{x}) = \tau^0_{\alpha\alpha}(\mathbf{x}) + \tau^s_{\alpha\alpha}(\mathbf{x}) = T + 2(\lambda_1 + \mu_1)u^s_{\alpha,\alpha}(\mathbf{x}),$$
$$\xi > \xi_0. \quad (8.17)$$

The value of $u^s_{\alpha,\alpha}(\mathbf{x})$ needed in the above formula follows, from Eq. (8.14), to be

$$u^s_{\alpha,\alpha}(\mathbf{x}) = \frac{c}{8\pi M_1}(\Delta\lambda u_{\delta\delta}(0) + 2\Delta\mu u_{11}(0))\frac{\partial}{\partial x}(\nabla^2 I_1(\mathbf{x})) + \frac{c}{8\pi M_1}(\Delta\lambda u_{\delta\delta}(0) + 2\Delta\mu u_{22}(0))\frac{\partial}{\partial y}(\nabla^2 I_2(\mathbf{x})) + \frac{c\Delta\mu u_{12}(0)}{4\pi M_1}\left\{\frac{\partial}{\partial y}(\nabla^2 I_1(\mathbf{x})) + \frac{\partial}{\partial x}(\nabla^2 I_2(\mathbf{x}))\right\} \quad \xi > \xi_0. \quad (8.18)$$

But, by using (8.15) and (8.16), we have

$$\nabla^2(I_1(\mathbf{x})) = -4\pi \sinh 2\xi_0\, e^{-\xi}\cos\eta,$$
$$\nabla^2(I_2(\mathbf{x})) = -4\pi \sinh 2\xi_0 e^{-\xi}\sin\eta, \quad \xi > \xi_0. \quad (8.19)$$

Substituting relations (8.19) in (8.18) and again using formulas (8.16) we have

$$u^s_{\alpha,\alpha}(\mathbf{x}) = \frac{(\Delta\mu \sinh 2\xi_0)[(\cos 2\eta - e^{-2\xi})(u_{11}(0) - u_{22}(0)) + 2\sin 2\eta u_{12}(0)]}{M_1[\cosh 2\xi - \cos 2\eta]}, \quad \xi > \xi_0. \quad (8.20)$$

Furthermore, the values of the coefficients $u_{11}(0)$, $u_{22}(0)$ occurring in this result are given by Eq. (3.15), which yield

$$u_{11}(0) - u_{22}(0) = \frac{\{[\alpha_1 + \gamma_1 - \alpha_2 + \gamma_2 + 2(1 + \Delta\lambda/2\Delta\mu)\delta_1]u^0_{11}(0) + [\alpha_1 - \gamma_1 - \alpha_2 - \gamma_2 + 2(1 + (\Delta\lambda/2\Delta\mu)\delta_1)]u^0_{22}(0)\}}{2(\alpha_1\gamma_1 + \beta_1\delta_1)}, \quad (8.21)$$

where $u^0_{11}(0)$ and $u^0_{22}(0)$ are given by (8.10d) and (8.10e) while $\alpha_1, \alpha_2, \beta_1, \gamma_1, \gamma_2, \delta_1$ are given by (5.12). Similarly, the value of $u_{12}(0)$ in (8.20) is given by (5.14) in terms of $u^0_{12}(0)$ which, in turn, are defined in (8.10f). Thereby, all the terms in (8.20) are known and $u^s_{\alpha,\alpha}(\mathbf{x})$ is completely determined. When this value is substituted in formula (8.17), we obtain the required value of $\tau_{\alpha\alpha}(\mathbf{x})$, $\xi > \xi_0$.

The expression for $\tau_{\alpha\alpha}(\mathbf{x})$ is further simplified when the guest elliptic cylinder is a cavity. In this case, $\lambda_2 = \mu_2 = 0$, so that $\Delta\lambda = -\lambda_1$, $\Delta\mu = -\mu_1$ and we find from the above results that

$$u_{11}(0) - u_{22}(0) = \frac{TM_1[e^{2\xi_0}\cos 2\alpha - 1]}{2\mu_1(\lambda_1 + \mu_1)\sinh 2\xi_0}, \quad (8.22a)$$

$$u_{12}(0) = \frac{TM_1 e^{2\xi_0}\sin 2\alpha}{4\mu_1(\lambda_1 + \mu_1)\sinh 2\xi_0}, \quad (8.22b)$$

$$u^s_{\alpha,\alpha}(\mathbf{x}) = \frac{T[(\cos 2\eta - e^{-2\xi})(1 - e^{2\xi_0}\cos 2\alpha) - e^{2\xi_0}\sin 2\alpha \sin 2\eta]}{2(\lambda_1 + \mu_1)(\cosh 2\xi - \cos 2\eta)}, \quad \xi > \xi_0, \quad (8.22c)$$

and consequently Eq. (8.17) yields

$$\tau_{\alpha\alpha}(\mathbf{x}) = \tau_{\xi\xi}(\mathbf{x}) + \tau_{\eta\eta}(\mathbf{x}) = \frac{T[\sinh 2\xi - e^{2\xi_0}\cos 2(\eta - \alpha) + e^{-2(\xi - \xi_0)}\cos 2\alpha]}{[\cosh 2\xi - \cos 2\eta]}, \quad \xi > \xi_0. \quad (8.22d)$$

Formula (8.22d) agrees with the one given by Muskhelisvili[4] and serves as a check on our analysis.

Since, in the case of the elliptic cylindrical cavity $\tau_{\xi\xi}(\xi_0, \eta) = 0$, relation (8.22d) yields

$$\tau_{\eta\eta}(\xi_0, \eta) = \frac{T[\sinh 2\xi_0 - e^{2\xi_0}\cos 2(\eta - \alpha) + \cos 2\alpha]}{[\cosh 2\xi_0 - \cos 2\eta]}, \quad 0 \leqslant \eta \leqslant 2\pi. \quad (8.23)$$

Hence, the intensity factor $\tau^*_{\eta\eta}(\xi_0, \eta)$ for the elliptic cylindrical cavity is

$$\tau^*_{\eta\eta}(\xi_0, \eta) = \frac{\tau_{\eta\eta}(\xi_0, \eta)}{T} = \frac{[\sinh 2\xi_0 - e^{2\xi_0}\cos 2(\eta - \alpha) + \cos 2\alpha]}{[\cosh 2\xi_0 - \cos 2\eta]}, \quad 0 \leqslant \eta \leqslant 2\pi. \quad (8.24)$$

When we let $c \to 0$, $\xi_0 \to \infty$, $\xi \to \infty$ such that $c \sinh\xi_0$, $c\cosh\xi_0 \to a$, $c\sinh\xi$, $c\cosh\xi \to r$, then formulas (8.22d) and (8.24) reduce to the corresponding relations for the circular cylindrical cavity,

$$\tau_{rr}(r,\vartheta) + \tau_{\vartheta\vartheta}(r,\vartheta) = T[1 - (2a^2/r^2)\cos 2(\vartheta - \alpha)], r > a, \quad (8.25a)$$

$$\tau^*_{\vartheta\vartheta}(a,\vartheta) = \tau_{\vartheta\vartheta}(a,\vartheta)/T = \{1 - 2\cos 2(\vartheta - \alpha)\}, \quad 0 \geqslant \vartheta \geqslant 2\pi, \quad (8.25b)$$

which agree with the known results.[9]

When $\xi_0 \to 0$ in (8.24) we get the corresponding results for an infinite crack $|x| < c$, $y = 0$, $|z| < \infty$,

$$\tau^*_{xx}(x,0) = \frac{[\cos 2\alpha - \cos 2(\eta - \alpha)]}{[1 - \cos 2\eta]}, \quad \eta = \arccos(x/c), \quad |x| < c. \quad (8.26)$$

[1]F. C. Chen and K. Young, J. Math. Phys. **18**, 1412 (1977).
[2]P. C. Waterman, J. Acoust. Soc. Am. **45**, 1417 (1969).
[3]L. Eyges, Ann. Phys. (NY) **81**, 567 (1973).
[4]N. I. Muskhelisvili, *Some Basic Problems of Mathematical Theory of Elasticity* (Noordhoff, Groningen, Netherlands, 1953).
[5]I. S. Sokolnikoff, *Mathermatical Theory of Elasticity*, 2nd ed. (McGraw-Hill, New York, 1956).
[6]A. I. Lure, *Three-dimensional Problems of the Theory of Elasticity*, 2nd ed. (Interscience, New York, 1978).
[7]A. E. H. Love, *Mathematical Theory of Elasticity* (Dover, New York, 1944).
[8]D. L. Jain and R. P. Kanwal, J. Integ. Equat. **4**, 31 (1982).
[9]J. N. Goodier, J. Appl. Mech. **55**, 39 (1933).
[10]D. L. Jain and R. P. Kanwal, J. Appl. Phys. **50**, 4067 (1979).

Time-dependent scattering by a bounded obstacle in three dimensions

Anders Boström
Institute of Theoretical Physics, S-412 96 Göteborg, Sweden

(Received 3 April 1981; accepted for publication 29 May 1981)

In the present paper we introduce a new method of treating the scattering of transient fields by a bounded obstacle in three-dimensional space. The method is a generalization to the time domain of the null field approach first given by Waterman. We define new sets of time-dependent basis functions, and use these to expand the free space Green's function and the incoming and scattered fields. The scattering problem is then reduced to the problem of solving a system of ordinary differential equations. One way of solving these equations is by means of Fourier transformation, and this leads to an efficient way of obtaining the natural frequencies of the obstacle. Finally, we have calculated the natural frequencies numerically for both a spheroid and a peanut-shaped obstacle for various ratios of the axes.

PACS numbers: 03.40.Kf, 02.30.Jr

I. INTRODUCTION

The radiation and scattering of transient waves have attracted much past and present interest, and many important applications can be found, for instance, in electromagnetics or geophysics. The standard approach to such problems is to apply a Fourier (or Laplace) transform. The transformation back to the time domain may then be cumbersome—often asymptotic or numerical techniques must be employed. Other approaches have also been proposed working directly in the time domain. Thus one can formulate and solve various types of integral equations or one can solve the wave equation numerically by a finite difference approach.

In the present paper we introduce a new method to treat the scattering of transient waves by a bounded obstacle in three-dimensional space. The method has many features in common with the null field approach (also called the extended boundary condition or T matrix method) first introduced by Waterman.[1] This approach has so far been applied mainly to stationary problems (but also to static ones), though by integrating in frequency it is of course possible to obtain solutions also to transient problems.[2]

The essential and novel in our method is the introduction of two new sets of basis functions containing a timelike parameter instead of the frequency. These functions have a rather elementary appearance, but it seems that they are not to be found in the literature. Once we have the sets of basis functions and the expansion of the Green's function in these sets, the same ideas as in the ordinary null field approach can be used. This leads to a set of coupled ordinary differential equations in the timelike parameter with constant coefficients that are integrals over the surface of the obstacle.

These differential equations can be solved by various techniques. One possibility is to apply a Fourier transform, and we then have an efficient way of obtaining the natural frequencies of the obstacle. By this technique we can also calculate the surface and scattered fields. We have performed numerical computations of the natural frequencies for both a spheroid and a peanut-shaped obstacle for various ratios of the axes.

II. BASIS FUNCTIONS

Consider the scattering of a transient wave by a bounded obstacle in three-dimensional space. We treat the case of a scalar field u, which thus satisfies the wave equation

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}, \tag{2.1}$$

where c is the wave velocity. The surface S of the obstacle must be sufficiently smooth to allow an application of the divergence theorem. To simplify the formulas, we take homogeneous Dirichlet's boundary condition on S:

$$u(\mathbf{r}) = 0, \quad \mathbf{r} \text{ on } S. \tag{2.2}$$

Our origin is chosen somewhere inside S.

By introducing the free space Green's function G it is easy to derive the following integral representation[3]

$$\begin{aligned}\int_{-\infty}^{\infty} dt' \int_S dS' \Big\{ & u_+(\mathbf{r}',t')\frac{\partial}{\partial n'}G(\mathbf{r},t;\mathbf{r}',t') \\ & - G(\mathbf{r},t;\mathbf{r}',t')\Big[\frac{\partial}{\partial n'}u(\mathbf{r}',t')\Big]_+\Big\} \\ & + u^i(\mathbf{r},t) = \begin{cases} u(\mathbf{r},t), & \mathbf{r} \text{ outside } S, \\ 0, & \mathbf{r} \text{ inside } S. \end{cases}\end{aligned} \tag{2.3}$$

Here $\partial/\partial n$ is the normal derivative on S and u^i is the prescribed incident field. As in the ordinary null field approach, (2.3) is the basic formula from which we derive our equations. The idea is to expand all fields in some global sets of functions and then compute the surface field from the second line of (2.3) and thereafter the scattered field from the first line.

The explicit form of the Green's functions is

$$G(\mathbf{r},t;\mathbf{r}',t') = \frac{\delta(|\mathbf{r}-\mathbf{r}'|/c-(t-t'))}{4\pi|\mathbf{r}-\mathbf{r}'|}, \tag{2.4}$$

and its expansion as a Fourier integral and in spherical waves is

$$\begin{aligned}G(\mathbf{r},t;\mathbf{r}',t') = {} & \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega(t-t')}\frac{\omega}{c} \\ & \times \sum_n j_l(\omega r_</c)h_l^{(1)}(\omega r_>/c)Y_n(\hat{r})Y_n(\hat{r}').\end{aligned} \tag{2.5}$$

 0022-2488/82/081444-07$02.50

Here j_l and $h_l^{(1)}$ are the spherical Bessel and Hankel functions, respectively, and $r_<$ ($r_>$) means the smaller (bigger) of r and r'. We use a normalized real spherical harmonic,

$$Y_n(\hat{r}) \equiv Y_{\sigma ml}(\theta,\varphi) = \left[\frac{\epsilon_m}{2\pi}\frac{2l+1}{2}\frac{(l+m)!}{(l-m)!}\right]^{1/2} \times P_l^m(\cos\theta)\binom{\cos m\varphi}{\sin m\varphi}, \tag{2.6}$$

where $\epsilon_m = 2 - \delta_{m0}$ and P_l^m is a Legendre function; $\sigma = \mathrm{e,o}$ (even,odd), $m = 0,1,...,l$, and $l = 0,1,\cdots$. The summation over n in (2.5), of course, denotes a triple summation over σ, m, l.

The expansion (2.5) is not quite the form of the Green's function that we want. Using it, we would simply obtain the standard formulas of the stationary null field approach, but with a frequency integral in the solution to obtain the transient behavior. We keep the expansion in spherical harmonics in (2.5), but the frequency expansion is now modified. Multiply by the factor

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau \int_{-\infty}^{\infty} d\omega' \exp[i(\tau - t')(\omega + \omega')] = 1, \tag{2.7}$$

and change the order of integrations to get

$$G(\mathbf{r},t;\mathbf{r}',t') = \frac{1}{a}\sum_n \int_{-\infty}^{\infty} d\tau \, \mathrm{Re}\psi_n(\tau;\mathbf{r}',t')\psi_n(\tau;\mathbf{r},t), \quad r > r'. \tag{2.8}$$

We have here assumed that $r > r'$. The basis functions are defined by the following integrals:

$$\mathrm{Re}\psi_n(\tau;\mathbf{r},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \, e^{-i\omega(t-\tau)}(\omega a/c)^{-l} j_l(\omega r/c) Y_n(\hat{r}) \tag{2.9}$$

and

$$\psi_n(\tau;\mathbf{r},t) = \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \, e^{-i\omega(t-\tau)}(\omega a/c)^{l+1} h_l^{(1)}(\omega r/c) Y_n(\hat{r}). \tag{2.10}$$

The extra factors of ω are introduced to ensure the convergence of the integrals at the origin, and c and the length a (which is at our disposal) are inserted for dimensional purposes. The integrals can in fact be performed analytically. For the regular function we obtain the following very explicit form[4]:

$$\mathrm{Re}\psi_n(\tau;\mathbf{r},t) = \begin{cases} 0, & r < c|t-\tau|, \\ \dfrac{c}{a 2^{l+1} l!}\left(\dfrac{r}{a}\right)^{l-1}[1 - c^2(t-\tau)^2/r^2]^l Y_n(\hat{r}), & r > c|t-\tau|. \end{cases} \tag{2.11}$$

That this is indeed a solution of the wave equation is easily demonstrated, but it seems that this fact has not been noted in the literature. Differentiating $l-k$ times with respect to t, we easily see that $r^{k-1}(1-(ct/r)^2)^{k/2} P_l^k(ct/r) Y_n(\hat{r})$ is also a solution of the wave equation. Because of the connection between the r and t dependence $\mathrm{Re}\psi_n(\tau;\mathbf{r},t)$ is not that elementary, though. We see that it is unbounded as $r \to \infty$, except for $l = 0$ and 1. On the other hand, it is regular at the origin. The function with $l = 0$ is exceptional: It goes to zero at $r \to \infty$, it has a jump discontinuity at $r = c|t-\tau|$ [whereas for $l \neq 0$ $\mathrm{Re}\psi_n(\tau;\mathbf{r},t)$ is continuous] and it has a $\delta(t-\tau)$ singularity at $r = 0$ [cf. (2.9)]. Except for the restriction on the region of validity, $\mathrm{Re}\psi_n(\tau;\mathbf{r},t)$ for $l = 0$ is in fact a static solution of the wave equation.

For the outgoing basis function the integration in (2.10) gives[5]

$$\psi_n(\tau;\mathbf{r},t) = \frac{a}{r}\left(\frac{a}{c}\right)^l \sum_{k=0}^{l} (-1)^{l-k}\xi_{lk}\left(\frac{c}{r}\right)^k \times \delta^{(l-k)}(\tau - t + r/c) Y_n(\hat{r}), \tag{2.12}$$

where

$$\xi_{lk} = (l+k)!/2^k k!(l-k)!. \tag{2.13}$$

That this "function" represents an outgoing spherical wave is immediately clear. As it will always appear in an integral over τ or t the appearance of $\delta^{(l-k)}$ will lead to simplifications in the following. We note that, apart from normalization, $\psi_{e00}(0;\mathbf{r}-\mathbf{r}',t-t')$ is just the free space Green's function.

By inverting the Fourier transform in (2.9) we obtain

$$e^{-i\omega t} j_l(\omega r/c) Y_n(\hat{r}) = (\omega a/c)^l \int_{-\infty}^{\infty} d\tau \, e^{-i\omega\tau} \mathrm{Re}\psi_n(\tau;\mathbf{r},t). \tag{2.14}$$

This gives us an easy way to relate the expansion coefficients of the two types of expansions. Assume the following expansions of some sufficiently regular function $f(\mathbf{r},t)$ (which must satisfy the wave equation):

$$f(\mathbf{r},t) = \sum_n \int_{-\infty}^{\infty} d\omega \, \tilde{f}_n(\omega) e^{-i\omega t} j_l(\omega r/c) Y_n(\hat{r}) = \sum_n \int_{-\infty}^{\infty} d\tau \, \hat{f}_n(\tau) \, \mathrm{Re}\psi_n(\tau;\mathbf{r},t). \tag{2.15}$$

We then have the relations

$$\hat{f}_n(\tau) = \int_{-\infty}^{\infty} d\omega \, \tilde{f}_n(\omega)(\omega a/c)^l e^{-i\omega\tau} \tag{2.16}$$

and

$$\tilde{f}_n(\omega) = (1/2\pi)(\omega a/c)^{-l}\int_{-\infty}^{\infty} d\tau \, \hat{f}_n(\tau) e^{i\omega\tau}. \tag{2.17}$$

Thus we see that the expansion in our regular basis functions can be computed with the help of the ordinary Fourier expansion. In this way it should also be possible to show the completeness of our regular functions. Similar considerations as for the regular functions hold, of course, also for the outgoing ones, so there is no need to write down the corresponding equations.

The expansion (2.8) of the Green's function holds only for $r > r'$. By using the reciprocity relationship for the Green's function we have for $r < r'$

$$G(\mathbf{r},t;\mathbf{r}',t') = G(\mathbf{r}',-t';\mathbf{r},-t) = \frac{1}{a}\sum_n \int_{-\infty}^{\infty} d\tau \, \mathrm{Re}\psi_n(\tau;\mathbf{r},t)\psi_n(-\tau;\mathbf{r}',-t'), \quad r < r', \tag{2.18}$$

where we have also used the fact that $\mathrm{Re}\psi_n(\tau;\mathbf{r},t)$ is even in $\tau - t$.

III. REDUCTION OF THE SCATTERING PROBLEM

We now return to our scattering problem as stated in the beginning of the previous section. If we insert the boundary condition (2.2) in the integral representation (2.3), we obtain

$$-\int_{-\infty}^{\infty} dt' \int_S dS'\, G(\mathbf{r},t;\mathbf{r}',t')v(\mathbf{r},t') + u^i(\mathbf{r},t) = \begin{cases} u(\mathbf{r},t), & \mathbf{r} \text{ outside } S, \\ 0, & \mathbf{r} \text{ inside } S, \end{cases} \tag{3.1}$$

where we have defined

$$v(\mathbf{r},t) = \left[\frac{\partial}{\partial n} u(\mathbf{r},t)\right]_+ . \tag{3.2}$$

To calculate the surface field $v(\mathbf{r},t)$ we assume that $\mathbf{r}$ in (3.1) is inside the inscribed sphere to S. Inserting the expansion (2.18) of the Green's function, we obtain an expansion of the incident field,

$$u^i(\mathbf{r},t) = \sum_n \int_{-\infty}^{\infty} d\tau\, a_n(\tau)\, \mathrm{Re}\psi_n(\tau;\mathbf{r},t), \tag{3.3}$$

where we have the relation

$$a_n(\tau) = (1/a)\int_{-\infty}^{\infty} dt \int_S dS\, \psi_n(-\tau;\mathbf{r},-t)v(\mathbf{r},t). \tag{3.4}$$

The coefficients $a_n(\tau)$ of the incident field are, of course, regarded as known. To compute $v(\mathbf{r},t)$, we therefore only have to invert (3.4). Employing the explicit form (2.12) of the outgoing basis function, we obtain

$$a_n(\tau) = (1/a)\int_S dS\, Y_n(\hat{r})D_t^{(l)}(r)v(\mathbf{r},t)\Big|_{t=\tau-r/c}, \tag{3.5}$$

where the differential operator is

$$D_t^{(l)}(r) = \frac{a}{r}\left(\frac{a}{c}\right)^l \sum_{k=0}^{l} \xi_{lk}\left(\frac{c}{r}\right)^k \frac{\partial^{(l-k)}}{\partial t^{(l-k)}}. \tag{3.6}$$

To proceed, we must make some kind of expansion of the surface field $v(\mathbf{r},t)$. There are several alternatives for this, the regular or outgoing basis functions, for instance. For reasons which we will soon discuss we make the following expansion in spherical harmonics:

$$v(\mathbf{r},t) = (1/a)\sum_n v_n(t + r/c)Y_n(\hat{r}), \tag{3.7}$$

where the functions $v_n(t + r/c)$ are to be determined. Inserting this in (3.5) gives

$$a_n(\tau) = \sum_{n'}\sum_{k=0}^{l} \xi_{lk}\left(\frac{a}{c}\right)^{l-k} Q_{nn'}^{k} \frac{d^{(l-k)}}{d\tau^{(l-k)}} v_{n'}(\tau), \tag{3.8}$$

where we have introduced the matrix

$$Q_{nn'}^{k} = (1/a^2)\int_S dS\, (a/r)^{k-1} Y_n(\hat{r})Y_{n'}(\hat{r}). \tag{3.9}$$

It is now apparent why we choose the argument in v_n in (3.7) to be $t + r/c$ (a more immediate choice would be simply t). Another choice would have given an r-dependent argument in v_n in (3.8), and thus v_n would have to appear in the integrand in the surface integral. For a sphere it is clear that the expansion in (3.7) is complete, but for a nonspherical obstacle this is not evident, though certainly to be expected with suitable restrictions on the surface S. Furthermore, our numerical results (see the next section) indicate that this expansion is useful also for nonspherical obstacles.

The problem of calculating the surface field is thus reduced to the simpler problem of solving a system of coupled linear differential equations with constant coefficients. This is a thoroughly studied field, and there exist many methods for obtaining the solution. One possibility is to apply a Fourier transform, a method which we will study in more detail in the next section.

Once we have calculated the surface field, there remains the calculation of the scattered field $u^s = u - u^i$. Taking an $\mathbf{r}$ outside the circumscribed sphere to S in (3.1) and inserting the expansion (2.8) of the Green's function, we obtain

$$u^s(\mathbf{r},t) = \sum_n \int_{-\infty}^{\infty} d\tau\, f_n(\tau)\psi_n(\tau;\mathbf{r},t), \tag{3.10}$$

where the expansion coefficients are given by

$$f_n(\tau) = -\frac{1}{a}\int_{-\infty}^{\infty} dt \int_S dS\, \mathrm{Re}\psi_n(\tau;\mathbf{r},t)v(\mathbf{r},t). \tag{3.11}$$

If we use the explicit form (2.12) of the outgoing basis function, we have

$$u^s(\mathbf{r},t) = \sum_n Y_n(\hat{r})D_\tau^{(l)}(r)f_n(\tau)\Big|_{\tau=t-r/c}, \tag{3.12}$$

where the differential operator is defined in (3.6) above. Here τ plays the role of retarded time. Formulas similar to (3.12) have been given by Granzow.[6]

The expansion of the scattered field can in fact be written in a more explicit form. From (3.11) and (3.12) we have

$$u^s(\mathbf{r},t) = (a/r)\sum_n b_n(r,t - r/c)Y_n(\hat{r}), \tag{3.13}$$

where [using the explicit form (2.11) of the regular basis function]

$$\begin{aligned} b_n(r,\tau) &= -\frac{a^{l-1}}{c^l}\sum_{k=0}^{l}\xi_{lk}\left(\frac{c}{r}\right)^k \int_{-\infty}^{\infty} dt \int_S dS' \\ &\quad\times\left[\frac{\partial^{(l-k)}}{\partial\tau^{(l-k)}}\mathrm{Re}\psi_n(\tau;\mathbf{r}',t)\right]v(\mathbf{r}',t) \\ &= \frac{c}{a}\sum_{k=0}^{l}\frac{(-1)^{l+k+1}}{2^{k+1}k!}r^{-k} \\ &\quad\times\int_S dS'\,(r')^{k-1}\int_{\tau-r'/c}^{\tau+r'/c} dt\left[1-\frac{c^2(t-\tau)^2}{r'^2}\right]^{k/2} \\ &\quad\times P_l^k\left(\frac{c(\tau-t)}{r'}\right)Y_n(\hat{r}')v(\mathbf{r}',t) \\ &= \frac{1}{a}\sum_{k=0}^{l}\frac{(-1)^{l+k+1}}{2^{k+1}k!} \\ &\quad\times\int_{-1}^{1} dx\,(1-x^2)^{k/2}P_l^k(x) \\ &\quad\times\int_S dS'\left(\frac{r'}{r}\right)^k Y_n(\hat{r}')v\left(\mathbf{r}',\tau-\frac{xr'}{c}\right). \end{aligned} \tag{3.14}$$

We note that the solution of the wave equation mentioned below (2.11) enters here. The scattered field is thus obtained from the surface field by an integration over the surface and a further integration over the time interval that can contribute. In the far field only the first term in the summation over k in (3.14) contributes.

Another, and more traditional, way of obtaining the scattered field is to use the explicit form of the Green's function in the integral representation. Thus (2.2), (2.3), and (2.4) give

$$u^s(\mathbf{r},t) = \frac{1}{4\pi}\int_S dS' \frac{1}{|\mathbf{r}-\mathbf{r}'|} v\left(\mathbf{r}', t - \frac{|\mathbf{r}-\mathbf{r}'|}{c}\right), \quad (3.15)$$

which in the far field reduces to

$$u^s(\mathbf{r},t) = \frac{1}{4\pi r}\int_S dS'\, v\left(\mathbf{r}', t - \frac{r}{c} + \frac{\hat{r}\cdot\mathbf{r}'}{c}\right). \quad (3.16)$$

As these equations contain no integration in time, they are possibly more advantageous to use than (3.13) and (3.14). On the other hand, (3.13) and (3.14) give the scattered field as a sum of spherical harmonics, and this can sometimes be a desirable feature. The surface integral in (3.14) is, furthermore, simpler than that in (3.15).

IV. CALCULATION OF NATURAL FREQUENCIES AND NUMERICAL EXAMPLES

We now study one way to solve the equations of the previous section. Multiply (3.8) by $e^{i\omega\tau}$ and integrate:

$$\frac{1}{2\pi}\left(\frac{\omega a}{c}\right)^{-l}\int_{-\infty}^{\infty} d\tau\, a_n(\tau)e^{i\omega\tau}$$
$$= i^{-l}\sum_{n'}\sum_{k=0}^{l}\xi_{lk}\left(\frac{-i\omega a}{c}\right)^{-k} Q^k_{nn'}\frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, v_{n'}(\tau)e^{i\omega\tau}. \quad (4.1)$$

The left-hand side of this equation is just the expansion coefficient in a Fourier expansion of the incident field, cf. (2.17). Thus we have

$$a_n(\omega) = \sum_{n'} W_{nn'}(\omega)\frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, v_{n'}(\tau)e^{i\omega\tau}, \quad (4.2)$$

where

$$W_{nn'}(\omega) = i^{-l}\sum_{k=0}^{l}\xi_{lk}(-i\omega a/c)^{-k}Q^k_{nn'}. \quad (4.3)$$

Formally solving for the expansion function of the surface field yields

$$v_n(\tau) = \sum_{n'}\int_{-\infty}^{\infty} d\omega\, W^{-1}_{nn'}(\omega)a_{n'}(\omega)e^{-i\omega\tau}. \quad (4.4)$$

If all the singularities of $W^{-1}_{nn'}(\omega)$ are known, we can cast this equation into a more convenient form. For a sphere with radius a we have

$$W_{nn'}(\omega) = i\frac{\omega a}{c}e^{-i\omega a/c}h_l^{(1)}\left(\frac{\omega a}{c}\right)\delta_{nn'}, \quad (4.5)$$

so there are only simple poles in this case. We likewise assume that only simple poles of $W^{-1}_{nn'}(\omega)$ in the third and fourth quadrants (symmetrically situated) will contribute when the integral in (4.4) is closed in the lower half-plane. The surface field then becomes

$$v(\mathbf{r},t) = -\frac{i}{a}\sum_{ns} a_n(\omega_s)\varphi_{ns}(\mathbf{r})\exp(-i\omega_s t), \quad (4.6)$$

where

$$\varphi_{ns}(\mathbf{r}) = 2\pi\sum_{n'}\operatorname{Res} W^{-1}_{n'n}(\omega)\Big|_{\omega=\omega_s} Y_{n'}(\hat{r})\exp(-i\omega_s r/c) \quad (4.7)$$

and ω_s are the zeros (assumed simple) of

$$\det[W_{nn'}(\omega)] = 0. \quad (4.8)$$

Of course, (4.6) is only valid for times such that the closing of the integral in the lower half-plane is legitimate—for earlier times the integral must be closed in the upper half-plane and we then expect the surface field to vanish. When calculating (4.6), we assumed that the expansion coefficient $a_n(\omega)$ of the incident field has no poles. If it has, the corresponding contributions must be added to (4.6).

Once the surface field is computed by means of (4.6) [or (4.4)], the scattered field can be obtained by employing (3.13)–(3.16), and we have thereby solved our scattering problem. However, the computational aspects of our theory and comparisons with other methods still remain to investigate.

To represent the solution in terms of natural frequencies ω_s and corresponding natural modes as is done in (4.6) is not new. Derived in a completely different manner, it is often referred to as the singularity expansion method (SEM).[7]

There exists another way of deriving the equation that determines the natural frequencies of the obstacle. If we use the ordinary null field approach, we have in analogy with (3.4) an equation relating the expansion coefficients for the incoming field and the surface field (in the frequency domain)

$$a_n(\omega) = \frac{i\omega}{c}\int_S dS\, h_l^{(1)}\left(\frac{\omega r}{c}\right)Y_n(\hat{r})v(\mathbf{r},\omega). \quad (4.9)$$

Expanding the surface field in spherical harmonics,

$$v(\mathbf{r},\omega) = e^{-i\omega r/c}\sum_n \alpha_n(\omega)Y_n(\hat{r}), \quad (4.10)$$

and inserting this and the exact sum for the spherical Hankel function, we arrive at the conclusion that the matrix relating $a_n(\omega)$ and $\alpha_n(\omega)$ is essentially $W_{nn'}(\omega)$. Thus we get the same

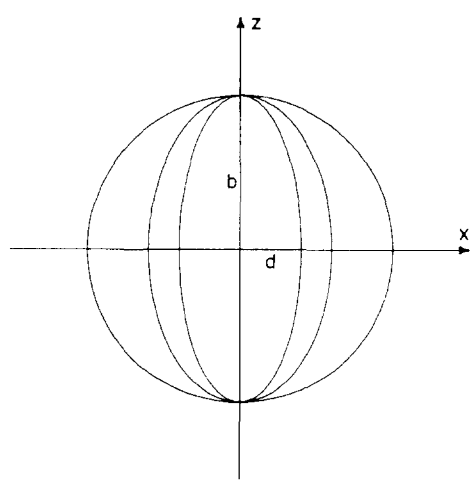


FIG. 1. Spheroidal obstacles with $d/b = 1, 0.6, 0.4$.

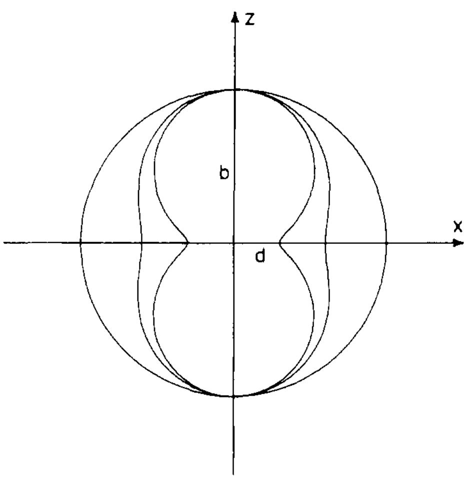


FIG. 2. Peanut-shaped obstacles with $d/b = 1, 0.6, 0.3$.

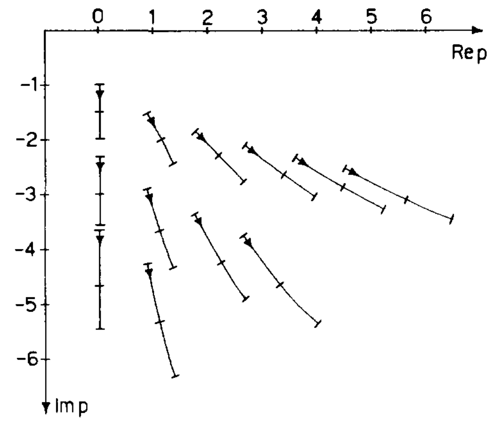


FIG. 4. Same as in Fig. 3 but for $m = 1$.

form for the equation determining the natural frequencies as with our more elaborate procedure in the time domain.

To illustrate the applicability of our formalism, we now turn to a few numerical examples. We will compute the natural frequecies for some simple obstacles by using Eq. (4.8). By inspecting (4.3) we note one very important fact about the matrix $W_{nn'}(\omega)$, namely that the frequency dependence is explicit and not in the surface integral defining $Q^k_{nn'}$. For a given obstacle we therefore first compute (by numerical integration) and store $Q^k_{nn'}$, and then it is an easy and fast task to compute $\det[W_{nn'}(\omega)]$ for various values of ω.

We consider rotationally symmetric obstacles, but we compute natural frequencies also for other values of azimuthal order m than $m = 0$. We first take an origin-shifted spherical obstacle with radius a and shift of origin b. We must then use our full formalism, and this is a good check on both the analytical and numerical performance of the method. For a shift $b/a = 0.5$ and truncation $l_{\max} = 15$ [the maximum value of l used in (3.9), (4.3), and (4.8)] we obtain the six first natural modes (the zeros of $h_l^{(1)}$, $l = 1,2,3,4$) to seven correct figures.

Numerical computations of natural frequencies have been performed for two different kinds of obstacles. The first is a spheroid (see Fig. 1), where the equation for the surface is

$$r(\theta) = bd\,(b^2 \sin^2\theta + d^2 \cos^2\theta)^{-1/2}, \tag{4.11}$$

and the second is a "peanut" (see Fig. 2), where the equation is

$$r(\theta) = (d^2 \sin^2\theta + b^2 \cos^2\theta)^{1/2}. \tag{4.12}$$

The peanut has the merits of a simple defining equation and of being an example of a body that is not wholly convex. In both (4.11) and (4.12) we can have $b > d$ or $d > b$; $b > d$ gives a prolate spheroid or a peanut and $d > b$ gives an oblate spheroid or an "apple" (the body one gets by rotating the curve in Fig. 2 around the x axis). The parameter a is chosen equal to the radius of the sphere circumscribing the obstacle, i.e., $a = b$ or $a = d$ depending on $b > d$ or $d > b$, respectively.

Twelve natural frequencies have been computed for our obstacles. For the limit of a sphere this corresponds to the zeros of $h_l^{(1)}$ for $l = 1,...,6$. As the zeros appear in pairs symmetrically located in the third and fourth quadrants (except when they are purely imaginary), we only give the locations

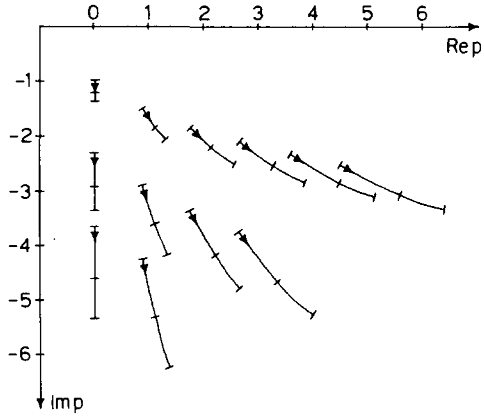


FIG. 3. The locus of the natural frequencies for a prolate spheroid for $m = 0$ when $0.4 \leqslant d/b \leqslant 1$. The natural frequencies for $d/b = 0.4, 0.6, 1$ are indicated on the curves with the arrows pointing away from $d/b = 1$.

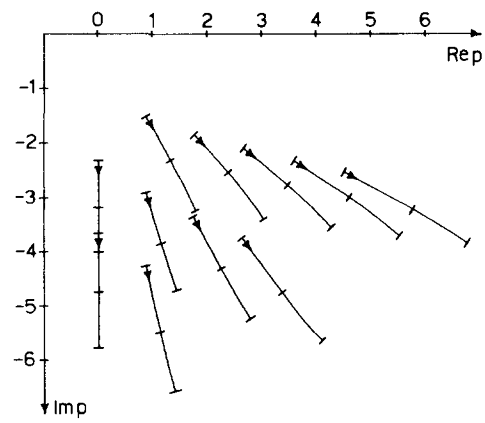


FIG. 5. Same as in Fig. 3 but for $m = 2$.

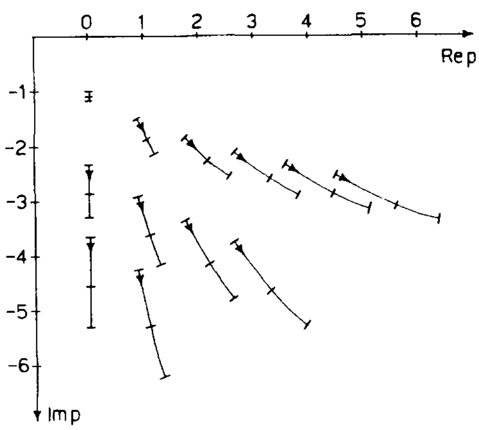


FIG. 6. The locus of the natural frequencies for an oblate spheroid for $m = 1$ when $0.4 \leqslant b/d \leqslant 1$. The natural frequencies for $b/d = 0.4, 0.6, 1$ are indicated on the curves with the arrows pointing away from $b/d = 1$.

in the fourth quadrant of the dimensionless frequency $p = \omega a/c$. Depending on how much the obstacle deviates from a sphere, we have used truncations $l_{\max}$ ranging from 15 to 31 [as the even and odd l values decouple in (4.8), the maximum matrix size is thus 16×16]. This gives errors in the natural frequencies that are expected to be less than 1%.

In Figs. 3, 4, and 5 we show the natural frequencies for a prolate spheroid for $m = 0$, 1, and 2, respectively. The ratio between the axes varies from $d/b = 1$ to $d/b = 0.4$, and the locations for $d/b = 1$, 0.6, and 0.4 are indicated on the curves (with an arrow pointing away from $d/b = 1$). Comparing the figures, we note that the m dependence of the natural frequencies is not very pronounced, but that at least for the first few modes the frequencies grow with increasing m. For $m = 2$ the first mode is, of course, missing.

To compare different kinds of obstacles we show in Figs. 6, 7, and 8 the natural frequencies for $m = 1$ for an oblate spheroid, a peanut, and an apple, respectively. In Fig. 6 the axes vary from $b/d = 1$ to $b/d = 0.4$, in Fig. 7 from $d/b = 1$ to $d/b = 0.3$, and in Fig. 8 from $b/d = 1$ to $b/d = 0.3$. Comparing spheroids (Figs. 4 and 6) with peanuts (Figs. 7 and 8), we observe that the natural frequencies for the former changes more from the spherical ones. At least partly, this may be a volume effect as a peanut has a larger

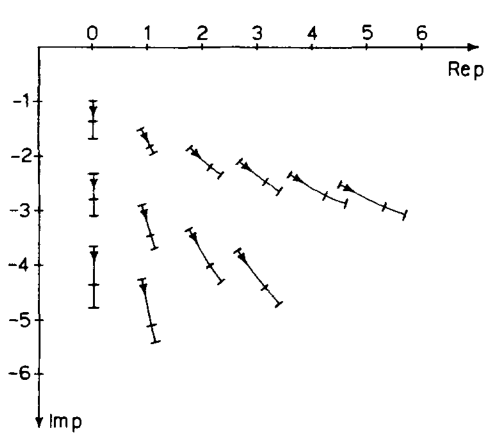


FIG. 7. The locus of the natural frequencies for a peanut for $m = 1$ when $0.3 \leqslant d/b \leqslant 1$. The natural frequencies for $d/b = 0.3, 0.6, 1$ are indicated on the curves with the arrows pointing away from $d/b = 1$.

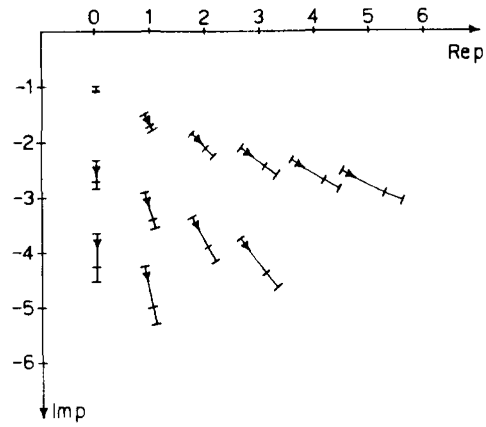


FIG. 8. The locus of the natural frequencies for an apple for $m = 1$ when $0.3 \leqslant b/d \leqslant 1$. The natural frequencies for $b/d = 0.3, 0.6, 1$ are indicated on the curves with the arrows pointing away from $b/d = 1$.

volume than a spheroid when the axes are equal. The most clear difference between prolate (Figs. 4 and 7) and oblate (Figs. 6 and 8) obstacles is the much smaller change in the first natural frequency for the oblate obstacles.

V. CONCLUDING REMARKS

So far we have considered the scattering of a time-dependent scalar field by an obstacle on which the field obeys a homogeneous Dirichlet's boundary condition. There are, however, some aspects of the present approach that should be further investigated. Thus the completeness of our expansion of the surface field should be proven, and there are several unsolved questions concerning the poles of $W_{nn'}^{-1}(\omega)$.[7,8] Furthermore, more computations should be performed, including computations of both surface and scattered fields. Also other methods of solving the ordinary differential equations (3.8) should be tried, and this could be useful, especially for early times.

Other boundary conditions on the surface of the obstacle are of interest. Homogeneous Neumann boundary conditions will not involve any particular difficulties, but if we consider a penetrable obstacle, the problem becomes more intricate. The integral representation for the interior must then be used also, and this may lead to useful formulas.

Another very interesting generalization is to treat the more complex electromagnetic and elastic cases. The complication then is the vector character of the fields. The usual way of defining the vector basis functions from the scalar ones can still be used, however, so these cases should be possible to handle.

A more far-reaching generalization would be to consider multiple-scattering problems. For stationary waves many cases can be handled efficiently by the null field approach, and it seems possible that these ideas in conjunction with the present work could be applied to treat time-dependent multiple scattering.

ACKNOWLEDGMENT

The present work is supported by the National Swedish Board for Technical Development (STU), and this is gratefully acknowledged.

[1]See, e.g., P. C. Waterman, J. Acoust. Soc. Am. **45**, 1417 (1969); for the electromagnetic case see, e.g., P. C. Waterman, Phys. Rev. D **3**, 825 (1971), and for the elastic case see P. C. Waterman, J. Acoust. Soc. Am. **60**, 567 (1976) and V. Varadan and Y.-H Pao, J. Acoust. Soc. Am. **60**, 556 (1976).
[2]A. Boström and G. Kristensson, "Scattering of a pulsed Rayleigh wave by a spherical cavity in an elastic half space," Report 80-27, Institute of Theoretical Physics, Göteborg, Sweden (1980).
[3]This is a slight generalization of a formula given in Chap. 7 of P. M. Morse and H. F. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953). Note that we use a different normalization of the Green's function and that we have extended the time integration.
[4]Formula 6.699.2 in I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965), gives the integral as a hypergeometric function, which can then be reduced to the given form by formulas 15.4.23, 8.6.6, and 8.6.18 in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, National Bureau of Standards, Washington, D. C., 1970).
[5]Use formula 10.1.16 in Abramowitz and Stegun, Ref. 4, and the representation $2\pi\delta^{(n)}(u) = i^n\int_{-\infty}^{\infty} e^{iux}x^n\,dx$ of the δ function and its derivatives.
[6]K. D. Granzow, J. Appl. Phys. **39**, 3435 (1968); for the electromagnetic case see also K. D. Granzow, J. Math. Phys. **7**, 634 (1966).
[7]Both the mathematical and applied literature in the field can be traced through contributions by C. L. Dolph, R. A. Scott, and C. E. Baum, in *Electromagnetic Scattering*, edited by P. L. E. Uslenghi (Academic, New York, 1978).
[8]For a few recent results, see A. G. Ramm, *Theory and Applications of Some New Classes of Integral Equations* (Springer, New York, 1980), especially Appendix 10.

Quantum evolution as a parallel transport

M. Asorey, J. F. Cariñena, and M. Paramio
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain

(Received 12 June 1981; accepted for publication 19 March 1982)

We point out that the evolution of a quantum system can be considered as a parallel transport of unitary operators in Hilbert spaces along the time with respect to a generalized connection. The different quantum representations of the system are shown to correspond to the choices of cross sections in the principal fiber bundle where the generalized connection is defined. This interpretation of time evolution allows us to solve the problem of the formulation of the evolution of a quantum particle in a four-dimensional gauge field.

PACS numbers: 03.65.Bz, 11.10.Np

I. INTRODUCTION

The motion of a nonrelativistic particle in the space-time $\mathbb{R}^4$ under the action of a four-dimensional electromagnetic gauge field A_μ is governed by the Schrödinger equation

$$i\frac{\partial}{\partial t}\psi(\mathbf{x},t) = -\frac{1}{2m}\sum_{j=1}^{3}(\partial_j - ieA_j(\mathbf{x},t))^2\psi(\mathbf{x},t) - eA_0(\mathbf{x},t)\psi(\mathbf{x},t), \tag{1.1}$$

ψ being the wave function of the particle, e and m its electrical charge and mass. We will use natural units $c = \hbar = 1$.

The quantum meaning of this equation is that the evolution of the system is governed by a time-dependent Hamiltonian $H(t)$ which is at each time $t \geqslant 0$ a self-adjoint extension in $L^2(\mathbb{R}^3)$ of the elliptic operator

$$-\frac{1}{2m}\sum_{j=1}^{3}(\partial_j - ieA_j(\mathbf{x},t))^2 - eA_0(\mathbf{x},t). \tag{1.2}$$

Such a self-adjoint extension is usually obtained by considering first a self-adjoint extension of the positive symmetric differential operator

$$-\frac{1}{2m}\sum_{j=1}^{3}(\partial_j - ieA_j(\mathbf{x},t))^2$$

acting on functions with compact support and then switching on the potential term $-eA_0(\mathbf{x},t)$ as a perturbation.[1] The first step is always possible because of the Friedrich theorem about extensions of positive symmetric operators. But the second one requires that the nonpositive part of $A_0(\mathbf{x},t)$ belongs to $L^\infty(\mathbb{R}^3) + L^2(\mathbb{R}^3)$ for any time t, and this is not always the case. Therefore, for a wide class of differentiable electromagnetic gauge fields, this standard mechanism, due to Kato, to give a quantum sense to Eq. (1.1), does not work. Moreover, if a differentiable gauge field for which such a mechanism works is given, then by a simple change of gauge we can obtain gauge potentials representing the same gauge field for which the mechanism does not work. However, quantum mechanics must be gauge invariant. Furthermore, if the particle and the electromagnetic gauge field are not defined in $\mathbb{R}^4$, but in a more general space-time of the type $M\times\mathbb{R}$, M being a three-dimensional spacelike manifold, the equation corresponding to (1.1) has only a local sense[2] when the gauge field is defined in a nontrivial fiber bundle $P(M\times\mathbb{R}, U(1))$. This difficulty can be avoided for the part corresponding to the first term of (1.2) by considering the wave functions, not as ordinary functions, but as cross sections of the bundle associated to $P(M\times\{0\}, U(1))$ by means of the natural action of $U(1)$ on $\mathbb{C}$, which gives an intrinsic formulation to this first term. On the contrary, the perturbation term has only a local sense and we do not know how to apply the standard approach in such a case.

The aim of this paper is to try to avoid all these problems by giving a new geometrical sense to the quantum evolution.

This geometrical sense was suggested to us by the similarity existing between the expression for the time-evolution operator of a quantum system with a bounded time-differentiable Hamiltonian $H(t)$ given by

$$U(t) = P\exp\left\{-i\int_0^t H(t')dt'\right\} = I + \sum_{k=1}^{\infty}(-i)^k\int_0^t dt_1\int_0^{t_1}\cdots\int_0^{t_{k-1}}dt_k\, H(t_1)\cdots H(t_k) \tag{1.3}$$

and that of the parallel transport operator in $\mathbb{R}^n\times\mathbb{R}^m$ along a curve $\gamma:[0,t]\to\mathbb{R}^n$ with respect to a connection Γ in $\mathbb{R}^n\times GL(m,\mathbb{R})$, which is given by[3,4]

$$\tau_\gamma^t = P\exp\left\{-\int_0^t \Gamma_\mu dx^\mu(\gamma)\right\} = I + \sum_{k=1}^{\infty}\int_0^t dt_1\int_0^{t_1}\cdots\int_0^{t_{k-1}}dt_k\Gamma_{\mu_1}(t_1)\cdots \times\Gamma_{\mu_k}(t_k)\dot\gamma^{\mu_1}(t_1)\cdots\dot\gamma^{\mu_k}(t_k), \tag{1.4}$$

where $\sigma^*(\omega) = \Gamma_\mu dx^\mu$, ω being the 1-form of the connection Γ, and $\sigma:\mathbb{R}^4\to\mathbb{R}^4\times GL(m,\mathbb{R})$ being the section $\sigma(x) = (x,I)$.

This analogy suggested to us that we could interpret quantum evolution of a system with bounded Hamiltonian as parallel transport along time. One can hope to exploit this interpretation in order to generalize the geometrical setting of quantum evolution in a way that makes possible the physical description of the evolution of a quantum particle in a four-dimensional gauge field. This is possible, as we will show in this paper. We will find a geometrical approach to quantum mechanics in the fiber bundle language, where the quantum evolution is interpreted as the parallel transport with respect to a generalized connection, and this permits us to define the quantum evolution of a particle in a gauge field in that setting. Moreover, this leads to a deeper understanding of the meaning of the evolution postulate.

The structure of the paper is as follows. In Sec. 2 and 3 we set up the mathematical tools necessary for our purposes. We define the concept of generalized connection. In Sec. 4 we state the geometrical formulation of time evolution. Section 5 is devoted to displaying in this geometrical setting two special systems: a nonrelativistic particle in a scalar potential (case a) and a nonrelativistic particle in a four-dimensional gauge field (case b). In the latter case, we show explicitly that although there is no canonical identification between the state spaces at different times, it is possible to define the dynamics intrinsically.

II. DIFFERENTIABILITY ON UNITARY GROUPS

Let $\mathscr{H}$ be a complex separable Hilbert space, and $\mathscr{B}(\mathscr{H})$ be the Banach space of bounded operators in $\mathscr{H}$. The set of bounded invertible operators in $\mathscr{H}$ with bounded inverse, $\mathscr{L}(\mathscr{H}) = \{A\in\mathscr{B}(\mathscr{H}); A^{-1}\in\mathscr{B}(\mathscr{H})\}$, is an open Banach submanifold of $\mathscr{B}(\mathscr{H})$. Furthermore $\mathscr{L}(\mathscr{H})$ endowed with the usual composition law, becomes an infinite-dimensional Lie group. Its Lie algebra is $\mathscr{B}(\mathscr{H})$ where the Lie bracket product is defined by $[A,B] = i(AB - BA)$.

The set $\mathscr{A}_{\mathscr{B}}$ of bounded self-adjoint operators in $\mathscr{H}$, $\mathscr{A}_{\mathscr{B}} = \{A\in\mathscr{B}(\mathscr{H}); A^* = A\}$, is a Lie subalgebra of $\mathscr{B}(\mathscr{H})$ which generates the unitary group $\mathscr{U}(\mathscr{H})$ through the exponential map. Therefore the unitary group $\mathscr{U}(\mathscr{H})$ is a closed Lie subgroup of $\mathscr{L}(\mathscr{H})$.[5]

Now, since the curves described in $\mathscr{U}(\mathscr{H})$ by evolution operators in quantum mechanics are not always differentiable, we must consider another topological structure in the unitary group which is more relevant for quantum mechanics. It is the s-topology whose open sets are generated by the sets

$$B_{U,\epsilon}^{x_1,\dots,x_n} = \{U'\in\mathscr{U}(\mathscr{H}); \|Ux_i - U'x_i\| < \epsilon, \quad i = 1,\dots,n\},$$

with $U\in\mathscr{U}(\mathscr{H})$, $\epsilon > 0$ and $x_i\in\mathscr{H}$. If $\mathscr{H}$ is finite-dimensional, the s-topology and norm topology coincide. The unitary group $\mathscr{U}(\mathscr{H})$ is also a topological group when endowed with this topology, because of the joint continuity of the group operation. Notice that we could define an s-topology in $\mathscr{B}(\mathscr{H})$ and $\mathscr{L}(\mathscr{H})$ in a similar way, but in such cases the composition law is not jointly continuous; therefore $\mathscr{L}(\mathscr{H})$ would not be an s-topological group.

Now we are going to introduce a concept of s^+-differentiable curve in $\mathscr{U}(\mathscr{H})$ which is compatible with the s-topology.

Definition: An s-continuous curve $U:[0,\infty)\to\mathscr{U}(\mathscr{H})$ is s^+-differentiable if $\mathscr{D}(U) = \{x\in\mathscr{H}; \forall t\in[0,\infty),$

$\exists s\text{-}\lim_{\Delta\to 0^+} \Delta^{-1}[U(t+\Delta)x - U(t)x]\}$, is a dense subset in $\mathscr{H}$.

In $\mathscr{D}(U)$ we define the operator $D^+U(t)$ by

$$D^+U(t)x = s-\lim_{\Delta\to 0^+} \Delta^{-1}[U(t+\Delta)x - U(t)x].$$

The group $\mathscr{U}(\mathscr{H})$ acts on the set of curves in $\mathscr{H}$ by left (right) translation. The set of s^+-differentiable curves is invariant under this action.

Proposition: Let $U(t)$ be an s^+-differentiable curve in $\mathscr{U}(\mathscr{H})$. For every $t\in[0,\infty)$, the operator $A(t)$ defined in $\mathscr{H}$ by $iA(t) = U^*(t)D^+U(t)$, is an essentially self-adjoint operator.

Proof: Let x be a point in $\mathscr{D}(U)$. Then,

$$\lim_{\Delta\to 0^+} \Delta^{-1}\{\|U(t+\Delta)x\|^2 - \|U(t)x\|^2\}$$

$$= \lim_{\Delta\to 0^+} \Delta^{-1}\{\|x\|^2 - \|x\|^2\} = 0,$$

and therefore $(D^+U(t)x, U(t)x) + (U(t)x, D^+U(t)x) = 0$; so that, $(Ax,x) = (x,Ax)$.

This relation holds for any $x\in\mathscr{D}(U)$, which implies that A is a symmetric operator.

Now we must proof that the only vector $u\in\mathscr{D}(A^*)$ satisfying $A^*u = \pm iu$ is the zero vector in $\mathscr{H}$. This proof may be carried out following standard proofs for the particular case of $U(t)$ being a semigroup of unitary operators (see, e.g., Ref. 6).

According to this proposition any s^+-differentiable curve in $\mathscr{U}(\mathscr{H})$ defines for each $t\geqslant 0$ a self-adjoint operator in $\mathscr{H}$, the closure of $A(t)$, which will also be denoted by $A(t)$.

Definition: Two s^+-differentiable curves U_1 and U_2 in $\mathscr{U}(\mathscr{H})$ are said to be tangent at $U_0\in\mathscr{U}(\mathscr{H})$ if there exist $t_1,t_2\in[0,\infty)$ such that $U_1(t_1) = U_2(t_2) = U_0$ and $A_1(t_1) = A_2(t_2)$.

This definition gives an equivalence relation in the set of all the s^+-differentiable curves passing through U_0. We define the s-tangent space of $\mathscr{U}(\mathscr{H})$ at U_0 as the set of the equivalence classes (tangent vectors) in that relation. Then, there is a bijective canonical correspondence between the s-tangent space at any point U_0 of $\mathscr{U}(\mathscr{H})$ and the set $\mathscr{A}(\mathscr{H})$ of self-adjoint operators in $\mathscr{H}$. The onto character of this relation is obvious, because if A is a self-adjoint operator, U_0e^{itA} is an s^+-differentiable curve whose s-tangent vector at U_0 corresponds to A.

The left (right) translation of the curves in $\mathscr{U}(\mathscr{H})$ by the left (right) product by any element U of $\mathscr{U}(\mathscr{H})$ induces an isomorphism between the tangent space of $\mathscr{U}(\mathscr{H})$ at any point $U_0\in\mathscr{U}(\mathscr{H})$ and the tangent space of $\mathscr{U}(\mathscr{H})$ at U_0U. Since the tangent spaces of $\mathscr{U}(\mathscr{H})$ at U_0 and at U_0U are canonically identified with $\mathscr{A}(\mathscr{H})$, this isomorphism induces a transformation in $\mathscr{A}(\mathscr{H})$. It is easy to show that this transformation does not depend on U_0. Hence, $\mathscr{U}(\mathscr{H})$ also acts canonically on $\mathscr{A}(\mathscr{H})$ on the left (right). It is easy to prove that this left action is trivial, because the field of s-tangent vectors of $\mathscr{U}(\mathscr{H})$ corresponding to the same element of $\mathscr{A}(\mathscr{H})$ is invariant under left translations. In the same way we can show that the right action coincides with the action of $\mathscr{U}(\mathscr{H})$ on $\mathscr{A}(\mathscr{H})$ defined by

$$\mathrm{ad}(U^*)A = U^*AU$$

for any $U\in\mathscr{U}(\mathscr{H})$, $A\in\mathscr{A}(\mathscr{H})$.

Accordingly, $\mathscr{A}(\mathscr{H})$ has some of the properties of a Lie algebra of a Lie group, i.e., it is a kind of "Lie space" of the topological group $\mathscr{U}(\mathscr{H})$.

III. UNITARY PARALLEL TRANSPORT AND GENERALIZED CONNECTIONS

The norm and strong topologies defined in $\mathscr{U}(\mathscr{H})$ give rise to two different concepts of topological principal fiber bundle with structural group $\mathscr{U}(\mathscr{H})$ and base space $\mathbb{R}^+$, the set of nonnegative real numbers. Since the s-topology of $\mathscr{U}(\mathscr{H})$ is the relevant one for quantum mechanics, we have to consider principal fiber bundles with this topology. However, $\mathscr{U}(\mathscr{H})$ does not have a differentiable structure compatible with the s-topology, and, the problem is how to define a concept of connection in such a fiber bundle. Since there is no differentiable structure in $\mathscr{P}(\mathbb{R}^+,\mathscr{U}(\mathscr{H}))$ there are neither tangent spaces nor horizontal spaces and consequently the only way to introduce a connection is by means of that of parallel transport. The base space being contractible and one-dimensional, $\mathscr{P}$ is trivial and the parallel transport is path-independent and can be formulated in terms of cross sections.

Definition: A generalized connection Θ in $\mathscr{P}$ is a family $\{\sigma_\alpha\}_{\alpha\in I}$ of continuous sections of $\mathscr{P}$ such that

(i) for every $u_0\in\mathscr{P}$ there exist one $\delta\in I$ such that $\sigma_\delta(t_0)=u_0$, where $t_0=\pi_{\mathscr{P}}(u_0)$.

(ii) given any pair $\alpha,\beta\in I$, there exists one $U\in\mathscr{U}(\mathscr{H})$ with $\sigma_\alpha(t)=\sigma_\beta(t)U$ for any $t\geqslant 0$.

Notice that condition (ii) implies that the family $\{\sigma_\alpha\}_{\alpha\in I}$ is fully determined by any one section of the family.

Definition: Two cross sections σ and σ' of $\mathscr{P}$ are said to be s^+-differentiable equivalent if the curve U in $\mathscr{U}(\mathscr{H})$ defined by $\sigma'(t)=\sigma(t)U(t)$, is s^+-differentiable.

The cross sections σ_α of the family Θ are s^+-differentiable equivalent and will be said to be parallel with respect to Θ.

Let $\rho=\mathscr{U}(\mathscr{H})\times V\to V$ be a continuous left action of $\mathscr{U}(\mathscr{H})$ on a topological space V and $\mathscr{E}(\mathbb{R}^+,V,\rho)$ the corresponding associated fiber bundle. For every σ_α in the family Θ and any $v\in V$, we can define a section ψ_v^α in $\mathscr{E}$ by $\psi_v^\alpha(t)=[\sigma_\alpha(t),v]$, where the bracket means the element of $\mathscr{E}$ corresponding to the equivalence class of $(\sigma_\alpha(t),v)$. Then, it is easy to check that $\psi_v^\beta=\psi^\alpha_{\rho(U,v)}$ when σ_β and σ_α are related by $\sigma_\beta(t)=\sigma_\alpha(t)U$. Thus, the set $\{\psi_v^\alpha\}_{v\in V}$ does not depend on α and the sections of such a set will be said to be parallel sections in $\mathscr{E}$ with respect to Θ. Note that for any $\phi_{t_0}\in\pi_{\mathscr{E}}^{-1}(t_0)$ there is one parallel section $\psi(t)$ with $\psi(t_0)=\phi_{t_0}$.

Definition: A cross section σ of $\mathscr{P}$ is s^+-differentiable with respect to Θ if σ is s^+-differentiable equivalent to the parallel cross sections σ_α.

To every cross section σ in $\mathscr{P}$ s^+-differentiable with respect to Θ we can associated an $\mathscr{A}$-valued 1-form ω_σ of $\mathbb{R}^+$ by means of the expression

$$\omega_\sigma\left(\frac{\partial}{\partial t}\right)=A(t), \tag{3.1}$$

where $A(t)$ is the self-adjoint operator corresponding to the tangent vector at $\sigma(t)$ to the curve $U(s)$ in $\mathscr{U}(\mathscr{H})$ such that $\sigma(s)=\sigma_\alpha(s)U(s)$, σ_α being the only parallel section of Θ with $\sigma_\alpha(t)=\sigma(t)$. The 1-form ω_σ is said to be the connection 1-form of Θ in the cross section σ.

Note that Θ does not define a connection 1-form in $\mathscr{P}$ but only a 1-form ω_σ on $\mathbb{R}^+$ for every cross section σ in $\mathscr{P}$.

IV. THE GEOMETRICAL SETTING FOR QUANTUM MECHANICS

According to the first postulate of quantum mechanics, every state of a quantum system at a time t is described by a ray of a Hilbert space $\mathscr{H}_t$. We will assume that the correspondence physical states–rays of $\mathscr{H}_t$ is bijective, i.e., we will neglect the possibility of the existence of supeselection rules, for the sake of simplicity. On the other hand, the postulate referring to quantum evolution is not so clearly stated but it contains implicitly two points: the first one is the assumption that there is a canonical identification of the Hilbert spaces corresponding to different times; the second one that the evolution is described by the Schrödinger equation, namely, it is possible to choose at every time t a vector representative $\psi(t)$ of the ray such that with the above identification the differential equation describing the evolution is $i\partial\psi/\partial t=H(t)\psi(t)$, where $H(t)$ is the Hamiltonian operator. Notice that in the particular case of $H:\mathbb{R}^+\to\mathscr{A}_{\mathscr{B}}$ being a differentiable function, the expression (1.3) is meaningful and the solution of the evolution equation is given by $\psi(t)=U(t)\psi(0)$.

This way of introducing the evolution postulate deserves some comments: first of all, if the Hilbert spaces corresponding to different times are really different, the space of states in not a Hilbert space but a Hilbert bundle which seems to be a Hilbert space because of the trivialization provided by the "canonical" identification of the different fibers. Such identification is needed in order to compare vectors at different times (e.g, to define $\partial\psi/\partial t$). Now, it is well known from elementary differential geometry that vectors in different fibers of a vector bundle can be intrinsically compared by means of a parallel transport, i.e., by means of a connection. This way of comparing vectors at different times is intrinsic and has no need of a "canonical" identification between fibers. For instance, expressions like $\partial\psi/\partial t$ can be replaced by covariant derivatives of sections along the time. In this setting the space of states will not be a Hilbert space but a Hilbert bundle and the evolution of a state will be described by a cross section in such a Hilbert bundle, rather than by a curve in a Hilbert space. Then, the evolution will become a parallel transport in the Hilbert bundle, which explains the formal analogy existing between (1.3) and (1.4).

Accordingly, we propose to formulate the postulates of quantum mechanics as follows:

Postulate: A quantum system is described by means of a differentiable Hilbert bundle $\mathscr{E}(\mathbb{R}^+,\mathscr{H})$ with base $\mathbb{R}^+$, typical fiber $\mathscr{H}$, and Hermitian structure h. At every time $t\geqslant 0$, the state of the system is described by a ray in the fiber $\pi_{\mathscr{E}}^{-1}(t)$.

Next, we proceed to build up a principal fiber bundle $\mathscr{P}(\mathbb{R}^+,\mathscr{U}(\mathscr{H}))$ for which $\mathscr{E}(\mathbb{R}^+,\mathscr{H})$ is an associated Hilbert bundle. Let $\mathscr{P}_t$ denote the set of orthonormal bases of $\pi_{\mathscr{E}}^{-1}(t)$, and $\mathscr{P}=\bigcup_{t\geqslant 0}\mathscr{P}_t$ be the disjoint union of sets $\mathscr{P}_t$. Once an arbitrary but fixed orthonormal basis $\{e_i\}_{i\in\mathbb{N}}$ of $\mathscr{H}$ is given, we can define a right action of the unitary group $\mathscr{U}(\mathscr{H})$ on $\mathscr{P}$ as follows. For any orthonormal basis $u=\{u_i\}_{i\in\mathbb{N}}$ of $\pi_{\mathscr{E}}^{-1}(t)$ and every $U\in\mathscr{U}(\mathscr{H})$, we define the orthonormal basis $v=uU$ by

$$v_i=\sum_{j=1}^{\infty}(e_j,Ue_i)u_j.$$

The right action of $\mathscr{U}(\mathscr{H})$ on $\mathscr{P}$ endowes it with an s-principal fiber bundle structure $\mathscr{P}(\mathbb{R}^+,\mathscr{U}(\mathscr{H}))$. Every cross-section of $\mathscr{P}$ will be called a gauge of the quantum system. When the natural action of $\mathscr{U}(\mathscr{H})$ on $\mathscr{H}$ is considered, it is easy to see that the fiber bundle associated to $\mathscr{P}$ is $\mathscr{E}$.

The left adjoint action of $\mathscr{U}(\mathscr{H})$ on $\mathscr{B}(\mathscr{H})$ endowed with the s-topology generates a vector bundle $\mathscr{B}(\mathbb{R}^+,\mathscr{B}(\mathscr{H}))$ associated to $\mathscr{P}$. Every element $R\in\mathscr{B}$ can be considered as a bounded operator of the Hilbert space $\pi_{\mathscr{E}}^{-1}(\pi_{\mathscr{B}}(R))$ of the Hilbert bundle $\mathscr{E}$ as follows. Let u be any point in $\pi_{\mathscr{P}}^{-1}(\pi_{\mathscr{B}}(R))$ and let B be the element of $\mathscr{B}(\mathscr{H})$ such that $R=[u,B]_{\mathscr{B}}$. For any $\xi\in\pi_{\mathscr{E}}^{-1}(\pi_{\mathscr{B}}(R))$ there exists a $\psi\in\mathscr{H}$ with $\xi=[u,\psi]_{\mathscr{E}}$. Then, we define $R\xi=[u,B\psi]_{\mathscr{E}}$.

Postulate: Each observable of the quantum system is described by a cross section ρ of the fiber bundle $\mathscr{B}(\mathbb{R}^+,\mathscr{B}(\mathscr{H}))$ such that $\rho(t)$ is a bounded self-adjoint operator in $\pi_{\mathscr{E}}^{-1}(t)$ for any $t\geqslant 0$.

Notice that for unbounded observables we can consider at each time $t\geqslant 0$ their imaginary exponentials, which are bounded unitary operators, and hence describe them by continuous sections of $\mathscr{B}$ too.

The translation to this new frame of the postulates concerning the correspondence rules, expressed in terms of probability, between the mathematical model and experimental facts is straightforward.

As we are considering the absence of superselection rules, there is at every time t a complete system of commuting observables. For the more general case with superselection rules see, e.g., Ref. (7) and references therein. A choice of such a complete system provides a "basis" of physical states at every time t. This gives an identification of states spaces at different times up to the choice of relative phases on the vectors representing the basic states. The observables of the complete system become constant after this identification. Next, we formalize these facts in the geometrical framework we are proposing.

As it is well known, the existence of a complete system of commuting observables is equivalent to the existence of a maximal abelian subalgebra of observables.

A maximal abelian algebra Σ of observables of the system is described by a family $\{\rho_i\}_{i\in J}$ of sections of $\mathscr{B}$ such that for every $t\geqslant 0$, $\{\rho_i(t)\}_{i\in J}$ is a maximal abelian algebra of bounded normal operators in $\pi_{\mathscr{E}}^{-1}(t)$.

Definition: A cross section σ of $\mathscr{P}$ is said to be a gauge tied to Σ if for any $\rho_i\in\Sigma$ there exists a bounded normal operator $R_i\in\mathscr{B}(\mathscr{H})$ such that for any $t\geqslant 0$, $[\sigma(t),R_i]=\rho_i(t)$, i.e., the sections ρ_i are constant in the trivialization of $\mathscr{B}$ generated by σ.

Every maximal abelian subalgebra of observables has gauges tied to it. However, this gauge tied to Σ is not unique. In fact, if σ is such a gauge, then the section $\bar\sigma$ of $\mathscr{P}$, defined by

$$\bar\sigma(t)=\sigma(t)U, \tag{4.1}$$

U being any operator in $\mathscr{U}(\mathscr{H})$, is also a gauge tied to Σ. Moreover, two gauges tied to Σ may be not differentiable-equivalent. Indeed, if f is any continuous, nondifferentiable real function of $\mathbb{R}^+$ and σ is a gauge tied to Σ, the gauge σ' defined by

$$\sigma'(t)=\sigma(t)e^{if(t)} \tag{4.2}$$

is also tied to Σ, and is not differentiable-equivalent to σ. This fact is a consequence of the projective character of the quantum states, and it points out that the identification of states spaces at different times by means of a gauge tied to Σ is not canonical, as is assumed in standard formulation of quantum mechanics.

We shall now consider the dynamics of the system. It describes the evolution of the states from time $t=0$ to time $t>0$ in a continuous and unitary way. More precisely, the time evolution of the system will be given by a sheaf of nonintersecting continuous sections in $\mathscr{E}$ which preserves linearity and the Hermitian structure of $\mathscr{E}$, and such that reach any point of $\mathscr{E}$. In our framework the evolution postulate can be stated as follows.

Postulate: The evolution of the quantum system is given by a generalized connection Θ in $\mathscr{P}$. The evolution of a state ψ_0 of $\pi_{\mathscr{E}}^{-1}(0)$ is given by the parallel cross section $\psi(t)$ of $\mathscr{E}$ with $\psi(0)=\psi_0$.

Following the standard formulation of quantum mechanics, we assume that any maximal abelian algebra of observables of the system has tied to it a gauge differentiable with respect to Θ. In particular, this implies that for any two maximal abelian algebras of observables Σ,Σ' there exist two gauges tied to Σ and Σ', respectively, which are differentiable equivalent.

Let us consider a fixed gauge σ tied to a maximal abelian algebra of observables Σ which is differentiable with respect to Θ.

Proposition: Let U be the differentiable curve in $\mathscr{U}(\mathscr{H})$ given by $\sigma(t)=\sigma(t)U(t)$ where σ_0 is the parallel cross section of $\mathscr{P}$ with $\sigma_0(0)=\sigma(0)$. Then U is the solution of

$$D^+U(t)=-iA(t)U(t) \tag{4.3}$$

satisfying $U(0)=I$.

Proof: Let $\sigma_t(s)$ be the parallel section of Θ with $\sigma_t(t)=\sigma(t)$ and V_t be the curve in $\mathscr{U}(\mathscr{H})$ defined by $\sigma(s)=\sigma_t(s)V_t(s)$. Then, $\omega_\sigma(\partial/\partial t)=iA(t)=V_t^*(t)D^+V_t(t)$. Since σ_t and σ_0 are parallel, $\sigma_0(s)=\sigma_t(s)V_t(0)$. Therefore, $V_t(s)U(s)=V_t(0)$ and

$$V_t(D^+U)+(D^+V_t)U=0.$$

Hence, $iA(t)=-D^+U(t)U^*(t)$, from which (4.3) follows.

Any section ξ of $\mathscr{E}$ has associated with it a continuous curve $\psi_\xi(t)$ in $\mathscr{H}$ in such a way that $\xi(t)=[\sigma(t),\psi_\xi(t)]$. When ξ is parallel with respect to Θ, we have

$$\begin{aligned}\xi(t)=[\sigma_0(t),\psi_\xi(0)]&=[\sigma(t)U(t),\psi_\xi(0)]\\&=[\sigma(t),U(t)\psi_\xi(0)],\end{aligned}$$

U and σ_0 being as in the proposition above. Then, $\psi_\xi(t)=U(t)\psi_\xi(0)$ and if $\psi_\xi(0)\in\mathscr{D}(U)$, from (4.3) it follows that

$$D^+\psi_\xi(t)=-iA(t)\psi_\xi(t). \tag{4.4}$$

This equation is the expression in the gauge σ of the parallel character of section ξ and the differentiable operator $\nabla_{\partial/\partial t}=D^++iA(t)$ can be considered as the covariant derivative operator associated to Θ in the gauge σ. Notice that equation (4.4) can be re-written as

$$\nabla_{\partial/\partial t}\psi_\xi(t)=(D^++iA(t))\psi_\xi(t)=0. \tag{4.5}$$

This interpretation gives a geometrical meaning to Eq. (4.4), which is the differential evolution equation of the system in the gauge σ, the Hamiltonian in that gauge being $A(t)$.

It follows from (3.2) that the transformation law for the Hamiltonian $A(t)$ in a differentiable change of gauge is given by

$$\overline{A}(t) = V^*(t)A(t)V(t) - iV^*D^+V(t) \tag{4.6}$$

when σ is changed to $\bar{\sigma}$ with $\bar{\sigma}(t) = \sigma(t)V(t)$. Thus, although the evolution is gauge independent, the explicit form of the Hamiltonian is strongly dependent on the choice of the differentiable gauge.

If we choose a gauge σ_S tied to a maximal abelian algebra of bounded observables which contains the observables of position in the classical configuration space, we give the description of the evolution corresponding to the Schrödinger representation. In such a case σ_S is said to be a Schrödinger gauge; Eq. (4.4) in this case is the Schrödinger equation.

The Heisenberg representation for Θ corresponds to the choice of a parallel gauge σ_H with respect to Θ (Heisenberg gauge). Notice that in the Heisenberg gauge the Hamiltonian is the null operator. Given a fixed generalized connection Θ_0 in $\mathscr{P}$, we define the interaction gauge with respect to Θ_0 by the choice of a section σ_I in $\mathscr{P}$ parallel with respect to Θ_0. The Hamiltonian in the Schrödinger gauge H_s is related to the Hamiltonian in the interaction gauge H_I by

$$H_I(t) = V_0^*(t)H_s(t)V_0(t) - iV_0^*(t)D^+V_0(t), \tag{4.7}$$

where V_0 is the curve in $\mathscr{U}(\mathscr{H})$ defined by $\sigma_I(t) = \sigma_S(t)V_0(t)$. The second term in the right-hand side of (4.7) being the Hamiltonian of Θ_0 in the Schrödinger gauge H_0, we obtain the usual equality

$$H_I(t) = V_0^*(t)W(t)V_0(t),$$

with $W = H_s - H_0$.

Let $\rho(t)$ be any bounded observable of the system. For any gauge σ there is a curve $R(t)$ in $\mathscr{B}(\mathscr{H})$ such that $[\sigma(t),R(t)] = \rho(t)$ for any $t \geqslant 0$. The curves R and $\overline{R}$ corresponding to ρ in two differentiable equivalent sections σ, $\bar{\sigma}$ are related by the expression

$$\overline{R}(t) = V^*(t)R(t)V(t), \tag{4.8}$$

where $\bar{\sigma}(t) = \sigma(t)$.

It is noteworthy that the transformation law for the Hamiltonian (4.6) does not reduce to that of observables (4.8) but it contains an additional term displaying the connection feature of Θ. This fact also occurs in classical mechanics, where the Hamiltonians transform in an inhomogeneous way under time-dependent canonical transformations, while observables transform homogeneously.

The concept of conservative system must be stated in a precise way because of the dependence of the Hamiltonian on the gauge. We will say that a system is conservative when there exists a Schrödinger gauge σ_S, where the Hamiltonian H_s is time independent.

The description of the evolution of mixed states could also be stated in this geometrical framework, by making use of the bundle associated to $\mathscr{P}$ by the left adjoint action of $\mathscr{U}(\mathscr{H})$ in the set of trace-class positive operators in $\mathscr{H}$.

V. APPLICATIONS

A. A nonrelativistic particle in a scalar potential

This is a very familiar system in quantum mechanics. In the standard framework of quantum mechanics in the coordinate representation the states of the system are described by rays of $L^2(\mathbb{R}^3)$ and time evolution by the rays' trajectories of the solutions of Schrödinger equation

$$(-\tfrac{1}{2}\Delta + \mathscr{V}(\mathbf{x},t))\psi(\mathbf{x},t) = i\frac{\partial\psi(\mathbf{x},t)}{\partial t} \tag{5.1}$$

where $\mathscr{V}$ is the potential. For simplicity, the mass of the particle is taken to be the unit.

The meaning of Eq. (5.1) is the following. If $\mathscr{V}$ is a real function such that $\mathscr{V}(\mathbf{x},t) \in L^2(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$ for every $t \geqslant 0$, then the Kato–Rellich theorem[1] implies that the symmetric operator

$$H(t) = -1/2\Delta + \mathscr{V}(x,t)$$

is essentially self-adjoint in $C_0^\infty(\mathbb{R})$. Thus, $H(t)$ has a unique self-adjoint extension $A(t)$ in a domain dense in $L^2(\mathbb{R}^3)$ independent of t. By a solution $\psi(\mathbf{x},t)$ of (5.1) we mean that the curve ξ of $L^2(\mathbb{R}^3)$ with $\xi(t) = \psi(\mathbf{x},t) \in L^2(\mathbb{R}^3)$ satisfy

$$iD^+\xi = A\xi. \tag{5.2}$$

Now, assuming that $\mathscr{V}$ is a continuously differentiable curve in $L^2(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$ Kato's theorem[1] assures the existence of an s^+-differentiable curve U in $\mathscr{U}(L^2(\mathbb{R}^3))$ satisfying

$$D^+U = -iAU, \tag{5.3}$$

with $U(0) = I$. This curve furnishes the complete solution of the evolution problem of the system; the time evolution of any initial state $\xi_0 \in L^2(\mathbb{R}^3)$ at $t = 0$ is given by $\xi(t) = U(t)\xi_0$.

We can consider the system in our geometrical approach as follows. The bundle of physical states $\mathscr{E}$ is $\mathbb{R}^+ \times L^2(\mathbb{R}^3)$ endowed with the constant Hermitian $L^2(\mathbb{R}^3)$ product. Therefore the corresponding fiber bundles $\mathscr{P}(\mathbb{R}^+, \mathscr{U}(L^2(\mathbb{R})))$ and $\mathscr{B}(\mathbb{R}^+, \mathscr{B}(L^2(\mathbb{R}^3)))$ are built up from $\mathscr{E}$ as indicated in Sec. 4. To every bounded observable $B(t)$ in $L^2(\mathbb{R}^3)$ we associate at each time t the operator $B(t)$ acting on $\mathscr{E}_t = \{t\} \times L^2(\mathbb{R}^3)$, which defines a section ρ_B in $\mathscr{B}(\mathbb{R}^+, L^2(\mathbb{R}^3))$. The section ρ_B is the representation in our framework of the observable $B(t)$. Once a basis $\{e_i\}_{i\in\mathbb{N}}$ in $L^2(\mathbb{R}^3)$ is fixed, there is a trivialization of $\mathscr{P}$ induced by the cross section of $\mathscr{P}$ defined by $\sigma(t) = (t, \{e_i\}_{i\in\mathbb{N}})$. This cross section is a gauge tied to the position observables. In fact, for any bounded observable ρ_B that in the standard formulation of quantum mechanics is time independent, there exists a bounded operator R in $L^2(\mathbb{R}^3)$ such that $[\sigma(t),R] = \rho_B(t)$. Hence any position observable is constant in the trivialization induced by the section σ. The same result holds for observables depending only on the linear momentum $\mathbf{p}$, the unique self-adjoint extension of the essentially self-adjoint operator $-i\nabla$ defined in $C_0^\infty(\mathbb{R}^3)$. This gauge σ corresponds to the Schrödinger representation of the system, but it is not unique. If we change σ to $\sigma' = \sigma U_0$, U_0 being a fixed unitary operator, the position and momentum observables remain as constant sections in the new trivialization of $\mathscr{B}$. For instance, we can change from the Schrödinger coordinate representation to the Schrödinger momentum representation by the Fourier transformation, i.e., to take U_0 as the unitary operator of $L^2(\mathbb{R}^3)$ defined by

$$(U_0\psi)(\mathbf{x}) = (2\pi)^{-3/2}\int d\mathbf{y}\, e^{-i\mathbf{x}\cdot\mathbf{y}}\psi(y)$$

for any $\psi\in L^2(\mathbb{R}^3)$. This new gauge is also tied to the position and momentum observables. We only change the multiplicative representation of position observables to a differential representation and vice versa for the momentum observables.

The dynamics of the system is given by the generalized connection Θ in $\mathscr{P}$ whose parallel cross sections are of the form

$$\sigma_\alpha(t) = (t,\{U(t)U_\alpha e_i\}_{i\in\mathbb{N}})$$

for any $t>0$, where $U(t)$ is the solution of Eq. (5.3) with $U(0)=I$, and U_α is an arbitrary element of $\mathscr{U}(L^2(\mathbb{R}^3))$. Notice that there are a lot of gauges tied to the position observables which are not differentiable with respect to Θ. For the usual change of gauge σ to $\sigma' = \sigma e^{-ict}$, the expression of the Hamiltonian is changed to $A+c$. For a slightly more general change $\bar\sigma = \sigma e^{-i\mathscr{W}(\mathbf{x},t)}$ the new Hamiltonian is obtained by changing $\mathscr{V}$ to $\mathscr{V}+\mathscr{W}$ i.e., the classical potential acting on the particle is only seen in the Hamiltonian in the gauge tied to the maximal abelian algebra of position observables in its spectral representation.

It is worthwhile to note that we started by considering the standard formulation in the coordinate representation. This explains why in our framework the position observables play a relevant role in the gauge σ. Had we started with another representation, then the corresponding gauge would have been associated to another maximal abelian algebra of observables.

B. A particle in a gauge field

Although it is not necessary to use the geometrical approach which we are describing in order to understand the meaning of the physical systems considered in the case (a), we will present another interesting case for which understanding this geometrical approach is essential.

Indeed, as we remarked in Sec. 1, the dynamics of a nonrelativistic quantum particle moving in the space time $\mathbb{R}^4$ under the action of a four-dimensional gauge field cannot be described in the standard framework of quantum mechanics. Moreover, when the space-time is nonflat or its topology is not trivial the operator corresponding to (1.2) has only a local sense.[2] How shall we define the dynamics in such a case? We are going to show how this can be done in the geometrical setting proposed above for quantum dynamics.

Let $\mathscr{M} = M\times\mathbb{R}$ be a space-time manifold, M being a connected orientable Riemannian manifold. We shall assume M to be compact for technical reasons, but this is not essential. Let $P(M,G)$ be the principal fiber bundle where the classical gauge field acting on the particle lies. Let Γ be a connection of P associated to this gauge field. G is usually taken to be a connected, compact, simple Lie group. Let $E(M,\alpha,\mathbb{C}^n)$ be the Hermitian bundle associated to an n-dimensional unitary representation α of G in the space $\mathbb{C}^n$ of internal degrees of freedom of the particle. We will consider the evolution of the system from an initial time t_0, and we take $t_0=0$ for simplicity. For each $t\geqslant 0$, $\pi_P^{-1}(M\times\{t\})$ is a principal fiber bundle with base M and structural group G, and $\pi_E^{-1}(M\times\{t\})$ is its associated vector bundle by means of α. Let $\mathscr{E}_t$ denote the set of sections ξ of $\pi_E^{-1}(M\times\{t\})$ such that

$$\int_M d\mu(\mathbf{x})h(\xi(\mathbf{x}),\xi(\mathbf{x})) < \infty,$$

where $d\mu$ is the Riemannian measure of M and h is the product in the Hermitian structure of E. It is trivial to show that $\mathscr{E}_t$ is a Hilbert space with the inner product defined by

$$h(\xi,\eta) = \int_M h(\xi(\mathbf{x}),\eta(\mathbf{x}))d\mu(x)$$

for any $\xi,\eta\in\mathscr{E}_t$. Let us define $\mathscr{E} = \cup_{t>0}\mathscr{E}_t$. For any given connection Γ_0 in P, the parallel transport with respect to Γ_0 in E induces a Hilbert isomorphism between $\mathscr{E}_0$ and $\mathscr{E}_t$ for any $t\geqslant 0$. These isomorphisms generate a vector bundle structure in $\mathscr{E}$, with typical fiber $\mathscr{E}_0$ and base $\mathbb{R}^+$, which does not depend on the choice of Γ_0. Therefore, as was shown in Sec. 4, once a basis $e=\{e_i\}_{i\in\mathbb{N}}$ in $\mathscr{E}_0$ is chosen there exists a principal bundle $\mathscr{P}(\mathbb{R}^+,\mathscr{U}(\mathscr{E}_0))$ such that $\mathscr{E}$ is the associated vector bundle of $\mathscr{P}$ by the natural action of $\mathscr{U}(\mathscr{E}_0)$ on $\mathscr{E}_0$.

If we choose another element $u=\{u_i\}_{i\in\mathbb{N}}$ in $\pi_P^{-1}(0)$, the isomorphisms between the fibers of $\mathscr{E}$ generated by the parallel transport with respect to Γ_0 define a cross section σ_0 in $\mathscr{P}$ as follows. Let $u_i(t)$ be the element of $\mathscr{E}_t$ corresponding to the parallel transport of u_i along the curves $c_{\mathbf{x},t}:[0,t]\to M\times\mathbb{R}^+$ with $c_{\mathbf{x},t}(s) = (\mathbf{x},s)$. Then, for any $t\geqslant 0$, $\sigma_0(t)$ is defined by $\sigma_0(t) = u(t) = \{u_i(t)\}_{i\in\mathbb{N}}$.

On the other hand, each $f\in C(M)$ has associated with a continuous cross section ρ_f of the vector bundle $\mathscr{B}(\mathbb{R}^+,\mathscr{B}(\mathscr{E}_0))$ in such a way that for each $t\in\mathbb{R}^+$, $\rho_f(t)$ is the self-adjoint operator of $\mathscr{E}_t$ defined by

$$(\rho_f\xi)(\mathbf{x}) = f(\mathbf{x})\xi(\mathbf{x})$$

for any $\xi\in\mathscr{E}_t$. The family $\{\rho_f\}_{f\in C(M)}$ is a commutative algebra Σ_S, the Schrödinger algebra of the continuous position observables of the system.

Since the parallel transport preserves linearity in E, the constant multiplicative self-adjoint operator $\rho_f(0)$ of $\mathscr{E}_0$ corresponds to $\rho_f(t)$ for any $t\geqslant 0$ in the trivialization of $\mathscr{B}$ induced by the σ_0 with $\sigma_0(0)=e$, $\rho_f(t) = [\sigma_0(t),\rho_f(0)]_{\mathscr{B}}$. Thus, σ_0 is a gauge tied to Σ_S. The cross section σ_0 depends on u and Γ_0. Notice that since there is no priviledged Γ_0 in $\mathscr{P}$ the trivialization of $\mathscr{P}$ is not so canonical as in case (a). Even if we take P and Γ_0 to be trivial, there are a lot of connections in P satisfying that condition and defining different gauges in $\mathscr{P}$.

In order to define the dynamics of the particles under the action of the gauge field defined by connection Γ, we construct a connection Γ_1 in P as follows. Let H_u^0, H_u be the horizontal subspaces with respect to Γ_0 and Γ, respectively, of the tangent space to $P(M\times\mathbb{R},G)$ at u. Then, we can define the splittings

$$H_u^0 = M_u^0 + T_u^0,\quad H_u = M_u + T_u,$$

of H_u^0, H_u in such a way that

$\pi_{P*}(M_u^0) = \pi_{P*}(M_u) = T_{\pi_P(u)}(M)$ and

$\pi_{P*}(T_u^0) = \pi_{P*}(T_u) = T_{\pi_P(u)}(\mathbb{R})$. Let us define Γ_1 as the connection whose horizontal subspaces are given by

$$H_u^1 = M_u^0 + T_u. \tag{5.4}$$

We construct the cross section σ_1 in $\mathscr{P}$ associated to Γ_1 with $\sigma_1(0) = \sigma_0(0)$ in the same way as above for σ_0 and Γ_0. We are now going to consider the dynamics in this gauge σ_1. The parallel transport with respect to Γ_1 along the curves $c_{\mathbf{x},t}$ establishes the isomorphism τ_t of $P_t = \pi_P^{-1}(M \times \{t\})$ in P_0. Let Γ_t be the connection images in P_0 through τ_t of the restriction of Γ to P_t. The family $\{\Gamma_t\}_{t\in\mathbb{R}^+}$ describes a differentiable curve in the space of sections of the bundle $\Lambda^1(P_0)\times\mathscr{g}$, $\mathscr{g}$ being the Lie algebra of G.

Let Δ_t be the elliptic differential operator defined in the space of differentiable sections in E_0 in such a way that, given a local chart (θ,ϕ) of M and a cross section $\beta:\theta\rightarrow P_0$,

$$(\Delta_t\xi)_\beta = g^{jk}(\partial_j - iA_j)(\partial_k - iA_k)\xi_\beta - g^{jk}\Gamma_{jk}^m(\partial_m - iA_m)\xi_\beta \tag{5.5}$$

for any differentiable section ξ of E_0, where g^{jk} is the local expression in (θ,ϕ) of the Riemannian metric in T^{*m}, Γ_{jk}^m are the corresponding Chistoffel symbols, ξ_β and $(\Delta_t\xi)_\beta$ are the functions of $\varphi(\theta)$ in $\mathbb{C}^n$ corresponding to the sections ξ and $\Delta_t\xi$ in the trivialization of E_0 induced by β, and

$$A_j = i\alpha_*\left[\omega_t\left((\beta\cdot\varphi^{-1})_*\left(\frac{\partial}{\partial x^j}\right)\right)\right],$$

ω_t being the 1-form of connection Γ_t and α_* the Lie homomorphism of $\mathscr{g}$ in $M_n(\mathbb{C})$ induced by $\alpha: G\rightarrow \mathrm{GL}(n;\mathbb{C})$.

Now, since $-\Delta_t$ is a symmetric positive operator and M is compact, it has one self-adjoint extension, which will be denoted by $-\Delta_t$ too.

Proposition: There exist an s^+-differentiable curve U in $\mathscr{U}(\mathscr{E}_0)$ such that

$$D^+U(t) = (i/2)\Delta_t U(t) \tag{5.6}$$

for any $t\geqslant 0$ and $U(0) = 1$.

Proof: This is a consequence of a Kato's theorem.[8] Since M is compact the domain of $-\frac{1}{2}\Delta_t$ is t-independent. Because $-\Delta_t$ is self-adjoint, 1 is in the resolvent set of $(i/2)\Delta_t$ for any $t\geqslant 0$, and since $\{\Gamma_t\}_{t\in\mathbb{R}^+}$ is a differential curve in the space of sections $\Lambda^1(P_0)\otimes\mathscr{g}$, for each given $s\in\mathbb{R}^+$ the bounded operator $B_s(t) = (I-(i/2)\Delta_t)(I-(i/2)\Delta_s)^{-1}$ is strongly continuously differentiable for any $t\geqslant 0$. Therefore, by Kato's theorem, there exists an s^+-differentiable curve V in $\mathscr{U}(\mathscr{E}_0)$ satisfying (5.6) with $U(0) = I$.

We define the dynamics of the particle with mass 1 under the action of Γ by the generalized connection $\Theta = \{\sigma_Z\}_{Z\in\mathscr{U}(\mathscr{E}_0)}$ in $\mathscr{P}$, σ_Z being the cross section $\mathscr{P}$ with $\sigma_Z(t) = \sigma_1(t)U(t)Z$. From this definition it follows that the cross sections σ_Z are differentiable equivalent to σ_1, which is a gauge tied to the commutative algebra Σ_S. This gauge corresponds to a temporal gauge formulation of the dynamics of the particle in the gauge field Γ, because the Hamiltonian in this gauge is $-(1/2)\Delta_t$. Fixing $u\in\pi_{\mathscr{P}}^{-1}(0)$ the generalized connection Θ depends only on Γ. In fact, had we considered another reference connection $\overline{\Gamma_0}$ in P, the isomorphisms $\bar\tau_t$ would be the same as τ_t, because the parallel transports along the curves $c_{\mathbf{x},t}$ with respect to $\overline{\Gamma}$ and Γ coincide as a consequence of the fact that in the corresponding splitting (5.4) the temporal parts $\overline{T}_u$ and T_u do so. Then, $\Gamma_t = \overline{\Gamma}_t$ and $\sigma_1 = \bar\sigma_1$. Thus, $\Theta = \overline{\Theta}$.

Next, we will show that the gauges σ_1 and σ_0 are differentiable equivalent. This permits us to display the dynamics in the gauge σ_0 too.

Proposition: The gauges σ_1 and σ_0 are differentiable equivalent.

Proof: Let $\{\xi_i^t\}_{i\in\mathbb{N}}$ and $\{\eta_i^t\}_{i\in\mathbb{N}}$ be the bases of $\pi_{\mathscr{E}}^{-1}(t)$ given by $\sigma_0(t) = \{\xi_i^t\}_{i\in\mathbb{N}}$ and $\sigma_1(t) = \{\eta_i^t\}_{i\in\mathbb{N}}$. Then, by the definition of σ_0 and σ_1, $\xi_i^t(\mathbf{x}) = \tau_{c_{\mathbf{x},t}}^0\xi_i(\mathbf{x})$ and $\eta_i^t(\mathbf{x}) = \tau_{c_{\mathbf{x},t}}^1\eta_i(\mathbf{x})$, where $\tau_{c_{\mathbf{x},t}}^i$ $(i = 0,1)$ is the parallel transport map of $\pi_E^{-1}(0)$ in $\pi_E^{-1}(t)$ with respect to Γ_i $(i = 0,1)$ along the curves $c_{\mathbf{x},t}:[0,t]\rightarrow M\times\mathbb{R}^+$. Therefore, we have

$$\xi_i^t(\mathbf{x}) = \tau_{c_{\mathbf{x},t}}^0(\tau_{c_{\mathbf{x},t}}^1)^{-1}\eta_i^t(\mathbf{x}). \tag{5.7}$$

Let $W_t:\mathscr{E}_0\rightarrow\mathscr{E}_0$ be the operator defined by

$$(W_t\xi)(\mathbf{x}) = (\tau_{c_{\mathbf{x},t}}^1)^{-1}(\tau_{c_{\mathbf{x},t}}^0)\xi(\mathbf{x})$$

for every $\xi\in\mathscr{E}_0$. Since α is a unitary representation, the parallel transports $\tau_{c_{\mathbf{x},t}}^0$ and $\tau_{c_{\mathbf{x},t}}^1$ preserve the Hermitian structure h of E. Thus,

$$h(\xi,W_t\eta) = \int d\mu(\mathbf{x})h(\xi(\mathbf{x}),(\tau_{c_{\mathbf{x},t}}^1)^{-1}\tau_{c_{\mathbf{x},t}}^0\eta(\mathbf{x}))$$

$$= \int d\mu(\mathbf{x})h(\xi(\mathbf{x})\cdot\eta(\mathbf{x})) = h(\xi,\eta)$$

i.e., $W_t\in\mathscr{U}(\mathscr{E}_0)$. Furthermore, the curve W with $W(t) = W_t$ is s^+-differentiable in $\mathscr{U}(\mathscr{E}_0)$ and

$$D^+W(t) = iW(t)A(t),$$

where $A(t):\mathscr{E}_0\rightarrow\mathscr{E}_0$ is the self-adjoint operator defined as follows. Let ω_0 and ω be the 1-forms of connections Γ_0 and Γ, respectively, and $\widetilde{X}$ any vector field in P with $\pi_{P*}(\widetilde{X}_{u(\mathbf{x},t)}) = \partial/\partial t$ for any $u(\mathbf{x},t)$ in $\pi_P^{-1}(\mathbf{x},t)$. Then, $A(t)$ is defined by

$$(A(t)\xi)(\mathbf{x}) = \left[u(\mathbf{x},0), -i\alpha_*((\omega-\omega_0)(\widetilde{X}_{\tau_{c_{\mathbf{x},t}}^0 u(\mathbf{x},0)}))\psi(\mathbf{x})\right]_{E_0} \tag{5.8}$$

for any $\xi\in\mathscr{E}_0$ with $\xi(\mathbf{x}) = [u(\mathbf{x}),0),\psi(\mathbf{x})]_{E_0}$. Notice that in (5.8) we have used the fact that $\omega(\widetilde{X}) = \omega_1(\widetilde{X})$, because of the splitting (5.4). This proves the proposition, because from (5.7) and (5.8) it follows that

$$\sigma_0(t) = \sigma_1(t)W(t)$$

for any $t\geqslant 0$.

According to the proposition above, σ_0 is differentiable with respect to Θ. The Hamiltonian in this gauge $H^0(t)$ is obtained from the transformation law (4.6),

$$H^0(t) = W(t)^*(-1/2\Delta_t)W(t) - iW(t)^*D^+W(t), \tag{5.9}$$

i.e.,

$$H^0(t) = -\tfrac{1}{2}\Delta_t^0 + A(t), \tag{5.10}$$

where Δ_t^0 is the elliptic differential operator defined by (5.5) from the connection Γ_t^0 of P_0 obtained by parallel transport with respect to Γ_0 of the restriction of Γ to P_t.

Let (θ,φ) be any local chart of M and $\beta:\theta\rightarrow P_0$ be any local cross section of P_0. Then, for any differential section ξ of E_0 with $\xi(x) = [\beta(\mathbf{x}),\psi(\mathbf{x})]_{E_0}$ and support in θ,

$$(H^0(t)\xi)_\beta(\psi(\mathbf{x})) = -\tfrac{1}{2}(\Delta_t^0\xi)_\beta(\varphi(\mathbf{x})) - i\alpha_* \times \left\{(\omega-\omega_0)(\widetilde{X}_{\tau^0_{c_{\mathbf{x},t}}\beta(\mathbf{x})})\right\}\psi(\mathbf{x}). \quad (5.11)$$

It is worth remarking that the expression (5.11) defines a self-adjoint extension of the operator corresponding to (1.2) in the case of M being compact, when Γ_0 is trivial and β is a global section of P parallel with respect to Γ_0. This fact guarantees that the evolution we have defined in a slightly abstract way corresponds to the desired dynamics for the system.

The gauge σ_0 in $\mathscr{P}$ is a temporal gauge iff $(\omega-\omega_0)\tilde{x}=0$ for any vector field $\widetilde{X}$ of P with $\pi_P(\widetilde{X}_t)=\partial/\partial t$. Furthermore, for any temporal gauges σ, $\bar{\sigma}$ with $\sigma(0)=\bar{\sigma}(0)$, we have

$$\overline{H}(t) = U(t)^* H(t) U(t) = H(t),$$

U being the differentiable curve of $\mathscr{U}(\mathscr{E}_0)$ with $\bar{\sigma}(t)=\sigma(t)U(t)$, because $\bar{\tau}_{\mathbf{x},t}(\tau_{\mathbf{x},t})^{-1} = I_{\pi_E^{-1}(\mathbf{x},t)}$ for any $\mathbf{x}\in M$.

In consequence, since there is no natural gauge as in case (a), this points out how the geometric description we are proposing for the quantum systems is the natural and intrinsic setting for the study of the dynamics of such a system (b). We could recover the quantum dynamics in a standard setting by fixing a gauge σ. But without such a choice of σ it is not possible to formulate the dynamics in the standard setting.

However, one could argue that we have assumed M to be compact and therefore we have not solved the problem in the general case, for instance, the dynamics of the particles under a gauge field defined in the whole space-time $\mathbb{R}^4$. We have restricted ourselves to the compact case in order to have a global dynamics, i.e., a global solution of (5.6). But, in the general case, we can also build a self-adjoint extension $H^0(t)$ of the operator (1.2) in the same way. Indeed, if we define Δ_t as in (5.6) in the space of sections of E_0 with support compact, there exists a Friedrich self-adjoint extension of $-\Delta_t$. Then by making use of (5.9) we find the wanted self-adjoint Hamiltonian. However, the differential evolution, equation corresponding to the Hamiltonian $-\frac{1}{2}\Delta_t$ is not always integrable. In consequence, in the noncompact case, the evolution of the system always cannot be interpreted as a parallel transport. Only systems with globally defined evolution have a physical meaning. Therefore, there is a restriction on the four-dimensional gauge fields which give rise to reasonable quantum systems. This restriction is gauge independent, because if the evolution equation is not integrable in one gauge neither is it integrable in any other gauge. However, most of the four-dimensional gauge fields on noncompact manifolds give rise to globally defined evolutions and in this case these evolutions are to be understood as parallel transports. In fact, the set of such fields is dense in the set of all gauge fields.

Finally, notice that more general dynamical systems, such as classical Hamiltonian systems or differential dynamical systems, could be studied in a similar geometrical setting, by taking as structural groups the groups of simplectomorphisms and the group of diffeomorphisms, respectively. The differential structure of those groups is stronger than that of $\mathscr{U}(\mathscr{H})$.[9] However, in such cases the corresponding descriptions of their evolutions for noncompact manifolds do not cover all physical situations. For instance, the case of a dynamical system given by a complete vector field which is not complete at each time, can not be considered in that way.

ACKNOWLEDGMENTS

We thank Professor L. J. Boya for critical reading of the manuscript.

[1]M. Reed and B. Simon, *Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[2]T.T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[3]H. G. Loos, J. Math. Phys. **8**, 2114 (1967).
[4]I. Ya Aref'eva, Theor. Math. Phys. **43**, 353 (1980).
[5]S. Lang, *Introduction to Differentiable Manifolds* (Interscience, New York, 1972).
[6]A. V., Balakrishnan, *Applied Functional Analysis* (Springer, New York, 1976).
[7]A. Galindo, A. Morales, and R. Nuñez-Lagos, J. Math. Phys. **3**, 324 (1962).
[8]T. Kato, J. Math. Soc. Japan, **5**, 208 (1953).
[9]H. Omori, *Infinite Dimensional Lie Transformation Groups, Lecture Notes in Mathematics*, Vol. **427** (Springer, New York, 1974).

Feynman path integrals of operator-valued maps

K. R. Parthasarathy and K. B. Sinha
Indian Statistical Institute, Delhi Centre, New Delhi 110016, India

(Received 17 December 1980; accepted for publication 1 May 1981)

Feynman integral is defined for operator-valued maps which are Fourier transforms of an operator-valued measure of bounded variation. Such an integral is then used to describe perturbation of certain unitary groups by certain cocycles.

PACS numbers: 03.65.Db

I. INTRODUCTION

In the 1940s, Feynman proposed a new formalism of quantum mechanics, in contrast to the well-used canonical or Hamiltonian formalism. According to this formalism, the probability amplitudes of mutually exclusive events should be added (unlike the addition of probabilities themselves in classical probability theory) and the probability amplitude associated with each of the possible "paths" connecting the space–time points $(x,0)$ to (y,t) is given by

$$\Phi(\gamma)\equiv\exp[(i/\hbar)S(\gamma)], \tag{1}$$

where $S(\gamma)$ is the classical expression for the action functional corresponding to the path γ and $\hbar$ is $(2\pi)^{-1}$ times Planck's constant. In the simplest of situations when the particle is moving in one dimension under the influence of a conservative force, $S(\gamma)=\int_t^0[(1/2m)\dot{\gamma}(s)^2-V(\gamma(s))]ds$, where m is the mass of the particle and $\gamma(0)=x$, $\gamma(t)=y$. Combining these two basic hypotheses, one arrives at Feynman's expression for $K(y,t;x)$ the probability amplitude for the particle to move from x at time 0 to y at time t, called the propagator:

$$K(y,t;x)=\int_{\substack{\gamma(0)=x\\ \gamma(t)=y}}\Phi(\gamma)\mathscr{D}\gamma. \tag{2}$$

The above integral symbol attempts to "sum" the contributions for each of the paths γ subject to the restrictions $\gamma(0)=x$ and $\gamma(t)=y$. In the following, we shall set $\hbar=1$ for simplicity of discussion. For a detailed discussion of applications of the above ideas to various physical problems, the reader is referred to Ref. 1.

One of the earlier attempts to understand expression (2) mathematically is due to Nelson.[2] One considers a heat equation

$$\frac{\partial f}{\partial t}=D\Delta f-iVf,\quad f(0)=f_0, \tag{3}$$

with $D=i/2m$, $\operatorname{Im} m>0$. Then (3) can be solved using the Wiener integral with the complex parameter D. It can be shown that the limit of the integral as $\operatorname{Im} m\to 0+$ exists for almost all $\operatorname{Re} m$, and one defines the result as the "Feynman path integral." However, the limit cannot be given as an integral with respect to a signed measure on the space of Brownian paths.[3]

II. FEYNMAN INTEGRALS

Here we shall employ the method of Fresnel integrals on infinite dimensional Hilbert spaces as done by Ito[4] and Albeverio and Hoegh-Krohn.[5] First, we note the well known result that as an improper Riemann integral on $\mathbb{R}^k$ one has

$$(2\pi i)^{-k/2}\int\exp\Big(-\frac{i}{2}\,\|x\|^2\Big)d^kx=1.$$

Such integrals are called Fresnel integrals after Fresnel, who used these study diffraction patterns in light scattering. Then it is an easy computation to show that for every $\Phi\in\mathscr{S}(\mathbf{R}^k)$ (the Schwartz space)

$$\int\exp\Big(\frac{-i}{2}\,\|x\|^2\Big)\tilde{\phi}(x)dx=(i)^{k/2}\int\exp\Big(\frac{i}{2}\,\|y\|^2\Big)\phi(y)dy,$$

where $\tilde{\phi}$ is the Fourier transform of ϕ. Note that $\tilde{\phi}\in\mathscr{S}(\mathbf{R}^k)$.

Definition: Let $\mathscr{H}$ be a real separable Hilbert space. Then $f:\mathscr{H}\to\mathbb{C}$ is said to be Fresnel integrable if $f(x)=\int\exp(-i(x,y))\,d\mu(y)$, where $\mu\in\mathscr{M}(\mathscr{H})\equiv$the set of all bounded complex measures on $\mathscr{H}$. We denote the class of all Fresnel integrable functions by $\mathscr{F}(\mathscr{H})$ and define "Feynman integral" F as a map: $\mathscr{F}(\mathscr{H})\to\mathbb{C}$ by

$$F(f)\equiv\int\exp\Big(-\frac{i}{2}\,\|x\|^2\Big)d\mu(x)\quad\forall f\in\mathscr{F}(\mathscr{H}). \tag{4}$$

If f is normed by the variation norm of the corresponding μ, then $\mathscr{F}(\mathscr{H})$ is an abelian Banach $*$-algebra (under pointwise multiplication) and $f\rightleftarrows\mu$ correspondence is 1–1. Then it is clear that F defined by (4) is a linear bounded functional on $\mathscr{F}(\mathscr{H})$.

For the definition of Feynman path integrals, we now need to equip the space of paths with Hilbert structure. Let $\mathscr{H}\equiv\{\gamma:\mathbb{R}^+\to\mathbb{R}$ locally absolutely continuous $|\int_0^\infty\dot{\gamma}(s)^2ds<\infty,\ \gamma(0)=0\}$. Then $\mathscr{H}$ is a reproducing Kernel Hilbert space with reproducing kernel G such that $\langle G(\sigma,\cdot),\gamma\rangle=\gamma(\sigma)$ for all $\sigma\in\mathbb{R}^+$ and all $\gamma\in\mathscr{H}$, $G(\sigma,\tau)$ is the Green's function for the differential operator $-d^2/ds^2$ on $\mathbb{R}^+$ with boundary conditions $(dG/d\tau)(\sigma,\infty)=G(\sigma,0)=0$. Then $(dG/d\tau)(\sigma,\tau)=\theta(\sigma-\tau)$.

Remark: We have taken one-dimensional space for convenience and three-dimensional space can be easily accommodated. However, if the configuration space is more general, say a C^∞ manifold (e.g., for a particle moving under some constraints), then the above construction fails.

So we have made the space of classical paths into a Hilbert space $\mathscr{H}$. Now a quantum mechanical particle lives in a Hilbert space of its own [e.g., $L^2(\mathbb{R})$ for a single particle is one-dimensional without spin]. Let this Hilbert space be denoted by $\mathfrak{H}$, and let A be the self-adjoint generator of a one-parameter strongly continuous unitary group U_α, with its

0022-2488/82/081459-04$02.50

spectral family $\{E_\lambda\}$. Define for $0\leqslant s,\tau$

$$J_{s,\tau}(\gamma)\equiv U_{\gamma(\tau)-\gamma(s)} = \int \exp\{-i[\gamma(\tau)-\gamma(s)]\lambda\}E(d\lambda). \tag{5}$$

Next theorem summarizes the properties of $J_{\tau,s}$.

Theorem 1: (i) For every fixed $\tau,s,J_{\tau,s}$ is a $\mathfrak{H}$-unitary operator valued function $\mathscr{H}$. (ii) For a fixed $\gamma\in\mathscr{H},J_{\tau,s}(\gamma)$ is a propagator, i.e., for $0\leqslant s\leqslant t\leqslant u$,

$$J_{s,t}(\gamma)J_{t,u}(\gamma)=J_{s,u}(\gamma). \tag{6}$$

(iii) For every $f,g\in\mathfrak{H}$, $\langle f,J_{\tau,s}(\cdot)g\rangle$ is Fresnel integrable.
(iv) The Feynman path integral of $J_{\tau,s}$ is well defined and

$$F(J_{s,t})=\exp[-(i/2)(t-s)A^2]. \tag{7}$$

Proof: (i) and (ii) are obvious from definition (5). For fixed τ,s, define a map $\pi_{s,\tau}:\mathbb{R}\rightarrow\mathscr{H}$ by

$$\pi_{s,\tau}(\lambda)=\lambda\,[G(\tau,\cdot)-G(s,\cdot)]. \tag{8}$$

$\pi_{s,\tau}$ maps $\mathbb{R}$ into a certain one-dimensional subspace of $\mathscr{H}$ and is a Borel map. Setting $\mu_{f,g}(\Delta)\equiv\langle f,E(\Delta)g\rangle$ for every Borel set $\Delta\subseteq\mathbb{R}$, we find

$$\langle f,J_{s,\tau}(\gamma)g\rangle = \int \exp\{-i[\gamma(\tau)-\gamma(s)]\lambda\}\,\mu_{f,g}(d\lambda)$$
$$= \int \exp[-i(\gamma,\gamma')]\,\mu_{f,g}\circ\pi_{s,\tau}^{-1}(d\gamma'),$$

where we have used the fact that $(\gamma,\pi_{s,\tau}(\lambda))=\lambda\,(\gamma,[G(\tau,\cdot)-G(s,\cdot)])=\lambda\,[\gamma(\tau)-\gamma(s)]$. Clearly the $\mu_{f,g}\circ\pi_{s,\tau}^{-1}\in\mathscr{M}(\mathscr{H})$, and its total variation is equal to that of $\mu_{f,g}$ which is majorized by $\|f\|\,\|g\|$. Thus $\langle f,J_{s,\tau}(\cdot)g\rangle\in\mathscr{F}(\mathscr{H})$ and we can compute its Feynman integral by the definition (4) to obtain

$$F\langle f,J_{s,t}g\rangle = \int \exp\Big(-\frac{i}{2}\|\gamma\|^2\Big)\mu_{f,g}\circ\pi_{s,t}^{-1}(d\gamma)$$
$$=\int \exp\Big(-\frac{i}{2}\lambda^2\|G(t,\cdot)-G(s,\cdot)\|^2\Big)\mu_{f,g}(\lambda)$$
$$=\int \exp\Big(-\frac{i}{2}\lambda^2|t-s|\Big)\langle f,E(d\lambda)g\rangle$$
$$=\Big\langle f,\exp\Big(-\frac{i}{2}A^2|t-s|\Big)g\Big\rangle.$$

Thus by Riesz' representation theorem, $F(J_{s,t})\in\mathscr{B}(\mathfrak{H})$ and

$$F(J_{s,t})=\exp\Big[-\Big(\frac{i}{2}\Big)|t-s|A^2\Big]. \qquad \square$$

As an example, we consider the translation group in one-dimension $U_\alpha=\exp(-iP\alpha)$ with P the momentum operator in $L^2(\mathbb{R})$. Then $F(\exp\{-i[\gamma(\tau)-\gamma(s)]P\})$ $=\exp[-(i/2)|\tau-s|P^2]=\exp(-iH_0|\tau-s|)$, which we identify as the free evolution operator with generator H_0. Thus, if $f\in L^2(\mathbb{R})$ is the state of a free particle at time 0, then its state f_τ at a later time τ is given as

$$f_\tau=\exp(-iH_0\tau)f=F(J_{\tau,0}f)=F(f(\cdot-\gamma(\tau)))$$

or $f_\tau(x)=F(f(x-\gamma(\tau)))$ for almost all (a.a.) x. This gives a mathematical meaning to the free evolution operator as a Feynman path integral.

Let $\mathscr{F}(\mathscr{H},\mathscr{B}(\mathfrak{H}))\equiv\{A:\mathscr{H}\rightarrow\mathscr{B}(\mathfrak{H})|A(\gamma)=\int\exp[-i(\gamma,\gamma')]E(d\gamma)$, variation norm of $\langle f,E(\cdot)g\rangle\leqslant M(E)\|f\|\,\|g\|$ for all $f,g\in\mathfrak{H}\}$. As before the correspondence, $A\rightleftarrows E$ is one-to-one. If the measure E depends on values of γ in the interval $[s,\tau]$, we say the corresponding A has support in $[s,\tau]$. It does not follow in general that $\mathscr{F}(\mathscr{H},\mathscr{B}(\mathfrak{H}))$ is an algebra, though it is clearly a normed linear space.

Lemma 2: Let A,B and AB belong to $\mathscr{F}(\mathscr{H},\mathscr{B}(\mathfrak{H}))$ and assume that supp$A$ and supp$B$ are disjoint. Then

$$F(AB)=F(A)F(B).$$

Proof: $(AB)(\gamma)=\int\exp[-i(\gamma,\gamma')]E_A*E_B(d\gamma')$, and, by hypothesis, E_A*E_B is of weak-bounded variation. Therefore by definition (4)

$$F(AB)=\int \exp\Big(-\frac{i}{2}\|\gamma\|^2\Big)E_A*E_B(d\gamma)$$
$$=\int \exp\Big(-\frac{i}{2}\|\gamma+\gamma'\|^2\Big)E_A(d\gamma)E_B(d\gamma')$$
$$=\int \exp\Big(-\frac{i}{2}\|\gamma\|^2\Big)E_A(d\gamma)\int \exp\Big(-\frac{i}{2}\|\gamma'\|^2\Big)E_B(d\gamma')$$
$$=F(A)F(B),$$

since $(\gamma,\gamma')=0$ for $\gamma\in\operatorname{supp}E_A$ and $\gamma'\in\operatorname{supp}E_B$.$\square$

This result corresponds to what one calls processes with independent increments in the theory of stochastic processes.

III. PERTURBATION OF A UNITARY GROUP

Next we introduce $\mathscr{B}(\mathfrak{H})$-values cocycles with respect to the unitary propagator $J_{t,s}$, defined in (5). For a fixed $\gamma\in\mathscr{H}$, let α be a map from $\mathbb{R}^+\times\mathbb{R}^+\rightarrow\mathscr{B}(\mathfrak{H})$ such that

$$\alpha(s,u)=\alpha(s,t)J_{s,t}\alpha(t,u)J_{s,t}^{-1},\quad 0\leqslant s\leqslant t\leqslant u<\infty,$$

and

$$\alpha(s,s)=I. \tag{9}$$

It is easy to see that, for any such α, $T_{s,t}\equiv\alpha(s,t)J_{s,t}$ is again a propagator, i.e., $T_{s,u}=T_{s,t}T_{t,u}$. For a given $\mathscr{B}(\mathfrak{H})$-valued function M on $\mathbb{R}^+$, if the differential equation

$$\frac{d\alpha}{dt}(s,t)=-i\alpha(s,t)J_{s,t}M(t)J_{s,t}^{-1} \tag{10}$$

with initial condition $\alpha(s,s)=I$, $s\leqslant t$, has a solution α, then that solution will certainly satisfy the relation (9). Such an α is called a $\mathscr{B}(\mathfrak{H})$-valued cocycle or just a cocycle with respect to $J_{s,t}$. Under various conditions on M, e.g., if M is norm-uniformly continuous, then (10) has a unique (unitary) solution, and this follows by essentially mimicking the Cauchy–Picard method for ordinary differential equation. The associated series is called the Dyson series in the physics literature, and for further reference on this point, we refer to Ref. 6. We note that any solution α of (10) depends on the path γ via J.

The next set of theorems show how cocycles can be used to generate new evolutions.

Theorem 3: Let $M(\tau)\equiv M$ in (10) be such that either (i) the pair $\{M,A\}$ form an irreducible imprimitivity system, or (ii) $M\in\mathscr{B}(\mathfrak{H})$. Then (10) has a unique solution in $\mathscr{B}(\mathfrak{H})$.

Remark: The hypothesis (i) of the above theorem means M is a selfadjoint operator with its spectral measure E^M with the property: $\exp(-iA\alpha)E^M(\Delta)\exp(iA\alpha)=E^M(\Delta+\alpha)\ \forall\alpha\in\mathbb{R}$. It is a result of Mackey[7] that such an imprimitivity pair can be represented as a Schrödinger pair, i.e., $\mathfrak{H}$ is isometrically isomorphic to $L^2(X)$ such that $(Mf)(x) = M(x)f(x)$ and $[\exp(-iA\alpha)f](x) = f(x-\alpha)$ for a.a. $x \in X$.

Proof: (i) In view of the preceding remark, Eq. (10) takes the following form in $L^2(X)$:

$$\frac{d\alpha(s,t)}{dt} = -i\alpha(s,t)M(\cdot - \gamma(s)) \quad \text{and} \quad \alpha(s,s) = I.$$

This can be easily solved to give $\alpha(s,t) = \exp[-i\int_s^t M(\cdot - \gamma(\tau) + \gamma(s))d\tau]$, as a bounded multiplication operator in $L^2(X)$. This is the so-called Feynman–Kac cocycle. The case (ii) has already been discussed and in this case the solution is given by the Dyson series.□

Theorem 4: Assume the hypotheses of Theorem 3. Assume furthermore that either (i) the function M in the remark is the Fourier transform of a bounded measure ν on X, or (ii) $M \in \mathscr{B}_2(\mathfrak{H})$, the class of Hilbert Schmidt operators in $\mathfrak{H}$. Then

(a) $T_{s,t} \equiv \alpha(s,t)J_{s,t} \in \mathscr{F}(\mathscr{H}, \mathscr{B}(\mathfrak{H}))$,

(b) $V_{s,t} \equiv F(T_{s,t})$

$$= \exp[-i(\tfrac{1}{2}A^2 + M)(t-s)], \quad 0 \leqslant s \leqslant t. \tag{11}$$

Proof: (i) Considering the expansion of $\alpha(t,s) = \exp[-i\int_s^t M(Q - \gamma(\tau) + \gamma(s))d\tau]$ [Q is the self-adjoint operator of multiplication by x in $L^2(X)$] and taking the first nontrivial term in $T_{s,t}$, we have

$$\int_s^t M(Q - \gamma(\tau) + \gamma(s))d\tau J_{s,t}$$
$$= \int_s^t d\tau \int d\nu(\beta) \exp\{-i[Q - \gamma(\tau) + \gamma(s)]\beta\}$$
$$\times \int \exp[-i(\gamma,\gamma')] E \circ \pi_{s,t}^{-1}(d\gamma').$$

The last expression can be written as a Fourier transform of the measure $d\tau d\nu(\beta)\exp(iQ\beta)E\circ\pi_{s,t}^{-1}(d\gamma')$ on $[s,t] \times X \times \mathscr{H}$. This measure is easily seen to be of weak bounded variation. Similarly one can consider other terms of the series and convince oneself that $\alpha(s,t)J_{s,t} \in \mathscr{F}(\mathscr{H}, \mathscr{B}(\mathfrak{H}))$.

(ii) In this case Dyson series gives the solution of α

$$\alpha(s,t) = I + \sum_{n=1}^{\infty} (-i)^n \int_s^t dt_1 \int_s^{t_1} dt_2 \cdots \int_s^{t_{n-1}} dt_n$$
$$\times M(s,t_1)M(s,t_2)\cdots M(s,t_n),$$

where we have written $M(s,t) \equiv J_{s,t}MJ_{s,t}^{-1}$ and we note that the series converges in norm. It is clear that $\alpha(s,t)J_{s,t}$ will be in $\mathscr{F}(\mathscr{H}, \mathscr{B}(\mathfrak{H}))$ if $E(\Delta_1)ME(\Delta_2)ME(\Delta_3)\cdots ME(\Delta_n)$ $= \Gamma(\Delta_1 \times \Delta_2 \times \cdots \times \Delta_n)$ for some spectral measure Γ on $\mathbb{R}^n$.

Since M is Hilbert–Schmidt, it admits the canonical decomposition

$$M = \sum_{j=1}^{\infty} \lambda_j \langle e_j, \cdot \rangle h_j \quad \text{with} \sum |\lambda_j|^2 < \infty.$$

Let $\mathfrak{H}^*$ be the antilinear dual of $\mathfrak{H}$ so that for $f^* \in \mathfrak{H}^*$ and $g \in \mathfrak{H}$ we have $f^*(g) = \langle f, g \rangle$. Then it is easy to see that $\mathscr{B}_2(\mathfrak{H})$ is in 1–1 correspondence with $\mathfrak{H}^* \otimes \mathfrak{H}$. In fact, $M = \Sigma \lambda_j e_j^* \otimes h_j \in \mathfrak{H}^* \otimes \mathfrak{H}$ since $\{e_j^* \otimes h_j\}$ is an orthonormal (o.n.) basis in $\mathfrak{H}^* \otimes \mathfrak{H}$ and $\Sigma |\lambda_j|^2 < \infty$. Define a spectral measure Γ on $\mathbb{R} \otimes \mathbb{R}$ taking values in $\mathfrak{H}^* \otimes \mathfrak{H}$ by $\Gamma(\Delta_1 \times \Delta_2) = E(\Delta_2)^* \otimes E(\Delta_1)$. Then

$$\langle \Gamma(\Delta_1 \times \Delta_2) f^* \otimes g, M \rangle_{\mathfrak{H}^* \otimes \mathfrak{H}}$$
$$= \sum \lambda_j \langle (E(\Delta_2)f)^* \otimes E(\Delta_1)g, e_j^* \otimes h_j \rangle$$
$$= \sum \lambda_j \langle e_j, E(\Delta_2)f \rangle \langle E(\Delta_1)g, h_j \rangle$$
$$= \langle g, E(\Delta_1)ME(\Delta_2)f \rangle.$$

Thus set $\Gamma(\Delta_1 \times \Delta_2 \times \cdots \times \Delta_n) = E(\Delta_n)^* \otimes E(\Delta_{n-1}) \otimes \cdots \otimes E(\Delta_2)^* \otimes E(\Delta_1)$ if n is even and $= E(\Delta_n) \otimes E(\Delta_{n-1})^* \otimes \cdots \otimes E(\Delta_2)^* \otimes P(\Delta_1)$, if n odd. If n is even, then

$$\langle \Gamma(\Delta_1 \times \Delta_2 \times \cdots \times \Delta_n) f^* \otimes M^* \otimes M^* \otimes \cdots \otimes g, M \otimes M \otimes \cdots \rangle$$
$$= \sum \lambda_{j_1} \lambda_{j_2} \cdots \lambda_{j_{n-1}} \langle e_{j_1}, E(\Delta_n) f \rangle$$
$$\times \langle e_{j_2}, E(\Delta_{n-1}) h_{j_1} \rangle \langle e_{j_3}, E(\Delta_{n-2}) h_{j_2} \rangle$$
$$\times \langle e_{j_{n-1}}, E(\Delta_2) h_{j_{n-2}} \rangle \langle g, E(\Delta_1) h_{j_{n-1}} \rangle$$
$$= \langle g, E(\Delta_1) E(\Delta_2) M \cdots ME(\Delta_n) f \rangle.$$

A similar calculation is valid for n odd. Clearly such a measure Γ on $\mathbb{R}_n$ is a spectral measure, taking values in a large Hilbert space $\mathfrak{H}^* \otimes \mathfrak{H} \otimes \mathfrak{H}^* \otimes \cdots$. Therefore, the measure

$$\underbrace{E(\cdot)ME(\cdot)\cdots ME(\cdot)}_{n \text{ times}}$$

is a measure of weak-bounded variation in $\mathfrak{H}$, and we have the result (a).

The Feynman integral of $T_{s,t}$ exists by (a) and defines a bounded operator in $\mathscr{B}(\mathfrak{H})$. From the cocycle property (9) of α it follows that $T_{t,s}$ is a propagator, i.e., $T_{s,u} = T_{s,t}T_{t,u}$ when $0 \leqslant s \leqslant t \leqslant u < \infty$. Since α is a solution of (10), it is obvious by the remarks following Eq. (10) that $\alpha(s,t)$ depends on the values of γ only in the interval $[s,t]$. Thus the support of $T_{t,u}$ and $T_{s,t}$ are disjoint, and hence, by Lemma 2, we have

$$V_{s,u} \equiv F(T_{s,u}) = F(T_{s,t}T_{t,u}) = F(T_{s,t})F(T_{t,u}) = V_{s,t}V_{t,u},$$

i.e., V is a propagator.

Next we compute the strong derivatives of $V_{s,t}$ w.r.t. t. Let $f \in D(A^2)$. Then it follows from the propagation property of V that

$$(1/h)(V_{s,t+h}f - V_{s,t}f) = V_{s,t}(1/h)(V_{t,t+h}f - f)$$
$$= V_{s,t}F[(1/h)(\alpha(t,t+h) - I)J_{t,t+h}f]$$
$$+ V_{s,t}(1/h)\{\exp[(-ih/2)A^2] - I\}f. \tag{12}$$

In the above, we have observed that $I \in \mathscr{F}(\mathscr{H}, \mathscr{B}(\mathfrak{H}))$ and $F(I) = I$, and also used (7). Clearly the second term in (12) converges strongly to $-(i/2)V_{s,t}A^2 f$ as $h \to 0$.

In either case (i) or case (ii), we expand $\alpha(t,t+h) - I$ by the appropriate Dyson series and conclude that the first term converges to $-iV_{s,t}Mf$ strongly as $h \to 0$, leading to the result (b).□

To conclude, we give the example of a single quantum mechanical particle without spin moving in one-dimensional space under the influence of a static potential $V(x)$. Its evolution is given by the Schrödinger equation: $f_t = \exp(-iHt)f$, where $H = P^2/2 + V$, which is a self-adjoint operator in $L^2(\mathbb{R})$ under a wide range of assumptions on $V(x)$. Since the pair $\{Q,P\}$ forms an irreducible imprimitivity system in $L^2(\mathbb{R})$, we can apply Theorems 3 and 4, if, furthermore, $V(x)$ satisfies $V(x) = \int \exp(-ix\beta)d\nu(\beta)$ for some ν of bounded variation. Thus we obtain

$$f_t \equiv \exp(-iHt)f$$
$$= F\left(\exp\left[-i\int_0^t V(Q-\gamma(\tau))d\tau\right]\exp[-i\gamma(t)P]\,f\right)$$
$$= F\left(\exp\left[-i\int_0^t V(Q-\gamma(\tau))d\tau\right]f(\cdot-\gamma(t))\right),$$

which is the Feynman-Ito formula.[8]

Remark: From the proof of Theorem 4 it is clear that its extension to any larger class of potentials will depend on how smooth is the Feynman path integral map F.

[1]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[2]E. Nelson, J. Math. Phys. **5**, 332 (1964).
[3]R. H. Cameron, J. Math. Phys. **39**, 126 (1960).
[4]K. Itô, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. of California Press, Berkeley, CA, 1961), Vol. 2, p. 227.
[5]S. Albeverio and R. Hoegh-Krohn, *Lecture Notes in Mathematics* 523 [Springer-Verlag, Berlin, 1976).
[6]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[7]G. W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[8]A. Truman, *Lectures Notes in Mathematics* 644 (Springer, Dublin, 1977).

Solutions of type D possessing a group with null orbits as contractions of the seven-parameter solution

Alberto García D. and Jerzy F. Plebański[a)]
Centro de Investigación y Estudios Avanzados del I.P.N., Apartado Postal 14-740, México 14, D. F., México

(Received 7 December 1981; accepted for publication 26 March 1982)

It is shown that several type D solutions with null group orbits of local isometries are limiting contractions of the seven-parameter solution.

PACS numbers: 04.20.Jb

The aim of this work is, first, to construct explicitly some limiting contractions of the seven-parameter solution of Plebański and Demiański[1] to the family of all null orbit divergenceless solutions, and second, to provide a specific answer to a question posed by Debever and McLenaghan in their recent paper[2] concerned with type D fields—is the Leroy null orbit solution[3] just a contraction of the seven-parameter solution?

The families of solutions studied are those in which the electromagnetic field, if present, has real eigenvectors aligned along the double Debever–Penrose (D–P) directions.

The starting point of the present paper is an alternative representation of the P–D solution which can be obtained by subjecting Eq. (3.30) of Ref. 1 to the transformation

$$d\tau \to d\tau + \frac{q^2}{Q}dq\,, \quad d\sigma \to d\sigma - \frac{1}{Q}dq\,,$$

which brings the seven-parameter metric to the form

$$ds^2 = \frac{1}{(1-pq)^2}\left\{\frac{p^2+q^2}{\mathscr{P}} + dp^2\frac{\mathscr{P}}{p^2+q^2}(d\tau + q^2 d\sigma)^2 - 2dq(d\tau - p^2 d\sigma) - \frac{Q}{p^2+q^2}(d\tau - p^2 d\sigma)^2\right\}, \quad (1)$$

where

$$\begin{aligned}\mathscr{P} &= (-\tfrac{1}{6}\lambda - g^2 + \gamma) + 2np - \varepsilon p^2 + 2mp^3 + (-\tfrac{1}{6}\lambda - e^2 - \gamma)p^4\,,\\ Q &= (-\tfrac{1}{6}\lambda + e^2 + \gamma) - 2mq + \varepsilon q^2 - 2nq^3 + (-\tfrac{1}{6}\lambda + g^2 - \gamma)q^4\,. \end{aligned}\quad (2)$$

The conformal curvature invariant,

$$\delta := \tfrac{2}{3}\overset{2}{C} = C^{AB}{}_{CD}\, C^{CD}{}_{AB} = (C^{(3)})^2 - \tfrac{4}{3}C^{(4)}C^{(2)} + \tfrac{1}{3}C^{(5)}C^{(1)}$$

in the null tetrad formalism and signature $(+\ +\ +\ -)$, and the 2-form of the electromagnetic field which characterize this solution, are

$$\delta = 4\left(\frac{1-pq}{q+ip}\right)^6\left\{-(m+in) + (e^2+g^2)\frac{1+pq}{q-ip}\right\}^2 \quad (3)$$

and

$$\omega = -d\left\{\frac{e+ig}{q+ip}(d\tau - ipqd\sigma)\right\}, \quad (4)$$

respectively.

By scaling the coordinates according to

$$p \to \epsilon^{-1}p\,, \quad q \to \epsilon^{-1}q\,, \quad \tau \to \epsilon\tau\,, \quad \sigma \to \epsilon^3\sigma\,,$$

and simultaneously adjusting the constants

$$e + ig \to \epsilon^{-2}(e+ig)\,, \quad m \to \epsilon^{-3}m\,, \quad \varepsilon \to \epsilon^{-2}\varepsilon\,,$$
$$n \to \epsilon^{-3}n\,, \quad \gamma \to \epsilon^{-4}\gamma + \tfrac{1}{6}\lambda\,, \quad \lambda \to \lambda\,,$$

and then taking in (1) the limit $\epsilon \to \infty$, one arrives at

$$ds^2 = \frac{p^2+q^2}{\mathscr{P}}dp^2 + \frac{\mathscr{P}}{p^2+q^2}(d\tau + q^2 d\sigma)^2 - 2dq(d\tau - p^2 d\sigma) - \frac{Q}{p^2+q^2}(d\tau - p^2 d\sigma)^2\,, \quad (5)$$

with

$$\begin{aligned}\mathscr{P} &= \gamma - g^2 + 2np - \varepsilon p^2 - \tfrac{1}{3}\lambda p^4\,,\\ Q &= \gamma + e^2 - 2mq + \varepsilon q^2 - \tfrac{1}{3}\lambda q^4\,,\\ \delta &= \frac{4}{(q+ip)^6}\left\{-(m+in) + (e^2+g^2)\frac{1}{q-ip}\right\}^2,\\ \omega &= -d\left\{\frac{e+ig}{q+ip}(d\tau - ipqd\sigma)\right\}. \end{aligned}\quad (6)$$

This solution, studied exhaustively in Ref. 4, is an alternative representation of the Carter solution.[5]

A limiting contraction of the above solution leads in turn to the "anti-NUT" branch of type D solutions (see Ref. 4, Secs. 10 and 11), which was shown in Ref. 6 (Sec. 5) to be equivalent to *all* divergenceless type D solutions with λ and with the electromagnetic field aligned along the double D–P vectors. In terms of the present representation of the solution via (5) and (6), the contraction consists in replacing

$$p \to p\,, \quad q \to q_0 + \epsilon q\,, \quad \sigma \to \epsilon^{-1}\sigma\,, \quad \tau \to \tau - q_0^2\epsilon^{-1}\sigma\,, \quad (7)$$

and adjusting the parameters according to

$$e + ig \to e + ig\,, \quad n \to n\,, \quad \lambda \to \lambda\,,$$
$$\varepsilon = \zeta_0 + 2\lambda q_0^2\,, \quad m = -\eta_0\epsilon + \zeta_0 q_0 + \tfrac{4}{3}\lambda q_0^3\,,$$
$$\gamma_0 = -2\eta_0 q_0\epsilon + \xi_0\epsilon^2 - e^2 + \zeta_0 q_0^2 + \lambda q_0^4\,,$$

($q_0, \xi_0, \eta_0, \zeta_0$ are arbitrary constants independent on ϵ), and then taking $\epsilon \to 0$. One obtains

$$ds^2 = \frac{p^2+q_0^2}{\mathscr{P}}dp^2 + \frac{\mathscr{P}}{p^2+q_0^2}(d\tau + 2q_0 q d\sigma)^2 + 2(p^2+q_0^2)dqd\sigma - (p^2+q_0^2)\mathscr{S}d\sigma^2\,, \quad (8)$$

where

$$\begin{aligned}\mathscr{P} &= -(e^2+g^2) + \zeta_0 q_0^2 + \lambda q_0^4 + 2np - (\zeta_0 + 2\lambda q_0^2)p^2 - \tfrac{1}{3}\lambda p^4\,,\\ \mathscr{S} &= \xi_0 + 2\eta_0 q + \zeta_0 q^2\,. \end{aligned}\quad (9)$$

However, in the present version of the anti-NUT family, one

[a)] On leave of absence from the University of Warsaw, Warsaw, Poland.

is permitted in particular to set $\xi_0 = \eta_0 = \zeta_0 = 0$, obtaining thus a five-parameter solution, the family of all divergenceless null orbit type D solutions,[7] given explicitly by

$$ds^2 = \frac{p^2 + q_0^2}{\mathscr{P}} dp^2 + \frac{\mathscr{P}}{p^2 + q_0^2} (d\tau + 2q_0 q d\sigma)^2 + 2(p^2 + q_0^2) dq d\sigma ,$$
$$\mathscr{P} = -(e^2 + g^2) + 2np - \lambda (\tfrac{1}{3} p^4 + 2q_0^2 p^2 - q_0^4), \quad (10)$$

with the nonvanishing curvature quantities and the electromagnetic field given by

$$C^{(3)} = \frac{2}{(p + iq_0)^3} \left\{ n + \tfrac{4}{3} i\lambda q_0^3 - \frac{e^2 + g^2}{p - iq_0} \right\},$$
$$R = -4\lambda ,$$
$$C_{12} = -\frac{e^2 + g^2}{(p^2 + q_0^2)^2} ,$$
$$\omega = -(e + ig) d \left\{ \frac{1}{p + iq_0} (i d\tau - (p - iq_0) q d\sigma) \right\}, \quad (11)$$

when referred to the null tetrad

$$\left.\begin{matrix} e^1 \\ e^2 \end{matrix}\right\} = \frac{1}{\sqrt{2}} \left\{ \left(\frac{\mathscr{P}}{p^2 + q_0^2} \right)^{1/2} (d\tau + 2q_0 q d\sigma) \pm i \left(\frac{\mathscr{P}}{p^2 + q_0^2} \right)^{-1/2} dp \right\},$$
$$e^3 = (p^2 + q_0^2) dq , \quad e^4 = d\sigma . \quad (12)$$

In order to bring the above solution into the canonical form of the nondiverging and nontwisting type D solutions, formulas (3.34)–(3.40) of Ref. 6, one replaces

$$q_0 \to l , \quad \sigma \to u , \quad q \to v , \quad \tau \to \sigma - luv , \quad (13)$$

thus arriving at

$$ds^2 = \frac{p^2 + l^2}{\mathscr{P}} dp^2 + \frac{\mathscr{P}}{p^2 + l^2} \times [d\sigma + l(v du - u dv)]^2 + 2(p^2 + l^2) du dv , \quad (14)$$

with

$$\mathscr{P} = -(e^2 + g^2) + 2np - \lambda (\tfrac{1}{3} p^4 + 2l^2 p^2 - l^4),$$
$$\delta = \frac{4}{(p + il)^6} \left\{ n + i\tfrac{4}{3}\lambda l^3 - (e^2 + g^2) \frac{1}{p - il} \right\}^2 ,$$
$$\omega = -(e + ig) d \left\{ \tfrac{1}{2}(u dv - v du) + \frac{i}{p + il} (d\sigma + l(v du - u dv)) \right\} . \quad (15)$$

Therefore, the general non-null orbit divergenceless type D solutions reduce simply to the solutions with null orbit, provided that the parameter ϵ in (3.34)–(3.40) of Ref. 6 assumes value zero.

This null orbit solution contains, among others, the vacuum solutions obtained by Bampi and Cianci[8] and Melvin's[9] magnetic universe ($l = \lambda = 0$).

Within the whole class of divergenceless type D solutions it remains still to obtain the null orbit limit of the Bertotti–Robinson[10,11] branch which turns out to be the only exceptional null orbit type D solution.[12,13] This can be easily done starting from (14)–(15) with

$$l = 0 , \quad n = e^2 + g^2 - \tfrac{1}{3}\lambda + \epsilon^2 , \quad \lambda = e^2 + g^2 ,$$
$$p \to 1 + \epsilon p , \quad \sigma \to \epsilon^{-1} \sigma . \quad (16)$$

In the limit $\epsilon \to 0$ one obtains

$$ds^2 = \frac{1}{\mathscr{P}} dp^2 + \mathscr{P} d\sigma^2 + 2 du dv ,$$
$$\mathscr{P} = 1 - (e^2 + g^2) p^2 . \quad (17)$$

This metric always can be brought to the form

$$ds^2 = 2\phi^{-2} d\xi d\bar{\xi} + 2 du dv , \quad (18)$$

with

$$\phi = 1 + (E^2 + \check{B}^2) \xi\bar{\xi} , \quad \lambda = E^2 + \check{B}^2 , \quad C^{(3)} = -\tfrac{2}{3}\lambda ,$$
$$\omega = \tfrac{1}{2}(E + i\check{B}) d \{ \phi^{-1} (\xi d\bar{\xi} - \bar{\xi} d\xi) + u dv - v du \} . \quad (19)$$

Alternatively, one arrives at this special B–R solution by subjecting Eqs. (10.22)–(10.30) of Ref. 4 to the transformation

$$d\tau \to d\tau + \mathscr{S}^{-1} dq ,$$

and by setting

$$\xi_0 = \eta_0 = 0 , \quad \lambda = \frac{e_0^2 + g_0^2}{(q_0^2 + p_0^2)^2} = E^2 + \check{H}^2 .$$

Having derived some null orbit type D solutions via contractions of the seven-parameter family, we can now state that as far as the second objective of this work is concerned, i.e., deriving the null orbit solution of Leroy, formulas (2.26) in Ref. 2 and (3.36) with $b = 0$ in Ref. 3, modulo minor redefinitions, reduce just to (5) and (6) with the particular values of constants

$$m = \lambda = \epsilon = 0 , \quad \gamma = -e^2 , \quad (20)$$

which leads to the metric

$$ds^2 = \frac{p^2 + q^2}{\mathscr{P}} dp^2 + \frac{\mathscr{P}}{p^2 + q^2} (d\tau + q^2 d\sigma)^2 - 2dq(d\tau - p^2 d\sigma) , \quad (21)$$

where

$$\mathscr{P} = -(e^2 + g^2) + 2np .$$

Notice also that applying then, to the so-constructed solution, the transformation (7), and taking the limit $\epsilon \to 0$, one obtains a null orbit solution, being a special case of (15), with vanishing cosmological constant.

A formal null orbit metric can be derived from (1) by specializing the parameters, in the case of $\lambda > 0$, to particular values

$$m = n = \epsilon = 0 , \quad \gamma = -\tfrac{1}{2}(e^2 - g^2) , \quad \lambda/3 = e^2 + g^2 ,$$

one arrives at a metric with

$$Q = 0 , \quad \mathscr{P} = -\tfrac{1}{3}\lambda (1 + p^4) . \quad (22)$$

Nevertheless, working initially with signature $(+ + + -)$, the so-constructed metric, with $P < 0$, has a wrong signature. In this respect, the metric derived by Debever and Kamran,[14] formulas (7.15)–(7.17), starting from their version of the D's as described in Kinnersley coordinates with signature $(+ - - -)$ (there the Leroy solution was also constructed), is not self-consistent; this fact becomes evident because of the impossibility of equality (7.17) of Ref. 14.

We conclude this paper conjecturing that very likely *all* null orbit solutions of type D with electromagnetic field aligned along the D–P vectors are derivable by limiting transitions and real cuts from the (complexified) seven-parameter solution.

[1]J. F. Plebański and M. Demiański, Ann. Phys. (NY) **98**, 98–127 (1976).
[2]R. Debever and R. G. McLenaghan, J. Math. Phys. **8**, 1711–1726 (1981); related results are also in Bull. Acad. R. Belg. Cl. Sci. **66**, 204–222 (1980). In the first cited publication one finds a complete description of the non-null and null orbit type D fields, modulo Einstein equations.
[3]J. Leroy, Bull. Acad. R. Belg. Cl. Sci. **64**, 130 (1978).
[4]J. F. Plebański, Ann Phys. (NY) **90**, 196–255 (1975).
[5]B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[6]J. F. Plebański, J. Math. Phys. **9**, 1946–1962 (1979).
[7]A. García D. and H. Salazar I., "All Divergenceless Null Orbit Type D Solutions", CINVESTAV, Apartado Postal 14-740, 07000 México, D.F., preprint (1981).
[8]F. Bampi and R. Cianci, Commun. Math. Phys. **70**, 69–74 (1979).
[9]M. A. Melvin, Phys. Lett. **8**, 65 (1964).
[10]B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[11]I. Robinson, Bull. Acad. Pol. Sci. Ser. Math. Phys. **7**, 352 (1959).
[12]J. F. Plabański and S. Hacyan, J. Math. Phys. **20**, 1004–1010 (1979).
[13]A. García D. and J. F. Plebański, J. Math. Phys.**23**, 123 (1982).
[14]R. Debever and N. Kamran, Bull. Acad. R. Belg. Cl. Sci. **66**, 585 (1980).

Explicit solutions of the conformal scalar equations in arbitrary dimensions

José M. Cerveró
Departamento de Fisica Teórica, Facultad de Ciencias, University of Salamanca, Salamanca, Spain

(Received 2 February 1982; accepted for publication 26 March 1982)

Solutions of the equations of motion derived from the scalar conformal invariant Lagrangian in arbitrary dimensions are found. The solutions are invariant under the maximal compact subgroup of the corresponding conformal group. They have finite energy and action. In the case $N = 2$, we also find noticeable topological properties of the solutions.

PACS numbers: 04.20.Jb, 04.20.Me

INTRODUCTION

The number of new—and very often surprising—features appearing in the study of nonlinear equations in mathematical physics has increased in the last few years. The appearance of soliton solutions, topological quantum numbers, and so forth, is not only interesting from the mathematical point of view but also leads us to contemplate possible alternative explanations of several physical effects which are still not too well understood. The most relevant of such an effect is, no doubt, the problem of confinement in nonabelian gauge theories. It is also possible that those properties belonging specifically to nonlinear dynamics might well induce a drastic change in our knowledge of the physical evolution and other basic concepts such as space-time and Lorentz invariance. The success of conformal invariance, as a mathematical tool for studying specific properties of some of those nonlinear equations, suggests that something more fundamental than simply group-theoretical analysis may be hidden behind the door. This paper is an attempt in this direction. From a considerably more modest point of view we show here several surprising properties of the conformal scalar equation in an arbitrary dimension. The goal is achieved by exploiting conformal invariance and constructing solutions with interesting properties by means of the systematic use of the hypertoroidal formalism, a very natural framework whenever we deal with conformal invariance. The paper is organized as follows. Section 1 is devoted to the detailed description of the above mentioned formalism in a N-dimensional "Minkowski space": $(N-1)$-spatial dimensions and one time dimension. Since the cases $N = 1,2$ are rather pathological we confine ourselves to the case $N > 2$ in Secs. 2 and 3, where we give explicit solutions with finite energy and action. Interesting properties which concern the energy momentum tensor are described in Sec. 3. Section 4 deals with the $N = 1$ case. Section 5 is entirely devoted to the $N = 2$ case (the Liouville equation) and new solutions of this equation are presented with several topological properties arising precisely from the use of the toroidal formalism. Finally, Sec. 6 is one of conclusions. There, we show that all our solutions are stable under the new "time-parameter" θ and we speculate about the possibility of quantization using this θ-parameter. Also, the quantum field theory based on the θ-evolution is shown to be a rather suitable approach to quantization of conformally invariant field theories.

1. THE SPACE AND THE EQUATIONS OF MOTION

We consider first a generalized Minkowski space in N dimensions. This means a pseudo-Euclidean metric flat space of signature $(N-1,1)$. The $(N-1)$ spatial variables will be labeled by x_i $(i{:}1,\ldots,N-1)$. The time dimension will be called t throughout the paper. The flat metric $g_{\mu\nu}$ $:(\mu,\nu{:}1,2,\ldots,N-1,t)$ is

$$g_{11} = g_{22} = \cdots = g_{N-1,N-1} = -g_{tt} = 1. \tag{1}$$

The conformal group corresponding to this generalized Minkowski space is $SO(N,2)$[1], and it acts linearly on the pseudo-Euclidean space of signature $(1,2,\ldots,N-1,N+1;\ t,N+2)$. The flat metric in this space is

$$g_{11} = g_{22} = \cdots = g_{N-1,N-1} = g_{N+1,N+1} = +1, \tag{2a}$$

$$g_{tt} = g_{N+2,N+2} = -1. \tag{2b}$$

The conformal invariant Lagrangian for a scalar field in the generalized Minkowski space requires that all terms have to have overall scale dimensionality $-N$. Since the field has scale dimension $l = -(N-2)/2$, the only allowed terms are[1]

$$\mathscr{L} = \tfrac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - g\,\frac{N-2}{2N}\,\phi^{2N/N-2}, \tag{3}$$

and g is, of course, a dimensionless coupling constant. The equations of motion are readily obtained from (3)

$$\Box\phi + g\phi^{N+2/N-2} = 0, \tag{4}$$

where $\Box$ is the generalized d'Alambertian

$$\Box = \partial_+^2 + \partial_2^2 + \cdots + \partial_{N-1}^2 - \partial_t^2. \tag{5}$$

We are dealing with Eq. (4) for arbitrary N. In the physical case, $N = 4$ and we recover the celebrated $g\phi^4$ Lagrangian. However, since cases $N = 1,2$ are slightly pathological they will be discussed separately at the end of the paper. Thus, we confine ourselves to the general case for $N > 2$.

Although the Lagrangian (3) is invariant under the full conformal group $SO(N,2)$, we look for solutions invariant under the maximal compact subgroup of $SO(N,2)$, namely $O(N)\times O(2)$. In conformal coordinates ξ, this compact subgroup leaves the $(N+2)$-light-cone invariant

$$\xi_1^2 + \xi_2^2 + \cdots + \xi_{N-1} - \xi_t^2 + \xi_{N+2}^2 - \xi_{N+2}^2 = 0, \tag{6}$$

which is a general property of the conformal group. Besides, since we are confined to the $O(N)\times O(2)$ subgroup, the ξ's

 0022-2488/82/081466-05$02.50

also have to satisfy:

$$\xi_1^2+\xi_2^2+\cdots+\xi_{N-1}^2+\xi_{N+1}^2=1, \tag{7a}$$

$$\xi_t^2+\xi_{N+2}^2=1. \tag{7b}$$

The last condition defines the submanifold $S^{N-1}\times S^1$, a generalized hypertorus. In fact, it has been proved by the author[2] that the generalized Minkowski space is homeomorphic to this manifold in the following way:

$$\mathscr{M}_N \simeq (S^{N-1}\times S^1)/Z_a,$$

where Z_a is an abelian group which depends on the topology of the S^{N-1} sphere. Since this is a compact manifold, it can be described by N angles and it represents the natural compactification of $\mathscr{M}_N$. In order to parametrize the $(S^{N-1}\times S^1)/Z_a$ manifold we take the $(N-2)$ spatial angles of $\mathscr{M}_N:(\theta_1,\theta_2,\ldots,\theta_{N-2})$. Next, we introduce light cone coordinates

$$t_\pm = t\pm r, \tag{8a}$$

$$r=[x_1^2+x_2^2+\ldots+x_{N-1}^2]^{1/2} \tag{8b}$$

and define the angles

$$\theta=\arctan\frac{t_+ + t_-}{1-t_+t_-},\quad \theta_0=\arctan\frac{t_+ - t_-}{1+t_+t_-}, \tag{9}$$

where

$$\sin\theta=\frac{t_+ + t_-}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}},$$

$$\sin\theta_0=\frac{t_+ - t_-}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}}, \tag{10a}$$

$$\cos\theta=\frac{1-t_+t_-}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}},$$

$$\cos\theta_0=\frac{1+t_+t_-}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}}. \tag{10b}$$

The connection between the ξ's and the generalized Minkowski coordinates is

$$\xi_i=\frac{2x_i}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}};\ (1:1,\ldots,N-1),$$

$$\xi_{N+1}=\frac{1+t_+t_-}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}}, \tag{11a}$$

$$\xi_t=\frac{t_+ + t_-}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}},$$

$$\xi_{N+2}=\frac{1-t_+t_-}{(1+t_+^2)^{1/2}(1+t_-^2)^{1/2}}. \tag{11b}$$

It is easy to check that (11a) and (11b) fulfill the condition (6) and (7a) and (7b). Then θ parametrizes S^1 and $(\theta_0,\theta_1,\theta_2,\ldots,\theta_{N-2})$ parametrize S^{N-1}.

The generalized Minkowski line element is

$$ds^2=[dr^2+r^2d\Omega^2-dt^2], \tag{12}$$

where $d\Omega^2$ contains the angular dependence in the $(N-2)$ spatial angles $(\theta_1,\theta_2,\ldots,\theta_{N-2})$. In the new coordinates, (12) becomes

$$ds^2=\frac{1}{(\cos\theta+\cos\theta_0)^2}[d\theta_0^2+\sin^2\theta_0 d\Omega^2-d\theta^2]. \tag{13}$$

The only requirement needed in order to obtain identification between (12) and (13) is

$$(\cos\theta+\cos\theta_0)>0, \tag{14}$$

which represents analytically the condition imposed by the abelian quotient $(S^{N-1}\times S^1)/Z_a$.

It is useful to write the generalized d'Alambertian (5) in $t_\pm$ coordinates. We obtain

$$\Box=-4\frac{\partial^2}{\partial t_+\partial t_-}+2\frac{N-2}{t_+-t_-}\left[\frac{\partial}{\partial t_+}-\frac{\partial}{\partial t_-}\right]+\text{angular part}. \tag{15}$$

The solutions that will be investigated are isotropic. Thus, we can neglect the last term in (15).

2. THE SOLUTIONS

Consider the field configuration

$$\phi_c=\frac{A}{(1+t_+^2)^\alpha(1+t_-^2)^\alpha},\quad A,\alpha\ \text{constants}. \tag{16}$$

Looking at the expressions (6), (7), (11a) and (11b), we can easily see that ϕ_c is $O(N)\times O(2)$ invariant. The d'Alambertian (15) acting on (16) yields the relationship:

$$\Box\phi_c=-\phi_c^{N+2/N-2} \tag{17}$$

if

$$2\ln A=(N-2)\ln(N-2), \tag{18a}$$

$$\alpha=(N-2)/4. \tag{18b}$$

Then

$$\phi_c=g^{-(N-2)/4}\frac{(N-2)^{(N-2)/2}}{(1+t_+^2)^{(N-2)/4}(1+t_-^2)^{(N-2)/4}} \tag{19}$$

is an $O(N)\times O(2)$ solution of (4).

The corresponding $O(N)$ solutions can be found if we multiply ϕ_c by a function of the angle θ [see Eq. (9)]. The d'Alambertian acting on the product $\phi_c f(\theta)$ gives the following result:

$$\Box\{\phi_c f(\theta)\}=-\phi_c^{N+2/N-2}\left\{\frac{4}{(N-2)^2}f''(\theta)+f(\theta)\right\}. \tag{20}$$

where $f''(\theta)=d^2f(\theta)/d\theta^2$. Comparing (20) with (4) see that $\phi_c f(\theta)$ is a $O(N)$ solution if $f(\theta)$ verifies the nonlinear ordinary differential equation

$$\frac{4}{(N-2)^2}f''(\theta)+f(\theta)-f(\theta)^{N+2/N-2}=0. \tag{21}$$

The only cases in which $(N+2)/(N-2)$ is an integer are $N=3,4,6$. In those cases we find solutions of (21) in terms of elliptical and hyperelliptical functions. The general case when $(N+2)/(N-2)$ is an arbitrary rational number is at present under investigation.

3. ENERGY MOMENTUM TENSOR

We generalize the improved energy momentum tensor of Callan, Coleman, and Jackiw[3] to arbitrary dimension in this generalized Minkowski space. We find

$$\theta_{\mu\nu} = (\partial_\mu \phi)(\partial_\nu \phi) - g_{\mu\nu}\mathscr{L} - \frac{1}{M}[\partial_\mu\partial_\nu - g_{\mu\nu}\Box]\phi^2. \tag{22}$$

where M is:

$$M = 4(N-1)/(N-2). \tag{23}$$

The tensor $\phi_{\mu\nu}$ (22) can be rewritten as

$$\phi_{\mu\nu} = \frac{N}{2(N-1)}\partial_\mu\phi\partial_\nu\phi - \frac{N-2}{2(N-1)}\phi\partial_\mu\partial_\nu\phi - \frac{1}{2}g_{\mu\nu}\left[\frac{1}{N-1}(\partial\phi)^2 + g\frac{N-2}{N(N-1)}\phi^{2N/N-2}\right], \tag{24}$$

exhibiting its obvious property of being traceless.

Using (19), the energy momentum tensor for ϕ_c becomes

$$\phi_{\mu\nu} = \phi_c^2\frac{(N-2)^2}{(N-1)}\left[\omega_\mu\omega_\nu - \frac{1}{N}g_{\mu\nu}\omega^\alpha\omega_\alpha\right] \tag{25}$$

and ω_μ is defined as

$$\omega_\mu = \frac{1}{2}\left[\frac{L_\mu^+}{1+t_+^2} + \frac{L_\mu^-}{1+t_-^2}\right] \tag{26}$$

and

$$L_\mu^\pm = \left(\pm\frac{x_i}{r}, 1\right).$$

The vector field ω_μ verifies

$$\omega^\nu\overline{D}_\nu\omega_\mu = 0, \tag{27a}$$

$$\overline{D}_\mu\omega_\nu = 0, \tag{27b}$$

where the covariant derivative $\overline{D}_\mu$ is defined with respect to the conformally flat Weyl invariant metric

$$\overline{g}_{\mu\nu} = \phi_x^2 g_{\mu\nu}. \tag{28}$$

The first condition (27a) says that ω_μ is a geodesic field. If $x_\mu = x_\mu(s)$ is a geodesic corresponding to the metric (29) then, $\omega_\mu = dx_\mu(s)/ds$. The condition (27b) shows that in addition ω_μ represents a vector field of parallel vectors.

From the expression (25) we can find the energy of the solution ϕ_c. We have

$$E = \int dv\theta_{tt}, \tag{29}$$

where dV is the spatial volume element in N dimensions. The action of the solution is

$$A = \int dVdt\mathscr{L}(\phi_c), \tag{30}$$

where $\mathscr{L}(\phi_c)$ is the Lagrangian (3) particularized to the solution ϕ_c. After some lengthy calculations we can find expressions for the energy and action as functions of N. They are

$$E = (N-2)^N\pi^{(N-1)/2}\Gamma\left(\frac{N+1}{2}\right)\quad \Gamma(N+1), \tag{31a}$$

$$A = (N-2)^N\pi^{(N+1)/2}\Gamma\left(\frac{N+1}{2}\right)\quad \Gamma(N+1), \tag{31b}$$

we see that both quantities are indeed finite for an arbitrary (but not infinite) number of dimensions N.

Let us calculate now, the energy-momentum tensor for the O(N)-symmetric solutions $\phi \equiv \phi_c f(\theta)$, where $f(\theta)$ verifies (21). It turns out that the energy momentum tensor is *independent* of the explicit form of $f(\theta)$. This curious fact was first observed by the author,[4,5] in the $N=4$ case. However, it also holds for general $N>2$. To see this, we insert $\phi \equiv \phi_c f(\theta)$ in (24). The energy-momentum tensor for $\phi_c f(\theta)$ is thus obtained. After some algebra, we obtain

$$\phi_{\mu\nu} = E_0 N\frac{(N-2)^2}{(N-1)}\phi_c^2\left[\omega_\mu\omega_\nu - \frac{1}{N}\omega^\alpha\omega_\alpha\right], \tag{32}$$

where E_0 stands for the quantity

$$E_0 = \frac{2}{(N-2)^2}f'^2(\theta) + \frac{1}{N}f^2(\theta) - \frac{2}{N(N-2)}f(\theta)f''(\theta). \tag{33}$$

It is easy to see that E_0 is a constant. Integrating (21), we get

$$\frac{2}{(N-2)^2}f'^2(\theta) + \frac{1}{2}f^2(\theta) - \frac{N-2}{2N}f(\theta)^{2N/N-2} = E_0. \tag{34}$$

Multiplying (21) by $[(N-2)/2N]f(\theta)$ and subtracting from (34), we obtain (33). The constant E_0 can be identified as the mechanical energy of a particle of mass $4/(N-2)^2$ moving in a potential

$$V(f) = \frac{1}{2}f^2 - \frac{N-2}{2N}f^{2N/N-2}.$$

The interesting fact is that even if we don't know the explicit form of $f(\theta)$, the energy of the solution can be calculated explicitly. From (32) we see that

$$E^{O(N)} = NE_0E,$$

where E is given by (31a). Therefore, we have

$$E^{O(N)} = E_0(N-2)^N\pi^{(N-1)/2}\Gamma\left(\frac{N+1}{2}\right)\Big/\Gamma(N). \tag{35}$$

4. THE $N=1$ CASE

This case is not even a field theory but simply one-dimensional particle mechanics. The integration of the equations of motion is, however, interesting since it can be performed in our formalism, and we shall present it here for the sake of completeness. In addition, the quantum theory of this system has been analyzed by De Alfaro, Fubini, and Furlan,[6] and their results are very interesting, too.

The Lagrangian is

$$\mathscr{L} = \frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 - \frac{1}{2}\frac{g}{\phi^2}, \tag{36}$$

where t represents the only independent dimension of the system. The equation of motion is

$$\frac{d^2\phi}{dt^2} = \frac{g}{\phi^3}. \tag{37}$$

The corresponding conformal group is SO(2,1) and the maximal compact subgroup is, of course, SO(2). In light cone coordinates

$$\xi_1^2 + \xi_2^2 - \xi_3^2 = 0,$$

with the restriction $\xi_1^2 + \xi_2^2 = 1, \xi_3^2 = 1$, we obtain

$$\xi_1 = \sin\theta = \frac{2t}{1+t^2}, \tag{38a}$$

$$\xi_2 = \cos\theta = \frac{1-t^2}{1+t^2}. \tag{38b}$$

In those coordinates, the general solution can be written as

$$\phi = g^{1/4}\left[\frac{2C + 2(C^2-1)^{1/2}\sin\theta}{1+\cos\theta}\right]^{1/2}, \tag{39}$$

where C is an integration constant. The action is divergent for any value of C. It is, however, interesting to consider the cases $C = 0$ and $C = 1$. In the first case we obtain a *complex* solution

$$\phi_0 = g^{1/4}\left[\frac{2i\sin\theta}{1+\cos\theta}\right]^{1/2}, \tag{40}$$

which has a logarithmically divergent action and *zero* energy. In the case $C = 1$ we obtain

$$\phi_1 = g^{1/4}\left[\frac{2}{1+\cos\theta}\right]^{1/2}, \tag{41}$$

with action that diverges as tan $(\pi/2)$, and energy $E = \frac{1}{2}g^{1/2}$, a constant proportional to the square root of the dimensionless coupling.

The zero energy case is reminiscent of the *instanton* configuration in Euclidean space[5] and the finite energy case to the *meron* configuration[4] in Yang–Mills fields and ϕ^4-theory. If we consider the quantum evolution in the "compact-time" θ, we find a rich discrete spectrum which has been analyzed exhaustively in Ref. 6. Whether or not quantization in θ-time might have a more profound physical significance will be discussed in Sec. 6.

5. THE $N = 2$ CASE

As it stands, the Lagrangian (3) is not defined for $N = 2$. This is due to the well-known fact that the conformal scalar equation for $N = 2$ is the Liouville equation, as can be seen from the differential geometry of conformal spaces.[7] The Lagrangian is then

$$\mathscr{L} = \tfrac{1}{2}(\partial_\alpha\phi)(\partial^\alpha\phi) - ge^\phi, \tag{42}$$

where α:(1,0):(x,t). The equations of motion are

$$(\partial_x^2 - \partial_t^2)\phi = ge^\phi. \tag{43}$$

The general solution of (43) was found by Liouville more than a hundred years ago.[8] This equation has also been recently studied by several authors.[9–12] In our formalism, however, some particular solutions can be obtained which exhibit interesting behavior from the topological point of view. This is due to the particular kind of boundary conditions introduced by the toroidal formalism. We can still use the conventions of Sec. 1. The only change that we shall introduce is the definition of light-cone variables. From now on, we define

$$t_+ = t + x, \tag{44a}$$

$$t_- = t - x. \tag{44b}$$

Since $-\infty < t < +\infty$ and $-\infty < x < \infty$, the new variables t_+ and t_- range now within a different open subset of the real line. However, this minor change only affects the range of integration and we will take it into account only in our calculations.

Apart from this technical point, the definiton (9) and Eqs. (10a) and (10b) for the angles θ and θ_0 will be the same. The conformal group is now SO(2,2) and the compact manifold $(S^1 \otimes S^1)/Z_2$ is the *torus*. The maximal compact subgroup is, indeed, O(2)×O(2), which might lead us to the erroneous conclusion that θ and θ_0 play a dual role. This is, in fact, not the case as we shall see below. The angle θ is a *real evolution variable* as opposed to the "spatial" angle θ_0. The situation would be different for a Euclidean metric in (43). We will go back to this point later on.

No O(2)×O(2) symmetric solution exists. This can be easily seen from the form of Eq. (43). Let us consider $O(2)_s$ invariant solutions (where $O(2)_s$ is the spatial subgroup of O(2)×O(2)) of the following form:

$$\phi_1 = \ln[(B/g)(\cos\theta + \cos\theta_0)^2\sec^2\{(B/2)^{1/2}(\theta - \bar\theta)\}], \tag{45a}$$

$$\phi_2 = \ln[(B/g)(\cos\theta + \cos\theta_0)^2\mathrm{csec}^2\{(B/2)^{1/2}(\theta - \bar\theta)\}], \tag{45b}$$

$$\phi_3 = \ln[(B/g)(\cos\theta + \cos\theta_0)^2\mathrm{cosech}^2\{(B/2)^{1/2}(\theta - \bar\theta)\}], \tag{45c}$$

where B is a *positive* constant and $\bar\theta$ a trivial constant shift of the variable θ. The configurations (45a)–(45c) are solutions of (43) for *positive g*. If we consider the case of *negative g*, then the only solution would be

$$\phi_4 = \ln[(B/|g|)(\cos\theta + \cos\theta_0)\mathrm{sech}^2\{(B/2)^{1/2}(\theta - \bar\theta)\}]. \tag{45d}$$

We have to address ourselves to the task of constructing an energy-momentum tensor for the Lagrangian (42) satisfying the requirements of conformal invariance. Tracelessness will be introduced through the equation of motion. The only possible choice for this "improved" energy momentum tensor is

$$\theta_{\alpha\beta} = \partial_\alpha\phi\partial_\beta\phi - g_{\mu\nu}\mathscr{L} - 2\epsilon_{\alpha\gamma}\epsilon_{\beta\delta}\partial^\gamma\partial^\delta\phi, \tag{46}$$

where $\epsilon_{\alpha\beta}$:($\epsilon_{01} = -\epsilon_{10} = 1$ and $\epsilon_{00} = \epsilon_{11} = 0$). We can calculate the energy of the configuration (45a)–(45d) through the θ_{tt} component of (46). The answer is

$$E = 4\pi(1 - B/2), \text{ for the configurations (45a) and (45b)}, \tag{47a}$$

$$E = 4\pi(1 + B/2), \text{ for the configurations (45c) and (45d)}. \tag{47b}$$

Therefore, configurations (45a) and (45b) have *zero* energy for the case $B = 2$. This is again reminiscent of the properties of instantons. That simple intuition is indeed confirmed by the following observation. For the configurations

(45a) and (45d) which are solutions of the equations of motion, we can replace in the Lagrangian (42) the potential $-ge^{\phi}$ by the left-hand side of (43). But this is a total divergence which should not have any effect in the Lagrangian. However, upon integration of $(\partial_x^2 - \partial_t^2)\phi$ over x and t this gives a nonvanishing contribution. For instance, for configuration (45a) we obtain

$$-g \int dx\, dt\, e^{\phi_1} = 4\pi(B/2)^{1/2}\tan\left\{\left(\frac{B}{2}\right)^{1/2}\pi\right\}, \tag{48a}$$

and for configurations ϕ_2, ϕ_3, and ϕ_4 we also have

$$g \int dx\, dt\, e^{\phi_2} = 4\pi(B/2)^{1/2}\cot\left\{\left(\frac{B}{2}\right)^{1/2}\pi\right\}, \tag{48b}$$

$$g \int dx\, dt\, e^{\phi_3} = 4\pi(B/2)^{1/2}\coth\left\{\left(\frac{B}{2}\right)^{1/2}\pi\right\}, \tag{48c}$$

$$-g \int dx\, dt\, e^{\phi_4} = 4\pi(B/2)^{1/2}\tanh\left\{\left(\frac{B}{2}\right)^{1/2}\pi\right\}. \tag{48d}$$

Those continuous quantities represent a sort of topological indices for the Liouville equation. Notice that (48a) and (48b) blow up for $B = 2$ and $B = \frac{1}{2}$, respectively. However, (48c) and (48d) are everywhere continuous and regular for any value of B.

The action is also finite (although sometimes complex) for those field configurations. The regular piece of the action for ϕ_1 is

$$A = 4\pi^2[1 - (B/2)^{1/2}]^2,$$

with similar expression for the others. Again $A = 0$ if $B = 2$, which suggests a minimum for the action as in the instanton case.

The physical applications of our solutions are clear. Recently, Polyakov[13] has suggested that the two-dimensional Liouville system may be crucial for the quantization of the four-dimensional string in arbitrary dimension. Obviously the solutions that we present (when continued to Euclidean space) are perfect candidates for semiclassical calculation of the Euclidean Green function since they represent, in fact, minima of the action. The functional integral can be expanded around those minima through the conventional saddle-point approximation. Work in this direction is now in progress.

Concerning the topological significance of (45a)–(45d), we would like to point out that similar continuous topological indices appear in Yang–Mills fields, took, when working in Minkowski space.[14] Much further work is needed, however, in order to understand the topological properties of Minkowskian field theory in this hypertoroidal framework.

6. CONCLUSIONS

The hypertoroidal formalism represents a really fruitful mathematical tool in the search for classical solutions within the framework of conformally invariant field theories. This formalism has already been successfully used in the context of sourceless Yang–Mills fields.[4,14] In this paper we have been able to solve the conformal scalar equation in arbitrary dimension using hypertoroidal coordinates. The fact that the solutions possess not only finite energy and action but also (as in Liouville's case, $N = 2$) interesting topological properties is, from our point of view, a clear sign of the fruitfulness of this approach. We would like to end with some comments about the stability of our solutions. As we have stated in Ref. 4, our configuration are dissipative in the usual time variable t; that is to say, that from the point of view of usual time, t-evolution, the solutions should not be stable. However θ-evolution is also possible.[14] In fact, it is the evolution that we should use in any conformally invariant quantum theory. As has been proved by Lüscher and Mack,[15] the analytic continuation of the Euclidean Green's functions is only free of kinemetic singularities if we use, instead of the conventional Hamiltonian H, the "conformal Hamiltonian" $H_\theta = \frac{1}{2}(P_0 + K_0)$. The evolution parameter corresponding to such a Hamiltonian is precisely our θ-variable. Besides, the energy H and H_θ *coincide* for hypertoroidal configurations, as do those presented in this work. We should seriously consider the possibility of quantizing in θ rather than in t, any conformally invariant field theory.

From the point of view of θ-stability, our configurations are perfectly stable since they are defined in a compact manifold; although *not t-stable*, our solutions *are θ-stable*. An exciting example of θ-quantization has been given in reference.[6] Whereas the physical interpretation is not totally clear, we consider the possibility of analyzing the θ-evolution a rather promising alternative to the quantization of conformally invariant field theories.

ACKNOWLEDGMENTS

I would like to thank A. Actor for correspondence concerning this paper and for pointing out Ref. 16 to me. Some of the results obtained here slightly overlap with the work of this author, although in a different context. I would also like to thank J. Mateos for some comments about the topological properties of the Liouville equation.

[1] S. Fubini, Nuovo Cimento **34**, 521–54 (1976).
[2] J. Cervero, Ph.D. thesis (1973) (unpublished). Some pieces of information can be found in J. Cervero and L. J. Boya, Int. J. Theor. Phys. **12**, 47–55; 55–61 (1975) and **16**, 333–39; 339–45 (1977).
[3] C. C. Callan, S. Coleman, and R. Jackiw, Ann. Phys. **59**, 42–73 (1970).
[4] J. Cervero, L. Jacobs, and C. Nohl, Phys. Lett. B **69**, 351–54 (1977).
[5] A. Actor, Rev. Mod. Phys. **51**, 461–525 (1979).
[6] V. De Alfaro, S. Fubini, and G. Furlan, Nuovo Cimento **34**, 569–612 (1976).
[7] See L. P. Eisenhart, *Riemannian Geometry* (Princeton, U.P., Princeton, N. J., 1953), pp. 89–96.
[8] J. Liouville, J. Math. Pures Appl. **18**, 71 (1853).
[9] V. A. Andreev, Theor. Math. Phys. **29**, 213 (1976).
[10] B. M. Barbashov, V. V. Nesterenko, and A. M. Chervyakov, Theor. Math. Phys. **40**, 15 (1979).
[11] G. P. Dzhordzhadze, A. K. Pogrebkov, and M. K. Polivanov, Theor. Math. Phys. **40**, 221 (1979).
[12] A. K. Pogrebkov, Theor. Math. Phys. **45**, 161 (1980).
[13] A. M. Polyakov, Phys. Lett. **B103**, 207 (1981).
[14] J. Cervero and C. Gomez, Phys. Lett. **B104**, 467 (1981).
[15] M. Lüscher and G. Mack, Comm. Math. Phys. **41**, 203 (1975).
[16] A. Actor, Ann. Phys. **131**, 269 (1981).

Twistor bundles and gauge action principles of gravitation

C. P. Luehr
Department of Mathematics, University of Florida, Gainesville, Florida 32611
M. Rosenbaum P.[a)]
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, C. U., México 20, D. F., Mexico

(Received 17 December 1980; accepted for publication 26 February 1982)

A gauge theory of gravitation is constructed with a twistor bundle as the starting point. Each fiber is a twistor space, acted upon by the Poincaré group, which forms an internal symmetry group. The formalism leads to a twistor action principle which overcomes difficulties encountered in previous attempts in the literature to formulate a true spinorial variational principle.

PACS numbers: 04.50. + h

I. INTRODUCTION

Gauge theories play a fundamental role in the description of the basic interactions of nature and form part of present-day attempts at grand unification. The essential features which characterize all gauge theories are the appearance of arbitrary functions in the description of the fields and the existence of constraints. General relativity contains all these features, and many papers have been written presenting different approaches to its treatment as a gauge theory of the Poincaré group.[1]

One such formalism was developed by the authors in a previous paper[2] (hereafter referred as I), in which we dealt with the Poincaré group as an internal gauge group acting on the fibers of an appropriately constructed vector bundle (rather than on the space–time base manifold itself) and thus arrived at an unambiguous gauge theory of gravitation. The essential features of the theory are the use of fiber bundle techniques which provide a convenient framework for a geometric and coordinate-free discussion and the introduction of a five-dimensional faithful representation space of the Poincaré group as the typical fiber. This allows the treatment of the group as an internal group. The possible functional form of the free Lagrangian, which must be included in the theory in order to determine the equations of motion for the connections (gauge fields), encompasses general relativity and the Einstein–Cartan theory as special cases, as well as other gravitational theories with torsion which have been proposed recently.

In addition to the inherent advantage of gauge theories for the systematic construction and study of a wide range of gravitational Lagrangians, they also afford a natural structure for the formulation of variational principles of the Palatini type. A discussion of the general features of the variational procedure was given in I in the context of the tensor formalism there developed.

It is well known, however, that spinors fit in with general relativity in a most natural way; a fact which leads to the belief that spinors are essentially simpler and more fundamental than 4-vectors.[3–5] Thus, it appears desirable in this context to develop a gravitational gauge theory based on spinors rather than on tensors.[6]

The formulation of such a spinor gauge theory will be the main objective of this paper. In addition, our results will also serve to shed further light on the structure of the theory developed in I and some of its features which will be shown to originate within the spinor formalism in a most natural manner.

One further result of our spinor gauge theory, which we believe merits consideration by itself is the development of a truly spinorial variational principle. Although gravitational Lagrangians and field equations constructed in tensor language have been readily translated into spinor formalism, no action principle based on purely spinorial entities and in the strict sense of gauge theories has been obtained so far. There exist in fact some "hybrid" approaches in the literature,[7] but these involve variations of the Hermitian mixed quantities[8] $(I_4)_{\mu A'B}$ and simultaneous variations of tensorial affine connections and spinorial affine connections for the Palatini principles.[9] If we recall that the tensorial Palatini action principles involve simultaneous variations of the 4-vector inner product (or metric tensor) and the vector connections, we clearly see that what was actually done in the papers referred to above is a mere renaming and variation of still intrinsically tensorial quantities. In fact, from[10]

$$\mathbf{I}_4 = (I_4)_{\mu A'B}\mathbf{E}^\mu \mathbf{h}^{A'} \mathbf{h}^B, \tag{1.1}$$

we see that the $(I_4)_{\mu A'B}$ are nothing but the hybrid 4-vector and spinor components of the identity tensor in $\mathscr{M}_4$. Thus, a variation of $(I_4)_{\mu A'B}$ is intrinsically the same operation as the variation of $\mathbf{I}_4$ (i.e., the variation of the 4-vector inner product).

One naive attempt at constructing a variational principle from truly spinorial entities would be to vary simultaneously the spinor inner product and the spinor connections. Note, however, that since the spinor space $\mathscr{S}_2$ and its conjugate $\bar{\mathscr{S}}_2$ are two-dimensional symplectic spaces[10] for which there is only one independent unit dyadic $\mathbf{I}_2$ (metric spin-tensor) and one independent conjugate unit dyadic $\bar{\mathbf{I}}_2$, variation of $\mathbf{I}_2$ and $\bar{\mathbf{I}}_2$ would be proportional to $\mathbf{I}_2$ and $\bar{\mathbf{I}}_2$, respectively, and the resulting equations would be scalar and the theory trivial. The reason why this approach fails becomes

[a)] Supported in part by International Scientific Exchange Program Grants, National Science Foundation OIP75-09783A01 and Consejo Nacional de Cienca y Tecnología No. 955.

evident when we note that the gauge group needed in I for constructing gravitational theories is the Poincaré group, while spinors form the representation space of SL(2, C), which is homomorphic to the Lorentz group.

Therefore, in order to construct a proper spinor gauge theory, we must make use of twistor algebra.[11,12] In fact, twistors form the representation space of SU(2, 2), which is (4–1) homomorphic to the conformal group C(3, 1). By incorporating the vertex of the null cone at infinity explicitly into the formalism, we can break the conformal invariance while retaining the Poincaré invariance[13] required for our theory.

In order to make the discussion of the following sections more self-contained, we review in Sec. II some of the various spinor and spin-tensor spaces from an abstract point of view developed by us previously.[10,14] We also present in tabular form the twistor and twist-tensor spaces and their basic properties that will be utilized later in the paper together with a description of our notation and its relation to Penrose's twistor formalism. The reader familiar with twistor algebra in that author's notation should have no problem in following the discussion in the rest of the paper by making use of Tables I, II, and III.

For the more mathematically oriented reader we have included in the Appendix a summary of the essential features of twistor algebra obtained from an axiomatic point of view, in which we arrive at the realization of the homomorphism:

$$\mathrm{SU}(2,2)\to \mathrm{O}(4,2)\to \mathrm{C}(3,1).$$

We show in this Appendix how the essentially coordinate-free approach to general relativity adopted by Penrose[3] can be made further intrinsic, thus emphasizing spinors and twistors as geometrical objects subject to formal rules of manipulation rather than seeing them as sets of components. For notation, we rely to a large extent on the one we have developed previously,[10,14,15] since it seems to fit best the purposes mentioned above.

Thus, we hope that this section and the Appendix will serve not only as a summary of twistor algebra but also as an introduction to an abstract approach to twistor algebra from the axiomatic point of view of modern coordinate-free and component-free tensor analysis.

In Sec. III we make use of these structures to construct the appropriate twistor bundles on which we can base the formulation of our gauge field theory of gravitation. Several results in this section serve to elucidate some points of the theory we presented in I.

In Sec. IV we apply our formalism to the development of a truly spinorial variational principle for the Einstein–Cartan theory. We show that the full theory emerges if we vary only the twistor connections (gauge variables) on the bundle space. This is what we ought to expect from a properly constructed gauge theory.

II. SPINORS, TWISTORS AND TWIST-TENSOR SPACES

Spinors: Since we previously[14] presented a systematic discussion of an intrinsic formulation of spinor theory and some of its computational advantages over the component notation, here we only summarize in Table I some of the spinor and spin-tensor spaces as well as their basic properties, which we will require in later sections. Also, because intrinsic notation is not commonly utilized, we give in the same table a comparison with the component notation employed by Penrose[13] in his theory of twistors. Making use of the above structures, we can define additional twist-tensor spaces by taking tensorial and exterior powers of $\mathscr{U}$ and $\mathscr{U}'$. The essential features of these spaces are contained in Tables II and III together with a comparison (when appropriate) with the notation of Penrose.[13] A more complete discussion of these ideas, intended for the more mathematically oriented reader, is given in the Appendix.

Twistors: Flat twistor space $\mathscr{U}\equiv\mathscr{U}_{2,2}$ is essentially a space of Dirac bispinors in which a nondegenerate Hermitian inner product $\langle \mathbf{s}|\mathbf{t}\rangle$ is defined. This product is antilinear in the Dirac bispinor $\mathbf{s}$ and linear in the Dirac bispinor $\mathbf{t}$, has the signature $(+\ +\ -\ -)$, and is invariant under SU(2, 2).

We can relate the algebra of twistors to the algebra of Dirac bispinors by noting [see Eq. (A63) in the Appendix] that

$$\langle \mathbf{s}|\mathbf{t}\rangle = \bar{\mathbf{s}}\blacktriangle(\mathbf{I}_2-\bar{\mathbf{I}}_2)\blacktriangle\mathbf{t}, \tag{2.1}$$

where $\bar{\mathbf{s}}$ is a Dirac adjoint bispinor (also adjoint twistor) and $\mathbf{I}_2$ and $\bar{\mathbf{I}}_2$ are the unit spin-tensors in $\mathscr{S}_2\otimes\mathscr{S}_2$, and $\bar{\mathscr{S}}_2\otimes\bar{\mathscr{S}}_2$, respectively, defined in Table I.

The inner product in $\mathscr{U}$ gives a complex number which can be reexpressed in terms of linear functionals on $\mathscr{U}$, by introducing a dual twistor space $\mathscr{U}'$. Explicitly, we define conjugate twistors $\hat{\mathbf{l}}$ by the antilinear map $\mathbf{l}\in\mathscr{U}\to\hat{\mathbf{l}}\in\mathscr{U}'$ such that

$$\hat{\mathbf{l}}\circ\mathbf{m} = (\mathbf{l}\circ\hat{\mathbf{m}})^* = \langle \mathbf{l}|\mathbf{m}\rangle. \tag{2.2}$$

We also use the symbol $\circ$ to denote the action of a cotangent vector on a tangent vector at a point and to denote contraction operations on tensors constructed from cotangent and tangent vector spaces.

Note that conjugate twistors are related to Dirac adjoint bispinors by means of Eqs. (A54) and (A55) given in the Appendix. In particular, we shall require the totally antisymmetric alternating twist-tensor $\boldsymbol{\Lambda}\in\mathscr{U}^{\wedge 4}$ ($\epsilon^{\alpha\beta\gamma\delta}$ in Penrose's notation) and $\hat{\boldsymbol{\Lambda}}\in\mathscr{U}'^{\wedge 4}$ ($\epsilon_{\alpha\beta\gamma\delta}$ in Penrose's notation) defined by the requirement

$$\hat{\boldsymbol{\Lambda}} \substack{\circ\circ\\ \circ\circ} \boldsymbol{\Lambda} = 4!. \tag{2.3}$$

With $\hat{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Lambda}$ we can form duals of antisymmetric twist-tensors in $\mathscr{U}^{\wedge 2}$ and $\mathscr{U}'^{\wedge 2}$, respectively, by means of the following operations:

$$\mathbf{B}\in\mathscr{U}^{\wedge 2}\to\star\mathbf{B}\in\mathscr{U}'^{\wedge 2}:\quad \star\mathbf{B}=\tfrac{1}{2}\,\hat{\boldsymbol{\Lambda}} \substack{\circ\\ \circ} \mathbf{B}, \tag{2.4}$$

$$\mathbf{C}'\in\mathscr{U}'^{\wedge 2}\to\star\mathbf{C}'\in\mathscr{U}^{\wedge 2}:\quad \star\mathbf{C}'=\tfrac{1}{2}\,\boldsymbol{\Lambda} \substack{\circ\\ \circ} \mathbf{C}'. \tag{2.5}$$

We will denote the space $\mathscr{U}$ with the element $\boldsymbol{\Lambda}\in\mathscr{U}^{\wedge 4}$ given as part of its structure by $(\mathscr{U},\boldsymbol{\Lambda})$. With the aid of $\hat{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Lambda}$ we can now define inner products in $\mathscr{U}^{\wedge 2}$ and $\mathscr{U}'^{\wedge 2}$ by means of the equations

$$\mathbf{A}\odot\mathbf{B}=\tfrac{1}{2}\mathbf{A} \substack{\circ\\ \circ} \hat{\boldsymbol{\Lambda}} \substack{\circ\\ \circ} \mathbf{B}=\star\mathbf{A} \substack{\circ\\ \circ} \mathbf{B}=\mathbf{A} \substack{\circ\\ \circ} \star\mathbf{B}$$

$$(=A^{[\alpha\beta]}B_{[\alpha\beta]} \text{ in Penrose's notation}), \tag{2.6}$$

$$\hat{\mathbf{A}}\odot\hat{\mathbf{B}}=\tfrac{1}{2}\,\hat{\mathbf{A}} \substack{\circ\\ \circ} \boldsymbol{\Lambda} \substack{\circ\\ \circ} \hat{\mathbf{B}}=\star\hat{\mathbf{A}} \substack{\circ\\ \circ} \hat{\mathbf{B}}=\hat{\mathbf{A}} \substack{\circ\\ \circ} \star\hat{\mathbf{B}}$$

$$(=\bar{A}_{[\alpha\beta]}\bar{B}^{[\alpha\beta]} \text{ in Penrose's notation}). \tag{2.7}$$

TABLE I. Spinor and spin-tensor spaces.

Space	Notation and expressions of elements of space: Ours[10,14]	Notation and expressions of elements of space: Penrose[13]	Inner products and properties: Ours[10,14]	Inner products and properties: Penrose[13]	Bases and reciprocal bases: Ours[10,14]
Spinor space $\mathscr{S}_2$	$\boldsymbol{\omega},\boldsymbol{\pi},\boldsymbol{\lambda},\boldsymbol{\mu},\cdots$ $\boldsymbol{\omega}^*,\boldsymbol{\pi}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*,\cdots$ $\boldsymbol{\omega}=\omega^A\mathbf{h}_A=\omega_A\mathbf{h}^A$ $\boldsymbol{\omega}^*=\bar{\omega}^A\mathbf{h}_A=\bar{\omega}_A\mathbf{h}^A$ $(A=1,2)$	$\omega^A,\pi^A,\lambda^A,\mu^A,\cdots \quad (A=0,1)$ $\bar{\omega}^A,\bar{\pi}^A,\bar{\lambda}^A,\bar{\mu}^A,\cdots$	$\boldsymbol{\omega}\blacktriangle\boldsymbol{\pi}=-\boldsymbol{\pi}\blacktriangle\boldsymbol{\omega}$ $\boldsymbol{\omega}^*\blacktriangle\boldsymbol{\pi}^*=-\boldsymbol{\pi}^*\blacktriangle\boldsymbol{\omega}^*$	$\omega_A\pi^A=-\omega^A\pi_A$ $\bar{\omega}_A\bar{\pi}^A=-\bar{\omega}^A\bar{\pi}_A$	$\mathbf{h}_1,\mathbf{h}_2$ and $\mathbf{h}^1,\mathbf{h}^2$ $\mathbf{h}^A\blacktriangle\mathbf{h}_B=\delta^A_B$
Conjugate spinor space $\overline{\mathscr{S}_2}$	$\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\pi}},\bar{\boldsymbol{\lambda}},\bar{\boldsymbol{\mu}},\cdots$ $\bar{\boldsymbol{\omega}}^*,\bar{\boldsymbol{\pi}}^*,\bar{\boldsymbol{\lambda}}^*,\bar{\boldsymbol{\mu}}^*,\cdots$ $\bar{\boldsymbol{\omega}}=\bar{\omega}^{A'}\bar{\mathbf{h}}_{A'}=\bar{\omega}_{A'}\bar{\mathbf{h}}^{A'}$ $\bar{\boldsymbol{\omega}}^*=\omega^{A'}\bar{\mathbf{h}}_{A'}=\omega_{A'}\bar{\mathbf{h}}^{A'}$ $\overline{(\alpha\boldsymbol{\omega}+\beta\boldsymbol{\pi})}=\bar{\alpha}\bar{\boldsymbol{\omega}}+\bar{\beta}\bar{\boldsymbol{\pi}}$ for $\alpha,\beta\in\mathrm{C}$	$\bar{\omega}^{A'},\bar{\pi}^{A'},\bar{\lambda}^{A'},\bar{\mu}^{A'},\cdots$ $\omega^{A'},\pi^{A'},\lambda^{A'},\mu^{A'},\cdots$ $\overline{(\alpha\omega^A+\beta\pi^A)}=\bar{\alpha}\bar{\omega}^{A'}+\bar{\beta}\bar{\pi}^{A'}$	$\bar{\boldsymbol{\omega}}\blacktriangle\bar{\boldsymbol{\pi}}=-\bar{\boldsymbol{\pi}}\blacktriangle\bar{\boldsymbol{\omega}}$ $\bar{\boldsymbol{\omega}}^*\blacktriangle\bar{\boldsymbol{\pi}}^*=-\bar{\boldsymbol{\pi}}^*\blacktriangle\bar{\boldsymbol{\omega}}^*$	$\bar{\omega}_{A'}\bar{\pi}^{A'}=-\bar{\omega}^{A'}\bar{\pi}_{A'}$ $\omega_{A'}\pi^{A'}=-\omega^{A'}\pi_{A'}$ $\overline{\omega_A}=\bar{\omega}_{A'},\ \overline{\omega^A}=\bar{\omega}^{A'}$ $\overline{\omega^*_A}=\omega_{A'},\ \overline{\omega^{*A}}=\omega^{A'}$	$\bar{\mathbf{h}}_1',\bar{\mathbf{h}}_2'$ and $\bar{\mathbf{h}}^{1'},\bar{\mathbf{h}}^{2'}$ $\bar{\mathbf{h}}^{A'}\blacktriangle\bar{\mathbf{h}}_{B'}=\delta^{A'}_{B'}$
Bispinors $\mathscr{S}_4=\mathscr{S}_2\oplus\overline{\mathscr{S}_2}$	$\mathbf{u}=\boldsymbol{\omega}+\bar{\boldsymbol{\pi}}$ $=\omega^A\mathbf{h}_A+\bar{\pi}^{A'}\bar{\mathbf{h}}_{A'}$ $=\omega_A\mathbf{h}^A+\bar{\pi}_{A'}\bar{\mathbf{h}}^{A'}$ $=\omega^A\mathbf{h}_A+\pi_{A'}\bar{\mathbf{h}}^{A'}$ $\mathbf{s}=\boldsymbol{\lambda}+\bar{\boldsymbol{\mu}}=\lambda^A\mathbf{h}_A+\bar{\mu}_{A'}\bar{\mathbf{h}}^{A'}$ $\mathbf{v}=\boldsymbol{\lambda}+\bar{\boldsymbol{\mu}}^*=\lambda^A\mathbf{h}_A+\mu_{A'}\bar{\mathbf{h}}^{A'}$ $\bar{\mathbf{u}}=\bar{\boldsymbol{\omega}}+\boldsymbol{\pi};\bar{\bar{\mathbf{u}}}=\mathbf{u}$	$u^\alpha=(\omega^0,\omega^1,\bar{\pi}^{0'},\bar{\pi}^{1'})$ $u_\alpha=(\omega_0,\omega_1,\bar{\pi}_{0'},\bar{\pi}_{1'})$ $\overline{u^\alpha}=(\overline{\omega^0},\overline{\omega^1},\overline{\bar{\pi}^{0'}},\overline{\bar{\pi}^{1'}})$ $=(\bar{\omega}^{0'},\bar{\omega}^{1'},\pi^0,\pi^1)$	$\mathbf{u}\blacktriangle\mathbf{s}=-\mathbf{s}\blacktriangle\mathbf{u}=\boldsymbol{\omega}\blacktriangle\boldsymbol{\lambda}$ $+\bar{\boldsymbol{\pi}}\blacktriangle\bar{\boldsymbol{\mu}}$ $\bar{\mathbf{u}}\blacktriangle\bar{\mathbf{s}}=\overline{\mathbf{u}\blacktriangle\mathbf{s}}$	$u_\alpha s^\alpha=-u^\alpha s_\alpha$ $=\omega_A\lambda^A+\bar{\pi}_{A'}\bar{\mu}^{A'}$ $\overline{u_\alpha}\,\overline{s^\alpha}=\overline{u_\alpha s^\alpha}$ $=\bar{\omega}_{A'}\bar{\lambda}^{A'}+\pi_A\mu^A$	$\mathbf{l}_\alpha$ and $\mathbf{l}^\alpha$ $(\alpha=1,\ldots,4)$ $\mathbf{l}^\alpha\blacktriangle\mathbf{l}_\beta=\delta^\alpha_\beta$ for example $\mathbf{l}_1=\mathbf{h}_1,\mathbf{l}_2=\mathbf{h}_2,\mathbf{l}_3=\bar{\mathbf{h}}^{1'},\mathbf{l}_4=\bar{\mathbf{h}}^{2'}$ $\mathbf{l}^1=\mathbf{h}^1,\mathbf{l}^2=\mathbf{h}^2,\mathbf{l}^3=-\bar{\mathbf{h}}_1',\mathbf{l}^4=-$
$\mathscr{S}_2\otimes\mathscr{S}_2$	$\boldsymbol{\Psi}=\boldsymbol{\omega}\boldsymbol{\pi}+\cdots$ $\boldsymbol{\Phi}=\boldsymbol{\lambda}\boldsymbol{\mu}+\cdots$ $\mathbf{I}_2=\mathbf{h}_A\mathbf{h}^A=-\mathbf{h}^A\mathbf{h}_A$ $\mathbf{I}_2\blacktriangle\boldsymbol{\omega}=\boldsymbol{\omega},\ \mathbf{I}_2\blacktriangle\bar{\boldsymbol{\pi}}=0$ ($\mathbf{I}_2$ unit tensor in $\mathscr{S}_2\otimes\mathscr{S}_2$)	$\Psi^{AB}=\omega^A\pi^B+\cdots;\Psi_{AB}=\omega_A\pi_B+\cdots$ $\epsilon^{AB}=\epsilon_{AB},\ \epsilon_A{}^B=\delta^A_B$ $\omega^A=\epsilon^{AB}\omega_B$ $\omega_A=\omega^B\epsilon_{BA}$	$\boldsymbol{\Psi}\stackrel{\blacktriangle}{\blacktriangle}\boldsymbol{\Phi}=(\boldsymbol{\omega}\blacktriangle\boldsymbol{\lambda})(\boldsymbol{\pi}\blacktriangle\boldsymbol{\mu})+\cdots$ $\mathbf{I}_2\stackrel{\blacktriangle}{\blacktriangle}\mathbf{I}_2=2$	$\Psi^{AB}\Phi_{AB}=(\omega^A\lambda_A)$ $(\pi^B\mu_B)+\cdots$ $\epsilon^{AB}\epsilon_{AB}=-\epsilon^A{}_A=\epsilon_A{}^A=2$	$\mathbf{h}_A\mathbf{h}_B$ and $\mathbf{h}^A\mathbf{h}^B$ $\mathbf{h}_A\mathbf{h}^B$ and $\mathbf{h}^A\mathbf{h}_B$
$\overline{\mathscr{S}_2}\otimes\overline{\mathscr{S}_2}$	$\bar{\boldsymbol{\Psi}}=\bar{\boldsymbol{\omega}}\bar{\boldsymbol{\pi}}+\cdots$ $\bar{\boldsymbol{\Phi}}=\bar{\boldsymbol{\lambda}}\bar{\boldsymbol{\mu}}+\cdots$ $\bar{\mathbf{I}}_2=\bar{\mathbf{h}}_{A'}\bar{\mathbf{h}}^{A'}=-\bar{\mathbf{h}}^{A'}\mathbf{h}_{A'}$ $\bar{\mathbf{I}}_2\blacktriangle\boldsymbol{\omega}=0,\ \bar{\mathbf{I}}_2\blacktriangle\bar{\boldsymbol{\pi}}=\bar{\boldsymbol{\pi}}$ ($\bar{\mathbf{I}}_2$ unit tensor in $\overline{\mathscr{S}_2}\otimes\overline{\mathscr{S}_2}$)	$\bar{\Psi}^{A'B'}=\bar{\omega}^{A'}\bar{\pi}^{B'}+\cdots$ $\bar{\Phi}_{A'B'}=\bar{\lambda}_{A'}\bar{\mu}_{B'}+\cdots$ $\epsilon^{A'B'}=\epsilon_{A'B'},\ \epsilon_{A'}{}^{B'}=\delta^{B'}_{A'}$ $\bar{\omega}^{A'}=\epsilon^{A'B'}\bar{\omega}_{B'}$ $\bar{\omega}_{A'}=\bar{\omega}^{B'}\epsilon_{B'A'}$	$\bar{\boldsymbol{\Psi}}\stackrel{\blacktriangle}{\blacktriangle}\bar{\boldsymbol{\Phi}}=(\bar{\boldsymbol{\omega}}\blacktriangle\bar{\boldsymbol{\lambda}})(\bar{\boldsymbol{\pi}}\blacktriangle\bar{\boldsymbol{\mu}})+\cdots$ $\bar{\mathbf{I}}_2\stackrel{\blacktriangle}{\blacktriangle}\bar{\mathbf{I}}_2=2$	$\bar{\Psi}^{A'B'}\bar{\Phi}_{A'B'}$ $=(\bar{\omega}^{A'}\bar{\lambda}_{A'})(\bar{\pi}^{B'}\bar{\mu}_{B'})$ $+\cdots$ $\epsilon^{A'B'}\epsilon_{A'B'}$ $=-\epsilon^{A'}{}_{A'}=\epsilon_{A'}{}^{A'}=2$	$\bar{\mathbf{h}}_{A'}\bar{\mathbf{h}}_{B'}$ and $\bar{\mathbf{h}}^{A'}\bar{\mathbf{h}}^{B'}$ $\bar{\mathbf{h}}_{A'}\bar{\mathbf{h}}^{B'}$ and $\bar{\mathbf{h}}^{A'}\bar{\mathbf{h}}_{B'}$
$\overline{\mathscr{S}_2}\otimes\mathscr{S}_2$	$\mathbf{A}=\bar{\boldsymbol{\omega}}\boldsymbol{\pi}+\cdots;\ \mathbf{B}=\bar{\boldsymbol{\lambda}}\boldsymbol{\mu}+\cdots$	$A^{A'B}=\bar{\omega}^{A'}\pi^B+\cdots$ $B_{A'B}=\bar{\lambda}_{A'}\mu_B+\cdots$	$\mathbf{A}\stackrel{\blacktriangle}{\blacktriangle}\mathbf{B}=(\bar{\boldsymbol{\omega}}\blacktriangle\bar{\boldsymbol{\lambda}})(\boldsymbol{\pi}\blacktriangle\boldsymbol{\mu})+\cdots$		$\bar{\mathbf{h}}_{A'}\mathbf{h}_B$ and $\bar{\mathbf{h}}^{A'}\mathbf{h}_B$ $\bar{\mathbf{h}}_{A'}\mathbf{h}^B$ and $\bar{\mathbf{h}}^{A'}\mathbf{h}_B$
$\mathscr{S}_2\otimes\overline{\mathscr{S}_2}$	$\tilde{\mathbf{A}}=\boldsymbol{\pi}\bar{\boldsymbol{\omega}}+\cdots;\tilde{\mathbf{B}}=\boldsymbol{\mu}\bar{\boldsymbol{\lambda}}+\cdots$	$A^{BA'}=\pi^B\bar{\omega}^{A'}+\cdots$ $B_{BA'}=\mu_B\bar{\lambda}_{A'}+\cdots$	$\tilde{\mathbf{A}}\stackrel{\blacktriangle}{\blacktriangle}\tilde{\mathbf{B}}=\mathbf{A}\stackrel{\blacktriangle}{\blacktriangle}\mathbf{B}$ $\tilde{\mathbf{A}}\stackrel{\blacktriangle}{\blacktriangle}\mathbf{B}=0,\ \mathbf{A}\stackrel{\blacktriangle}{\blacktriangle}\tilde{\mathbf{B}}=0$		$\mathbf{h}_A\bar{\mathbf{h}}_{B'}$ and $\mathbf{h}^A\bar{\mathbf{h}}^{B'}$ $\mathbf{h}_A\bar{\mathbf{h}}^{B'}$ and $\mathbf{h}^A\bar{\mathbf{h}}_{B'}$
$\mathscr{M}_4=\overline{\mathscr{S}_2}\otimes_H\mathscr{S}_2$ subspace of Hermitian tensors in $\overline{\mathscr{S}_2}\otimes\mathscr{S}_2$	$\mathbf{x}=\mathbf{X}=\mathbf{X}^\dagger=\bar{\boldsymbol{\omega}}\boldsymbol{\omega}+\cdots$ $=\bar{\boldsymbol{\pi}}\boldsymbol{\lambda}+\bar{\boldsymbol{\lambda}}\boldsymbol{\pi}+\cdots$	$X^{A'B}=\bar{\omega}^{A'}\omega^B+\cdots$ $=\bar{\pi}^{A'}\lambda^B+\bar{\lambda}^{A'}\pi^B+\cdots$ $Y_{A'B}=\bar{\rho}_{A'}\mu_B+\bar{\mu}_{A'}\rho_B+\cdots$	$\mathbf{x}\cdot\mathbf{y}=-\mathbf{X}\stackrel{\blacktriangle}{\blacktriangle}\mathbf{Y}$ $=-(\bar{\boldsymbol{\pi}}\blacktriangle\bar{\boldsymbol{\rho}})(\boldsymbol{\lambda}\blacktriangle\boldsymbol{\mu})$ $-(\bar{\boldsymbol{\lambda}}\blacktriangle\bar{\boldsymbol{\mu}})(\boldsymbol{\pi}\blacktriangle\boldsymbol{\rho})+\cdots$	$X^{A'B}Y_{A'B}=(\bar{\pi}^{A'}\bar{\rho}_{A'})$ $\times(\lambda^A\mu_A)+(\pi^A\rho_A)$ $\times(\bar{\lambda}^{A'}\bar{\mu}_{A'})+\cdots$	$\mathbf{e}_\mu\equiv\mathbf{E}_\mu$ and $\mathbf{e}^\mu\equiv\mathbf{E}^\mu(\mu=0,\ldots,3)$ $\mathbf{e}^\mu\cdot\mathbf{e}_\nu=\delta^\mu_\nu=-\mathbf{E}^\mu\stackrel{\blacktriangle}{\blacktriangle}\mathbf{E}_\nu$ $\mathbf{e}_\mu\cdot\mathbf{e}_\nu=-\mathbf{E}_\mu\stackrel{\blacktriangle}{\blacktriangle}\mathbf{E}_\nu=g_{\mu\nu}$ $\mathbf{E}_0=(2)^{-1/2}(\bar{\mathbf{h}}_1'\mathbf{h}_1+\bar{\mathbf{h}}_2'\mathbf{h}_2)$ $\mathbf{E}_1=-(2)^{-1/2}(\bar{\mathbf{h}}_1'\mathbf{h}_2+\bar{\mathbf{h}}_2'\mathbf{h}_1)$ $\mathbf{E}_2=-i(2)^{-1/2}(\bar{\mathbf{h}}_1'\mathbf{h}_2-\bar{\mathbf{h}}_2'\mathbf{h}_1)$ $\mathbf{E}_3=-(2)^{-1/2}(\bar{\mathbf{h}}_1'\mathbf{h}_1-\bar{\mathbf{h}}_2'\mathbf{h}_2)$
$\mathscr{S}_4\otimes\mathscr{S}_4$	$\mathbf{S}=\mathbf{u}\mathbf{v}+\cdots$ $\mathbf{T}=\mathbf{s}\mathbf{t}+\cdots$ $\bar{\mathbf{S}}=\bar{\mathbf{u}}\bar{\mathbf{v}}+\cdots$ $S^\dagger=\tilde{\bar{S}}=\bar{v}\bar{u}+\cdots$	$S^{\alpha\beta}=u^\alpha v^\beta+\cdots$ $T_{\alpha\beta}=s_\alpha t_\beta+\cdots$ $\overline{S^{\alpha\beta}}=\overline{u^\alpha v^\beta}+\cdots$ $\overline{\tilde{S}^{\alpha\beta}}=\overline{v^\beta u^\alpha}+\cdots$	$\mathbf{S}\stackrel{\blacktriangle}{\blacktriangle}\mathbf{T}=(\mathbf{u}\blacktriangle\mathbf{s})(\mathbf{v}\blacktriangle\mathbf{t})+\cdots$	$S^{\alpha\beta}T_{\alpha\beta}=(u^\alpha s_\alpha)$ $\times(v^\beta t_\beta)+\cdots$	$\mathbf{l}_\alpha\mathbf{l}_\beta$ and $\mathbf{l}^\alpha\mathbf{l}^\beta$ $\mathbf{l}_\alpha\mathbf{l}^\beta$ and $\mathbf{l}^\alpha\mathbf{l}_\beta$

TABLE II. Twistor and twist-tensor spaces.

Space	Notation and expressions of elements of space: Ours[10,14]	Penrose[13]
Twistor space $\mathscr{U} \equiv \mathscr{U}_{2,2}$	$\mathbf{a},\mathbf{b},\ldots,\mathbf{m},\mathbf{l},\ldots,\mathbf{s},\mathbf{t},\ldots$ $\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_m,\ldots$; Dirac bispinors of the form $\mathbf{u} = \boldsymbol{\omega} + \bar{\boldsymbol{\pi}}^* = \omega^A \mathbf{h}_A + \pi_{A'} \bar{\mathbf{h}}^{A'}$	$Z^\alpha \Leftrightarrow (\omega^A, \pi_{A'}) = (\omega^0, \omega^1, \pi'_0, \pi'_1)$
	adjoint twistor $\bar{\mathbf{a}},\bar{\mathbf{b}},\ldots,\bar{\mathbf{m}},\bar{\mathbf{l}},\ldots$; adjoint Dirac bispinors of the form $\bar{\mathbf{u}} = \bar{\boldsymbol{\omega}} + \boldsymbol{\pi}^* = \bar{\omega}^{A'} \bar{\mathbf{h}}_{A'} + \bar{\pi}_A \mathbf{h}^A$	not defined
	not defined	$\bar{Z}^{\bar{\alpha}} = (\bar{\omega}^{0'}, \bar{\omega}^{1'}, \bar{\pi}_0, \bar{\pi}_1)$
Dual twistor space $\mathscr{U}' \equiv \mathscr{U}'_{2,2}$ (the set of linear	$\mathbf{a}',\mathbf{b}',\ldots,\mathbf{m}',\mathbf{l}',\ldots$ $\mathbf{t}'_1,\mathbf{t}'_2,\ldots,\mathbf{t}'_m,\ldots$; dual Dirac bispinors of the form $\mathbf{u}' = \omega^A \hat{\mathbf{h}}_A + \bar{\pi}_{A'} \hat{\bar{\mathbf{h}}}^{A'}$	not defined
functionals on $\mathscr{U}$)	conjugate twistor $\hat{\mathbf{a}},\hat{\mathbf{b}},\ldots,\hat{\mathbf{l}},\hat{\mathbf{m}},\ldots$; dual Dirac bispinors of the form $\hat{\mathbf{u}} = \bar{\omega}^{A'} \hat{\bar{\mathbf{h}}}_{A'} - \bar{\pi}_A \hat{\mathbf{h}}^A$ given by the antilinear map $\mathbf{u} \in \mathscr{U} \to \hat{\mathbf{u}} \in \mathscr{U}'$	$\bar{Z}_\alpha = (\bar{\pi}_0, \bar{\pi}_1, \bar{\omega}^{0'}, \bar{\omega}^{1'})$
$\mathscr{U}^{\otimes m} = \otimes^m \mathscr{U}$ (mth tensorial power of $\mathscr{U}$)	$\mathbf{M} = \mathbf{t}_1 \otimes \mathbf{t}_2 \otimes \cdots \otimes \mathbf{t}_m + \cdots$	$M^{\alpha_1\alpha_2\cdots\alpha_m} = Z^{\alpha_1} Z^{\alpha_2} \cdots Z^{\alpha_m} + \cdots$
$\mathscr{U}'^{\otimes k} = \otimes^k \mathscr{U}'$	$\mathbf{K}' = \mathbf{k}'_1 \otimes \mathbf{k}'_2 \otimes \cdots \otimes \mathbf{k}'_k + \cdots$	not defined
	$\hat{\mathbf{K}} = \hat{\mathbf{k}}_1 \otimes \hat{\mathbf{k}}_2 \otimes \cdots \otimes \hat{\mathbf{k}}_k + \cdots$	$\bar{K}_{\beta_1\beta_2\cdots\beta_k} = \bar{Y}_{\beta_1} \bar{Y}_{\beta_2} \cdots \bar{Y}_{\beta_k} + \cdots$
$\mathscr{U}^{\wedge m} \equiv \wedge^m \mathscr{U}$ (mth exterior power of $\mathscr{U}$)	$\mathbf{B} = \mathbf{t}_1 \wedge \cdots \wedge \mathbf{t}_m + \cdots$ $\Lambda \in \mathscr{U}^{\wedge 4}$ (alternating twistors)	$B^{[\alpha_1\alpha_2\cdots\alpha_m]} = Z^{[\alpha_1} Z^{\alpha_2} \cdots Z^{\alpha_m]} + \cdots$ $\epsilon^{\alpha\beta\gamma\delta} = \epsilon^{[\alpha\beta\gamma\delta]}, \epsilon^{0123} = 1$
$\mathscr{U}'^{\wedge k} = \wedge^k \mathscr{U}'$	$\mathbf{C}' = \mathbf{k}'_1 \wedge \mathbf{k}'_2 \wedge \cdots \wedge \mathbf{k}'_k$ $\Lambda' \in \mathscr{U}'^{\wedge 4}$ (alternating twistors)	$P_{[\alpha\beta]} = \frac{1}{2}\epsilon_{\alpha\beta\gamma\delta} P^{\gamma\delta}$ $\epsilon_{\alpha\beta\gamma\delta} = \epsilon_{[\alpha\beta\gamma\delta]}, \epsilon_{0123} = 1$
	$\hat{\mathbf{C}} = \hat{\mathbf{k}}_1 \wedge \hat{\mathbf{k}}_2 \wedge \cdots \wedge \hat{\mathbf{k}}_k + \cdots$ $\hat{\Lambda} = \Lambda' \in \mathscr{U}'^{\wedge 4}$ (alternating twistors)	$\bar{C}_{[\beta_1\beta_2\cdots\beta_k]} = \bar{Y}_{[\beta_1} \bar{Y}_{\beta_2} \cdots \bar{Y}_{\beta_k]} + \cdots$ $\epsilon_{\alpha\beta\gamma\delta} = \epsilon_{[\alpha\beta\gamma\delta]}, \epsilon_{0123} = 1$
$\mathscr{E} = \mathscr{E}_{2,4}$	${}^*\hat{\mathbf{P}} = \mathbf{P}$	$\bar{P}^{[\alpha\beta]} \equiv \frac{1}{2}\epsilon^{\alpha\beta\gamma\delta}\bar{P}_{\gamma\delta} = P^{[\alpha\beta]}$
$= \{P \mid \mathbf{P} \in \mathscr{U}^{\wedge 2}, \mathbf{P} \text{ real}\}$	${}^*\hat{\mathbf{I}} = \mathbf{I} = \mathbf{I}_2 = \mathbf{h}_A \mathbf{h}^A$ (infinity twistor)	$I^{\alpha\beta} \leftrightarrow \begin{bmatrix} \epsilon^{AB} & 0 \\ 0 & 0 \end{bmatrix}, I_{\alpha\beta} \leftrightarrow \begin{bmatrix} 0 & 0 \\ 0 & \epsilon^{A'B'} \end{bmatrix}$
	${}^*\hat{\mathbf{O}} = \mathbf{O} = \bar{\mathbf{I}}_2 = \bar{\mathbf{h}}_{A'} \bar{\mathbf{h}}^{A'}$ (origin twistor)	$\mathbf{O}^{\alpha\beta} \leftrightarrow \begin{bmatrix} 0 & 0 \\ 0 & \epsilon^{A'B'} \end{bmatrix}, \mathbf{O}_{\alpha\beta} = \begin{bmatrix} \epsilon^{AB} & 0 \\ 0 & 0 \end{bmatrix}$

A twist-tensor $\mathbf{B} \in \mathscr{U}^2$ for which

$$\mathbf{B} \odot \mathbf{B} = 0 \tag{2.8}$$

is called null, and twist-tensors in $\mathscr{U}^{\wedge 2}$ of the form $\mathbf{B} = \mathbf{l} \wedge \mathbf{m}$ are called simple and satisfy the property

$$\mathbf{B} \circ \star\mathbf{B} = 0. \tag{2.9}$$

Of special importance to establish a correspondence between twistors and Minkowski space is the subspace of twist-tensors in $\mathscr{U}^{\wedge 2}$ with inner product $\odot$ and for which the condition

$$\star\hat{\mathbf{P}} = \mathbf{P} \leftrightarrow \hat{\mathbf{P}} = \tfrac{1}{2}\hat{\Lambda} \mathbin{\overset{\circ}{\circ}} \mathbf{P} \ (\bar{P}_{[\alpha\beta]} = P_{[\alpha\beta]} \text{ in Penrose's notation}) \tag{2.10}$$

is satisfied. These twist-tensors are called real and the subspace will be denoted by $\mathscr{E} = \mathscr{E}_{2,4} = \{\mathbf{P} \mid \mathbf{P} \in \mathscr{U}^{\wedge 2}, \mathbf{P} \text{ real}\}$. The inner product $\odot$ in $\mathscr{E}$ has signature $(+ + - - - -)$.

Null, real twist-tensors form a null cone in the six-dimensional $\mathscr{U}^{\wedge 2}$ space. Among these, we distinguish a privileged element $\mathbf{I}$, which is simple and invariant under the action of the Poincaré group. $\mathbf{I}$ can then be identified with the vertex of the null cone at infinity and is known as the infinity twistor (or metric twistor). It can be readily shown that

$$\mathbf{I} = \mathbf{I}_2 = \mathbf{h}_A \mathbf{h}^A = -\mathbf{h}_1 \wedge \mathbf{h}_2$$
$$(= \text{the unit tensor in } \mathscr{S}_2 \otimes \mathscr{S}_2). \tag{2.11}$$

Correspondence with Minkowski space: Introducing $\mathbf{I}$ as part of the structure of $(\mathscr{U}, \Lambda)$, i.e., retaining only that group of linear transformations which leave $\mathbf{I}$ invariant, breaks the conformal invariance of $\mathscr{U}$ and leads to a faithful representation space of the Poincaré group $\mathscr{P}$. This subspace will be denoted by $(\mathscr{U}, \Lambda, \mathbf{I})$.

Moreover, introducing $\mathbf{I}$ as part of the structure of $\mathscr{E}$, we can construct a hypersurface

$$\mathscr{W} = \{\mathbf{P} \mid \mathbf{P} \in \mathscr{E},\ \mathbf{P} \odot \mathbf{P} = 0,\ \mathbf{I} \odot \mathbf{P} = 2\},$$

which has a one-to-one correspondence with the elements of Minkowski space–time. This hypersurface is the intersection of the null cone in $\mathscr{E}$ and the plane $\mathbf{I} \odot \mathbf{P} = 2$, and is invariant under the action of $\mathscr{P} \otimes \mathscr{P}$.

The tangent space $\mathscr{W}_{\mathbf{P}}$ at a given element $\mathbf{P} \in \mathscr{W}$ is the set of elements $\mathbf{T} \in \mathscr{E}$ which are tangent to the hypersurface

TABLE III. Scalar and inner products in twistor and twist-tensor spaces.

Ours[10,14]	Penrose[13]
	Inner product in $\mathscr{U}$
$\langle \mathbf{s}\mid\mathbf{t}\rangle = \langle \mathbf{t}\mid\mathbf{s}\rangle^*$ (Hermiticity) signature $(+\ +\ -\ -)$	not defined
	Scalar product between $\mathscr{U}'$ and $\mathscr{U}$
$\mathbf{k}'\circ\mathbf{l} = \mathbf{l}\circ\mathbf{k}'$ = complex number $\hat{\mathbf{l}}\circ\mathbf{m} = \mathbf{m}\circ\hat{\mathbf{l}} \equiv \langle \mathbf{l}\mid\mathbf{m}\rangle$ $\hat{\mathbf{l}}\circ\mathbf{m} = (\mathbf{l}\circ\hat{\mathbf{m}})^* = \bar{\mathbf{l}}\blacktriangle(\mathbf{I}-\mathbf{O})\blacktriangle\mathbf{m}$	not defined $Z^\alpha\bar{Y}_\alpha = \omega^A\bar{\mu}_A + \pi_{A'}\bar{\lambda}^{A'}$ $Z^\alpha\bar{Y}_\alpha = (\overline{Y^\alpha\bar{Z}_\alpha})$
$\hat{\mathbf{h}}_{B'}\circ\bar{\mathbf{h}}_{A'} = \langle \mathbf{h}_{B'}\mid\bar{\mathbf{h}}_{A'}\rangle = -\epsilon_{B'A'}$ $\hat{\mathbf{h}}_B\circ\mathbf{h}_A = \langle \mathbf{h}_B\mid\bar{\mathbf{h}}_A\rangle^* = -\epsilon_{BA}$ $\hat{\mathbf{h}}_{B'}\circ\mathbf{h}_A = \langle \mathbf{h}_{B'}\mid\mathbf{h}_A\rangle = 0$ $\hat{\mathbf{h}}_B\circ\bar{\mathbf{h}}_{A'} = \langle \mathbf{h}_B\mid\mathbf{h}_{A'}\rangle^* = 0$	
	Inner product in $\mathscr{U}^{\otimes m}$
$\langle \mathbf{M}\mid\mathbf{N}\rangle = \langle \mathbf{t}_1\mid\mathbf{s}_1\rangle\langle \mathbf{t}_2\mid\mathbf{s}_2\rangle\cdots\langle \mathbf{t}_m\mid\mathbf{s}_m\rangle + \cdots$	not defined
	Contraction operations between twist-tensors in $\mathscr{U}'^{\otimes k}$ and twist-tensors in $\mathscr{U}^{\otimes m}$
$\mathbf{M}\circ\mathbf{K}' = \mathbf{t}_1\otimes\cdots\otimes(\mathbf{t}_m\circ\mathbf{k}'_1)\otimes\cdots\otimes\mathbf{k}'_k + \cdots$ $\mathbf{M}\mathbin{\substack{\circ\\\circ}}\mathbf{K}' = \mathbf{t}_1\otimes\cdots\otimes(\mathbf{t}_{m-1}\circ\mathbf{k}'_1)(\mathbf{t}_m\circ\mathbf{k}'_2)\otimes\cdots\otimes\mathbf{k}'_k$	not defined
$\mathbf{M}\circ\hat{\mathbf{K}} = \mathbf{t}_1\otimes\cdots\otimes(\mathbf{t}_m\circ\hat{\mathbf{k}}_1)\otimes\cdots\otimes\hat{\mathbf{k}}_k + \cdots$ $\mathbf{M}\mathbin{\substack{\circ\\\circ}}\hat{\mathbf{K}} = \mathbf{t}_1\otimes\cdots\otimes(\mathbf{t}_{m-1}\circ\hat{\mathbf{k}}_1)(\mathbf{t}_m\circ\hat{\mathbf{k}}_2)\otimes\cdots\otimes\hat{\mathbf{k}}_k$ $\hat{\mathbf{\Lambda}}\mathbin{\substack{\circ\circ\\\circ\circ}}\mathbf{\Lambda} = 4!$ ${}^*\mathbf{B} = \frac{1}{2}\hat{\mathbf{\Lambda}}\mathbin{\substack{\circ\\\circ}}B$ ${}^{**}\mathbf{B} = \frac{1}{4}\mathbf{\Lambda}\mathbin{\substack{\circ\\\circ}}\hat{\mathbf{\Lambda}}\mathbin{\substack{\circ\\\circ}}\mathbf{B}$	$M^{\alpha_1\cdots\alpha_m}\bar{K}_{\alpha_m\beta_1\cdots\beta_k} = Z^{\alpha_1}Z^{\alpha_2}\cdots Z^{\alpha_m}\bar{Y}_{\alpha_m}\bar{Y}_{\beta_1}\cdots\bar{Y}_{\beta_k} + \cdots$ $M^{\alpha_1\cdots\alpha_{m-1}\alpha_m}\bar{K}_{\alpha_{m-1}\alpha_m\beta_1\cdots\beta_k} = Z^{\alpha_1}Z^{\alpha_2}\cdots Z^{\alpha_{m-1}}\bar{Y}_{\alpha_{m-1}}Z^{\alpha_m}\bar{Y}_{\alpha_m}\bar{Y}_{\beta_1}\cdots\bar{Y}_{\beta_k} + \cdots$ $\epsilon_{\alpha\beta\gamma\delta}\epsilon^{\alpha\beta\gamma\delta} = 4!$ $B_{[\alpha\beta]} = \frac{1}{2}\epsilon_{\alpha\beta\gamma\delta}B^{\gamma\delta}$ $B^{[\alpha\beta]} = \frac{1}{4}\epsilon^{\alpha\beta\gamma\delta}\epsilon_{\alpha\delta\lambda\tau}B^{\lambda\tau}$
	Inner product in $\mathscr{U}^{\wedge 2}$
$\mathbf{A}\odot\mathbf{B} = \frac{1}{2}\mathbf{A}\mathbin{\substack{\circ\\\circ}}\hat{\mathbf{\Lambda}}\mathbin{\substack{\circ\\\circ}}\mathbf{B} = {}^*\mathbf{A}\mathbin{\substack{\circ\\\circ}}\mathbf{B} = \mathbf{A}\mathbin{\substack{\circ\\\circ}}{}^*\mathbf{B}$	$A^{[\alpha\beta]}B_{[\alpha\beta]} = \frac{1}{2}A^{\alpha\beta}\epsilon_{\alpha\beta\gamma\delta}B^{\gamma\delta} = A_{[\alpha\beta]}B^{[\alpha\beta]}$
	Contraction operations between twist-tensors in $\mathscr{U}^{\wedge 2}$ and twist-tensor in $\mathscr{U}'^{\wedge k}$
real twistors ${}^*\hat{\mathbf{P}} = \mathbf{P} \leftrightarrow \hat{\mathbf{P}} = \frac{1}{2}\hat{\mathbf{\Lambda}}\mathbin{\substack{\circ\\\circ}}\mathbf{P}$	$\bar{P}_{[\alpha\beta]} = \frac{1}{2}\epsilon_{\alpha\beta\gamma\delta}P^{\gamma\delta} = P_{[\alpha\beta]}$

$\mathscr{W}$ at $\mathbf{P}$. It follows that $\mathscr{W}_{\mathbf{P}} = \{\mathbf{T}\mid\mathbf{T}\in\mathscr{E},\ \mathbf{I}\odot\mathbf{T} = 0,\ \mathbf{P}\odot\mathbf{T} = 0\}$. The inner product $\odot$ in $\mathscr{W}_{\mathbf{P}}$ is the Minkowski inner product with signature $(+\ -\ -\ -)$. With this inner product, $\mathscr{W}$ becomes a pseudo-Riemannian space with a curvature tensor which is zero everywhere, i.e., $\mathscr{W}$ is intrinsically flat.

Note that if we further choose another element $\mathbf{O}$ of $\mathscr{W}$ as a reference point (origin twistor), and we require it to be a part of the structure of $(\mathscr{U}, \mathbf{\Lambda}, \mathbf{I})$, i.e., we retain only the group of linear transformations that leave $\mathbf{I}$ and $\mathbf{O}$ invariant, then the Poincaré invariance is broken and we are left with the subspace $(\mathscr{U}, \mathbf{\Lambda}, \mathbf{I}, \mathbf{O})$, which is invariant under the action of the Lorentz group. In the Appendix we show how the structure of this space leads to a unique bilinear antisymmetric inner product in $\mathscr{U}$, an adjoint operation in $\mathscr{U}$, and a unique decomposition $\mathscr{U} = \mathscr{S}_2 \oplus \bar{\mathscr{S}}_2$, which serve to relate twistors to Dirac bispinors.

It is also easy to show that

$$\mathbf{O} = \bar{\mathbf{I}}_2 = \bar{\mathbf{h}}_{A'}\bar{\mathbf{h}}^{A'} = -\bar{\mathbf{h}}'_1 \wedge \bar{\mathbf{h}}'_2 \quad \text{(the unit tensor in } \bar{\mathscr{S}}_2 \otimes \bar{\mathscr{S}}_2\text{)}. \tag{2.12}$$

Some twist-tensor spaces: To conclude, we now list some of the twist-tensor spaces that will be required in the next section.

First, we use the space $\mathscr{U} = (\mathscr{U}, \mathbf{\Lambda}, \mathbf{I},)$ as a representation space for the Poincaré group $\mathscr{P}$, which preserves the given structure of $\mathscr{U}$.

The group $\mathscr{P}$ of transformations on $\mathscr{U}$ gives rise to the tensor product group $\mathscr{P}\otimes\mathscr{P}$ acting on $\mathscr{U}^{\otimes 2}$. The space $\mathscr{E} = (\mathscr{E}, \odot, \mathbf{I})$ of real twist-tensors in $\mathscr{U}^{\wedge 2}$ is invariant, and its structure is preserved under the action of $\mathscr{P}\otimes\mathscr{P}$, and thus $\mathscr{E}$ is a representation space for the group $\mathscr{P}$.

We define the two parallel five-dimensional planes $\mathscr{H} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{E}, \mathbf{I}\odot\mathbf{P} = 0\}$ and $\mathscr{K} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{E}, \mathbf{I}\odot\mathbf{P} = 2\}$, the two parallel four-dimensional planes $\mathscr{F} \equiv \mathscr{W}_{\mathbf{O}} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{H},\ \mathbf{O}\odot\mathbf{P} = 0\} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{E},\ \mathbf{I}\odot\mathbf{P} = 0,\ \mathbf{O}\odot\mathbf{P} = 0\}$ and $\mathscr{L} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{K},\ \mathbf{O}\odot\mathbf{P} = 0\} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{E},\ \mathbf{I}\odot\mathbf{P} = 2,\ \mathbf{O}\odot\mathbf{P} = 0\}$, and the null cone $\mathscr{N} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{E}, \mathbf{P}\odot\mathbf{P} = 0\}$. The surface we discussed previously is given by $\mathscr{W} = \mathscr{N}\cap\mathscr{K} = \{\mathbf{P}\mid\mathbf{P}\in\mathscr{E},\ \mathbf{P}\odot\mathbf{P} = 0,\ \mathbf{I}\odot\mathbf{P} = 2\}$. The planes $\mathscr{H}$ and $\mathscr{F}$ are closed under vector addition of their elements; thus they are vector subspaces. Figure 1 shows the relations among these surfaces.

Under the action of $\mathscr{P}\otimes\mathscr{P}$, the element $\mathbf{I}$ and the surfaces $\mathscr{H}$, $\mathscr{K}$, $\mathscr{N}$, and $\mathscr{W}$ are invariant. The element $\mathbf{O}$ is not invariant, but it remains on the invariant surface $\mathscr{W}$. The plane $\mathscr{L}$ is not invariant, but it remains contained in the invariant plane $\mathscr{K}$. The plane $\mathscr{F}$ is not invariant, but it

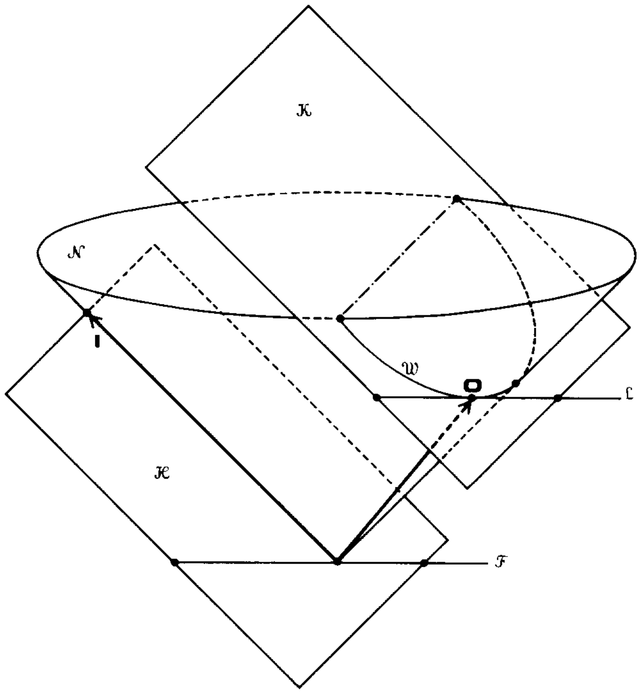


FIG. 1. Twist-tensor hypersurfaces and relation to Minkowski space–time.

remains contained in the invariant plane $\mathscr{H}$. Also note that $\mathscr{H} = \mathscr{F} \oplus \{\lambda \mathbf{I}\}$, where $\{\lambda \mathbf{I}\}$ is the one-dimensional subspace spanned by $\mathbf{I}$.

The plane $\mathscr{L}$ is tangent to the invariant surface $\mathscr{W}$ at $\mathbf{O}$. Since $\mathscr{F} \equiv \mathscr{W}_{\mathbf{O}}$ is parallel to $\mathscr{L}$ and is also a vector subspace, $\mathscr{F}$ acts as the tangent vector space to the surface $\mathscr{W}$ at $\mathbf{O}$. Likewise, at each point of $\mathscr{W}$ there is a similar tangent vector space, and the inner product in these tangent spaces provides $\mathscr{W}$ with the above-mentioned pseudo-Riemannian intrinsic structure [with signature $(+\;-\;-\;-)$ and zero curvature tensor] so that $\mathscr{W}$ has the structure of Minkowski space and thus provides a model of space–time.

III. GAUGE THEORY WITH A TWISTOR BUNDLE

The twistor space $\mathscr{U} = (\mathscr{U}, \mathbf{\Lambda}, \mathbf{I})$ will now serve as a representation space in terms of which the Poincaré group will be treated as an internal symmetry group for a gauge theory of gravitation.

We first summarize the material of this section. Starting with a four-dimensional manifold $\mathscr{M}$, we set up the twistor bundle $\mathscr{U}(\mathscr{M})$ and the twist-tensor bundle $\mathscr{E}(\mathscr{M})$. A given cross section $\mathbf{O}$ of $\mathscr{E}(\mathscr{M})$, called the origin twist-tensor field, is assumed. Given $\mathbf{O}$, a bundle $\mathscr{F}(\mathscr{M})$ is uniquely specified. Here, the spaces $\mathscr{U}$, $\mathscr{E}$, and $\mathscr{F}$ are the typical fibers of the $\mathscr{U}(\mathscr{M})$, $\mathscr{E}(\mathscr{M})$ and $\mathscr{F}(\mathscr{M})$ bundles respectively. Next a connection $\mathbf{D}$ on the $\mathscr{U}(\mathscr{M})$ bundle is assumed to be given. This connection naturally gives rise to a connection $\mathbf{D}$ (same symbol) on $\mathscr{E}(\mathscr{M})$, which in turn can be projected to give a connection $\mathbf{D}^{\mathscr{F}}$ on $\mathscr{F}(\mathscr{M})$. The action of $\mathbf{D}$ on $\mathbf{O}$ gives a field $\mathbf{J}$, which can be utilized as a map that takes cross sections of the tangent bundle $\mathscr{T}(\mathscr{M})$ into cross sections of $\mathscr{F}(\mathscr{M})$. The $\mathbf{J}$ field makes possible the mapping of other objects from $\mathscr{F}(\mathscr{M})$ such as the inner product, the connection, and the curvature tensor into corresponding objects on $\mathscr{T}(\mathscr{M})$. Since these objects are already uniquely specified on $\mathscr{F}(\mathscr{M})$, but not on $\mathscr{T}(\mathscr{M})$, the application of this map imposes a unique metric structure and connection on $\mathscr{T}(\mathscr{M})$.

The curvature tensor $\mathbf{R}_{\mathscr{E}}$ for the $\mathscr{E}(\mathscr{M})$ connection $\mathbf{D}$ is decomposed uniquely into the curvature tensor $\mathbf{R}_{\mathscr{F}}$ for the $\mathscr{F}(\mathscr{M})$ connection $\mathbf{D}^{\mathscr{F}}$ plus another term constructed out of a uniquely defined tensor $\mathbf{T}_{\mathscr{F}}$. Under the above map, $\mathbf{T}_{\mathscr{F}}$ goes into the torsion tensor for the imposed connection $\mathbf{\nabla}$ for the $\mathscr{T}(\mathscr{M})$ bundle. Finally the curvature invariant for the $\mathbf{D}^{\mathscr{F}}$ connection and for the $\mathbf{\nabla}$ connection are shown to be equal. With the imposed metric structure and connection on $\mathscr{T}(\mathscr{M})$, one is ready to set up Lagrangians for the gauge field theory. However, the Lagrangians can be equivalently expressed directly in terms of the metric structure on the $\mathscr{F}(\mathscr{M})$ bundle and the connections on the $\mathscr{U}(\mathscr{M})$ and $\mathscr{E}(\mathscr{M})$ bundles.

We construct the twistor bundle $(\mathscr{U}(\mathscr{M}), \mathscr{M}, \mathscr{U}, \pi, \mathscr{P}, \phi)$, where $\mathscr{U}(\mathscr{M})$ is the bundle space, the base space $\mathscr{M}$ is a four-dimensional manifold, the typical fiber is the twistor space $\mathscr{U} = (\mathscr{U}, \mathbf{\Lambda}, \mathbf{I})$, π is the surjective projection of $\mathscr{U}(\mathscr{M})$ onto $\mathscr{M}$, the Poincaré group $\mathscr{P}$ is the structure group of the bundle, and ϕ is a set of homeomorphisms that establishes the local triviality condition.[2] At each $q \in \mathscr{M}$, the fiber above q is $\pi^{-1} q = \mathscr{U}_q = (\mathscr{U}_q, \mathbf{\Lambda}(q), \mathbf{I}(q))$, a space with structure isomorphic to that of $\mathscr{U} = (\mathscr{U}, \mathbf{\Lambda}, \mathbf{I})$. The action of $\mathscr{P}$ on each fiber $\mathscr{U}_q$ is represented by the group of linear transformations $\mathscr{P}_q$ which preserves the structure of $\mathscr{U}_q$.

Let $\Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$ denote the space of smooth cross sections of the bundle. An element $\mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$ is a twistor field, which associates a twistor $\mathbf{u}(q) \in \mathscr{U}_q$ with each point $q \in \mathscr{M}$. Note that $\mathbf{\Lambda} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 4}(\mathscr{M}))$ is a twist-tensor field, having its value $\mathbf{\Lambda}(q)$ at q in $\mathscr{U}_q{}^{\wedge 4}$. Also $\mathbf{I} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 2}(\mathscr{M}))$ is a twist-tensor field having its value $\mathbf{I}(q)$ at q in $\mathscr{U}_q{}^{\wedge 2}$.

A twistor connection, i.e., a connection $\mathbf{D}$ on $\mathscr{U}(\mathscr{M})$, is a map $\mathbf{D}$: $\Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M})) \rightarrow \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}) \otimes \mathscr{U}(\mathscr{M}))$, where $\mathscr{T}'(\mathscr{M})$ is the cotangent bundle over $\mathscr{M}$, by means of which each $\mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$ goes into $\mathbf{D} \otimes \mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}) \otimes \mathscr{U}(\mathscr{M}))$. Defining

$$D_X \mathbf{u} = \mathbf{x} \circ (\mathbf{D} \otimes \mathbf{u}), \tag{3.1}$$

where $\mathbf{x} \equiv X \in \Gamma(\mathscr{M}, \mathscr{T}(\mathscr{M}))$, and $D_X \mathbf{u} \in \Gamma(\mathscr{M}, \mathscr{U}(\mathscr{M}))$, where $\mathscr{T}(\mathscr{M})$ is the tangent bundle over $\mathscr{M}$, then D_X satisfies the following axioms:

$$D_X(\mathbf{u} + \mathbf{v}) = D_X \mathbf{u} + D_X \mathbf{v}, \tag{3.2a}$$

$$D_X(f\mathbf{u}) = (Xf)\mathbf{u} + f(D_X \mathbf{u}), \tag{3.2b}$$

$$D_{X+Y} \mathbf{u} = D_X \mathbf{u} + D_Y \mathbf{u}, \tag{3.2c}$$

$$D_{gX} \mathbf{u} = g(D_X \mathbf{u}), \tag{3.2d}$$

where $f \in \Gamma(\mathscr{M}, \mathbb{C})$ is any smooth complex scalar field and $g \in \Gamma(\mathscr{M}, \mathbb{R})$ is any smooth real scalar field. Also

$$X(\langle \mathbf{u} | \mathbf{v} \rangle) = \langle D_X \mathbf{u} | \mathbf{v} \rangle + \langle \mathbf{u} | D_X \mathbf{v} \rangle, \tag{3.3}$$

$$D_X \mathbf{\Lambda} = 0, \tag{3.4}$$

$$D_X \mathbf{I} = 0, \tag{3.5}$$

because the structure of the fibers have to be preserved. Note that once D_X is defined on twistor fields, its action on twist-tensor fields is determined. Thus, in particular, $D_X \mathbf{\Lambda}$ and

$D_X \mathbf{I}$ are well defined.

Other various twist-tensor bundles may now be formed. One of particular interest here is $(\mathscr{E}(\mathscr{M}), \mathscr{M}, \mathscr{E}, \Pi, \mathscr{P}_{\mathscr{E}}, \Phi)$, where $\mathscr{E}(\mathscr{M})$ is the bundle space, the base space $\mathscr{M}$ is the previously mentioned four-dimensional manifold, the typical fiber is the twist-tensor space $\mathscr{E} = (\mathscr{E}, \odot, \mathbf{I})$, Π is the surjective projection of $\mathscr{E}(\mathscr{M})$ onto $\mathscr{M}$, the structure group $\mathscr{P}_{\mathscr{E}}$ is the tensor product group $\mathscr{P} \otimes \mathscr{P}$ acting on $\mathscr{E}$, and Φ is a set of homeomorphisms constructed with the aid of the set ϕ to establish the local triviality condition. At each $q \in \mathscr{M}$, the fiber above q is $\Pi^{-1} q = \mathscr{E}_q = (\mathscr{E}_q, \odot, \mathbf{I}(q))$, which is the space of real twist-tensors in $\mathscr{U}_q{}^{\wedge 2}$, and consequently its structure is isomorphic to that of $\mathscr{E} = (\mathscr{E}, \odot, \mathbf{I})$. The action of $\mathscr{P}_{\mathscr{E}}$ in each fiber $\mathscr{E}_q$ is given by the tensor product group $\mathscr{P}_q \otimes \mathscr{P}_q$.

At each $q \in \mathscr{M}$, we shall be interested in the hyperplanes $\mathscr{H}_q = \{\mathbf{P}_q | \mathbf{P}_q \in \mathscr{E}_q, \mathbf{I}(q) \odot \mathbf{P}_q = 0\}$ and $\mathscr{K}_q = \{\mathbf{P}_q | \mathbf{P}_q \in \mathscr{E}_q, \mathbf{I}(q) \odot \mathbf{P}_q = 2\}$, the null cone $\mathscr{N}_q = \{\mathbf{P}_q | \mathbf{P}_q \in \mathscr{E}_q, \mathbf{P}_q \odot \mathbf{P}_q = 0\}$, the surface $\mathscr{W}_q = \mathscr{N}_q \cap \mathscr{K}_q = \{\mathbf{P}_q | \mathbf{P}_q \in \mathscr{E}_q, \mathbf{P}_q \odot \mathbf{P}_q = 0, \mathbf{I}(q) \odot \mathbf{P}_q = 2\}$, and the plane $\mathscr{F}_q = \{\mathbf{P}_q | \mathbf{P}_q \in \mathscr{H}_q, \mathbf{O}(q) \odot \mathbf{P}_q = 0\} = \{\mathbf{P}_q | \mathbf{P}_q \in \mathscr{E}_q, \mathbf{I}(q) \odot \mathbf{P}_q = 0, \mathbf{O}(q) \odot \mathbf{P}_q = 0\}$. A choice of an origin $\mathbf{O}(q)$ in each $\mathscr{W}_q$ gives the twist-tensor field $\mathbf{O}$ appearing in the definition of $\mathscr{F}_q$.

Let $\Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ denote the space of smooth cross sections of the bundle $\mathscr{E}(\mathscr{M})$. An element $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ is a twist-tensor field, which associates a twist-tensor $\mathbf{V}(q) \in \mathscr{E}_q$ with each point $q \in \mathscr{M}$. We shall call elements of the cross section $\Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ vector fields.

The twistor-connection $\mathbf{D}$, by its action on twist-tensor fields, gives rise to a twist-tensor connection $\mathbf{D}$ on $\mathscr{E}(\mathscr{M})$. It is a map $\mathbf{D}: \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M})) \rightarrow \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}) \otimes \mathscr{E}(\mathscr{M}))$ by which each $\mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$ goes into $\mathbf{D} \otimes \mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{T}'(\mathscr{M}) \otimes \mathscr{E}(\mathscr{M}))$. Defining

$$D_X \mathbf{V} = \mathbf{x} \circ (\mathbf{D} \otimes \mathbf{V}), \tag{3.6}$$

where $D_X \mathbf{V} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$, we can show that D_X acting on the bundle $\mathscr{E}(\mathscr{M})$ has the properties

$$D_X(\mathbf{V} + \mathbf{W}) = D_X \mathbf{V} + D_X \mathbf{W}, \tag{3.7a}$$

$$D_X(g\mathbf{V}) = (Xg)\mathbf{V} + g(D_X \mathbf{V}), \tag{3.7b}$$

$$D_{X+Y} \mathbf{V} = D_X \mathbf{V} + D_Y \mathbf{V}, \tag{3.7c}$$

$$D_{gX} \mathbf{V} = g(D_X \mathbf{V}), \tag{3.7d}$$

where $g \in \Gamma(\mathscr{M}, \mathbb{R})$ is any smooth real scalar field. Also, because of (3.4),

$$X(\mathbf{V} \odot \mathbf{W}) = (D_X \mathbf{V}) \odot \mathbf{W} + \mathbf{V} \odot (D_X \mathbf{W}). \tag{3.8}$$

Note that for arbitrary twist-tensor fields $\mathbf{L} \in \Gamma(\mathscr{M}, \mathscr{U}^{\wedge 2}(\mathscr{M}))$, we have

$$\star [(D_X \mathbf{L})^{\wedge}] = D_X \star \hat{\mathbf{L}}. \tag{3.9}$$

This property leads to the result that, for real twist-tensor fields $\mathbf{L} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$, the covariant derivative $D_X \mathbf{L}$ is also real, i.e., $D_X \mathbf{L} \in \Gamma(\mathscr{M}, \mathscr{E}(\mathscr{M}))$.

Induced structure on $\mathscr{T}(\mathscr{M})$ from $\mathscr{E}(\mathscr{M})$: The selection of an origin twist-tensor field $\mathbf{O}$ makes it possible to define a unique map from $\mathscr{T}_q$ to $\mathscr{F}_q$ for each q in $\mathscr{M}$. This map leads to a unique way of imposing a metric structure and connection on the tangent bundle $\mathscr{T}(\mathscr{M})$.

Note that, although we are introducing an origin twist-tensor field $\mathbf{O}$ in the theory, it is not regarded as a privileged field to be included in the specification of the structure of the fibers. Furthermore, it will not be invariant under parallel transportation. As pointed out in a similar situation discussed in I, this field $\mathbf{O}$ imposes no special restriction on the theory since a change in the choice of $\mathbf{O}$ can be compensated by a corresponding change in the connection $\mathbf{D}$ such that a completely equivalent theory is obtained. Arbitrary changes in $\mathbf{O}$ can be generated by the action of the Poincaré group. The connection $\mathbf{D}$ also changes according to a definition of the action of the Poincaré group on twistor connections. The $\mathbf{J}$ field (to be defined next) is also changed; however, the resulting theory is equivalent to the original one in that the induced metric structure and connection on the tangent bundle remain unchanged.

Now we introduce a tensor field $\mathbf{J}$, with value $\mathbf{J}(q) \in \mathscr{T}'_q \otimes \mathscr{E}_q$ at point q, defined as

$$\mathbf{J} = \mathbf{D} \otimes \mathbf{O}. \tag{3.10}$$

We shall need the following theorems:

Theorem 3.1: If $\mathbf{P}$ is an $\mathscr{H}_q$-valued vector field, then $D_X \mathbf{P}$ is an $\mathscr{H}_q$-valued vector field. If $\mathbf{P}$ is a $\mathscr{K}_q$-valued vector field, then $D_X \mathbf{P}$ is an $\mathscr{H}_q$-valued vector field.

Proof: Assume $\mathbf{I} \odot \mathbf{P} = c$, where the constant $c = 0$ or 2 in the respective cases of $\mathbf{P}$ being $\mathscr{H}_q$-valued or $\mathscr{K}_q$-valued. Then

$$X(\mathbf{I} \odot \mathbf{P}) = (D_X \mathbf{I}) \odot \mathbf{P} + \mathbf{I} \odot (D_X \mathbf{P}), \tag{3.11}$$

from which we immediately get

$$\mathbf{I} \odot (D_X \mathbf{P}) = 0. \tag{3.12}$$

Thus $D_X \mathbf{P}$ is $\mathscr{H}_q$-valued.

Theorem 3.2: $\mathbf{J}(q) \in \mathscr{T}'_q \otimes \mathscr{F}_q \subset \mathscr{T}'_q \otimes \mathscr{E}_q$.

Proof: From $\mathbf{O} \odot \mathbf{O} = 0$ we get

$$0 = X(\mathbf{O} \odot \mathbf{O}) = 2(D_X \mathbf{O}) \odot \mathbf{O}. \tag{3.13}$$

Then

$$\mathbf{x} \circ (\mathbf{D} \otimes \mathbf{O}) \odot \mathbf{O} = (D_X \mathbf{O}) \odot \mathbf{O} = 0. \tag{3.14}$$

Also, from Theorem 3.1, we have

$$\mathbf{x} \circ (\mathbf{D} \otimes \mathbf{O}) \odot \mathbf{I} = (D_X \mathbf{O}) \odot \mathbf{I} = 0. \tag{3.15}$$

Thus we have $\mathbf{J}(q) = (\mathbf{D} \otimes \mathbf{O})_q \in \mathscr{T}'_q \otimes \mathscr{F}_q$ since $\mathscr{F}_q$ is the subspace of $\mathscr{E}_q$ orthogonal to both $\mathbf{O}(q)$ and $\mathbf{I}(q)$.

As a consequence of this theorem, at each q, $\mathbf{J}(q)$ maps $\mathscr{T}_q$ into $\mathscr{F}_q$ as follows:

$$\mathbf{x}_q \in \mathscr{T}_q \rightarrow \mathbf{x}_q \circ \mathbf{J}(q) \in \mathscr{F}_q.$$

With the additional assumption that $\mathbf{J}(q)$ for each q is nonsingular, this map is a bijection of $\mathscr{T}_q$ on $\mathscr{F}_q$.

The inverse map $\mathbf{L}_q \in \mathscr{F}_q \rightarrow \mathbf{L}_q \odot \mathbf{F}(q) \in \mathscr{T}_q$ from $\mathscr{F}_q$ onto $\mathscr{T}_q$ is given by a unique $\mathscr{F}_q \otimes \mathscr{T}_q$-valued tensor field $\mathbf{F}$ satisfying the requirements

$$\mathbf{z} \circ \mathbf{J} \odot \mathbf{F} = \mathbf{z} \tag{3.16}$$

for every $\mathscr{T}_q$-valued vector field $\mathbf{z}$ and

$$\mathbf{L} \odot \mathbf{F} \circ \mathbf{J} = \mathbf{L} \tag{3.17}$$

for every $\mathscr{F}_q$-valued vector field $\mathbf{L}$. Note that $\mathbf{J}$ maps each $\mathscr{T}_q$-valued vector field $\mathbf{z}$ onto an $\mathscr{F}_q$-valued vector field

$\mathbf{z}\circ\mathbf{J}$, and $\mathbf{F}$ maps each $\mathscr{F}_q$-valued vector field $\mathbf{L}$ onto a $\mathscr{T}_q$-valued vector field $\mathbf{L}\odot\mathbf{F}$. We also, have

$$\mathbf{O}\odot\mathbf{F} = 0, \tag{3.18}$$

$$\mathbf{I}\odot\mathbf{F} = 0. \tag{3.19}$$

The inner product in $\mathscr{F}_q$ is mapped into an inner product in $\mathscr{T}_q$ by means of the $\mathbf{J}$ map according to the equation

$$\mathbf{x}_q\cdot\mathbf{y}_q = (\mathbf{x}_q\circ\mathbf{J}_q)\odot(\mathbf{y}_q\circ\mathbf{J}_q), \tag{3.20}$$

where $\mathbf{x}_q, \mathbf{y}_q\in\mathscr{T}_q$ and $\mathbf{J}_q\equiv\mathbf{J}(q)$. It follows that the unit tensors $\mathbf{I}_{\mathscr{F}}(q)\in\mathscr{F}_q\otimes\mathscr{F}_q$ and $\mathbf{I}_{\mathscr{T}}(q)\in\mathscr{T}_q\otimes\mathscr{T}_q$, defined by the equations $\mathbf{I}_{\mathscr{F}}(q)\odot\mathbf{L}_q = \mathbf{L}_q$ and $\mathbf{I}_{\mathscr{T}}(q)\cdot\mathbf{x}_q = \mathbf{x}_q$ for $\mathbf{L}_q\in\mathscr{F}_q$ and $\mathbf{x}_q\in\mathscr{T}_q$, are related by

$$\mathbf{I}_{\mathscr{T}} = \tilde{\mathbf{J}}\circ\mathbf{I}_{\mathscr{F}}\circ\mathbf{J}, \tag{3.21}$$

where $\tilde{\mathbf{J}}\in\mathscr{F}_q\otimes\mathscr{T}'_q$ is the transpose of $\mathbf{J}$.

Given the method of mapping vectors from $\mathscr{F}_q$ into $\mathscr{T}_q$, the mapping of tensors from $\mathscr{F}_q^{\otimes r}$ into $\mathscr{T}_q^{\otimes r}$ is straightforward. For convenience in subsequent calculations, we use the notation $(\circ\mathbf{J})_k$, $(\odot\mathbf{F})_k$, $(\mathbf{J}\odot)_k$, and $(\mathbf{F}\circ)_k$ to define linear maps in analogy to what was done in I. Thus (3.21) can be rewritten as

$$\mathbf{I}_{\mathscr{T}} = \mathbf{I}_{\mathscr{F}}(\circ\mathbf{J})_2(\circ\mathbf{J})_1, \tag{3.22}$$

and making use of (3.16) it can be inverted to yield

$$\mathbf{I}_{\mathscr{F}} = \mathbf{I}_{\mathscr{T}}(\odot\mathbf{F})_2(\odot\mathbf{F})_1. \tag{3.23}$$

Mapping of connections with $\mathbf{J}$: If $\mathbf{V}$ is an $\mathscr{F}_q$-valued vector field, $D_X\mathbf{V}$ is not necessarily another $\mathscr{F}_q$-valued vector field. However, according to Theorem 3.1, we do know that it is an $\mathscr{H}_q$-valued vector field. If we project $D_X\mathbf{V}$ onto $\mathscr{F}_q$ by means of $\mathbf{I}_{\mathscr{F}}$, the result defines an $\mathscr{F}$-connection $D^{\mathscr{F}}{}_X$ (i.e., it operates on $\mathscr{F}_q$-valued vector fields to produce $\mathscr{F}_q$-valued vector fields) according to

$$D^{\mathscr{F}}{}_X\mathbf{V} = \mathbf{I}_{\mathscr{F}}\odot(D_X\mathbf{V}). \tag{3.24}$$

To reexpress this in a different form, we write

$$\begin{aligned} D^{\mathscr{F}}{}_X\mathbf{V} &= D_X\mathbf{V} - \tfrac{1}{2}(\mathbf{I}\otimes\mathbf{O} + \mathbf{O}\otimes\mathbf{I})\odot(D_X\mathbf{V}) \\ &= D_X\mathbf{V} - \tfrac{1}{2}(\mathbf{I}\otimes\mathbf{O})\odot(D_X\mathbf{V}) \\ &= D_X\mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X\mathbf{O})\odot\mathbf{V}. \end{aligned} \tag{3.25}$$

Now define the connection $D^{\mathscr{E}}{}_X$ for $\mathscr{E}_q$-valued vector fields $\mathbf{V}$, by this new expression, i.e.,

$$D^{\mathscr{E}}{}_X\mathbf{V} = D_X\mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X\mathbf{O})\odot\mathbf{V}. \tag{3.26}$$

Note, in particular, that

$$D^{\mathscr{E}}{}_X\mathbf{O} = D_X\mathbf{O}, \tag{3.27}$$

$$D^{\mathscr{E}}{}_X\mathbf{I} = D_X\mathbf{I} = 0, \tag{3.28}$$

and for $\mathscr{F}_q$-valued vector fields $\mathbf{V}$,

$$D^{\mathscr{E}}{}_X\mathbf{V} = D^{\mathscr{F}}{}_X\mathbf{V}. \tag{3.29}$$

Theorem 3.3: The $\mathscr{F}$-connection $\mathbf{D}^{\mathscr{F}}$ is compatible with the inner product, i.e.,

$$X(\mathbf{V}\odot\mathbf{W}) = (D^{\mathscr{F}}{}_X\mathbf{V})\odot\mathbf{W} + \mathbf{V}\odot(D^{\mathscr{F}}{}_X\mathbf{W}) \tag{3.30}$$

for $\mathscr{F}_q$-valued vector fields $\mathbf{V}$ and $\mathbf{W}$.

Proof: For $\mathscr{F}_q$-valued vector fields $\mathbf{V}$ and $\mathbf{W}$, we have

$$\begin{aligned} &(D^{\mathscr{F}}{}_X\mathbf{V})\odot\mathbf{W} + \mathbf{V}\odot(D^{\mathscr{F}}{}_X\mathbf{W}) \\ &\quad= [D_X\mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X\mathbf{O})\odot\mathbf{V}]\odot\mathbf{W} \\ &\qquad+ \mathbf{V}\odot[D_X\mathbf{W} + \tfrac{1}{2}\mathbf{I}(D_X\mathbf{O})\odot\mathbf{W}] \\ &\quad= (D_X\mathbf{V})\odot\mathbf{W} + \mathbf{V}\odot(D_X\mathbf{W}) = X(\mathbf{V}\odot\mathbf{W}) \end{aligned} \tag{3.31}$$

since $\mathbf{I}\odot\mathbf{W} = 0$, $\mathbf{V}\odot\mathbf{I} = 0$, and D_X is compatible with the inner product.

Note that the connection $D^{\mathscr{E}}{}_X$ is not compatible with the inner product.

The $\mathscr{F}$-connection $\mathbf{D}^{\mathscr{F}}$ is mapped onto a connection $\boldsymbol{\nabla}$ on the tangent bundle $\mathscr{T}(\mathscr{M})$ by means of the equation

$$\nabla_X\mathbf{z} = [D^{\mathscr{F}}{}_X(\mathbf{z}\circ\mathbf{J})]\odot\mathbf{F} \tag{3.32}$$

for $\mathscr{T}_q$-valued vector fields $\mathbf{z}$. Equivalently,

$$(\nabla_X\mathbf{z})\circ\mathbf{J} = D^{\mathscr{F}}{}_X(\mathbf{z}\circ\mathbf{J}). \tag{3.33}$$

Since $\mathbf{I}_{\mathscr{F}}\odot\mathbf{F} = \mathbf{F}$, we also get from (3.24) the result

$$\nabla_X\mathbf{z} = [D^{\mathscr{F}}{}_X(\mathbf{z}\circ\mathbf{J})]\odot\mathbf{F} = [D_X(\mathbf{z}\circ\mathbf{J})]\odot\mathbf{F}. \tag{3.34}$$

Theorem 3.4: The connection $\boldsymbol{\nabla}$ is compatible with the inner product in $\mathscr{T}(\mathscr{M})$.

Proof: Let $\mathbf{y}$ and $\mathbf{z}$ be aribtrary $\mathscr{T}_q$-valued vector fields. Then, making use of (3.20), (3.30), and (3.33), we get

$$\begin{aligned} X(\mathbf{y}\cdot\mathbf{z}) &= X[(\mathbf{y}\circ\mathbf{J})\odot(\mathbf{z}\circ\mathbf{J})] \\ &= [D^{\mathscr{F}}{}_X(\mathbf{y}\circ\mathbf{J})]\odot(\mathbf{z}\circ\mathbf{J}) + (\mathbf{y}\circ\mathbf{J})\odot[D^{\mathscr{F}}{}_X(\mathbf{z}\circ\mathbf{J})] \\ &= [(\nabla_X\mathbf{y})\circ\mathbf{J}]\odot(\mathbf{z}\circ\mathbf{J}) + (\mathbf{y}\circ\mathbf{J})\odot[(\nabla_X\mathbf{z})\circ\mathbf{J}] \\ &= (\nabla_X\mathbf{y})\cdot\mathbf{z} + \mathbf{y}\cdot(\nabla_X\mathbf{z}). \end{aligned} \tag{3.35}$$

Curvature tensor: For the $\mathscr{E}(\mathscr{M})$ bundle, define the curvature tensor $\mathbf{R}_{\mathscr{E}}$ with values in $\mathscr{T}'_q\otimes\mathscr{T}'_q\otimes\mathscr{E}_q\otimes\mathscr{E}_q$ by

$$\mathbf{xy} \overset{\circ}{\circ} \mathbf{R}_{\mathscr{E}}\odot\mathbf{V} = (D_XD_Y - D_YD_X - D_{[X,Y]})\mathbf{V} \tag{3.36}$$

for vector fields $\mathbf{x}$, $\mathbf{y}$, $\mathbf{V}$ with values $\mathbf{x}(q)$, $\mathbf{y}(q)\in\mathscr{T}_q$ and $\mathbf{V}(q)\in\mathscr{E}_q$. We prove the antisymmetry property of $\mathbf{R}_{\mathscr{E}}$ under the transposition **(34)**, which exchanges the 3rd and 4th vector files.[16]

Theorem 3.5: **(34)** $\mathbf{R}_{\mathscr{E}} = -\mathbf{R}_{\mathscr{E}}$

Proof:

$$\begin{aligned} XY(\mathbf{V}\odot\mathbf{W}) &= X[(D_Y\mathbf{V})\odot\mathbf{W} + \mathbf{V}\odot(D_Y\mathbf{W})] \\ &= (D_XD_Y\mathbf{V})\odot\mathbf{W} + (D_Y\mathbf{V})\odot(D_X\mathbf{W}) \\ &\qquad+ (D_X\mathbf{V})\odot(D_Y\mathbf{W}) + \mathbf{V}\odot(D_XD_Y\mathbf{W}); \end{aligned} \tag{3.37}$$

also,

$$\begin{aligned} &[X, Y](\mathbf{V}\odot\mathbf{W}) \\ &\quad= (D_{[X,Y]}\mathbf{V})\odot\mathbf{W} + \mathbf{V}\odot(D_{[X,Y]}\mathbf{W}). \end{aligned} \tag{3.38}$$

Taking (3.37) minus a similar equation with X and Y interchanged and subtracting (3.38) gives

$$\begin{aligned} &[(D_XD_Y - D_YD_X - D_{[X,Y]})\mathbf{V}]\odot\mathbf{W} \\ &\quad+ \mathbf{V}\odot[(D_XD_Y - D_YD_X - D_{[X,Y]})\mathbf{W}] = 0, \end{aligned}$$

in which the equation $(XY - YX - [X, Y])(\mathbf{V}\odot\mathbf{W}) = 0$ was used. This results in

$$\mathbf{xy} \overset{\circ}{\circ} \mathbf{R}_{\mathscr{E}} \overset{\odot}{\odot} (\mathbf{WV} + \mathbf{VW}) = 0,$$

$$\mathbf{xy} \overset{\circ}{\circ} \{[1 + \mathbf{(34)}]\mathbf{R}_{\mathscr{E}}\} \overset{\odot}{\odot} \mathbf{WV} = 0$$

for arbitrary $\mathbf{x}$, $\mathbf{y}$, $\mathbf{V}$, and $\mathbf{W}$. Therefore,

$$[1 + \mathbf{(34)}]\mathbf{R}_{\mathscr{E}} = 0. \tag{3.39}$$

We also have

Theorem 3.6: $\mathbf{R}_{\mathscr{G}}$ has its value at q in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{H}_q \otimes \mathscr{H}_q$.

Proof: The equation

$$\mathbf{xy} \underset{\circ}{:} \mathbf{R}_{\mathscr{G}} \odot \mathbf{I} = 0$$

follows from the fact that $D_X \mathbf{I} = 0$ for all X. Since $\mathbf{x}$ and $\mathbf{y}$ are arbitrary, we have

$$\mathbf{R}_{\mathscr{G}} \odot \mathbf{I} = 0. \tag{3.40}$$

Also, from (3.39),

$$[(\mathbf{34})\mathbf{R}_{\mathscr{G}}] \odot \mathbf{I} = -\mathbf{R}_{\mathscr{G}} \odot \mathbf{I} = 0. \tag{3.41}$$

These two equations imply that $\mathbf{R}_{\mathscr{G}}$ has its values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{H}_q \otimes \mathscr{H}_q$.

Now define the tensor $\mathbf{T}_{\mathscr{F}}$ by

$$\mathbf{T}_{\mathscr{F}} = \mathbf{R}_{\mathscr{G}} \odot \mathbf{O}. \tag{3.42}$$

For this tensor, we have

$$\begin{aligned}\mathbf{T}_{\mathscr{F}} \odot \mathbf{O} &= \mathbf{R}_{\mathscr{G}} \overset{\odot}{\odot} \mathbf{OO} \\ &= \{\tfrac{1}{2}[1 + (\mathbf{34})]\mathbf{R}_{\mathscr{G}}\} \overset{\odot}{\odot} \mathbf{O} = 0.\end{aligned} \tag{3.43}$$

Thus $\mathbf{T}_{\mathscr{F}}$ has its values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{F}_q$.

Now define the tensor $\mathbf{R}_{\mathscr{F}}$ by

$$\mathbf{R}_{\mathscr{F}} = \mathbf{R}_{\mathscr{G}} - \tfrac{1}{2}[1 - (\mathbf{34})](\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}). \tag{3.44}$$

For this tensor,

$$\begin{aligned}\mathbf{R}_{\mathscr{F}} \odot \mathbf{O} &= \mathbf{R}_{\mathscr{G}} \odot \mathbf{O} - \tfrac{1}{2}\mathbf{T}_{\mathscr{F}} \mathbf{I} \odot \mathbf{O} \\ &\quad + \tfrac{1}{2}(\mathbf{T}_{\mathscr{F}} \odot \mathbf{O}) \otimes \mathbf{I} \\ &= \mathbf{T}_{\mathscr{F}} - \mathbf{T}_{\mathscr{F}} + 0 = 0.\end{aligned} \tag{3.45}$$

Also, due to the antisymmetry of $\mathbf{R}_{\mathscr{F}}$ under the $(\mathbf{34})$ transposition, we have

$$[(\mathbf{34})\mathbf{R}_{\mathscr{F}}] \odot \mathbf{O} = -\mathbf{R}_{\mathscr{F}} \odot \mathbf{O} = 0. \tag{3.46}$$

The last two equations, along with the easily proven facts that $\mathbf{R}_{\mathscr{F}} \odot \mathbf{I} = 0$ and $[(\mathbf{34})\mathbf{R}_{\mathscr{F}}] \odot \mathbf{I} = 0$ gives us the result that $\mathbf{R}_{\mathscr{F}}$ has its values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{F}_q \otimes \mathscr{F}_q$. Rearranging the terms in (3.44) gives

$$\mathbf{R}_{\mathscr{G}} = \mathbf{R}_{\mathscr{F}} + \tfrac{1}{2}[1 - (\mathbf{34})]\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}, \tag{3.47}$$

which is a unique decomposition of $\mathbf{R}_{\mathscr{G}}$.

For the connection $D^{\mathscr{E}}{}_X$, we have the curvature tensor $\mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}}$ with values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{E}_q \otimes \mathscr{E}_q$ defined by

$$\begin{aligned}&\mathbf{xy} \underset{\circ}{:} \mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \mathbf{V} \\ &\quad = (D^{\mathscr{E}}{}_X D^{\mathscr{E}}{}_Y - D^{\mathscr{E}}{}_Y D^{\mathscr{E}}{}_X - D^{\mathscr{E}}{}_{[X,Y]})\mathbf{V}\end{aligned} \tag{3.48}$$

for $\mathscr{E}_q$-valued vector fields $\mathbf{V}$. Using (3.26), we have

$$\begin{aligned}\mathbf{xy} \underset{\circ}{:} \mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \mathbf{V} &= D^{\mathscr{E}}{}_X [D_Y \mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_Y \mathbf{O}) \odot \mathbf{V}] \\ &\quad - D^{\mathscr{E}}{}_Y [D_X \mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X \mathbf{O}) \odot \mathbf{V}] \\ &\quad - D_{[X,Y]}\mathbf{V} - \tfrac{1}{2}\mathbf{I}(D_{[X,Y]}\mathbf{O}) \odot \mathbf{V} \\ &= D_X [D_Y \mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_Y \mathbf{O}) \odot \mathbf{V}] \\ &\quad + \tfrac{1}{2}\mathbf{I}(D_X \mathbf{O}) \odot [D_Y \mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_Y \mathbf{O}) \odot \mathbf{V}] \\ &\quad - D_Y [D_X \mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X \mathbf{O}) \odot \mathbf{V}] \\ &\quad - \tfrac{1}{2}\mathbf{I}(D_Y \mathbf{O}) \odot [D_X \mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X \mathbf{O}) \odot \mathbf{V}] \\ &\quad - D_{[X,Y]}\mathbf{V} - \tfrac{1}{2}\mathbf{I}(D_{[X,Y]}\mathbf{O}) \odot \mathbf{V} \\ &= (D_X D_Y - D_Y D_X - D_{[X,Y]})\mathbf{V} \\ &\quad + \tfrac{1}{2}\mathbf{I}(D_X D_Y \mathbf{O} - D_Y D_X \odot - D_{[X,Y]}\mathbf{O}) \odot \mathbf{V} \\ &= \mathbf{xy} \underset{\circ}{:} \mathbf{R}_{\mathscr{G}} \odot \mathbf{V} + \tfrac{1}{2}\mathbf{xy} \underset{\circ}{:} \mathbf{T}_{\mathscr{F}} \odot \mathbf{VI}.\end{aligned}$$

Therefore,

$$\mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} = \mathbf{R}_{\mathscr{G}} + \tfrac{1}{2}(\mathbf{34})(\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}). \tag{3.49}$$

Similarly to (3.42), we define for the connection $D^{\mathscr{E}}{}_X$

$$\mathbf{T}^{\mathscr{E}}{}_{\mathscr{F}} = \mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \mathbf{O}. \tag{3.50}$$

Thus it follows from (3.49) and (3.43) that

$$\mathbf{T}^{\mathscr{E}}{}_{\mathscr{F}} = \mathbf{T}_{\mathscr{F}}, \tag{3.51}$$

i.e., $\mathbf{T}^{\mathscr{E}}{}_{\mathscr{F}}$ has its values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{F}_q$. Also, substitution of (3.47) into (3.49) results in

$$\begin{aligned}\mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} &= \mathbf{R}_{\mathscr{F}} + \tfrac{1}{2}[1 - (\mathbf{34})](\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}) \\ &\quad + \tfrac{1}{2}(\mathbf{34})(\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}) \\ &= \mathbf{R}_{\mathscr{F}} + \tfrac{1}{2}\mathbf{T}_{\mathscr{F}} \otimes \mathbf{I}.\end{aligned} \tag{3.52}$$

This is a unique decomposition of $\mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}}$.

Because of (3.48), (3.27), and (3.29), we have

$$\begin{aligned}&\mathbf{xy} \underset{\circ}{:} \mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \mathbf{O} \\ &\quad = (D^{\mathscr{E}}{}_X D_Y - D^{\mathscr{E}}{}_Y D_X - D_{[X,Y]})\mathbf{O} \\ &\quad = D^{\mathscr{F}}{}_X(\mathbf{y} \circ \mathbf{D} \otimes \mathbf{O}) - D^{\mathscr{F}}{}_Y(\mathbf{x} \circ \mathbf{D} \otimes \mathbf{O}) - [\mathbf{x}, \mathbf{y}] \circ \mathbf{D} \otimes \mathbf{O} \\ &\quad = D^{\mathscr{F}}{}_X(\mathbf{y} \circ \mathbf{J}) - D^{\mathscr{F}}{}_Y(\mathbf{x} \circ \mathbf{J}) - [\mathbf{x}, \mathbf{y}] \circ \mathbf{J} \\ &\quad = (\nabla_X \mathbf{y} - \nabla_Y \mathbf{x} - [\mathbf{x}, \mathbf{y}]) \circ \mathbf{J} = \mathbf{xy} \underset{\circ}{:} \mathbf{T}_{\mathscr{T}} \circ \mathbf{J},\end{aligned}$$

where $\mathbf{T}_{\mathscr{T}}$ is the torsion tensor on the tangent bundle with its values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{T}_q$. Therefore,

$$\mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \mathbf{O} = \mathbf{T}_{\mathscr{T}} \circ \mathbf{J}. \tag{3.53}$$

In view of (3.50) and (3.51), this becomes

$$\mathbf{T}_{\mathscr{F}} = \mathbf{T}_{\mathscr{T}} \circ \mathbf{J}, \tag{3.54}$$

or, equivalently,

$$\mathbf{T}_{\mathscr{T}} = \mathbf{T}_{\mathscr{F}} \odot \mathbf{F}. \tag{3.55}$$

With the aid of (3.48), (3.29), and (3.33), we have

$$\begin{aligned}&\mathbf{xy} \underset{\circ}{:} \mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot (\mathbf{z} \circ \mathbf{J}) \\ &\quad = (D^{\mathscr{E}}{}_X D^{\mathscr{E}}{}_Y - D^{\mathscr{E}}{}_Y D^{\mathscr{E}}{}_X - D^{\mathscr{E}}{}_{[X,Y]})(\mathbf{z} \circ \mathbf{J}) \\ &\quad = (D^{\mathscr{F}}{}_X D^{\mathscr{F}}{}_Y - D^{\mathscr{F}}{}_Y D^{\mathscr{F}}{}_X - D^{\mathscr{F}}{}_{[X,Y]})(\mathbf{z} \circ \mathbf{J}) \\ &\quad = [(\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]})\mathbf{z}] \circ \mathbf{J} \\ &\quad = \mathbf{xy} \underset{\circ}{:} (\mathbf{R}_{\mathscr{T}} \circ \mathbf{z}) \circ \mathbf{J},\end{aligned}$$

where $\mathbf{R}_{\mathscr{T}}$ is the curvature tensor on the tangent bundle with its values in $\mathscr{T}'_q \otimes \mathscr{T}'_q \otimes \mathscr{T}_q \otimes \mathscr{T}'_q$. Thus,

$$\mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \tilde{\mathbf{J}} = \mathbf{R}_{\mathscr{T}}(\circ \mathbf{J})_2. \tag{3.56}$$

Moreover, noting that, by virtue of (3.52),

$$\mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \tilde{\mathbf{J}} \circ \tilde{\mathbf{F}} = \mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot (\mathbf{F} \circ \mathbf{J}) = \mathbf{R}^{\mathscr{E}}{}_{\mathscr{G}} \odot \tilde{\mathbf{I}}_{\mathscr{F}} = \mathbf{R}_{\mathscr{F}}, \tag{3.57}$$

(3.56) yields

$$\mathbf{R}_{\mathscr{F}} = \mathbf{R}_{\mathscr{T}}(\circ \mathbf{J})_2(\circ \tilde{\mathbf{F}})_1 \tag{3.58}$$

or, equivalently,

$$\mathbf{R}_{\mathscr{T}} = \mathbf{R}_{\mathscr{F}}(\odot \mathbf{F})_2(\odot \tilde{\mathbf{J}})_1. \tag{3.59}$$

Performing contractions on $\mathbf{R}_{\mathscr{T}}$ gives the Ricci tensor

$$R_{\mathscr{T}} \equiv C(\mathbf{13})\mathbf{R}_{\mathscr{T}}, \tag{3.60}$$

with values in $\mathscr{T}'_q \otimes \mathscr{T}'_q$, and the curvature invariant

$$(R_{\mathscr{T}})_s \equiv C(\mathbf{13;24})(\mathbf{R}_{\mathscr{T}} \circ \mathbf{I}_{\mathscr{T}}), \tag{3.61}$$

where the symbol $C(\ )$ denotes contraction on the designated files, for example, $C(\mathbf{12})(\mathbf{m}'\otimes\mathbf{v}) = \mathbf{m}'\circ\mathbf{v}$ for $\mathbf{m}'\in\mathscr{T}'_q$ at q and $\mathbf{v}\in\mathscr{T}_q$ at q. Similarly, from $\mathbf{R}_{\mathscr{T}}$ we may also get a second-order tensor

$$R_{\mathscr{T}} = C(\mathbf{13},\odot)(\mathbf{F}\circ\mathbf{R}_{\mathscr{T}}) \tag{3.62}$$

with values in $\mathscr{T}'_q\otimes\mathscr{F}_q$, and a scalar

$$\begin{aligned}(R_{\mathscr{T}})_s &= C(\mathbf{13;24},\odot)[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{T}}]\\ &= C(\mathbf{13;24},\odot)[(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{E}}],\end{aligned} \tag{3.63}$$

where $C(\ ,\odot)$ in these two cases denotes contraction again but with respect to the $\odot$ product; for example, $C(\mathbf{12},\odot)(\mathbf{L}\otimes\mathbf{M}) = \mathbf{L}\odot\mathbf{M}$ for $\mathbf{L},\mathbf{M}\in\mathscr{F}_q$ at q. These quantities are related as follows:

$$\mathbf{R}_{\mathscr{T}} = \mathbf{R}_{\mathscr{T}}\odot\tilde{\mathbf{J}}, \tag{3.64}$$

$$(R_{\mathscr{T}})_s = (R_{\mathscr{T}})_s. \tag{3.65}$$

Local Poincaré transformations: Let $\gamma(P)$ and $\Gamma(P)$ represent the action of an arbitrary local Poincaré transformation $\mathscr{P}$ on the $\mathscr{U}(\mathscr{M})$ and $\mathscr{E}(\mathscr{M})$ bundles, respectively. Then

$$\mathbf{u}\to\mathbf{u}^{(P)} = \gamma(P)\mathbf{u},$$

$$\mathbf{V}\to\mathbf{V}^{(P)} = \Gamma(P)\mathbf{V},$$

where $\mathbf{u}$ and $\mathbf{u}^{(P)}$ are $\mathscr{U}_q$-valued twistor fields and $\mathbf{V}$ and $\mathbf{V}^{(P)}$ are $\mathscr{E}_q$-valued twist-tensor fields. Let the action of $\mathscr{P}$ on twistor connections

$$D_X\to D_X{}^{(P)}$$

be defined by

$$D_X{}^{(P)}\mathbf{u} = \gamma(P)D_X[\gamma(P^{-1})\mathbf{u}] \tag{3.66}$$

for arbitrary $\mathbf{u}$. To the twistor connection $D_X{}^{(P)}$ on the $\mathscr{U}(\mathscr{M})$ bundle, there corresponds the twist-tensor connection $D_X{}^{(P)}$ (same symbol) on the $\mathscr{E}(\mathscr{M})$ bundle. It follows that

$$D_X{}^{(P)}\mathbf{V} = \Gamma(P)D_X[\Gamma(P^{-1})\mathbf{V}] \tag{3.67}$$

for arbitrary $\mathbf{V}$. The covariant gradients $\mathbf{D}\otimes\mathbf{V}$ and $\mathbf{D}^{(P)}\otimes\mathbf{V}$ corresponding to the D_X and $D_X{}^{(P)}$ connections are $\mathscr{T}'_q\otimes\mathscr{E}_q$-valued fields defined by

$$\begin{aligned}\mathbf{x}\circ(\mathbf{D}\otimes\mathbf{V}) &= D_X\mathbf{V},\\ \mathbf{x}\circ(\mathbf{D}^{(P)}\otimes\mathbf{V}) &= D_X{}^{(P)}\mathbf{V},\end{aligned} \tag{3.68}$$

respectively, for arbitrary tangent vector fields $\mathbf{x}\equiv\mathbf{X}$.

Now we compare the inner products

$$\mathbf{x}\cdot\mathbf{y} = (\mathbf{x}\circ\mathbf{J})\odot(\mathbf{y}\circ\mathbf{J})$$

and

$$\mathbf{x}(\cdot)^{(P)}\mathbf{y} = (\mathbf{x}\circ\mathbf{J}^{(P)})\odot(\mathbf{y}\circ\mathbf{J}^{(P)})$$

of the tangent vector fields $\mathbf{x}\equiv\mathbf{X}$ and $\mathbf{y}\equiv\mathbf{Y}$ induced on the tangent bundle by the maps generated by the fields $\mathbf{J} = \mathbf{D}\otimes\mathbf{O}$ and $\mathbf{J}^{(P)} = \mathbf{D}^{(P)}\otimes\mathbf{O}^{(P)}$. Let

$$\begin{aligned}\mathscr{F}_q{}^{(P)} &= \{\mathbf{R}_q\,|\,\mathbf{R}_q\in\mathscr{H}_q,\ \mathbf{O}^{(P)}(q)\odot\mathbf{R}_q = 0\}\\ &= \{\mathbf{R}_q\,|\,\mathbf{R}_q\in\mathscr{E}_q,\ \mathbf{I}(q)\odot\mathbf{R}_q = 0,\\ &\qquad\qquad\qquad \mathbf{O}^{(P)}(q)\odot\mathbf{R}_q = 0\}.\end{aligned}$$

It follows that $\mathscr{F}_q{}^{(P)}$ is the image of $\mathscr{F}_q$ under the map $\Gamma(P)$. Note that $\mathbf{x}\circ\mathbf{J}$ is $\mathscr{F}_q$-valued and $\mathbf{x}\circ\mathbf{J}^{(P)}$ is $\mathscr{F}_q{}^{(P)}$-valued. Furthermore,

$$\begin{aligned}\mathbf{x}\circ\mathbf{J}^{(P)} &= \mathbf{x}\circ(\mathbf{D}^{(P)}\otimes\mathbf{O}^{(P)}) = D_X^{(P)}\mathbf{O}^{(P)}\\ &= \Gamma(P)D_X[\Gamma(P^{-1})\Gamma(P)\mathbf{O}] = \Gamma(P)D_X\mathbf{O}\\ &= \Gamma(P)[\mathbf{x}\circ(\mathbf{D}\otimes\mathbf{O})] = \Gamma(P)(\mathbf{x}\circ\mathbf{J}).\end{aligned} \tag{3.69}$$

Now the inner products are

$$\begin{aligned}\mathbf{x}(\cdot)^{(P)}\mathbf{y} &= (\mathbf{x}\circ\mathbf{J}^{(P)})\odot(\mathbf{y}\circ\mathbf{J}^{(P)})\\ &= [\Gamma(P)(\mathbf{x}\circ\mathbf{J})]\odot[\Gamma(P)(\mathbf{y}\circ\mathbf{J})]\\ &= (\mathbf{x}\circ\mathbf{J})\odot(\mathbf{y}\circ\mathbf{J}) = \mathbf{x}\cdot\mathbf{y}.\end{aligned}$$

Thus the inner product induced in the tangent bundle is invariant under $\mathscr{P}$.

Now we compare the connections ∇_X and $\nabla_X{}^{(P)}$ induced on the tangent bundle by the connections D_X and $D_X{}^{(P)}$, respectively, by means of the maps generated by $\mathbf{J}$ and $\mathbf{J}^{(P)}$, respectively. For tangent vector fields $\mathbf{x}\equiv\mathbf{X}$ and $\mathbf{z}$, we have

$$(\nabla_X\mathbf{z})\circ\mathbf{J} = D_X{}^{\mathscr{F}}(\mathbf{z}\circ\mathbf{J}), \tag{3.70}$$

where

$$D_X{}^{\mathscr{F}}\mathbf{V} = D_X\mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X\mathbf{O})\odot\mathbf{V} \tag{3.71}$$

for $\mathscr{F}_q$-valued twist-tensor fields $\mathbf{V}$. Also we have

$$(\nabla_X{}^{(P)}\mathbf{z})\circ\mathbf{J}^{(P)} = D_X{}^{(P)\mathscr{F}^{(P)}}(\mathbf{z}\circ\mathbf{J}^{(P)}), \tag{3.72}$$

where

$$D_X{}^{(P)\mathscr{F}^{(P)}}\mathbf{V} = D_X{}^{(P)}\mathbf{V} + \tfrac{1}{2}\mathbf{I}(D_X{}^{(P)}\mathbf{O}^{(P)})\odot\mathbf{V} \tag{3.73}$$

for $\mathscr{T}_q$-valued twist-tensor fields $\mathbf{V}$. Then

$$\begin{aligned}&(\nabla_X{}^{(P)}\mathbf{z})\circ\mathbf{J}^{(P)}\\ &\quad= D_X{}^{(P)\mathscr{F}^{(P)}}(\mathbf{z}\circ\mathbf{J}^{(P)})\\ &\quad= D_X{}^{(P)}(\mathbf{z}\circ\mathbf{J}^{(P)}) + \tfrac{1}{2}\mathbf{I}(D_X{}^{(P)}\mathbf{O}^{(P)})\odot(\mathbf{z}\circ\mathbf{J}^{(P)})\\ &\quad= \Gamma(P)D_X[\Gamma(P^{-1})\Gamma(P)(\mathbf{z}\circ\mathbf{J})]\\ &\qquad+ \tfrac{1}{2}\mathbf{I}\{\Gamma(P)D_X[\Gamma(P^{-1})\Gamma(P)\mathbf{O}]\}\odot[\Gamma(P)(\mathbf{z}\circ\mathbf{J})]\\ &\quad= \Gamma(P)D_X(\mathbf{z}\circ\mathbf{J}) + \tfrac{1}{2}\mathbf{I}(D_X\mathbf{O})\odot(\mathbf{z}\circ\mathbf{J})\\ &\quad= \Gamma(P)[D_X(\mathbf{z}\circ\mathbf{J}) + \tfrac{1}{2}\mathbf{I}(D_X\mathbf{O})\odot(\mathbf{z}\circ\mathbf{J})]\\ &\quad= \Gamma(P)D_X{}^{\mathscr{F}}(\mathbf{z}\circ\mathbf{J}) = \Gamma(P)[(\nabla_X\mathbf{z})\circ\mathbf{J}]\\ &\quad= (\nabla_X\mathbf{z})\circ\mathbf{J}^{(P)},\end{aligned} \tag{3.74}$$

where we have used $\Gamma(P)\mathbf{I} = \mathbf{I}$ and Eq. (3.69). Applying the inverse of the map generated by $\mathbf{J}^{(P)}$ gives

$$\nabla_X{}^{(P)}\mathbf{z} = \nabla_X\mathbf{z}. \tag{3.75}$$

Thus the connection induced in the tangent bundle is invariant under $\mathscr{P}$.

IV. TWISTOR VARIATIONAL PRINCIPLES

As we have shown in I, given a Lagrangian constructed from the gauge fields (connections), we can obtain the equations of motion for these fields by means of an action principle in which the fundamental quantities to be varied are the D_X.

Moreover, since any two linear connections may differ only by a linear transformation, we have

$$(\delta D_X)\mathbf{V} = (\delta\mathbf{B}_X)\odot\mathbf{V} \tag{4.1}$$

for each $\mathscr{E}_q$-valued twist-tensor field $\mathbf{V}$, where $\delta\mathbf{B}_X(q)\in\mathscr{E}_q$

$\otimes \mathscr{E}_q$. If we now vary (3.8), we get

$$0 = (\delta D_X \mathbf{V}) \odot \mathbf{W} + \mathbf{V} \odot (\delta D_X \mathbf{W}) = \mathbf{V} \odot \{ [1 + (\mathbf{12})] \delta \mathbf{B}_X \} \odot \mathbf{W}. \quad (4.2)$$

Since $\mathbf{V}$, $\mathbf{W}$ are arbitrary, it follows immediately that

$$[1 + (\mathbf{12})] \delta \mathbf{B}_X = 0. \quad (4.3)$$

In addition, setting $\mathbf{V} = \mathbf{I}$ in the first part of (4.2) leads to

$$\mathbf{I} \odot \delta \mathbf{B}_X = 0. \quad (4.4)$$

Thus, (4.3) and (4.4) combined imply that $\delta \mathbf{B}_X(q) \in \mathscr{H}_q \otimes \mathscr{H}_q$ and $\delta \mathbf{B}(q) \in \mathscr{T}'_q \otimes \mathscr{H}_q \otimes \mathscr{H}_q$, where $\delta \mathbf{B}(q)$ is defined by $\delta \mathbf{B}_X = \mathbf{x} \circ \delta \mathbf{B}$, in analogy to what we did in I.

To relate this variation of a connection on a twist-tensor field to the variation of the twistor connection

$$(\delta D_X)\mathbf{l} = (\delta \mathbf{M}_X) \blacktriangle \mathbf{l}, \quad (4.5)$$

where $\delta \mathbf{M}_X \in \mathscr{U}^{\otimes 2}$ and $\mathbf{l}$ is a $\mathscr{U}_q$-valued twistor field, observe that

$$(\delta D_X)\mathbf{V} = (\delta \mathbf{M}_X) \blacktriangle \mathbf{V} - \mathbf{V} \blacktriangle (\delta \widetilde{\mathbf{M}}_X) = \{ (23) [(\delta \mathbf{M}_X)(\mathbf{I} + \mathbf{O}) + (\mathbf{I} + \mathbf{O})(\delta \mathbf{M}_X)] \} \overset{\blacktriangle}{\blacktriangle} \mathbf{V}. \quad (4.6)$$

Substituting (4.1) on the left side of (4.6) and using (2.27) on the right side of (4.6) yields

$$\delta \mathbf{B}_X \odot \mathbf{V} = \tfrac{1}{2} \{ (23) [(\delta \mathbf{M}_X)(\mathbf{I} + \mathbf{O}) + (\mathbf{I} + \mathbf{O})(\delta \mathbf{M}_X)] \} \overset{\blacktriangle}{\blacktriangle} \mathbf{\Lambda} \odot \mathbf{V}, \quad (4.7)$$

i.e.,

$$\delta \mathbf{B}_X = \tfrac{1}{2} \{ (23) [\delta \mathbf{M}_X (\mathbf{I} + \mathbf{O}) + (\mathbf{I} + \mathbf{O})(\delta \mathbf{M}_X)] \} \overset{\blacktriangle}{\blacktriangle} \mathbf{\Lambda}. \quad (4.8)$$

Note that by putting $\mathbf{V} = \mathbf{O}$ in the first part of (4.6) leads, in particular, to

$$\delta \mathbf{J}_X = \delta D_X \mathbf{O} = \delta \mathbf{B}_X \odot \mathbf{O} = \delta \mathbf{M}_X \blacktriangle \mathbf{O} - \mathbf{O} \blacktriangle \delta \widetilde{\mathbf{M}}_X = \delta \mathbf{M}_X \blacktriangle \mathbf{O} - (\delta \mathbf{M}_X \blacktriangle \mathbf{O})\widetilde{\ }, \quad (4.9)$$

which relates the variation of $\mathbf{J}_X$ with the variation of the twistor connection.

The space in which $\delta \mathbf{M}_X$ has values at each point $q \in \mathscr{M}$ is given by the following:

Theorem 4.1:

$$\delta \mathbf{M}_X \in \left[\left(\mathscr{S}_2 \underset{s}{\otimes} \mathscr{S}_2 \right) \underset{R}{\oplus} \left(\overline{\mathscr{S}}_2 \underset{s}{\otimes} \overline{\mathscr{S}}_2 \right) \right] \oplus \left(\mathscr{S}_2 \underset{aH}{\otimes} \overline{\mathscr{S}}_2 \right) \quad (4.10)$$

where $(\mathscr{S}_2 \otimes_s \mathscr{S}_2)$ is the subset of all symmetric tensors in $\mathscr{S}_2 \otimes \mathscr{S}_2$, $(\overline{\mathscr{S}}_2 \otimes_s \overline{\mathscr{S}}_2)$ is the subset of all symmetric tensors in $\overline{\mathscr{S}}_2 \otimes \overline{\mathscr{S}}_2$, $(\mathscr{S}_2 \otimes_s \mathscr{S}_2) \oplus_R (\overline{\mathscr{S}}_2 \otimes_s \overline{\mathscr{S}}_2)$ denotes the real part (with respect to the adjoint operation) of the direct sum of these subspaces, and $\mathscr{S}_2 \otimes_{aH} \overline{\mathscr{S}}_2$ is the subset of all anti-Hermitian tensors in $\mathscr{S}_2 \otimes \overline{\mathscr{S}}_2$.

Proof: From (3.5) and the first part of (4.6) we have

$$0 = (\delta D_X)\mathbf{I} = (\delta \mathbf{M}_X) \blacktriangle \mathbf{I} - \mathbf{I} \blacktriangle (\delta \widetilde{\mathbf{M}}_X).$$

Thus

$$\delta \mathbf{M}_X \blacktriangle \mathbf{I} = (\delta \mathbf{M}_X \blacktriangle \mathbf{I})\widetilde{\ }, \quad (4.11)$$

i.e., $\delta \mathbf{M}_X \blacktriangle \mathbf{I} \in \mathscr{S}_2 \otimes_s \mathscr{S}_2$. Equation (4.11) implies the direct sum decomposition

$$\delta \mathbf{M}_X \in \left(\mathscr{S}_2 \underset{s}{\otimes} \mathscr{S}_2 \right) \oplus (\mathscr{S}_4 \otimes \overline{\mathscr{S}}_2).$$

Now, from (3.4) and (4.5) we get

$$0 = (\delta D_X)\mathbf{\Lambda} = [(\delta \mathbf{M}_X \blacktriangle)_1 + (\delta \mathbf{M}_X \blacktriangle)_2 + (\delta \mathbf{M}_X \blacktriangle)_3 + (\delta \mathbf{M}_X \blacktriangle)_4] \mathbf{\Lambda} = -(\delta M_X)_s \mathbf{\Lambda} \quad (4.12)$$

[where again we are making use of the notation introduced in I, Eqs. (3.14) and (3.15)]. Equation (4.12) follows from a well-known result in linear algebra with the symbol $(\delta M_X)_s$ denoting the scalar invariant of $\delta \mathbf{M}_X$.

Therefore, (4.12) results in

$$(\delta M_X)_s \equiv -(\delta \mathbf{M}_X) \overset{\blacktriangle}{\blacktriangle} (\mathbf{I} + \mathbf{O}) = 0, \quad (4.13)$$

and (4.11) together with (4.13) implies

$$\delta \mathbf{M}_X \in \left(\mathscr{S}_2 \underset{s}{\otimes} \mathscr{S}_2 \right) \oplus \left(\overline{\mathscr{S}}_2 \underset{s}{\otimes} \overline{\mathscr{S}}_2 \right) \oplus (\mathscr{S}_2 \otimes \overline{\mathscr{S}}_2). \quad (4.14)$$

Finally, from (3.3) and (2.63) we can write

$$X \langle \mathbf{l} | \mathbf{m} \rangle = \langle D_X \mathbf{l} | \mathbf{m} \rangle + \langle \mathbf{l} | D_X \mathbf{m} \rangle$$

or

$$X [\bar{\mathbf{l}} \blacktriangle (\bar{\mathbf{I}} - \mathbf{O}) \blacktriangle \mathbf{m}] = (\overline{D_X \mathbf{l}}) \blacktriangle (\mathbf{I} - \mathbf{O}) \blacktriangle \mathbf{m} + \bar{\mathbf{l}} \blacktriangle (\mathbf{I} - \mathbf{O}) \blacktriangle D_X \mathbf{m}.$$

Varying this last expression gives

$$0 = (\overline{\delta \mathbf{M}_X \blacktriangle \mathbf{l}}) \blacktriangle (\mathbf{I} - \mathbf{O}) \blacktriangle \mathbf{m} + \bar{\mathbf{l}} \blacktriangle (\mathbf{I} - \mathbf{O}) \blacktriangle \delta \mathbf{M}_X \blacktriangle \mathbf{m}) = -\bar{\mathbf{l}} \blacktriangle \delta \widetilde{\overline{\mathbf{M}}}_X \blacktriangle (\mathbf{I} - \mathbf{O}) \blacktriangle \mathbf{m} + \bar{\mathbf{l}} \blacktriangle (\mathbf{I} - \mathbf{O}) \blacktriangle \delta \mathbf{M}_X \blacktriangle \mathbf{m},$$

i.e.,

$$0 = (\mathbf{I} - \mathbf{O}) \blacktriangle \delta \mathbf{M}_X + [\overline{(\mathbf{I} - \mathbf{O}) \blacktriangle \delta \mathbf{M}_X}]\widetilde{\ }. \quad (4.15)$$

Equation (4.15) together with (4.14) leads to

$$\delta \mathbf{M}_X \in \left[\left(\mathscr{S}_2 \underset{s}{\otimes} \mathscr{S}_2 \right) \underset{R}{\oplus} \left(\overline{\mathscr{S}}_2 \underset{s}{\otimes} \overline{\mathscr{S}}_2 \right) \right] \oplus (\mathscr{S}_2 \underset{aH}{\otimes} \overline{\mathscr{S}}_2).$$

Q.E.D.

We can now apply these results to specific Lagrangians, constructed from (3.44), to obtain twistorial equations of motion for the corresponding gauge fields. In what follows, we shall concentrate on the Lagrangian which will result in a spinorial formulation of the Einstein–Cartan theory. The extension of the procedure to other permissible Lagrangians is suggested by the approach here adopted and is rather straightforward. Thus we take

$$L_0 = \frac{1}{2k} \int (R_{\mathscr{F}})_s \, d\rho = \frac{1}{2k} \int C(13;24,\odot) [(\mathbf{F}\circ)_1 (\mathbf{F}\circ)_2 \mathbf{R}_{\mathscr{C}}] \, d\rho, \quad (4.16)$$

where $k = 8\pi G / c^4$ and $C(13;24,\odot)$ denotes contraction of the 1st with the 3rd and 2nd with the 4th twist-tensor files via the $\odot$ operation, and

$$d\rho = [d\mathbf{\Omega} (\circ \mathbf{J})_4 (\circ \mathbf{J})_3 (\circ \mathbf{J})_2 (\circ \mathbf{J})_1] \overset{\odot\odot}{\odot\odot} \mathbf{N} \quad (4.17)$$

is the scalar element of volume on $\mathscr{M}$ defined by Eq. (3.45) of I, except that here $\mathbf{N}(q) \in \mathscr{F}_q \wedge \mathscr{F}_q \wedge \mathscr{F}_q \wedge \mathscr{F}_q$,

$$\mathbf{N} \overset{\odot\odot}{\odot\odot} \mathbf{N} = -4!, \quad \text{and} \quad \mathbf{J}(q) \in \mathscr{T}'_q \otimes \mathscr{F}_q.$$

We consider first the variation of $(R_{\mathscr{F}})_s$. This is given by

$$\delta (R_{\mathscr{F}})_s = C(13;24,\odot) [(\delta \mathbf{F}\circ)_1 (\mathbf{F}\circ)_2 \mathbf{R}_{\mathscr{C}} + (\mathbf{F}\circ)_1 (\delta \mathbf{F}\circ)_2 \mathbf{R}_{\mathscr{C}} + (\mathbf{F}\circ)_1 (\mathbf{F}\circ)_2 \delta \mathbf{R}_{\mathscr{C}}]. \quad (4.18)$$

If we now recall Eq. (3.17), we have

$$\mathbf{L} \odot (\delta \mathbf{F} \circ \mathbf{J} + \mathbf{F} \circ \delta \mathbf{J}) = 0,$$

from which we get

$$\delta \mathbf{F} = -\mathbf{F}\circ\delta\mathbf{J}\odot\mathbf{F}. \tag{4.19}$$

Substituting (4.19) into (4.18) results in

$$\begin{aligned}\delta(R_{\mathscr{T}})_s = C(13;24,\odot)[&-(\mathbf{F}\circ\delta\mathbf{J}\odot\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{E}}\\ &-(\mathbf{F}\circ)_1(\mathbf{F}\circ\delta\mathbf{J}\odot\mathbf{F}\circ)_2\mathbf{R}_{\mathscr{E}}\\ &+(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\delta\mathbf{R}_{\mathscr{E}}].\end{aligned} \tag{4.20}$$

The variation of the curvature tensor $\mathbf{R}_E$ is given by

$$\begin{aligned}\mathbf{xy}\overset{\circ}{\circ}\delta\mathbf{R}_{\mathscr{E}}\odot\mathbf{L} &= \delta(D_XD_Y - D_YD_X - D_{[X,Y]})\mathbf{L}\\ &= \delta\mathbf{B}_X\odot D_Y\mathbf{L} + D_X(\delta\mathbf{B}_Y\odot\mathbf{L})\\ &\quad -\delta\mathbf{B}_Y\odot D_X\mathbf{L} - D_Y(\delta\mathbf{B}_X\odot\mathbf{L})\\ &\quad -\delta\mathbf{B}_{[X,Y]}\odot\mathbf{L}\\ &= D_X(\delta\mathbf{B}_Y)\odot\mathbf{L}\\ &\quad - D_Y(\delta\mathbf{B}_X)\odot\mathbf{L} - \delta\mathbf{B}_{[X,Y]}\odot\mathbf{L}.\end{aligned} \tag{4.21}$$

Furthermore, making use of the relation (3.34), we have

$$\begin{aligned}D_X(\delta\mathbf{B}_Y) &= D_X(\mathbf{y}\circ\mathbf{J}\odot\mathbf{F}\circ\delta\mathbf{B})\\ &= [D_X(\mathbf{y}\circ\mathbf{J})]\odot\mathbf{F}\circ\delta\mathbf{B} + \mathbf{y}\circ\mathbf{J}\odot[D_X(\mathbf{F}\circ\delta\mathbf{B})]\\ &= (\nabla_X\mathbf{y})\circ\delta\mathbf{B} + \mathbf{y}\circ\mathbf{J}\odot[D_X(\mathbf{F}\circ\delta\mathbf{B})].\end{aligned} \tag{4.22}$$

Consequently,

$$\begin{aligned}&\mathbf{xy}\overset{\circ}{\circ}\delta\mathbf{R}_{\mathscr{E}}\odot\mathbf{L}\\ &= \mathbf{xy}\overset{\circ}{\circ}\mathbf{T}_{\mathscr{T}}\circ\delta\mathbf{B}\odot\mathbf{L}\\ &+[\mathbf{y}\circ\mathbf{J}\odot D_X(\mathbf{F}\circ\delta\mathbf{B}) - \mathbf{x}\circ\mathbf{J}\odot D_Y(\mathbf{F}\circ\delta\mathbf{B})]\odot\mathbf{L},\end{aligned}$$

i.e.,

$$\delta\mathbf{R}_{\mathscr{E}} = \mathbf{T}_{\mathscr{T}}\circ\delta\mathbf{B} - [1-(\mathbf{12})][\mathbf{J}\odot(\mathbf{12})\mathbf{D}(\mathbf{F}\circ\delta\mathbf{B})]. \tag{4.23}$$

Inserting (4.23) into (4.20) and carrying out the contractions yields

$$\begin{aligned}\delta(R_{\mathscr{T}})_s = &-(\mathbf{F}\circ\delta\mathbf{J}\odot\mathbf{F})\overset{\odot}{\circ}[(\mathbf{1243})\mathbf{R}_{\mathscr{E}} + (\mathbf{14})\mathbf{R}_{\mathscr{E}}]\overset{\odot}{\circ}\mathbf{F}\\ &+[(\mathbf{123})(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{T}_{\mathscr{T}}]\overset{\circ}{\underset{\odot}{\odot}}\delta\mathbf{B} - C(\mathbf{12})\\ &\times[\mathbf{D}(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}].\end{aligned} \tag{4.24}$$

Note now that the last term in (4.24), which is being acted by the operator of covariant differentiation, can be expressed as

$$\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B} = \mathbf{z}\circ\mathbf{J}, \tag{4.25}$$

that is, $\exists\ \mathbf{z}\in\mathscr{T}_q$ such that mapping with $\mathbf{J}$ generates the twist-tensor in $\mathscr{F}_q$ given by the left-hand side of (4.25). Equivalently, we have

$$(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F} = \mathbf{z}. \tag{4.26}$$

Recalling (3.34), we can then write

$$\begin{aligned}&C(\mathbf{12})[\mathbf{D}(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}]\\ &= C(\mathbf{12})\nabla[(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}]\\ &= \nabla\circ[\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}].\end{aligned} \tag{4.27}$$

Hence

$$\begin{aligned}\delta(R_{\mathscr{T}})_s = &-2(\mathbf{F}\circ\delta\mathbf{J}\odot\mathbf{F})\overset{\odot}{\circ}[(\mathbf{14})\mathbf{R}_{\mathscr{E}}]\overset{\odot}{\circ}\mathbf{F}\\ &+[(\mathbf{123})(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{T}_{\mathscr{T}}]\overset{\circ}{\underset{\odot}{\odot}}\delta\mathbf{B}\\ &-\nabla\circ[(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}].\end{aligned} \tag{4.28}$$

The divergence term in (4.28) cannot be integrated out directly in the Lagrangian because it is a nonsymmetric connection in the tangent bundle which allows for torsion. We can, nevertheless, decompose this term as follows[14]:

$$\begin{aligned}&\nabla\circ[(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}]\\ &= \overset{s}{\nabla}\circ[(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}]\\ &\quad -[C(\mathbf{23})\mathbf{T}_{\mathscr{T}}]\circ[(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}],\end{aligned} \tag{4.29}$$

where the part with the standard connection $\overset{s}{\nabla}$ does, in fact, vanish upon integration and may, therefore, be neglected. Rewriting the second term in (4.29) and substituting into (4.28) leads to

$$\begin{aligned}\delta(R_{\mathscr{T}})_s = &-2(\mathbf{F}\circ\delta\mathbf{J}\odot\mathbf{F})\overset{\odot}{\circ}[(\mathbf{14})\mathbf{R}_{\mathscr{E}}]\overset{\odot}{\circ}\mathbf{F}\\ &+[(\mathbf{123})(\mathbf{F}\circ)_1(\mathbf{F}\circ)_2\mathbf{T}_{\mathscr{T}}]\overset{\circ}{\underset{\odot}{\odot}}\delta\mathbf{B}\\ &+[C(\mathbf{23})\mathbf{T}_{\mathscr{T}}]\circ[(\widetilde{\mathbf{F}}\overset{\circ}{\odot}[1-(\mathbf{23})]\delta\mathbf{B})\odot\mathbf{F}].\end{aligned} \tag{4.30}$$

To express this variation in terms of the variation of the twistor connection, we make use of (4.8) and (4.9). Thus, we obtain

$$\begin{aligned}\delta(R_{\mathscr{T}})_s = 4\delta\mathbf{M}&\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}\{[(\widetilde{\mathbf{F}}\odot(\mathbf{14})\mathbf{R}_{\mathscr{E}})\overset{\odot}{\circ}\mathbf{F}]\bullet(\widetilde{\mathbf{F}}\blacktriangle\mathbf{O})\}\\ &-\delta\mathbf{M}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}([(\mathbf{F}\circ)_2(\mathbf{F}\circ)_3(\mathbf{13})\mathbf{T}_{\mathscr{T}}]\overset{\blacktriangle}{\underset{\blacktriangle}{\blacktriangle}}\Lambda)\\ &+\delta\mathbf{M}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}([(C(\mathbf{23})\mathbf{T}_{\mathscr{T}})\circ\\ &\times((\mathbf{24})[1-(\mathbf{23})]\widetilde{\mathbf{F}}\mathbf{F})]\overset{\blacktriangle}{\underset{\blacktriangle}{\blacktriangle}}\Lambda).\end{aligned} \tag{4.31}$$

Because of (4.10), the variation $\delta\mathbf{M}$ in (4.31) cannot be treated as completely arbitrary. In order to be able to do this, we have to project the terms with which $\delta\mathbf{M}$ is contracted onto the same subspace in which $\delta\mathbf{M}$ has values at each point $q\in\mathscr{M}$. Consider a typical term in (4.31) which has the form $[\ ]\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}\delta\mathbf{M}$. Projecting onto the subspace of $\delta\mathbf{M}$ yields

$$\begin{aligned}\{\tfrac{1}{4}[[1+(23)]([\ ]\\ &+\overline{[\ ]})][(\blacktriangle\mathbf{I})_1(\blacktriangle\mathbf{I})_2\\ &+(\blacktriangle\mathbf{O})_1(\blacktriangle\mathbf{O})_2]\\ &+\tfrac{1}{2}([\ ]-(23)\overline{[\ ]})(\blacktriangle\mathbf{O})_1(\blacktriangle\mathbf{I})_2\}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}\delta\mathbf{M},\end{aligned} \tag{4.32}$$

or, equivalently,

$$\begin{aligned}[-\widetilde{\mathbf{J}}\circ\{\tfrac{1}{4}[1+(23)]([\ ]+\overline{[\ ]})[(\blacktriangle\mathbf{I})_1(\blacktriangle\mathbf{I})_2\\ &+(\blacktriangle\mathbf{O})_1(\blacktriangle\mathbf{O})_2]\\ &+\tfrac{1}{2}([\ ]-(23)\overline{[\ ]})(\blacktriangle\mathbf{O})_1(\blacktriangle\mathbf{I})_2]\}\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}[\mathbf{F}\circ\delta\mathbf{M}].\end{aligned} \tag{4.33}$$

If we now apply (4.33) to the first term in (4.31), take into account $\mathbf{I}\odot\mathbf{L} = \mathbf{O}\overset{\blacktriangle}{\blacktriangle}\mathbf{L} = 0$ and $\mathbf{O}\odot\mathbf{L} = \mathbf{I}\overset{\blacktriangle}{\blacktriangle}\mathbf{L} = 0$ for twist-tensors with values in $\mathscr{F}_q$, and write symbolically $\mathbf{R}_{\mathscr{E}} = \mathbf{xyAB}$ (where $\mathbf{x}, \mathbf{y}\in\mathscr{T}'_q$ and $\mathbf{A}, \mathbf{B}, \in\mathscr{H}_q$) to stress the spaces of the files in a typical term of $\mathbf{R}_{\mathscr{G}}$, we get

$$4\delta\mathbf{M}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}\{[(\tilde{\mathbf{F}}\odot(\mathbf{14})\mathbf{R}_{\mathscr{G}})\overset{\circ}{\underset{\circ}{\circ}}\mathbf{F}]\circ(\tilde{\mathbf{F}}\blacktriangle\mathbf{O})\}$$
$$= -4[\mathbf{B}(\mathbf{A}\overset{\blacktriangle}{\blacktriangle}\mathbf{F}\circ\mathbf{x})$$
$$\times(\mathbf{O}\blacktriangle(\mathbf{F}\circ\mathbf{y})\blacktriangle\mathbf{I})]\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(34)(\mathbf{F}\circ\delta\mathbf{M}). \tag{4.34}$$

Observe that for any $\mathbf{A}\in\mathscr{H}_q$ and $\mathbf{C}\in\mathscr{F}_q$

$$\mathbf{A}\overset{\blacktriangle}{\blacktriangle}\mathbf{C} = (\mathbf{I}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I} + \mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{O} + \mathbf{I}\blacktriangle\mathbf{A}\blacktriangle\mathbf{O}$$
$$+ \mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I})\overset{\blacktriangle}{\blacktriangle}(\mathbf{I}\blacktriangle\mathbf{C}\blacktriangle\mathbf{O} + \mathbf{O}\blacktriangle\mathbf{C}\blacktriangle\mathbf{I})$$
$$= (\mathbf{I}\blacktriangle\mathbf{A}\blacktriangle\mathbf{O} + \mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I})\overset{\blacktriangle}{\blacktriangle}$$
$$\times(\mathbf{I}\blacktriangle\mathbf{C}\blacktriangle\mathbf{O} + \mathbf{O}\blacktriangle\mathbf{C}\blacktriangle\mathbf{I})$$
$$= (\mathbf{I}\blacktriangle\mathbf{A}\blacktriangle\mathbf{O})\overset{\blacktriangle}{\blacktriangle}(\mathbf{I}\blacktriangle\mathbf{C}\blacktriangle\mathbf{O})$$
$$+ (\mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I})\overset{\blacktriangle}{\blacktriangle}(\mathbf{O}\blacktriangle\mathbf{C}\blacktriangle\mathbf{I})$$
$$= 2(\mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I})\overset{\blacktriangle}{\blacktriangle}(\mathbf{O}\blacktriangle\mathbf{C}\blacktriangle\mathbf{I}). \tag{4.35}$$

Consequently, (4.34) becomes

$$4\delta\mathbf{M}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}\{[(\tilde{\mathbf{F}}\odot(\mathbf{14})\mathbf{R}_{\mathscr{G}})\overset{\circ}{\underset{\circ}{\circ}}\mathbf{F}]\circ(\tilde{\mathbf{F}}\blacktriangle\mathbf{O})\}$$
$$= -8\{(\mathbf{I}\blacktriangle\mathbf{B}\blacktriangle\mathbf{O} + \mathbf{O}\blacktriangle\mathbf{B}\blacktriangle\mathbf{I})(\mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I})\overset{\blacktriangle}{\blacktriangle}$$
$$(\mathbf{O}\blacktriangle(\mathbf{F}\circ\mathbf{x})\blacktriangle\mathbf{I})(\mathbf{O}\blacktriangle(\mathbf{F}\circ\mathbf{y})\blacktriangle\mathbf{I})\}$$
$$\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(34)(\mathbf{F}\circ\delta\mathbf{M})$$
$$= -8\{(13)(24)[1-(34)](\mathbf{O}\blacktriangle(\mathbf{F}\circ\mathbf{x})\blacktriangle\mathbf{I})$$
$$\overset{\blacktriangle}{\blacktriangle}(\mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I})$$
$$\times(\mathbf{O}\blacktriangle(\mathbf{F}\circ\mathbf{y})\blacktriangle\mathbf{I})(\mathbf{O}\blacktriangle\mathbf{B}\blacktriangle\mathbf{I})\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(34)(\mathbf{F}\circ\delta\mathbf{M})\}. \tag{4.36}$$

The terms in (4.36) are all of the form $(\mathbf{O}\blacktriangle\mathbf{C}\blacktriangle\mathbf{I})$ with values in $\mathscr{F}_2 \underset{\mathrm{aH}}{\otimes} \mathscr{S}_2$. To convert to elements in $\mathscr{F}_2 \underset{\mathrm{H}}{\otimes} \mathscr{S}_2$, thus having the usual isomorphism with Minkowski space[10,14] and also preserving inner products, we have to make each term $\mathbf{C}\in\mathscr{F}$ correspond to $\sqrt{2}i\,(\mathbf{O}\blacktriangle\mathbf{C}\blacktriangle\mathbf{I})\in\mathscr{F}\underset{\mathrm{H}}{\otimes}\mathscr{S}_2$. Consequently,

$$4\delta\mathbf{M}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}\{[(\tilde{\mathbf{F}}\odot(\mathbf{14})\mathbf{R}_{\mathscr{G}})\overset{\circ}{\underset{\circ}{\circ}}\mathbf{F}]\circ(\tilde{\mathbf{F}}\blacktriangle\mathbf{O})\}$$
$$= -2\{(13)(24)[1-(34)](\sqrt{2}i\mathbf{O}\blacktriangle(\mathbf{F}\circ\mathbf{x})\blacktriangle\mathbf{I})$$
$$\overset{\blacktriangle}{\blacktriangle}(\sqrt{2}(i\mathbf{O}\blacktriangle\mathbf{A}\blacktriangle\mathbf{I})(\mathbf{F}\circ\mathbf{y})\blacktriangle\mathbf{I})$$
$$(\sqrt{2}i\mathbf{O}\blacktriangle\mathbf{B}\blacktriangle\mathbf{I})\}\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(34)(\mathbf{F}\bullet\delta\mathbf{M})$$
$$= 2\{(13)(24)[1-(34)]\mathbf{R}\}\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(34)(\mathbf{F}\circ\delta\mathbf{M}), \tag{4.37}$$

where $\mathbf{R}$ is the spinor equivalent of the Ricci tensor for a U^4 space (torsion allowed). Following Witten[17] and Penrose,[3] we have for this case[18]

$$\mathbf{R} = -\tfrac{1}{4}(23)\mathbf{OI}(R)_s + (23)\mathbf{O}[C(23)\mathbf{A}]$$
$$+ (23)\,\overline{[C(23)\mathbf{A}]}\mathbf{I} + 2(14)\mathbf{G}. \tag{4.38}$$

Here $(R)_s \equiv -[(23)\mathbf{OI}]\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}\mathbf{R}$ is the usual curvature scalar,

$$\mathbf{A} = \tfrac{1}{2}[1-(13)(24)]\mathbf{B}, \tag{4.39a}$$

$$\mathbf{B} = (12)\mathbf{B} = (34)\mathbf{B} = \tfrac{1}{4}C(13;57)\mathscr{R}$$

$$(\mathscr{R}\in \text{spinor equivalent of curvature tensor}), \tag{4.39b}$$

$$\mathbf{G} = \tfrac{1}{2}[\mathbf{C}+\mathbf{C}^\dagger], \tag{4.39c}$$

$$\mathbf{C} = (12)\mathbf{C} = (34)\mathbf{C} = \tfrac{1}{4}C(13;68)\mathscr{R}, \tag{4.39d}$$

and

$$(12)\mathbf{C}(23)\mathbf{A} = C(23)\mathbf{A}. \tag{4.39e}$$

Applying similar arguments to the other two terms in (4.31), we find

$$-\delta\mathbf{M}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}([(\mathbf{F}\circ)_2(\mathbf{F}\circ)_3(\mathbf{13})\mathbf{T}_{\mathscr{T}}]\overset{\blacktriangle}{\underset{\blacktriangle}{\blacktriangle}}\mathbf{A})$$
$$= -(i/\sqrt{2})\{[1-(12)](13)(24)[C(13)]$$
$$+ C(24)]\mathbf{T}]\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(\mathbf{F}\circ\delta\mathbf{M}) \tag{4.40}$$

and

$$\delta\mathbf{M}\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}([(C(\mathbf{23})\mathbf{T}_{\mathscr{T}})\circ((\mathbf{24})[1-(\mathbf{23})]\tilde{\mathbf{F}}\mathbf{F})]\overset{\blacktriangle}{\underset{\blacktriangle}{\blacktriangle}}\mathbf{A})$$
$$= -i/\sqrt{2}[1-(12)](13)(24)[([(13)-(23)]\mathbf{O}$$
$$+ [(14)-(24)]\mathbf{I})C(35;46)\mathbf{T}]\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(\mathbf{F}\circ\delta\mathbf{M}), \tag{4.41}$$

where

$$\mathbf{T} = -(13)(24)\mathbf{T} = -\tfrac{1}{2}(23)[\mathbf{O}C(13)\mathbf{T}$$
$$+ (35)(46)(34)\,\overline{(C(13)\mathbf{T})}\mathbf{I}] \tag{4.42}$$

is the spinor equivalent of the torsion tensor.

Combining (4.37), (4.40), and (4.41) yields

$$\delta(R_{\mathscr{T}})_s = \{[1-(12)](13)(24)(2\mathbf{R}$$
$$-[i(12)/\sqrt{2}][C(13)+C(24)]\mathbf{T}$$
$$-[i(12)/\sqrt{2}]([(13)-(23)]\mathbf{O}$$
$$+[(14)-(24)]\mathbf{I})C(35;46)\mathbf{T})\}$$
$$\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}[(34)\mathbf{F}\circ\delta\mathbf{M}]. \tag{4.43}$$

Next we evaluate the variation of the volume element in (4.16). First note that (4.17) may be written as

$$d\rho = d\mathbf{\Omega}\overset{\circ\circ}{\circ\circ}[(\mathbf{J}\odot)_1(\mathbf{J}\odot)_2(\mathbf{J}\odot)_3(\mathbf{J}\odot)_4\mathbf{N}]$$
$$= d\mathbf{\Omega}\overset{\circ\circ}{\circ\circ}\mathbf{\Gamma}', \tag{4.44}$$

where

$$\mathbf{\Gamma}' = (\mathbf{J}\odot)_1(\mathbf{J}\odot)_2(\mathbf{J}\odot)_3(\mathbf{J}\odot)_4\mathbf{N}. \tag{4.45}$$

Varying (4.45) gives

$$\delta\mathbf{\Gamma}' = [(\delta\mathbf{J}\odot)_1(\mathbf{J}\odot)_2(\mathbf{J}\odot)_3(\mathbf{J}\odot)_4$$
$$+ (\mathbf{J}\odot)_1(\delta\mathbf{J}\odot)_2(\mathbf{J}\odot)_3(\mathbf{J}\odot)_4 + \cdots]\mathbf{N}$$
$$= [(\delta\mathbf{J}\odot\mathbf{F}\circ)_1(\mathbf{J}\odot)_1(\mathbf{J}\odot)_2(\mathbf{J}\odot)_3(\mathbf{J}\odot)_4 + \cdots]\mathbf{N}$$
$$= \left[\sum_{k=1}^{4}(\delta\mathbf{J}\odot\mathbf{F}\circ)_k\right]\mathbf{\Gamma}' = (\delta\mathbf{J}\odot\mathbf{F})_s\mathbf{\Gamma}', \tag{4.46}$$

where the scalar invariant $(\delta\mathbf{J}\odot\mathbf{F})_s$ is here defined by

$$(\delta\mathbf{J}\odot\mathbf{F})_s = C(\mathbf{12})(\delta\mathbf{J}\odot\mathbf{F}) = \tilde{\mathbf{F}}\overset{\circ}{\odot}\delta\mathbf{J}. \tag{4.47}$$

If we now vary (4.44) and make use of (4.46) and (4.47), we immediately get

$$\delta\, d\rho = d\mathbf{\Omega}\overset{\circ\circ}{\circ\circ}\delta\mathbf{\Gamma}' = \tilde{\mathbf{F}}\overset{\circ}{\odot}\delta\mathbf{J}\, d\rho, \tag{4.48}$$

or, taking into account (4.9),

$$\delta\, d\rho = -2(\tilde{\mathbf{F}}\blacktriangle\mathbf{O})\overset{\circ}{\underset{\blacktriangle}{\blacktriangle}}\delta\mathbf{M}\, d\rho. \tag{4.49}$$

Again, in order to be able to treat $\delta\mathbf{M}$ arbitrarily, we need to project into the appropriate space. Thus, applying (4.33), we obtain after some straightforward operations

$$\delta\, d\rho = 2[(\tilde{\mathbf{J}}\circ\tilde{\mathbf{F}})\blacktriangle\mathbf{O}]\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(\mathbf{F}\circ\delta\mathbf{M})\, d\rho$$
$$= 2[\mathbf{I}_{\mathscr{T}}\blacktriangle\mathbf{O}]\overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle}(\mathbf{F}\circ\delta\mathbf{M})\, d\rho. \tag{4.50}$$

Moreover, recalling (A27) and subtracting the projection operator into the space spanned by **I** and **O**, we can write

$$\begin{aligned}\mathbf{I}_{\mathscr{F}} &= \tfrac{1}{2}\mathbf{A} - \tfrac{1}{2}\mathbf{I}\otimes\mathbf{O} - \tfrac{1}{2}\mathbf{O}\otimes\mathbf{I}\\ &= \tfrac{1}{2}(\mathbf{I}\wedge\mathbf{O}) - \tfrac{1}{2}\mathbf{I}\otimes\mathbf{O} - \tfrac{1}{2}\mathbf{O}\otimes\mathbf{I}\\ &= -\tfrac{1}{2}[(23)+(24)](\mathbf{I}\otimes\mathbf{O}+\mathbf{O}\otimes\mathbf{I}).\end{aligned} \tag{4.51}$$

Consequently, (4.50) becomes

$$\delta\, d\rho = ([1-(12)](23)\mathbf{OI}) \overset{\blacktriangle\blacktriangle}{\blacktriangle\blacktriangle} [(34)\mathbf{F}\circ\delta\mathbf{M}]\, d\rho. \tag{4.52}$$

Finally, making use of (3.65), (4.43), and (4.52) in the variation of (4.16) leads to the field equations

$$\begin{aligned}(34)\widetilde{\mathbf{J}}\circ\frac{\delta L_0}{\delta\mathbf{M}} &\equiv \frac{1}{2k}[1-(12)](13)(24)\{\tfrac{1}{2}(23)\mathbf{OI}(R\,)_s\\ &\quad + 2(23)(\mathbf{O}[C(23)\mathbf{A}]\\ &\quad + [\,\overline{C(23)\mathbf{A}}\,]\mathbf{I}) + 4(14)\mathbf{G}\\ &\quad - [i(12)/\sqrt{2}\,][C(13)+C(24)]\mathbf{T}\\ &\quad - [i(12)/\sqrt{2}\,]([(13)-(23)]\mathbf{O}\\ &\quad + [(14)-(24)]\mathbf{I})C(35;46)\mathbf{T}\}\\ &= -(34)\widetilde{\mathbf{J}}\circ\frac{\delta L_m}{\delta\mathbf{M}},\end{aligned} \tag{4.53}$$

where $\delta L_0/\delta\mathbf{M}$ and $\delta L_m/\delta\mathbf{M}$ stand symbolically for the variation of the free field and matter Lagrangians minimally coupled to the gauge fields (connections).

Projecting (4.53) from the left with $(\mathbf{O}\blacktriangle)_1(\mathbf{I}\blacktriangle)_2$ and from the right with $(\blacktriangle\mathbf{I})_1(\blacktriangle\mathbf{O})_2$ results in

$$\begin{aligned}&\tfrac{1}{2}(R\,)_s(23)\mathbf{OI} - 2[(23)\mathbf{O}C(23)\mathbf{A}\\ &+ (23)(\,\overline{C(23)\mathbf{A}})\mathbf{I} - 2(14)\mathbf{G}]\\ &= +2k(13)(24)\mathbf{\Sigma},\end{aligned} \tag{4.54}$$

where

$$(\mathbf{O}\blacktriangle)_1(\mathbf{I}\blacktriangle)_2\left[-(34)\widetilde{\mathbf{J}}\circ\frac{\delta L_m}{\delta\mathbf{M}}\right](\blacktriangle\mathbf{I})_1(\blacktriangle\mathbf{O})_2 \equiv (13)(24)\mathbf{\Sigma} \tag{4.55}$$

and $\mathbf{\Sigma}$ is related to the asymmetric total energy–momentum of the matter field.[1]

Taking the symmetric and antisymmetric part of (4.54), gives

$$\tfrac{1}{2}(R\,)_s(23)\mathbf{OI} + 4(14)\mathbf{G} = k[1+(13)(24)]\mathbf{\Sigma} \tag{4.56}$$

and

$$\begin{aligned}&2[(23)\mathbf{O}(C(23)\mathbf{A})\\ &+ (23)(\,\overline{C(23)\mathbf{A}})\mathbf{I}] = k[1-(13)(24)]\mathbf{\Sigma},\end{aligned} \tag{4.57}$$

respectively.

Now consider the projection from the right of (4.53) with $[(\blacktriangle\mathbf{I})_1(\blacktriangle\mathbf{I})_2 + (\blacktriangle\mathbf{O})_1(\blacktriangle\mathbf{O})_2]$. We get

$$\begin{aligned}&-(i/\sqrt{2})(13)(24)[C(13)+C(24)]\mathbf{T}\\ &-(i/\sqrt{2}[(13)-(14)][C(35;46)\mathbf{T}]\mathbf{O}\\ &-(i/\sqrt{2})[(23)-(24)][C(35;46)\mathbf{T}]\mathbf{I}\\ &= 2k(\mathbf{O}\blacktriangle)_1(\mathbf{I}\blacktriangle)_2\left[-(34)\widetilde{\mathbf{J}}\circ\frac{\delta L_m}{\delta\mathbf{M}}\right]\\ &\quad\times[(\blacktriangle\mathbf{I})_1(\blacktriangle\mathbf{I})_2 + (\blacktriangle\mathbf{O})_1(\blacktriangle\mathbf{O})_2]\\ &\equiv -(ik/\sqrt{2})[C(35)+C(46)]\boldsymbol{\tau},\end{aligned} \tag{4.58}$$

where $\boldsymbol{\tau}$ is related to the spin angular momentum of the matter field.

Equation (4.58) may be simplified further by making use of the unique torsion decomposition which we derived in a previous paper[15] [cf. Eq. (37) therein]:

$$\begin{aligned}\mathbf{T} = [1-(13)(24)]\{&-2[\boldsymbol{\theta}+i(24)\boldsymbol{\sigma}](23)\mathbf{OI}\\ &+(45)(12)\overline{\mathbf{H}}\mathbf{I} - (36)\mathbf{HO}\}.\end{aligned} \tag{4.59}$$

Substituting (4.59) into (4.58) yields

$$\begin{aligned}&2[(13)-(14)][(i\boldsymbol{\sigma}-2\boldsymbol{\theta})\mathbf{O}]\\ &\quad+2[(23)-(24)][(i\boldsymbol{\sigma}-2\boldsymbol{\theta})\mathbf{I}] + 2\mathbf{H} + 2(12)\overline{\mathbf{H}}\\ &\quad= k[C(35)+C(46)]\boldsymbol{\tau}.\end{aligned} \tag{4.60}$$

Contracting on (14) files in the above equation gives

$$-6(12)(i\boldsymbol{\sigma}-2\boldsymbol{\theta}) = k[C(14;35)+C(15;46)]\boldsymbol{\tau}, \tag{4.61}$$

and, contracting on (24) files in (4.60), we get

$$-6(i\boldsymbol{\sigma}-2\boldsymbol{\theta}) = k[C(24;35)+C(25;46)]\boldsymbol{\tau}. \tag{4.62}$$

To obtain separate equations for $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$, we take the complex conjugate of (4.61) and make use of the Hermitian property $\boldsymbol{\sigma}=\boldsymbol{\sigma}^\dagger$, $\boldsymbol{\theta}=\boldsymbol{\theta}^\dagger$. It then follows that

$$6(i\boldsymbol{\sigma}+2\boldsymbol{\theta}) = k[C(14;35)+C(15;46)]\overline{\boldsymbol{\tau}}. \tag{4.63}$$

Adding and subtracting (4.62) and (4.63) yields

$$\begin{aligned}\boldsymbol{\theta} = (k/24)\{&[C(14;35)+C(15;46)]\overline{\boldsymbol{\tau}}\\ &+[C(24;35)+C(25;46)]\boldsymbol{\tau}\}\end{aligned} \tag{4.64}$$

and

$$\begin{aligned}\boldsymbol{\sigma} = -(ik/12)\{&[C(14;35)+C(15;46)]\overline{\boldsymbol{\tau}}\\ &-[C(24;35)+C(25;46)]\boldsymbol{\tau}\},\end{aligned} \tag{4.65}$$

respectively.

Inserting (4.64) and (4.65) into (4.60) and projecting with **I** and **O**, we get corresponding equations for **H** and $\overline{\mathbf{H}}$. The physical interpretation of these results and the application to specific matter fields will be the subject of a forthcoming paper.

ACKNOWLEDGMENT

We wish to thank L. F. Urrutia and M. Ryan for fruitful discussions on the subject. We are also grateful to the referee for helpful comments and suggestions. Help with the drawing by Mr. José Luis Torres is also recognized.

APPENDIX: SUMMARY OF TWISTOR ALGEBRA

Twistor space $\mathscr{U}\equiv\mathscr{U}_{2,2}$ is a four-dimensional complex vector space with a Hermitian-type inner product $\langle \mathbf{s}|\mathbf{t}\rangle$, antilinear in $\mathbf{s}\in\mathscr{U}$ and linear in $\mathbf{t}\in\mathscr{U}$, having the signature $(+\,+\,-\,-)$. The *dual twistor space* $\mathscr{U}'\equiv\mathscr{U}'_{2,2}$ is the set of linear functionals on $\mathscr{U}$; thus each element $\mathbf{k}'\in\mathscr{U}'$ acts on each element $\mathbf{l}\in\mathscr{U}$ to produce a complex number, denoted $\mathbf{k}'\circ\mathbf{l}$ or $\mathbf{l}\circ\mathbf{k}'$.

The symbol $\circ$ will also denote contraction operations (inner multiplication) between twist-tensors; for example, $\mathbf{M}\circ\mathbf{K}'$, $\mathbf{M}\overset{\circ}{\circ}\mathbf{K}'$, $\mathbf{M}\overset{\circ\circ}{\circ\circ}\mathbf{K}'$, $\mathbf{K}'\circ\mathbf{M}$, $\mathbf{K}'\overset{\circ}{\circ}\mathbf{M}$, $\mathbf{K}'\overset{\circ\circ}{\circ\circ}\mathbf{M}$, etc., for $\mathbf{M}\in\mathscr{U}^{\otimes m}$ and $\mathbf{K}'\in\mathscr{U}'^{\otimes k}$. The number of circles $\circ$ is the number c of contractions performed, and $m\geqslant c$, $k\geqslant c$. The case with **M** followed by $\mathbf{K}'$ with c circles in between indicates the contractions of the last c files of **M** with the first c files of $\mathbf{K}'$

in sequence, e.g., $\mathbf{M}\overset{\circ}{\circ}\mathbf{K}'$ has the second to last file of $\mathbf{M}$ contracted with the first file of $\mathbf{K}'$ and the last file of $\mathbf{M}$ contracted with the second file of $\mathbf{K}'$. Similarly, $\mathbf{K}'$ followed by $\mathbf{M}$ with c circles is defined. The above definition also applies to other cases, for example, $\mathbf{B}\circ\mathbf{C}$, where $\mathbf{B}\in\mathscr{U}\otimes\mathscr{U}'$ and $\mathbf{C}\in\mathscr{U}\otimes\mathscr{U}'$ or $\mathbf{C}\in\mathscr{U}$.

Conjugation operations: The conjugation operation $\mathbf{l}\in\mathscr{U}\to\hat{\mathbf{l}}\in\mathscr{U}'$ is an antilinear map defined by the equation

$$\hat{\mathbf{l}}\circ\mathbf{m} = \langle\mathbf{l}|\mathbf{m}\rangle^* \tag{A1}$$

for all $\mathbf{m}\in\mathscr{U}$. An important property, which follows from the Hermiticity of the inner product, is

$$\hat{\mathbf{l}}\circ\mathbf{m} = (\mathbf{l}\circ\hat{\mathbf{m}})^*, \tag{A2}$$

where * denotes complex conjugation of a complex number.

The conjugation operation $\mathbf{B}\in\mathscr{U}^{\otimes 2}\to\hat{\mathbf{B}}\in\mathscr{U}'^{\otimes 2}$ is an antilinear map defined by the equation

$$\hat{\mathbf{B}}\overset{\circ}{\circ}\mathbf{lm} = (\mathbf{B}\overset{\circ}{\circ}\hat{\mathbf{l}}\hat{\mathbf{m}})^* \tag{A3}$$

for all $\mathbf{l},\mathbf{m}\in\mathscr{U}$, where the notation $\mathbf{B}\overset{\circ}{\circ}\mathbf{s}'\mathbf{t}'$ and $\mathbf{C}'\overset{\circ}{\circ}\mathbf{lm}$ denotes the action of $\mathbf{B}\in\mathscr{U}^{\otimes 2}$ as a bilinear functional on $\mathbf{s}',\mathbf{t}'\in\mathscr{U}'$ and the action of $\mathbf{C}'\in\mathscr{U}'^{\otimes 2}$ on $\mathbf{l},\mathbf{m}\in\mathscr{U}$. Some important properties which follow from (A3) are

$$(\mathbf{l}\otimes\mathbf{m})\hat{} = \hat{\mathbf{l}}\otimes\hat{\mathbf{m}}, \tag{A4}$$

$$\hat{\mathbf{B}}\overset{\circ}{\circ}\mathbf{A} = (\mathbf{B}\overset{\circ}{\circ}\hat{\mathbf{A}})^*, \tag{A5}$$

$$(\mathbf{B}\circ\hat{\mathbf{l}})\hat{} = \hat{\mathbf{B}}\circ\mathbf{l}, \tag{A6}$$

for $\mathbf{l},\mathbf{m}\in\mathscr{U}$ and $\mathbf{A},\mathbf{B}\in\mathscr{U}^{\otimes 2}$, where the notation $\mathbf{B}\hat{}\equiv\hat{\mathbf{B}}$ and $\mathbf{l}\hat{}\equiv\hat{\mathbf{l}}$ was used for convenience. We also have the result

$$\mathbf{B}\in\mathscr{U}^{\wedge 2}\Rightarrow\hat{\mathbf{B}}\in\mathscr{U}'^{\wedge 2},$$

where $\mathscr{U}^{\wedge 2}$ and $\mathscr{U}'^{\wedge 2}$ are the subspaces of antisymmetric twist-tensors in $\mathscr{U}^{\otimes 2}$ and $\mathscr{U}'^{\otimes 2}$ respectively.

The conjugation operation can be readily generalized to higher order twist-tensor spaces; for example, $\mathbf{M}\in\mathscr{U}^{\otimes 4}\to\hat{\mathbf{M}}\in\mathscr{U}'^{\otimes 4}$ is defined by the equation

$$\hat{\mathbf{M}}\overset{\circ\circ}{\circ\circ}\mathbf{lmpr} = (\mathbf{M}\overset{\circ\circ}{\circ\circ}\hat{\mathbf{l}}\hat{\mathbf{m}}\hat{\mathbf{p}}\hat{\mathbf{r}})^* \tag{A7}$$

for all $\mathbf{l},\mathbf{m},\mathbf{p},\mathbf{r}\in\mathscr{U}$, where the notation $\mathbf{M}\overset{\circ\circ}{\circ\circ}\mathbf{s}'\mathbf{t}'\mathbf{u}'\mathbf{v}'$ and $\mathbf{K}'\overset{\circ\circ}{\circ\circ}\mathbf{lmpr}$ denotes the action of $\mathbf{M}\in\mathscr{U}^{\otimes 4}$ on $\mathbf{s}',\mathbf{t}',\mathbf{u}',\mathbf{v}'\in\mathscr{U}'$ and the action of $\mathbf{K}'\in\mathscr{U}'^{\otimes 4}$ on $\mathbf{l},\mathbf{m},\mathbf{p},\mathbf{r},\in\mathscr{U}$. The properties (A4), (A5), (A6) can be then extended as follows:

$$(\mathbf{l}\otimes\mathbf{m}\otimes\mathbf{r}\otimes\mathbf{s})\hat{} = \hat{\mathbf{l}}\otimes\hat{\mathbf{m}}\otimes\hat{\mathbf{r}}\otimes\hat{\mathbf{s}}, \tag{A8}$$

$$(\mathbf{A}\otimes\mathbf{B})\hat{} = \hat{\mathbf{A}}\otimes\hat{\mathbf{B}}, \tag{A9}$$

$$\hat{\mathbf{M}}\overset{\circ\circ}{\circ\circ}\mathbf{N} = (\mathbf{M}\overset{\circ\circ}{\circ\circ}\hat{\mathbf{N}})^*, \tag{A10}$$

$$(\mathbf{M}\overset{\circ}{\circ}\hat{\mathbf{B}})\hat{} = \hat{\mathbf{M}}\overset{\circ}{\circ}\mathbf{B} \tag{A11}$$

for $\mathbf{l},\mathbf{m},\mathbf{r},\mathbf{s},\in\mathscr{U}$ and $\mathbf{A},\mathbf{B}\in\mathscr{U}^{\otimes 2}$ and $\mathbf{M},\mathbf{N}\in\mathscr{U}^{\otimes 4}$. Note also that

$$\mathbf{M}\in\mathscr{U}^{\wedge 4}\Rightarrow\hat{\mathbf{M}}\in\mathscr{U}'^{\wedge 4}$$

$$\mathbf{M}\in\mathscr{U}^{\wedge 2}\otimes\mathscr{U}^{\wedge 2}\Rightarrow\hat{\mathbf{M}}\in\mathscr{U}'^{\wedge 2}\otimes\mathscr{U}'^{\wedge 2}$$

Duals of antisymmetric twist-tensors: Assume that we have a given privileged element $\mathbf{\Lambda}\in\mathscr{U}^{\wedge 4}$ satisfying the requirement

$$\hat{\mathbf{\Lambda}}\overset{\circ\circ}{\circ\circ}\mathbf{\Lambda} = 4!. \tag{A12}$$

Since the subspace $\mathscr{U}^{\wedge 4}$ has dimension one, it follows that any element of $\mathscr{U}^{\wedge 4}$ satisfying the requirement (A12) must be of the form $e^{i\phi}\mathbf{\Lambda}$ for some angle ϕ. Let $\mathbf{\Lambda}'\in\mathscr{U}'^{\wedge 4}$ be defined as $\mathbf{\Lambda}'\equiv\hat{\mathbf{\Lambda}}$. Now let $(\mathscr{U},\mathbf{\Lambda})$ denote the space $\mathscr{U}$ with the element $\mathbf{\Lambda}\in\mathscr{U}^{\wedge 4}$ given as a part of the structure. From now on we shall abbreviate this notation with the symbol $\mathscr{U}$, i.e., $\mathscr{U}\equiv(\mathscr{U},\mathbf{\Lambda})$.

The operations of forming the dual, $\mathbf{B}\in\mathscr{U}^{\wedge 2}\to{\star}\mathbf{B}\in\mathscr{U}'^{\wedge 2}$ and $\mathbf{C}'\in\mathscr{U}'^{\wedge 2}\to{\star}\mathbf{C}'\in\mathscr{U}^{\wedge 2}$ are defined by the equations

$$\star\mathbf{B} = \tfrac{1}{2}\mathbf{\Lambda}'\overset{\circ}{\circ}\mathbf{B}, \tag{A13}$$

$$\star\mathbf{C}' = \tfrac{1}{2}\mathbf{\Lambda}\overset{\circ}{\circ}\mathbf{C}'. \tag{A14}$$

Some properties of this operation are

$$\star\star\mathbf{B} = \mathbf{B}, \tag{A15}$$

$$\star\star\mathbf{C}' = \mathbf{C}', \tag{A16}$$

$$\star\mathbf{A}\overset{\circ}{\circ}\mathbf{B} = \mathbf{A}\overset{\circ}{\circ}\star\mathbf{B}, \tag{A17}$$

for $\mathbf{A},\mathbf{B}\in\mathscr{U}^{\wedge 2}$ and $\mathbf{C}'\in\mathscr{U}'^{\wedge 2}$.

The double dual operations $\mathbf{M}\in\mathscr{U}^{\wedge 2}\otimes\mathscr{U}^{\wedge 2}\to\star\mathbf{M}\star\in\mathscr{U}'^{\wedge 2}\otimes\mathscr{U}'^{\wedge 2}$ and $\mathbf{K}'\in\mathscr{U}'^{\wedge 2}\otimes\mathscr{U}'^{\wedge 2}\to\star\mathbf{K}'\star\in\mathscr{U}^{\wedge 2}\otimes\mathscr{U}^{\wedge 2}$ are defined by

$$\star\mathbf{M}\star = \tfrac{1}{4}\mathbf{\Lambda}'\overset{\circ}{\circ}\mathbf{M}\overset{\circ}{\circ}\mathbf{\Lambda}', \tag{A18}$$

$$\star\mathbf{K}'\star = \tfrac{1}{4}\mathbf{\Lambda}\overset{\circ}{\circ}\mathbf{K}'\overset{\circ}{\circ}\mathbf{\Lambda}. \tag{A19}$$

Inner product in $\mathscr{U}^{\wedge 2}$. The inner product $\mathbf{A},\mathbf{B}\in\mathscr{U}^{\wedge 2}\to\mathbf{A}\odot\mathbf{B}\in\mathbb{C}$, where $\mathbb{C}$ is the complex number system, is a bilinear, symmetric map defined by the equation

$$\mathbf{A}\odot\mathbf{B} = \tfrac{1}{2}\mathbf{A}\overset{\circ}{\circ}\mathbf{\Lambda}'\overset{\circ}{\circ}\mathbf{B}. \tag{A20}$$

It follows from (A20) that

$$\mathbf{A}\odot\mathbf{B} = \star\mathbf{A}\overset{\circ}{\circ}\mathbf{B} = \mathbf{A}\overset{\circ}{\circ}\star\mathbf{B}, \tag{A21}$$

$$\mathbf{A}\wedge\mathbf{B} = \tfrac{1}{2}(\mathbf{A}\odot\mathbf{B})\mathbf{\Lambda}, \tag{A22}$$

where $\mathbf{A}\wedge\mathbf{B}$ is defined as

$$\mathbf{A}\wedge\mathbf{B} = \tfrac{1}{4}\sum_P \operatorname{sgn}(P)P(\mathbf{A}\otimes\mathbf{B}). \tag{A23}$$

The sum in (A23) goes over all permutations P acting on the four files of the tensor $\mathbf{A}\otimes\mathbf{B}$; sgn $(P)=1$ if P is even, and $\operatorname{sgn}(P)=-1$ if P is odd.

Related to this inner product are the following tensor contraction operations:

$$\mathbf{M}\in\mathscr{U}^{\wedge 2}\otimes\mathscr{U}^{\wedge 2},\quad \mathbf{B}\in\mathscr{U}^{\wedge 2}\to\mathbf{M}\odot\mathbf{B}\in\mathscr{U}^{\wedge 2},$$

$$\mathbf{B}\in\mathscr{U}^{\wedge 2},\quad \mathbf{M}\in\mathscr{U}^{\wedge 2}\otimes\mathscr{U}^{\wedge 2}\to\mathbf{B}\odot\mathbf{M}\in\mathscr{U}^{\wedge 2},$$

$$\mathbf{M},\mathbf{N}\in\mathscr{U}^{\wedge 2}\otimes\mathscr{U}^{\wedge 2}\to\mathbf{M}\odot\mathbf{N}\in\mathscr{U}^{\wedge 2}\otimes\mathscr{U}^{\wedge 2}.$$

These are defined by

$$\mathbf{M}\odot\mathbf{B} = \tfrac{1}{2}\mathbf{M}\overset{\circ}{\circ}\mathbf{\Lambda}'\overset{\circ}{\circ}\mathbf{B}, \tag{A24}$$

$$\mathbf{B}\odot\mathbf{M} = \tfrac{1}{2}\mathbf{B}\overset{\circ}{\circ}\mathbf{\Lambda}'\overset{\circ}{\circ}\mathbf{M}, \tag{A25}$$

$$\mathbf{M}\odot\mathbf{N} = \tfrac{1}{2}\mathbf{M}\overset{\circ}{\circ}\mathbf{\Lambda}'\overset{\circ}{\circ}\mathbf{N}. \tag{A26}$$

It follows from (A24) and (A15) that

$$\tfrac{1}{2}\mathbf{\Lambda}\odot\mathbf{B} = \tfrac{1}{2}\mathbf{B}\odot\mathbf{\Lambda} = \mathbf{B} \tag{A27}$$

for all $\mathbf{B}\in\mathscr{U}^{\wedge 2}$, so $\tfrac{1}{2}\mathbf{\Lambda}$ acts as an identity operator on $\mathscr{U}^{\wedge 2}$ via the $\odot$ inner product.

Definition: $B\in\mathscr{U}^{\wedge 2}$ is *null* iff $B\odot B = 0$.

Simple elements of $\mathscr{U}^{\wedge 2}$ *and* $\mathscr{U}'^{\wedge 2}$: A twist-tensor $\mathbf{B}\in\mathscr{U}^{\wedge 2}$ is *simple* if it has the form $\mathbf{B}=\mathbf{l}\wedge\mathbf{m}$, where $\mathbf{l},\mathbf{m}\in\mathscr{U}$. Similarly, $\mathbf{C}'\in\mathscr{U}'^{\wedge 2}$ is *simple* if it has the form $\mathbf{C}'=\mathbf{p}'\wedge\mathbf{q}'$, where $\mathbf{p}', \mathbf{q}' \in \mathscr{U}'$.

We construct the subspace $\mathscr{S}(\mathbf{B})$ of $\mathscr{U}$ and the subspace $\mathscr{S}'(\mathbf{C}')$ of $\mathscr{U}'$ as

$$\mathscr{S}(\mathbf{B}) = \{\mathbf{B}\circ\mathbf{n}' | \mathbf{n}' \in \mathscr{U}'\}, \tag{A28}$$

$$\mathscr{S}'(\mathbf{C}') = \{\mathbf{C}'\circ\mathbf{r} | \mathbf{r} \in \mathscr{U}\}. \tag{A29}$$

Relevant to some of the subsequent discussion are the following easily proved theorems about simple twist-tensors:

Theorem A.1: If $\mathbf{B} \neq 0$ and $\mathbf{B} = \mathbf{l} \wedge \mathbf{m}$ (i.e., $\mathbf{B}$ is simple), then $\mathbf{l}, \mathbf{m}$ is a basis for $\mathscr{S}(\mathbf{B})$. Similarly, if $\mathbf{C}' \neq 0'$ and $\mathbf{C}' = \mathbf{p}' \wedge \mathbf{q}'$ (i.e., $\mathbf{C}'$ is simple), then $\mathbf{p}', \mathbf{q}'$ is a basis for $\mathscr{S}'(\mathbf{C}')$.

Theorem A.2: $\mathbf{B}$ is simple and not 0 iff $\mathscr{S}(\mathbf{B})$ has dimension 2. Similarly, $\mathbf{C}'$ is simple and not $0'$ iff $\mathscr{S}'(\mathbf{C}')$ has dimension 2.

Theorem A.3: $\mathbf{B}$ is simple iff $\star\mathbf{B}$ is simple. Also $\mathbf{B} \neq 0$ iff $\star\mathbf{B} \neq 0'$.

Theorem A.4: Suppose $\mathbf{B}$ is simple. Then $\mathscr{S}(\mathbf{B})$ and $\mathscr{S}'(\star\mathbf{B})$ are perpendicular, i.e., $\mathbf{l}\circ\mathbf{q}' = 0$ for all $\mathbf{l} \in \mathscr{S}(\mathbf{B})$ and all $\mathbf{q}' \in \mathscr{S}'(\star\mathbf{B})$.

Theorem A.5: $\mathbf{B}$ is simple iff $\mathbf{B}\circ\star\mathbf{B} = 0$.

Theorem A.6: $\mathbf{B}$ is simple iff $\mathbf{B}$ is null.

Theorem A.7: Suppose $\mathbf{A} \neq 0$ and $\mathbf{B} \neq 0 \in \mathscr{U}^{\wedge 2}$ are simple. Then $\mathbf{A} \odot \mathbf{B} \neq 0$ iff $\mathscr{U} = \mathscr{S}(\mathbf{A}) \oplus \mathscr{S}(\mathbf{B})$.

Theorem A.8: Suppose $\mathbf{A} \neq 0$ and $\mathbf{B} \neq 0 \in \mathscr{U}^{\wedge 2}$ are simple. Then $\mathscr{U} = \mathscr{S}(\mathbf{A}) \oplus \mathscr{S}(\mathbf{B})$ iff $\mathscr{U}' = \mathscr{S}'(\star\mathbf{A}) \oplus \mathscr{S}'(\star\mathbf{B})$.

Projection operators: Let $\mathbf{A}, \mathbf{B} \in \mathscr{U}^{\wedge 2}$. Then $\mathbf{A}\circ\star\mathbf{B}$ maps $\mathscr{U}$ into $\mathscr{U}$ by the operation $(\mathbf{A}\circ\star\mathbf{B})\circ\mathbf{l}$ for $\mathbf{l} \in \mathscr{U}$. The *range* $\mathscr{S}(\mathbf{A}\circ\star\mathbf{B})$ of this map is a subspace of $\mathscr{U}$ defined as

$$\mathscr{S}(\mathbf{A}\circ\star\mathbf{B}) = \{\mathbf{A}\circ\star\mathbf{B}\circ\mathbf{m} | \mathbf{m} \in \mathscr{U}\}. \tag{A30}$$

Theorem A.9: Suppose $\mathbf{A} \neq 0$ and $\mathbf{B} \neq 0$ are simple and that $\mathbf{A} \odot \mathbf{B} \neq 0$. Then $\mathscr{S}(\mathbf{A}) = \mathscr{S}(\mathbf{A}\circ\star\mathbf{B})$ and $\mathscr{S}(\mathbf{B}) = \mathscr{S}(\mathbf{B}\circ\star\mathbf{A})$.

Theorem A.10: Suppose $\mathbf{A} \neq 0$ and $\mathbf{B} \neq 0$ are simple, and that $\mathbf{A} \odot \mathbf{B} = 2$. Then $S_\mathbf{A} = -\mathbf{A}\circ\star\mathbf{B}$ and $S_\mathbf{B} = -\mathbf{B}\circ\star\mathbf{A}$ are mutually orthogonal projection operators onto $\mathscr{S}(\mathbf{A})$ and $\mathscr{S}(\mathbf{B})$, respectively.

According to this theorem, we have

$$S_\mathbf{A}\circ S_\mathbf{A} = S_\mathbf{A}, \tag{A31a}$$

$$S_\mathbf{B}\circ S_\mathbf{B} = S_\mathbf{B}, \tag{A31b}$$

$$S_\mathbf{A}\circ S_\mathbf{B} = S_\mathbf{B}\circ S_\mathbf{A} = 0, \tag{A31c}$$

$$\mathbf{l} \in \mathscr{S}(\mathbf{A}) \Leftrightarrow S_\mathbf{A}\circ\mathbf{l} = \mathbf{l}, \tag{A31d}$$

$$\mathbf{l} \in \mathscr{S}(\mathbf{B}) \Leftrightarrow S_\mathbf{B}\circ\mathbf{l} = \mathbf{l}, \tag{A31e}$$

$$\mathbf{l} \in \mathscr{S}(\mathbf{A}) \Leftrightarrow S_\mathbf{B}\circ\mathbf{l} = 0, \tag{A31f}$$

$$\mathbf{l} \in \mathscr{S}(\mathbf{B}) \Leftrightarrow S_\mathbf{A}\circ\mathbf{l} = 0. \tag{A31g}$$

Note that $S_\mathbf{A} + S_\mathbf{B}$ is the unit operator on $\mathscr{U}$, i.e.,

$$(S_\mathbf{A} + S_\mathbf{B})\circ\mathbf{l} = \mathbf{l} \tag{A32}$$

for all $\mathbf{l} \in \mathscr{U}$.

Real twist-tensors: A twist-tensor $\mathbf{B} \in \mathscr{U}^{\wedge 2}$ is real iff $\star\hat{\mathbf{B}} = \mathbf{B}$. Also, $\mathbf{M} \in \mathscr{U}^{\wedge 2} \otimes \mathscr{U}^{\wedge 2}$ is real iff $\star\hat{\mathbf{M}}\star = \mathbf{M}$. In particular, it follows from (A15) that $\mathbf{A}$ is real.

Let $\mathscr{E} \equiv \mathscr{E}_{2,4} = \{\mathbf{P} | \mathbf{P} \in \mathscr{U}^{\wedge 2}, \mathbf{P} \text{ real}\}$. Then $\mathscr{E}$ is a vector space over the reals. The inner product $\odot$ in $\mathscr{E}$ has signature $(+ + - - - -)$.

Correspondence with Minkowski space: The *infinity twist-tensor* $\mathbf{I} \in \mathscr{E}$ is a given privileged element such that $\mathbf{I} \neq 0$, and $\mathbf{I} \odot \mathbf{I} = 0$. The space $(\mathscr{U}, \mathbf{A}, \mathbf{I})$, which has $\mathbf{I}$ as part of the structure, has the property that the group of linear transformations on it that preserves its structure is a faithful representation of the Poincaré group $\mathscr{P}$.

The set $\mathscr{W} = \{\mathbf{P} | \mathbf{P} \in \mathscr{E}, \mathbf{P} \odot \mathbf{P} = 0, \mathbf{I} \odot \mathbf{P} = 2\}$ is a hypersurface in $\mathscr{E}$ which has a one-to-one correspondence with the elements of Minkowski space–time. This hypersurface is invariant under the action of $\mathscr{P}$.

The *tangent space* $\mathscr{W}_\mathbf{P}$ at a given element $\mathbf{P} \in \mathscr{W}$ is the set of elements $\mathbf{T} \in \mathscr{E}$ which are tangent to the hypersurface $\mathscr{W}$ at $\mathbf{P}$. It follows that $\mathscr{W}_\mathbf{P} = \{\mathbf{T} | \mathbf{T} \in \mathscr{E}, \mathbf{I} \odot \mathbf{T} = 0, \mathbf{P} \odot \mathbf{T} = 0\}$. The inner product $\odot$ in $\mathscr{W}_\mathbf{P}$ is the Minkowski inner product with signature $(+ - - -)$. This inner product makes $\mathscr{W}$ a pseudo-Riemannian space. The standard connection on $\mathscr{W}$ leads to a curvature tensor that is zero everywhere on $\mathscr{W}$. Thus $\mathscr{W}$ is intrinsically a flat space.

The *origin twist-tensor* $\mathbf{O} \in \mathscr{W}$ (note, therefore, that $\mathbf{O} \odot \mathbf{O} = 0$ and $\mathbf{I} \odot \mathbf{O} = 2$) is any chosen element of $\mathscr{W}$ regarded as a reference point. The space $(\mathscr{U}, \mathbf{A}, \mathbf{I}, \mathbf{O})$ has the property that the group of linear transformations on it that preserves its structure is a faithful representation of the Lorentz group. The structure of this space leads uniquely to a bilinear antisymmetric inner product on $\mathscr{U}$, an adjoint operation on $\mathscr{U}$, and a unique decomposition $\mathscr{U} = \mathscr{S}(\mathbf{I}) \oplus \mathscr{S}(\mathbf{O})$, which serve to relate twistors to the Dirac bispinor space.

Inner product on $(\mathscr{U}, \mathbf{A}, \mathbf{I}, \mathbf{O})$: First note that by Theorem A.7 the subspaces $\mathscr{S}(\mathbf{I})$ and $\mathscr{S}(\mathbf{O})$ have the property that $\mathscr{U} = \mathscr{S}(\mathbf{I}) \oplus \mathscr{S}(\mathbf{O})$. Also, by Theorem A.10, $S_\mathbf{I} = -\mathbf{I}\circ\star\mathbf{O}$ and $S_\mathbf{O} = -\mathbf{O}\circ\star\mathbf{I}$ are mutually orthogonal projection operators onto the subspaces $\mathscr{S}(\mathbf{I})$ and $\mathscr{S}(\mathbf{O})$, respectively.

The inner product $\mathbf{l}, \mathbf{m} \in \mathscr{U} \rightarrow \mathbf{l}_\blacktriangle\mathbf{m} \in \mathbb{C}$ is a bilinear, antisymmetric map defined by the equation

$$\mathbf{l}_\blacktriangle\mathbf{m} = -\mathbf{l}\circ(\star\mathbf{I} + \star\mathbf{O})\circ\mathbf{m}. \tag{A33}$$

Related to this inner product are the following tensor contraction operations:

$$\mathbf{B} \in \mathscr{U}^{\otimes 2}, \quad \mathbf{l} \in \mathscr{U} \rightarrow \mathbf{B}_\blacktriangle\mathbf{l} \in \mathscr{U},$$

$$\mathbf{l} \in \mathscr{U}, \quad \mathbf{B} \in \mathscr{U}^{\otimes 2} \rightarrow \mathbf{l}_\blacktriangle\mathbf{B} \in \mathscr{U},$$

$$\mathbf{A} \in \mathscr{U}^{\otimes 2}, \quad \mathbf{B} \in \mathscr{U}^{\otimes 2} \rightarrow \mathbf{A}_\blacktriangle\mathbf{B} \in \mathscr{U}^{\otimes 2}.$$

These are defined by the equations

$$\mathbf{B}_\blacktriangle\mathbf{l} = -\mathbf{B}\circ(\star\mathbf{I} + \star\mathbf{O})\circ\mathbf{l}, \tag{A34}$$

$$\mathbf{l}_\blacktriangle\mathbf{B} = -\mathbf{l}\circ(\star\mathbf{I} + \star\mathbf{O})\circ\mathbf{B}, \tag{A35}$$

$$\mathbf{A}_\blacktriangle\mathbf{B} = -\mathbf{A}\circ(\star\mathbf{I} + \star\mathbf{O})\circ\mathbf{B}. \tag{A36}$$

It also follows directly from (A33)–(A36) that

$$\mathbf{m}_\blacktriangle(\mathbf{B}_\blacktriangle\mathbf{l}) = (\mathbf{m}_\blacktriangle\mathbf{B})_\blacktriangle\mathbf{l}, \tag{A37}$$

$$(\mathbf{A}_\blacktriangle\mathbf{B})_\blacktriangle\mathbf{l} = \mathbf{A}_\blacktriangle(\mathbf{B}_\blacktriangle\mathbf{l}), \tag{A38}$$

$$\mathbf{A}_\blacktriangle\mathbf{l} = -\mathbf{l}_\blacktriangle\tilde{\mathbf{A}}, \tag{A39}$$

$$(\mathbf{A}_\blacktriangle\mathbf{B})^{\sim} = -\tilde{\mathbf{B}}_\blacktriangle\tilde{\mathbf{A}} \tag{A40}$$

for $\mathbf{l}, \mathbf{m} \in \mathscr{U}$ and $\mathbf{A}, \mathbf{B} \in \mathscr{U}^{\otimes 2}$.

As a consequence of Theorems A5 and A10, we have

Theorem A.11: Acting via the $\blacktriangle$ inner product, $\mathbf{I}$ and $\mathbf{O}$ are mutually orthogonal projection operators onto $\mathscr{S}(\mathbf{I})$ and

$\mathscr{S}(\mathbf{O})$, respectively.

Thus, according to this theorem, we have

$$\mathbf{I}\blacktriangle\mathbf{I} = \mathbf{I}, \tag{A41a}$$

$$\mathbf{O}\blacktriangle\mathbf{O} = \mathbf{O}, \tag{A41b}$$

$$\mathbf{I}\blacktriangle\mathbf{O} = \mathbf{O}\blacktriangle\mathbf{I} = 0, \tag{A41c}$$

$$\mathbf{l}\in\mathscr{S}(\mathbf{I})\Leftrightarrow\mathbf{I}\blacktriangle\mathbf{l} = \mathbf{l}, \tag{A41d}$$

$$\mathbf{l}\in\mathscr{S}(\mathbf{O})\Leftrightarrow\mathbf{O}\blacktriangle\mathbf{l} = \mathbf{l}, \tag{A41e}$$

$$\mathbf{l}\in\mathscr{S}(\mathbf{I})\Leftrightarrow\mathbf{O}\blacktriangle\mathbf{l} = 0, \tag{A41f}$$

$$\mathbf{l}\in\mathscr{S}(\mathbf{O})\Leftrightarrow\mathbf{I}\blacktriangle\mathbf{l} = 0. \tag{A41g}$$

Note that $\mathbf{I}+\mathbf{O}$, acting via the $\blacktriangle$ inner product, is the unit operator on $\mathscr{U}$, i.e.,

$$(\mathbf{I}+\mathbf{O})\blacktriangle\mathbf{l} = \mathbf{l}, \tag{A42}$$

for all $\mathbf{l}\in\mathscr{U}$.

The basic properties of this inner product are

$\mathbf{l}\blacktriangle\mathbf{m}$ is bilinear in $\mathbf{l}$, $\mathbf{m}\in\mathscr{U}$,

$\mathbf{l}\blacktriangle\mathbf{m} = -\mathbf{m}\blacktriangle\mathbf{l}$ for all $\mathbf{l}$, $\mathbf{m}\in\mathscr{U}$ (antisymmetry),

$\mathbf{l}$, $\mathbf{m}$ independent elements of $\mathscr{S}(\mathbf{I})\Rightarrow\mathbf{l}\blacktriangle\mathbf{m}\neq 0$,

$\mathbf{l}$, $\mathbf{m}$ independent elements of $\mathscr{S}(\mathbf{O})\Rightarrow\mathbf{l}\blacktriangle\mathbf{m}\neq 0$,

$\mathbf{l}\in\mathscr{S}(\mathbf{I})$ and $\mathbf{m}\in\mathscr{S}(\mathbf{O})\Rightarrow\mathbf{l}\blacktriangle\mathbf{m} = 0$,

$\mathbf{l}, \mathbf{m}\in\mathscr{S}(\mathbf{I})\Rightarrow\mathbf{l}\wedge\mathbf{m} = -(\mathbf{l}\blacktriangle\mathbf{m})\mathbf{I}$,

$\mathbf{l}, \mathbf{m}\in\mathscr{S}(\mathbf{O})\Rightarrow\mathbf{l}\wedge\mathbf{m} = -(\mathbf{l}\blacktriangle\mathbf{m})\mathbf{O}$.

As a natural extension of (A33) we define the *double product* $\overset{\blacktriangle}{\blacktriangle}$, or *double contraction* with the $\blacktriangle$ product, $\mathbf{T}$, $\mathbf{W}\in\mathscr{U}^{\otimes 2}\rightarrow\mathbf{T}\overset{\blacktriangle}{\blacktriangle}\mathbf{W}\in\mathbb{C}$ as a bilinear, symmetric map given by the equation

$$\begin{aligned}\mathbf{T}\overset{\blacktriangle}{\blacktriangle}\mathbf{W} &= [\mathbf{T}\circ({}^\star\mathbf{I}+{}^\star\mathbf{O})]\overset{\circ}{\circ}[({}^\star\mathbf{I}+{}^\star\mathbf{O})\circ\mathbf{W}]\\ &= -[({}^\star\mathbf{I}+{}^\star\mathbf{O})\circ\mathbf{T}\circ({}^\star\mathbf{I}+{}^\star\mathbf{O})]\overset{\circ}{\circ}\mathbf{W}.\end{aligned} \tag{A43}$$

An important property, which follows from (A43), is

$$(\mathbf{l}\otimes\mathbf{m})\overset{\blacktriangle}{\blacktriangle}(\mathbf{r}\otimes\mathbf{s}) = (\mathbf{l}\blacktriangle\mathbf{r})(\mathbf{m}\blacktriangle\mathbf{s}) \tag{A44}$$

for $\mathbf{l}$, $\mathbf{m}$, $\mathbf{r}$, $\mathbf{s}\in\mathscr{U}$.

Making use of some of our previous results, we can relate the $\overset{\blacktriangle}{\blacktriangle}$ and $\odot$ products in $\mathscr{U}^{\wedge 2}$ by

Theorem A.12: For $\mathbf{T}$, $\mathbf{W}\in\mathscr{U}^{\wedge 2}$

$$\mathbf{T}\overset{\blacktriangle}{\blacktriangle}\mathbf{W} = \tfrac{1}{2}\mathbf{T}\odot(\mathbf{I}+\mathbf{O})(\mathbf{I}+\mathbf{O})\odot\mathbf{W} - \mathbf{T}\odot\mathbf{W}. \tag{A45}$$

As a consequence of this theorem, we have

$$\mathbf{I}\overset{\blacktriangle}{\blacktriangle}\mathbf{I} = \mathbf{O}\overset{\blacktriangle}{\blacktriangle}\mathbf{O} = 2, \tag{A46}$$

$$\mathbf{I}\overset{\blacktriangle}{\blacktriangle}\mathbf{O} = \mathbf{O}\overset{\blacktriangle}{\blacktriangle}\mathbf{I} = 0. \tag{A47}$$

It also follows from (A45) that for $\mathbf{T}\in\mathscr{U}^{\wedge 2}$

$$\mathbf{I}\overset{\blacktriangle}{\blacktriangle}\mathbf{T} = \mathbf{O}\odot\mathbf{T}, \tag{A48}$$

$$\mathbf{O}\overset{\blacktriangle}{\blacktriangle}\mathbf{T} = \mathbf{I}\odot\mathbf{T}. \tag{A49}$$

Hence, $\mathbf{T}\in\mathscr{W}_{\mathbf{O}}\Leftrightarrow\mathbf{I}\overset{\blacktriangle}{\blacktriangle}\mathbf{T} = \mathbf{O}\overset{\blacktriangle}{\blacktriangle}\mathbf{T} = 0$. Furthermore, for $\mathbf{T}$, $\mathbf{W}\in\mathscr{W}_{\mathbf{O}}$ we have

$$\mathbf{T}\overset{\blacktriangle}{\blacktriangle}\mathbf{W} = -\mathbf{T}\odot\mathbf{W}. \tag{A50}$$

Equations (A48) and (A49) allow us to invert (A45) to obtain an expression for the $\odot$ in terms of the $\overset{\blacktriangle}{\blacktriangle}$ product:

$$\mathbf{T}\odot\mathbf{W} = \tfrac{1}{2}\mathbf{T}\overset{\blacktriangle}{\blacktriangle}(\mathbf{I}+\mathbf{O})(\mathbf{I}+\mathbf{O})\overset{\blacktriangle}{\blacktriangle}\mathbf{W} - \mathbf{T}\overset{\blacktriangle}{\blacktriangle}\mathbf{W}. \tag{A51}$$

Adjoint operation: We shall make use of the fact that the quantity

$$\begin{aligned}S_{\mathbf{I}}+S_{\mathbf{O}} &= -(\mathbf{I}\circ{}^\star\mathbf{O}+\mathbf{O}\circ{}^\star\mathbf{I})\\ &= (\mathbf{I}-\mathbf{O})\circ({}^\star\mathbf{I}-{}^\star\mathbf{O}) = -(\mathbf{I}+\mathbf{O})\circ({}^\star\mathbf{I}+{}^\star\mathbf{O})\end{aligned} \tag{A52}$$

is the unit operator on $\mathscr{U}$, and

$$\begin{aligned}\widetilde{S}_{\mathbf{I}}+\widetilde{S}_{\mathbf{O}} &= -({}^\star\mathbf{O}\circ\mathbf{I}+{}^\star\mathbf{I}\circ\mathbf{O})\\ &= ({}^\star\mathbf{I}-{}^\star\mathbf{O})\circ(\mathbf{I}-\mathbf{O}) = -({}^\star\mathbf{I}+{}^\star\mathbf{O})\circ(\mathbf{I}+\mathbf{O})\end{aligned} \tag{A53}$$

is the unit operator on $\mathscr{U}'$, acting by the $\circ$ operation in both cases.

The adjoint operation $\mathbf{l}\in\mathscr{U}\rightarrow\bar{\mathbf{l}}\in\mathscr{U}$ is an antilinear map defined by the equation

$$\bar{\mathbf{l}} = \hat{\mathbf{l}}\circ(\mathbf{I}-\mathbf{O}) = -(\mathbf{I}-\mathbf{O})\circ\hat{\mathbf{l}}. \tag{A54}$$

It follows from (A52) and (A53) that

$$\hat{\mathbf{l}} = \bar{\mathbf{l}}\circ({}^\star\mathbf{I}-{}^\star\mathbf{O}) = -({}^\star\mathbf{I}-{}^\star\mathbf{O})\circ\bar{\mathbf{l}}. \tag{A55}$$

Similarly, we define the adjoint operation $\mathbf{B}\in\mathscr{U}^{\otimes 2}\rightarrow\overline{\mathbf{B}}\in\mathscr{U}^{\otimes 2}$ as the antilinear map given by the equation

$$\overline{\mathbf{B}} = -(\mathbf{I}-\mathbf{O})\circ\widehat{\overline{\mathbf{B}}}\circ(\mathbf{I}-\mathbf{O}). \tag{A56}$$

From this we obtain

$$\widehat{\mathbf{B}} = -({}^\star\mathbf{I}-{}^\star\mathbf{O})\circ\overline{\mathbf{B}}\circ({}^\star\mathbf{I}-{}^\star\mathbf{O}). \tag{A57}$$

Some basic properties of the adjoint operation are

$\mathbf{l}\in\mathscr{U}\rightarrow\bar{\mathbf{l}}\in\mathscr{U}$ is one-to-one, $\mathbf{l}\in\mathscr{U}\rightarrow\bar{\mathbf{l}}\in\mathscr{U}$ is antilinear,

$\bar{\bar{\mathbf{l}}} = \mathbf{l}$ for all $\mathbf{l}\in\mathscr{U}$, $(\mathbf{l}\blacktriangle\mathbf{m})^* = \bar{\mathbf{l}}\blacktriangle\bar{\mathbf{m}}$ for all $\mathbf{l}$, $\mathbf{m}\in\mathscr{U}$,

$\mathbf{l}\in\mathscr{S}(\mathbf{I})\Rightarrow\bar{\mathbf{l}}\in\mathscr{S}(\mathbf{O})$, $\mathbf{l}\in\mathscr{S}(\mathbf{O})\Rightarrow\bar{\mathbf{l}}\in\mathscr{S}(\mathbf{I})$,

$\bar{\mathbf{I}} = \mathbf{O}$, $\bar{\mathbf{O}} = \mathbf{I}$,

$\mathbf{l}\blacktriangle\overline{\mathbf{B}}\blacktriangle\mathbf{m} = (\bar{\mathbf{l}}\blacktriangle\mathbf{B}\blacktriangle\bar{\mathbf{m}})^*$, $\overline{\mathbf{B}\blacktriangle\mathbf{l}} = \overline{\mathbf{B}}\blacktriangle\bar{\mathbf{l}}$,

$$\overline{\overline{\mathbf{B}}} = \mathbf{B}, \quad \text{for } \mathbf{B}\in\mathscr{U}^{\otimes 2} \text{ and } \mathbf{l}, \mathbf{m}\in\mathscr{U}. \tag{A58}$$

The adjoint operation can be readily extended to higher order twist tensors. Thus $\mathbf{M}\in\mathscr{U}^{\otimes 4}\rightarrow\overline{\mathbf{M}}\in\mathscr{U}^{\otimes 4}$ is defined by

$$\overline{\mathbf{M}} = \widehat{\mathbf{M}}[\circ(\mathbf{I}-\mathbf{O})]_1[\circ(\mathbf{I}-\mathbf{O})]_2[\circ(\mathbf{I}-\mathbf{O})]_3[\circ(\mathbf{I}-\mathbf{O})]_4, \tag{A59}$$

where $[\circ(\mathbf{I}-\mathbf{O})]_k$ for $k = 1, 2, 3, 4$ are linear operators acting to the left on the k th twistor file of $\widehat{\mathbf{M}}$ numbered from right to left (in analogy to the notation introduced in I).

Theorem A.13: The adjoint of $\mathbf{B}\in\mathscr{E}$ (real twist-tensor) is related to $\mathbf{B}$ by

$$\begin{aligned}\overline{\mathbf{B}} &= \mathbf{I}\blacktriangle\mathbf{B}\blacktriangle\mathbf{O} + \mathbf{O}\blacktriangle\mathbf{B}\blacktriangle\mathbf{I} + \tfrac{1}{2}(\mathbf{I}\otimes\mathbf{O}+\mathbf{O}\otimes\mathbf{I})\overset{\blacktriangle}{\blacktriangle}\mathbf{B}\\ &= -\tfrac{1}{2}[(\mathbf{I}-\mathbf{O})\otimes(\mathbf{I}-\mathbf{O})]\overset{\blacktriangle}{\blacktriangle}\mathbf{B}+\mathbf{B}.\end{aligned} \tag{A60}$$

Thus, if $\mathbf{B}\in\mathscr{W}_{\mathbf{O}}$, it follows directly that

$$\overline{\mathbf{B}} = \mathbf{B}. \tag{A61}$$

Relation to the Dirac bispinor space: The space ($\mathscr{U}$, $\mathbf{\Lambda}$, $\mathbf{I}$, $\mathbf{O}$) with the bilinear inner product $\mathbf{l}\blacktriangle\mathbf{m}$ and the adjoint operation $\bar{\mathbf{l}}$, together with the given decomposition $\mathscr{U} = \mathscr{S}_2\oplus\overline{\mathscr{S}}_2$, where $\mathscr{S}_2\equiv\mathscr{S}(\mathbf{I})$ and $\overline{\mathscr{S}}_2\equiv\mathscr{S}(\mathbf{O})$, is isomorphic to the Dirac bispinor space. However, the Hermitian type inner product

$$(\mathbf{l}|\mathbf{m}) = -i\,\bar{\mathbf{l}}\blacktriangle\mathbf{m} \tag{A62}$$

is also occasionally used for Dirac spinors instead of $\langle \mathbf{l}|\mathbf{m}\rangle$. The latter can be reexpressed by means of the operations of the Dirac bispinor space as

$$\langle \mathbf{l}|\mathbf{m}\rangle = \bar{\mathbf{l}}\blacktriangle(\mathbf{I}_2 - \bar{\mathbf{I}}_2)\blacktriangle\mathbf{m}, \tag{A63}$$

where $\mathbf{I}_2 = \mathbf{I}$ and $\bar{\mathbf{I}}_2 = \mathbf{O}$.

Minkowski 4-vector space, as it is usually constructed from Dirac spin-tensors, is the subspace $\bar{\mathscr{S}}_2 \otimes_H \mathscr{S}_2$ of Hermitian spin-tensors in $\bar{\mathscr{S}}_2 \otimes \mathscr{S}_2$. Here, $\mathbf{A} \in \bar{\mathscr{S}}_2 \otimes \mathscr{S}_2$ is called *Hermitian* iff $\bar{\bar{\mathbf{A}}} = \mathbf{A}$. If we antisymmetrize the elements of $\bar{\mathscr{S}}_2 \otimes_H \mathscr{S}_2$ we get a subspace $\bar{\mathscr{S}}_2 \wedge_H \mathscr{S}_2$ of antisymmetric Hermitian tensors. This subspace coincides with $i\mathscr{W}_O$, the space obtained by multiplying each element of $\mathscr{W}_O$ by i. The elements of $i\mathscr{W}_O$ are pure imaginary with respect to the definition $\star\hat{\mathbf{A}} = \mathbf{A}$ of reality for $\mathbf{A} \in \mathscr{U}^{\wedge 2}$. The inner product $\odot$ in $i\mathscr{W}_O$ has signature $(+ + + -)$. Also $\mathbf{A}\odot\mathbf{B} = -\mathbf{A} \overset{\blacktriangle}{\blacktriangle} \mathbf{B}$ for $\mathbf{A}$, $\mathbf{B}$ in $\mathscr{W}_O$ and $i\mathscr{W}_O$.

In order to construct a basis for $i\mathscr{W}_O$, let $\mathbf{h}_1$, $\mathbf{h}_2$ be a basis for $\mathscr{S}_2$ such that $\mathbf{h}_1\blacktriangle\mathbf{h}_2 = 1$. Then $\bar{\mathbf{h}}_{1'}$, $\bar{\mathbf{h}}_{2'}$, is a basis for $\bar{\mathscr{S}}_2$ with $\bar{\mathbf{h}}_{1'}\blacktriangle\bar{\mathbf{h}}_{2'} = 1$. We can therefore write

$$\mathbf{T}_0 = \tfrac{1}{2}(\bar{\mathbf{h}}_{1'} \wedge \mathbf{h}_1 + \bar{\mathbf{h}}_{2'} \wedge \mathbf{h}_2), \tag{A64a}$$

$$\mathbf{T}_1 = -\tfrac{1}{2}(\bar{\mathbf{h}}_{1'} \wedge \mathbf{h}_2 + \bar{\mathbf{h}}_{2'} \wedge \mathbf{h}_1), \tag{A64b}$$

$$\mathbf{T}_2 = -\tfrac{1}{2}i(\bar{\mathbf{h}}_{1'} \wedge \mathbf{h}_2 - \bar{\mathbf{h}}_{2'} \wedge \mathbf{h}_1), \tag{A64c}$$

$$\mathbf{T}_3 = -\tfrac{1}{2}(\bar{\mathbf{h}}_{1'} \wedge \mathbf{h}_1 - \bar{\mathbf{h}}_{2'} \wedge \mathbf{h}_2), \tag{A64d}$$

which constitute an orthonormal basis for $i\mathscr{W}_O$ satisfying the relations $\mathbf{T}_\mu \odot \mathbf{T}_\nu = g_{\mu\nu}$, where $g_{\mu\nu} = 0$ for $\mu \neq \nu$ and $g_{11} = g_{22} = g_{33} = -g_{00} = 1$. The set $i\,\mathbf{T}_\mu$ for $\mu = 0, 1, 2, 3$ is a basis for $\mathscr{W}_O$. Including two more independent elements, for example, $\frac{1}{2}(\mathbf{I} + \mathbf{O})$ and $\frac{1}{2}(\mathbf{I} - \mathbf{O})$, together with the $i\,\mathbf{T}_\mu$ ($\mu = 0, 1, 2, 3$), gives an orthogonal basis for $\mathscr{E}$.

[1] R. Utiyama, Phys. Rev. **101**, 1597 (1956); T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961); K. Hayashi and A. Bregman, Ann. Phys. **75**, 562 (1973); H. P. Durr, Nuovo Cimento A **4**, 187 (1971); R. Utiyama and T. Fukuyama, Prog. Theor. Phys. **45**, 612 (1971); F. W. Hehl, P. Van Der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).

[2] C. P. Luehr and M. Rosenbaum, J. Math. Phys. **21**, 1432 (1980).

[3] R. Penrose, Ann. Phys. **10**, 171 (1960).

[4] R. Penrose, in *Quantum Theory and Beyond*, edited by Ted Bastin (Cambridge U. P., Cambridge, 1971).

[5] C. F. Von Weiszacker, in *Quantum Theory and Beyond*, edited by Ted Bastin (Cambridge U.P., Cambridge, 1971).

[6] M. Carmeli [see, e.g., *Group Theory and General Relativity*, (McGraw-Hill, New York, 1977), Chap. 9, and references therein] has developed an SL $(2,C)$ theory of the gravitational field. This theory, however, is not a gauge theory in the Utiyama sense since it begins with curved space–time and subsequently introduces a complex three-dimensional linear space, the basis vectors of which are the infinitesimal generators of the group SL$(2,C)$ as a tangent space at each space–time point. Moreover, as a consequence of this formalism, the Lagrangian density required to obtain the complete set of field equations mixes gravitational field variables with matter field variables in a way which does not seem to be in accord with the minimal coupling principle.

[7] E. Schmutzer, Z. Naturforsch. **19a**, 1027 (1964).

[8] We use the notation of Ref. 10. Other authors use the symbols $\sigma_\mu{}^{AB'}$ (Ref. 3), $\sigma_{j\dot{A}B}$ (Ref. 7), or $g_{\mu\dot{A}B}$ (Ref. 9) to denote these quantities.

[9] A. Jamiolkowski, Rep. Math. Phys. **2**, 1 (1971).

[10] C. P. Luehr and M. Rosenbaum, J. Math. Phys. **9**, 2225 (1968).

[11] R. Penrose, J. Math. Phys. **8**, 345 (1967).

[12] R. O. Wells, Jr., Bull. Amer. Math. Soc. **1**, 296 (1979).

[13] R. Penrose, in *Quantum Gravity an Oxford Symposium*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1975); R. Penrose, in *Group Theory in Non-Linear Problems*, edited A. O. Barut (Reidel, Dordrecht, 1974).

[14] C. P. Luehr and M. Rosenbaum, J. Math. Phys. **15**, 1120 (1974).

[15] C. P. Luehr, M. Rosenbaum, M. Ryan, and L. C. Shepley, J. Math. Phys. **18**, 965 (1977).

[16] Boldface numerals in the symbol (**mn**) will be used to denote permutation of vector files (composed of elements in $\mathscr{E}_q$ or $\mathscr{T}_q$) to distinguish from permutation or twistor files for which we use lightface numerals.

[17] L. Witten, Phys. Rev. **113**, 357 (1959).

[18] R. Gambini and L. Herrera, J. Math. Phys. **21**, 1449 (1980).

Properties of Mayer cluster expansion

Udayan Mohanty
Department of Chemistry, B-014, University of California at San Diego, La Jolla, California 92093

(Received 15 December 1981; accepted for publication 19 March 1982)

The Umbral algebra developed by Rota and his co-workers is used to show that the Mayer cluster expansion of the canonical partition function is related to the Bell polynomials. The algebra is also used to find a representation of the partition function and a rederivation of Mayer's first theorem. Finally, it is shown that in the "tree approximation" for the cluster integrals, the summation of Mayer's expression for the canonical-ensemble partition function for a finite number of particles could be performed using Dénes' and Rényi's theorems in graph theory.

PACS numbers: 05.20.Gg, 02.30. + g

INTRODUCTION

This article shows that the Umbral algebra developed by Rota and his collaborators[1,2] can be used to shed light on properties of the Mayer[3] cluster expansion of the canonical partition function. Recently Gibbs, Bagchi, and Mohanty (GBM) and Donoghue and Gibbs[4] were successful in developing a theory of vapor condensation along the lines set forth by Mayer several decades ago. The results of GBM motivated this work.

This article is organized as follows. The Mayer equations are briefly reviewed. Then the Umbral algebra is introduced. This algebra is used to derive a recursion formula for the cluster expansion of the partition function. For volume independent cluster integrals, the cluster expansion of the partition function is shown to be expressible in terms of A^{-1}, where A is an invertible shift invariant operator, and powers of volume V. Finally, it is pointed out that in the tree approximation for the cluster integrals b_l, the complete summation of the Mayer's expression for the canonical-ensemble partition function for a finite number of particles, could be performed with the aid of some theorems in graph theory.

MAYER EQUATIONS

Consider a classical system consisting of N particles enclosed in a volume V. Let the temperature of the system be T. The configuration partition function for the system is given by

$$Z_N(T, V) = \frac{1}{N!} \int \cdots \int W_N(\mathbf{r}_1, \ldots, \mathbf{r}_N) d\mathbf{r}_1 \cdots d\mathbf{r}_N. \tag{1}$$

The quantity $W_N(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is defined as

$$e^{-V(\mathbf{r}_1,\ldots,\mathbf{r}_N)/k_BT}, \tag{2}$$

where $V(\mathbf{r}_1,\ldots,\mathbf{r}_N)$ is the interaction energy between the particles, and k_B is the Boltzmann constant.

A set of cluster functions $U_1(\mathbf{r}_1)$, $U_2(\mathbf{r}_1, \mathbf{r}_2), \ldots$, is introduced by the equations

$$\begin{aligned}
W_1(\mathbf{r}_1) &= U_1(\mathbf{r}_1) = 1,\\
W_2(\mathbf{r}_1, \mathbf{r}_2) &= U(1, 2) + U(1)U(2)\\
&\equiv U_2(\mathbf{r}_1, \mathbf{r}_2) + U_1(\mathbf{r}_1)U_1(\mathbf{r}_2),\\
W_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) &= U(1, 2, 3) + U(1, 2)U(3) + U(2, 3)U(1)\\
&\quad + U(3, 1)U(2) + U(1)U(2)U(3).
\end{aligned} \tag{3}$$

The total number of terms in U_l is

$$\sum_{k=l}^{(1/2)l(l-1)} C(l, k),$$

where $C(l, k)$ is the number of connected graphs of l labelled points and k lines. The jth connected cluster integral $b_j(T, V)$ is defined[3] in terms of the cluster function U_j as

$$Vj!\, b_j(T, V) = \int \cdots \int U(1, 2, \ldots, j) d\mathbf{r}_1 \cdots d\mathbf{r}_j. \tag{4}$$

The left-hand side of the above equation denotes the sum over the "weights" of all connected graphs of j labelled vertices.

What Mayer[3] showed is that the partition function is expressible directly in terms of the connected cluster integrals

$$Z_N(T, V) = \sum_{\substack{\{m_l\}\\ \sum_l l m_l = N}} \prod_{l=1}^{N} \frac{[Vb_l(T, V)]^{m_l}}{m_l!} \equiv \frac{Q_N(T, V)}{N!}. \tag{5}$$

The grand partition function is defined in terms of $Q_N(T, V)$ as follows:

$$\Xi(T, V, z) = \sum_{N=0}^{\infty} \frac{Q_N(T, V)}{N!} z^N, \tag{6}$$

where z is the fugacity of the system. The generating function of the grand canonical partition function is given by

$$\sum_{N=0}^{\infty} \frac{Q_N(T, V)}{N!} z^N = \exp\left[V \sum_{l=1}^{\infty} l!\, b_l(T, V) \frac{z^l}{l!}\right]. \tag{7a}$$

That the above expression holds over a very wide class of weight functions is known as Mayer's First Theorem.[5] To be precise, let $G(N)$ be a graph with N labelled points and let $W[G(N)]$ be the weight associated with graph $G(N)$. Let the weights have the properties that (a) $W[G(N)]$ depends on the topology of $G(N)$, but not on the labelling of the points, and (b) $W[G(N)] = \Pi W[C(l)]$, where the product goes over all disjoint connected graphs $C(l)$ of $G(N)$. Then, Eq. (7a) continues to hold if one makes the replacement

$$\begin{aligned}
Vl!\, b_l(T, V) &\to \sum_{C(l)} W[C(l)],\\
\frac{Q_N}{N!} &\to \sum_{G(N)} W[G(N)].
\end{aligned} \tag{7b}$$

UMBRAL ALGEBRA

Let B denote the commutative algebra of all polynomials in a single variable x with coefficients in a field of characteristic zero. Let B^* be the vector space of all linear functions of B. Denote the action of a linear functional L on a polynomial $q(x)$ by $\langle L|q(x)\rangle$. The vector space B^* is made into an algebra by defining the product of two linear functionals L and M as

$$\langle LM|x^n\rangle = \sum_{k=0}^{n} \binom{n}{k} \langle L|x^k\rangle\langle M|x^{n-k}\rangle. \tag{8}$$

The identity for the product defined above is given by the action of the linear function ω_c on $q(x)$,

$$\langle \omega_c|q(x)\rangle = q(c), \tag{9}$$

where c is an arbitrary constant.

The vector space of linear functions B^* with the above product and identity is called the Umbral algebra.[2]

A delta functional is a linear functional L with the property that $\langle L|1\rangle = 0$ and $\langle L|x\rangle \neq 0$. A polynomial sequence is a sequence of polynomials $q_n(x)$, $n = 0,1,2,\ldots$, where $q_n(x)$ is exactly of degree n for all n. However, a sequence of polynomials $p_n(x)$ is of binomial type if it satisfies the identity

$$p_n(x+a) = \sum_{k=0}^{n} \binom{n}{k} p_k(a)p_{n-k}(x), \qquad p_0(x) = 1. \tag{10}$$

A polynomial sequence $q_n(x)$ is the associated sequence for a delta functional L if

$$\langle L^k|q_n(x)\rangle = n!\,\delta_{n,k} \quad \forall n, k \geqslant 0. \tag{11}$$

The product of any number of linear functionals can be computed by using sequences of binomial type, in place of x^n. We have the following proposition[2]:

Proposition 1: If $q_n(x)$ is a sequence of binomial type, and if $L_1, L_2, \ldots, L_k$ are linear functionals, then

$$\langle L_1, L_2, \ldots, L_k|q_n(x)\rangle = \sum_{\sum_{l=1}^{k} j_l = n} \binom{n}{j_1,\ldots,j_k} \langle L_1|q_{j_1}(x)\rangle \cdots \langle L_k|q_{j_k}(x)\rangle. \tag{12}$$

One of the key properties of the product of linear functions follows from Proposition 1. If

$$\langle L|1\rangle = \langle L|x\rangle = \cdots = \langle L|x^{m-1}\rangle = 0, \tag{13}$$

then $\langle L^k|x^n\rangle = 0$ for $n < km$ and $\langle L^k|x^{km}\rangle = (km)!/(m!)^k \langle L|x^m\rangle^k$.

BELL POLYNOMIALS AND RECURSION RELATION

Now, we have all the tools necessary to study the algebraic structure of the Bell recursion formula and many of its properties. Over a field of characteristic zero define the generic delta functional L by

$$\langle L|x^n\rangle = x_n, \quad n \geqslant 1, \qquad \langle L|1\rangle = 0. \tag{14}$$

The *conjugate polynomials for the generic delta functional* are the Bell polynomials[6]

$$Y_n(x; x_1, x_2, \ldots, x_n) = \sum_{k=0}^{n} B_{n,k} x^k. \tag{15}$$

The coefficients of x^k in (15) follow from Proposition 1:

$$B_{n,k} = \frac{\langle L^k|x^n\rangle}{k!} = n! \sum_{\substack{\{m_j\} \\ \Sigma jm_j = n \\ \Sigma m_j = k}} \prod_{j=1}^{n} \left(\frac{x_j}{j!}\right)^{m_j} \frac{1}{m_j!}. \tag{16}$$

From (16) one can see that

$$Y_n(1; x_1, x_2, \ldots, x_n) = n! \sum_{\substack{\{m_j\} \\ \Sigma jm_j = n}} \prod_{j=1}^{n} \left(\frac{x_j}{j!}\right)^{m_j} \frac{1}{m_j!} \equiv Y_n(x_1, x_2, \ldots, x_n). \tag{17}$$

A comparison of (17) with (1) shows that the partition function $Q_N(T, V)$ for a system of N particles enclosed in a volume V and at a temperature T is related to the Nth Bell polynomial

$$Q_N(T, V) = Y_N(1!Vb_1, 2!Vb_2, \ldots, N!\,Vb_N). \tag{18}$$

An equivalent representation of Y_n is given by

$$Y_n(g_1, \ldots, g_n) = e^{g(x)} \frac{d^n}{dx^n} e^{-g(x)} \equiv e^{-g} D_x^n e^g. \tag{19}$$

From (19) it is a simple exercise to show that $Y_{n+1}(g_1, \ldots, g_{n+1})$ satisfies the recursion formula

$$Y_{n+1}(g_1, \ldots, g_{n+1}) = \sum_{k=0}^{n} \binom{n}{k} Y_{n-k}(g_1, \ldots, g_{n-k}) g_{k+1}. \tag{20}$$

On combining Eqs. (18) with (20) one gets a useful recursion formula for the partition function,

$$\frac{Q_{N+1}(T, V)}{(N+1)!} = \sum_{k=0}^{N} \left(\frac{k+1}{N+1}\right) Vb_{k+1}(T, V) \frac{Q_{N-k}(T, V)}{(N-k)!}. \tag{21}$$

The generating function $B(t)$ of the Bell Polynomials is defined as

$$B(t) = \sum_{n=0}^{\infty} \frac{Y_n}{n!} t^n. \tag{22}$$

By the use of Eq. (17), one can show that

$$\ln B(t) = \sum_{n=1}^{\infty} \frac{x_n}{n!} t^n. \tag{23}$$

Since $Q_n(T, V) = Y_n(1!Vb_1, \ldots, n!Vb_n)$, Eq. (23) leads us to Mayer's First Theorem. In the Appendix we will give a different proof of this theorem.

A REPRESENTATION FOR THE PARTITION FUNCTION

Every linear functional L defines a multiplication operator $\tilde{\theta}(L)^*$ on B^*. The operator $\tilde{\theta}(L)^*$ maps the linear functional M to the linear functional L as $\tilde{\theta}(L)^*M = L \cdot M$. $\tilde{\theta}(L)$ is called a shift invariant operator.

A delta operator is defined as an operator of the form $P = \tilde{\theta}(L)$, where L is a delta functional. This operator has many of the properties of the derivative operator. For instance, $Pa = 0$ for every constant a.

The generalization of the relationship between the derivative operator D and the sequence $q_n(x) = x_n$ is that between a delta operator and its associated sequence $q_n(x)$. We now state a powerful theorem—the so-called Transfer formula[1] that enables one to compute the associated sequence for a delta operator.

Theorem 1: If $P = AD$ is a delta operator, where A is an invertible shift invariant operator, and if $q_n(x)$ is the associated sequence of polynomials for P, then

$$q_n(x) = P'A^{-n-1}x^n, \quad \forall n \geqslant 0, \tag{24}$$

where P' is the Pincherle derivative[1] of the operator P.

Corollary 1: If $P = AD$ is a delta operator with associated sequence $q_n(x)$, then

$$q_n(x) = xA^{-n}x^{-n-1}, \quad \forall n \geqslant 1. \tag{25}$$

The proof[1] of this corollary follows from Theorem 1. By definition one has

$$\begin{aligned} P'A^{-n-1}x^n &= (A + DA')A^{-n-1}x^n \\ &= A^{-n}x^n + nA'A^{-n-1}x^{n-1}. \end{aligned} \tag{26}$$

The above equation can be further simplified by recalling the Pincherle derivative[1] of $(A^{-n})'$,

$$A^{-n}x^n - (A^{-n}X - XA^{-n})x^{n-1} = xA^{-n}x^{-n-1}. \tag{27}$$

This completes the proof. Note that in Eq. (27), X is the multiplicative operator. Thus X: $p(x) \to xp(x)$.

We can now make the connection with the partition function $Q_N(T, V)$. Let $Q_N(T, V)$ be a sequence of basic polynomials for a shift-invariant delta operator P. The delta operator is determined by $b_l(T)$; in fact, $P = (d/dV)A$, where A is invertible. Then, Theorem 1 and Corollary 1 allow us to express Q_N in terms of A^{-1} and powers of V:

$$Q_N(T, V) = VA^{-N}V^{N-1}. \tag{28}$$

PARTITION FUNCTION IN CAYLEY TREE APPROXIMATION

In the Cayley Tree approximation, the connected cluster integrals are given in terms of the first irreducible star integrals $\beta_1(T)$ as[4]

$$b_l(T) = \frac{l^{l-2}}{l!}\beta_1^{l-1}(T). \tag{29}$$

The configuration partition function (5) then becomes

$$Q_N(T, V) = N! \sum_{\substack{\{m_l\} \\ \sum_l lm_l = N}} \prod_{l=1}^{N} (V\beta_1^{l-1})^{m_l} \left(\frac{l^{l-2}}{l!}\right)^{m_l} \frac{1}{m_l!}. \tag{30}$$

We can rewrite Eq. (30) as

$$\begin{aligned} Q_N(T, V) &= \sum_{M=1}^{N} \left(\frac{V}{\beta_1}\right)^M (\beta_1)^N N! \sum_{\substack{\{m_l\} \\ \Sigma lm_l = N \\ \Sigma m_l = M}} \prod_{l=1}^{N} \left(\frac{l^{l-2}}{l!}\right)^{m_l} \frac{1}{m_l!} \end{aligned} \tag{31a}$$

$$\equiv \sum_{M=1}^{N} \left(\frac{V}{\beta_1}\right)^M \beta_1^N G_M(N). \tag{31b}$$

The last equation defines $G_M(N)$. The quantity $G_M(N)$ can be interpreted as the number of all graphs with N labelled vertices consisting of M disjoint trees. This was first proven by Dénes.[7] However, Rényi[8] showed that $G_M(N)$ can be expressed in a simpler form,

$$G_M(N) = \frac{1}{M!}\sum_{j=0}^{M} \left(-\frac{1}{2}\right)^j \binom{M}{j} \binom{N-1}{M+j-1} N^{N-M-j}(M+j)!. \tag{32}$$

The main advantage[9] of Eq. (32) is that it provides a simple way to numerically evaluate the partition function. However, if higher order star integrals are used to approximate b_l, then the recursion formula (21) is indispensible in performing a complete summation of the canonical ensemble partition function.

CONCLUDING REMARKS

Bell polynomials are related to the Mayer cluster expansion of the canonical partition function. This observation results in a recursion formula for the partition function. The recursion formula was used by Gibbs, Bagchi, and Mohanty[4,10] to develop a theory of vapor condensation.

For volume independent cluster integrals, the Umbral algebra was exploited to show that the partition function $Q_N(T, V)$ can be written in terms of an invertible shift invariant operator and powers of volume V. Whether this representation for the partition function is possible for volume dependent cluster integrals needs further investigation.

ACKNOWLEDGMENTS

I thank Dr. Julian Gibbs for his encouragement to publish this work. I thank Dr. E. Donoghue for a useful discussion. I am indebted to Professor K. E. Shuler for his hospitality and support at the University of California, San Diego. This work was supported in part by NIH Grant No. ROI GM10906-18.

APPENDIX

In this Appendix we briefly outline a different proof of the Mayer first theorem. The proof will be based on theorems in Umbral algebra developed by Rota and collaborators.[1,2]

Let us indicate that the shift invariant operator P corresponds to the formal power series $p(t)$ under the isomorphism theorem by $P = p(Q)$. The isomorphism theorem[1] states that there exists an isomorphism from the ring of formal power series in the variable t over the same field as the delta operator Q onto the ring of shift invariant operator, which carries

$$f(t) = \sum_{n \geqslant 0} \frac{a_n t^n}{n!} \quad \text{into} \quad \sum_{n \geqslant 0} \frac{a_n}{n!} Q^n. \tag{A1}$$

Proposition 2: Let P be a delta operator with basic polynomials $q_n(x)$, and let $P = p(D)$. Let $p^{-1}(t)$ be the inverse formal power series. Then

$$\sum_{n \geqslant 0} \frac{q_n(x)}{n!} u^n = e^{xp^{-1}(u)}. \tag{A2}$$

Proof: Any shift invariant operator can be expanded in terms of the delta operator[1] (Expansion Theorem). In particular, expand the translation operator E^a [E^a: $E^a q(x) = q(x+a)$] in terms of P as

$$E^a = \sum_{n>0} \frac{q_n(a)}{n!} P^n. \tag{A3}$$

We can now use the isomorphism theorem with D as delta operator to get

$$e^{at} = \sum_{n=0} \frac{q_n(a)}{n!} (p(t))^n. \tag{A4}$$

Now set $a = x$ and $u = p(t)$ in the above equation.

Corollary 2: Given any sequence of constants $\alpha_{n,1}$, $n = 1,2,\ldots$, with $\alpha_{1,k} \neq 0$ there exists a unique sequence of basic polynomials $q_n(x)$ such that

$$\alpha_{n,1} = (x^{-1} q_n(x))_{x=0},$$

i.e., (A5)

$$q_n(x) = \sum_{k>1} \alpha_{n,k} x^k, \quad n = 1, 2, \ldots .$$

Corollary 3: Let $g(x)$ be the formal power series corresponding to P in Corollary 2. Then $g = f^{-1}$, where

$$f(t) = \sum_{k>1} \alpha_{k,1} \frac{t^k}{k!}. \tag{A6}$$

The preceding corollary gives an explicit interpretation to the generating function of a sequence of basic polynomials, which can be restated as

$$\sum_{n>0} \frac{q_n(x)}{n!} t^n = \exp\left(x \sum_{k>1} \alpha_{k,1} \frac{t^k}{k!}\right). \tag{A7}$$

If one identifies $q_n(x)$ with $Q_N(T, V)$, x with V, and $\alpha_{k,1}$ with $k!\, b_k(T)$, then Eq. (A7) becomes Mayer's first theorem.

[1]G. C. Rota, *Finite Operator Calculus* (Academic, New York, 1975).

[2]S. M. Roman and G. C. Rota, Adv. Math. **27**, 95 (1978).

[3]J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1977), 2nd ed.

[4]J. H. Gibbs, B. Bagchi, and U. Mohanty, Phys. Rev. B **24**, 2893 (1981). See also Ref. 10, and E. Donoghue and J. H. Gibbs, J. Chem. Phys. **74**, 2975 (1981).

[5]G. E. Uhlenbeck and G. W. Ford, *Lectures in Statistical Mechanics* (American Mathematical Society, Providence, RI, 1963), Vol. 1.

[6]E. T. Bell, Ann. Math. **35**, 258 (1938).

[7]J. Dénes, Magyar Tudományos Akadémia Matematikai Kutató Intézetenek Közleményei **4**, 63 (1959).

[8]A. Rényi, Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei **4**, 73 (1959).

[9]The partition function as given by Eq. (31) can also be expressed as a contour integral,

$$\frac{Q_N(T, V)}{(N-1)!} = \sum_{M=1}^{N} \frac{V^M}{M!} \beta_1^{N-M} \times \left\{ \frac{1}{2\pi i} \oint_C \frac{[(d/dZ)(Z - \frac{1}{2}Z^2)^M] e^{NZ}}{Z^N} dZ \right\},$$

where C is a closed contour in the Z plane encircling and sufficiently close to $Z = 0$.

[10]U. Mohanty, A. Yang, and J. H. Gibbs (in preparation). In this work a physical cluster theory of vapor condensation is developed.

Continuity of sample paths and weak compactness of measures in Euclidean field theory

Z. Haba
Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Poland

(Received 5 September 1980; accepted for publication 15 May 1981)

Under an assumption on the asymptotic $u \to \infty$ behavior of the generating functional $\int \exp[u\varphi(g)]\, d\mu(\varphi)$, an estimate is obtained on $|\chi(x) - \chi(x')|$, where $\chi(x) = \varphi_g(x) \equiv \varphi(g(\cdot - x))$, showing Hölder continuity of $\varphi_g(x)$ for a class of g. It is proved that the family of measures υ_γ, with $d\upsilon_\gamma(\chi) = d\mu_\gamma(\varphi_g)$ and $\int \exp\chi(c)\, d\upsilon_\gamma(\chi)$ bounded in γ, is weakly conditionally compact. In application to the infinite volume cutoff μ_κ measures in $P(\varphi)_d$, these results imply the continuity of $\mu_\kappa(C)$ when $\kappa \to \infty$ for a class of sets C. Such a property allows distinguishing the interacting measure from the free ones.

PACS numbers: 11.10. – z

I. INTRODUCTION

It is of some importance to know the supports of measures we are dealing with in the Euclidean field theory.[1,2] In Ref. 3 we have established some continuity properties of $\varphi(\delta_t f)$ for φ from the support of the Euclidean measure μ on $\mathscr{S}'(R^d)$ under some assumptions, which are fulfilled in $P(\varphi)_2$ and $(\varphi^4)_3$. In this paper we prove more general results concerning the continuity properties of $\varphi(g(\cdot - x))$ for a certain class of g under much weaker assumptions. In fact, only the asymptotic behavior of $E[\exp u\varphi(g)]$ for $|u| \to \infty$ is needed as a technical tool. The continuity properties of sample paths depend then on the properties of the covariance.

Using our results on the continuity of sample paths and the Prokhorov criterion[4] (concerning measures of certain sets of continuous functions) for weak compactness of a family $\{\upsilon_\gamma\}$ of measures, we can give a criterion for weak compactness involving only the properties of $E[\exp u\varphi(g)]$. The weak compactness allows one to determine the measure $\upsilon_\infty(C)$ of a certain class of sets C if the measures $\upsilon_\gamma(C)$ are known and if $\upsilon_\gamma \to \upsilon_\infty$ in the sense of convergence of the characteristic functions. Finally, an application to the infinite volume regularized $P(\varphi)_d$ theory is discussed. We consider the infinite volume measures μ_κ with an ultraviolet cutoff κ on the interaction. The measure μ_κ on $\mathscr{S}'(R^d)$ induces a measure υ_κ with the support on $C(R^d)$ by means of the transformation $\varphi(x) \to \varphi(g(\cdot - x))$. It is shown that if $E_\kappa[\exp\varphi(g)]$ are bounded in κ, then the measures υ_κ on $C(R^d)$ are weakly compact with the usual topology on $C(R^d)$ of almost uniform convergence. In an earlier paper[5] we have extended some results of Rosen and Simon[6] on the behavior of the sample paths (fluctuations) of $\varphi(g(\cdot - x))$ for $x \to \infty$. We have found some sets in $C(R^d)$ describing the fluctuations of φ, which are closed in the topology of the almost uniform convergence. Our results on weak compactness imply now, that if the characteristic functions of μ_κ converge, then the measure of a closed set C is 1 if $\upsilon_\kappa(C) = 1$. This result means that some properties of the sample paths distinguishing Gaussian and non-Gaussian random fields are preserved under the limit $\kappa \to \infty$.

The weak convergence of measures on nuclear and more general linear topological spaces is discussed in Refs. 7 and 8. It is proved there that μ_κ converge weakly, if the characteristic functions converge. The topology in the space of measures is then related (Ref. 8 in the Appendix) to the topology of the set of the characteristic functions. However, these general results cannot be directly applied to the sample properties we are interested in. In order to get some conclusions about the support of the limit measure μ we need to consider either closed or open subsets in the topology of $\mathscr{S}'$. We were unable to describe fluctuations of φ in terms of such sets.

II. CONTINUITY OF SAMPLE PATHS OF EUCLIDEAN RANDOM FIELDS

We begin with a generalized random field φ over the test-function space $\mathscr{S}(R^d)$.[9] In applications, it is often necessary to enlarge the test-function space. The generalized random field φ can be defined for a larger test-function space $\mathscr{F}$ if $\mathscr{S} \ni f_n \to f$ in $\mathscr{F}$ implies that $\varphi(f_n)$ is convergent in measure (see Ref. 10). We shall restrict ourselves to a subspace L^2_σ of $\mathscr{F}$ consisting of functions f such that

$$\int d\sigma(p) |f(p)|^2 < \infty , \tag{II.1}$$

where σ is the spectral function of the covariance,

$$E[\varphi(x)\varphi(y)] = \int d\sigma(p)\, e^{ip(x-y)} .$$

We assume that the random field φ is *translation invariant*. If f fulfills the condition (II.1) and $\mathscr{S} \ni f_n \to f$ in L^2_σ then $\varphi(f)$ can be defined as a limit of $\varphi(f_n)$ in the mean. As the random field $\varphi(x)$ is translation invariant we may define the (random) spectral measure ω such that

$$\varphi_g(x) \equiv \varphi(g(\cdot - x)) = \tilde{\varphi}(\tilde{g}(p)\, e^{ipx}) = \int e^{ipx}\, d\omega_g(p) ; \tag{II.2}$$

here $\tilde{\varphi}$ means the Fourier transform. We shall also introduce some auxiliary random fields $\psi_j(x)$,

$$\psi_j(x) = \tilde{\varphi}(\tilde{g}(p)\, |p_j|^\alpha\, e^{ipx}) = \int e^{ipx} |p_j|^\alpha\, d\omega_g(p) . \tag{II.3}$$

In Osterwalder–Schrader field theory[11] $d\sigma(p)$ should have the form

$$d\sigma(p) = d^d p \int_{m_0}^{\infty} d\rho(m) \frac{1}{p^2 + m^2}$$

[if $\rho(m)$ grows as m^2 or faster some subtractions are needed in this formula].

In our estimates on the sample paths we shall assume

(i) for a certain $0 < \alpha < \frac{1}{2}$ there exists $\rho > 1$ such that $\lim \sup_{|u| \to \infty} |u|^{-\rho} \ln |E[\exp u\psi_j(0)]| < M(g)$.

If $g \notin \mathscr{S}(R^d)$ we assume, in addition to (i),

(ii) either $d\sigma(p) = d^d p\, \rho(p)$ and $|\tilde{g}(p)|^2 \rho(p) |p_i|^\alpha \in L_q \cap L_1$ for certain $1 < q \leqslant 2$, $i = 1,2,...d$ and $\alpha < \frac{1}{2} - 1/2q$, or $\tilde{g}(p) = f(\mathbf{p}_{d-1})$ [where $\mathbf{p}_{d-1} = (p_2,...,p_d)$] with $f \in \mathscr{S}(R^{d-1})$, $|\tilde{f}(\mathbf{p}_{d-1})|^2 \rho(p) |p_1|^\alpha \in L_1$, and $m_0 > 0$ in Eq. (II.3).

This section is mainly devoted to the proof of

Theorem II.1. Assuming (i), and if $g \notin \mathscr{S}(R^d)$ also (ii), we have the following estimate ($V_{x,x'}$ means "there exist x and x' such that") with certain numerical constants K_1, K_2:

$$P\{ V_{x,x'} | \varphi_g(x) - \varphi_g(x') | \geqslant R\, [M(g)]^{1/\rho} | x - x' |^{\alpha'} \} \leqslant R^{-1} \left\{ K_1 + K_2 \sum_j E\left[\left(\ln \frac{\tilde{B}^j_{2T+1}(0)}{T+1} \right)^{(\rho-1)/\rho} \right] \right\} \quad \text{(II.4)}$$

for $|x|$, $|x'| \leqslant T$ and any $\alpha' < \alpha$, where

$$\tilde{B}^j_T(x) = \int_{-T}^{T} ds \exp |r\psi_j(x_j + s, \mathbf{x}_{d-1})|^{\rho/(\rho-1)} \quad \text{(II.5)}$$

with certain $r \leqslant (a(\rho))^{-1} (M(g))^{-1/\rho}$, where $a(\rho)$ is a numerical constant. In particular there exists a random variable K_T finite almost surely (a.s.) such that

$$| \varphi_g(x) - \varphi_g(x') | \leqslant K_T | x - x' |^{\alpha'} \quad \text{(II.6)}$$

with probability 1.

Remark: (1) Theorem 1 of my previous paper (Ref. 3) is a special case of the present Theorem II.1 corresponding to $\tilde{g}(p) = \tilde{f}(\mathbf{p}_{d-1})$ depending only on $d-1$ momenta $\mathbf{p}_{d-1}$. We have proved this special case before under an additional assumption on higher order moments (assumption d of Ref. 3). This assumption is not needed at all. We have used it only in the proof of Lemma 5 of Ref. 3. We will prove this lemma here (Lemma II.6) assuming only (i) and (ii).

(2) Let us note that the estimate (II.6) is trivial if $g \in \mathscr{S}(R^d)$ but (II.4) is not even in this case.

In terms of $\psi_j(x)$ [Eq. (II.2)] we can now define the stochastic process depending on $\mathbf{x}_{d-1} = (x_2,...,x_d)$ as a parameter,

$$\mathscr{C}_t(\mathbf{x}_{d-1}) = \int_0^t \psi_1(x_1, \mathbf{x}_{d-1})\, dx_1 \quad \text{(II.7)}$$

and

$$\xi^n_t(\mathbf{x}_{d-1}) = \frac{\Gamma(1-\alpha)}{\pi} \sin\frac{\alpha\pi}{2} \int_{-n^\rho}^{n^\rho} |s|^{\alpha-1} e^{-n^{\gamma\rho}|s|^\delta} \times \psi_1(t-s, \mathbf{x}_{d-1}) ds \quad \text{(II.8)}$$

with $0 < \gamma < \delta$ and $\gamma p > 1/2$. These are the analogs of $\mathscr{C}_t$ and ξ^n_t of Ref. 3. The only difference is that we replaced n of Ref. 3 by n^ρ, which improves the convergence of $\xi^n_t(\mathbf{x}_{d-1})$ to $\varphi_g(t, \mathbf{x}_{d-1})$. We need several lemmas before we start the proof of Theorem II.1.

Lemma II.2. For any $\rho > 1$ there exists a constant $b(a)$ such that

$$\exp |x|^{\rho/(\rho-1)} \leqslant b \int dy \exp(-|y|^\rho) \exp a|x|y, \quad \text{(II.9)}$$

where $a \leqslant a(\rho) = \rho(\rho-1)^{-1+1/\rho}$.

Proof: We write

$$-|y|^\rho + \rho|z|y = -r|y-v|^\rho + v^\rho + v^\rho[r|1 - y/v|^\rho - |y/v|^\rho - 1 + \rho y/v], \quad \text{(II.10)}$$

with $v = |z|^{1/(\rho-1)}$. For $|y/v| \geqslant \epsilon > 0$ we can choose (if $\rho > 1$) r such that the expression in square brackets in (II.10) will be larger than $\rho - 2$. Then, if $|y/v| < \epsilon$ we may expand the function in the square brackets and show that it is not less than $\rho - 2$. By exponentiation of such an inequality we get (II.9).

Remark: For $\rho \leqslant 2$ we could get a similar bound from below in Eq. (II.9) with different a and b.

Lemma II.3. The integral (II.5) exists a.s. and

$$E[\tilde{B}^j_T(x)] = E[\tilde{B}^j_T(0)] < \infty$$

if $r^{-1} \geqslant (M(g))^{1/\rho} a(\rho)$.

Proof: From Lemma II.2, assumption (i), and translation invariance

$$\begin{aligned}
E&[\exp|r\varphi_g(x_j + s_j, \mathbf{x}_{d-1})|^{\rho/(\rho-1)}] \\
&= E[\exp|r\varphi_g(0)|^{\rho/(\rho-1)}] \\
&\leqslant b \int dy \exp(-|y|^\rho) E[\exp ar|\varphi_g(0)|y] \\
&\leqslant b \int dy \exp(-|y|^\rho) \{ E[\exp(ar\varphi_g(0)y)] + E[\exp(-ar\varphi_g(0)y)] \} \\
&\leqslant C + \int_{|y| > N} dy \exp(-|y|^\rho) \exp M(g) |ary|^\rho (1-\epsilon) < \infty .
\end{aligned}$$

This is also sufficient for the integral (II.5) to exist (Doob's theorem[12], Sec. II).

Lemma II.4.

$$|\mathscr{C}_{x_1+h}(\mathbf{x}_{d-1}) - \mathscr{C}_{x_1}(\mathbf{x}_{d-1})| \leqslant (M(g))^{1/\rho} a(\rho) |h| \left(\ln \frac{\tilde{B}^1_T(x)}{|h|} \right)^{(\rho-1)/\rho}$$

for $|h| < T$ with probability 1.

Proof: For $t,h > 0$ from the Jensen inequality,

$$F\left(\frac{1}{h} \int_t^{t+h} \psi_1(s) ds \right) \leqslant \frac{1}{h} \int_t^{t+h} F(\psi_1(s)) ds \leqslant \frac{1}{h} \int_t^{t+T} F(\psi_1(s)) ds$$

for any convex function F. For $F(x) = \exp|x|^{\rho/(\rho-1)}$, $\mathscr{C}_t \times$ (II.7), and $\tilde{B}^1_T(x)$ (II.5) with $r^{-1}(M(g))^{1/\rho} a(\rho)$, we get the inequality in the Lemma (the case t,h nonpositive can be treated similarly).

Lemma II.5. The inequality

$$|\xi^n_{x_1+h}(\mathbf{x}_{d-1}) - \xi^n_{x_1}(\mathbf{x}_{d-1})| \leqslant (M(g))^{1/\rho} \mathscr{A}_n(x,h) |h|^\alpha \quad \text{(II.11)}$$

holds true with probability 1 for each $\mathbf{x}_{d-1}$ and $|x_1|,\ |h|<T$,

where

$$E\,[\mathscr{A}_n(x,h)]\leqslant\left(\ln\frac{2T}{|h|}\right)^{\rho-1/\rho}$$
$$\times(\mathscr{A}_1+\mathscr{A}_2E\,[(\ln(\widetilde{B}^{\,1}_{2T+1}(0)T^{-1}))^{(\rho-1)/\rho}]),\quad\text{(II.12)}$$

$\mathscr{A}_1$, $\mathscr{A}_2$ being some numerical constants. For each $\mathbf{x}_{d-1}$ there exists an a.s. finite radnom variable $B_T(\mathbf{x}_{d-1})$ such that

$$\mathscr{A}_n(x,h)\leqslant\mathscr{A}\left(\ln\frac{B_T(\mathbf{x}_{d-1})}{2T|h|}\right)^{\rho-1/\rho},\qquad\text{(II.13)}$$

with $\mathscr{A}$ independent of n.

Proof: By means of the integration by parts we can express $\xi^{\,n}_t$ in terms of an integral over $\mathscr{C}_{t-s}$. Next, we use Lemma II.4 in order to estimate $|\mathscr{C}_{t+h-s}-\mathscr{C}_{t-s}|$ for $|s|\leqslant n^p$. Still, the integrals over s have to be estimated. This has been done in Ref. 3 (Lemma 4). It can be seen from the calculations performed there that we may take as $\mathscr{A}_n(x,h)$ in Eq. (II.11) the following expression:

$$\begin{aligned}\mathscr{A}_n(x,h)&=\mathscr{A}_1(\ln\widetilde{B}^{\,1}_{2T+1}(0,\mathbf{x}_{d-1})|h|^{-1})^{(\rho-1)/\rho}\\&+\mathscr{A}_2(\ln2T|h|^{-1})^{(\rho-1)/\rho}\\&+\mathscr{A}_3\int_1^{n^p}ds\,s^{\alpha-2+\delta}e^{-n^{-\gamma p}s^\delta}\\&\times\{(\ln\widetilde{B}^{\,1}_T(x_1-s,\mathbf{x}_{d-1})|h|^{-1})^{(\rho-1)/\rho}\\&+(\ln\widetilde{B}^{\,1}_T(x_1+s,\mathbf{x}_{d-1})|h|^{-1})^{(\rho-1)/\rho}\}\\&+\mathscr{A}_4\left\{\left[\ln\frac{\widetilde{B}^{\,1}_T(x_1-n^p,\mathbf{x}_{d-1})}{(|x_1|+n^p)|h|}\right]^{(\rho-1)/\rho}\right.\\&\left.+\left[\ln\frac{\widetilde{B}^{\,1}_T(x_1+n^p,\mathbf{x}_{d-1})}{(|x_1|+n^p)|h|}\right]^{(\rho-1)/\rho}\right\}.\end{aligned}\qquad\text{(II.14)}$$

In deriving (II.14) we have decomposed the integration over s into the parts $[-1,1]$, $[1,n^p]$ $[-n^p,-1]$. The first term in Eq. (II.14) comes from the $[-1,+1]$ part. The last term from the integral $\int_1^{n^p}\psi_1(x_1-s)ds$. The estimate (II.12) is a consequence of Eq. (II.14) and the inequalities

$$\widetilde{B}^{\,1}_T\leqslant\widetilde{B}^{\,1}_{2T+1},\quad(a+b)^{(\rho-1)/\rho}\leqslant K\,(a^{(\rho-1)/\rho}+b^{(\rho-1)/\rho})$$

for certain K. Next, the existence of $B_T(\mathbf{x}_{d-1})$ such that (II.13) holds true is a consequence of the ergodic theorem. To see this let us consider

$$\begin{aligned}&\widetilde{B}^{\,1}_T(t)\\&=\int_{t-T}^{t+T}ds\,\exp|r\psi_1(s)|^{\rho/(\rho-1)}\\&\leqslant\int_0^{t+T}ds\,\exp|r\psi_1(s)|^{\rho/(\rho-1)}+\int_0^{t-T}ds\,\exp|r\psi_1(s)|^{\rho/(\rho-1)}\\&\leqslant\int_0^{n+T}ds\,\exp|r\psi_1(s)|^{\rho/(\rho-1)}+\int_0^{n-T}ds\,\exp|r\psi_1(s)|^{\rho/(\rho-1)}\end{aligned}$$

for $t>T$ and $n=[t]+1$ ([ ] means the integer part). From the ergodic theorem,

$$\frac{1}{n+T}\int_0^{n+T}ds\,\exp|r\psi_1(s)|^{\rho/(\rho-1)}$$

is a.s. convergent as $n\to\infty$, hence a.s. bounded. The case $t<T$ can be treated similarly. So, $\widetilde{B}^{\,1}_T(t,\mathbf{x}_{d-1})(1+|t|)^{-1}$ is bounded by certain a.s. finite random variable $B^{\,1}_T(\mathbf{x}_{d-1})$. Using this we can estimate the integral over s in Eq. (II.14) by

$$C\,[(\ln B^{\,1}_T(\mathbf{x}_{d-1})|h|^{-1})^{(\rho-1)/\rho}+(\ln2T|h|^{-1})^{(\rho-1)/\rho}]$$
$$\leqslant C_2(\ln C_1B^{\,1}_T(\mathbf{x}_{d-1})|h|^{-1})^{(\rho-1)/\rho},$$

arriving finally at Eq. (II.13) with certain $\widetilde{B}_T$,

Lemma II.6. Under the assumption (i) if $g\in\mathscr{S}(R^d)$ and also (ii) if $g\notin\mathscr{S}(R^d)$, $\xi^{\,n}_{x_1}(\mathbf{x}_{d-1})$ converges a.s. to $\bar{\varphi}(x_1,\mathbf{x}_{d-1})$ for each x, where $\bar{\varphi}(x)$ is a random field equivalent to $\varphi(x)$.

Proof: $\mathbf{x}_{d-1}$ plays no role in the proof so we shall conceal it. It is known that the condition

$$\sum_{n=1}^{\infty}a_n^{-2}E\,[(\xi^{\,n+1}_t-\xi^{\,n}_t)^2]<\infty\qquad\text{(II.15)}$$

for a sequence $\{a_n\}$, such that $\Sigma_{n=1}^\infty|a_n|<\infty$, is sufficient for a.s. convergence. In our case we have from the definition (II.8)

$$\begin{aligned}\xi^{\,n+1}_t-\xi^{\,n}_t&=K\int_{-n^p}^{n^p}|s|^{\alpha-1}(\exp[-(n+1)^{-p\gamma}|s|^\delta]\\&-\exp[-n^{-p\gamma}|s|^\delta])\psi_1(t-s)ds\\&+K\int_{n^p}^{(n+1)^p}s^{\alpha-1}\exp[-(n+1)^{-\gamma p}s^\delta]\psi_1(t-s)ds\\&+K\int_{n^p}^{(n+1)^p}s^{\alpha-1}\exp[-(n+1)^{-\gamma p}s^\delta]\psi_1(t+s)ds.\end{aligned}\qquad\text{(II.16)}$$

Using $|E\,[\psi_1(t-s)\psi_1(t-s')]|\leqslant E\,[\psi_1^2(0)]$ we get

$$\begin{aligned}&E\left[\left(\int_{n^p}^{(n+1)^p}s^{\alpha-1}\exp[-(n+1)^{-\gamma p}s^\delta]\psi_1(t-s)ds\right)^2\right]\\&\leqslant E\,[\psi_1^2(0)]\left(\int_{n^p}^{(n+1)^p}s^{\alpha-1}\exp[-(n+1)^{-\gamma p}s^\delta]ds\right)^2\\&<Cn^{2\alpha p-2}\exp[-2(n+1)^{-\gamma p}n^{\delta p}].\end{aligned}\qquad\text{(II.17)}$$

Next,

$$\exp[-(n+1)^{-\gamma p}|s|^\delta]-\exp[-n^{-\gamma p}|s|^\delta]$$
$$\leqslant|s|^\delta(n^{-\gamma p}-(n+1)^{-\gamma p})\leqslant\gamma p|s|^\delta n^{-1-\gamma p}.$$

Using this inequality we get

$$\begin{aligned}&E\left[\left(\int_{-n^p}^{-n^p}|s|^{\alpha-1}(e^{-(n+1)^{-\gamma p}|s|^\delta}-e^{-n^{-\gamma p}|s|^\delta})\psi_1(t-s)ds\right)^2\right]\\&\leqslant\gamma^2p^2n^{-2-2\gamma p}\int_{-n^p}^{n^p}ds\int_{-n^p}^{n^p}ds'|s|^{\alpha+\delta-1}|s'|^{\alpha+\delta-1}\\&\quad|E\,[\psi_1(t-s)\psi_1(t-s')]|\\&\leqslant\gamma^2p^2n^{-2-2\gamma p}\int_{-\infty}^{\infty}ds\int_{-\infty}^{\infty}ds'|s|^{\alpha+\delta-1}|s'|^{\alpha+\delta-1}\\&\quad|E\,[\psi_1(s)\psi_1(s')]|\\&\leqslant C_1n^{-2-2\gamma p}\int_{-\infty}^{\infty}dx_1|x_1|^{2(\alpha+\delta)-1}|E\,[\psi_1(0)\psi_1(x_1)]|.\end{aligned}\qquad\text{(II.18)}$$

In the last step we have changed variables, setting $x_1=s-s'$, and computed the integral

$$\int_{-\infty}^{\infty}ds'|s'+x_1|^{\alpha+\delta-1}|s'|^{\alpha+\delta-1}=C\,|x_1|^{2(\alpha+\delta)-1}$$

If $g\in\mathscr{S}(R^d)$, then $E\,[\psi_1(0)\psi_1(x_1)]$

is the Fourier transform of a function from $\mathscr{S}$; therefore, it decreases rapidly. Hence the last integral in Eq. (II.18) is finite for $\alpha+\delta<\frac{1}{2}$. If $g\notin\mathscr{S}(R^d)$ we have to apply assumption (ii). Let us denote

$$v(x)=E\,[\psi_1(0)\psi_1(x)]=\int d\,p^d\rho(p)|\,p_1|^{2\alpha}|\tilde{g}(p)|^2\,e^{ipx}.\qquad\text{(II.19)}$$

If $\rho(p)|p_1|^{2\alpha}|\tilde{g}(p)|^2 \in L_q$, then $v(x)$ is continuous and $v \in L_r$ with $r = q/(q-1)$ from the Hausdorff–Young inequality. Applying the Hölder inequality to

$$\int dx_1 |x_1|^{2(\alpha+\delta)-1}\left(\frac{|x_1|}{1+|x_1|}\right)^{\lambda}\left(\frac{1+|x_1|}{|x_1|}\right)^{\lambda}|v(x_1,0)|,$$

with appropriate $\lambda < 1/r$, we get that the integral (II.18) is finite if $r(1-2(\alpha+\delta)) > 1$. Next, if

$$\tilde{g}(p) = \tilde{f}(\mathbf{p}_{d-1}) \in \mathscr{S}(R^{d-1})$$

and

$$\rho(p) = \int_{m_0}^{\infty} d\rho(m)\, \frac{1}{p^2+m^2},$$

we can compute the integral

$$\int dp_1\, \frac{e^{ip_1x_1}|p_1|^{2\alpha}}{p_1^2 + (m^2 + \mathbf{p}_{d-1}^2)}$$

and show that it behaves as $(m^2+\mathbf{p}_{d-1}^2)^{-1}|x_1|^{-1-2\alpha}$ for $|x_1| \to \infty$. Therefore, the last integral in Eq. (II.18) is finite. With the estimates (II. 17) and (II.18) we get

$$E\,[(\xi_t^{n+1} - \xi_t^n)^2] \leqslant K_1 n^{-2-2\gamma p} + K_2 n^{2\alpha p - 2} \times \exp[-2(n+1)^{-\gamma p} n^{\delta p}]. \qquad \text{(II.20)}$$

Choosing $p > 1/2\gamma$ and $a_n = n^{-1-\epsilon}$ with ϵ small enough we get the convergence of the series (II.15). That $\xi_t^n(\mathbf{x}_{d-1})$ converges to a process equivalent to $\varphi_g(t,\mathbf{x}_{d-1})$ follows from the convergence of $\xi_t^n(\mathbf{x}_{d-1})$ to $\varphi_g(t,\mathbf{x}_{d-1})$ in the mean, which is a consequence of the point-wise (i.e., for each φ convergence of ξ_t^n (II.8) to $\tilde{\varphi}(\tilde{g}(p)e^{ip_1t})$ and the Fatou lemma.

As the consequence of the last two lemmas we get

Theorem II.7. Assume (i) and (ii) with $\tilde{g}(p) = \tilde{f}(\mathbf{p}_{d-1}) \in \mathscr{S}(R^{d-1})$. Then, with $\varphi(\delta_t f) = \tilde{\varphi}(e^{ip_1t}\tilde{f}(\mathbf{p}_{d-1}))$, the following inequality holds true with probability 1:

$$|\varphi(\delta_t f) - \varphi(\delta_{t'} f)| \leqslant \mathscr{A}|t-t'|^{\alpha}\left|\ln\frac{B_T}{|t-t'|}\right|^{(p-1)/p}$$

for $|t|, |t'| < T$, a certain constant $\mathscr{A}$, and an a.s. finite random variable B_T (depending on f).

Proof: This follows from Lemma II.5 [Eqs. (II.11) and (II.13) with $\mathbf{x}_{d-1} = 0$] and the a.s. convergence of $\xi_t^n(0)$ (Lemma II.6).

Remark: Theorem II.7 has been proved in Ref. 3 under an additional assumption (assumption d) on higher order Schwinger functions. Note, however, that it was *incorrectly* claimed in Ref. 3 that B_T can always be chosen to be integrable.

Lemma II.8.

$$P\left\{\bigvee_{x,x'}|\varphi_g(x_1,\mathbf{x}_{d-1}) - \varphi_g(x_1',\mathbf{x}_{d-1})| \geqslant R\,(M(g))^{1/p}|x_1 - x_1'|^{\alpha}\left(\ln\frac{2T}{|x_1-x_1'|}\right)^{(p-1)/p}\right\}$$
$$\leqslant R^{-1}\{K_1 + K_2 E\,[(\ln T^{-1}\tilde{B}^{\,1}_{2T+1}(0))^{(p-1)/p}]\}$$

for $|x_1|,|x_1'| < T/2$, with certain numerical constants K_1, K_2.

Proof: The inequality (II.11) holds true with probability 1 for each n, $\mathbf{x}_{d-1}$ and $|x_1|,|h| < T$. Hence the probability that for certain n, x, x' an inverse inequality holds,

$$P\left\{\bigvee_{n,x,x'}|\xi^n_{x_1'}(\mathbf{x}_{d-1}) - \xi^n_{x_1}(\mathbf{x}_{d-1})| > R\,(M(g))^{1/p}|x_1-x_1'|^{\alpha}\left(\ln\frac{2T}{|x-x_1'|}\right)^{(p-1)/p}\right\}$$
$$\leqslant P\left\{\mathscr{A}_n(x,x')\left(\ln\frac{2T}{|x_1-x_1'|}\right)^{(1-p)/p} > R\right\}$$
$$\leqslant R^{-1}E\left[\mathscr{A}_n(x,x')\left(\ln\frac{2T}{|x_1-x_1'|}\right)^{(1-p)/p}\right]$$
$$\leqslant R^{-1}\{\mathscr{A}_1 + \mathscr{A}_2 E\,[(\ln\tilde{B}^{\,1}_{2T+1}(0)T^{-1})^{(p-1)/p}]\}, \qquad \text{(II.21)}$$

where in the last step the estimate (II.12) was applied.

Now, as $\xi^n_{x_1}(\mathbf{x}_{d-1}) \to \varphi(x_1,\mathbf{x}_{d-1})$ a.s. (Lemma II.6) Eq. (II.21) implies the inequality formulated in the lemma. We come finally to

Proof of Theorem II.1: We could define $\mathscr{C}_t$ and ξ_t^n for each j as integrals over $\psi_j(x_j,\mathbf{x}_{d-1})$ and prove Lemmas (II.4)–(II.8) for them [there is an asymmetry in j when $\tilde{g}(p) = \tilde{f}(\mathbf{p}_{d-1})$; nevertheless, the estimate of Lemma II.8, which we need, also remains true for each j in this case]. Next, we repeatedly use the inequality

$$|\varphi_g(x_1,\mathbf{x}_{d-1}) - \varphi_g(x_1',\mathbf{x}_{d-1}')| \leqslant |\varphi_g(x_1,\mathbf{x}_{d-1}) - \varphi_g(x_1',\mathbf{x}_{d-1})| + |\varphi_g(x_1',\mathbf{x}_{d-1}) - \varphi_g(x_1',\mathbf{x}_{d-1}')| \qquad \text{(II.22)}$$

till we get on the right side of Eq. (II.22), a sum of terms $\Sigma|\varphi_g(y) - \varphi_g(y')|$ with the vectors y, y' composed from components of x and x', with only one component of y' different from y. For each $|\varphi_g(y) - \varphi_g(y')|$ we have the estimate of Lemma II.8. Using the inequalities

$$\sum_j |x_j - x_j'|^{\alpha}\left(\ln\frac{2T}{|x_j-x_j'|}\right)^{(p-1)/p} \leqslant K\,|x-x'|^{\alpha'}$$

and $P(\cup\mathscr{A}_j) \leqslant \Sigma_j P(\mathscr{A}_j)$, we get Theorem II.1 [Eq. (II.6) follows from Eq. (II.4) if we let $R \to \infty$].

III. WEAK COMPACTNESS OF MEASURES WITH SUPPORTS ON CONTINUOUS FUNCTIONS

Let us specify our probability space of Sec. II as $(\mathscr{S}', \mathscr{Z}, \mu)$, where $\mathscr{S}'$ is the dual to the nuclear Schwartz space

$\mathscr{S}(R^d)$, $\mathscr{Z}$ is the σ-algebra generated by the cylinder sets,[13] and μ is a cylindrical probability measure on $\mathscr{Z}$. Now, for each $g\in\mathscr{S}(R^d)$ we can define a map $g':\mathscr{S}'(R^d)\to C(R^d)$ into the space of continuous functions on R^d by means of the formula

$$g'(\varphi)(x) = \varphi(T_x g) = \varphi(g(\cdot - x)) = \varphi_g(x)$$

(T_x means the translation). The image of the cylinder set

$$Z_A(x_1,\ldots,x_n) = \{\varphi\in\mathscr{S}':(\varphi_g(x_1),\ldots,\varphi_g(x_n))\in \mathscr{A}\subset R^n\}\subset\mathscr{Z}$$

under g' is what is usually called the finite dimensional (cylinder) set in $C(R^d)$ and denoted $\Pi^{-1}_{x_1\ldots x_n}\mathscr{A}$. Denote by $\mathscr{Z}_c$ the minimal σ-algebra generated by such finite dimensional sets. Then g' defines a map of the probability space $(\mathscr{S}',\mathscr{Z},\mu)$ into $(C, \mathscr{Z}_c, v)$, where v is the measure $\mu g'^{-1}$ defined by the formula

$$v(S) = \mu(g'^{-1}S) \quad \text{for } S = g'B,\ B\subset\mathscr{Z}.$$

The image of the topological space $\mathscr{S}'$ (with the usual weak topology) under the transformation g' is a topological space of continuous (moreover C^∞) functions on R^d. We may always consider v as defined on a larger topological space Φ with support on a certain measurable subset of Φ. We take Φ to be the space of continuous functions on R^d with the topology (weaker on $\Phi\cap g'\mathscr{S}'$ than the image of the weak topology of $\mathscr{S}'$) defined by the family of open sets

$$\mathscr{U}_{K,a} = \left\{\chi\in\Phi:\sup_{x\in K}|\chi(x)| < a\right\},$$

where K are compact subsets of R^d. It can be shown that with this definition open (closed) sets in Φ are v measureable. To see this it is sufficient to note that the set $\{\sup_{x\in K}|\chi(x)|\leqslant a\}$ is an intersection of the finite dimensional (cylinder) sets

$$\{|\chi(x_i)|\leqslant a,\quad x_i\in K,\ i = 1,\ldots,n\}.$$

So, v can be extended to all Borel sets. Let us note finally that the transformation g' is a continuous transformation of $\mathscr{S}'$ into Φ.

Let $\mathscr{C}(\Phi)$ be the space of continuous bounded functions on Φ. We define the weak topology in the space $\mathscr{M}(\Phi)$ of measures on Φ (see Refs. 4 and 14) saying that $v_n\to v$ if

$$\int_\Phi F(\chi)dv_n(\chi)\to\int F(\chi)dv(\chi) \tag{III.1}$$

for each $F\in\mathscr{C}(\Phi)$. We say that a set Σ of measures $v\in\mathscr{M}(\Phi)$ is weakly conditionally compact if each sequence of measures $\{v_n\}\subset\Sigma$ has a weakly convergent subsequence. It can be seen that Φ is a limit of an increasing sequence of closed sets. For such topological spaces Φ the following criterion of weak conditional compactness is known[14] (Φ is metrizable so the results of Prokhorov[4] would suffice).

Theorem III.1 (Prokhorov, Varadarajan). Let $\Sigma\subset\mathscr{M}(\Phi)$ be a set of probability measures on Φ. If for each ϵ there exists a compact $\mathscr{K}\subset\Phi$ such that for all $v\in\Sigma$,

$$v(\Phi - \mathscr{K}) < \epsilon, \tag{III.2}$$

then Σ is weakly conditionally compact.

By finite dimensional distributions we shall mean

$$\mathscr{P}(u_1x_1,\ldots,u_nx_n) = \int_\Phi \prod_{j=1}^n \exp[iu_j\chi(x_j)]dv(\chi). \tag{III.3}$$

They completely determine the cylinder measure v.

Theorem III.2. Assume we have a sequence of probability measures $\{v_\gamma\}\in\mathscr{M}(\Phi)$, such that their finite dimensional distributions converge and the set $\{v_\gamma\}$ is weakly conditionally compact. Then the sequence $\{v_\gamma\}$ is weakly convergent to certain probability measure v on Φ.

Proof: From the convergence of the finite dimensional distributions it follows that all the convergent subsequences of the weakly compact set $\{v_\gamma\}$ have the same limit, which determines the measure v by its finite dimensional distributions.

The next theorem, proved in Refs. 4 and 14, shows that for certain sets we can get $v(\mathscr{A})$ as a limit of $v_\gamma(\mathscr{A})$.

Theorem III.3. The weak convergence $v_\gamma\to v$ is equivalent to each of the following statements:

(i) $\limsup v_\gamma(C)\leqslant v(C)$ for all closed sets $C\subset\Phi$,

(ii) $\liminf v_\gamma(\mathscr{U})\geqslant v(\mathscr{U})$ for all open sets $\mathscr{U}\subset\Phi$,

(iii) $\lim v_\gamma(\mathscr{A}) = v(\mathscr{A})$ for all $\mathscr{A}\subset\Phi$ such that

there exist $\mathscr{U}, C, \mathscr{U}\subset\mathscr{A}\subset C$ and $v(C-\mathscr{U}) = 0$ (in a metric space this means that the boundary of $\mathscr{A}$ has v measure zero).

Theorem III.1 gives a criterion for the weak conditional compactness in terms of compact sets in Φ.

The following generalization of the Arzela–Ascoli theorem (see Refs. 15 and 16) allows one to determine compact sets in Φ.

Theorem III.4. If a set $\mathscr{F}\subset\Phi$ consists of functions which at each point $x\in R^d$ are equally bounded and equally continuous, then the closure of $\mathscr{F}$ is compact.

So, to establish compactness of $\mathscr{F}$ we have to check for every $x_0\in R^d$ that the set $|\chi(x_0)|$ is bounded when $\chi\in\mathscr{F}$ and that for every $x_0\in R^d$ and $\epsilon > 0$ there exists δ such that $|x_0 - x| < \delta\Rightarrow|\chi(x) - \chi(x_0)| < \epsilon$ for all $\chi\in\mathscr{F}$.

Now consider a family μ_γ of measures on $\mathscr{S}'(R^d)$, which is uniformly bounded in the sense that (iii)

$$\int_{\mathscr{S}'} d\mu_\gamma(\varphi)\exp\varphi(g) = E_\gamma[\exp\varphi(g)]\leqslant I(g)$$

for $g\in\mathscr{S}(R^d)$ and certain $I(g)$ independent of γ.

We are going to show that the set $v_\gamma = \mu_\gamma g'^{-1}$ of (iii) fulfills the Prokhorov criterion of Theorem III.1.

Let us prove first

Lemma III.5. Under assumption (iii) the set $\{|\chi(x)| > L\}\subset\Phi$ has for each x a v_γ measure less than $e^{-L}[I(g) + I(-g)]$.

Proof:

$$\int\exp\varphi(g(\cdot - x))d\mu_\gamma(\varphi) = \int\exp\chi(x)dv_\gamma(\chi) \geqslant \exp L\, v_\gamma\{\chi(x) > L\}$$

Similarly

$$\int \exp[-\varphi(g(\cdot - x))]\, d\mu_\gamma(\varphi)$$
$$\geqslant \exp L\, v_\gamma\{-\chi(x) > L\}.$$

The estimate (II.4) of Theorem II.1, Lemma III.5, Arzela–Ascoli theorem, and Prokhorov criterion allow one to prove.

Theorem III.6. Assume we have a family $\{\mu_\gamma\}$ of translation invariant measures on $\mathscr{S}'(R^d)$ fulfilling (iii) such that the constant $M(g)$ in assumption (i) of Sec. II (with $E_\gamma[\quad] = \int d\mu_\gamma[\quad]$) does not depend on γ, then the set of measures $v_\gamma = \mu_\gamma g'^{-1}$ on Φ is weakly conditionally compact in $\mathscr{M}(\Phi)$.

Proof: We have proved in Lemma III.5 that $v_\gamma\{|\chi(x)| > L\} < Ce^{-L}$. So, to show that the condition (III.2) is fulfilled for a set of equally bounded and continuous functions it is sufficient to show that the right hand side of the estimate (II.4) of Theorem II.1 is bounded by $R^{-1}K$, where K does not depend on γ. This will prove that a certain set H_R of equally (Hölder) continuous functions has a measure not less than $1 - K/R$. In this way the set of equally bounded and continuous functions will have the measure not less than $1 - K/R - Ce^{-L} > 1 - \epsilon$. As K_1 and K_2 in Eq. (II.4) are numerical constants independent of the measure μ_γ the following lemma completes the proof of Theorem III.6.

Lemma III.7. Assume (iii) and that $M(g)$ in assumption (i) does not depend on γ, then

$$E_\gamma[(\ln(\widetilde{B}^j_{2T+1}(0)/T+1))^{(\rho-1)/\rho}]$$

is bounded by a constant independent of γ.

Proof: Let us note first that $\psi_j(x_j)\,[x_j = (0,\ldots,x_j,\ldots 0)]$ as defined in Eq. (II.2) has a realization a.s. continuous in x_j. This continuity could be shown in the same way as the continuity of $\varphi_g(x_j, x_{d-1})$ in x_j. We would define a process $\tilde{\psi}$ having the spectral measure $|p_j|^{\alpha+\epsilon} d\omega_g(p)$, with ϵ such that $\alpha + \epsilon < \frac{1}{2}$, and expressed $\psi_j(x_j)$ as an a.s. limit of an integral over $\tilde{\psi}$ being Hölder continuous with index ϵ. From this continuity and the definition of $\widetilde{B}_T$ [Eq. (II.5)] we get that the inequality

$$\widetilde{B}_{2T+1}(0) \leqslant 2(2T+1) \sup_{|x_j| \leqslant 2T+1} \exp|r\psi_j(x_j)|^{\rho/(\rho-1)}$$
$$= 2(2T+1)\exp\Big[\sup_{|x_j|\leqslant 2T+1} r|\psi_j(x_j)|\Big]^{\rho/(\rho-1)} \tag{III.3}$$

holds true for almost every ψ. Therefore (with the value of r coming from Lemma II.3)

$$E_\gamma[(\ln \widetilde{B}_{2T+1}(0)/T+1)^{\rho-1/\rho}]$$
$$\leqslant \frac{4(M(g))^{-1/\rho}}{a(\rho)} E_\gamma\Big[\sup_{|x_j|\leqslant 2T+1} |\psi_j(x_j)|\Big]$$
$$= -K_1 \int_o^\infty y\, dP_\gamma\big(\sup_{|x_j|\leqslant 2T+1} |\psi_j(x_j)| > y\big). \tag{III.4}$$

Next

$$|\psi_j(x_j)| \leqslant |\psi_j(0)| + |\psi_j(x_j) - \psi_j(0)|$$

and

$$P_\gamma\Big(\sup_{|x_j|\leqslant 2T+1} |\psi_j(x_j)| > y\Big) \leqslant P_\gamma\Big(|\psi_j(0)| > \frac{y}{2}\Big)$$
$$+ P_\gamma\big(\sup_{|x_j|\leqslant 2T+1} |\psi_j(0) - \psi_j(x_j)| > \frac{y}{2}\big),$$
$$P_\gamma\Big(|\psi_j(0)| > \frac{y}{2}\Big) \leqslant e^{-y/2} K_1.$$

With K_1 independent of γ follows from assumption (iii) by the same argument which was used in the proof of Lemma III.5. It remains to estimate the second term in Eq. (III.5). For this purpose, due to the fact that $g \in \mathscr{S}(R^d)$, the knowledge of the covariance of ψ will be sufficient. This follows from the following (Gikhman and Skorohod,[17] Sec. III.5).

Lemma: Let ψ_s be a stochastic process such that

$$E[(\psi_s - \psi_0)^2] \leqslant \mathscr{A}|s|^{1+\epsilon}, \quad \epsilon > 0, |s| < 2T+1. \tag{III.5}$$

Then there exists a constant K depending only on ϵ and T such that

$$P\Big(\sup_{|s|\leqslant 2T+1} |\psi_s - \psi_0| > \frac{y}{2}\Big) \leqslant \frac{K\mathscr{A}}{y^2}. \tag{III.6}$$

Our $\psi_j(x_j)$ depends only on one coordinate and [cf. Eq. (II.19)]

$$E_\gamma[(\psi_j(x_j) - \psi_j(0))^2]$$
$$= 2\int d\sigma(p) |\tilde{g}(p)|^2 |p_j|^{2\alpha}(1 - \cos p_j x_j)$$
$$\leqslant x_j^2 \int d\sigma(p) |\tilde{g}(p)|^2 |p_j|^{2\alpha+2} = x_j^2 E_\gamma[(\tilde{\varphi}(\tilde{g}(p)|p_j|^{1+\alpha}))^2]. \tag{III.7}$$

But

$$E_\gamma[\varphi(h)^2] \leqslant E_\gamma[\exp\varphi(h) + \exp(-\varphi(h))]$$
$$\leqslant I(h) + I(-h)$$

from assumption (iii). Hence, $\mathscr{A}$ in Eqs. (III.5) and (III.6) of the Lemma can be chosen independent of γ. This leads to the estimate

$$P_\gamma\Big(\sup_{|x_j|\leqslant 2T+1} |\psi_j(x_j)| > y\Big) \leqslant K_2\Big(e^{-y/2} + \frac{1}{y^2}\Big), \tag{III.8}$$

with K_2 independent of γ. Integrating by parts in Eq. (III.4) we get that

$$E_\gamma\Big[\Big(\ln\frac{\widetilde{B}^j_{2T+1}(0)}{T+1}\Big)^{(\rho-1)/\rho}\Big]$$
$$= K_1 \int_0^\infty P_\gamma(\sup|\chi_j(x_j)| > y)\, dy \leqslant K,$$

K being independent of γ. This completes the proof of Lemma III.7 and Theorem III.6.

Remark: We could also consider in this section transformations $(g'\varphi)(x) = \varphi_g(x) (\varphi \in \mathscr{S}')$ with $g \notin \mathscr{S}(R^d)$ such that the condition (ii) is fulfilled. A minor complication then arises from the fact that not all $\varphi_g(x)$ are continuous, but almost all. In such a case v could be defined by means of the finite dimensional distributions (III.3)

$$\mathscr{P}(u_1 x_1, \ldots, u_n x_n) = \int \prod_{j=1}^n \exp i\varphi_g(x_j)\, d\mu(\varphi).$$

The results of Sec. II show that v can be realized as a measure on the space of continuous functions with the σ-algebra of measurable sets (including Borel sets of Φ) generated by the finite dimensional (cylinder) sets. Then the weak compactness of the set of measures $\{v_\gamma\}$ follows if only we are able to prove Lemma III.7. The proof of this lemma is based on the uniform integrability in γ of sup $|\psi_j(0) - \psi_j(x_j)|$. This uniform integrability is true[12] if $E\,[(\psi_0 - \psi_s)^r] \leqslant A\,|s|^{1+\epsilon}$ for certain r and A independent of γ. For such an estimate to be fulfilled [for all g of (i)] an additional assumption is needed, e.g., $E_\gamma\,[\varphi\,(g)^{2r}] \leqslant C\,(E_\gamma\,[\varphi\,(g)^2])^r$ would be sufficient [and holds true in $P(\varphi)_2$ and $(\varphi^4)_3$]. Then Theorem III.3 leads to the conclusion that the Hölder continuity (with index α) of paths remains preserved in the limit $\gamma \to \infty$ if it is true for each γ (this is of some relevance to the problem of existence of noncanonical field theories).

IV. AN APPLICATION: WEAK COMPACTNESS OF MEASURES IN THE REGULARIZED $P(\varphi)_d$

In the constructive field theory (see, e.g., Refs. 1 and 2) one considers measures of the form

$$d\mu_\kappa^\Lambda\,(\varphi) = \left\{\exp\left[-\int_\Lambda :P(\varphi_\kappa(x)):dx\right] d\mu_0(\varphi)\right\}^{-1} \times \exp\left[-\int_\Lambda :P(\varphi_\kappa(x)):dx\right] d\mu_0(\varphi), \qquad \text{(IV.1)}$$

where μ_0 is the Gaussian measure with the covariance $(-\Delta + m_0^2)^{-1}(x,y)$, Λ is a bounded region in R^d, $\varphi_\kappa(x) = \tilde{\varphi}\,(e^{ipx}\omega_\kappa(p))$, with ω_κ such that $\langle \varphi_\kappa^2 \rangle = \int d\mu_0 \varphi_\kappa^2(x) < \infty$ and $:P(\varphi_\kappa):$ is the normal ordered [1,2] polynomial of order $2n$. We are then interested in the limits $\Lambda \to R^d$, $\omega_\kappa \to 1 (\kappa \to \infty)$ of the characteristic function $E_\kappa^\Lambda\,[\exp\, i\varphi_\kappa(g)] \equiv \int d\mu_\kappa^\Lambda(\varphi)$ $\exp\, i\varphi_\kappa(g)$ or of the moments of μ_κ^Λ. It is known that the limit $\Lambda \to R^d$ cannot exist in the sense of the convergence of $\mu_\kappa^\Lambda(\mathscr{A})$ for all sets $\mathscr{A} \subset \mathscr{S}'$, because the infinite volume and finite volume measures are mutually singular. This also seems to be true regarding the $\kappa \to \infty$ limit if $d > 2$.

We will choose $\omega_\kappa(p)$ independent of p_1. In such a case the Osterwalder–Schrader (O.S.) positivity[11] is preserved (but the Euclidean invariance violated). Then we can get a useful bound on $E\,[e^{\varphi(g)}]$ coming from the chessboard estimates[2,18]

$$E_\kappa^\Lambda\,[\exp\, u\varphi_\kappa(g)] \equiv \int \exp u\varphi_\kappa(g) d\mu_\kappa^\Lambda(\varphi) \leqslant \exp \int [\alpha_\infty^\kappa\,(P - ug(x)\varphi) - \alpha_\infty^\kappa\,(P)]dx \qquad \text{(IV.2)}$$

if Λ is invariant under the reflection $x_1 \to -x_1$, and $g(x)$ has its support in Λ, where

$$\alpha_\infty^\kappa\,(P) = \lim_{\Lambda \to R^d} \frac{1}{|\Lambda|} \ln \int \exp\left[-\int_\Lambda :P\,(\varphi_\kappa(x)):dx\right] d\mu_0(\varphi) \qquad \text{(IV.3)}$$

and $P - ug\varphi$ means that the interaction polynomial $P(\varphi)$ has been replaced by $P(\varphi) - ug\varphi$ [the existence of the limit in (IV.3) is again the result of chessboard estimates, i.e., a consequence of O.S. positivity]. If g has a compact support then the right side of the inequality (IV.3) does not depend on Λ, because it involves integration only over the support of g. By compactness arguments a limit $\Lambda \to R^d$ exists, although it may be nonunique. It seems that for small couplings (and $m_0 > 0$) the cluster expansion [19] would work for the ultraviolet regularized theory, proving the existence of the unique infinite volume limit. Further on it will be assumed that a translation invariant infinite volume limit exists and that a measure μ_κ exists such that

$$\lim_{\Lambda \to R^d} E_\kappa^\Lambda\,[\exp u\varphi_\kappa(g)] = E_\kappa\,[\exp u\varphi_\kappa(g)] = \int \exp u\varphi_\kappa(g) d\mu_\kappa(\varphi).$$

Equation (IV.2) shows that it is sufficient to get a bound on the pressure α_∞^κ in order to establish (i) of Sec. II and (iii) of Sec. III. Let us note first the following lemma proved in Ref. 20 (Lemma VII.11).

Lemma IV. 1 (Guerra, Rosen, Simon). Let $P(x) = \Sigma_{j=2}^{2n} a_j x^j (a_{2n} > 0)$, then there exists a constant A (depending only on n) such that

$$:P(\varphi_\kappa(x)): - ug\varphi_\kappa(x) \geqslant -a_{2n} A\,[\langle \varphi_\kappa^2 \rangle^n + \sigma(a,ug)], \qquad \text{(IV.4)}$$

where

$$\sigma(a,ug) = \sum_{j=1}^{2n-2} \left(\frac{a_{2n-j}}{|a_{2n}|}\right)^{2n/j} + |u|^{2n/(2n-1)} \times a_{2n}^{-2n/(2n-1)} |g|^{2n/(2n-1)}.$$

As a consequence of Eqs. (IV.3) and (IV.4) we get

Lemma IV.2. With certain constant $D(\kappa, a_j)$

$$\alpha_\infty^\kappa\,(P - ug\varphi) \leqslant D(\kappa, a_j) + A\,|u|^{2n/(2n-1)} (a_{2n})^{-1/(2n-1)} |g|^{2n/(2n-1)}. \qquad \text{(IV.5)}$$

Lemma IV.3. Assume g has a compact support, then for the polynomial $P(x) = \Sigma_{j=2}^{2n} a_j x^j$,

$$E_\kappa\,[\exp\, u\varphi_\kappa(g)] \leqslant \exp\,[(D(\kappa, a_j) - \alpha_\infty^\kappa\,(P))|\,\mathrm{Supp}\, g|] \times \exp\left[A\,|u|^{2n/(2n-1)} (a_{2n})^{-1/(2n-1)} \int dx |g(x)|^{2n/(2n-1)}\right]. \qquad \text{(IV.6)}$$

Proof: From Eqs. (IV.2) and (IV.5)

$$\begin{aligned} E_\kappa\,[\exp\, u\varphi_\kappa(g)] &\leqslant \exp(-\alpha_\infty^\kappa\,(P)|\,\mathrm{Supp}\, g|] \times \exp \int_{\mathrm{Supp}\, g} \alpha_\infty^\kappa\,(P - ug(x)\varphi)\,dx \\ &\leqslant \exp\,[(D(\kappa, a_j) - \alpha_\infty^\kappa\,(P))|\mathrm{Supp}\, g|] \\ &\quad \times \exp\left[A\,(a_{2n})^{-1/(2n-1)} |u|^{2n/(2n-1)} \int dx |g(x)|^{2n/(2n-1)}\right]. \end{aligned}$$

Theorem IV.4. Assume that in a translation invariant $P(\varphi_\kappa)_d$, α_∞^κ is bounded in κ from above and that $a_{2n}(\kappa) > \tilde{a}_{2n} > 0$, then the set of measures $\{v_\kappa\}$, $v_\kappa = g^{-1}\mu_\kappa$ ($g \in C_0^\infty(R^d)$) is weakly conditionally compact.

Proof: The assumptions of this theorem together with Lemma IV.3 and Eq. (IV.2) ensure the fulfillment of the conditions of Theorem III.3.

Remarks: (1) We allow here an arbitrary dependence of a_j on κ in order to take into account the counterterms.
(2) Instead of introducing φ_κ as $\tilde{\varphi}\,(e^{ipx}\omega_\kappa(p))$ we may equivalently leave φ unregularized, replacing μ_0 by the Gaussian

measure with the covariance being the Fourier transform of $(p^2+m_0^2)^{-1}\omega_\kappa^2(p)$.

V. DISCUSSION

The investigation of properties of the Euclidean fields deals with measures of certain measurable subsets of $\mathscr{S}'$. We have discussed here in detail the continuity properties (see also Refs. 1, 21, and 22 for other results in this field). The weak compactness of measures established in Sec. III allows, on the basis of Theorem III.3, one to determine the measure $\mu(A)$ of a set A if $\mu_\gamma(A)$ are known and if $\mu_\gamma \to \mu$ in the sense of the convergence of the characteristic functions. The sets A, which are considered in Sec. III, are of a special form. They should belong to the σ-algebra $\mathscr{Z}_c$ generated by the cylinder sets

$$\{\varphi\in\mathscr{S}' : (\varphi_g(x_1),\dots,\varphi_g(x_n))\subset B\subset R^n\}. \tag{V.1}$$

If now for a set $K\in\mathscr{Z}_c$, closed in the Φ topology (Sec. III), $\mu_\gamma(K)=1$, then also $\mu(K)=1$. We have discussed some closed sets K describing the behavior (fluctuations[6]) of $\varphi_g(x)$ at $x\to\infty$ in ultraviolet regularized $P(\varphi)_d$ in our earlier paper.[5] The behavior at infinity of a $P(\varphi)_d$ random field and the Gaussian field are different. Due to the weak compactness of μ_κ from Sec. IV, if K is closed and $\mu_\kappa(K)=1$, then also $\mu(K)=1$. If for a Gaussian measure $\mu_0(K)\neq 1$, we could conclude that μ is non-Gaussian. We were unable in Ref. 5 to exclude the Gaussian limits without additional (rather strong) assumptions. The possible improvement of these results could possibly be obtained with the use of κ dependent sets K and some results of Topsoe[23] on uniformity in the weak convergence.

We have restricted ourselves in this paper and in Ref. 5 to the σ-algebra $\mathscr{Z}_c$ generated by the sets (V.1). We have based our approach on the results of Prokhorov[4] on weak convergence in metric spaces. It appears that the weak convergence of measures on nuclear[7] and even more general Suslin spaces[8] (Appendix) follows from the pointwise convergence of the characteristic functions. Then Theorem III.3 remains true. We can conclude, e.g., that $\lim\sup\mu_\gamma(K)\leqslant\mu(K)$ if K is closed in the weak topology of $\mathscr{S}'$. Unfortunately, we were unable to describe the fluctuations of φ in terms of sets K closed (or open) in this topology. The difficulty stems from the fact that in order to describe φ at infinity, we first regularize the distribution φ. Then, the sets generated by $\varphi_g(x)$, with fixed g, are not closed in the weak topology of $\mathscr{S}'$.

ACKNOWLEDGMENT

The author wishes to thank the referee for showing him the relevance of the results of Fernique and Schwartz on weak convergence.

[1] *Constructive Quantum Field Theory, Lecture Notes in Physics, Vol.* 25, edited by G. Velo and A. Wightman (Springer, New York, 1973).
[2] F. Guerra, L. Rosen, and B. Simon, Ann. Math. **101**, 111 (1975).
[3] Z. Haba, Commun. Math. Phys. **69**, 247 (1979).
[4] Yu. V. Prokhorov, Theor. Probab. Appl. **1**, 157 (1956).
[5] Z. Haba, "Fluctuations of $P(\varphi)$ random fields," J. Math. Phys. (1981), in print.
[6] J. Rosen and B. Simon, Ann. Probab. **4**, 155 (1974).
[7] P. A. Meyer, Le théoréme de continuité de P. Lévy sur les spaces nucléaires (d' aprés X. Fernique), Seminaire Bourbaki, Exposés 295-312, 1965/66, p. 311.
[8] L. Schwartz, *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures* (Oxford University, London, 1973).
[9] E. Nelson, J. Funct. Anal. **12**, 97 (1973).
[10] G. Hegerfeldt and J. R. Klauder, Commun. Math. Phys. **16**, 329 (1970).
[11] K. Osterwalder and R. Schrader, Commun. Math. Phys. **33**, 83 (1973); **42**, 281 (1975).
[12] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[13] I. M. Gel'fand and N. I. Vilenkin, *Generalized Functions, Vol. IV, Some applications of harmonic analysis* (Academic, New York, 1964).
[14] V. S. Varadarajan, Mat. Sb. **55**, 35 (1961) (English translation in Am. Math. Soc. Transl. Ser. 2 **45**, 161 (1965).
[15] K. Maurin, *Analysis*, Vol. 2 (PWN– D. Reidel, Warszawa, 1980).
[16] S. B. Myers, Ann. Math. **47**, 496 (1946).
[17] Y. Y. Gikhman and A. V. Skorohod, *Theory of Stochastic Processes*, Vol. 1 (Moscow, 1971) (English translation in Springer Verlag).
[18] J. Fröhlich and B. Simon, Ann. Math. **105**, 493 (1977).
[19] J. Glimm and A. Jaffe, "The cluster expansion," in Ref. 1.
[20] F. Guerra, L. Rosen, and B. Simon, Ann. Inst. Henri Poincaré **25A**, 232 (1976).
[21] J. T. Cannon, Commun. Math. Phys. **35**, 215 (1974).
[22] G. Benfatto, G. Gallavotti, and F. Nicolo, J. Funct. Anal. **36**, (1980).
[23] F. Topsoe, *Topology and Measure, Lecture Notes in Math.* 133 (Springer, New York, 1970).

Comment on calculations of higher-order tadpoles and triangle vertex integral

M. W. Kalinowski
Institute of Philosophy and Sociology, Polish Academy of Sciences Nowy Świat 72,00-330, Warsaw, Poland
M. Seweryński [a] and L. Szymanowski [b]
Institute of Nuclear Research, Hoża 69,00-681 Warsaw, Poland

(Received 25 February 1981; accepted for publication 1 May 1981)

It is found that the calculations of higher-order tadpoles and some integrals associated with the triangle diagram reported by Capper and Leibbrandt are incorrect within their dimensional regularization scheme.

PACS numbers: 11.10.Gh

In this paper we will demonstrate that the calculations of higher-order tadpole integrals and the integral associated with pure graviton triangle diagram reported in Ref. 7 are incorrect. The authors of Ref. 1 considered the integrals

$$J_1 = \int d^{2w}q_1 d^{2w}q_2\, [q_1^2 q_2^2 (q_1 - q_2)^2]^{-1},$$

$$J_2 = \int d^{2w}q d^{2w}p\, [p^2(q-p)^2 q^4]^{-1},$$

$$J_3 = \int d^{2w}k\, [k^2(k-p_2)^2(k+p_3)^2]^{-1},$$

and they claimed that within the dimensional regularization scheme proposed previously in Ref. 2

$$J_1 = \pi^{2w}\Gamma(1-w)\Gamma(2-w) f^{2w-3}, \tag{1}$$

$$J_2 = \pi^{2w}\Gamma(2-w)\Gamma(2-w) f^{2w-4}, \tag{2}$$

$$\begin{aligned} J_3 = {} & \frac{1}{2}\pi^{w}(p_1^2)^{w-3}[\Gamma(3-w)/(w-2)] \\ & \times \int_0^1 d\xi\, (c_0 + c_1\xi + c_2\xi^2)^{-1/2} \\ & \times [z_1^{w-2}\, {}_2F_1(w-2, \tfrac{1}{2}; w-1; z_1/R^2) \\ & - z_0^{w-2}\, {}_2F_1(w-2, \tfrac{1}{2}; w-1; z_0/R^2)], \end{aligned} \tag{3}$$

where

$$R = (c_0 + c_1\xi + c_2\xi^2)^{1/2}, \quad p_2 + p_3 = -p_1,$$
$$f_0 = f(w)/p_1^2,$$
$$c_0 = (1+4f_0)/4, \quad z_0 = f_0 + \xi(1-\xi)p_2^2/p_1^2,$$
$$c_1 = -(p_2 p_3)/p_1^2, \quad z_1 = f_0 + \xi(1-\xi)p_3^2/p_1^2,$$
$$c_2 = (p_2p_3)^2 - p_2^2 p_3^2/p_1^4$$

[see Ref. 1, Eqs. (8), (13), (19), (20)].

An examination of the method applied for calculations of J_1 and J_2 reveals that the formulae (8) and (13) given in Ref. 1 have been obtained by a misuse of the t' Hooft–Veltman formula

$$\int \frac{d^{2w}q}{(2\pi)^{2w}} (q^2)^n = 0, \quad w \in \mathbb{C}, \; n \in N_0^\infty \tag{4}$$

[see, e.g. Ref. 3, Eq. (3.13), and Ref. 4, Eq. (2.6)]. To prove our statement, let us repeat the calculations along the lines presented in Ref. 1. First we rewrite J_1 in the form

$$J_1 = \pi^w \Gamma(2-w) \int_0^1 d\xi d^{2w}q_1\, (q_1^2)^{-1} [q_1^2 \xi(1-\xi) + f]^{w-2}$$

(see Ref. 1 for an explanation of symbols) and then expand the term in the square bracket in power series about q_1^2. This yields

$$\begin{aligned} J_1 = \pi^w\Gamma(2-w)\int_0^1 d\xi \Big\{ \Big[\int d^{2w}q_1\, (q_1^2)^{-1} \times [f^{w-2} + q_1^2\xi(1-\xi)(w-2) f^{w-3} + \cdots \\ + (1/n!)(1-\xi)^n \xi^n (w-2)(w-3)\cdots(w-1-n)(q_1^2)^n f^{w-2-n} + \cdots \Big] \Big\} \end{aligned} \tag{5}$$

[see Ref. 1, Eq. (7)]. Now it is claimed in Ref. 1 that, due to the H-V formula (4), all terms in (5) vanish except the first one. The point to note is that the last statement is equivalent to the presumption that

$$\int d^{2w}q \lim_{n\to\infty} (q^2)^n = \lim_{n\to\infty} \int d^{2w}q\, (q^2)^n, \tag{6}$$

and, since the last integral vanishes [see (4)] for any finite $n \in N_0^\infty$, then the conclusion $\int d^{2w}q\,(q^2)^n = 0$ follows. However, one observes that:

(a) Eq. (6) does not follow from (4);

(b) if (6) is taken at its face value, then it quickly leads to the following contradiction with the Capper–Leibbrandt extension of the Gaussian integral:

$$\int \frac{d^{2w}q}{(2\pi)^{2w}} \exp(-xq^2 + 2bq) = (4\pi)^{-w} x^{-w} \exp(b^2/x - fx) \tag{7}$$

(see Ref. 2).

First it is obvious that our formula (2.6) given in Ref. 4 has been proven for $(w, z) \in \mathbb{C}^2$ and does not hold when (w, z) belongs to compactified $\mathbb{C}^2$. And an analysis of the Capper–Leibbrandt proof of (4) (see Ref. 3) reveals that their proof

[a] Supported by MNSzWT Grant No. 04.3.14.02.05.2A — 1K8E and partially supported by INT Grant No. 73 — 20002 A01.
[b] Supported by MSzWNT Grant No. 04.03.14.02.05.2A — 1K7E.

 0022-2488/82/081501-03$02.50

also does not cover the limit case $n\to\infty$ [the series $\Sigma_{j=0}^{n}(-1)^j\binom{n}{j}$ makes no sense when $n\to\infty$]. Morever, if (6) were true, then

$$\int d^{2w}q\, e^{-q^2}=\int d^{2w}q\sum_{n=0}^{\infty}\frac{(-1)^n(q^2)^n}{(1)_n}$$
$$=\sum_{n=0}^{\infty}(-1)^n\int d^{2w}q\,(q^2)^n/(1)_n$$

and one gets $\int d^{2w}q\, e^{-q^2}=0$ in such a case. But from (7) one gets immediately that $\int d^{2w}\exp(-q^2)=\pi^w\exp(-f)$. So (6) cannot be considered even as an additional assumption an (4) holds for finite n, $n\in N_0^\infty$, only. (This fact is usually reflected by the statement "integrals over *polynomials* give zero within the dimensional regularization scheme; see, e.g., Ref. 5, p. 107). Thus the final formulae for J_1 and J_2 reported in Ref. 1 remain to be proven, and (4) is not too useful in the treatment of multiloop massless integrals (see Ref. 3, Sec. 4).

Now let us consider J_3. It follows from (3) that $J_3=0$ provided $p_2^2=p_3^2$ and $p_1^2\neq 0$. However, we will demonstrate that if this integral is calculated correctly, then, in general, $J_3\neq 0$ when $p_1^2\neq 0$ and $p_2^2=p_3^2$. It is obvious that J_3 can be rewritten in the form

$$J_3=C\int\frac{d^{2w}k}{ab}+B\int\frac{d^{2w}k}{ac}+A\int\frac{d^{2w}k}{bc},$$

where $a=(k+p_3)^2$, $b=(k-p_2)^2$, $c=k^2$, and the coefficients A, B, and C obey the relation

$$Ay+Bz+(A+B+C)k^2+2(Ap_3-Bp_2)k=1,$$
$$y=p_3^2,\quad z=p_2^2. \tag{8}$$

If we assume that

$$Ap_3-Bp_2=0,\quad A+B+C=0,\quad Ay+Bz=1 \tag{9}$$

then (8) is evidently satisfied. Further, we will assume that this is the case and additionally suppose that $p_2/(p_2^2)^{1/2}=p_3/(p_3^2)^{1/2}$ (we presume that $p_1^2\neq 0$, $p_2^2\neq 0$, $p_3^2\neq 0$). After some algebra, one finds

$$A=\frac{\sqrt{z}}{y\sqrt{z}+z\sqrt{y}},\quad B=\frac{\sqrt{y}}{y\sqrt{z}+z\sqrt{y}},$$

$$C=-\frac{\sqrt{y}+\sqrt{z}}{y\sqrt{z}+z\sqrt{y}}.$$

Shifting the variable $k\to k+p_2$ in the first integral and using the relation $p_1=-(p_2+p_3)$, one obtains

$$J_3=C\int\frac{d^{2w}k}{k^2(k-p_1)^2}+B\int\frac{d^{2w}k}{k^2(k+p_3)^2}+C\int\frac{d^{2w}k}{k^2(k-p_2)^2}.$$

The integrals of this type have been calculated already in (Ref. 4 see Appendix A), and we give below the explicit formula only for the case $N=1$ (to compare the result with given in Ref. 1). One finds

$$J_3=\pi^w\Gamma(2-w)\frac{\Gamma(w-1)\Gamma(w-1)}{\Gamma[2(w-1)]}$$
$$\times\Big[C(4f+p_1^2)^{w-2}\,{}_2F_1\Big(\tfrac12,2-w;\tfrac32;\frac{p_1^2}{4f+p_1^2}\Big)$$
$$+B(4f+p_3^2)^{w-2}\,{}_2F_1\Big(\tfrac12,2-w;\tfrac32;\frac{p_3^2}{4f+p_3^2}\Big)$$
$$+A(4f+p_2^2)^{w-2}\,{}_2F_1\Big(\tfrac12,2-w;\tfrac32;\frac{p_2^2}{f+p_2^2}\Big)\Big]\Big/$$
$${}_2F_1(\tfrac12,2-w;\tfrac32;1). \tag{10}$$

It is easy to see that when $p_2=p_3$ (i.e., $p_2^2=p_3^2$), then (10) yields

$$J_3=\frac{\pi^w}{p_2^2}\Gamma(2-w)\frac{\Gamma(w-1)\Gamma(w-1)}{\Gamma[2(w-1)]}$$
$$\times\Big[(4f+p_2^2)^{w-2}\,{}_2F_1\Big(\frac12,2-w;3/2;\frac{p_2^2}{4f+p_2^2}\Big)$$
$$-(4f+4p_2^2)^{w-2}\,{}_2F_1\Big(1/2,2-w;3/2;\frac{p_2^2}{f+p_2^2}\Big)\Big]\Big/$$
$${}_2F_1(\tfrac12,2-w;3/2;1). \tag{11}$$

Obviously, the right-hand side of (11) does not vanish identically; thus our conclusion that, in general, $J_3\equiv 0$ when $p_2^2=p_3^2$ follows. Since the sum in the square bracket in (10) vanishes at $w=2$, one easily finds that $\lim_{w=2}(2-w)J_3=0$; hence J_3 has no pole at $w=2$ in this particular case.

In the general case $N\neq 0$ [see Ref. 4, Eq. (2.2)], one has

$$J_3=\pi^w\int\int\int dx\,dy\,dz\,(x+y+z)^{-w}$$
$$\times\exp\Big[-\frac{\alpha yz+\beta xy+\gamma xz}{x+y+z}-(x+y+z)^Nf\Big]$$
$$=\pi^w\int_0^1dx\int_0^{1-x}dy\int_0^\infty dz\,z^{2-w}$$
$$\times\exp\{-z^Nf-z[\alpha y(1-x-y)+\beta x(1-x)+\gamma xy]\}$$
$$=\pi^w\int_0^1dx(1-x)\int_0^1dt\int_0^\infty dz\,z^{2-w}$$
$$\times\exp\{-z^Nf-z(1-x)[\alpha(1-x)t(1-t)$$
$$+\beta x(1-t)+\gamma xt]\}$$
$$=\pi^w\int_0^\infty dz\,z^{2-w}\exp(-z^Nf)\int_0^1dx\,x\int_0^1dy\exp(-zA),$$

where

$$A=x\{(1-x)[-\alpha y(1-y)+\beta(1-y)+\gamma y]+\alpha y(1-y)\},$$
$$\alpha=p_1^2,\quad \beta=p_2^2,\quad \gamma=p_3^2.$$

Unfortunately, such a complicated structure of J_3 prevented us from performing all integrations explicitly. Although J_3, when $N=1$, can be reduced to the form

$$J_3=\pi^w\frac{\Gamma(3-w)}{\Gamma(3)}\int_0^1dx$$
$$\times[f+x(1-x)\beta]^{w-3}\cdot x\cdot F_1\Big(1,3-w,2;\frac{1}{y_1},\frac{1}{y_2}\Big),$$

where

$$y_{1/2} = [\alpha x + (\gamma - \beta)(1-x) \pm \sqrt{\Delta}\,]/2\alpha x,$$

$$\Delta = [\alpha x + (\gamma - \beta)(1-x)]^2 + 4\alpha[\,\beta x(1-x) + f\,],$$

the last integrand is still very complicated. However, J_3 can be calculated explicitly also when, e.g., $p_2 = -p_3$, $p_2^2 = 0, p_3^2 = 0$. In this particular case one gets

$$J_3 = \frac{\pi^w}{2}\int_0^\infty d\, z^{(3-w)-1} e^{-z^N f} = \frac{\pi^w}{2}\,\frac{\mathrm{sgn}N}{N} f^{(w-3)/N} \times \Gamma\left(\frac{3-w}{N}\right).$$

Hence J_3 has a pole of a high order when $N = 1$ (and vanishes when $N < 0$) at the physical point $w = 2$. It is obvious now that a singularity structure of J_3 in the w plane is a very complicated function of external momenta p_1, p_2, p_3. Moreover, it is not a continuous function of p_1^2, p_2^2, p_3^2 (see also Ref. 4).

Summarizing, we have demonstrated that the calculations of J_1, J_2, J_3 reported in Ref. 1 are unreliable and the final conclusion concerning J_3 reached in Ref. 1 that J_3 is free of divergences at $w = 2$ is, in general, incorrect.

[1] D. M. Capper and G. Leibbrandt, J. Math. Phys. **15**, 795 (1974).
[2] D. M. Capper and G. Leibbrandt, J. Math. Phys. **15**, 82 (1974).
[3] D. M. Capper and G. Leibbrandt, J. Math. Phys. **15**, 86 (1974).
[4] M. W. Kalinowski, M. Seweryński, and L. Szymanowski, "On some generalizations of Gaussian integral and the dimensional regularization," Preprint IFT/19/1980.
[5] G. t' Hooft and M. Veltman, "Diagrammar," CERN Preprint R/17172, August 1973.

Superfiber bundle structure of gauge theories with Faddeev–Popov fields

J. Hoyos
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Matemáticas, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

M. Quirós, J. Ramírez Mittelbrunn, and F. J. de Urríes
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Fisica, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

(Received 22 June 1981; accepted for publication 5 March 1982)

In this work the mathematical structure of principal superfiber bundle P_s is used to give a geometrical description of gauge theories. The base space of P_s is an $S^{4,2}$-supermanifold X_s (four commuting and two anticommuting variables), and the structure group an $S^{n,0}$ supergroup G_s, where n is the dimension of the gauge group for the classical theory. The body of P_s is the usual principal fiber bundle P of gauge theories. Gauge and Faddeev–Popov fields arise as superfields, components of the connections in P_s, in a local coordinate system. BRS (Becchi, Rouet, and Stora) and anti-BRS transformations are gauge transformations, in P_s, of parameters the ghost and antighost superfields, respectively. In the case of soul-flat connections, which are connections in P_s coming from connections in P, the BRS and anti-BRS transformations are finite translations along the anticommuting directions of X_s, and generate an $S^{0,2}$-supergroup.

PACS numbers: 11.10.Np, 11.30.Pb

I. INTRODUCTION

The mathematical structure of gauge theories was displayed in 1975 by Wu and Yang,[1] who established a dictionary of equivalences between physical terms and mathematical concepts. Thus a pure gauge theory is described by a principal fiber bundle $P(X, G)$, with base space X (the space–time manifold) and structure group G (the gauge group). The gauge fields A_μ and the strength field tensor $F_{\mu\nu}$ are the coefficients of a connection (1-form) in the principal fiber bundle and its curvature (2-form), respectively, in a local coordinate system in P. In this scheme the gauge transformations are fiber bundle equivalences.

When a gauge theory is quantized, one needs to introduce in the Lagrangian a gauge fixing term so that the effect it produces is twofold: (a) New (spinless and anticommuting) fields appear, the so-called Faddeev–Popov (FP) fields; (b) the original gauge invariance is broken and a new invariance, first discovered by Becchi, Rouet, and Stora[2] (BRS) arises. FP fields and BRS transformations do not appear in the Wu and Yang's dictionary. Nevertheless, several attempts have been recently done in order to provide them with a precise geometrical interpretation.

In our knowledge, the first essay in this sense was achieved by Thierry-Mieg,[3] associating the FP ghost with the vertical component of the connection 1-form in the principal fiber bundle $P(X, G)$ and the BRS transformation with the exterior differential along the fiber. In this way the FP ghosts, as 1-forms, have the physically required anticommutativity properties. However, in this scheme, FP ghosts get a space–time dependence only by means of changes of coordinates in the fiber bundle, which cannot be interpreted as gauge transformations. On the other hand, neither do FP antighosts have any geometrical interpretation nor do they transform under BRS transformations.

Following the idea of interpreting FP fields as 1-forms we introduced, in a previous work,[4] an enlarged principal fiber bundle allowing one to interpret geometrically FP ghosts and antighosts. It was thus possible to give both of them a space–time dependence by means of true gauge transformations. BRS transformations have, in this scheme, an interpretation similar to Thierry-Mieg's one. As a consequence of our enlarged mathematical structure, new anti-BRS transformations did appear together with the usual BRS transformations. The new transformations had already been introduced in the literature by Curci and Ferrari[5] and, more recently, by Ojima,[6] without any relation to the geometrical structure of gauge theories.

The interpretation of FP fields as 1–forms on a finite-dimension manifold is not, however, fully satisfactory since it would lead to the vanishing of Green functions containing a number of FP fields higher than the dimension of the manifold.[7] If we wish to save the former problem, casting the interpretation of FP fields as 1-forms away, we need, as usual, to introduce an infinite-dimensional Grassmann algebra in order to give the anticommuting character to FP fields. On the other hand BRS transformations map commuting into anticommuting fields, and this fact led Ferrara, Piguet, and Schweda[8] to interpret them as supersymmetric transformations. Recently, Tonin and Bonora[9] have extended these ideas also to anti-BRS transformations. These authors introduce[9] a superspace, with two anticommuting variables, and define a 1-form on it, to be identified with a certain connection, whose coefficients are superfields involving gauge and FP fields. Furthermore, by imposing some conditions of null curvature, they interpret the BRS and anti-BRS transformations as translations along the anticommuting variables. In a later work,[10] Bonora, Pasti, and Tonin searched for a geometrical structure where the former 1-form would correspond to a connection form. They do not find a structure of principal fiber bundle, as would be expected in a gauge theory, but they are constrained to introduce "ad hoc" a much more complicated structure, called by them quasifiber bundle.

 0022-2488/82/081504-07$02.50

The aim of this work is to find out a natural geometrical interpretation of FP fields and BRS and anti-BRS transformations within the framework of the structure of principal fiber bundle. To this end, we use the concepts of supermanifold, supergroup, and superfiber bundle which have been developed in another paper.[11]

In Sec. II we present a summary of concepts and results of the differential supergeometry,[11] which will be useful in the present work. In Sec. III we build a principal superfiber bundle whose body (real part) is the principal fiber bundle $P(X, G)$ of unquantized gauge theories (that is the theory without FP and gauge-fixing terms). We define the connection 1-form, on the principal superfiber bundle, whose coefficients are superfields as in Bonora and Tonin's heuristic construction.[9] Among the connection 1-forms we stress those (soul–flat connections) coming from the pullback of a connection in the body $P(X, G)$ and having automatically the required property of null curvature along any anticommuting direction. In Sec. IV we interpret BRS and anti-BRS transformations as mappings, in the space of connections, determined by true gauge transformations, in the superfiber bundle, whose parameters are the superfields corresponding to ghost and antighost fields. We also analyze the mathematical conditions which must be satisfied by the principal superfiber bundle. For soul–flat connections we recover, for the BRS and anti-BRS transformations, the interpretation of translations along the anticommuting variables in agreement with the results of Ref. 9. It is easily proven that the set of these transformations generates an additive supergroup, with two anti commuting parameters, isomorphic to the superspace $S^{0,2}$. The parameters of this supergroup are global (they do not depend on the point of the supermanifold) so that the total invariance of gauge theories (BRS and anti-BRS) is global-like, and the introduction of new ghost fields is no longer necessary.

II. A SUMMARY OF DIFFERENTIAL SUPERGEOMETRY

Since FP fields are anticommuting quantities and the charges generating BRS and anti-BRS transformations anticommute with each other, the suitable framework to analyze the mathematical structure of these objects is differential supergeometry. That is, in short, a geometry where the real, or complex, numbers are replaced by Grassmann numbers.

Analysis on superspaces has recently been considered by Rogers[11] and Jadczyk and Pilch.[12] In this section we give a brief summary of some mathematical concepts and results obtained in a previous work,[13] which will be used later in this paper. The originality of our approach[13] can be mainly stated as follows:

(a) We introduced in Ref. 13 the concept of generalized supermanifold, where different coordinates may belong to different Grassmann algebras: This has been proven very useful, in this paper, for the geometrical interpretation of extended BRS symmetry and FP fields.

(b) Jadczyk and Pilch's approach[12] only applies to infinite-dimensional Grassmann algebras, so that the superfield expansion is lacking because one is forced to handle it with analytic functions. Our approach[13] is free of this failure: In particular, the superspace constructed in Sec. III has its even coordinates belonging to finite-dimensional Grassmann algebras, so that the superfield expansion holds for it.

A. Superspace

We present here a generalization of the concept of superspace introduced in Ref. 13. Let B be a Grassmann algebra generated by $\{\beta_i\}_{i\in I}$ (where I is a finite or countably infinite set of indices) with $\beta_\emptyset = 1$. For $M\in F(I)$, the set of finite parts of I, we shall define B_M as the Grassmann subalgebra generated by $\{\beta_i\}_{i\in M}$ if $M\neq\emptyset$, and $B_\emptyset = R$ (in particular $B_I = B$). B_M is Z_2-graded, $B_M = B^0_M \oplus B^1_M$, with $B^i_M = B^i\cap B_M (i = 0, 1)$. We can endow B with the structure of Banach algebra and then B_M is a Banach subalgebra.

A superspace associated to the triple (Λ, m, n), with $\Lambda = (K_j)_{j=1}^{m+n}$, $K_j\in F(I)$, $m>0$, $n>0$, is the direct product

$$B^0_{K_1}\times\cdots\times B^0_{K_m}\times B^1_{K_{m+1}}\times\cdots\times B^1_{K_{m+n}}$$

which will be indicated by $S^{m,n}_\Lambda$ or $S^{m,n}$. As a product of Banach spaces, $S^{m,n}$ is also a Banach space. The heterogeneity of the K_i will allow a minimal superspace in the construction, which will be worked out in the next sections.

Given a superspace $S^{m,n}$, we shall consider B-valued functions defined on an open set U of $S^{m,n}$. A function f: $U\to B$ is G^0 if it is continuous. f is G^1 if there exists a family $G_i f$: $U\to B$ of G_0 functions such that for all

$$\mathbf{x} = (x_i)_{i=1}^{m+n}\in U, \quad \mathbf{h} = (h_i)_{i=1}^{m+n}\in S^{m,n} \quad \text{with } \mathbf{x}+\mathbf{h}\in U;$$

then

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \sum_{i=1}^{m+n} h_i G_i f(\mathbf{x}) + o(\|\mathbf{h}\|). \tag{1}$$

If $I_m f\subset B_L$, $L\subset I$, a consequence of the definition of G^1 is that a G^1-function f does not depend on the ith coordinate if $K_i\not\subset L$. In the same way, G^k functions $(0\leqslant k<\infty)$ can be defined as usual. The class of G^k-functions will be indicated by $G^k(U)$, which is a Z_2-graded algebra: $f\in G^k(U)^i$ iff $I_m f\subset B^i$ $(i = 0, 1)$.

If $f\in G^\infty(U)$, the following expansion can be written:

$$f(\mathbf{x}) = \sum_{M\in F_{m,n}} \Pi_M(\mathbf{x}) g_M(\mathbf{x}), \tag{2}$$

where $F_{m,n} = F(\{m+1, ..., m+n\})$, $\Pi_M(\mathbf{x}) = x_{i_1}x_{i_2}\cdots x_{i_k}$ if $M = \{i_1,...,i_k\}$, $i_1<i_2<\cdots<i_k$ and $\Pi_\emptyset(\mathbf{x}) = 1$, and $g_M\in G^\infty(U)$ only depends on the even coordinates.

The expansion (2) shows that G^∞-functions coincide with the superfields which appear usually in supersymmetric field theories.

Let us finally note that $G^\infty(U)\subset C^\infty(U, B)$.

We define the body of the Grassmann algebra as the mapping r: $B\to R$ which maps all elements of B into its real part The generalization to the body of the superspace r: $S^{m,n}\to R^m$ is obvious from the former definition. Nevertheless, we shall only define the body of a particular class of open sets in $S^{m,n}$, satisfying certain connectedness properties, and which will be referred to as "good opens."[13]

B. Supermanifold

Let (X_s, X, r) be the triple where X_s and X are C^∞-manifolds and r: $X_s\to X$ a surjective, differentiable mapping.

A structure of bodied supermanifold is

(a) an atlas $\{U_\alpha, \phi_\alpha\}$ in X,

(b) a family $\{\psi_\alpha\}$, $\psi_\alpha: r^{-1}(U_\alpha) \to S^{m,n}$, such that

(i) $\{r^{-1}(U_\alpha), \psi_\alpha\}$ is an atlas of X_s,

(ii) $\psi_\alpha(r^{-1}U_\alpha)$ is a "good open" of $S^{m,n}$, and $r \cdot \psi_\alpha = \phi_\alpha \cdot r$ or, homologically,

$$\begin{tikzcd} r^{-1}(U_\alpha) \arrow[r, "\psi_\alpha"] \arrow[d, "r"'] & \psi_\alpha(r^{-1}U_\alpha) \arrow[d, "r"'] \\ U_\alpha \arrow[r, "\phi_\alpha"] & \phi_\alpha(U_\alpha) \subset R^m \end{tikzcd}$$

(iii) $\psi_{\alpha\beta} = \psi_\alpha \cdot \psi_\beta^{-1}: \psi_\beta(r^{-1}U_\alpha \cap r^{-1}U_\beta) \to \psi_\alpha(r^{-1}U_\alpha \cap r^{-1}U_\beta)$ is a G^∞-diffeomorphism,

(iv) the family $\{U_\alpha, \phi_\alpha, \psi_\alpha\}$ is maximal, and satisfies (i), (ii), and (iii).

The manifold X is called body of the supermanifold X_s. The concept of G^∞ function between two supermanifolds is induced from the definition given in subsection IIA through a local coordinate system.

We define the tangent space $T(X_s)$ to the supermanifold X_s as the space of tangent vectors to curves in X_s. The space $T(X_s)$ has the structure of vector bundle with fiber $S^{m,n}$ and it is a bodied supermanifold, with body $T(X)$. However, we cannot define $T(X_s)$ as the space of derivations of functions defined on X_s. Actually, the space of derivations of $G^\infty(X_s)$-functions is a different tangent space $T_d(X_s)$. The tangent space to a point $p \in X_s$, $T_d(X_s)_p$, is a B-module. To verify this property, let us consider a local coordinate system (x_i) in X_s; every element belonging to $T_d(X_s)_p$ can be written as $\Sigma_{i=1}^{m+n} a_i(p)\partial/\partial x_i$, where $a_i(p) \in B$, while the elements of the tangent space $T(X_s)_p$ have the expression $\Sigma_{i=1}^{m+n} a_i(p)\partial/\partial x_i$, where $\mathbf{a}(p) = (a_i(p)) \in S^{m,n}$.

The construction of the V (graded B-module)-valued graded exterior algebra $D(X_s)$ over the tangent spaces $T_d(X_s)$ has been carried out in detail in Ref. 13, and we shall not dwell upon it. In particular, the exterior product of two forms $\omega_1 \in D^{r_1}$ and $\omega_2 \in D^{r_2}$ satisfies the property

$$\omega_2 \wedge \omega_1 = (-1)^{r_1 r_2 + |\omega_1||\omega_2|} \omega_1 \wedge \omega_2, \tag{3}$$

where ω_i is the Grassmann grade and r_i the grade of the form ω_i, while the exterior differential of a 1-form ω is given by

$$d\omega(X_1, X_2) = (-1)^{|X_1||\omega|} X_1 \omega(X_2) + (-1)^{|X_1||X_2| + 1 + |X_2||\omega|} \times X_2 \omega(X_1) - \omega([X_1, X_2]) \tag{4}$$

where $|X_i|$ is the Grassmann grade of the vector field X_i.

C. Supergroup

A Lie supergroup G_s is a bodied supermanifold with a product law providing the structure of group and such that mapping $(a, b) \mapsto ab^{-1}$, $a, b \in G_s$ is G^∞. As a consequence of the definition we have that $G = r(G_s)$, the body of G_s, is a Lie group and r is a group homomorphism. Let us note that the particular realizations of supergroups studied in the literature[14] do satisfy to our definition.

The Lie algebra $\mathscr{G}_s$ of the supergroup G_s can be identified, in the usual manner, with the tangent space to G_s at the identity e, $T(G_s)_e$. Furthermore, we have the Lie superalgebra $T_d(G_s)_e$, which is graded B-module and normally referred in the literature to as a graded Lie algebra (GLA).[15] Let us finally note that $T_d(G_s)_e^0$ (even part) generates, through the exponential mapping, a Lie supergroup. In particular $T_d(G)_e^0$ generates a Lie supergroup with body G.

D. Superfiber bundles

A principal superfiber bundle $P_s(X_s, G_s)$, with base space X_s and structure group G_s, is a bodied supermanifold, with body $P(X, G)$–principal fiber bundle with base space X, the body of X_s, and structure group G, the body of G_s, satisfying the usual properties of the principal fiber bundle,[16] but where the C^∞-condition for the involved functions is replaced by the G^∞-condition. Indeed, since every G^∞-function is automatically C^∞, as we saw above, a principal superfiber bundle has also the structure of the principal fiber bundle.

A connection in P_s is defined as follows: Let $\omega: T_d(P_s) \to T_d(G_s)_e$ be an even 1-form such that

(i) $\omega(A^*) = A$, where $A \in T_d(G_s)_e$ and A^* is the fundamental vector field in P_s associated with A,

(ii) $R_a^* \omega = \mathrm{ad}(a^{-1})\omega$, $a \in G_s$.

Using the property $\omega(T(P_s)) \subset T(G_s)_e$, we define the connection in P_s as the restriction $\omega: T(P_s) \to T(G_s)_e$. The curvature is defined from the connection in the usual way, satisfying the structure equation $\Omega = d\omega + \frac{1}{2}[\omega, \omega]$.

III. THE MATHEMATICAL SCHEME

In this section we shall proceed to construct a suitable mathematical scheme where BRS and anti-BRS transformations acquire a geometrical interpretation. In a recent work,[9] Bonora and Tonin succeeded to interpret the BRS transformations as sypersymmetric transformations using heuristically the concepts of connections and curvatures in superfiber bundles. In Bonora and Tonin's construction,[9] it is essential to impose the vanishing of the curvature components whenever a direction corresponding to an anticommuting coordinate does appear, in a similar way to what happens in the construction of Ref. 4. An attempt to give a mathematical rigor to the heuristic construction of Ref. 9 has recently been made,[10] but the price to pay was to leave the structure of the principal fiber bundle and replace it by the ill-defined structure of the quasifiber bundle. Here we get that mathematical rigor in the context of the structure of the principal superfiber bundle. That is, we build a structure of principal superfiber bundle whose connection is that heuristically built in Ref. 9 and where the condition of null curvature along the anticommuting variables arises in a natural way.

A. General construction

Let X be the space–time manifold $(R^4, S^4, \cdots)$ and G an n–dimensional internal symmetry compact Lie group. Let $P(X, G)$ be the usual principal fiber bundle in gauge theories, in which the gauge potentials $A_\mu(x)$ are the components of the connection projected over X. Let X_s be an $S^{4,2}$-supermanifold, with $\Lambda = (\emptyset, \emptyset, \emptyset, \emptyset; I, I)$ and I an countably infinite set of indices, whose body is the manifold $X = r(X_s)$ and such that a global section $\tau: X \to X_s$ does exist. Let G_s be an $S_{\Lambda^1}^{n;0}$-supergroup, with $\Lambda^1 = (I, I, \ldots, I)$, whose body is G.

Let us consider the following commutative diagram:

$$\begin{tikzcd} X_s \arrow[d, "r"'] & P_s(X_s, G_s) \arrow[l, "\Pi_s"'] \arrow[d, "\tau"] \\ X & P(X, G_s) \arrow[l, "\Pi"] \end{tikzcd}$$

(that is, $r\cdot\Pi_s = \Pi\cdot\tau$), where

(a) $P_s(X_s, G_s)$ is a principal superfiber bundle with fiber G_s, base space X_s, and projection Π_s, whose body is the principal bundle $P(X, G)$.

(b) $P(X, G_s)$ is a principal superfiber bundle with fiber G_s, base X, and projection Π, whose body is also $P(X, G)$.

(c) τ is a superfiber bundle homomorphism: C^∞ and such that $\tau R_a = R_a\tau$ for all $a\in G_s$.

Since $P(X, G_s)$ is included into $P_s(X_s, G_s)$, via the global section t, every homomorphism τ induces a gauge transformation τ_0 through its restriction to $P(X, G_s)$. The composition $\tau_0^{-1}\cdot\tau$ is another homomorphism inducing the identity gauge transformation. We shall restrict ourselves, hereafter, to this class of homomorphisms.

The class of connections we shall consider in P_s will be the pullback, by the homomorphisms τ, of connections in $P(X, G_s)$. We shall study these connections through families of 1–forms defined on the respective base spaces. Let us take a local trivialization $\{(U_i, \sigma_i)\}$ in $P(X, G_s)$ and the local trivialization $\{(V_i, \sigma_i^s)\}$ in $P_s(X_s, G_s)$, where $\{U_i\}$ is an open covering of X, $V_i = r^{-1}(U_i)$, σ_i^s: $V_i \to \Pi_s^{-1}(V_i)$ is the preferred local section in $P_s(X_s, G_s)$, and $\sigma_i = \tau\cdot\sigma_i^s\cdot t$: $U_i \to \Pi^{-1}(U_i)$ a local section in $P(X, G_s)$. In this trivialization the homomorphism τ is determined by the family γ_i: $V_i \to G_s$ of G^∞-functions satisfying the compatibility relations $\gamma_j = (\psi_{ji}\cdot r)^{-1}\gamma_i\psi_{ij}^s$, where ψ_{ji} and ψ_{ji}^s are the transition functions in the fiber bundles $P(X, G_s)$ and $P_s(X_s, G_s)$, respectively, for the above defined trivializations. In particular, we have $\psi_{ij} = \psi_{ij}^s\cdot t$ and $\Psi_{ji} = \Psi_{ij}^{-1}$. Furthermore, it is easy to see that, in this trivialization, the homomorphism τ_0 given by $\gamma_{0i} = \gamma_i\cdot t$: $U_i \to G_s$ is the identity. In general, the existence of homomorphism τ is not *a priori* guaranteed for any couple of fiber bundles $P(X, G_s)$ and $P_s(X_s, G_s)$. Nevertheless, given the bundle $P(X, G_s)$ with transition functions ψ_{ij}, we can build a bundle $P_s(X_s, G_s)$ with transition functions $\psi_{ij}^s = \psi_{ij}\cdot r$ satisfying, trivially, the cocycle condition. The homomorphism τ which maps P_s into $P(X, G_s)$ should be given by the family of functions $\gamma_i \equiv e$. On the other hand, given two homomorphisms τ and τ' of $P_s(X_s, G_s)$ into $P(X, G_s)$, there exists a gauge transformation α in $P_s(X_s, G_s)$ such that $\tau' = \tau\cdot\alpha$, with α determined by the family $\alpha_i = \gamma_i^{-1}\cdot\gamma_i'$, where γ_i and γ_i' are the functions defining τ and τ', respectively, in a given trivialization.

A connection in $P(X, G_s)$ is determined by the family of 1-forms $\{\alpha_i\}$, defined on U_i, satisfying the compatibility relations $\alpha_j = \psi_{ji}\alpha_i\psi_{ij} + \psi_{ji}d\psi_{ij}$. The pullback by τ of this connection is expressed, likewise, by the family of 1-forms $\{\alpha_i^s\}$ defined on V_i, where

$$\alpha_i^s = \mathrm{ad}(\gamma_i^{-1})\epsilon^*\alpha_i + \gamma_i^{-1}d\gamma_i \tag{5}$$

with the corresponding compatibility relation.

The curvature corresponding to the connection in $P(X, G_s)$ is given by the family of 2-forms $\{R_i\}$ defined on U_i, satisfying the compatibility condition $R_j = \mathrm{ad}(\psi_{ij}^{-1})R_i$. On the other hand, the curvature corresponding to the connection in $P_s(X_s, G_s)$ is given by the family R_i^s of 2-forms on V_i, where

$$R_i^s = \mathrm{ad}(\gamma_i^{-1})r^*R_i. \tag{6}$$

Equation (6) shows clearly that the components of the curvature R_i^s with respect to any odd coordinate vanish identically due to the presence of the mapping r^*.

B. The pullback of a connection in *P(X, G)*

We shall start from a connection ω in $P(X, G)$ whose coefficients are, in a local coordinate system, the gauge fields $A_\mu(x)$ corresponding to the symmetry group G. Since our goal will be to build a structure with room for FP fields, we shall induce, from ω, a connection in $P(X, G_s)$ and then apply the mathematical scheme described throughout Sec. IIIA.

The connection in $P(X, G_s)$ induced by ω is given by $r^*\omega$. We can choose locally, with the aid of the inclusion t, the coordinate neighborhood in X_s, $V = U\times W$, where U is an open in R^4 and W an open in $S^{0,2}_{(I,I)}$. U is chosen to be a coordinate neighborhood in X and t: $U \to V$ is given by $t(x_\mu) = (x_\mu, 0, 0)$. Since $\mathscr{Y}\subset\mathscr{Y}_s$, both ω and $r^*\omega$ are expressed, in these coordinates, by

$$\alpha_i = A^i_\mu(x)dx^\mu. \tag{7}$$

(Hereafter we shall suppress, unless explicitly mentioned, the family index i from all connection and curvature forms.) Let us remark that, while the coefficients $A_\mu(x)$ of a general connection in $P(X, G_s)$ should be arbitrary, even elements of $\mathscr{Y}_s$ (the connection is an even form in the Grassmann algebra), the coefficients in (7) belong to $\mathscr{Y}$, so they are real.

In the local coordinate system we are using, the homomorphism τ is expressed by[17]

$$\gamma(x_\mu, \theta, \bar\theta) = \exp[\theta\bar c(x) + \bar\theta c(x) + \theta\bar\theta(B(x) + \tfrac12\{c(x), \bar c(x)\})], \tag{8}$$

and the connection ω_s is obtained by pulling back $r^*\omega$ by τ, as indicated in Eq. (5), so that it can be written as

$$\alpha_s = \phi_\mu(x, \theta, \bar\theta)dx^\mu + d\theta\,\bar\eta(x, \theta, \bar\theta) + d\bar\theta\,\eta(x, \theta, \bar\theta), \tag{9}$$

where $\phi_\mu(x, \theta, \bar\theta)$, $\eta(x, \theta, \bar\theta)$, and $\bar\eta(x, \theta, \theta, \bar\theta)$ are G^∞-functions which can be expanded as

$$\phi_\mu(x, \theta, \bar\theta) = A_\mu(x) + \theta D_\mu\bar c(x) + \bar\theta D_\mu c(x) + \theta\bar\theta(D_\mu B(x) + \{D_\mu c(x), \bar c(x)\}), \tag{10a}$$

$$\eta(x, \theta, \bar\theta) = c(x) - \theta(B(x) + \{c(x), \bar c(x)\}) - \tfrac12\bar\theta\{c(x), c(x)\} + \theta\bar\theta[\bar B(x), c(x)], \tag{10b}$$

$$\bar\eta(x, \theta, \bar\theta) = \bar c(x) - \tfrac12\theta\{\bar c(x), \bar c(x)\} + \bar\theta B(x) - \theta\bar\theta[B(x), \bar c(x)], \tag{10c}$$

with $D_\mu = \partial_\mu + [A_\mu(x),\ \ ]$, the usual covariant derivative, and $\bar B(x)$ a function of $B(x)$, $c(x)$, and $\bar c(x)$ given by the relation

$$B(x) + \bar B(x) + \{c(x), \bar c(x)\} = 0. \tag{11}$$

Let us note that ϕ_μ is a $\mathscr{Y}_s$-valued G^∞-function, where $\mathscr{Y}_s = T(G_s)_e$ is the Lie algebra of G_s and also coincides with the even part of the Lie superalgebra $T_d(G_s)_e^0$. On the other hand, η and $\bar\eta$ are $T_d(G_s)_e^1$-valued, the odd part of the Lie superalgebra. In short, the 1-form α_s is even.

Next, we shall discuss the null curvature conditions, introducing a more compact notation, which will be used later on. Let us denote by u_i the coordinates of V, with $u_i = x_i$ $(1 \leqslant i \leqslant 4)$, $u_5 = \theta$, and $u_6 = \bar{\theta}$. In this notation a connection in $P_s(X_s, G_s)$ is written, in V, as

$$\alpha_s = du_i\, \rho_i(u), \tag{12}$$

with $\rho_i(u)$ even for $1 \leqslant i \leqslant 4$ and odd for $i = 5,6$. The components of its curvature are[18]

$$F_{ij}(u) = \partial_i \rho_j - (-1)^{|\rho_i||\rho_j|} \partial_j \rho_i + [\rho_i, \rho_j]_{\mp} \tag{13}$$

where the symbol $[\ ,\ ]_-$ ($[\ ,\ ]_+$) means the commutator (anticommutator) which is used whenever ρ_i and/or ρ_j are even (ρ_i and ρ_j are odd).

It is easily proved that the curvature of α_s,[13] satisfies the condition $F_{ij} = 0$ whenever i or j are equal to 5 or 6. This condition characterizes the connections in $P_s(X_s, G_s)$ coming from a connection in $P(X, G_s)$ through homomorphism τ. To give this statement in a more precise way, we shall introduce the following concepts.

Definition 1: A local coordinate system in X_s is called t-trivial if

$$t(x_\mu) = (x_\mu, 0, 0).$$

Definition 2: A connection ω_s in $P_s(X_s, G_s)$ is called soul-flat if, for any t-trivial local coordinate system, $F_{ij} = 0$ for $\{i, j\} \cap \{5, 6\} \neq \emptyset$.

Next, we shall state a theorem to give a global characterization of soul-flat connections.

Theorem 1: A connection ω_s in $P_s(X_s, G_s)$ is soul-flat if and only if there is a connection ω in $P(X, G_s)$ and a homomorphism τ: $P_s(X_s, G_s) \to P(X, G_s)$ such that $\omega_s = \tau^* \omega$.

The proof of this theorem will be relegated to Appendix A.

Thus, every soul-flat connection is given, in a t-trivial local coordinate system, by Eqs. (10). It is therefore depending on seven independent fields, $A_\mu(x)$, $c(x)$, $\bar{c}(x)$, and $B(x)$, which are interpreted as the gauge potentials, ghost, antighost, and auxiliary fields, respectively, in agreement with the heuristic construction of Ref. 9. So it is possible to have room, inside a geometrical object, for the fundamental fields which appear in a quantized gauge field theory.

IV. BRS AND ANTI-BRS TRANSFORMATIONS

We shall analyze in this section, using the geometrical scheme built in Sec. III, the BRS and anti-BRS transformations.

The classical action corresponding to a gauge invariant theory quantized in the covariant gauge $\partial^\mu A_\mu = B$ is given by the expression

$$S = \int d^4x\, [-\tfrac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a - \bar{c}^a \partial^\mu D_\mu c_a - \tfrac{1}{2} B^a B_a + (\partial^\mu A^a_\mu) B_a]. \tag{14}$$

As it is well known, although this action is not gauge-invariant—since the gauge invariance has been broken by the quantization procedure (gauge fixing + FP term)—it is indeed invariant under BRS transformations[2] of the fields

$$\delta A_\mu(x) = \bar{\zeta} D_\mu c(x), \quad \delta c(x) = -\tfrac{1}{2}\bar{\zeta}\{c(x), c(x)\},$$
$$\delta \bar{c}(x) = \bar{\zeta} B(x), \quad \delta B(x) = 0. \tag{15}$$

Furthermore, the action (14) is invariant under the anti-BRS transformations[4,6] of the fields

$$\bar{\delta} A_\mu(x) = \zeta D_\mu \bar{c}(x), \quad \bar{\delta}\, \bar{c}(x) = -\tfrac{1}{2}\zeta \{\bar{c}(x), \bar{c}(x)\},$$
$$\bar{\delta} c(x) = \zeta \bar{B}(x), \quad \bar{\delta}\, \bar{B}(x) = 0, \tag{16}$$

where the parameters ζ and $\bar{\zeta}$ are constant odd elements of a Grassmann algebra. The shape of transformations (15) and (16) looks like gauge transformations of parameters $\bar{\zeta} c(x)$ and $\zeta \bar{c}(x)$, respectively. This property has been used to introduce them heuristically in the physical literature. In the following we shall see that BRS and anti-BRS transformations are represented in our formalism, acting over each connection, by true gauge transformations of parameters $\bar{\zeta}\eta(x, \theta, \bar{\theta})$ and $\zeta\bar{\eta}(x, \theta, \bar{\theta})$, the ghost and antighost superfields, respectively.

Let $\mathscr{C}$ be the set of connections in $P_s(X_s, G_s)$. We defined the transformations $T_{\bar{\zeta}}, \bar{T}_\zeta$: $\mathscr{C} \to \mathscr{C}$ by

$$T_{\bar{\zeta}}(\omega_s) = \gamma(\omega_s)[\omega_s], \quad \bar{T}_\zeta(\omega_s) = \bar{\gamma}(\omega_s)[\omega_s], \tag{17}$$

where $\gamma(\omega_s)$ and $\bar{\gamma}(\omega_s)$ are gauge transformations in P_s which, in a t-trivial local coordinate system, are expressed by

$$\gamma(\omega_s) = \exp[\bar{\zeta}\eta(x, \theta, \bar{\theta})], \quad \bar{\gamma}(\omega_s) = \exp[\zeta\bar{\eta}(x, \theta, \bar{\theta})]. \tag{18}$$

Next, we shall restrict the class of superfiber bundles $P_s(X_s, G_s)$ for which Eqs. (18) define true gauge transformations by the following theorem.

Theorem 2: Equations (18) define gauge transformations in $P_s(X_s, G_s)$ if and only if:

(i) The transition functions of the superfiber bundle $P_s(X_s, G_s)$ can be written as

$$\psi^s_{ij} = \psi_{ij} \cdot r \tag{19}$$

in some family of trivializations of $P_s(X_s, G_s)$.

(ii) The base space of the superfiber bundle $P_s(X_s, G_s)$ is

$$X_s = X \times S^{0,2}. \tag{20}$$

The proof of this theorm will be given in Appendix B.

Let us now verify that the transformations $T_{\bar{\zeta}}$ and $\bar{T}_\zeta$, introduced in Eq. (17), induce over the physical fields $A_\mu(x)$, $c(x)$, $\bar{c}(x)$, and $B(x)$ the transformation laws given by Eqs. (15) and (16). Let α_s be an arbitrary connection in P_s given by Eq. (9), where the superfields $\phi_\mu(x, \theta, \bar{\theta})$, $\eta(x, \theta, \bar{\theta})$, and $\bar{\eta}(x, \theta, \bar{\theta})$ are arbitrary G^∞-functions not necessarily given by (10). Using the transformation law of connections under gauge transformations,[13] we can write

$$T_{\bar{\zeta}}(\alpha_s) \equiv \alpha_s + \delta\alpha_s = \gamma^{-1}\alpha_s\gamma + \gamma^{-1}d\gamma,$$
$$\bar{T}_\zeta(\alpha_s) \equiv \alpha_s + \bar{\delta}\alpha_s = \bar{\gamma}^{-1}\alpha_s\bar{\gamma} + \bar{\gamma}^{-1}d\bar{\gamma} \tag{21}$$

or, in terms of the superfields,

$$\begin{aligned}
&\phi_\mu + \delta\phi_\mu = \gamma^{-1}\phi_\mu\gamma + \gamma^{-1}\partial_\mu\gamma, && \phi_\mu + \bar\delta\phi_\mu = \bar\gamma^{-1}\phi_\mu\bar\gamma + \bar\gamma^{-1}\partial_\mu\bar\gamma,\\
&\eta + \delta\eta = \gamma^{-1}\eta\gamma + \gamma^{-1}\partial_{\bar\theta}\gamma, && \eta + \bar\delta\eta = \bar\gamma^{-1}\eta\bar\gamma + \bar\gamma^{-1}\partial_{\bar\theta}\bar\gamma,\\
&\bar\eta + \delta\bar\eta = \gamma^{-1}\bar\eta\gamma + \gamma^{-1}\partial_\theta\gamma, && \bar\eta + \bar\delta\,\bar\eta = \bar\gamma^{-1}\bar\eta\,\bar\gamma + \bar\gamma^{-1}\partial_\theta\bar\gamma.
\end{aligned} \tag{22}$$

Equations (22) read, in the notation of Eq. (12), for the superfields $\rho_i(u)$, as

$$\begin{aligned}
&\delta\,\rho_i(u) = D_i\bar\zeta\,\eta,\\
&\bar\delta\,\rho_i(u) = D_i\zeta\,\bar\eta,
\end{aligned} \tag{23}$$

where

$$D_i = \partial_i + [\,\rho_i,\ \,] \tag{24}$$

is the covariant derivative along the direction u_i.

In particular, for soul-flat connections, that is, ϕ_μ, η, and $\bar\eta$ given by (10), we have for the covariant derivatives (23) the following results:

$$\begin{aligned}
&D_\mu\,\bar\zeta\eta = \bar\zeta\,\partial_{\bar\theta}\phi_\mu, && D_\mu\zeta\bar\eta = \zeta\partial_\theta\phi_\mu,\\
&D_{\bar\theta}\bar\zeta\eta = \bar\zeta\partial_{\bar\theta}\eta, && D_{\bar\theta}\zeta\bar\eta = \zeta\partial_\theta\eta,\\
&D_\theta\bar\zeta\eta = \bar\zeta\partial_{\bar\theta}\bar\eta, && D_\theta\zeta\bar\eta = \zeta\partial_\theta\bar\eta.
\end{aligned} \tag{25}$$

These equations lead, for the physical components A_μ, c, $\bar c$, and B of the superfields, to the variations given by Eqs. (15) and (16). It is thus proven that the transformations $T_{\bar\zeta}$ and $\bar T_\zeta$ defined in (17), restricted to soul-flat connections, induce the BRS and anti-BRS transformations, respectively.

Equations (23) and (25) allow us to recover Bonora and Tonin's interpretation[9] for BRS and anti-BRS transformations as translations along the directions θ and $\bar\theta$ of the superspace X_s. Although (25) seems to correspond to infinitesimal translations, they are really finite translations of parameters ζ and $\bar\zeta$ (this is due to the fact that $\zeta^2 = \bar\zeta^2 = 0$), which can be expressed as

$$\begin{aligned}
&T_{\bar\zeta}\,\rho_i(x,\theta,\bar\theta)\\
&\quad = \rho_i(x,\theta,\bar\theta+\bar\zeta) = \rho_i(x,\theta,\bar\theta) + \bar\zeta\,\partial_{\bar\theta}\,\rho_i(x,\theta,\bar\theta)\\
&\quad \equiv \rho_i(x,\theta,\bar\theta) + \delta\,\rho_i(x,\theta,\bar\theta),\\
&\bar T_\zeta\,\rho_i(x,\theta,\bar\theta)\\
&\quad = \rho_i(x,\theta+\zeta,\bar\theta) = \rho_i(x,\theta,\bar\theta) + \zeta\partial_\theta\,\rho_i(x,\theta,\bar\theta)\\
&\quad \equiv \rho_i(x,\theta,\bar\theta) + \bar\delta\,\rho_i(x,\theta,\bar\theta).
\end{aligned} \tag{26}$$

From (26) it is straightforward to prove that the following product law holds:

$$\begin{aligned}
&T_{\bar\zeta_1}T_{\bar\zeta_2} = T_{\bar\zeta_1+\bar\zeta_2},\\
&\bar T_{\zeta_1}\bar T_{\zeta_2} = \bar T_{\zeta_1+\zeta_2},\\
&\bar T_\zeta T_{\bar\zeta} = T_{\bar\zeta}\bar T_\zeta.
\end{aligned} \tag{27}$$

Defining

$$T_{(\zeta,\bar\zeta)} \equiv \bar T_\zeta T_{\bar\zeta}, \quad T_{(0,\bar\zeta)} \equiv T_{\bar\zeta}, \quad T_{(\zeta,0)} \equiv \bar T_\zeta, \tag{28}$$

we obtain that the transformation $T_{(\zeta,\bar\zeta)}$ are a representation of the additive supergroup $S^{0,2}$, with the product law

$$T_{(\zeta_1,\bar\zeta_1)}T_{(\zeta_2,\bar\zeta_2)} = T_{(\zeta_1+\zeta_2,\bar\zeta_1+\bar\zeta_2)}. \tag{29}$$

Thus, we have made clear from the above construction that, for soul-flat connections, the BRS and anti-BRS transformations generate a global (nonlocal) supergroup with two anticommuting parameters. However, acting over each connection, they coincide with a particular gauge transformation of parameters $\zeta\bar\eta$ and $\bar\zeta\eta$. The last discussion cannot be extended to general connections in P_s since (25), and hence (27), do not hold.

V. CONCLUSION

In this work we have found that a suitable structure for giving a geometrical interpretation to quantized gauge theories is that of the principal superfiber bundle. Let us remark that every superfiber bundle is also a fiber bundle in such a way that the structure of the principal fiber bundle is valuable not only for the unquantized, but also for the quantized gauge theory. Yang–Mills and Faddeev–Popov fields appear as superfields which are coefficients of a particular kind of connections in the superfiber bundle—soul-flat connections—coming from the usual connections in the principal fiber bundle describing the unquantized theory. The BRS and anti-BRS transformations are, acting over a given connection, true gauge transformations, in the superfiber bundle, of parameters, the ghost and antighost superfields, respectively, which appear in the particular connection on which they act. For the case of soul-flat connections the BRS and anti-BRS transformations are also translations along the anticommuting variables which act in the base supermanifold. Moreover, they generate an $S^{0,2}$-supergroup.

In short, we see that Bonora and Tonin's heuristic results[9] can be geometrically interpreted within the framework of the theory of principal fiber bundles by means of the concepts of supermanifold, supergroup, and superfiber bundle lately developed in the literature.[11–13]

We leave as an open problem, in this work the geometrical interpretation of the Lagrangian of quantized gauge theories, as some invariant of the supergroup $S^{0,2}$, as well as its relation with the geometrical interpretation given to the Lagrangian of the unquantized gauge theory.

APPENDIX A

In this appendix we shall prove Theorem 1.

(i) A connection in P_s, in t-trivial coordinates, can be expressed as

$$\omega_s = dx^\mu\,\phi_\mu(x,\theta,\bar\theta) + d\bar\theta\,\eta(x,\theta,\bar\theta) + d\theta\,\bar\eta(x,\theta,\bar\theta), \tag{A1}$$

where

$$\begin{aligned}
&\phi_\mu(x,\theta,\bar\theta) = A_\mu(x) + \theta\bar R_\mu(x) + \bar\theta R_\mu(x) + \theta\bar\theta S_\mu(x),\\
&\eta(x,\theta,\bar\theta) = c(x) + \theta\bar B(x) + \bar\theta r(x) + \theta\bar\theta s(x),\\
&\bar\eta(x,\theta,\bar\theta) = \bar c(x) + \theta\bar r(x) + \bar\theta B(x) + \theta\bar\theta\,\bar s(x).
\end{aligned} \tag{A2}$$

The connection ω_s is soul-flat if and only if the following relations hold:

$$
\begin{aligned}
R_\mu(x) &= D_\mu c(x), \quad r(x) = -\tfrac{1}{2}\{c(x), c(x)\}, \quad s(x) = [\bar{B}(x), c(x)],\\
\bar{R}_\mu(x) &= D_\mu \bar{c}(x), \quad \bar{r}(x) = -\tfrac{1}{2}\{\bar{c}(x), \bar{c}(x)\}, \quad \bar{s}(x) = -[B(x), \bar{c}(x)],\\
S^\mu(x) &= D^\mu B(x) + \{R^\mu(x), \bar{c}\}.
\end{aligned}
\tag{A3}
$$

The proof is straightforward,[9] using the components of the curvature given by (13). Thus ω_s only depends on the fields $A_\mu(x)$, $c(x)$, $\bar{c}(x)$, and $B(x)$.

(ii) If $\omega_s = \tau^*\omega$, from (6) we immediately deduce that ω_s is soul-flat.

(iii) Conversely, let ω_s be soul-flat and let $(\phi^i_\mu, \eta^i, \bar{\eta}^i)$ and $(\phi^j_\mu, \eta^j, \bar{\eta}^j)$ be the components of ω_s with respect to the same system of t-trivial coordinates but for two different trivializations of P_s whose transition function ψ^s_{ij} is given by

$$\psi^s_{ij}(x, \theta, \bar{\theta}) = e^{a(x)} e^{\theta \bar{b}(x) + \bar{\theta} b(x) + \theta\bar{\theta}(d(x) + \{b(x), \bar{b}(x)\}/2)}, \tag{A4}$$

and let the transition function in $P(X, G_s)$, ψ_{ij}, be given by

$$\psi_{ij} = \psi^s_{ij} \cdot t = e^{a(x)}. \tag{A5}$$

Let us define the 1-form ω in $P(X, G_s)$ by

$$\omega_k = A^k_\mu \, dx^\mu \quad (k = i, j) \tag{A6}$$

and τ by

$$\tau_k = e^{\bar{\theta} c_k(x) + \theta \bar{c}_k(x) + \theta\bar{\theta}(B_k(x) + \{c_k(x), \bar{c}_k(x)\}/2)}. \tag{A7}$$

Then, if ω is a connection in $P(X, G_s)$ and τ a homomorphism, the following compatibility conditions must hold:

$$
\begin{aligned}
\omega_j &= \mathrm{ad}(\psi_{ij}^{-1})\omega_i + \psi_{ij}^{-1}\, d\psi_{ij},\\
\tau_j &= \psi_{ij}^{-1}\, \tau_i\, \psi^s_{ij}.
\end{aligned}
\tag{A8}
$$

Since ω_s is a connection in $P_s(X_s, G_s)$, the following compatibility conditions must also hold:

$$\omega_{sj} = \mathrm{ad}(\psi^{s^{-1}}_{ij})\omega_{si} + \psi^{s^{-1}}_{ij}\, d\psi^s_{ij}. \tag{A9}$$

Thus, if ω_s is soul-flat, by (i) and (A9) we deduce that

$$
\begin{aligned}
A^j_\mu &= e^{-a} A^i_\mu e^a + e^{-a}\partial_\mu e^a, \quad c^j = e^{-a} c^i e^a + b,\\
\bar{c}^j &= e^{-a}\bar{c}^i e^a + \bar{b},\\
B^j &= e^{-a} B^i e^a + d + [e^{-a}\bar{c}^i e^a, b].
\end{aligned}
\tag{A10}
$$

It can be straightforwardly proven that Eqs. (A10) are equivalent to (A8), so that ω is a connection in $P(X, G_s)$ and τ is a homomorphism. Q.E.D.

APPENDIX B

In this appendix we shall prove Theorem 2 for the case of BRS transformations $\gamma(\omega_s) = \exp[\bar{\zeta}\eta(x, \theta, \bar{\theta})]$. The proof for anti-BRS transformations follows along identical lines.

(i) Let $\gamma(\omega_s)$, given by (18), be a gauge transformation, so that the compatibility condition

$$\psi^{s^{-1}}_{ij}\, e^{\bar{\zeta}\eta_i}\, \psi^s_{ij} = e^{\bar{\zeta}\eta_j} \tag{B1}$$

holds. On the other hand, since η_i and η_j are components of the connection ω_s, they must verify the compatibility condition

$$\eta_j = \mathrm{ad}(\psi^{s^{-1}}_{ij})\eta_i + \psi^{s^{-1}}_{ij}\, \partial_{\bar{\theta}}\psi^s_{ij}. \tag{B2}$$

Expanding (B1) and using (B2), we get

$$b - \theta(d + \tfrac{1}{2}[b, \bar{b}]) = 0, \quad \forall\, \theta. \tag{B3}$$

From (B3) we immediately deduce $b = d = 0$, and thus Eq. (19) holds. Q.E.D.

(ii) It is a straightforward calculation to verify that, under a change of coordinates in X_s, $(x, \theta, \bar{\theta}) \to (x', \theta', \bar{\theta}')$, the BRS transformation $e^{\bar{\zeta}\eta} \to e^{\bar{\zeta}a_1\eta'} e^{\bar{\zeta}a_2\bar{\eta}'}$, where a_1 and a_2 are even functions of the new coordinates $(x', \theta', \bar{\theta}')$. Only if the change of coordinates is $\theta' = \theta$, $\bar{\theta}' = \bar{\theta}$, do we find $e^{\bar{\zeta}\eta} \to e^{\bar{\zeta}\eta'}$. In this way, if we want to define BRS transformations as gauge transformations, we need an atlas for X_s, where all the changes of coordinates have the form

$$x'_\mu = x'_\mu(x), \quad \theta' = \theta, \quad \bar{\theta}' = \bar{\theta}. \tag{B4}$$

But Eq. (B4) is equivalent to Eq. (20). Q.E.D.

[1] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).

[2] C. Becchi, A. Rouet, and R. Stora, Commun. Math. Phys. **42**, 127 (1975).

[3] J. Thierry-Mieg, Nuovo Cimento A **56**, 396 (1980); J. Math. Phys. **21**, 2834 (1981).

[4] M. Quirós, F. J. de Urríes, J. Hoyos, M. L. Mazón, and E. Rodríguez, J. Math. Phys. **22**, 1767 (1981).

[5] G. Curci and R. Ferrari, Phys. Lett. B. **63**, 91 (1976).

[6] I. Ojima, Prog. Theor. Phys. **64**, 625 (1980).

[7] This remark was recently pointed out in J. M. Leinaas and K. Olaussen, "Ghosts and Geometry," CERN-TH 3061, 10 April 1981 and Phys. Lett. (to be published).

[8] S. Ferrara, O. Piguet, and M. Schweda, Nucl. Phys. B **119**, 493 (1977).

[9] L. Bonora and M. Tonin, Phys. Lett. B. **98**, 48 (1981).

[10] L. Bonora, P. Pasti, and M. Tonin, Nuovo Cimento A **63**, 353 (1981).

[11] A. Rogers, J. Math. Phys. **21**, 1352 (1980).

[12] A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).

[13] J. Hoyos, M. Quirós, J. Ramírez Mittelbrunn, and F. J. de Urríes, "Differential Supergeometry" (in preparation).

[14] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977); V. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 709 (1978).

[15] See, for instance, L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).

[16] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.

[17] A word of explanation of our notation is in order here. The overbar does not refer to any kind of complex adjunction since we are working with real Grassmann algebras. Instead, it is related to the ghost number G: If A is any quantity with ghost number equal to N, $G(A) = N$, $\bar{A}$ will be referred to as an independent quantity such that $G(\bar{A}) = -N$. In particular, θ and $\bar{\theta}$ indicate independent coordinates with $G(\theta) = 1$, $G(\bar{\theta}) = -1$. The same can be said about the component fields (A2), R_μ, c, r, s, B and their partners $\bar{R}_\mu$, $\bar{c}$, $\bar{r}$, $\bar{s}$, $\bar{B}$, where $G(R_\mu) = G(c) = G(s) = 1$, $G(r) = 2$, $G(B) = 0$, while, $G(\bar{R}_\mu) = G(\bar{c}) = G(\bar{s}) = -1$, $G(\bar{r}) = -2$, and $G(\bar{B}) = 0$.

[18] The use of anticommutators for anticommuting coefficients in (13) needs some clarification. Although the connection and curvature forms, α_s and Ω_s, are even, their coefficients ρ_i and F_{ij} are even or odd, depending on the grade of u_i and u_j. In particular, if $A = A^a t_a$, $B = B^a t_a$ (t_a being the matrices of the Lie algebra G), then the Lie bracket of A and B is the anticommutator (instead of the commutator, as usually), i.e., $\{A, B\}^a = f^a_{bc} A^b B^c$, where f^a_{bc} are the structure constants of the Lie algebra G.

Relativistic field theories in three dimensions

Birne Binegar[a)]
Department of Physics, University of California, Los Angeles, California 90024

(Received 10 March 1981; accepted for publication 22 May 1981)

We provide a complete classification of the unitary irreducible representations of the (2 + 1)-dimensional Poincaré group. We show, in particular, that only two types of "spin" are available for massless field theories. We also construct generalized Foldy–Wouthuysen transformations which connect the physical UIR's with covariant field theories in three dimensions.

PACS numbers: 11.10.Qr, 11.30.Cp, 02.20. + b

I. INTRODUCTION

The complexities of field theory in four dimensions have very often forced theorists to test models in unphysical spaces of lower dimension. To our knowledge, however, the field theoretical foundations of such models have been obtained only by "projection" from the physical dimension. Such projections, *ipso facto*, cannot reveal the subtleties inherent to a particular dimension. Certainly omniscience, which is now so often demanded of a theorist, requires the independent development of theories in their native dimension.

Because of its proximity to "reality," three-dimensional field theory deserves particular scrutiny. In this paper we shall examine the group theoretical foundations of any relativistic field theory in 2 + 1 dimensions.[1] Such a study reveals a theory which mimics very well the structure of its physical counterpart, and yet it also possesses some rather startling simplifications.

For instance, we shall show that there are only two types of massless particles in three dimensions. This circumstance implies that the pathology of particles with high spin may be avoided, and yet it still allows one to make a distinction between tensor and spinor fields. It is even more surprising to see how three-dimensional gravity accommodates this situation.

The recent work of Flato and Fronsdal implies that this study of three-dimensional field theory may also have direct physical significance.[2] These authors have demonstrated the equivalence between a theory of massless particles in a four-dimensional de Sitter space and a field theory of two (and only two) interacting fields on the three-dimensional hypersurface at spatial infinity. Indeed, the two massless representations we found must correspond to these primordial fields.

The work itself is divided into three parts. In the first we shall examine the underlying group of three-dimensional space–time. We next provide a complete classification of the unitary irreducible representations of this group and relate these UIR's to the subject of elementary particles. We then connect the physical UIR's with their respective field theories by means of generalized Foldy–Wouthuysen transformations.

II. THE GROUP

The three-dimensional Poincaré group π is defined as the group of real transformations

$$(a,\Lambda):\ x^\mu \to \Lambda_\nu{}^\mu x^\nu + a^\mu \tag{1}$$

in a 2 + 1 pseudo-Euclidean space which leave

$$|x-y|^2 = (x^0-y^0)^2 - (x'-y')^2 - (x^2-y^2)^2 \tag{2}$$

invariant. Applying two successive Poincaré transformations on x, we find

$$(a',\Lambda')(a,\Lambda) = (a' + \Lambda' a, \Lambda'\Lambda), \tag{3}$$

which indicates that π is, in fact, a semidirect product of the (2 + 1)-dimensional translation and Lorentz groups:

$$\pi = N \circledS L. \tag{4}$$

The Lorentz subgroup L is the group of transformations $x \to \Lambda x$ leaving x^2 unchanged. Every such transformation falls into one of four disjoint sets:

$$\begin{aligned}
L_+{}^\uparrow &= \{\Lambda \in L;\ \det\Lambda = +1,\ \Lambda_0{}^0 > 0\},\\
L_-{}^\uparrow &= \{\Lambda \in L;\ \det\Lambda = -1,\ \Lambda_0{}^0 > 0\},\\
L_+{}^\downarrow &= \{\Lambda \in L;\ \det\Lambda = +1,\ \Lambda_0{}^0 < 0\},\\
L_-{}^\downarrow &= \{\Lambda \in L;\ \det\Lambda = -1,\ \Lambda_0{}^0 < 0\}.
\end{aligned} \tag{5}$$

We shall henceforth restrict our attention to $L_+{}^\uparrow$ (and $\pi_+{}^\uparrow = N \circledS L_+{}^\uparrow$), remarking only that it is the largest connected subgroup of L (the other subsets $L_-{}^\uparrow$, $L_+{}^\downarrow$, and $L_-{}^\downarrow$ are not connected to the identity).

We now relate $L_+{}^\uparrow$ to the group SL(2,R) of real unimodular 2×2 matrices. Consider the space $\mathscr{M}$ of real Hermitian matrices with basis $\{\tau_\alpha\}$:

$$\tau_0 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad \tau_1 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad \tau_2 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}. \tag{6}$$

With each vector x in our (2 + 1)-dimensional space M, we can associate a matrix $\chi \in \mathscr{M}$ by the map

$$T:\ M \to \mathscr{M} \to:\ x \to \chi = x^\mu \tau_\mu = \begin{pmatrix} x^0 + x' & x^2\\ x^2 & x^0 - x'\end{pmatrix} \tag{7}$$

or vice versa

$$T^{-1}:\ \mathscr{M} \to M:\ \chi \to x,\quad x^\mu = \tfrac{1}{2}\mathrm{Tr}\,(\chi\tau_\mu). \tag{8}$$

As is almost obvious, the map T provides a homeomorphism between M and $\mathscr{M}$. Further, noting that

$$\det\chi = (x^0)^2 = (x')^2 - (x^2)^2 = x^2, \tag{9}$$

we infer that a homomorphism should exist between $L_+{}^\uparrow$ and any group of automorphisms of $\mathscr{M}$ which leaves $\det\chi$ invariant. Thus, defining the action

$$\mathrm{SL}\,(2,R)\quad \Omega:\ \chi \to \chi' = \Omega\chi\Omega^T \tag{10}$$

and checking

[a)]This work is supported in part by The National Science Foundation under Grant No. PHY 78-21502.

$$\det\chi' = (\det\Omega)(\det\chi)(\det\Omega^T) = \det\chi, \tag{11}$$

we conclude

$$L_+{}^\uparrow \sim \mathrm{SL}(2,R).$$

In fact, denoting the "ineffective" subgroup $\{I, -I\}$ of $\mathrm{SL}(2,R)$ by Z, one may verify

$$L_+{}^\uparrow \approx \mathrm{SL}(2,R)/Z \tag{12}$$

and so $\mathrm{SL}(2,R)$ is a double covering group of $L_+{}^\uparrow$.

The translation subgroup N is an additive vector group and so every unitary irreducible representation (UIR) of N is of the form

$$N \quad a \to \langle a,p\rangle = \exp(ip\cdot a). \tag{13}$$

It follows that we can characterize the equivalence classes of UIR's of N by elements p of the vector space dual to N. We can also extend the adjoint action of $L_+{}^\uparrow$ to the dual space $\hat{N}$ by

$$\langle a,p\rangle = \exp(ip_\mu \Lambda^\mu{}_\nu a^\nu) = \langle a, \Lambda^{-1}p\rangle. \tag{14}$$

Let us associate with each point $\tilde{p}\in\hat{N}$ a subset of $\hat{N}$

$$\hat{O}_{\tilde{p}} = \{\Lambda\tilde{p};\ \Lambda\in L_+{}^\uparrow\}, \tag{15}$$

which is appropriately called the "orbit" of $\tilde{p}$ in $\hat{N}$. Noting that the elements of $L_+{}^\uparrow$ can neither change the value of p^2 nor the sign of p_0 if $p^2 \geqslant 0$, we see that we have six classes of orbits:

$$\begin{aligned}
&\hat{O}_m{}^+ = \{p\in N;\ p^2 = m^2,\ p_0 > 0\},\\
&\hat{O}_m{}^- = \{p\in N;\ p^2 = m^2,\ p_0 < 0\},\\
&\hat{O}_0{}^+ = \{p\in N;\ p^2 = 0,\ p_0 > 0\},\\
&\hat{O}_0{}^- = \{p\in N:\ p^2 = 0,\ p_0 < 0\},\\
&\hat{O}_{im} = \{p\in N;\ p^2 = -m^2\},\\
&\hat{O}_0 = p = (0,0,0).
\end{aligned} \tag{16}$$

As the notation suggests (and as our intuition demands), the elements p correspond to momenta (as the infinitesimal generators of translations) while the orbits correspond to the positive energy mass shell, the negative energy mass shell, the forward light cone, etc.

III. IRREDUCIBLE UNITARY REPRESENTATIONS OF $\pi_+{}^\uparrow$

The quantum mechanical description of an elementary particle inevitably entails a linear vector space of physical states wherein a positive definite inner product is defined. Invariance of probability amplitudes requires that the symmetry operators of such an elementary system (of physical states) be represented by unitary operators.[3] Thus the states themselves, or the wavefunctions which prescribe them, should carry a unitary representation of the symmetry group. Further, corresponding to the physical notion of an elementary particle as an object indivisible, is the requirement that this unitary representation be also irreducible. Therefore, in order to determine the "particle content" of three-dimensional field theories, it behooves us to first consider the UIR's of the three-dimensional Poincaré group.

To identify all the UIR's of $\pi_+{}^\uparrow$, we rely on the following mathematical result due to Mackey.[4] Every UIR of a group which is of a certain class of semidirect product (to which $\pi_+{}^\uparrow$ belongs) is induced by a UIR of a stability group associated with an orbit. The identification of the UIR's of $\pi_+{}^\uparrow$ then proceeds as follows:

1. We first classify all orbits $\hat{O}$ in the dual space of N.
2. From each orbit we select an element $\hat{p}$ and then determine the orbits associated stability group $S_{\hat{O}}$.
3. We identify all the UIR's of each stability group $S_{\hat{O}}$. By Mackey's result, this then amounts to a complete classification of the UIR's of $\pi_+{}^\uparrow$.
4. With each p in each orbit $\hat{O}$ we associate an element $\Omega(p)\in\mathrm{SL}(2,R)$ corresponding to a Lorentz transformation which takes p to the stability point $\hat{p}$.
5. Finally, from each UIR $D[\ ]$ of each stability group $S_{\hat{O}}$ we form the induced UIR of $\pi_+{}^\uparrow$ given by

$$\begin{aligned}U_D{}^{\hat{O}}(a,\Lambda)u(p) = \exp(ip\cdot a)D\,[\Omega^{-1}(p)\Omega_\Lambda\Omega(\Lambda^{-1}p)]\\ \times u(\Lambda^{-1}p),\end{aligned} \tag{17}$$

where Ω_Λ is an element of $\mathrm{SL}(2,R)$ corresponding to the element $\Lambda\in L_+{}^\uparrow$ and $\Omega^{-1}(p)\Omega_\Lambda\Omega(\Lambda^{-1}p)\in\mathrm{SL}(2,R)$ is to be interpreted as its projection on the stability subgroup $S_{\hat{O}}$.

The first step of this program was carried out at the end of the previous section. We shall now proceed to complete the classification of the UIR's of $\pi_+{}^\uparrow$ orbit by orbit.[5]

$\hat{O}_m{}^+$

In order to discover the associated stability subgroup, we choose as the stability point $\hat{p} = (m,0,0)$ and note that the stability subgroup must be homomorphic to the subgroup of $\mathrm{SL}(2,R)$ for which

$$\Omega p\cdot\tau\Omega^T = \Omega\begin{pmatrix} m & 0\\ 0 & m\end{pmatrix}\Omega^T = \begin{pmatrix} m & 0\\ 0 & m\end{pmatrix}. \tag{18}$$

This condition implies

$$\Omega^T = \Omega^{-1} \tag{19}$$

and so the stability subgroup associated with $\hat{O}_m{}^+$ is homomorphic to O(2), the group of real orthogonal 2×2 matrices. The matrices $\Omega\in\mathrm{O}(2)$ may be parametrized by a single variable θ

$$\Omega_\theta = \begin{pmatrix} \cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}. \tag{20}$$

Since the group O(2) is not simply connected (in fact, it is infinitely connected), we jump to its universal covering group R, the additive group of real numbers, in order to include all the multivalued representations of O(2).
The UIR's of R are of the form

$$D^j:\ \theta \to e^{ij\theta},\quad j\in R, \tag{21}$$

and so we can label the UIR's of $\pi_+{}^\uparrow$ associated with $\hat{O}_m{}^+$ by

$$U^{m,+,j},\quad j\in R.$$

$\hat{O}_m$

If we let $\hat{p} = (-m,0,0)$, we find, in the same manner as above, that the stability subgroup for $\hat{O}_m{}^-$ is O(2), and so we can label the UIR's associated with this orbit by

$U^{m,-,j}\quad j\in R.$

$\hat{O}_0{}^+$

In this case we take $\hat{p}=(\frac{1}{2},\frac{1}{2},0)$ and then infer that

$$\Omega\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\Omega^T=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\quad \Omega\in S_{\hat{O}_0{}^+}. \tag{22}$$

This implies that the matrices Ω have the form

$$\Omega=\begin{pmatrix}\pm 1 & a\\ 0 & \pm 1\end{pmatrix},\quad a\in R. \tag{23}$$

Let us parametrize $S_{\hat{O}_0{}^+}$ by

$$(a,\pm 1)=\pm\begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix},\quad a\in R. \tag{24}$$

We see that

$$S_{\hat{O}_0{}^+}\sim Z\otimes R, \tag{25}$$

where Z is the multiplicative group consisting of two elements $\{1,-1\}$ and R is again the additive group of real numbers. Z has just two UIR's (up to equivalence); one is the trivial representation

$$D^0(\pm 1)=1 \tag{26}$$

and the other is

$$D^1(\pm 1)=\pm 1. \tag{27}$$

Recalling the UIR's of R, we conclude that the UIR's of $Z\otimes R$ are of the form

$$D^{\epsilon,t}=D_Z{}^{\epsilon}\otimes D_R{}^t,\quad \epsilon=0,1,\ t\in R, \tag{28}$$

and so we can label the induced UIR's of $\pi_+{}^\uparrow$ by

$U^{0,+,\epsilon,0}\quad \epsilon=0,1,\quad t=0,$

$U^{0,+,\epsilon,t}\quad \epsilon=0,1,\quad t\in R-\{0\}.$

$\hat{O}_0{}^-$

If we let $\hat{p}=(-\frac{1}{2},-\frac{1}{2},0)$, we find, as above, that the stability subgroup of $\hat{O}_0{}^-$ is also $Z\otimes R$; hence we can label the UIR's associated with O by

$U^{0,-,\epsilon,0}\quad \epsilon=0,1,\quad t=0,$

$U^{0,-,\epsilon,t}\quad \epsilon=0,1,\quad t\in R-\{0\}.$

$\hat{O}_{im}$

Here we choose $\hat{p}=(0,m,0)$ and then look for all $\Omega\in SL(2,R)$ satisfying

$$\Omega\begin{pmatrix}m & 0\\ 0 & -m\end{pmatrix}\Omega^T=\begin{pmatrix}m & 0\\ 0 & -m\end{pmatrix}. \tag{29}$$

This implies that the $\Omega\in S_{\hat{O}_{im}}$ are of the form

$$\Omega=\begin{pmatrix}a & 0\\ 0 & a^{-1}\end{pmatrix},\quad a\in R-\{0\} \tag{30}$$

or that

$$S_{\hat{O}_{im}}\sim Z\otimes R^+, \tag{31}$$

where R^+ is the multiplicative group of positive real numbers. Noting that R^+ is isomorphic to R, we infer that

$$S_{\hat{O}_{im}}\sim Z\otimes R \tag{32}$$

and so we can label the UIR's associated with $\hat{O}_{im}$ by

$U^{im,\epsilon,0}\quad \epsilon=0,1,\quad t=0$

$U^{im,\epsilon,t}\quad \epsilon=0,1,\quad t\in R-\{0\}.$

$\hat{O}_0{}^0$

This orbit consists of a single point $\hat{p}=(0,0,0)$ and its stability subgroup is the entire group SL(2,R). SL(2,R) has three series of UIR's conventionally labeled by: $D^{i\sigma,\epsilon}$, $\sigma\in R$, $\epsilon=0,1$; D^n, $n=0,1$; and D^ρ, $0<|\rho|<1$.[6] We can therefore associate with $\hat{O}_0{}^0$ three UIR's of $\pi_+{}^\uparrow$:

$U^{0,0,i\sigma,t}\quad \sigma\in R;\quad \epsilon=0,1$

$U^{0,0,n}\quad n=0,1,$

$U^{0,0,\rho}\quad 0<|\rho|<1.$

In summary we have the following complete classification of the unitary irreducible representations of $\pi_+{}^\uparrow$:

$$\begin{array}{ll}
U^{m,+,j} & m>0,\quad p_0>0,\quad j\in R,\\
U^{m,-,j} & m>0,\quad p_0<0,\quad j\in R,\\
U^{0,+,\epsilon,0} & m=0,\quad p_0>0, \epsilon=0,1,\quad t=0,\\
U^{0,+,\epsilon,t} & m=0,\quad p_0>0,\quad \epsilon=0,1,\quad t\in R-\{0\},\\
U^{0,-,\epsilon,t} & m=0,\quad p<0,\quad \epsilon=0,1,\quad t\in R-\{0\},\\
U^{im,\epsilon,0} & p^2=-m^2,\quad \epsilon=0,1,\quad t=0,\\
U^{im,\epsilon,t} & p^2=-m^2,\quad \epsilon=0,1,\quad t\in R-\{0\},\\
U^{0,0,\sigma,\epsilon} & p=0,\quad \sigma\in R,\quad \epsilon=0,1,\\
U^{0,0,n} & p=0,\quad n=0,1,\\
U^{0,0,\rho} & p=0,\quad 0<|\rho|<1.
\end{array} \tag{33}$$

Note the three classes of positive energy representations: $U^{m,+,j}$, $U^{0,+,0,0}$, and $U^{0,+,1,0}$. These correspond to, respectively, massive particles with "spin" j, massless particles with discrete "spin" (of which there are two types), and massless particles with continuous "spin." We shall reject the massless continuous spin representations as "unphysical" since we suspect that, like their four-dimensional counterparts, they cannot be connected with a local field theory (except, perhaps, in the sense of Iverson and Mack[7]). In the final section we shall return to this allegation. We now give an explicit realization of the physical UIR's.

$U^{m,+,j}$

We again choose $\hat{p}=(m,0,0)$ and $\Omega(p)$ to correspond to the Lorentz rotation in the p–$\hat{p}$ plane which takes p to $\hat{p}$. If $R(\theta)\in SL(2,R)$ corresponds to a spatial rotation by an infinitesimal angle θ, one finds that

$$\Omega^{-1}(p)R(\theta)\Omega(R^{-1}(\theta)\,p)=R(\theta) \tag{34}$$

while if $L(\boldsymbol{\theta})$ corresponds to an infinitesimal Lorentz boost in the $\boldsymbol{\theta}$ direction, then

$$\Omega^{-1}(p)L(\boldsymbol{\theta})\Omega(L^{-1}(\theta)\,p)=R\left(\frac{\theta\wedge\mathbf{p}}{E+m}\right), \tag{35}$$

where $\boldsymbol{\theta}\wedge\mathbf{p}=\theta_1p_2-\theta_2p_1$. Thus the infinitesimal operators of rotations and Lorentz boosts are given by

$$\hat{R}^{m,+,j}(\theta)u(p)=D^j(\theta)u(R^{-1}(\theta)p)=(1+ij\theta-\theta\mathbf{p}\wedge\boldsymbol{\partial})u(p) \tag{36}$$

$$L^{m,+,j}(\boldsymbol{\theta})u(p) = D_j\left(\frac{\boldsymbol{\theta}\wedge\mathbf{p}}{E+m}\right)u(L^{-1}(\boldsymbol{\theta})p)$$
$$= \left(1 + ij\left(\frac{\boldsymbol{\theta}\wedge\mathbf{p}}{E+m}\right) - E\mathbf{p}\cdot\boldsymbol{\partial}\right)u(p). \quad (37)$$

$U^{0,+,0,0}$

From Eqs. (17) and (26) we see that the wavefunctions belonging to $U^{0,+,0,0}$ are scalar fields (since $U^{0,+,0,0}$ is induced by a trivial representation of the stability subgroup). Therefore, the operators associated with infinitesimal rotations $R(\theta)$ and infinitesimal Lorentz boosts $L(\boldsymbol{\theta})$ are

$$\hat{R}^{0,+,0,0}(\theta) = (1 - \theta\mathbf{p}\wedge\boldsymbol{\partial}), \quad (38)$$
$$\hat{L}^{0,+,0,0}(\boldsymbol{\theta}) = (1 - E\,\boldsymbol{\theta}\cdot\boldsymbol{\partial}). \quad (39)$$

$U^{0,+,1,0}$

The wavefunctions of this representation are "almost" scalar fields in that the infinitesimal operators $\hat{R}^{0,+,1,0}(\theta)$ and $\hat{L}^{0,+,1,0}(\boldsymbol{\theta})$ are identical to (38) and (39). However, under a finite rotation $R(\alpha)$ one finds that

$$D_Z[\Omega^{-1}(p)R(\alpha)\Omega(R^{-1}(\alpha)p)] = \begin{cases} D_Z[1] & \text{if } 0\leqslant\alpha\leqslant 2\pi, \\ D_Z[-1] & \text{if } 2\pi<\alpha<4\pi. \end{cases} \quad (40)$$

Thus a wavefunction $u(p)\in U^{0,+,1,0}$ transforms as

$$\hat{R}^{0,+,1,0}(\alpha)u(p) = \begin{cases} u(R^{-1}(\alpha)p) & \text{if } 0\leqslant\alpha\leqslant 2\pi, \\ -u(R^{-1}(\alpha)p) & \text{if } 2\pi<\alpha<4\pi. \end{cases} \quad (41)$$

We emphasize that this splitting is not interpretable as helicity; rather it corresponds to the double-valuedness of $U^{0,+,1,0}$ and so suggests that this representation describes massless spinor fields.

IV. COVARIANT FIELD THEORIES

The UIR's found in the previous section provide the foundation for any relativistic field theory in three dimensions. However, because the individual UIR's transform in such diverse and complicated ways, it is very difficult to introduce interactions between them. Therefore, one instead begins with "covariant" fields, i.e., fields transforming as

$$(a,\Lambda):\quad \phi(p)\to e^{ip\cdot a}D(\Lambda)\phi(\Lambda^{-1}p), \quad (42)$$

where D is a finite-dimensional representation of the full Lorentz group (as opposed to some stability subgroup).

These covariant fields are not, in general, irreducible, and so their prescription must be supplemented with field equations (and, if necessary, subsidiary conditions) which remove the unphysical degrees of freedom. One usually then verifies that the solutions of the field equations provide a UIR of π. An alternate approach is to display a suitable "Foldy–Wouthuysen transformation" which disentangles the UIR's lying within a covariant field.

Finite-dimensional representations of L

Let

$$\alpha_0 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad \alpha_1 = \begin{pmatrix}0 & i\\ i & 0\end{pmatrix},\quad \alpha_2 = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}; \quad (43)$$

then we have the following "Dirac algebra":

$$(\alpha_\mu,\alpha_\nu)_+ = 2g_{\mu\nu},$$
$$(\alpha_\mu,\alpha_\nu)_- = -2i\epsilon_{\mu\nu\lambda}g^{\lambda\rho}\alpha_\rho. \quad (44)$$

If we define

$$M_{\mu\nu} = \tfrac{1}{4}i(\alpha_\nu,\alpha_\mu)_- \quad (45)$$

then

$$(M_{\mu\nu},M_{\lambda\rho}) = i(g_{\mu\rho}M_{\nu\lambda} - g_{\nu\rho}M_{\mu\lambda} - g_{\mu\lambda}M_{\mu\nu} + g_{\nu\lambda}M_{\rho\mu}), \quad (46)$$

which is the Lie algebra of SO(2,1) $\sim L_+{}^{\uparrow}$. By exponentiating this algebra we obtain another (nonunitary) representation of $L_+{}^{\uparrow}$:

$$S(w)\equiv\exp(iw_{\mu\nu}M^{\mu\nu}) = \exp(\tfrac{1}{2}iw\cdot\alpha). \quad (47)$$

The fundamental representation of SO(2,1) is given by the action of $S(w)$ on a complex spinor

$$D_F(w):\ \xi_a\to\xi_a{}' = (S(w))_a{}^b\xi_b. \quad (48)$$

All other finite-dimensional representations of L can be constructed by forming tensor products and direct sums of this fundamental representation.

For example, let X be a traceless symmetric spinor–tensor of rank 2, i.e., transforming as

$$X_{ab}\to S_a{}^cS_b{}^dX_{cd}. \quad (49)$$

We can relate these objects to 3-vectors via

$$X_{ab} = x^\mu(\alpha_\mu)_a{}^cC_{cb}, \quad (50)$$

where C is the (charge conjugation) matrix defined by

$$\alpha_\mu{}^T = -C\alpha_\mu C^{-1} \quad (51)$$

and for which

$$CS^T(w) = S^{-1}(w)C. \quad (52)$$

Under a Lorentz transformation

$$\begin{aligned} X_{ab}{}' &= S_a{}^cS_b{}^dX_{cd} \\ &= x_\mu S_a{}^cS_b{}^d(\alpha^\mu)_c{}^eC_{ed} \\ &= x_\mu(S\alpha^\mu CS^T)_{ab} \\ &= x_\mu(S\alpha^\mu S^{-1}C)_{ab} \\ &= x_\mu(\Lambda^{-1})^\mu{}_\nu(\alpha^\nu)_a{}^cC_{cb}, \end{aligned} \quad (53)$$

where Λ is the matrix of the adjoint representation

$$\Lambda^\mu{}_\nu = \tfrac{1}{2}\mathrm{Tr}(S^{-1}\alpha^\mu S\alpha_\nu) = (\Lambda^{-1})_\nu{}^\mu. \quad (54)$$

Thus equivalently

$$x_\mu\to\Lambda_\mu{}^\nu x_\nu \quad (55)$$

under $S(w)$. Indeed, one may verify that the matrix is exactly that matrix which we would associate with a Lorentz rotation about the 3-vector w.

Scalar fields

We define a "covariant" scalar field by its Lorentz transformation properties

$$\phi(p)\rightarrow\phi(\Lambda^{-1}p) \tag{56}$$

and its field equation

$$(p^2-m^2)\phi(p)=0. \tag{57}$$

Because of their simple transformation properties we can readily infer that a massive scalar field belongs to $U^{m,+,0}$, while a massless scalar field belongs to $U^{0,+,0,0}$.

Dirac fields

Let $\psi(p)$ be a covariant field transforming according to the fundamental representation of SO(2,1)

$$\psi(p)\rightarrow\exp(\tfrac{1}{2}i\omega\cdot\alpha)\psi(\Lambda^{-1}p) \tag{58}$$

and satisfying

$$(p\cdot\alpha-m)\psi(p)=0. \tag{59}$$

We expect, as in the four-dimensional case, that this field is actually the direct sum of a positive energy and a negative energy UIR. Therefore, to disentangle these UIR's, we search for a Foldy–Wouthuysen (F–W) transformation which will separate the positive and negative energy components.

1. Massive case

The operator which diagonalizes the Hamiltonian

$$H=\alpha_0\alpha\cdot p+\alpha_0 m \tag{60}$$

is

$$U=\exp\left(\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|\mathbf{p}|}\frac{\lambda}{2}\right),\quad \lambda=\tan^{-1}\frac{|\mathbf{p}|}{m}. \tag{61}$$

The new Hamiltonian is then

$$\begin{aligned}H'=\mathrm{UHU}^{-1}&=(m^2+p^2)^{1/2}\alpha_0\\&=(m^2+p^2)^{1/2}\begin{pmatrix}1&0\\0&-1\end{pmatrix},\end{aligned} \tag{62}$$

and so U succeeds in separating ψ into positive and negative energy components:

$$U\psi(p)=\begin{pmatrix}\phi^+(p)\\\phi^-(p)\end{pmatrix}, \tag{63}$$

$$H\phi^\pm(p)=\pm(m^2+p^2)^{1/2}\phi^\pm(p). \tag{64}$$

In order to further specify the nature of the "canonical" fields ϕ^+ and ϕ^-, we examine the F–W transform of the operator associated with infinitesimal rotations

$$\begin{aligned}R'=\mathrm{URU}=&\exp\left(\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|p|}\frac{\lambda}{2}\right)(1+\tfrac{1}{2}i\theta\alpha_0-\theta\,\mathbf{p}\wedge\boldsymbol{\partial})\\&\times\exp\left(-\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|p|}\frac{\lambda}{2}\right).\end{aligned} \tag{65}$$

Using this relation

$$e^TAe^{-T}=A+(T,A)+\tfrac{1}{2}(T,(T,A))+\cdots \tag{66}$$

and the fact that

$$\left(\alpha_0,\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|p|}\right)=-\left(\mathbf{p}\wedge\boldsymbol{\partial},\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|p|}\right), \tag{67}$$

we find

$$R'=1+\tfrac{1}{2}i\theta\alpha_0+\theta\mathbf{p}\cdot\boldsymbol{\alpha}. \tag{68}$$

Therefore, the canonical fields transform as

$$\phi^\pm(p)\rightarrow(1\pm\tfrac{1}{2}i\theta+\theta\mathbf{p}\wedge\boldsymbol{\partial})\phi^\pm(p). \tag{69}$$

Equations (64) and (69) suggest that ϕ^+ and ϕ^- belong, respectively, to $U^{m,+,1/2}$ and $U^{m,-,-1/2}$. Indeed, if we apply an inverse F–W transformation to the operator

$$L=L^{m,+,1/2}\oplus L^{m,-,-1/2}=\left(1+\frac{i}{2}\frac{\boldsymbol{\Theta}\wedge\mathbf{p}}{Em}\alpha_0-E\alpha_0\boldsymbol{\Theta}\cdot\boldsymbol{\partial}\right), \tag{70}$$

we obtain

$$L=\left(1-\tfrac{1}{2}\boldsymbol{\Theta}\cdot\boldsymbol{\alpha}\,\alpha_0-E\,\boldsymbol{\Theta}\cdot\boldsymbol{\partial}-\tfrac{1}{2}\frac{\boldsymbol{\Theta}\cdot\mathbf{p}}{E}\right), \tag{71}$$

which, excluding the last term, is the appropriate "covariant" operator. However, this last term may be removed by simply redefining the canonical fields as

$$\Phi^\pm(p)=E^{1/2}\phi^\pm(p). \tag{72}$$

Thus, as in four dimensions, the massive Dirac field is a direct sum of positive energy and negative energy spin-$\frac{1}{2}$ UIR's.

2. Massless case

The operator which diagonalizes the massless Dirac Hamiltonian

$$H=\alpha_0\mathbf{p}\cdot\boldsymbol{\alpha} \tag{73}$$

is

$$U=\exp\left(\frac{\pi}{4}\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|\mathbf{p}|}\right). \tag{74}$$

The new Hamiltonian

$$H'=\mathrm{UHU}^{-1}=|\mathbf{p}|\alpha_0 \tag{75}$$

again separates the Dirac field ψ into positive and negative energy components

$$U\psi(p)=\begin{pmatrix}\phi^+(p)\\\phi^-(p)\end{pmatrix}, \tag{76}$$

$$H\phi^\pm(p)=\pm|\mathbf{p}|\phi^\pm(p). \tag{77}$$

To determine the transformation properties of the canonical fields ϕ^+ and ϕ^-, we again examine the canonical version of R:

$$\begin{aligned}R'&=\mathrm{URU}^{-1}\\&=\exp\left(\frac{\pi}{4}\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|\mathbf{p}|}\right)(1+\tfrac{1}{2}i\theta\alpha_0+\theta\mathbf{p}\wedge\boldsymbol{\partial})\exp\left(\frac{-\pi}{4}\frac{\mathbf{p}\cdot\boldsymbol{\alpha}}{|\mathbf{p}|}\right)\\&=(1+\tfrac{1}{2}i\theta\alpha_0+\theta\,\mathbf{p}\wedge\boldsymbol{\partial}).\end{aligned} \tag{78}$$

Therefore, under an infinitesimal rotation

$$\phi^\pm(p)\rightarrow(1\pm\tfrac{1}{2}i\theta+\theta\,\mathbf{p}\wedge\boldsymbol{\partial})\phi^\pm(p). \tag{79}$$

However, if we redefine the canonical fields as

$$\Phi^\pm(p)=(p_1\mp p_2)^{1/2}\phi^\pm(p) \tag{80}$$

we have under R

$$\Phi^\pm(p)\rightarrow\Phi^\pm(R^{-1}p) \tag{81}$$

while under finite rotations $R(\alpha)$

$$\Phi^\pm(p)\rightarrow\begin{cases}\Phi^\pm(R^{-1}p) & \text{if } 0\leqslant\alpha\leqslant2\pi,\\-\Phi^\pm(R^{-1}p) & \text{if } 2\pi<\alpha<\pi'\end{cases} \tag{82}$$

due to the doublevaluedness of the function $(p_1\mp ip_2)^{1/2}$.[8]

Further, one may verify that

$$L' = L^{0,+,1,0} \oplus L^{0,-,1,0} = (1 + \alpha_0 E\,\mathbf{\Theta}\cdot\mathbf{\partial}) \tag{83}$$

leads to the appropriate covariant operator

$$L(\mathbf{\Theta}) = (1 - \tfrac{1}{2}\mathbf{\Theta}\cdot\mathbf{\alpha}\alpha_0 - E\,\mathbf{\Theta}\cdot\mathbf{\partial}). \tag{84}$$

Thus, we can conclude that $\Phi^{+} \in U^{0,+,1,0}$ and $\Phi^{-} \in U^{0,-,1,0}$, confirming our suspicion that these double-valued representations would be connected with massless spinor fields.

Vector fields

As a final example of a covariant field theory in three dimensions, we now examine covariant vector fields, i.e., fields transforming as

$$A_\mu(p) \rightarrow \Lambda_\mu{}^\nu A_\nu(\Lambda^{-1}p). \tag{85}$$

1. Massive case

We begin with a covariant vector field $A^\lambda(p)$ satisfying

$$(p^2 - m^2)A^\lambda(p) = 0, \tag{86}$$

$$p_\lambda A^\lambda(p) = 0. \tag{87}$$

Following the example of the Dirac field, we look for a transformation which will permit an easy separation of the scalar part [which is anulled by the subsidiary condition (87)] from the vector field. Therefore, we take our F–W transformation to be a Lorentz transformation $\Lambda^{-1}(p)$ which takes the momentum p to its rest frame $(m,0,0)$. We can then define the canonical vector field as

$$B_\mu(p) = \Lambda^{-1}{}_\mu{}^\lambda(p)A_\lambda(p). \tag{88}$$

We see that the subsidiary condition implies

$$B_0(p) = 0. \tag{89}$$

Apparently, the canonical field transforms as

$$\begin{aligned} B(p) &\rightarrow \Lambda^{-1}(p)L_\Lambda \Lambda(p)B(p) \\ &= \Lambda^{-1}(p)\Lambda\Lambda(\Lambda^{-1}p)B(\Lambda^{-1}p) \end{aligned} \tag{90}$$

or infinitesimally [see Eqs. (34) and (35)]

$$R'(\theta)B_0 = L'(\mathbf{\theta})B_0 = B_0 = 0, \tag{91}$$

$$R(\theta)B_i(p) = (\delta_{ij} + \epsilon_{ij}\theta + \delta_{ij}\theta p \wedge \partial)B_j(p), \tag{92}$$

$$L(\mathbf{\theta})B_i(p) = \left(\delta_{ij} + \epsilon_{ij}\,\frac{\theta \wedge p}{E+m} - \delta_{ij}E\theta\cdot\partial\right)B_j(p). \tag{93}$$

Thus we may take

$$\begin{aligned} B_1(p) &= \mathrm{Re}[\Phi(p)], \\ B_2(p) &= \mathrm{Im}[\Phi(p)], \end{aligned} \tag{94}$$

where $\Phi(p) \in U^{m,+,1}$.

2. Massless case

Let $A^\lambda(p)$ be a covariant vector field satisfying

$$p^2 A^\lambda(p) - p^\lambda p_\mu A^\mu(p) = 0. \tag{95}$$

Unlike the previous examples, the unphysical degrees of freedom for this field reside in the gauge freedom of the theory. Therefore, instead of looking for an F–W transformation which separates the constituent UIR's, we endeavor to remove the gauge ambiguity from the field.

If we choose a gauge where the Lorentz condition

$$p_\lambda A^\lambda(p) = 0 \tag{96}$$

is satisfied, we find that $A^\lambda(p)$ must have the form

$$A^\lambda(p) = p^\lambda \xi(p) + p_\perp \eta(p) \tag{97}$$

since $p_\perp = (0, p_2, -p_1)$ is the only 3-vector "perpendicular" to p besides p itself. The first term $p^\lambda \xi(p)$ is obviously a residual gauge freedom, and so we conclude that the physical part of $A^\lambda(p)$ has but one degree of freedom.

Let us then define $\phi(p)$ by

$$p_\mu \phi(p) = \epsilon_{\mu\nu\lambda} p^\nu A^\lambda(p). \tag{98}$$

Apparently, ϕ is oblivious to any gauge transformation on A^λ. Further, this definition implies both

$$p^\mu p_\mu \phi(p) = p^\mu \epsilon_{\mu\nu\lambda} p^\nu A^\lambda(p) = 0 \tag{99}$$

and

$$\begin{aligned} p_\mu p^\mu A^\lambda(p) = p^\lambda p_\mu A^\mu(p) &= \epsilon^{\mu\sigma\lambda}\epsilon_{\mu\nu\rho}p_\sigma p^\nu A^\rho(p) \\ &= \epsilon^{\mu\sigma\lambda} p_\mu p_\sigma \phi(p) \\ &= 0. \end{aligned} \tag{100}$$

Thus, in three dimensions we can replace the theory of a massless vector field with that of a scalar field $\phi \in U^{0,+,0,0}$.

V. CONCLUSIONS

We have classified all the unitary irreducible representations of the three-dimensional Poincaré group. Although the wavefunctions of these representations have only one component, they are distinguishable by the phase factors which effect Lorentz transformations (and we have referred to this phenomena as three-dimensional "spin"). Because the manifold (in the literary sense) transformation formulae of these UIR's would pose a problem in constructing theories of interacting particles, we have displayed the connection between the physical UIR's and (the more wieldy) covariant field theories in three dimensions.

One might argue that we were too cavalier in choosing the physical UIR's and, in particular, in ruling out the case of massless particles with continuous spin. However, upon further investigation, this one would find that in any finite-dimensional representation of SO(2,1) the generator of the (continuous part of the) stability group of massless particles is nilpotent. Since nilpotent operators possess only eigenvectors with eigenvalue 0, he should conclude that only the scalar and "almost" scalar UIR's can appear in covariant field theories.

Massive particles in three dimensions were also seen to possess a continuous spectra of "spin." Even so, only UIR's with half-integral spin appeared in our examples of massive covariant field theories. In fact, since the finite dimensional representations of L, being constructable from the fundamental (spinor) representation of SO(2,1), are at most double-valued, we may conclude that *only* the integral and half-integral UIR's are relevant to covariant field theories.

We have seen that although a *massive* vector field is *spin* 1, the *massless* vector field in three dimensions is *spin* 0. In four dimensions an analogous situation is exhibited by antisymmetric 2-tensor fields.[9] In fact, in the light of a recent paper by Aurilia and Takahashi,[10] which treats antisymmetric tensor fields as generalizations of Abelian gauge fields, this analogy does not seem so coincidental and leads us to further speculate that a massless antisymmetric 2-tensor field in three dimensions is nonpropagating.

It can be shown that the two physical massless representations are the only UIR's which have extensions to the conformal group. In de Sitter space this trait can be taken as a definition of "massless-ness" since it provides a criterion which can be applied consistently through the flat space limit (when the background space is curved, "mass" is not so well defined). This then implies that these two representations can be identified with the "Di" and "Rac" fields of Flato and Fronsdal.

There is a school of thought which maintains that the differences between electromagnetism and gravity are due to the different spins of their quanta.[11] But in three dimensions there is really only one spin available (we exclude the massive and double-valued representations), and so it seems impossible to accommodate this viewpoint. However, upon further investigation one finds that in three dimensions the free field equation $R^{\mu\nu} = 0$ implies that spacetime is flat.[12] With the geometry so fixed, there are no physical degrees of freedom left for the metric field; therefore, it cannot propagate, and thus the inconsistency is eliminated.

Note added in proof: The existence and utility of parity and time reversal in three-dimensional gauge theories were recently demonstrated by Jackiw and Templeton.[13]

ACKNOWLEDGMENTS

The author gratefully acknowledges many useful conversations with C. Fronsdal, S. Salamó, and M. Flato.

[1]Although this study is pretty much self-contained, an interested reader may refer to the original papers on four-dimensional field theory. E. P. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann and E. P. Wigner, Proc. Nat. Acad. Sci. USA **34**, 211 (1948); L. L. Foldy, Phys. Rev. **102**, 568 (1956); C. Fronsdal, Phys. Rev. **113**, 1367 (1959).

[2]M. Flato and C. Fronsdal, Phys. Lett. B **97**, 236 (1980).

[3]In the case of discrete symmetries, such as time reversal, antiunitary operators may also occur.

[4]G. W. Mackey, Acta. Math. **99**, 265 (1958).

[5]For the four-dimensional case, see A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN-Polish Scientific Publishers, Warsaw, 1977).

[6]Ref. 5, p. 518.

[7]G. J. Iverson and G. Mack, Ann. Phys. **64**, 211 (1971).

[8]One might inquire as to why we didn't attempt such a redefinition in the massive case. The answer is that such a redefinition is "singular" when there exists a rest frame.

[9]K. Hayashi, Phys. Lett. B **44**, 497 (1973).

[10]A. Aurilia and Y. Takahashi, Imperial College Preprint ICTP/79-80/7, Imperial College, 1979–80.

[11]See, for instance, S. N. Gupta, Phys. Rev. **96**, 1684 (1954); J. Fang and C. Fronsdal, J. Math. Phys. **20**, 2264 (1979).

[12]A. Staruszkiewicz, Acta Phys. Pol. **24**, 735 (1963).

[13]R. Jackiw and S. Templeton, Phys. Rev. D **23**, 2291 (1981).

The nonlinear Schrödinger equation as a Galilean-invariant dynamical system

L. Martínez Alonso
Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense, Madrid-3, Spain

(Received 29 October 1981; accepted for publication 12 March 1982)

The invariance of the nonlinear Schrödinger equation under the Galilei group is analyzed from the point of view of the inverse scattering transform. It is shown that this group induces an infinite-dimensional nonlinear canonical realization which is locally equivalent to a direct product of the two well-known Galilean actions describing classical particles and the free Schrödinger equation.

PACS numbers: 11.30.Na

I. INTRODUCTION

Lie group-theoretical methods are an essential tool for analyzing and interpreting the mathematical models of classical and quantum mechanics. In this paper we use them in order to understand the meaning of the nonlinear Schrödinger equation

$$i\psi_t = -\frac{1}{2\mu}\psi_{xx} - g|\psi|^2\psi, \quad \mu>0, \quad g>0, \tag{1.1}$$

considered as a Galilean-invariant dynamical system. The quantization of this wave equation leads to an exactly soluble, completely finite, nonrelativistic field theory[1]: nonrelativistic bosons of mass μ interacting in pairs via an attractive δ-function potential, in one space dimension. But we are here interested in (1.1) as a nonlinear partial differential equation describing a nonrelativistic classical field. Our starting point is the invariance of (1.1) under the transformations

$$\psi(t,x)\rightarrow\psi'(t',x') = \exp[i(-\tfrac{1}{2}\mu v^2 t' + \mu v x' + C)]\psi(t,x), \tag{1.2}$$

where

$$t' = t+b, \quad x' = x+vt+a \tag{1.3}$$

are the Galilean transformations in two-dimensional space–time. As a consequence, (1.2) determines a realization of the extended Galilei group G whose elements are of the form

$$g = (\theta,b,a,v) = \exp(-\theta\widehat{M})\exp(-b\widehat{H})\exp(a\widehat{P})\exp(v\widehat{K}), \quad \theta,b,a,v\in\mathbb{R}, \tag{1.4}$$

with the composition law

$$g_1g_2 = (\theta_1+\theta_2+\tfrac{1}{2}v_1^2b_2+v_1a_2, b_1+b_2, a_1+a_2+v_1b_2, v_1+v_2). \tag{1.5}$$

Moreover, there is a symplectic structure which permits us to formulate (1.1) as a Hamiltonian system such that (1.2) defines a canonical realization of G.

The fundamental ingredient of our analysis is the use of the inverse scattering transform technique. As is well-known, after the discovery of this method of resolution for the Korteweg–deVries equation,[2] the nonlinear Schrödinger equation (1.1) was the second physically interesting nonlinear model solved by means of an inverse scattering transform.[3] In this way, we have at hand a simple complete picture of (1.1) in terms of the scattering data variables associated with the Zakharov–Shabat spectral problem. We find that these variables have simple transformation laws under the Galilei group. It allows us to perform a complete characterization of (1.1) in terms of well-known Galilean-invariant systems. Our main results are the following:

(1) The structure of the set of scattering data variables consisting of a discrete and a continuous part leads us to identify locally the phase space of (1.1) with an infinite-dimensional Euclidean space of the form $\mathbb{R}^{4N}\times L^2(\mathbb{R})$. This local decomposition of the phase space reduces both the evolution law and the Galilei action to a direct product of two components acting on $\mathbb{R}^{4N}$ and $L^2(\mathbb{R})$.

(2) The nonlinear Schrödinger equation can be described as the composition of two independent dynamical systems. One of them has a finite-dimensional phase space and represents a system of free classical particles. The other one is characterized by a field function $\phi=\phi(x)$ evolving according to the free Schrödinger equation

$$i\phi_t = -(1/2\mu)\phi_{xx}. \tag{1.6}$$

It is proved that $\psi(x)=\phi(x)$ in the linear limit of the inverse scattering transform.

(3) Under the constraint $\phi=0$, (1.1) reduces to a system of free classical particles. The dynamical state of such a system may be specified in two equivalent ways either as a point on $\mathbb{R}^{4N}$ or as a field function $\psi(x)$. It follows that the corresponding solutions $\psi(t,x)$ of (1.1) are the pure N-soliton solutions. In particular, we obtain that a free particle is described by a plane wave modulated by a pulse of permanent shape whose center moves with the free particle trajectory.

(4) We analyze the action of a uniform constant field over (1.1) by considering the modified equation

$$i\psi_t = -\frac{1}{2\mu}\psi_{xx} - g|\psi|^2\psi + V(x)\psi, \quad V(x)\equiv -f_0 x, \tag{1.7}$$

where the interaction term is introduced following the quantum-mechanical procedure. It is found that the classical particles of the model react to the external field as if all of them should have a mass equal to μ. In addition, the field $\phi(x)$ evolves now according to the linear Schrödinger equation

$$i\phi_t = -(1/2\mu)\phi_{xx} + V(x)\phi. \tag{1.8}$$

II. THE ACTION OF THE GALILEI GROUP

Let us denote by V the space of initial data for (1.1). We shall assume that the elements of V are rapidly decaying smooth functions $\psi = \psi(x)$. We can think of (1.1) as an infinite-dimensional Hamiltonian system by introducing on the set of functionals of the form $F = F[\psi,\psi^*]$ the Poisson bracket operation

$$\{F_1,F_2\} = i\int_{-\infty}^{\infty}\left(\frac{\delta F_1}{\delta\psi^*(x)}\frac{\delta F_2}{\delta\psi(x)} - \frac{\delta F_1}{\delta\psi(x)}\frac{\delta F_2}{\delta\psi^*(x)}\right)dx . \tag{2.1}$$

It allows us to write (1.1) in the Hamiltonian form

$$\partial_t\psi(x) = \{\psi(x),H\} , \quad \partial_t\psi^*(x) = \{\psi^*(x),H\} , \tag{2.2}$$

where

$$H = \int_{-\infty}^{\infty}\left(\frac{1}{2\mu}|\psi_x|^2 - \tfrac{1}{2}g|\psi|^4\right)dx . \tag{2.3}$$

In this way the evolution law $U(t):\psi(0)\rightarrow\psi(t)$ associated with the nonlinear Schrödinger equation is a one-parameter group of canonical transformations over V.

Let us now consider the action of the extended Galilei group G. From the active point of view the action (1.2) becomes

$$(\hat{R}(g)\psi)(t,x) = \exp\{i[-\tfrac{1}{2}\mu v^2 t + \mu v x + C(g)]\} \times\psi(t-b,x-v(t-b)-a) , \tag{2.4}$$

where $C(g)$, chosen so that $\hat{R}(g_1g_2) = \hat{R}(g_1)\hat{R}(g_2)$, is given by

$$C(g) = \mu(\tfrac{1}{2}bv^2 - av + \theta) .$$

From (2.4) we have that the initial data transform under G in the form

$$(R(g)\psi)(x) = \exp\{i[\mu v x + C(g)]\}(U(-b)\psi)(x+vb-a) , \tag{2.5}$$

and it determines a realization of G as a group of transformations over V. The presence of the evolution map $U(-b)$ in the definition (2.5) implies that R is a nonlinear realization of G.

From the physical point of view, if we think of (1.1) as a Galilean invariant system, the action of the Galilei group over physical states must reproduce the Galilei group law. However, the restriction of R to the Galilei group $\tilde{G}\equiv\{\tilde{g} = (0,b,a,v)\in G\}$ is not a true realization of $\tilde{G}$, but a realization of $\tilde{G}$ up to a constant phase factor. This means that two elements of V differing only in a constant phase factor must correspond to the same physical state. In addition, only those functionals $F[\psi,\psi^*]$ which are invariant under the transformation $\psi' = e^{i\alpha}\psi$ may represent physical observables.

The realization R of G is of a canonical character with respect to the symplectic structure (2.1). Indeed, the vector fields representing the Lie algebra generators of G are Hamiltonian fields of the form

$$X(\hat{M})\tilde{\psi}(x) = \{\tilde{\psi}(x),M\} , \quad M = \mu\int_{-\infty}^{\infty}|\psi|^2\,dx , \tag{2.6a}$$

$$X(\hat{H})\tilde{\psi}(x) = \{\tilde{\psi}(x),H\} , \qquad H = \int_{-\infty}^{\infty}\left(\frac{1}{2\mu}|\psi_x|^2 - \frac{g}{2}|\psi|^4\right)dx , \tag{2.6b}$$

$$X(\hat{P})\tilde{\psi}(x) = \{\tilde{\psi}(x),P\} , \quad P = \int_{-\infty}^{\infty}\psi^*(-i\psi_x) , \tag{2.6c}$$

$$X(\hat{K})\tilde{\psi}(x) = \{\tilde{\psi}(x),K\} , \quad K = -\mu\int_{-\infty}^{\infty}\psi^* x\psi\,dx , \tag{2.6d}$$

where $\tilde{\psi}(x)$ denotes either $\psi(x)$ or $\psi^*(x)$, and the vector field $X(A)$ corresponding to a given generator of G is defined by

$$X(A)\tilde{\psi}(x) = \frac{d}{da}\bigg|_{a=0}[R(\exp(aA))\psi]^{\sim}(x) . \tag{2.7}$$

Thus, we have a representation of the Lie algebra of G in terms of functionals depending on ψ and ψ^*. One may easily show the following Poisson bracket relations:

$$\begin{aligned}&\{M,H\} = \{M,P\} = \{M,K\} = \{H,P\} = 0 ,\\ &\{H,K\} = P, \quad \{P,K\} = M ,\end{aligned} \tag{2.8}$$

which reproduce the Lie algebra structure of G. As a consequence, the transformation properties of these functionals under the passive action (1.2) of G are

$$M' = M , \quad H' = H + \tfrac{1}{2}Mv^2 + \tfrac{1}{2}Pv , \tag{2.9a}$$

$$P' = P + Mv , \quad K' = K - Mvt - Ma . \tag{2.9b}$$

Clearly, M, H, and P transform like observables describing, respectively, the mass, the energy, and the momentum. On the other hand, $Q\equiv -K/M$ transforms like a position observable. We note also that a functional F is invariant under the transformation $\psi' = e^{i\alpha}\psi$ if and only if $\{F,M\} = 0$. That is to say, only those functionals which are in involution with M may represent observables of (1.1).

III. THE INVERSE SCATTERING TRANSFORM OF THE GALILEAN ACTION

A. Canonical variables associated with spectral data

Now we are going to describe a convenient coordinate system for our infinite-dimensional phase space V. First, we note that upon performing the transformation

$$u(t,x) = (g/2)^{1/2}\psi(t,(2\mu)^{-1/2}x) , \tag{3.1}$$

Eq. (1.1) becomes

$$iu_t = -u_{xx} - 2|u|^2u , \tag{3.2}$$

which is the standard form of the nonlinear Schrödinger equation as it appears in the literature about the inverse scattering method.

Given $\psi\in V$ we consider the Zakharov–Shabat spectral problem[3]

$$\left(i\sigma_3\partial_x + \begin{pmatrix}0 & u\\ -u^* & 0\end{pmatrix} - k\right)\varphi = 0 , \qquad u(x)\equiv(g/2)^{1/2}\psi((2\mu)^{-1/2}x) . \tag{3.3}$$

Let $\varphi_-(k,x)$ be the Jost solution of (3.3) with the properties

$$\begin{pmatrix}e^{-ikx}\\ 0\end{pmatrix} \underset{x\rightarrow-\infty}{\longleftarrow} \varphi_-(k,x) \underset{x\rightarrow+\infty}{\longrightarrow} \begin{pmatrix}a(k)e^{-ikx}\\ b(k)e^{ikx}\end{pmatrix} . \tag{3.4}$$

The function $a(k)$ is analytic in the upper half-plane $\mathrm{Im}k > 0$ and $a(k)\rightarrow 1$ as $k\rightarrow\infty$. The zeros k_l $(l = 1,\ldots,N)$ of $a(k)$ correspond to the eigenvalues of (3.1). Each of these zeros k_l determines a complex number c_l such that

$$\varphi_-(k_l,x) \underset{x\to+\infty}{\longrightarrow} \begin{pmatrix} 0 \\ c_l e^{ik_l x} \end{pmatrix}. \tag{3.5}$$

Moreover, the following relation holds:

$$|a(k)|^2 + |b(k)|^2 = 1\,, \quad k\in\mathbb{R}\,. \tag{3.6}$$

If we assume that the zeros of $a(k)$ are simple and that no zero lies on the real axis, then the inverse scattering theory of (3.3)[3,4] shows that the potential $u(x)$ can be uniquely recovered from the following set of scattering data:

$$(k_l,c_l,|a(k)|,\arg[b(k)])\,, \quad l=1,\dots,N\,, \quad k\in\mathbb{R}\,. \tag{3.7}$$

In this way, we may characterize every $\psi\in V$ in terms of the scattering data of the spectral problem (3.3). The Poisson bracket relations of the scattering data, considered as functionals depending on ψ and ψ^*, turn out to be very simple.[5,6] For our purposes we will find convenient to introduce the following set of scattering data variables:

$$\begin{aligned} q_l &= (2\mu)^{-1/2}\ln|c_l|/2\operatorname{Im}\{k_l\}\,, \\ P_l &= -16g^{-1}\operatorname{Re}\{k_l\}\operatorname{Im}\{k_l\}\,, \end{aligned} \tag{3.8a}$$

$$\begin{aligned} \tau_l &= \mu^{-1}(\arg c_l + \ln|c_l|\cdot\operatorname{Re}\{k_l\}/\operatorname{Im}\{k_l\})\,, \\ m_l &= 4g^{-1}(2\mu)^{1/2}\operatorname{Im}\{k_l\}\,, \end{aligned} \tag{3.8b}$$

$$\begin{aligned} q(k) &= \arg[b(-ck)]\,, \\ p(k) &= -(\pi g\mu)^{-1}\ln|a(-ck)|\,, \end{aligned} \tag{3.8c}$$

where $l=1,\dots,N$, $k\in\mathbb{R}$, and the constant c appearing in the expressions for $q(k)$ and $p(k)$ is given by

$$c\equiv[2(2\mu)^{1/2}]^{-1}\,. \tag{3.9}$$

As is shown in Appendix A, these variables are such that (q_1,p_1), (τ_1,m_1), and $(q(k),p(k))$ are pairs of canonically conjugate variables. That is to say, the Poisson bracket relations between two of these variables are all zero except for the following ones:

$$\{q_l,p_{l'}\} = \{\tau_l,m_{l'}\} = \delta_{ll'}\,, \quad \{q(k),p(k')\} = \delta(k-k')\,. \tag{3.10}$$

B. Galilean generators in terms of scattering data variables

The canonical variables we have just introduced are very appropriate for analyzing the Galilean action on the nonlinear Schrödinger equation (1.1). Indeed, we are going to show that the functionals defined in (2.6) which generate the canonical realization of G may be expressed in terms of these variables in the following form:

$$M = \sum_l m_l + \mu\int_{-\infty}^{\infty} p(k)\,dk\,, \tag{3.11a}$$

$$\mathrm{H} = \sum_l \left(\frac{p_l^2}{2m_l} - \frac{g^2}{24\mu^2}m_l^3\right) + \int_{-\infty}^{\infty}\frac{k^2}{2\mu}p(k)\,dk\,, \tag{3.11b}$$

$$P = \sum_l p_l + \int_{-\infty}^{\infty} kp(k)\,dk\,, \tag{3.11c}$$

$$K = -\sum_l m_l q_l + \mu\int_{-\infty}^{\infty} q(k)\frac{\partial p(k)}{\partial k}\,dk\,. \tag{3.11d}$$

The expressions for M, H, and P are a consequence of the so-called trace relations[3,5,6] associated with the Zakharov–Shabat spectral problem. This is based on the fact that the function $\ln[a(k)]$ has an asymptotic expansion for large $|k|$:

$$\ln[a(k)] = \sum_{n=1}^{\infty} C_n k^{-n}\,, \tag{3.12}$$

where the coefficients are given by

$$C_n = -\frac{1}{\pi}\sum_l (k_l^n - k_l^{*n}) + \frac{1}{\pi i}\int_{-\infty}^{\infty} k^{n-1}\ln|a(k)|\,dk\,. \tag{3.13}$$

On the other hand, these coefficients can be expressed as local functionals depending on the functions u and u^* arising in the Zakharov–Shabat spectral problem (3.3). It turns out that M, H, and P, are proportional to the first three coefficients of (3.12). Indeed, as was shown by Zakharov and Shabat,[3] we have

$$C_1 = -(i/2)\int_{-\infty}^{\infty}|u|^2\,dx\,, \tag{3.14a}$$

$$C_2 = (i/4)\int_{-\infty}^{\infty} u^*(-iu_x)\,dx\,, \tag{3.14b}$$

$$C_3 = -(i/8)\int_{-\infty}^{\infty}(|u_x|^2 - |u|^4)\,dx\,. \tag{3.14c}$$

From (3.13), (3.14), and after some elementary operations, it is straightforward to get the expressions (3.11a), (3.11b), and (3.11c).

The proof of (3.11d) follows from a different procedure and it requires the evaluation of the Poisson brackets between K and the canonical variables (3.8). In order to do that, we will find the action of pure Galilean transformations on the scattering data of (3.3). Given $\psi\in V$, let us denote

$$\psi'(x)\equiv[R(\exp(v\hat{K}))\psi](x) = \exp(i\mu vx)\psi(x) \tag{3.15}$$

and

$$u'(x)\equiv(g/2)^{1/2}\psi'((2\mu)^{-1/2}x) = \exp[i(\mu/2)^{1/2}vx]u(x)\,. \tag{3.16}$$

Now, let $\varphi'_-(k,x)$ be the Jost solution of (3.3) corresponding to the potential $u'(x)$. Then, one may easily show that the function

$$\exp[-i\tfrac{1}{4}(2\mu)^{1/2}vx\sigma_3]\varphi'_-(k,x) \tag{3.17}$$

is the Jost solution $\varphi_-(k+\frac{1}{4}(2\mu)^{1/2}v,x)$ of (3.3) corresponding to the initial potential $u(x)$. That is to say, under pure Galilean transformations, Jost solutions of (3.1) transform as

$$\begin{aligned} &\varphi_-(k,x)\mapsto\varphi'_-(k,x) \\ &\quad = \exp[i\tfrac{1}{4}(2\mu)^{1/2}vx\sigma_3]\varphi_-(k+\tfrac{1}{4}(2\mu)^{1/2}v,x)\,. \end{aligned} \tag{3.18}$$

This implies that

$$a'(k) = a(k+\tfrac{1}{4}(2\mu)^{1/2}v)\,, \quad b'(k) = b(k+\tfrac{1}{4}(2\mu)^{1/2}v)\,, \tag{3.19a}$$

$$k'_l = k_l - \tfrac{1}{4}(2\mu)^{1/2}v\,, \quad c'_l = c_l\,. \tag{3.19b}$$

In this way, we deduce at once that the variables (3.8) transform as follows:

$$q_l' = q_l\,,\quad p_l' = p_l + m_l v\,,\quad \tau_l' = \tau_l - q_l v\,,\quad m_l' = m_l\,, \tag{3.20a}$$

$$q'(k) = q(k - \mu v)\,,\quad p'(k) = p(k - \mu v)\,. \tag{3.20b}$$

At this point we use the fact that $R(\exp(v\hat{K}))$ is a one-parameter group of canonical transformations generated by K. Hence, for every functional F we have

$$\left.\frac{d}{dv}\right|_{v=0} F' = \{F,K\}\,.$$

Therefore, by making use of (3.20) and due to the canonical character of the variables (3.8), we obtain that the functional K verifies

$$\frac{\partial K}{\partial p_l} = 0\,,\quad \frac{\partial K}{\partial q_l} = -m_l\,, \tag{3.21a}$$

$$\frac{\partial K}{\partial m_l} = -q_l\,,\quad \frac{\partial K}{\partial \tau_l} = 0\,, \tag{3.21b}$$

$$\frac{\delta K}{\delta p(k)} = -\mu\frac{\partial q(k)}{\partial k}\,,\quad \frac{\delta K}{\delta q(k)} = \mu\frac{\partial p(k)}{\partial k}\,. \tag{3.21c}$$

Obviously, these relations lead to the expression (3.11d) for K.

C. A new field associated with the continuous scattering data variables

The continuous part of the system of variables (3.8) is given by two real functions $q(k)$ and $p(k)$ which satisfy

$$p(k) \geqslant 0\,,\quad q(k) \in \mathbb{R}\ (\mathrm{mod}\, 2\pi)\,. \tag{3.22}$$

Moreover, due to the form (2.6a) of M and (3.11a) we have that $p = p(k)$ belongs to $L^1(\mathbb{R})$. Therefore, we can describe the continuous scattering data variables by means of a square integrable complex function defined as

$$\hat{\phi}(k) \equiv -p(k)^{1/2}\exp[-iq(k)]\,, \tag{3.23}$$

or, equivalently, by the inverse Fourier transform of $\hat{\phi}$ given by

$$\phi(x) \equiv (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{ikx}\hat{\phi}(k)\,dk\,. \tag{3.24}$$

It is elementary to realize that the field $\phi(x)$, considered as a functional depending on the initial field functions ψ and ψ^*, verifies the Poisson bracket relations

$$\{\phi^*(x),\phi(y)\} = i\delta(x-y)\,,$$
$$\{\phi(x),\phi(y)\} = \{\phi^*(x),\phi^*(y)\} = 0\,. \tag{3.25}$$

Let us now suppose that $\psi \in V$ becomes infinitesimally small; in that case we are in the linear limit[4] of the inverse scattering transform, and, therefore, there are no bound states of the spectral problem (3.3). In addition, the potential $u(x)$ of (3.3) is in the first approximation given by[4,7]

$$u(x) = (-2\pi)^{-1}\int_{-\infty}^{\infty} b^*(-k/2)e^{ikx}dk\,. \tag{3.26}$$

On the other hand, since to first order $a(k) = 1 + \delta a(k)$ and $b(k) = \delta b(k)$, we have also in the first approximation that

$$\ln|a(k)| = \tfrac{1}{2}\ln(1 - |b(k)|^2) = -\tfrac{1}{2}|b(k)|^2\,. \tag{3.27}$$

From (3.8c), (3.23), and (3.27), it follows at once that

$$b^*(-k/2) = -(2\pi g\mu)^{1/2}\hat{\phi}((2\mu)^{1/2}k)\,. \tag{3.28}$$

Inserting this result into (3.26), we get

$$u(x) = (g/2)^{1/2}\phi((2\mu)^{-1/2}x)\,. \tag{3.29}$$

That is to say, when ψ becomes small then in the first approximation $\psi(x) = \phi(x)$.

D. Analysis of the Galilean action

We are now ready to analyze the dynamical meaning of the nonlinear Schrödinger equation (1.1) as seen in terms of the variables we have just introduced. In this sense it is convenient to think of our phase space V as being as infinite-dimensional manifold and to regard the inverse scattering transform as defining a chart on V. Suppose we take a given element ψ_0 of V, then in a sufficiently small neighborhood Ω of ψ_0 the number N of eigenvalues corresponding to the spectral problem (3.3) is the same for all $\psi \in \Omega$, and hence we can consider the map

$$\Omega \to \mathbb{R}^{4N}\times L^2(\mathbb{R})\,,\quad \psi = \psi(x) \to ((q_l,p_l,\tau_l,m_l), \phi = \phi(x))\,, \tag{3.30}$$

where $\phi = \phi(x)$ is the field defined in (3.24) which characterizes the continuous scattering data variables. The map (3.30) identifies V locally with the infinite-dimensional Euclidean space $\mathbb{R}^{4N}\times L^2(\mathbb{R})$. That is to say, (3.30) defines a local coordinate system on V. From (3.11), (3.23), and (3.24), we can immediately write down the expressions of the generators of the Galilean action in terms of this coordinate system. They are

$$M = \sum_l m_l + \mu\int_{-\infty}^{\infty} |\phi|^2\,dx\,, \tag{3.31a}$$

$$H = \sum_l\left(\frac{p_l^2}{2m_l} - \frac{g^2}{24\mu^2}m_l^3\right) + \int_{-\infty}^{\infty}\frac{1}{2\mu}|\phi_x|^2\,dx\,, \tag{3.31b}$$

$$P = \sum_l p_l + \int_{-\infty}^{\infty}\phi^*(-i\phi_x)\,dx\,, \tag{3.31c}$$

$$K = -\sum_l m_l q_l - \mu\int_{-\infty}^{\infty}\phi^* x\phi\,dx\,. \tag{3.31d}$$

We note that the symplectic structure (2.1) has a simple form when expressed in the coordinate system (3.30). It is given by

$$\{F_1,F_2\} = \sum_l\left[\left(\frac{\partial F_1}{\partial q_l}\frac{\partial F_2}{\partial p_l} - \frac{\partial F_1}{\partial p_l}\frac{\partial F_2}{\partial q_l}\right) + \left(\frac{\partial F_1}{\partial \tau_l}\frac{\partial F_2}{\partial m_l} - \frac{\partial F_1}{\partial m_l}\frac{\partial F_2}{\partial \tau_l}\right)\right] + i\int_{-\infty}^{\infty}\left(\frac{\delta F_1}{\delta\phi^*(x)}\frac{\delta F_2}{\delta\phi(x)} - \frac{\delta F_1}{\delta\phi(x)}\frac{\delta F_2}{\delta\phi^*(x)}\right)dx\,. \tag{3.32}$$

From (3.31) and (3.32) it follows that in the coordinate system (3.30) the group action R is a direct product

$$R = (R_1\otimes\cdots\otimes R_N)\otimes R^{(0)}\,, \tag{3.33}$$

where the components $R_l\,(l = 1,\ldots,N)$ act on $\mathbb{R}^4$ and $R^{(0)}$ on $L^2(\mathbb{R})$. Each component R_l describes a Galilean system whose states are specified by four coordinates (q_l,p_l,τ_l,m_l) which evolve in time according to the equations

$$\dot{q}_l = p_l/m_l\,,\quad \dot{p}_l = 0\,, \tag{3.34a}$$

$$\dot{\tau}_l = -\left(\frac{p_l^2}{2m_l^2} + \frac{g^2}{8\mu^2} m_l^2\right), \quad \dot{m}_l = 0 . \tag{3.34b}$$

Under the passive action (1.2) of G, these variables transform as

$$q_l'(t') = q_l(t) + vt + a , \quad p_l'(t') = p_l(t') + m_l v , \tag{3.35a}$$

$$\tau_l'(t') = \tau_l(t) - vq_l(t) - \tfrac{1}{2}v^2 t - \theta , \quad m_l'(t') = m_l(t) . \tag{3.35b}$$

This implies that R_l represents a free classical particle which has position and momentum observables with the correct Galilean transformation properties. Nevertheless, it is not an elementary Galilean particle[8] because the mass m_l is not a parameter but an additional variable in the phase space. The variable τ_l canonically conjugate to m_l does not represent any Galilean observable because it is not in involution with the total mass observable M. We notice that both the evolution law and the Galilean transformation law of (q_l,p_l,m_l) are independent of τ_l. That is to say, the presence of τ_l in the space of states does not perturb the physical interpretation of the remaining variables.

In what concerns the continuous component $R^{(0)}$, we see that the expression (3.31b) for the Hamiltonian H and (3.32) imply that $R^{(0)}$ describes a Galilean field $\phi = \phi(x)$ evolving according to the free Schrödinger equation

$$i\phi_t = -(1/2\mu)\phi_{xx} . \tag{3.36}$$

The form (3.31) for the generators of G implies also that ϕ transforms under the passive action of G in the following way:

$$\phi'(t',x') = \exp\{i[-\tfrac{1}{2}\mu v^2 t' + \mu v x' + C(g)]\}\phi(t,x) , \tag{3.37}$$

which is the same transformation law as the one satisfied by the field ψ [see (1.2)].

E. Wave description of the particles arising in the model

An important consequence of the above analysis is the existence inside the phase space V of finite-dimensional invariant submanifolds describing systems of free classical particles. Indeed, for every integer $N>0$ we define V_N as the set of elements $\psi \in V$ such that the number of bound states of the corresponding spectral problem (3.3) is exactly equal to N, and such that the function $\phi = \phi(x)$ vanishes. From (3.30) we have obviously that locally $V_N \simeq \mathbb{R}^{4N}$. In this way, the elements of V_N may be specified in two equivalent ways, either by a field function $\psi = \psi(x)$ or by a point $\{(q_l,p_l,\tau_l,m_l); l = 1,...,N\}$ of $\mathbb{R}^{4N}$. We observe that in terms of scattering data the condition $\phi = 0$ reads

$$b(k) = 0 . \tag{3.38}$$

This means[3] that the submanifold V_N is the phase space for the pure N-soliton solutions of the nonlinear Schrödinger equation. Therefore, our analysis provides a precise mathematical meaning to the similarity between solitons and particles.[9]

It is interesting to consider the one-soliton manifold V_1 which according to our analysis describes a free classical particle. Suppose an element $\psi \in V_1$, which in the coordinate system (3.30) is represented by a point (q,p,τ,m). By means of the inverse scattering transform (see Appendix B) we can reconstruct ψ from (q,p,τ,m). The result is

$$\psi(x) = -\tfrac{1}{2}(g/\mu)^{1/2} m \exp[i(\mu/m)(px - m\tau - qp)] \times \operatorname{sech}[\tfrac{1}{2}mg(x-q)] . \tag{3.39}$$

This wave evolves according to

$$\psi(t,x) = -\tfrac{1}{2}(g/\mu)^{1/2} m \times \exp\{i(\mu/m)[px - (p^2/2m - (g^2/8\mu^2)m^3)t]\} \times \operatorname{sech}\{\tfrac{1}{2}mg[x - q(t)]\} . \tag{3.40}$$

This is a plane wave modulated by a pulse of permanent shape whose center moves exactly with the free particle trajectory $q(t) = q(0) + tp/m$. Observe that the wavenumber k and the frequency ω of the plane wave factor are related to the particle variables in the form

$$k = (\mu/m)p , \quad \omega = (\mu/m)[p^2/2m - (g^2/8\mu^2)m^3] . \tag{3.41}$$

Incidentally, it is worth noting that (3.40) provides us with a wave description of the classical free particle which reminds us of the ideas suggested by the de Broglie theory of the double solution.[10]

IV. INTERACTION WITH A UNIFORM CONSTANT FIELD

Suppose that the action of a uniform constant field on the system described by the nonlinear Schrödinger equation is represented by the equation[11,12]

$$i\psi_t = -(1/2\mu)\psi_{xx} - g|\psi|^2\psi + V(x)\psi , \quad V(x) \equiv -f_0 x . \tag{4.1}$$

It is also a Hamiltonian system with respect to the simplectic form (2.1). The corresponding Hamiltonian is

$$H' = H + \int_{-\infty}^{\infty} \psi^* V(x)\psi \, dx = H + (f_0/\mu)K , \tag{4.2}$$

where H is the Hamiltonian (2.3) for the nonlinear Schrödinger equation and K is the functional (2.6d) representing the generator of pure Galilean transformations. By using the expressions (3.31b) and (3.31d) for H and K, respectively, we immediately obtain the form of H' in terms of the coordinate system (3.30),

$$H' = \sum_l \left(\frac{p_l^2}{2m_l} - \frac{g^2}{24\mu^2} m_l^3 - \frac{f_0}{\mu} m_l q_l\right) + \int_{-\infty}^{\infty} \left(\frac{1}{2\mu}|\phi_x|^2 - f_0 x \phi^* \phi\right) dx . \tag{4.3}$$

Clearly, the equations of motion are now given by

$$\dot{q}_l = p_l/m_l , \quad \dot{p}_l = (f_0/\mu)m_l , \tag{4.4a}$$

$$\dot{\tau}_l = -\left(\frac{p_l^2}{2m_l^2} + \frac{g^2}{8\mu^2} m_l^2 + \frac{f_0}{\mu} q_l\right), \quad \dot{m}_l = 0 , \tag{4.4b}$$

$$i\phi_t = -\frac{1}{2\mu}\phi_{xx} + V(x)\phi , \quad V(x) \equiv -f_0 x . \tag{4.4c}$$

Curiously, the field ϕ obeys the linear Schrödinger equation with potential $V(x)$. On the other hand, we observe that, according to (4.4a), the acceleration of the particles is

$$\ddot{q}_l = f_0/\mu . \tag{4.5}$$

If the force f_0 is used to determine the inertial mass of these particles as the quotient between force and acceleration, we must conclude that all of these particles have a mass equal to μ.

ACKNOWLEDGMENTS

I wish to thank A. Galindo and A. Fernández-Rañada for several helpful discussions during the course of this work.

APPENDIX A

In order to prove that the variables (3.8) form a canonical system we relate them to the variables used by Faddeev,[6] which are given by

$$\eta_l = -\ln|c_l|, \quad \xi_l = 4\mathrm{Re}\{k_l\}, \tag{A1a}$$

$$\beta_l = \arg c_l, \quad \alpha_l = 4\mathrm{Im}\{k_l\}, \tag{A1b}$$

$$Q(k) = \arg b(k), \quad P(k) = -\frac{2}{\pi}\ln|a(k)|. \tag{A1c}$$

The relationship between both systems of variables is

$$q_l = -2(2\mu)^{-1/2}\eta_l/\alpha_l, \quad p_l = -g^{-1}\xi_l\alpha_l, \tag{A2a}$$

$$\tau_l = \mu^{-1}(\beta_l - \eta_l\xi_l/\alpha_l), \quad m_l = (2\mu)^{1/2}g^{-1}\alpha_l, \tag{A2b}$$

$$q(k) = Q(-ck), \quad p(k) = (2\mu g)^{-1}P(-ck). \tag{A2c}$$

Faddeev proves that (η_l,ξ_l), (β_l,α_l), and $(Q(k),P(k))$, are pairs of canonically conjugate variables with respect to the symplectic structure

$$\{F_1,F_2\}' = i\int_{-\infty}^{\infty}\left(\frac{\delta F_1}{\delta u^*(x)}\frac{\delta F_2}{\delta u(x)} - \frac{\delta F_1}{\delta u(x)}\frac{\delta F_2}{\delta u^*(x)}\right)dx, \tag{A3}$$

where $u = u(x)$ is the potential of the Zakharov–Shabat spectral problem (3.3). But, due to the relation

$$u(x) = (g/2)^{1/2}\psi((2\mu)^{-1/2}x), \tag{A4}$$

we deduce easily that our Poisson bracket (2.1) is related to that of Faddeev in the form

$$\{F_1,F_2\} = (2\mu)^{1/2}(g/2)\{F_1,F_2\}'. \tag{A5}$$

From (A2) and (A5) and since the Faddeev variables are canonical with respect to $\{\ ,\ \}'$, it follows at once that our variables (3.8) form a canonical system with respect to the symplectic structure (2.1).

APPENDIX B

In this appendix we want to indicate how (3.39) derives from the application of the inverse scattering transform. Let us consider the Zakharov–Shabat spectral problem (3.3) and suppose we have a potential $u = u(x)$ such that $b(k) = 0$, then the explicit form of $u = u(x)$ can be obtained from its scattering data variables by solving a finite systems of linear algebraic equations.[3] The simplest situation occurs when there is only one eigenvalue k_1. In this case the scattering data reduce to the pair of complex numbers (k_1,c_1) and $u(x)$ is given by[3]

$$u(x) = -2i\lambda^*(x)\theta_2^*(x), \tag{B1}$$

where

$$\lambda(x) \equiv \left(\frac{c_1}{a_1'}\right)^{1/2} e^{ik_1x}, \quad a_1' \equiv \left.\frac{\partial a(k)}{\partial k}\right|_{k=k_1} = (2i\,\mathrm{Im}\{k_1\})^{-1}, \tag{B2}$$

and $\theta_2(x)$ is determined by means of the following system

$$\theta_1 + \frac{|\lambda|^2}{2i\,\mathrm{Im}\{k_1\}}\theta_2^* = 0, \quad \theta_2^* + \frac{|\lambda|^2}{2i\,\mathrm{Im}\{k_1\}} = \lambda^*. \tag{B3}$$

It is then easy to find that

$$u(x) = -2\mathrm{Im}\{k_1\}\cdot\exp[-i(2\mathrm{Re}\{k_1\}\cdot x + \arg c_1)] \cdot\mathrm{sech}[2\mathrm{Im}\{k_1\}x - \ln|c_1|]. \tag{B4}$$

Starting with this formula we get immediately $\psi(x) = (2/g)^{1/2}u((2\mu)^{-1/2}x)$ in terms of the variables (q,p,τ,m) related to (k_1,c_1) through (3.8a) and (3.8b).

[1] H. A. Bethe, Z. Phys. **71**, 205 (1931); C. N. Yang, Phys. Rev. **168**, 1920 (1968); J. B. McGuire, J. Math. Phys. **5**, 622 (1964).
[2] C. S. Gardner, J. M. Green, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[3] V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[4] M. J. Ablowitz, D. I. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 294 (1974).
[5] V. E. Zakharov and V. Manakov, Teoret. Mat. Fiz. **19**, 332 (1974).
[6] L. D. Faddeev, "A Hamiltonian Interpretation of the Inverse Scattering Method," in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer-Verlag, Berlin, 1980).
[7] A. C. Newell, "The Inverse Scattering Transform," in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer-Verlag, Berlin, 1980).
[8] L. Martínez Alonso, J. Math. Phys. **20**, 219 (1979).
[9] D. J. Kaup, "Solitons as Particles, and the Effects of Perturbations," in *The Significance of Nonlinearity in the Natural Sciences* (Plenum, New York, 1977).
[10] L. de Broglie, *Nonlinear Wave Mechanics* (Elsevier, New York, 1960).
[11] H. H. Chen and C. S. Liu, Phys. Rev. Lett. **37**, 693 (1976).
[12] A. C. Newell, Proc. R. Soc. London Ser. A **365**, 283 (1979).

An analysis of Taylor's theory of toroidal plasma relaxation

V. Faber, A. B. White, Jr., and G. M. Wing
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 4 November 1981; accepted for publication 19 March 1982)

The Taylor theory of toroidal plasma relaxation is considered as a nonlinear eigenvalue problem. The analysis is rigorous and applies to quite general toroidal cross sections. Emphasis is placed on the symmetric state case, where the existence of field reversal and flux free states is demonstrated. Certain anomalies are revealed by the mathematical treatment and their significance is studied. Existence of a solution to the Taylor problem in the symmetric state is proved and the location of the eigenvalue of this solution state relative to other states is examined. The question of the existence of helical states is not resolved, but it is shown that in many respects any helical states behave like the anomalous cases in the symmetric problem and hence do not significantly affect the theory.

PACS numbers: 52.55.Gb, 02.30. + g

I. INTRODUCTION

The theory of Taylor[1,2] has given valuable insight in describing experimentally observed phenomena in plasma relaxation in a z-pinch. For the most part investigations of this theory have been confined to a model which replaces the actual toroidal geometry by a more tractable approximating cylindrical geometry, although recent work of Reiman[3] does consider certain aspects of the theory for a torus of arbitrary cross section. It is our purpose to investigate a variety of questions associated with Taylor's theory from as rigorous a viewpoint as possible. In so doing we examine only the toroidal case and avoid any discussion of the cylindrical approximation. This is done, at least in part, because arguments based on the cylinder predict states which have never been observed experimentally. It, therefore, seems desirable to obtain as much information as possible about the Taylor theory in the realistic geometry. We do not succeed in proving or disproving the existence for the torus of the so-called "helical" states found in the cylinder. However, we are able to demonstrate a number of properties of the "symmetric" states which are unexpected. We also discuss the overall effect that any helical states might have on our results.

In the next section we pose the problem in complete generality and then discuss both those portions which will come under detailed investigation and those which remain items for speculation. Our reasons for emphasizing the symmetric states are outlined. In Sec. III we transform the problem from the customary partial differential equation form to an integral equation for one component of the B field. The properties of the solution to this equation and their physical meanings are discussed in the following two sections, and important expressions for flux and magnetic energy are derived using just this component. It is observed that the problem posed in Sec. II can be considered as the "intersection" of two somewhat easier problems. Each of these is then studied in detail, primarily in Sec. VI. A result of Reiman[3] is established in Sec. VII for symmetric states, and the existence of at least one such state which solves the Taylor Problem is established in the following section. Several rather surprising possibilities are revealed.

Section IX takes up the helical case and is primarily speculative. We do discuss some consequences of the existence of such states and call attention to the fact some of the anomalies they might introduce are somewhat analogous to the "surprising possibilities" mentioned above. We conclude with a summary and some suggestions for further research.

While the work of this paper (with the partial exception of Sec. IX) is mathematically rigorous, we shall avoid excessive mathematical detail, not entering into discussions of legitimacy of interchange of limiting processes and other such niceties. Ample references are provided for the reader who may wish to pursue such matters.

II. STATEMENT OF THE PROBLEM

We consider a torus T of arbitrary cross section Ω, with surface ∂T which is a perfect conductor. In the theory of Taylor the following equations must be satisfied (see Appendix I) (we consistently use an overbar to denote a vector):

$$\nabla\times\overline{B}=\mu\overline{B}, \tag{2.1}$$

$$\overline{B}\cdot\overline{n}=0 \text{ on } \partial T, \tag{2.2}$$

$$\iint_{\Omega}\overline{B}\cdot\overline{n}\,d\Omega=F_0>0, \tag{2.3}$$

$$\iiint_{T}\overline{A}\cdot\overline{B}\,dV=K_0>0, \tag{2.4}$$

$$\int_{C_T}\overline{A}\cdot d\overline{l}=A_T. \tag{2.5}$$

In (2.1), $\overline{B}=\overline{B}(r,z,\phi)$ (r,z,ϕ are cylindrical coordinates; see Fig. 1), the magnetic field inside T. The parameter μ is to be determined. (Clearly, $\overline{B}$ also depends on μ.) Equation (2.2) follows from the assumption that ∂T is a perfect conductor. The flux condition (2.3) states that the flux through any cross section Ω must be independent of ϕ; that is, independent of which cross section is considered. Equation (2.4) is the "magnetic helicity" constraint, $\overline{A}$ being the vector potential. It may be shown (Appendix A) that for $K_0>0$, only $\mu\geqslant 0$ is of

 0022-2488/82/081524-14$02.50

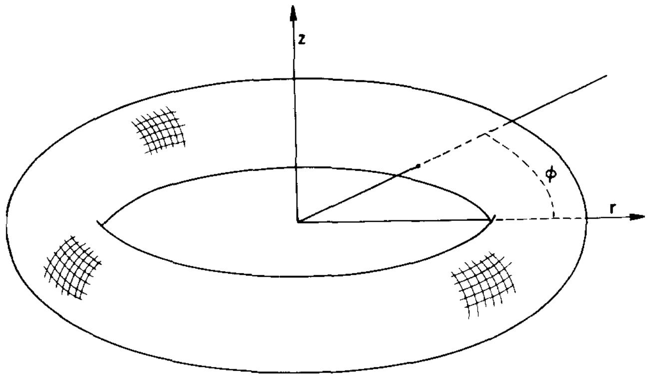


FIG. 1. Toroidal geometry.

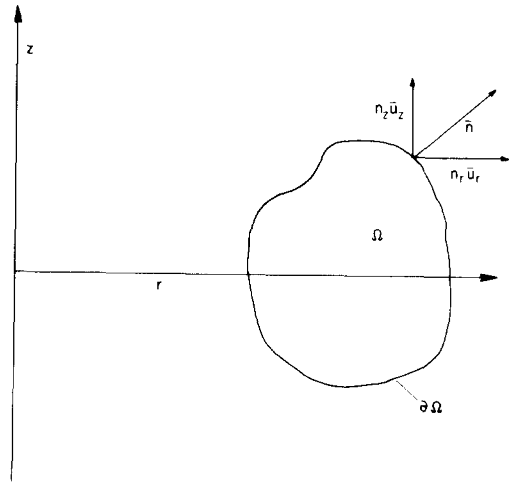


FIG. 2. Toroidal cross section.

interest. In (2.5) C_T is any closed path lying on ∂T which "encircles" the torus once but does not wrap around it. [For a physical interpretation of (2.5), see Ref. 3].

According to Taylor the relaxed plasma is associated with that $\overline{B}$ field subject to (2.1)–(2.5) and having lowest magnetic energy.

It is at once clear that (2.1)–(2.5) may or may not have a solution, though one expects that there is a (finite or infinite) set of μ's for which a solution does exist. If this is the case then one can hope to isolate that one (or those) which provide minimum energy.

Notation: We shall henceforth refer to the problem posed by (2.1)–(2.5) as TP (Taylor Problem).

Since $\overline{B}$ (r,z,ϕ) must be periodic in ϕ it is natural to attempt a formal expansion,

$$\overline{B}(r,z,\phi) = \sum_{l=-\infty}^{\infty} \overline{B}_l(r,z)e^{il\phi}. \tag{2.6}$$

Using this expansion and employing (2.3) we get

$$F_0 = \sum_{l=-\infty}^{\infty} e^{il\phi} \iint_{\Omega} \overline{B}_l(r,z)\cdot\bar{n}\, d\Omega. \tag{2.7}$$

Since F_0 is a constant, only the $l = 0$ term can appear in (2.7). Thus

$$\iint \overline{B}_l(r,z)\cdot\bar{n}\, d\Omega = 0, \quad l \neq 0. \tag{2.8}$$

That is, the states $\overline{B}_l$ are "flux free" for $l \neq 0$. These are the so-called helical states. That state, $\overline{B}_0$, which contributes all of the flux, is termed the symmetric state.

The foregoing should be considered as motivation. In fact, the existence of the helical states has not been established. We shall concentrate on $\overline{B}_0$ and prove that TP does indeed have a solution for $\overline{B}_0$. We shall also determine many of its properties. Because all further discussion (until Sec. IX) will involve this symmetric state we simplify notation and write simply $\overline{B}$ instead of $\overline{B}_0$.

Equation (2.1) may now be simplified:

$$\frac{-\partial B_\phi}{\partial z} = \mu B_r, \tag{2.9a}$$

$$\frac{\partial B_r}{\partial z} - \frac{\partial B_z}{\partial r} = \mu B_\phi, \tag{2.9b}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rB_\phi) = \mu B_z, \tag{2.9c}$$

where B_r, B_z, B_ϕ, all functions of just r and z, are the components of $\overline{B}$.

To understand (2.2) we note that we may study any cross section Ω we wish. For convenience, we choose that one in the plane $\phi = 0$. Since $\bar{n}$ can have no ϕ component we write (see Fig. 2)

$$\bar{n} = n_r\bar{u}_r + n_z\bar{u}_z, \tag{2.10}$$

and obtain

$$\overline{B}\cdot\bar{n} = n_rB_r + n_zB_z = 0 \quad \text{on } \partial\Omega. \tag{2.11}$$

From (2.9) this becomes

$$-n_r\frac{\partial B_\phi}{\partial z} + \frac{n_z}{r}\frac{\partial}{\partial r}(rB_\phi) = 0 \quad \text{on } \partial\Omega. \tag{2.12}$$

But

$$\nabla(rB_\phi) = \frac{\partial}{\partial r}(rB_\phi)\bar{u}_r + \frac{\partial}{\partial z}(rB_\phi)\bar{u}_z, \tag{2.13}$$

so that (2.12) can be written

$$[n_r\bar{u}_z - n_z\bar{u}_r]\cdot\nabla(rB_\phi) = \bar{t}\cdot\nabla(rB_\phi) = 0, \tag{2.14}$$

where $\bar{t}$ is a unit tangent to $\partial\Omega$. This states that rB_ϕ does not change in the tangential direction, or

$$B_\phi = c/r, \quad c \text{ any constant on } \partial\Omega. \tag{2.15}$$

We are now ready to proceed toward a solution of TP for the symmetric case.

III. TRANSFORMATIONS OF THE B PROBLEM

Some trivial manipulations applied to (2.9) result in

$$-\frac{\partial^2 B_\phi}{\partial z^2} - \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}(rB_\phi)\right) = \mu^2 B_\phi. \tag{3.1}$$

Set

$$\psi(r,z) = r^{1/2}B_\phi(r,z). \tag{3.2}$$

Then (3.1) becomes

$$L\psi = -\frac{\partial^2\psi}{\partial r^2} - \frac{\partial^2\psi}{\partial z^2} + \frac{3}{4r^2}\psi = \mu^2\psi. \tag{3.3}$$

Since r and z play the role of Cartesian coordinates in the plane ϕ = constant, (3.3) may be written

$$L\psi = -\nabla^2\psi + \frac{3}{4r^2}\psi = \mu^2\psi. \tag{3.4}$$

The boundary condition (2.15) becomes

$$\psi = c/\sqrt{r} \quad \text{on } \partial\Omega. \tag{3.5}$$

In Sec. V we shall find the form

$$\nabla\cdot\left\{\frac{1}{r}\nabla(rB_\phi)\right\} = -\mu^2 B_\phi \tag{3.6}$$

useful. Here, and elsewhere in the remainder of this paper, ∇ is the two-dimensional Cartesian gradient operator.

It is convenient to convert the differential system (3.4), (3.5) to an integral equation. Set

$$\psi(r,z) = \theta(r,z) + w(r,z) \tag{3.7}$$

and require

$$L\theta = \mu^2\psi \quad \text{on } \Omega, \tag{3.8a}$$

$$\theta = 0 \quad \text{on } \partial\Omega, \tag{3.8b}$$

and

$$Lw = 0, \tag{3.9a}$$

$$w = \frac{c}{\sqrt{r}} \quad \text{on } \partial\Omega. \tag{3.9b}$$

The solution to (3.9) is immediate,

$$w = c/\sqrt{r} \quad \text{on } \Omega. \tag{3.10}$$

To resolve (3.8) we consider

$$Lv = g \quad \text{on } \Omega, \tag{3.11a}$$

$$v = 0 \quad \text{on } \partial\Omega. \tag{3.11b}$$

It is known (see Ref. 4) that for $g \in L_2(\Omega)$ the solution to (3.11) is given by

$$v(r,z) = \iint_\Omega \Gamma(r,z;r',z')g(r',z')\, d\Omega', \tag{3.12}$$

where Γ is continuous on $\Omega \times \Omega$. Moreover, Γ is symmetric, positive definite, and pointwise positive on $\Omega - \partial\Omega$.

Thus, from (3.8) and (3.10)

$$\psi(r,z) = \frac{c}{\sqrt{r}} + \mu^2 \iint_\Omega \Gamma(r,z;r',z')\psi(r',z')\, d\Omega', \tag{3.13a}$$

which we shall often write as

$$\psi = \frac{c}{\sqrt{r}} + \mu^2\Gamma\psi. \tag{3.13b}$$

Almost all of our work will be based on (2.9), (3.6), and (3.13).

IV. SOME PROPERTIES OF THE SOLUTION OF THE INTEGRAL EQUATION

The structure of (3.13) and the specific properties of Γ allow a rather complete discussion of the properties of its solution, many of which have strong implications for the physics of the original problem. We shall state these with references to the classical literature.

It is convenient first to consider the corresponding eigenfunction problem, writing

$$\lambda\hat{\psi} = \Gamma\hat{\psi}. \tag{4.1}$$

From the symmetry and positive definiteness of Γ we find (see Refs. 5 and 6):

A. Equation (4.1) has a denumerably infinite set of eigenvalues λ_j, all positive. The corresponding eigenfunctions ψ_j may be taken orthonormal on Ω. The eigenvalues are of finite multiplicity. We write $0 < \cdots\lambda_n < \lambda_{n-1} < \cdots\lambda_2 < \lambda_1$ and $\lim_{n\to\infty}\lambda_n = 0$ and agree that to any λ_j there correspond finitely many eigenfunctions $\psi_j^{(1)}, \psi_j^{(2)},\ldots,\psi_j^{(m)}$.

B. Equation (3.13) is uniquely solvable for $c \neq 0$ provided $\mu^2\lambda_j \neq 1$ for all j. If $c \neq 0$ and $\mu^2\lambda_j = 1$ then (3.13) is solvable if and only if

$$0 = \left(\psi_j^{(k)}, \frac{c}{\sqrt{r}}\right) = c\iint_\Omega \frac{\psi_j^{(k)}(r,z)}{\sqrt{r}}\, d\Omega, \tag{4.2}$$

for *all* eigenfunctions $\psi_j^{(k)}$ belonging to λ_j.

The fact that $\Gamma(r,z;r',z')$ is pointwise positive on $\Omega - \partial\Omega$ allows us to say more (see Refs. 5 and 7).

C. To λ_1 there corresponds precisely *one* eigenfunction ψ_1 and it may be taken pointwise positive on $\Omega - \partial\Omega$.

From these properties of the operator Γ and its eigenfunctions we obtain:

Theorem 1a: If $\mu^2 < \lambda_1^{-1}$ then (3.13) with $c > 0$ has a unique positive solution on $\Omega - \partial\Omega$.

Proof: When $\mu^2 < \lambda_1^{-1}$ the solution may be written as a Neumann series

$$\psi = \frac{c}{\sqrt{r}} + \sum_{n=1}^{\infty}\mu^{2n}\Gamma^n\left(\frac{c}{\sqrt{r}}\right). \tag{4.3}$$

The result follows from the pointwise positivity of Γ.

Theorem 1b: If $\mu^2 = \lambda_1^{-1}$ and $c \neq 0$ then (3.13) has no solution.

Proof: By B and C there can be a solution only if $(\psi_1, c/\sqrt{r}) = 0$. Since $\psi_1 > 0$ on $\Omega - \partial\Omega$ this is impossible.

Theorem 1c: If $\mu^2 > \lambda_1^{-1}$, but $\mu^2 \neq \lambda_j^{-1}, j = 2,3,\ldots$, then (3.13) is uniquely solvable for $c \neq 0$ but the solution cannot be of one sign on $\Omega - \partial\Omega$.

Proof: The unique solvability follows from B. Take the inner product of ψ with ψ_1:

$$\begin{aligned}(\psi,\psi_1) &= c(\psi_1, 1/\sqrt{r}) + \mu^2(\Gamma\psi,\psi_1)\\ &= c(\psi_1, 1/\sqrt{r}) + \mu^2(\psi,\Gamma\psi_1)\\ &= c(\psi_1, 1/\sqrt{r}) + \mu^2\lambda_1(\psi,\psi_1).\end{aligned} \tag{4.4}$$

Here we have used the symmetry of Γ and the definition of ψ_1. From (4.4), and assuming, for convenience, that c is positive,

$$(\psi,\psi_1)[1 - \mu^2\lambda_1] = c(\psi_1, 1/\sqrt{r}) > 0. \tag{4.5}$$

Now $1 - \mu^2\lambda_1 < 0$. Thus $(\psi,\psi_1) < 0$ and so ψ, if it is of one sign on $\Omega - \partial\Omega$, must be negative. However, $\psi = c/\sqrt{r}$ on $\partial\Omega$ and, by continuity, ψ must be positive in a neighborhood of $\partial\Omega$. This is a contradiction.

We pause to note the physical significance of these results. For $\mu^2 \leqslant \lambda_1^{-1}$, the B_ϕ component is of one sign, but for $\mu^2 > \lambda_1^{-1}$, it cannot be. Thus "field reversal" occurs at the first eigenvalue λ_1.

Theorem 1d: If $\mu^2 \lambda_j = 1, j > 1$, and for one of the eigenfunctions $\psi_j^{(k)}$ we have $(\psi_j^{(k)}, 1/\sqrt{r}) \neq 0$, then (3.13) is solvable only if $c = 0$, in which case none of its solutions is of one sign on $\Omega - \partial\Omega$.

Proof: The result follows in part from B. If $c = 0$ then

$$\psi = \sum_{k=1}^{m} \alpha_k \psi_j^{(k)}, \tag{4.6}$$

where the α_k are arbitrary constants. Then, by orthogonality

$$(\psi, \psi_1) = \sum_{k=1}^{m} \alpha_k (\psi_j^{(k)}, \psi_1) = 0. \tag{4.7}$$

Since ψ_1 is of one sign, ψ cannot be.

Theorem 1e: If $\mu^2 \lambda_j = 1, j > 1$, and for all of the eigenfunctions $\psi_j^{(k)}$ we have $(\psi_j^{(k)}, 1/\sqrt{r}) = 0$, then (3.13) is solvable for $c \neq 0$. The solution is not unique. In fact, if ψ_P is a solution then so is

$$\psi = \psi_P + \sum_{k=1}^{m} \alpha_k \psi_j^{(k)},$$

where the α's are arbitrary. However, ψ cannot be of one sign.

Proof: Again ψ satisfies (3.13) and the proof of Theorem 1c applies.

We resist the temptation to summarize these results in one massive theorem. The important physical observation is that for $\mu^2 > \lambda_1^{-1}$ field reversal always occurs.

V. THE SOLUTION TO TWO "SUBPROBLEMS"

We are now in a position to attack the problem posed in Sec. 1. To this end we consider two subproblems:

P_1: the problem posed by (2.1), (2.2), and (2.3) with no constraint on the helicity or upon $\bar{A}$,

P_2: the problem posed by (2.1), (2,2), (2.4), and (2.5) with no constraint on the flux.

We shall eventually find for each of P_1 and P_2 an expression for the total magnetic field energy as a function of μ. If we denote these energies by $\widehat{W}_{\mathrm{I}}(\mu)$ and $\widehat{W}_{\mathrm{II}}(\mu)$, then any value $\hat{\mu}$ such that $\widehat{W}_{\mathrm{I}}(\hat{\mu}) = \widehat{W}_{\mathrm{II}}(\hat{\mu})$ clearly satisfies both P_1 and P_2, and hence such $\hat{\mu}$'s are candidates to solve the minimum energy problem.

A. Consideration of P_1

We first turn to the solution of (3.13) in terms of the eigenfunctions ψ_j. It is convenient at this point to reindex the eigenvalues so that if more than one eigenfunction belongs to a λ_j then that λ_j is repeated with a different index as often as needed. The eigenfunctions will be similarly indexed.

We suppose $c \neq 0$. The unique solution to (3.13) may be written (see Refs. 5 and 6)

$$\psi(r,z) = \frac{c}{\sqrt{r}} + c\mu^2 \sum_{j=1}^{\infty} \frac{\lambda_j (\psi_j, 1/\sqrt{r}) \psi_j(r,z)}{1 - \mu^2 \lambda_j}, \tag{5.1}$$

provided $\mu^2 \lambda_j \neq 1$ for all j. If $\mu^2 \lambda_j = 1$ for some $j = j_0$, then there is no solution to (3.13) unless B of Sec. IV applies. In our current notation that implies $\lambda_{j_0} = \lambda_{j_0+1} = \cdots = \lambda_{j_0+m-1}$ and $(1/\sqrt{r}, \psi_j) = 0, j = j_0, j_0 + 1, \ldots, j_0 + m - 1$. In that event (5.1) is replaced with

$$\psi(r,z) = \frac{c}{\sqrt{r}} + c\mu^2 \sum_{j=1}^{\infty}{}' \frac{\lambda_j (\psi_j, 1/\sqrt{r}) \psi_j(r,z)}{1 - \mu^2 \lambda_j} + \sum_{j=j_0}^{j_0+m-1} \alpha_j \psi_j(r,z). \tag{5.2}$$

Here the prime indicates that the indices $j_0, j_0 + 1, \ldots, j_0 + m - 1$ are deleted from the sum, and the α_j are arbitrary constants.

We again modify notation by isolating those subsets of $\{\lambda_j\}$ such that λ_{j_0} is of m-fold multiplicity and all corresponding eigenfunctions are orthogonal to $1/\sqrt{r}$. We remove these subsets from the λ_j sequence and call the remaining (appropriately reindexed) sequence $\{\Lambda_j\}$, with corresponding eigenfunctions ψ_j. Equations (5.1) and (5.2) become

$$\psi(r,z) = \frac{c}{\sqrt{r}} + c\mu^2 \sum_{j=1}^{\infty} \frac{\Lambda_j (\psi_j, 1/\sqrt{r}) \psi_j(r,z)}{1 - \mu^2 \Lambda_j} \tag{5.3}$$

and

$$\psi(r,z) = \frac{c}{\sqrt{r}} + c\mu^2 \sum_{j=1}^{\infty} \frac{\Lambda_j (\psi_j, 1/\sqrt{r}) \psi_j(r,z)}{1 - \mu^2 \Lambda_j} + \sum_{j=j_0}^{j_0+m-1} \alpha_j \psi_j(r,z). \tag{5.4}$$

We pause momentarily to comment on our preoccupation with ψ, and hence with B_ϕ. As we shall see, it is this component of $\bar{B}$ which plays the primary role in the theory. The components B_r and B_z can be obtained via (2.9). We shall later need the fact that B_r and B_z are zero when $\mu = 0$, a result which can easily be established using (5.3) and (5.4).

To find the flux associated with (5.3) and (5.4) note that

$$\begin{aligned} \text{Flux} &= F(\mu) = \iint_\Omega \bar{B} \cdot \bar{n}\, d\Omega \\ &= \iint_\Omega B_\phi d\Omega = \iint_\Omega \frac{\psi(r,z)}{\sqrt{r}} d\Omega \\ &= cF_1(\mu), \end{aligned} \tag{5.5}$$

where

$$F_1(\mu) = \iint_\Omega \frac{d\Omega}{\sqrt{r}} + \mu^2 \sum_{j=1}^{\infty} \frac{\Lambda_j (\psi_j, 1/\sqrt{r})^2}{1 - \mu^2 \Lambda_j}. \tag{5.6}$$

The behavior of this function is important in subsequent discussions. We note that $F_1(\mu)$ is meromorphic in μ with simple poles at $\mu = \Lambda_j^{-1/2}$. Moreover, for $\mu > 0$,

$$F_1'(\mu) = \mu \sum_{j=1}^{\infty} \frac{(\psi_j, 1/\sqrt{r})^2}{(1 - \mu^2 \Lambda_j)^2} > 0. \tag{5.7}$$

Thus F_1 is an increasing function on any domain in which it is continuous.

A schematic graph of $F_1(\mu)$ is shown in Fig. 3. We note that F_1 first becomes infinite at $\Lambda_1^{-1/2}$, which is actually

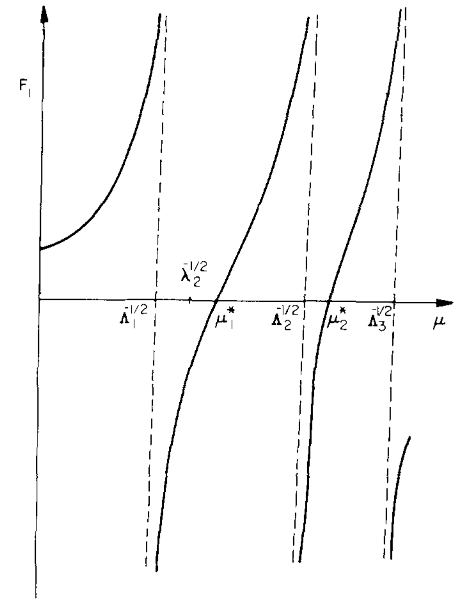


FIG. 3. Graph of $F_i(\mu)$.

$\lambda_1^{-1/2}$ (since ψ_1 cannot be orthogonal to $1/\sqrt{r}$). We also observe that between $\Lambda_1^{-1/2}$ and $\Lambda_2^{-1/2}$ $F_1(\mu)$ has a single simple zero, denoted μ_1^*. This represents a "flux free state." In general an infinite set of Λ_j's can be expected, and hence there are usually infinitely many flux free states, μ_j^*, $\Lambda_j^{-1/2} < \mu_j^* < \Lambda_{j+1}^{-1/2}$. It is shown in Appendix B that there is always one such. These states are also important for the subsequent discussion.

As yet we have not imposed the condition of constant flux [see (2.3)]. In order that

$$F(\mu) \equiv F_0 > 0 \tag{5.8}$$

it is necessary to choose [see (5.6)]

$$c = c(\mu) = \frac{F_0}{F_1(\mu)}. \tag{5.9}$$

Obviously c is not defined at the μ_j^*. The physical meaning of this is clear. There is no way to achieve a nonzero flux at a flux free state.

We note that throughout this discussion we have supposed $c \neq 0$. In case Theorem 1b or 1d applies we must choose $c = 0$. In that event $\mu = \Lambda_{j_0}^{-1/2}$ and

$$\psi(r,z) = \sum_{k=j_0}^{j_0+m-1} \beta_k \psi_k(r,z), \tag{5.10}$$

where the ψ's are the eigenfunctions belonging to Λ_{j_0} and the β's are arbitrary constants. The flux condition becomes

$$F_0 = \sum_{k=j_0}^{k_0+m-1} \beta_k (\psi_k, 1/\sqrt{r}). \tag{5.11}$$

Since at least one of the terms $(\psi_k, 1/\sqrt{r})$ is not zero, (5.11) always has a solution. If more than one such term is nonzero, the solution is not unique.

Finally, we compute the energy in the ϕ component of the magnetic field,

$$\begin{aligned} W_\phi(\mu) &= \frac{1}{2}\iiint_T B_\phi^2\, dV \\ &= \frac{1}{2}\int_0^{2\pi} d\phi \iint_\Omega B_\phi^2\, r dr dz \\ &= \pi \iint_\Omega \psi^2\, d\Omega. \end{aligned} \tag{5.12}$$

Assuming (5.3) holds

$$W_\phi(\mu) = c\pi \iint_\Omega d\Omega \left\{ \frac{1}{\sqrt{r}} + \mu \sum_{j=1}^{\infty} \frac{\Lambda_j(\psi_j, 1/\sqrt{r})\psi_j(r,z)}{1 - \mu^2 \Lambda_j} \right\}^2 \tag{5.13}$$

which yields, after considerable computation (see Appendix C),

$$\begin{aligned} W_\phi(\mu) &= c^2\pi \left\{ F_1(\mu) + \mu^2 \sum_{j=1}^{\infty} \frac{\Lambda_j(\psi_j, 1/\sqrt{r})^2}{(1 - \mu^2\Lambda_j)^2} \right\} \\ &\equiv c^2 W_1(\mu). \end{aligned} \tag{5.14}$$

If (5.4) holds then (5.14) must be augmented by a finite sum

$$c^2 \sum_{j=j_0}^{j_0+m-1} \alpha_j^2. \tag{5.15}$$

Since the α_j's are arbitrary, W_ϕ can obviously be made arbitrarily large (for fixed c) at $\mu = \lambda_{j_0}^{-1/2}$.

Finally, if (5.10) holds *and* if more than one term $(\psi_k, 1/\sqrt{r})$ is nonzero, then again W_ϕ can be made arbitrarily large.

More will be said about W_ϕ in the following section. This completes our discussion of Problem P_1.

B. Consideration of P_2

In the discussion of P_1 it was not necessary to introduce the vector potential $\bar{A}$. We must now discuss it. The equations

$$\operatorname{curl} \bar{B} = \mu \bar{B} \tag{5.16a}$$

and

$$\bar{B} = \operatorname{curl} \bar{A} \tag{5.16b}$$

suggest simply taking $\bar{A} = \bar{B}/\mu$. This choice, however, violates the constraint provided by (2.5).

To overcome this we select

$$A_\phi = B_\phi/\mu - c/\mu r, \tag{5.17a}$$

$$A_r = B_r/\mu, \tag{5.17b}$$

$$A_z = B_z/\mu. \tag{5.17c}$$

Here we imply the condition $rB_\phi = c$ on $\partial\Omega$ as in the discussion of P_1. In (2.5) we first choose C_T as a circle on ∂T such that r and z are constant. Then

$$\begin{aligned} \int_{C_T} \bar{A}\cdot d\bar{l} &= \int_0^{2\pi} A_\phi r\, d\phi \\ &= \int_0^{2\pi} \left[\frac{B_\phi}{\mu} - \frac{c}{\mu r} \right] r\, d\phi = 0. \end{aligned} \tag{5.18}$$

Thus the constraint (2.5) is satisfied for this path provided $A_T = 0$. However, the integral in (5.18) is actually independent of (admissable) paths C_T, according to Stokes' Theorem for multiply connected regions. Thus, (5.17) is a satisfactory vector potential. That $\overline{A}$ has been chosen in such a way as to make the value of the integral in (5.18) zero is no loss of generality (see Appendices D and I).

We may now calculate the helicity K [see (2.4)].

$$\begin{aligned} K(\mu) &= \iiint_T \overline{B}\cdot\overline{A}\, dV \\ &= \iiint_T \frac{\overline{B}\cdot\overline{B}}{\mu}\, dV - \frac{c}{\mu}\iiint_T \frac{B_\phi(r,z)}{r}\, dV \\ &= \frac{2\widehat{W}(\mu)}{\mu} - \frac{c}{\mu}\int_0^{2\pi} d\phi \iint_\Omega \frac{B_\phi}{r}\, r dr\, dz \\ &= \frac{2\widehat{W}(\mu)}{\mu} - \frac{2\pi c^2}{\mu} F_1(\mu), \end{aligned} \tag{5.19}$$

where $F_1(\mu)$ is given by (5.6) and

$$\widehat{W}(\mu) = \frac{1}{2}\iiint_T \overline{B}\cdot\overline{B}\, dV. \tag{5.20}$$

It is to be noted that $\widehat{W}$ is the total magnetic energy, not just the W_ϕ contribution. An important relation between $\widehat{W}$ and W_ϕ will be derived in the next section.

Finally, we impose the condition that $K(\mu) = K_0 > 0$, but note that this yields little information at this point.

VI. THE TOTAL MAGNETIC ENERGY

In this section, we shall compute the total energy $\widehat{W}$ of the general (symmetric) solution to TP as a function of μ. First let us show that $\widehat{W}$ is dependent only upon the energy in the ϕ component of the magnetic field and on the flux.

Theorem 2: $\widehat{W}(\mu) = 2W_\phi(\mu) - \pi c F(\mu)$, where $B_\phi = c/r$, c arbitrary, on $\partial\Omega$.

Proof: From (3.6), we have

$$\nabla\cdot \frac{1}{r}\nabla(rB_\phi) = -\mu^2 B_\phi. \tag{6.1}$$

Now

$$\nabla\cdot\Big[\frac{1}{r}(rB_\phi)\nabla rB_\phi\Big] = rB_\phi\Big[\nabla\cdot\frac{1}{r}\nabla rB_\phi\Big] + \frac{1}{r}\nabla rB_\phi\cdot\nabla rB_\phi. \tag{6.2}$$

Multiply (6.1) by c and integrate over Ω using (6.2) and the divergence theorem:

$$\begin{aligned} -\mu^2 cF(\mu) &= \int_{\partial\Omega} \frac{c}{r}\nabla(rB_\phi)\cdot\overline{n}\, ds \\ &= \int_{\partial\Omega} B_\phi \nabla(rB_\phi)\cdot\overline{n}\, ds = \iint_\Omega \nabla\cdot(B_\phi\nabla(rB_\phi))\, d\Omega \\ &= \iint_\Omega rB_\phi\Big[\nabla\cdot\frac{1}{r}\nabla rB_\phi\Big] d\Omega \\ &\quad + \iint_\Omega \frac{1}{r}(\nabla rB_\phi)\cdot(\nabla rB_\phi)\, d\Omega \\ &= -\mu^2\iint_\Omega rB_\phi^2 d\Omega \\ &\quad + \iint_\Omega \frac{1}{r}(\nabla rB_\phi)\cdot(\nabla rB_\phi)\, d\Omega. \end{aligned} \tag{6.3}$$

Now integrate (6.3) with respect to ϕ, use (2.9), and note that $d\Omega d\phi = dr\, dz\, d\phi = (1/r)dV$.

$$\begin{aligned} -2\pi cF(\mu) &= \iiint_T (-\mu^2 B_\phi^2) dV + \iiint_T \Big(\frac{1}{r}\nabla rB_\phi\Big)^2 dV \\ &= -\mu^2\iiint_T B_\phi^2 dV + \iiint_T \frac{1}{r^2}\Big[\Big(\frac{\partial}{\partial r}(rB_\phi)\Big)^2 \\ &\quad + \Big(\frac{\partial}{\partial z}(rB_\phi)\Big)^2\Big] dV \\ &= -\mu^2\iiint_T B_\phi^2 dV + \mu^2\iiint_T (B_r^2 + B_z^2) dV \\ &= \mu^2(2\widehat{W}(\mu) - 4W_\phi(\mu)). \end{aligned} \tag{6.4}$$

If $\mu \neq 0$, we may rearrange (6.4) to achieve the desired result:

$$\widehat{W}(\mu) = 2W_\phi(\mu) - \pi cF(\mu). \tag{6.5}$$

By (5.14), $W_\phi(0) = c\pi F(0)$. From the remarks immediately following (5.4), $B_r = B_z = 0$ at $\mu = 0$. Thus $\widehat{W}(0) = W_\phi(0) = c\pi F(0)$, so (5.6) holds at $\mu = 0$. This concludes the proof of Theorem 2.

Now we compute the total energy for each of the two problems P_1 and P_2. First let us consider P_1, that is, the flux is given to be $F_0 > 0$. As we have previously mentioned, there is no solution if $\mu = \mu_j^*$. In order to determine the total energy from (6.5), one must evaluate the constant c and the energy W_ϕ. If $\mu^2\lambda_j \neq 1$ for all j, then by (5.9), $c = F_0/F_1(\mu)$. We have from (5.14) that

$$W_\phi(\mu) = c^2 W_1(\mu) = \frac{F_0^2 W_1(\mu)}{F_1^2(\mu)}. \tag{6.6}$$

Thus (6.5) becomes

$$\widehat{W}(\mu) = \frac{F_0^2}{F_1^2(\mu)}(2W_1(\mu) - \pi F_1(\mu)); \quad \mu^2\lambda_j \neq 1 \text{ for all } j. \tag{6.7}$$

If $\mu^2\lambda_j = 1$, we have two alternatives. The first alternative, the situation in Theorem 1e, leads to Eq. (5.2) and the result that the flux is given by (5.5) again. In this case (5.9) still applies, but (6.6) must be augmented by the sum [see (5.15)]

$$\sum_{j=j_0}^{j_0+m-1} \alpha_j^2, \tag{6.8}$$

α_j arbitrary. Consequently, at such a $\mu = \lambda_{j_0}^{-1/2}$, the total energy has a fixed minimum value $\widehat{W}_m(\lambda_{j_0}^{-1/2})$ and can have any arbitrary larger value. We note that

$$\hat{W}_m(\lambda_{j_0}^{-1/2}) = \frac{F_0^2}{F_1^2(\lambda_{j_0}^{-1/2})}\,[2W_1(\lambda_{j_0}^{-1/2}) - \pi F_1(\lambda_{j_0}^{-1/2})]. \tag{6.9}$$

The second alternative, the situation in Theorem 1d, $\mu = \Lambda_{j_0}^{-1/2}$, leads to Eqs. (5.10) and (5.11). The energy W_ϕ associated with this solution is then given by

$$W_\phi(\mu) = \frac{1}{2}\iiint_T B_\phi^2\,dV = \frac{1}{2}\int_0^{2\pi} d\phi \iint_\Omega B_\phi^2(r,z)\,r\,dr\,dz$$

$$= \pi \iint_\Omega \psi^2 d\Omega = \pi \sum_{j=j_0}^{j_0+m-1} \beta_j^2, \tag{6.10}$$

subject to (5.11). Consequently, at such a $\mu = \Lambda_{j_0}^{-1/2}$, the total energy has a fixed minimum value $\hat{W}_m(\Lambda_{j_0}^{-1/2})$ and can have any arbitrary larger value unless the eigenvalue Λ_{j_0} has multiplicity one. In this case the energy has a unique value. (Note that, in fact, this is the case at $\mu = \hat{\Lambda}_1^{-1/2}$). A standard minimization technique gives the value for $\hat{W}_m$ (see Appendix E).

$$\hat{W}_m(\Lambda_{j_0}^{-1/2}) = 2\pi F_0^2 \left\{ \sum_{j=j_0}^{j_0+m-1} (\psi_j, 1/\sqrt{r})^2 \right\}^{-1}. \tag{6.11}$$

From Appendix F, we have

$$\lim_{\mu \to \Lambda_{j_0}^{-1/2}} \hat{W}(\mu) = \lim_{\mu \to \Lambda_{j_0}^{-1/2}} \frac{F_0^2}{F_1^2(\mu)}(2W_1(\mu) - \pi F_1(\mu)) = \hat{W}_m(\Lambda_{j_0}^{-1/2}). \tag{6.12}$$

Finally, we note that [see (5.6) and 6.6)]

$$\hat{W}(0) = \frac{\pi F_0^2}{F_1(0)} = \pi F_0^2 \left\{ \iint_\Omega \frac{d\Omega}{\sqrt{r}} \right\}^{-1}. \tag{6.13}$$

In Fig. 4, we show a schematic graph of $\hat{W}$ as a function of μ with fixed flux F_0. The graph summarizes the details of (6.7), (6.9), (6.12), and (6.13).

We have assumed for illustrative purposes that λ 's associated with the second alternatives do occur (vertical arrows). Their existence and location depend, of course, on the specific problem under investigation.

We turn now to the energy in problem P_2. We have from (5.19)

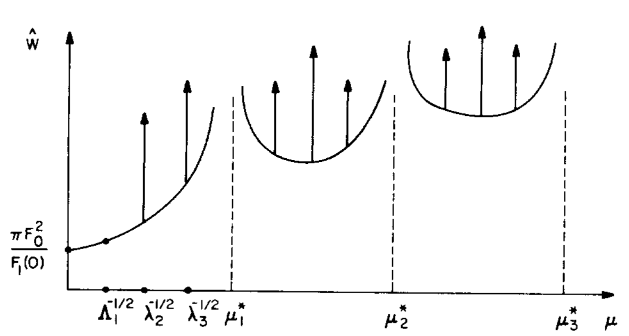


FIG. 4. Energy with fixed nonzero flux F_0.

$$\hat{W}(\mu) = \tfrac{1}{2}\mu K_0 + \pi c F(\mu) = \tfrac{1}{2}\mu K_0 + \pi c^2 F_1(\mu), \tag{6.14}$$
$$c = rB_\phi \quad \text{on } \partial\Omega.$$

If $\mu^2\lambda_j \neq 1$ for all j, the ϕ component of the energy is given by (6.6). From Theorem 2, the total energy is therefore

$$\hat{W}(\mu) = c^2(2W_1(\mu) - \pi F_1(\mu)). \tag{6.15}$$

Combine (6.14) and (6.15) and solve for c^2 to get

$$c^2 = \tfrac{1}{4}\mu K_0[W_1(\mu) - \pi F_1(\mu)]^{-1}. \tag{6.16}$$

From (6.15)

$$\hat{W}(\mu) = \tfrac{1}{2}\mu K_0\left[1 + \frac{\pi F_1(\mu)}{W_1(\mu) - \pi F_1(\mu)}\right]. \tag{6.17}$$

If $\mu^2\lambda_{j_0} = 1$, we again have two alternatives. The first alternative, the situation in Theorem 1e, leads once more to Eq. (5.2) and the result that the flux is given by (5.5), while the ϕ component of the energy is

$$W_\phi(\mu) = c^2 W_1(\mu) + \sum_{j=j_0}^{j_0+m-1} \alpha_j^2, \quad \alpha_j \text{ arbitrary}. \tag{6.18}$$

From Theorem 2, the total energy is thus

$$\hat{W}(\mu) = c^2(W_1(\mu) - \pi F_1(\mu)) + \sum_{j=j_0}^{j_0+m-1} \alpha_j^2. \tag{6.19}$$

Combine (6.19) and (6.14) to get

$$2c^2(W_1(\mu) - \pi F_1(\mu)) + \sum_{j=j_0}^{j_0+m-1} \alpha_j^2 = \tfrac{1}{2}\mu K_0. \tag{6.20}$$

From (5.14), $W_1(\mu) - \pi F_1(\mu)$ is a positive quantity. From this fact and (6.20), we see that the possible range of values of c^2 is

$$0 \leqslant c^2 \leqslant \tfrac{1}{4}\mu K_0[W_1 - \pi F_1]^{-1}. \tag{6.21}$$

It follows from (6.14) that if we set

$$m = \min\left\{\tfrac{1}{2}\mu K_0, \tfrac{1}{2}\mu K_0\left[1 + \frac{\pi F_1}{2(W_1 - \pi F_1)}\right]\right\} \tag{6.22a}$$

and

$$M = \max\left\{\tfrac{1}{2}\mu K_0, \tfrac{1}{2}\mu K_0\left[1 + \frac{\pi F_1}{2(W_1 - \pi F_1)}\right]\right\}, \tag{6.22b}$$

then

$$m \leqslant \hat{W}(\mu) \leqslant M. \tag{6.23}$$

It is important to note that

$$1 + \frac{\pi F_1}{2(W_1 - \pi F_1)} = \frac{1}{2} + \frac{W_1}{2(W_1 - \pi F_1)} > \frac{1}{2}, \tag{6.24}$$

so that $m \geqslant \tfrac{1}{4}\mu K_0 > 0$.

The second alternative, the situation in Theorem 1d, leads to $c = 0$, so from (6.14)

$$\hat{W}(\mu) = \tfrac{1}{2}\mu K_0, \quad \mu = \Lambda_j^{-1/2}. \tag{6.25}$$

Finally, we require the behavior of $\hat{W}(\mu)$ as $\mu \to 0^+$. From (6.17) we see that

$$\lim_{\mu \to 0^+} \hat{W}(\mu) = \lim_{\mu \to 0^+} \frac{K_0}{2}\,\pi\,\frac{\mu F_1(\mu)}{W_1(\mu) - \pi F_1(\mu)}. \tag{6.26}$$

From (5.14) we see that the denominator behaves like μ^2 for μ near zero. Since $F_1(0) > 0$ we obtain

$$\lim_{\mu \to 0^+} \hat{W}(\mu) = +\infty. \tag{6.27}$$

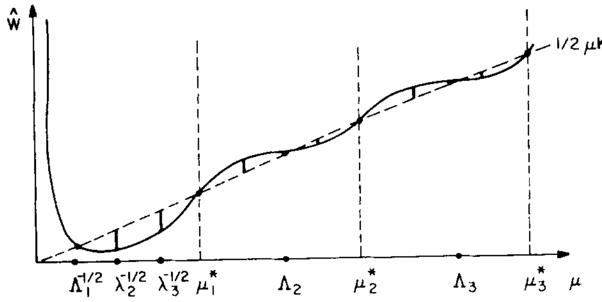


FIG. 5. Energy with fixed helicity K_0.

In Fig. 5, we show a schematic graph of $\hat{W}$ as a function of μ with fixed helicity K_0 which summarizes the details (6.16), (6.19), (6.23), (6.25), and (6.27) of the energy in problem P_2. (We suppose the same locations for "exceptional" λ 's as in Fig. 4.)

VII. THE RELATIONSHIP BETWEEN ANY TWO SOLUTIONS OF TP

Suppose $\bar{B}_1$ and $\bar{B}_2$ are two solutions to TP with $rB_{1\phi} = c_1$ and $rB_{2\phi} = c_2$ on $\partial\Omega$. (We shall show in Sec. VIII that there is at least one solution.) We have [using (2.9)]

$$
\begin{aligned}
(\mu_1-\mu_2)\iiint_T \bar{B}_1\cdot\bar{B}_2\,dV
&= \iiint_T [\bar{B}_2\cdot\nabla\times\bar{B}_1 - \bar{B}_1\cdot\nabla\times\bar{B}_2]\,dV \\
&= 2\pi\iint_\Omega \nabla\cdot(\bar{B}_1\times\bar{B}_2)\,r\,d\Omega \\
&= 2\pi\int_{\partial\Omega} r\bar{B}_1\times\bar{B}_2\cdot\bar{n}\,ds \\
&= 2\pi\int_{\partial\Omega} r[B_{1\phi}(B_{2z}\bar{u}_r - B_{2r}\bar{u}_z) \\
&\quad + B_{2\phi}(B_{1r}\bar{u}_z - B_{1z}\bar{u}_r)]\cdot\bar{n}\,ds \\
&= 2\pi\int_{\partial\Omega}\Big[\frac{1}{\mu_2}rB_{1\phi}\Big(\frac{1}{r}\nabla rB_{2\phi}\Big) \\
&\quad - \frac{1}{\mu_1}rB_{2\phi}\Big(\frac{1}{r}\nabla rB_{1\phi}\Big)\Big]\cdot\bar{n}\,ds \\
&= \frac{2\pi}{\mu_2}\int_{\partial\Omega}\frac{c_1}{r}\nabla rB_{2\phi}\cdot\bar{n}\,ds \\
&\quad - \frac{2\pi}{\mu_1}\int_{\partial\Omega}\frac{c_2}{r}\nabla rB_{1\phi}\cdot\bar{n}\,ds.
\end{aligned}
\tag{7.1}
$$

From the flux formula (6.3), we have, assuming that $\bar{B}_1$ and $\bar{B}_2$ have the same flux F_0,

$$(\mu_1-\mu_2)\iiint_T \bar{B}_1\cdot\bar{B}_2\,dV = (2\pi\mu_1c_2 - 2\pi\mu_2c_1)F_0. \tag{7.2}$$

Let the total energy associated with $\bar{B}_1$ and $\bar{B}_2$ be $\hat{W}_1$ and $\hat{W}_2$, respectively. We have, for $\bar{B}_1\neq\bar{B}_2$,

$$
\begin{aligned}
0<D &= \frac{1}{2}\iiint_T(\bar{B}_1-\bar{B}_2)^2\,dV \\
&= \frac{1}{2}\iiint_T(\bar{B}_1^2+\bar{B}_2^2-2\bar{B}_1\cdot\bar{B}_2)\,dV \\
&= \hat{W}_1+\hat{W}_2-\iiint_T\bar{B}_1\cdot\bar{B}_2\,dV.
\end{aligned}
\tag{7.3}
$$

Thus

$$(\mu_1-\mu_2)(\hat{W}_1-\hat{W}_2) = (\mu_1-\mu_2)D + 2\pi\mu_1c_2F_0 - 2\pi\mu_2c_1F_0. \tag{7.4}$$

If we further assume that K is constant, K_0, we have from (6.14) that

$$\pi c_iF_0 = \hat{W}_i - \tfrac{1}{2}\mu_iK_0,\quad i=1,2. \tag{7.5}$$

Substituting (7.5) into (7.4) yields

$$(\mu_1-\mu_2)(\hat{W}_1+\hat{W}_2) = (\mu_1-\mu_2)D + 2\mu_1\hat{W}_2 - 2\mu_2\hat{W}_1. \tag{7.6}$$

Hence, if $0\leqslant\mu_2<\mu_1$,

$$\hat{W}_1 = \frac{1}{2}\,\frac{\mu_1-\mu_2}{\mu_1+\mu_2}\iiint_T(\bar{B}_1-\bar{B}_2)^2\,dV + \hat{W}_2, \tag{7.7}$$

so that

$$\hat{W}_1 > \hat{W}_2 \tag{7.8}$$

Theorem 3: If $\bar{B}_1$ and $\bar{B}_2$ are two different symmetric solutions to TP, the one associated with the smaller μ is the one with the smaller energy.

Theorem 3 has been proved by Reiman.[3] While the basic ideas of our proof are similar to his, the details are rather different. Reiman's result, however, applies also to the nonsymmetric states. (For further comment, see Sec. 9).

VIII. EXISTENCE OF A SOLUTION TO TP

Theorem 4: There exists a symmetric solution to the problem TP in the interval $(0,\mu_1^*)$.

Proof: Consider the two curves $\hat{W}_{\rm I} = (F_0^2/F_1^2)\times(2W_1-\pi F_1)$, $0<\mu<\mu_1^*$, represented as the single-valued part of the first branch of the graph in Fig. 4, and $\hat{W}_{\rm II} = \tfrac{1}{2}\mu K_0[1+\pi F_1/(W_1-\pi F_1)]$, $0<\mu<\mu_1^*$, represented as the single-valued part of the first branch of the graph in Fig. 5. Since both $W_{\rm I}$ and $W_{\rm II}$ are continuous on $0<\mu<\mu_1^*$, the function $W_{\rm I}-W_{\rm II}$ is continuous on $(0,\mu_1^*)$. Also note that $\lim_{\mu\to0^+}(W_{\rm I}-W_{\rm II}) = -\infty$, while $\lim_{\mu\to\mu_1^{*-}}(W_{\rm I}-W_{\rm II}) = +\infty$, so $W_{\rm I}-W_{\rm II}$ must take the value zero in the interval $(0,\mu_1^*)$.

Suppose $\hat{W}_{\rm I}(\tilde{\mu}) = \hat{W}_{\rm II}(\tilde{\mu}) = \hat{W}$. If $\tilde{\mu}^2\lambda_j\neq1$ for all j, the energy uniquely determines the field $\bar{B}$. To see this note that $\hat{W}$ uniquely determines $F_0^2/F_1^2 = c^2$, by (6.7). But $\bar{B}$ is completely specified by the condition $rB_\phi = c$ on $\partial\Omega$. Since changing the sign of c merely replaces $\bar{B}$ by $(-\bar{B})$, we may always assume $c>0$. Hence $\bar{B}$ is determined. Next the flux must be F_0 (since $\bar{B}$ solves P_1) and K must have the given value K_0 (since $\bar{B}$ solves P_2).

If $\tilde{\mu}^2 \lambda_{j_0} = 1$, we have two alternatives. The first, the situation in Theorem 1e, leads to Eqs. (6.9) and (6.23). We see by the discussions that lead to these equations, that in both problems P_1 and P_2, there is a unique vector $\overline{B}$ which provides the solution which lies on the single-valued branch of the energy curves at this point, namely, the solution found by setting $\alpha_j = 0, j_0 \leqslant j \leqslant j_0 + m - 1$, in Eq. (5.15). As before, this solution has the desired flux and helicity.

The second alternative, the situation in Theorem 1d, leads to Eqs. (6.11) and (6.25). From the discussion preceding (6.11), there may be more than one solution with the correct flux and energy $\widehat{W}_1$, but since *all* solutions have $c = 0$, Eq. (6.14) shows that they *all* have the desired helicity. This concludes the proof of Theorem 4.

It should be observed that in *no* case have we proved that there is only one solution to TP in $0 < \mu < \mu_1^*$.

For further discussion we refer to Appendix G.

IX. HELICAL SOLUTIONS

We return to Sec. II and consider the formal expansion (2.6). This can be written

$$\overline{B}(r,z,\phi) = \overline{B}_0(r,z) + \overline{B}_h(r,z,\phi), \tag{9.1}$$

where $\overline{B}_0$ is the symmetric state (we now revert to the original notation) and $\overline{B}_h$ is a sum of helical states. Note that $\overline{B}_h$ has the following properties [see (2.8)]:

$$\iint_\Omega \overline{B}_h \cdot \bar{n} d\Omega = 0, \tag{9.2a}$$

$$\iiint_T \overline{B}_0 \cdot \overline{B}_h dV = 0, \tag{9.2b}$$

$$\int_{C_T} \overline{B}_h \cdot d\bar{l} = 0. \tag{9.2c}$$

[Actually, the expansion (2.6) may be bypassed completely by simply decomposing $\overline{B}$ into two parts $\overline{B}_0$ and $\overline{B}_h$ with properties (9.2). However, (2.6) provides motivation and is consistent with the literature.]

From (9.2b) it follows that the total energy in the magnetic field $\overline{B}$ is

$$W = \widehat{W} + W_h, \tag{9.3}$$

where $\widehat{W}$ is the energy in the symmetric state and W_h that in the helical states.

We now point out the modifications to Secs. VI, VII, and VIII, which are necessary in the event that $\overline{B}$ contains a (nonzero) helical part.

Suppose that for $\mu = \lambda$ such a solution exists. It is easily verified that (6.5) is simply replaced by

$$W(\lambda) = 2W_\phi(\lambda) - \pi c F(\lambda) + W_h(\lambda). \tag{9.4}$$

Also (6.7) readily becomes

$$W(\lambda) = \frac{F_0^2}{F_1^2(\mu)} (2W_1(\lambda) - \pi F_1(\lambda)) + W_h(\lambda) \tag{9.5}$$

and the augmentation by (6.8) is still required in the situation described.

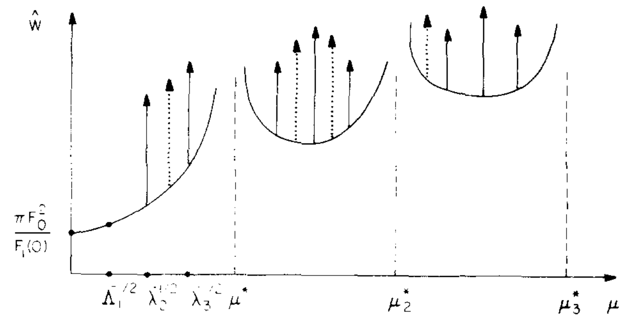


FIG. 6. Energy with fixed nonzero flux F_0;...indicates helical states.

To modify (6.14) we must select an appropriate $\overline{A}$. It is easy to see that we may choose, assuming $A_T = 0$ [see (5.17)]

$$A_\phi = \frac{B_{0,\phi}}{\lambda} - \frac{c}{\lambda r} + \frac{B_{h,\phi}}{\lambda}, \tag{9.6a}$$

$$A_r = \frac{B_{0,r}}{\lambda} + \frac{B_{h,r}}{\lambda}, \tag{9.6b}$$

$$A_z = \frac{B_{0,z}}{\lambda} + \frac{B_{h,z}}{\lambda}, \tag{9.6c}$$

since (9.2c) assures that (5.18), and hence (2.5), still hold. Thus (6.14) is replaced by

$$W(\lambda) = \tfrac{1}{2}\lambda K_0 + \pi c F(\lambda). \tag{9.7}$$

To mimic (6.20) we rewrite (9.7) as

$$\widehat{W}(\lambda) - \pi c F(\lambda) + W_h(\lambda) = \tfrac{1}{2}\lambda K_0. \tag{9.8}$$

Since $\overline{B}_h$ is completely unconstrained in Problem P_1, $W_h(\lambda)$ in (9.5) simply adds another spike to Fig. 4 at $\mu = \lambda$. See Fig. 6.

In problem P_2 the only constraint is that provided by (9.8). Hence the argument used to generate (6.23) applies and the helical solution just adds another segment to Fig. 5 at $\mu = \lambda$. See Fig. 7.

The basic open questions concern the existence, number, and location of these λ values. If the number is finite the situation differs little from the symmetric case. Even a denumerably infinite set of λ's does not significantly change things, unless there is an accumulation point at $\mu = 0$. It is conceivable that the λ set is nondenumerable. Even that would not create severe difficulties unless this set has an accumulation point at $\mu = 0$.

It may seem somewhat strange that a discrete set of λ's is even anticipated. Such a set certainly does not occur in the

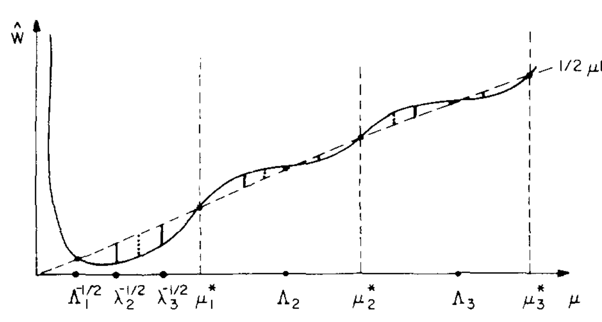


FIG. 7. Energy with fixed helicity K_0 (helical states shown by .....).

symmetric case. However, the helical case really consists of solving (2.1) and (2.2) subject to the *additional* constraint that F_0 in (2.3) be replaced by zero. We already have enough experience with flux-free states to know that they are special. Indeed, it can be shown that when $\overline{B}$ is so restricted the curl operation becomes self-adjoint (Appendix H). While self-adjointness in itself does not assure a discrete spectrum, this fact suggests that only very particular values of μ will solve the helical problem.

Theorem 3, however, continues to hold. (The consideration in Ref. 3 is not restricted to the symmetric case.) To see how our proof must be modified we remark that it may be shown (Appendix H) that if $\overline{B}^{(1)} = \overline{B}_0^{(1)} + \overline{B}_h^{(1)}$ belongs to μ_1 and $\overline{B}^{(2)} = \overline{B}_0^{(2)} + \overline{B}_h^{(2)}$ to $\mu_2 \neq \mu_1$ then

$$\iiint_T \overline{B}_h^{(1)} \cdot \overline{B}_h^{(2)} dV = 0. \tag{9.9}$$

Thus if $\overline{B}^{(1)}$ and $\overline{B}^{(2)}$ solve TP we have that (7.2) still holds, as do all subsequent equations through (7.8) provided $\widehat{W}$ is replaced by W.

X. SUMMARY AND CONCLUSIONS

We have examined in a rigorous fashion the Taylor Theory of plasma relaxation in tori with arbitrary smooth cross sections. Primary focus has been on the symmetric case, and a quite complete understanding of field-reversal and flux-free states has been achieved. The existence of a solution to the Taylor problem has been demonstrated in a torus, and the location of this solution has been isolated.

The helical states have not yielded to our methods. There is increasing numerical evidence that these states do, indeed, exist, at least for certain cross sections.[8] We have shown, however, that at any (possible) helical state the behavior of the plasma system is not significantly different from that found in the symmetric case. However, we believe it important to resolve the open question of the existence of helical states.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Don A. Baker, Los Alamos National Laboratory, for introducing us to the plasma relaxation problem and for his many insights and suggestions. We also wish to acknowledge the help and efforts of Allan Reiman, University of Maryland.

APPENDIX A

Theorem A-1: If K_0 is positive then there are no solutions to TP for $\mu \leqslant 0$.

Proof: By (9.7)

$$\begin{aligned} \tfrac{1}{2}\mu K_0 &= W(\mu) - \pi c F(\mu) \\ &\geqslant W_\phi(\mu) - \pi c F(\mu) \\ &= c^2(W_1(\mu) - \pi F_1(\mu)). \end{aligned} \tag{A1}$$

But (5.14) shows the last expression to be positive and the result is immediate.

APPENDIX B

We consider the assertion (Sec. V) that there is always a zero-flux state. There is surely always one unless $\Lambda_2 = \Lambda_3 = \cdots = 0$. From (5.6) we obtain in this event

$$F(\mu) = c \iint_\Omega \frac{d\Omega}{r} + \frac{c\mu^2 \Lambda_1 (\psi_1, 1/\sqrt{r})^2}{1-\mu^2\Lambda_1}. \tag{B1}$$

A zero of $F(\mu)$ occurs if

$$\iint_\Omega \frac{d\Omega}{r} = \mu^2 \Lambda_1 \left[\iint_\Omega \frac{d\Omega}{r} - (\psi_1, 1/\sqrt{r})^2 \right]. \tag{B2}$$

The only problem that arises is that the coefficient of $\mu^2 \Lambda_1$ might be nonpositive.

To see that this cannot happen we use the Schwarz inequality

$$\begin{aligned} (\psi_1, 1/\sqrt{r})^2 &= \left(\iint_\Omega \frac{\psi_1}{\sqrt{r}} d\Omega \right)^2 \\ &\leqslant \iint_\Omega \psi_1^2 d\Omega \iint_\Omega \frac{d\Omega}{r} \\ &= \iint_\Omega \frac{d\Omega}{r}. \end{aligned} \tag{B3}$$

Moreover, the strict inequality holds in (B3) unless $\psi_1 = k/\sqrt{r}$ for some constant k and for all $(r,z) \in \Omega$. But we know that $k/\sqrt{r}$ cannot be an eigenfunction since it fails to vanish on $\partial\Omega$.

We conclude that the coefficient of $\mu^2\Lambda_1$ in (B2) is positive and μ_1^* exists.

APPENDIX C

We present here the details of the calculations leading to (5.14). From (5.13)

$$\begin{aligned} W_\phi(\mu) &= c^2\pi \left\{ \iint_\Omega \frac{d\Omega}{r} + 2\mu^2 \sum_{j=1}^{\infty} \frac{\Lambda_j (\psi_j, 1/\sqrt{r})^2}{1-\mu^2\Lambda_j} \right. \\ &\quad + \mu^4 \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \frac{\Lambda_j \Lambda_k (\psi_j, 1/\sqrt{r})(\psi_k, 1/\sqrt{r})}{(1-\mu^2\Lambda_j)(1-\mu^2\Lambda_k)} \\ &\quad \left. \times \iint_\Omega \psi_j \psi_k d\Omega \right\} \\ &= c^2\pi F_1(\mu) + c^2\mu^2\pi \sum_{j=1}^{\infty} \frac{\Lambda_j (\psi_j, 1/\sqrt{r})^2}{1-\mu^2\Lambda_j} \\ &\quad + c^2\mu^4\pi \sum_{j=1}^{\infty} \frac{\Lambda_j^2 (\psi_j, 1/\sqrt{r})^2}{(1-\mu^2\Lambda_j)^2}. \end{aligned} \tag{C1}$$

Here we have used (5.6) and the orthonormality of the ψ's. Next we note

$$\frac{\Lambda_j}{1-\mu^2\Lambda_j} + \frac{\mu^2\Lambda_j^2}{(1-\mu^2\Lambda_j)^2} = \frac{\Lambda_j}{(1-\mu^2\Lambda_j)^2}. \tag{C2}$$

From (C1),

$$W_\phi = c\pi F(\mu) + c^2\mu^2\pi \sum_{j=1}^{\infty} \frac{A_j(\psi_j, 1/\sqrt{r})^2}{(1-\mu^2 A_j)^2} \tag{C3}$$

as in (5.14).

APPENDIX D

Suppose

$$\int_0^{2\pi} A_\phi r d\phi = \int_{C_T} \bar{A}\cdot d\bar{l} = A_T, \tag{D1}$$

a constant independent of μ. We may satisfy this constraint with

$$\begin{aligned} A_\phi &= \frac{B_\phi}{\mu} - \frac{c}{\mu r} + \frac{A_T}{2\pi r} + \frac{1}{r}\frac{\partial \chi}{\partial \phi}, \\ A_r &= \frac{B_r}{\mu} + \frac{\partial \chi}{\partial r}, \\ A_z &= \frac{B_z}{\mu} + \frac{\partial \chi}{\partial z}, \end{aligned} \tag{D2}$$

where χ is any single valued function independent of μ on T, and still have

$$\nabla \times \bar{A} = \bar{B}. \tag{D3}$$

However, now (see 5.19)

$$K(\mu) = \frac{2\widehat{W}(\mu)}{\mu} - \frac{2\pi c^2 F_1(\mu)}{\mu} + 2\pi c F_1(\mu) A_T. \tag{D4}$$

Since $cF_1(\mu) = F(\mu)$, this can be written

$$K(\mu) - 2\pi F(\mu) A_T = \frac{2\widehat{W}(\mu)}{\mu} - \frac{2\pi c^2 F_1(\mu)}{\mu}. \tag{D5}$$

When the flux $F(\mu) = F_0$ and the helicity $K(\mu) = K_0$ are fixed for a given A_T, the solutions to TP are precisely the same as the solutions to the TP in the text (with the assumption $A_T = 0$), provided the helicity is taken to be

$$K_0 - 2\pi F_0 A_T. \tag{D6}$$

This shows that the solution to TP is not unique for given K_0 and F_0 unless A_T is given.

However, the overall qualitative behaviors discovered for $\bar{B}$, W, etc., are not changed. In practice, if K_0 and A_T are given one simply defines a new K_0, $\tilde{K}_0$ by

$$\tilde{K}_0 = K_0 - 2\pi F_0 A_T \tag{D7}$$

and proceeds as in the body of the paper, using $\tilde{K}_0$ and (5.18). If $\tilde{K}_0 < 0$ then only negative μ's will occur (see Appendix A).

One might worry that further changes in the definition of $\bar{A}$ could lead to new phenomena. Consider any two vector potentials $\bar{A}_1$ and $\bar{A}_2$ which lead to the same $\bar{B}$ and which satisfy (D1). Let

$$\bar{V} = \bar{A}_1 - \bar{A}_2 \tag{D8}$$

so that

$$\nabla \times \bar{V} = 0 \tag{D9}$$

and

$$\int_{C_T} \bar{V}\cdot d\bar{l} = 0. \tag{D10}$$

Then

$$\bar{V} = \nabla \phi \tag{D11}$$

on T. Now

$$\int_{C_T} \nabla\phi \cdot d\bar{l} = 0, \tag{D12}$$

so ϕ is single valued on T by Stokes's Theorem for multiply connected regions. Thus all solutions to (D1) and (D3) are given by (D2). Hence all results of the main text are unchanged if K_0 is replaced by $\tilde{K}_0$.

APPENDIX E

Theorem E-1: The minimum of the function [see (6.10)]

$$\widehat{W}(\mu) = 2W_\phi(\mu) = 2\pi \sum_{j=j_0}^{j_0+m-1} \beta_j^2, \tag{E1}$$

subject to the constraint

$$\sum_{j=j_0}^{j_0+m-1} \beta_j(\psi_j, 1/\sqrt{r}) = F_0, \tag{E2}$$

is given by

$$\widehat{W} = 2\pi F_0^2 \left[\sum_{j=j_0}^{j_0+m-1} (\psi_j, 1/\sqrt{r})^2\right]^{-1}. \tag{E3}$$

Proof: We wish to minimize

$$f(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i^2 \tag{E4}$$

subject to

$$\sum_{i=1}^{n} a_i x_i = c. \tag{E5}$$

We form

$$Q(\mathbf{x},\lambda) = \sum_{i=1}^{n} x_i^2 + \lambda\left(\sum_{i=1}^{n} a_i x_i - c\right) \tag{E6}$$

and find its minimum. Equating partial derivatives to 0 yields

$$\begin{aligned} &2x_i + \lambda a_i = 0, \quad i = 1,2,\ldots,n \\ &\sum_{i=1}^{n} a_i x_i = c. \end{aligned} \tag{E7}$$

Thus $x_i = -\frac{1}{2}\lambda a_i$ so

$$-\tfrac{1}{2}\lambda \sum_{i=1}^{n} a_i^2 = c. \tag{E8}$$

This gives

$$x_i = a_i c / \sum_{i=1}^{n} a_i^2 \tag{E9}$$

and

$$\sum_{i=1}^{n} x_i^2 = c^2 / \sum_{i=1}^{n} a_i^2. \tag{E10}$$

The result in the theorem follows by direct substitution.

APPENDIX F

To verify Eq. (6.12), we recall (5.6) and (5.14). Clearly $F_1(\mu)$ and $W_1(\mu)$ are meromorphic functions with poles at $\mu = \Lambda_j^{-1/2}$. The flux has poles of order one at those points while the energy has poles of order 2. Near $\mu = \Lambda_{j_0}^{-1/2}$ we may write

$$W_1(\mu) = \left[\pi\mu^2 \Lambda_{j_0} \sum_{j=j_0}^{j_0+m-1} (\psi_j, 1/\sqrt{r})^2/(1-\mu^2\Lambda_{j_0})^2\right] \times \{1 + \widetilde{W}_1(\mu)\} \tag{F1}$$

and

$$F_1(\mu) = \left[\mu^2 \Lambda_{j_0} \sum_{j=j_0}^{j_0+m-1} (\psi_j, 1/\sqrt{r})^2/(1-\mu^2\Lambda_{j_0})\right] \times \{1 + \widetilde{F}_1(\mu)\}, \tag{F2}$$

where $\widetilde{W}_1$ and $\widetilde{F}_1$ are analytic near $\mu = \Lambda_{j_0}^{-1/2}$ and with zeros there. Thus,

$$\begin{aligned}\widehat{W}(\mu) &= \frac{F_0^2}{F_1^2(\mu)}(2W_1(\mu) - \pi F_1(\mu)) \\ &= \frac{2\pi F_0^2}{\mu^2 \Lambda_{j_0}} \left[\sum_{j=j_0}^{j_0+m-1} (\psi_j, 1/\sqrt{r})^2\right]^{-1} \{1 + \phi(\mu)\},\end{aligned} \tag{F3}$$

where $\phi(\mu) \to 0$ as $\mu \to \Lambda_{j_0}^{-1/2}$. The result is immediate.

APPENDIX G

We may explore the situation further by trying to find the solution to TP with least energy, that is, least μ by Theorem 3. The only difficulty that may arise comes when the two energy curves $\widehat{W}_{\rm I}$ and $\widehat{W}_{\rm II}$ intersect along the "spike" that may rise from the $\widehat{W}_{\rm I}$ curve at $\lambda_2^{-1/2}$. (If $\lambda_2^{-1/2}$ lies to the right of μ_1^* or if $\widehat{W}_{\rm I}$ and $\widehat{W}_{\rm II}$ have a previous point of intersection, this is of no consequence.) The problem occurs because of the need to find a *single* solution with energy $\widehat{W}$ which solves both P_1 and P_2 at this μ. The relevant equations are (4.6), (4.8), (6.8), (6.9), (6.20), and (6.21). We need only find a set of α_j's to satisfy (6.20). This will be possible if

$$\tfrac{1}{2}\mu K_0 \geqslant 2c^2(W_1 - \pi F_1) = 2\frac{F_0^2}{F_1^2}(W_1 - \pi F_1). \tag{G1}$$

Note from (6.23) that the $W_{\rm I}$ curve and $W_{\rm II}$ will have a previous intersection unless

$$\widehat{W}_m(\lambda_2^{-1/2}) \leqslant \tfrac{1}{2}\mu K_0\left[1 + \frac{\pi F_1}{2(W_1 - \pi F_1)}\right], \tag{G2}$$

since $F_1 < 0$ between Λ_1 and μ_1^*. (This expresses the fact that the $W_{\rm II}$ curve must lie above the single-valued part of the $W_{\rm I}$ curve between 0 and $\lambda_2^{-1/2}$.) Using (6.9), (G2) becomes

$$\frac{F_0}{F_1^2}(2W_1 - \pi F_1) \leqslant \mu K_0\left[\frac{2W_1 - \pi F_1}{2(W_1 - \pi F_1)}\right], \tag{G3}$$

which reduces to (G1). We have thus proved

Theorem G-1: The solution to TP with the least energy can be found at that μ at which $W_{\rm I}$ and $W_{\rm II}$ have their first intersection and, although the solution may not be unique, the value of $rB_\phi = c$ is unique and the total energy of any solution is given by

$$\widehat{W}(\mu) = \tfrac{1}{2}\mu K_0 + \pi c F_0. \tag{G4}$$

APPENDIX H

Lemma H-1: If $\overline{B}_1$ and $\overline{B}_2$ are flux-free, then

$$\iint_{\partial T} \overline{B}_1 \times \overline{B}_2 \cdot \overline{n}\, dS = 0. \tag{H1}$$

Proof: Since $\nabla \times \overline{B} \cdot \overline{n} = 0$ on ∂T, we may write

$$\overline{B} = \nabla \chi \text{ on } \partial T. \tag{H2}$$

Now let $\overline{u}_\tau$ be the unit vector on ∂T in the poloidal direction, $\overline{u}_\phi$ the unit vector in the toroidal direction. Then

$$\begin{aligned}\frac{1}{\mu}\int_{C_p} \overline{B}\cdot\overline{u}_\tau\, dl &= \frac{1}{\mu}\int_{\partial\Omega} \overline{B}\cdot d\overline{l} = \frac{1}{\mu}\iint_{\Omega} \nabla\times\overline{B}\cdot\overline{u}_\phi\, dl \\ &= \iint_{\Omega} \overline{B}\cdot\overline{n}\, d\Omega = F(\mu)\end{aligned} \tag{H3}$$

and

$$\frac{1}{\mu}\int_{C_T} \overline{B}\cdot\overline{u}_\phi\, dl = \frac{1}{\mu}\int_0^{2\pi} rB_\phi\, d\phi = \frac{2\pi c}{\mu}, \tag{H4}$$

where $B_\phi = c/r$ is the value of the ϕ component of the symmetric part of $\overline{B}$ on the boundary of T. [Note that no other components of $\overline{B}$ contribute to (H4) (see (2.6))]. Let

$$\omega = \left(\chi - \frac{F(\mu)\tau}{2\pi} - \frac{c\phi}{\mu}\right), \tag{H5}$$

where χ is as in (H2) and τ is the coordinate associated with $\overline{u}_\tau$. Then

$$\int_{C_p} \nabla\omega\cdot\overline{u}_\tau\, dl = 0 \tag{H6}$$

and

$$\int_{C_T} \nabla\omega\cdot\overline{u}_\phi\, dl = 0. \tag{H7}$$

Thus ω is periodic in τ and θ.

We have

$$\overline{B} = \mu\nabla\omega + \frac{F(\mu)}{2\pi}\nabla\tau + c\nabla\phi \quad \text{on} \quad \partial T. \tag{H8}$$

Hence,

$$\overline{B}_1 \times \overline{B}_2 = \mu_1\mu_2 \nabla\omega_1 \times \nabla\omega_2 + \left[\frac{F_1 c_2}{2\pi} - \frac{F_2 c_1}{2\pi}\right]\overline{u}_\phi. \tag{H9}$$

If $F_1 = F_2 = 0$,

$$\begin{aligned}\iint_{\partial T} (\overline{B}_1 \times \overline{B}_2)\cdot\overline{n}\, dS &= \frac{1}{\mu_1\mu_2}\iint_{\partial T} \nabla\omega_1 \times \nabla\omega_2 \cdot \overline{n}\, dS \\ &= \frac{1}{\mu_1\mu_2}\iiint_T \nabla\cdot\nabla\omega_1 \times \nabla\omega_2\, dV = 0,\end{aligned} \tag{H10}$$

since $\nabla\cdot\nabla\omega_1 \times \nabla\omega_2 = 0$. This establishes (H1).

Theorem H-1: The curl operator is self-adjoint when restricted to flux-free states. In other words,

$$\iiint_T [\overline{B}_1\cdot\nabla\times\overline{B}_2 - \overline{B}_2\cdot\nabla\times\overline{B}_1]\, dV = 0 \tag{H11}$$

if $\overline{B}_1$ and $\overline{B}_2$ are flux-free.

Proof: We use

$$\nabla\cdot(\overline{B}_1 \times \overline{B}_2) = \overline{B}_2\cdot\nabla\times\overline{B}_1 - \overline{B}_1\cdot\nabla\times\overline{B}_2. \tag{H12}$$

Thus

$$\iiint_T [\bar{B}_1\cdot\nabla\times\bar{B}_2 - \bar{B}_1\cdot\nabla\times\bar{B}_2]dV$$

$$= \iiint_T \nabla\cdot(\bar{B}_2\times\bar{B}_1)dV$$

$$= \iint_{\partial T} \bar{B}_2\times\bar{B}_1\cdot\bar{n}dS$$

$$= 0 \tag{H13}$$

by the lemma.

Corollary H-1: Flux-free states belonging to different λ's are orthogonal. In other words,

$$\iiint_T \bar{B}_1\cdot\bar{B}_2 dV = 0 \tag{H14}$$

if $\bar{B}_1$ and $\bar{B}_2$ are flux-free and satisfy $\nabla\times\bar{B}_i = \lambda_i\bar{B}_i$ with $\lambda_1\neq\lambda_2$.

Proof: From the theorem and $\nabla\times\bar{B}_i = \lambda_i\bar{B}_i$,

$$\iiint(\lambda_2\bar{B}_1\cdot\bar{B}_2 - \lambda_1\bar{B}_1\cdot\bar{B}_2)dV = 0. \tag{H15}$$

Thus

$$(\lambda_2-\lambda_1)\iiint\bar{B}_1\cdot\bar{B}_2 dV = 0. \tag{H16}$$

Corollary H-2: If $\bar{B}_h^{(1)}$ and $\bar{B}_h^{(2)}$ are helical solutions belonging to different eigenvalues, then

$$\iiint_T \bar{B}_h^{(1)}\cdot\bar{B}_h^{(2)}dV = 0. \tag{H17}$$

Proof: All helical states are flux-free, and the theorem applies.

APPENDIX I

In this final Appendix, we give an abbreviated derivation of the Taylor Problem, primarily due to Baker.[9] The presentation clarifies somewhat the results of Appendix D.

We seek the magnetic field $\bar{B}$ of a plasma inside a perfectly conducting toroidal shell. Woltjer[10] has shown that

$$\frac{\partial K}{\partial t} = -\iiint_T \bar{E}\cdot\bar{B}\,dV, \tag{I1}$$

where

$$K = \iiint_T \bar{A}\cdot\bar{B}\,dV. \tag{I2}$$

Since $\bar{E}$ is 0 inside a perfect conductor, we have

$$\frac{\partial F}{\partial t} = \frac{\partial}{\partial t}\iint_\Omega \bar{B}\cdot\bar{u}_\phi\,d\Omega$$

$$= -\iint_\Omega \nabla\times\bar{E}\cdot\bar{u}_\phi\,d\Omega = -\int_{C_p}\bar{E}\cdot d\bar{l} = 0 \tag{I3}$$

and

$$\frac{\partial A_T}{\partial t} = \frac{\partial}{\partial t}\int_{C_T}\bar{A}\cdot d\bar{l} = \frac{\partial}{\partial t}\iint_H \nabla\times\bar{A}\cdot\bar{n}\,dS \tag{I4}$$

$$= \frac{\partial}{\partial t}\iint_H \bar{B}\cdot\bar{n}\,dS$$

$$= -\iint_H \nabla\times\bar{E}\cdot\bar{n}\,dS = -\int_{C_T}\bar{E}\cdot d\bar{l} = 0,$$

where H is any surface spanning C_T. Now inside the conducting shell $\bar{B} = \nabla\times\bar{E} = 0$. Since $\bar{B}\cdot\bar{n}$ is continuous (Ref. 11, p. 16), we have

$$\bar{B}\cdot\bar{n} = 0 \quad \text{at } \partial T. \tag{I5}$$

Finally, Taylor[2] and Montgomery and Turner[12] have reasoned that $\partial W/\partial t$ is much greater than $\partial K/\partial t$ (see also Ref. 13 for further discussion). Because of this, Taylor[2] has made the assumption that

$$\frac{\partial K}{\partial t} \equiv 0, \tag{I6}$$

while W relaxes to the minimum energy:

$$W = \min \iiint_T \bar{B}\cdot\bar{B}\,dV. \tag{I7}$$

This leads us to the mathematical model which minimizes W subject to

$$\begin{cases} K \text{ given,} \\ F \text{ given,} \\ A_T \text{ given,} \\ \bar{B}\cdot\bar{n} = 0 \text{ at } \partial T. \end{cases} \tag{TP1}$$

A Lagrange multiplier argument leads to the problem of minimizing W subject to

$$\begin{cases} \nabla\times\bar{B} = \mu\bar{B}, \\ K \text{ given,} \\ F \text{ given,} \\ A_T \text{ given,} \\ \bar{B}\cdot\bar{n} = 0 \text{ at } \partial T. \end{cases} \tag{TP2}$$

We have shown in Appendix D that minimizing W subject to

$$\begin{cases} \nabla\times\bar{B} = \mu\bar{B}, \\ K - 2\pi A_T F \text{ given,} \\ F \text{ given,} \\ \bar{B}\cdot\bar{n} = 0 \text{ at } \partial T \end{cases} \tag{TP3}$$

has solutions which correspond to those of TP2. In fact, Taylor[14] has now proposed that $K - 2\pi A_T F$ be the new definition of helicity. [This definition of helicity makes the solution of (TP1) independent of the choice of $\bar{A}$].

[1] J. B. Taylor, "Relaxation of Toroidal Discharges," in *Third Topical Conference on Pulsed High Beta Plasmas*, edited by D. E. Evans (Pergamon, Oxford, 1975).

[2] J. B. Taylor, "Relaxation of Toroidal Plasma and Generation of Reverse Magnetic Fields," Phys. Rev. Lett. **33**, 1139–41 (1977).

[3]A. Reiman, "Taylor Relaxation in a Torus of Arbitrary Aspect Ratio and Cross Section," LPS 282, Laboratory of Plasma Studies (Cornell University, 1980) (to appear in Physics of Fluids).
[4]P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964).
[5]J. A. Cochran, *The Analysis of Linear Integral Equations* (McGraw-Hill, New York, 1972).
[6]F. Smithies, *Integral Equations* (Cambridge U.P., London, 1962).
[7]R. Jentzsch, "Über Integralgleichungen mit positivem Kern," J. Reine Ange. Math. **141**, 235–244 (1912).
[8]V. Faber, G. M. Wing, and A. B. White, Jr., "Numerical Solution of an Eigenvalue Problem Describing Relaxed Toroidal Plasmas," paper presented at the SIAM 1981 National Meeting, June 8–10, 1981, Troy, New York (LA-UR-81-566).
[9]D. A. Baker (private communication).
[10]L. Woltjer, "A Theorem on Force-free Magnetic Fields," Proc. Natl. Acad. Sci. (USA) **44**, 489–491 (1958).
[11]A. Sommerfeld, *Electrodynamics*, translated by E. G. Rambert (Academic, New York, 1952).
[12]D. Montgomery and L. Turner, "Selective Decays in MHD Turbulence," 1978 RFP Workshop, Padua.
[13]W. H. Matthews and D. C. Montgomery, "Why should Energy Decay while magnetic Helicity is conserved?: The Essentials of Turbulent Spectral Transfer," 1980 RFP Workshop, Los Alamos.
[14]J. B. Taylor, "Relaxation Revisited," 1980 RFP Workshop, Los Alamos.

Finite sum approximations to Brillouin zone integrals with symmetrized plane waves

N. O. Folland and Junaidah Osman[a)]
Physics Department, Kansas State University, Manhattan, Kansas 66506

(Received 30 June 1981; accepted for publication 2 April 1982)

Techniques are presented for expanding a periodic function of cubic translational symmetry and arbitrary rotational symmetry in a finite set of symmetry-adapted plane waves. Results for all cubic lattices are tabulated in a form convenient for use in computation. The role of "special points" in the sense of Chadi and Cohen is discussed in the extended context of symmetry-adapted plane waves.

PACS numbers: 71.10. + x, 02.20. + b, 02.30.Mv

I. INTRODUCTION

The basic idea for a finite sum approximation (FSA) of an integral over a finite region of space is to partition the space into identical volume elements and approximate the integral by a sum. Each term in the sum is the integrand evaluated at the center of a volume element multiplied by the volume element. If all volume elements approach zero in a limiting sense, then the resulting sum is a Riemann integral. The natural partition of a symmetric space into identical volume elements puts the center of the space at the center of a volume element. Suppose we partition the space by displacing all volume elements in the same way? Will anything be basically different? The answer, in certain circumstances can be dramatically, "yes!" The "certain circumstances" that are of interest here are when the integrand is invariant to operations of a point group. In such a case because of the symmetry, it is possible to use a significantly fewer number of points (equal volume elements), to approximate an integral when these points are positioned optimumly, than to simply place them so the center of the space is at the center of a volume element.

As a simple example, consider an integral of a periodic function where the finite region of space is a cube and the function is invariant with respect to each of the 48 cubic point group operations about the center of the cube. Now divide the cube into identical little cubes such that one little cube is centered about the origin and such that cubic point group operations leave the partition unaffected. Also require this partition to be such that if a volume element intersects the surface, then its center falls on the surface. Because of the point group symmetry the cube may also be partitioned into 48 symmetry-related volumes referred to as irreducible wedges (IW). Thus, using the symmetry an FSA can be expressed as 48 times a sum of terms consisting of the function evaluated at centers of volume elements which lie inside or on the surface of the IW multiplied by (weighted by) the fraction of the volume element contained inside the IW. Now, suppose all volume elements (points) are displaced by the same vector, but in such a way that the resulting set of points are invariant to cubic point group operations. The only way that this can be done is to displace all the original points by a vector which connects the center of a little cube to a corner. In this process all the points which originally fell inside or on the internal surfaces of the IW remain inside or on the internal surfaces. Points which originally fell on the external surface of the IW are now completely outside the IW in that their volume elements are completely excluded from the IW. Thus, in the case where the original volume partition places points on the external surface of the IW an FSA with the displaced points will involve fewer points than the original FSA. When the original volume partition places no points on the surface, then the original and displaced FSA involve exactly as many points in the IW. The procedure above may easily be extended to integrals over face-centered cubic (fcc) and body-centered cubic (bcc) symmetric regions of space. In all cases the effect, if any, of symmetrically displacing the volume elements is a surface effect. The displaced point FSA's described above are exactly the "special points" of Chadi and Cohen.[1] It is noteworthy that periodicity has no role in the preceding discussion.

In a previous paper[2] (hereafter referred to as I) an equivalent but different perspective as compared to the above discussion on FSA's was emphasized. Although this is quite arbitrary, for historical reasons the integration is taken to be in a reciprocal space or k-space and the volume of integration is the symmetric, Brillouin zone (BZ) in k-space. Now a periodic function, $f(\mathbf{k}) = f(\mathbf{k} + \mathbf{K})$, where $\mathbf{K}$ is any member of a reciprocal lattice, can be expanded in Fourier series or in terms of an infinite set of orthogonal plane waves of the form, $\exp(i\mathbf{k}\cdot\mathbf{R})$, where space lattice vectors $\mathbf{R}$ satisfy the property $\exp(i\mathbf{k}\cdot\mathbf{R}) = 1$. A BZ integral is directly related to the coefficient of the plane wave with $\mathbf{R} = \mathbf{0}$. If an FSA is made to a BZ integral, then the corresponding Fourier series is truncated to a finite series. For FSA's corresponding to symmetrically partitioned BZ's, the number of $\mathbf{R}$-vectors or planes waves which are inequivalent with respect to the FSA is exactly equal to the total number of points (or volume elements) into which the BZ is partitioned. For FSA's corresponding to nonsymmetrical displacements which are then symmetrized, the number of inequivalent plane waves as determined by the FSA is less than the number of points. Just as inequivalent k-vectors may be specified by a BZ, inequivalent $\mathbf{R}$-vectors with respect to an FSA may be specified by a

[a)] Present address: Physics Department, Universiti Sains Malaysia, Minden, Penang, Maylaysia.

symmetric zone in real space. When considering periodic functions $f(\mathbf{k})$ defined on the BZ which are invariant with respect to cubic point group operations, the plane waves may be reexpressed in terms of linear combinations which are invariant to the group operations. In the case of the example described in the preceding paragraph, both the FSA based on the symmetrically placed points including the origin and the corresponding FSA in which these points are symmetrically displaced the zone of inequivalent **R** vectors is exactly the same for each. The surface effect described above manifests itself in this case by the fact that all symmetrized linear combinations of plane waves formed from **R** vectors on the surface of the zone in **R**-space are identically zero when evaluated with respect to any point of the displaced FSA. For more details in this context the reader is referred to I. This surface effect must manifest itself for other symmetries besides the identity representation. The primary objective of the present work is to analyze completely these surface effects for periodic functions of arbitrary point symmetry.

One feature of the present work is to treat all FSA's on the same basis. A consequent practical observation, which is emphasized in Sec. IV, is that FSA's of comparable accuracy provide an estimate of the error. However, to treat all FSA's in the same manner requires a notation which precisely and compactly specifies all finite sets of points. The notation of I will be retained in this application. The notation is summarized in the next few paragraphs below for completeness. Then, the general results, Eq. (18) of I and Eq. (1) of this paper, will be stated in this notation. Note that the notation is designed to facilitate immediate implementation into a computer code. Also, the tables with their captions summarize the definitions and results which are needed to program the technique.

The basic idea for specifying finite point sets embodies the notion of an infinite, regular array of points from which a finite number are selected by including only those points contained in a region of space. Three steps are involved. (1) Define an infinite set of points; symbolize such a set by g. (2) Define a region of space or zone; let z symbolize such a zone. Then, (3) $g(z)$ symbolizes the set of all points g which are contained inside and on the surface of zone z. It is also necessary to distinguish between points in k-space and real space. This is done by symbolizing k-space points or point vectors and zones with lower case symbols and real space points and zones with upper case symbols. Point sets are defined in Table IA and zones in Table IB in terms of points and zones in a dimensionless space. All point sets and zones defined here are invariant to cubic point group operations. Actual space vectors are obtained from the dimensionless point vectors by attaching the unit $a/2$, where a is the lattice constant for a conventional cube. Similarly, k-vectors are obtained from the dimensionless point vectors by attaching the unit $2\pi/Na$, where N is a positive integer which specifies a zone.

For example, the space lattice point S defined by this notation consists of all space vectors of the form $2(i,j,k)a/2$, where i,j,k are any integers or zero. S is just a simple cubic lattice of space vectors. The symbol $S(F_N)$ specifies the finite set of simple cubic lattice vectors contained in the face-centered cubic zone F_N. To explicitly obtain these points, one must first find all points in the 1/48th zone wedge where $2i \geqslant 2j \geqslant 3k \geqslant 0$ and $2i + 2j \geqslant N$ and then add to these all additional points which may be obtained by applying the 48 cubic point group rotations. Thus, S, F, and B are the usual sc, fcc, and bcc space lattices, respectively, while BS and FS are point sets containing only body-centered and face-centered space vectors, respectively. The space vectors contained in ES are "edge" vectors and are not contained in any cubic space lattice. The symbols NS, NF, etc., simply mean that all space vectors in S, F, etc., are scaled by the positive integer N. It is useful to note that zones S_{2N}, F_N, and B_N are "primitive" cells for lattices, NS, NF, and NB, respectively.

In the case of k-vectors the unit $2\pi/Na$ is associated with the zone delimiting index N. The resulting zones s_N, b_N, and f_n in k-space are symmetric cell BZ's. The 1/48th zone wedges defined in Table IB are identical to the usual irreducible BZ's (IBZ's) for the cubic lattices. Although it leads to some awkwardness, the standard convention is followed for naming BZ's. For example, b_N is the BZ for k-vectors which are *reciprocal* to the body-centered cubic space lattice. This convention has been maintained in naming k-vector sets. For example, the set of k-vectors, fu consists of k-vectors of the form $(2i + 1, 2j + 1, 2k + 1)2\pi/Na$, where i,j,k are any integers or zero. These k-vectors are actually body-centered type vectors. For BZ z_N the FSA point sets $fu(z_N)$ are identical to the Chadi–Cohen points. The k-vector lattices s, b, and f are *reciprocal* to space lattices NS, NB, and NF, respectively. The k-vector points sets bu and fu involve k-vectors *reciprocal* to space lattices NB and NF, respectively. The k-vec-

TABLE I. Definitions of finite sets of vectors.[a]

A. Definitions of infinite sets of vectors.[b]

$u = S = \{(2i,2j,2k)\}$
$eu = ES = \{2i + 1,2j,2k)\} \cup \{(2i,2j + 1,2k)\} \cup \{(2i,2j,2k + 1)\}$
$bu = FS = \{(2i,2j + 1,2k + 1)\} \cup \{(2i + 1,2j,2k + 1)\} \cup \{(2i + 1,2j + 1,2k)\}$
$fu = BS = \{(2i + 1,2j + 1,2k + 1)\}; s = u \cup eu \cup bu \cup f$
$b = F = u \cup bu; f = B = u \cup fu; feu = fu \cup eu$
$beu = bu U eu$

B. Zonal restrictions, a point (i,j,k) is contained in the 1/48th zone wedge if integer or zero values of indices i,j,k satisfy the condition $i \geqslant j \geqslant k \geqslant 0$ and

k-vector BZ (symbol)	space vector zone (symbol)	Condition(s)
sc (s_N)	sc (S_N)	$2i \leqslant N$
bcc (b_N)	fcc (F_N)	$i + J \leqslant N$
fcc (f_N)	bcc (B_N)	$i \leqslant N$ and $2(i + j + k) \leqslant 3N$

[a] In specifying vectors, upper case symbols are associated with space lattice vectors and lower case symbols are associated with k-vectors. Actual space vectors **R** are formed from a dimensionless vector (i,j,k) by assigning to it the unit $a/2$, where a is the lattice constant for the conventional cube, $\mathbf{R} = (i,j,k)a/2$. Actual k-vectors **k** are formed from dimensionless vectors (i,j,k) by assigning the unit $2\pi/Na$, $\mathbf{k} = (i,j,k)2\pi/Na$. The same symbols that are used to define sets of dimensionless vectors will be used for space vectors and k-vectors. Zonal restrictions are specified in terms of the 1/48th zone wedge or irreducible wedge. The remaining vectors (points) in the zone are generated by the 48 operations of the cubic group. A restricted set of points $g(z)$ is denoted by the symbol for the infinite set of points, g, with the zonal restrictions z, in parentheses. For example, $fu(b_N)$ specifies the finite set of k-vectors contained in zone b_N.

[b] Indices i,j,k of vectors may take all integer and zero values. The symbol $\cup$ indicates a union of sets.

tor lattice u is a *universal* lattice in that it is a sublattice of all other lattices, s, b, and f.

The relevant results, Eq. (18) of I, are reproduced here:

$$\sum_{g(z_N)} w(\mathbf{k})\exp[i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')] = N^3[n_0\Delta(\mathbf{R},\mathbf{R}';NS)+n_1\Delta(\mathbf{R},\mathbf{R}';NES) + n_2\Delta(\mathbf{R},\mathbf{R}';NFS)+n_3\Delta(\mathbf{R},\mathbf{R}';NBS)]/M_z, \quad (1)$$

where the factor M_z is determined by the BZ: $M_s = 8$, $M_b = 4$, $M_f = 2$ for the three cubic BZ's. The sum includes k-vectors from the set g restricted to BZ z_N as discussed above and in Table I. The factors $w(\mathbf{k})$ are weights which allow all points in and on the surface of zone z_N to be included in a symmetric manner. Regarding the points as small spheres, the weights are equal to the fraction of the sphere contained inside zone z_N. Thus, all interior points $\mathbf{k}$ of z_N have weights $w(\mathbf{k}) = 1$. The weights for surface points are tabulated in Table V along with other information regarding surface points. In the right member of Eq. (1) the constants n_i, $i = 0,1,2,3$, are determined by the sum type, $n_i = n_i(g)$. When positive integer N is even, all sum types are meaningful in Eq. (1); but when N is odd, only certain sum types are meaningful. The sum dependent parameters n_i and the associated restrictions with respect to N are listed in Table II. It should be noted that only the first four rows of entries in Table II are basic in that the remaining entries correspond to sums which are additive combinations of the basic four.

The functions $\Delta(\mathbf{R},\mathbf{R}';NL) = 1$ if the space lattice vector $\mathbf{R}-\mathbf{R}'$ is contained in the set NL of space lattice vectors and is zero otherwise. Thus, in the same sense that inequivalent k-vectors are restricted to the BZ, the right member of Eq. (1) implies a zonal restriction for nonequivalent space lattice vectors. These zonal restrictions on inequivalent space lattice vectors depend on the sum type and are listed in Table II under the column heading Z_N. Thus, it is possible to have a set of inequivalent body-centered space lattice vectors restricted to simple cubic (S_N) or face-centered cubic (F_N) zones. Note that in the case $\mathbf{R} = \mathbf{R}'$ the left member of Eq. (1) is evaluated to be the number of *whole* points of sum type g in BZ z_N,

$$N_0(\mathbf{g}(z_n)) = N^3 n_0/M_z. \quad (2)$$

TABLE II. Brillouin zone sum parameters for Eq. (1).[a]

$P(S)$	$P(B)$	$P(F)$	g	n_0	n_1	n_2	n_3	Z_N
e	e	e	u	1	1	1	1	S_N
e	e	e	eu	3	1	-1	-3	S_N
e	e	e	bu	3	-1	-1	3	S_N
e	e	e	fu	1	-1	1	-1	S_N
o	o	o	s	8	8	8	8	S_{2N}
e	o	e	b	4	0	0	4	B_N
e	o	e	feu	4	0	0	-4	B_N
e	e	o	f	2	0	2	0	F_N
e	e	o	beu	6	0	-2	0	F_N

[a] The first three columns labeled $P(L)$ indicate restrictions on N for each of the space lattices. A (sub) lattice sum g which occurs nontrivially only for even N is indicated with an e and a sum g which occurs nontrivially for both even and odd N is indicated with an o. The last column Z_N specifies zonal restrictions on lattice vectors.

To illustrate its meaning and use, Eq. (1) is applied to approximate expansion coefficients for a function $f(\mathbf{k})$ which is invariant to translations by a reciprocal lattice of vectors $\mathbf{K}$, $f(\mathbf{k}+\mathbf{K}) = f(\mathbf{k})$,

$$f(\mathbf{k}) \cong \sum_{G(Z_N)}{}' C(\mathbf{R})\exp(i\mathbf{k}\cdot\mathbf{R}). \quad (3)$$

The reciprocal lattice $\{\mathbf{K}\}$ determines the BZ z_N and the appropriate space lattice G [$\exp(i\mathbf{k}\cdot\mathbf{R}) = 1$]. For a given positive integer N the finite set of k-vectors $g(z_N)$ determines the zone Z_N of inequivalent space lattice vectors. Space lattice vectors $\mathbf{R}$ and $\mathbf{R}+\mathbf{R}'$ are equivalent in the sense that for all k-vectors in $g(z_N)$ $\exp[i\mathbf{k}\cdot(\mathbf{R}+\mathbf{R}')] = \pm\exp(i\mathbf{k}\cdot\mathbf{R})$ for all $\mathbf{R}'$ contained in the extended space lattice NZ. Certain R-vectors *on the surface of zone* Z_N may also be equivalent in this sense. The prime on the sum in Eq. (3) is a reminder to exclude all but one of these surface points. With these restrictions on the space lattice vectors, Eq. (1) reduces to

$$\sum_{g(z_N)} w(\mathbf{k})\exp[i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')] = N_0(\mathbf{g}(z_N))\delta_{\mathbf{R},\mathbf{R}'}, \quad (4)$$

and the coefficients in Eq. (3) are

$$C(\mathbf{R}) \cong \sum_{g(z_N)} w(\mathbf{k})\exp(-i\mathbf{k}\cdot\mathbf{R})f(\mathbf{k})/N_0(\mathbf{g}(z_N)). \quad (5)$$

In Sec. II symmetry-adapted plane waves are defined, and their orthogonality relations obtained for BZ integrals. Then in Sec. III FSA's to the BZ integral orthogonality relations are made. The reason for proceeding in this manner is that it is desired to make very clear the relationship of FSA's to their BZ integral counterparts. The paper concludes in Sec. IV with a discussion of the results and a comparison of FSA's with symmetry-adapted plane waves.

II. SYMMETRY-ADAPTED PLANE WAVES

The use of irreducible symmetry operators (ISO's) to reorganize a basis set of functions into irreducible subspaces or symmetry-adapted functions, which transform according to the irreducible unitary representations (IUR's) of a finite group is well-known.[3] For a summary of the properties of ISO's and their representation in the factored form which will be exploited here the reader is referred to papers by one of us.[4,5]

Symmetry-adapted plane waves (SAPW's) are defined in terms of an ISO $P(R)_{ij}$ for IUR $R = \{D^R(s)\}$ as

$$f_m(R,\mathbf{k})_{ij} = P(R)_{ij}\exp(i\mathbf{k}\cdot\mathbf{R}_m), \quad (6)$$

where the ISO is defined:

$$P(R)_{ij} = (n_R/G^0)\sum_{s\in G} D^R(s^{-1})_{ij}s. \quad (7)$$

Indices $i,j = 1,\dots,n_R$ specify the rows and columns of the matrix representatives, $D^R(s)$. The vectors $\mathbf{R}_m = (i,j,k)a/2$ of the appropriate space lattice are restricted ony to the semi-infinite irreducible wedge, $i \geqslant j \geqslant k \geqslant 0$. In this section considerations are restricted to integrals over the BZ. The properties of the ISO's assure that a group element s operating on a SAPW transforms according to the IUR,

$$sf_m(R,\mathbf{k})_{ij} = \sum_{n=1}^{n_R} f_m(R,\mathbf{k})_{in}D^R(s)_{nj}. \quad (8)$$

It should be noted that one has no *a priori* assurance that the SAPW's defined by Eq. (6) are nonzero or linearly independent. These problems are analyzed below.

A point group element s operating on a plane wave

$$s\exp(i\mathbf{k}\cdot\mathbf{R}_m) \equiv \exp(is^{-1}\mathbf{k}\cdot\mathbf{R}_m) = \exp(i\mathbf{k}\cdot s\mathbf{R}_m) \tag{9}$$

either produces a new plane wave if $s\mathbf{R}_m \neq \mathbf{R}_m$ or recovers the same plane wave when $s\mathbf{R}_m = \mathbf{R}_m$. The subset of group elements which leave $\mathbf{R}_m$ invariant is a subgroup G_m of the point group. All new plane waves are determined by the left coset generators S_m of the point group G with respect to G_m, $G = S_m G_m$. Thus, the problem of finding SAPW's for $\exp(i\mathbf{k}\cdot\mathbf{R}_m)$ is equivalent to reorganizing the stable subspace $S_m \exp(i\mathbf{k}\cdot\mathbf{R}_m)$ into irreducible subspaces. The total number of linearly independent SAPW's equals the number of group elements in left coset S_m.

BZ integrals of products of functions $f(\mathbf{k})$ and $g(\mathbf{k})$ will be represented in the conventional manner $(f(\mathbf{k}),g(\mathbf{k})) = \int_{BZ} d^3k\, f(k)^* g(k)$. Thus, the orthogonality relations for plane waves is expressed

$$(\exp(i\mathbf{k}\cdot\mathbf{R}_m), \exp(i\mathbf{k}\cdot\mathbf{R}_{m'}) = \Omega_{\mathrm{BZ}}\delta_{m,m'}, \tag{10}$$

where Ω_{BZ} is the volume of the BZ. BZ integrals with SAPW's are

$$\begin{aligned}(f_m(R,\mathbf{k})_{ij}, f_{m'}(R',\mathbf{k})_{i'j'})\\ = \delta_{R,R'}\delta_{j,j'}\delta_{m,m'}(\exp(i\mathbf{k}\cdot\mathbf{R}_m), f_m(R,\mathbf{k})_{i'i}),\end{aligned} \tag{11}$$

where the adjoint and multiplication properties of ISO's are used to obtain the right member of Eq. (11). Also, the factor $\delta_{m,m'}$ is extracted because of the orthogonality of distinct plane waves and the requirement that space lattice vectors $\mathbf{R}_m$ be chosen from the irreducible wedge. If $\mathbf{R}_m$ and $\mathbf{R}_{m'}$ form distinct plane waves, then the plane waves generated by the point operations implicit in $f_{m'}(R,\mathbf{k})_{i'i}$ will also be distinct and therefore orthogonal.

Normalization integrals are a special case of Eq. (10). The equation

$$(f_m(R,\mathbf{k})_{ii}, f_m(R,\mathbf{k})_{ii}) = N(R,m)_{ii} \tag{12}$$

defines the diagonal components of the normalization factor. From Eq. (11) it is evident that the only contributions to the normalization integrals arise from the point group elements contained in the subgroup G_m defined above. Hence, using Eq. (7),

$$N(R,m)_{ii} = (n_R/G^0)\sum_{s\in G_m} D^R(s^{-1})_{ii}. \tag{13}$$

A zero value for a normalization integral implies that the SAPW is identically zero. From Eqs. (11) and (12) it is seen that

$$(f_m(R,\mathbf{k})_{ij}, f_m(R,\mathbf{k})_{ij'}) = \delta_{j,j'}N(R,m)_{ii}. \tag{14}$$

Thus, nonzero SAPW's with the same "row" index i, but differing column indices j, j' are orthogonal and have the same normalization. Such functions are called "partner" functions. From the defining Eq. (6) it is seen that there are n_R sets of partner functions. To complete the analysis of the right member of Eq. (11), an analysis of the relationship between partner function sets must be done.

In general a nonzero result may occur in the right member of Eq. (11) when $i' \neq i$. The factored ISO's defined in

TABLE III. Irreducible symmetry operators (ISO's) for the cubic point group.[a]

A. Definitions of cubic point group operators

$E(x,y,z) = (z,y,z)$	$T_1(x,y,z) = (z,x,y)$
$ID_x(x,y,z) = (x,z,y)$	$C_x(x,y,z) = (x,-y,-z)$
$C_y(x,y,z) = (-x,y,-z)$	$I(x,y,z) = (-x,-y,-z)$

B. ISO's corresponding to "pleasant" IUR's of the cubic point group

$P(Asp)_{11} = (E + T_1 + T_1^2)(E + C_x)(E + C_y)(E + sID_x)(E + pI)/48$
$P(Ep)_{11} = (E + \omega^2 T_1 + \omega T_1^2)(E + C_x)(E + C_y)(E + pI)/24$
$P(Ep)_{ij} = ID_x^{\,j-1}P(Ep)_{11}ID_x^{\,i-1}, \quad i,j = 1,2$
$P(Tsp)_{11} = (E + C_x)(E - C_y)(E + sID_x)(E + pI)/16$
$P(Tsp)_{ij} = T_1^{2(j-1)}P(Tsp)_{11}T_1^{\,i-1}, \quad i,j = 1,2,3$

C. Alternative, time-reversal invariant ISO's for the two-dimensional IUR

$P(E'p)_{11} = (2E - T_1 - T_1^2)(E + C_x)(E + C_y)(E + ID_x)(E + pI)/48$
$P(E'p)_{ij} = ((T_1 - T_1^2)/\sqrt{3})^{j-1}P(E'p)_{11}((T_1 - T_1^2)/\sqrt{3}) - (i-1),$
$i,j = 1,2$

[a] Only six of the 48 cubic point group elements appear explicitly in the factored form of the ISO's. These elements are defined by their effect on a 3-vector, (x, y, z), in part A. For multidimensional irreducible unitary representations only, a basic projection operator is listed explicitly along with the relationship to the remaining ISO's. The connection between the present notation and the notation of Bouckaert *et al.* is given in the caption to Table III of Ref. 5. Parameters s, p have values $+1$ or -1 and $\omega = \exp(i2\pi/3)$.

Table IIIB correspond to IUR's which are particularly suitable for analyzing relationships between partner sets of SAPW's. The ISO's are represented in terms of the identity element and inversion element and four more elements of the cubic point group whose choice was motivated by their close association with symmetry lines and planes of the cubic irreducible wedge. The six-point group elements which occur in the ISO's are defined by their transformation properties when acting on a 3-vector in Table IIIA. The nonzero SAPW's produced by these ISO's have the property that they are either orthogonal to one another or, in the case of related sets of partner functions, they differ by a phase factor. This observation is demonstrated below.

First, note that each projection operator $P(R)_{ii}$ in Table IIIB is a linear combination of a set of group elements which are a subgroup designated $G_{R(i)}$ of the cubic point group G. For these IUR's a general ISO is expressed in the form

$$P(R)_{ij} = s^{j-1}P(R)_{11}s^{-(i-1)}, \quad i,j = 1,\ldots,n_R, \tag{15}$$

where point group elements s^i are left or right coset generators for G with respect to the subgroups $G_{R(i)}$. Thus, the SAPW in the right member of Eq. (11) may be written as

$$f_m(R,\mathbf{k})_{i'i} = P(R)_{ii}s^{-(i'-i)}\exp(i\mathbf{k}\cdot\mathbf{R}_m). \tag{16}$$

Since $P(R)_{ii}$ is an ISO for a one-dimensional IUR of $G_{R(i)}$, any element of $G_{R(i)}$ satisfies the relation

$$hP(R)_{ii} = P(R)_{ii}D^R(h)_{ii}. \tag{17}$$

Using the adjoint property[4] of ISO's, Eq. (17) may be rewritten as

$$P(R)_{ii} = D^R(h)^*_{ii}P(R)_{ii}h. \tag{18}$$

Then, if any element h contained in $G_{R(i)}$ is such that element $s_m = hs^{-(i'-i)}$ is contained in group G_m, it follows from Eqs. (16) and (18) that

$$f_m(R,\mathbf{k})_{i'i} = D^R(h)^*_{ii}f_m(R,\mathbf{k})_{ii}. \tag{19}$$

In this case the BZ integral in the right member of Eq. (11) is evaluated as

TABLE IV. Normalization factors for cubic SAPW's formed with respect to the eight types (as distinguished by symmetry) of space lattice vectors, $\mathbf{R}_m = (x,y,z)$, occuring in the irreducible wedge $x \geqslant y \geqslant z \geqslant 0$.[a]

$R\,(ij)$	(0,0,0)	$(x,0,0)$	(x,x,x)	$(x,x,0)$	(x,y,y)	(x,x,z)	$(x,y,0)$	(x,y,z)
$Asp(11)$	$(1+s)(1+p)/4$	$(1+s)(1+p)/24$	$(1+s)/16$	$(1+s)(1+p)/48$	$(1+s)/48$	$(1+s)/48$	$(1+p)/48$	1/48
$Ep(11)$	0	$(1+p)/12$	0	$(1+p)/24$	1/24	1/24	$(1+p)/24$	1/24
$Ep(2i)$	0	$N_{21}=N_{11}$	0	$N_{21}=\omega^2 N_{11}$	$N_{21}=N_{11}$	$N_{21}=\omega^2 N_{11}$	$(1+p)/24$	1/24
$E'p(11)$	0	$(1+p)/6$	0	$(1+p)/48$	1/12	1/48	$(1+p)/24$	1/24
$E'p(2i)$	0	0	0	$N_{21}=-\sqrt{3}N_{11}$	0	$N_{21}=\sqrt{3}N_{11}$	$(1+p)/24$	1/24
$Tsp(11)$	0	$(1+s)(1-p)/8$	$N_{13}=N_{33}$	$(1-p)/16$	$(1+s)/16$	1/16	$(1-p)/16$	1/16
$Tsp(2i)$	0	0	$N_{23}=N_{33}$	$N_{21}=-sN_{11}$	1/16	$N_{21}=-sN_{11}$	$(1-p)/16$	1/16
$Tsp(3i)$	0	0	$(1+s)/16$	$(1+s)(1+p)/16$	$N_{32}=-sN_{22}$	$(1+s)/16$	$(1+p)/16$	1/16

[a] Independent normalization factors are given in terms of the parameters, s, p, etc. characterizing the ISO's of Table III. Relationships between sets of partner functions are indicated by specifying the relationship. For this purpose the notation for normalization matrices $N(R,m)_{ij}$ is abbreviated to N_{ij}.

$$(\exp(i\mathbf{k}\cdot\mathbf{R}_m), f_m(R,\mathbf{k})_{i'i}) = D^R(h)^*_{ii} N(R,m)_{ii}\Omega_{\mathrm{BZ}}. \quad (20)$$

On the other hand, if there is no element in the set $G_{R(i)}s^{-(i'-i)}$ contained in G_m, then the right member of Eq. (11) is zero.

The analysis of BZ integrals of SAPW's is summarized by

$$(f_m(R,\mathbf{k})_{ij}, f_{m'}(R',\mathbf{k})_{i'j'}) = \delta_{R,R'}\delta_{jj'}\delta_{m,m'}N(R,m)_{i'i}\Omega_{\mathrm{BZ}}, \quad (21)$$

where the relevant information for cubic point group normalization matrices is listed in Table IV. From Eq. (14) it is seen that the normalization $N(R,m)_{ii}$ for partner SAPW's is independent of the column index. Nonzero off-diagonal elements of the normalization matrix simply mean that a relationship, Eq. (19), exists between sets of partner functions. Thus, the relevant information consists of explicit results for the independent normalization factors in terms of the parameters of the ISO's as defined in Table III. When sets of partner functions are related, the entry for the normalization factor states the relationship,

$$N(R,m)_{i'i} = D^R(h)^*_{ii} N(R,m)_{ii}. \quad (22)$$

From Eqs. (13) and (19) it follows that the normalization for the related partner set is

$$N(R,m)_{i'i'} = |D^R(h)_{ii}|^2 N(R,m)_{ii}. \quad (23)$$

ISO's of the form shown in Table IIIB having IUR's with the properties in Eqs. (15)–(18) above may be found for many groups of physical interest, including all nonrelativistic point groups and space groups. More complicated, but analagous IUR's may be found for the point and space double groups.[4,5] The late Professor J. M. Keller suggested that these be called "pleasant" representations. It is not known what properties of finite groups might forbid "pleasant" IUR's.

For some purposes the time-reversal (complex conjugation in this context) properties of ISO's transforming according to pleasant IUR's may not be desirable. For example, in Table IIIB, ISO's for IUR's Asp and Tsp are invariant to time-reversal while the two-dimensional IUR Ep has relatively complicated time-reversal properties. Alternative, time-reversal invariant ISO's for the two-dimensional IUR's are given in Table IIIC and distinguished from IUR Ep by attaching a prime ($E'p$). These IUR's and the corresponding ISO's differ by a unitary transformation. It follows that the form of Eqs. (19)–(23) remains the same. However, for the IUR $E'p$ the proportionality constant in Eqs. (19), (20), (22), and (23) is not simply a matrix element as above, but is a combination of matrix elements. The normalization factors for both two-dimensional IUR's are listed in Table IV in the condensed form described below [Eq. (21)].

III. FINITE SUM APPROXIMATIONS

FSA's to BZ integrals may be viewed as a partition of the BZ volume into N_0 identical cells surrounding each point $\mathbf{k}$ in the sum. Thus, in the notation for finite sums summarized in Sec. I,

$$\int_{\mathrm{BZ}} d^3k\, f(\mathbf{k}) = \sum_{g(z_N)} w(\mathbf{k})(\Omega_{\mathrm{BZ}}/N_0) f(\mathbf{k}), \quad (24)$$

where the weight factor $w(\mathbf{k})$ is unity for points inside Ω_{BZ} and is such that $w(\mathbf{k})\Omega_{\mathrm{BZ}}/N_0$ is the fraction of the cell inside Ω_{BZ} in the case of points on the BZ surface. Also, note that $\Omega_{\mathrm{BZ}} = (2\pi)^3/\Omega_0$, where Ω_0 is the volume of a primitive cell in 3-space. Thus, the cell volume $\Omega_{\mathrm{BZ}}/N_0 = (2\pi)^3/\Omega$, where $\Omega = N_0\Omega_0$ is the volume of the primitive cell in 3-space containing space lattice vectors which are inequivalent with respect to the points in an FSA.

In the integral limit N_0 approaches infinity and no finite space lattice vectors can be on the surface. For finite sums and the consequent finite volume Ω, surface points $\mathbf{R}_{m'}$ may occur in an SAPW which differ from the base point $\mathbf{R}_m$ (in the irreducible wedge) by a space lattice vector which is equivalent to $\mathbf{0}$ relative to the summation points. Therefore, the normalization for SAPW's formed with respect to surface points is affected. This effect will be represented as a factor $M(R,g,m)$ as it is dependent on the IUR, the sum type, and the point $\mathbf{R}_m$ considered. Thus, the finite sum analog to Eq. (21) is

$$\sum_{g(z_N)} w(\mathbf{k}) f_m(R,\mathbf{k})^*_{ij} f_{m'}(R',\mathbf{k})_{i'j'}$$

$$= \delta_{R,R'}\delta_{j,j'}\delta_{m,m'}N_0(g(z_N))N(R,m)_{i'i}M(R,g,m), \quad (25)$$

where the finite point set $g(z_N)$ must be invariant to point group operations. Of course, one may obtain the same result directly by using Eq. (1) to evaluate the left member of Eq. (25). This is how the surface normalization factors $M(R,g,m)$ displayed in Table V were evaluated. The "derivation" of the finite sum form from the BZ integral is supplied for the in-

TABLE V. Weights and normalization factors for surface points.[a]

Points	$w(\mathbf{k})$	$Asp(11)$	$E'p(11)$	$E'p(2i)$	$Tsp(11)$	$Tsp(2i)$	$Tsp(3i)$
A. Zone $s_n = S_N$							
(4) $(N,N,N)/2$	1/8	$(1,3p,3,p)$	x	x	$(1,-p,-1,p)$	x	x
(3) $(N,N,2k)/2$, $N>2k$	1/4	$(1,2p,1,0)$	$(1,2p,1,0)$	x	$(1,0,-1,0)$	x	$(1,0,-1,0)$
(2) $(\frac{1}{2}N,j,k)$, $\frac{1}{2}N>j$	1/2	$(1,p,0,0)$	$(1,p,0,0)$	$(1,p,0,0)$	$(1,p,0,0)$	$(1,-p,0,0)$	$(1,-p,0,0)$
B. Zone $b_N = F_N$							
(5) $(N,0,0)$	1/6	$(2,0,4,0)$	$(2,0,-2,0)$	x	0	x	x
(4) $(N,N,N)/2$	1/4	$(1,0,3,0)$	x	x	$(1,0,-1,0)$	x	x
(3) $(i,N-i,N-i)$, $N>i>\frac{1}{2}N$	1/3	$(1,0,2,0)$	$(1,0,-1,0)$	x	$(1,0,0,0)$	$N_{21}\propto N_{11}$	x
(2) $(N,N,2k)/2$, $N>2k$	1/2	$(1,0,1,0)$	$(1,0,1,0)$	x	$(1,0,-1,0)$	x	$(1,0,1,0)$
(2′) $(i,N-i,k)$, $N>i>\frac{1}{2}N$; $N>i+k$	1/2	$(1,0,s,0)$	$(1,0,-\frac{1}{2},0)$	$N_{21}\propto N_{11}$	$(1,0,0,0)$	$N_{21}\propto N_{11}$	$(1,0,s,0)$
C. Zone $f_N = B_N$							
(5) $N,\frac{1}{2}N,0)$	1/4	$(2,0,0,2s)$	$(2,0,0,-1)$	$N_{21}\propto N_{11}$	0	$(2,0,0,2s)$	0
(4) (N,j,k), $\frac{1}{2}N>j$; (HF)	1/2	$(1+p,0,0,0)$	$(1+p,0,0,0)$	$(1+p,0,0,0)$	$(1+p,0,0,0)$	$(1-p,0,0,0)$	$(1-p,0,0,0)$
(3) $(i,\frac{1}{2},N,N-i)$, $N>i\geqslant\frac{1}{2}N$	1/2	$(1,0,0,s)$	$(1,0,0,-\frac{1}{2}p)$	$(1,0,0,\frac{1}{2}p)$	$(1,0,0,0)$	$(1,0,0,ps)$	$(N_{31}\propto N_{11})$
(2) (i,j,k), $N>i>\frac{1}{2}N>j>\frac{1}{4}N$; HF	1	$(1,0,0,0)$	$(1,0,0,0)$	$(1,0,0,0)$	$(1,0,0,0)$	$(1,0,0,0)$	$(1,0,0,0)$

[a] Points are specified by 3-tuples (i,j,k) where $i\geqslant j\geqslant k\geqslant 0$. A restriction to the hexagonal face $2(i+j+k)=3N$ is denoted by HF. Symmetry-related, distinct HF points have been combined. Weights are listed in the column $w(\mathbf{k})$. Surface normalization factors are represented $M(R,g,m)=(m_0n_0+m_1n_1+m_2n_2+m_3n_3)/n_0$, where the sum-dependent factors n_i are given in Table II. Only the 4-tuples (m_0,m_1,m_2,m_3) are listed for each irreducible representation R. Interior points have weights $w(\mathbf{k})=1$ and surface factor $M(R,g,m)=1$.

sight into the meaning of FSA's and to motivate the parametrization of Eq. (25).

The results for the surface normalization factors listed in Table V might appear deceptively simple. To obtain these results, it was necessary to analyze each distinct type of surface point which occurs for each type of zone. For example, in the case of zone B_N four entries appear in Table VC which represent the combined results for 15 distinct types of points. An example of the derivation for a particular point is given in the Appendix.

Six types of k-vectors on the hexagonal face (HF) of the BZf_N have been combined because they are symmetry-related. The points $(N,\frac{1}{4}N,\frac{1}{4}N)$ and $(N,j,\frac{1}{2}N-j)$ listed under the entry Table VC(4) of themselves have weight $w(k)=\frac{1}{4}$, but they are related by the symmetry operation D_y $[D_y(x,y,z) = (-z,-y,-x)]$ to points in the BZ which differ from independent points in the IBZ by a reciprocal lattice vector, (N,N,N). These have been listed as one point in Table VC(4) with weight $w(\mathbf{k})=\frac{1}{2}$. Also, note that the combined points of this type (N,j,k) include values $\frac{1}{2}N\geqslant j\geqslant k\geqslant 0$, which are not all HF points. The parentheses about the HF is a reminder that this is not an absolute restriction to the HF. Similarly, the points (i,j,k) of Table VC(2) of themselves have weight $w(k)=\frac{1}{2}$ and are restricted absolutely to the HF. They are related as above by the symmetry operation D_y plus a reciprocal lattice vector to inequivalent points on the surface of the IBZ and have been combined.

In Table V the entries "x" indicate that a relationship or zero value for the SAPW occurs without consideration of surface effects as shown in Table IV. Only zero values or relationships which arise from surface effects are recorded explicitly in Table V. In the case of relationships only a proportionality is indicated. This is because there is an ambiguity in sign between plane waves formed with respect to related surface points $\mathbf{R}_m$. This sign depends on the type of sum over k-vectors and is not of importance in applications of the results.

Obviously, it would be easy to overlook a relationship or a factor in the derivation of these results. The tabulated quantitites have been checked in the context of applications to expansions of periodic functions in terms of finite sums of SAPW's. For this purpose the basis set of SAPW's must exclude related SAPW partner sets and, of course, trivial SAPW's which are identically zero. With these provisions Eq. (25) becomes

$$\sum_{g(z_{N})} w(\mathbf{k}) f_m(R,\mathbf{k})^*_{ij} f_{m'}(R',\mathbf{k})_{i'j'} = \delta_{R,R'}\,\delta_{i,i'}\,\delta_{j,j'}\,\delta_{m,m'}\,N(g,z,R,m)_i, \tag{26}$$

where the combined normalization factor is

$$N(g,z,R,m)_i = N_0(g(z_N))N(R,m)_{ii}M(R,g,m). \tag{27}$$

The tables were verified by evaluating the left member of Eq. (26) directly and comparing this result to the right member as obtained using the tabulated information. This process was done completely only for $N=2$ as the direct evaluation is quite expensive on the computer. The special case of Eq. (26) where $m=m'$ and the space inversion parities are the same was verified for $N=2-8$. All cases included in the tables were checked in this way.

The coefficients in the expansion of a periodic function of arbitrary rotational symmetry in terms of a finite sum of SAPW's,

$$F(\mathbf{k}) = \sum_{G(RZ_N)} \sum_R \sum_{ij} C(R,m)_{ij} f_m(R,\mathbf{k})_{ij} \tag{28}$$

are obtained using Eqs. (26) and (27) to be

$$C(R,m)_{ij} = \sum_{g(z_N)} w(\mathbf{k}) f_m(R,\mathbf{k})^*_{ij} F(\mathbf{k})/N(g,z,R,m)_i. \tag{29}$$

Equations (28) and (29) are the SAPW analogs to the plane wave expansion, Eqs. (3) and (5). For a given finite set of k-vectors, $g(z_N)$ the expansion represented by Eqs. (3) and (5) is just a reorganized version of the expansion represented by Eqs. (28) and (29).

The BZ sums in Eqs. (29) may be reduced to sums over points in the irreducible wedge (IBZ). For this purpose, let $G_{\mathbf{k}}$ be the subgroup of elements h for which $h\mathbf{k} = \mathbf{k}$, and let $S_{\mathbf{k}}$ be the set of left coset generators for the entire group, $G = S_{\mathbf{k}} G_{\mathbf{k}}$. Then,

$$C(R,m)_{ij} = \sum_{g(rz_N)} \sum_{s \in G} (w(s\mathbf{k})/{}^0G_{\mathbf{k}}) f_m(R,\mathbf{k})^*_{ij} F(s\mathbf{k})/N(g, z, R, m)_i, \tag{30}$$

where $g(rz_N)$ represents those k-vectors of the set $g(z_N)$ contained in the IBZ. Using the definition and properties of ISO's given in Eqs. (6)–(8), Eq. (30) becomes

$$C(R,m)_{ij} = \sum_{g(rz_N)}{}' ({}^0d_{\mathbf{k}}/n_R) \sum_{n=1,n_R} f_m(R,\mathbf{k})_{in} {}^*P(R)_{jn} \times F(\mathbf{k})/N(g, z, R, m)_i, \tag{31}$$

where

$${}^0d_{\mathbf{k}} = {}^0s_{\mathbf{k}}\, w(\mathbf{k}) \tag{32}$$

is the order of the group of the k-vector $d_{\mathbf{k}}$, which is the set of group elements h such that $h\mathbf{k} = \mathbf{k} + \mathbf{K}$. The prime on the sum in Eq. (31) is a reminder to exlude symmetry-related hexagonal face (HF) points.

Since, the weights to be associated with HF points are complicated and may be an element of uncertainty to the user, they are summarized here.

In the cases [Eqs. (26) and (29)] of results where the sums extend over the entire BZ, the weights for the surface points are those given in Table V except for the HF points included in entries, Table VC (4) and (2). These weights must be divided by a factor 2. The corresponding symmetry-related HF points must be included in the sum and carry the same weight. The presentation here reflects the manner in which sets of points were generated to test Eq. (26) on a computer. First, points of a specified type were found within the IBZ according to the zonal restrictions of Table IB. HF points (i,j,k) of the type where $j > \frac{1}{2}N$ are exluded from this basic set of points. Interior points are assigned weight $w(\mathbf{k}) = 1$ and surface points are assigned weights as given in Table V. In performing the sum the module of points generated by the elements of $S_{\mathbf{k}}$ is found. In the case of HF points the weight from Table V is halved, and an additional module generated corresponding to the symmetry-related HF point.

For sums over points in the IBZ as in Eq. (31) the weights are taken exactly as given in Table V with the provision that they symmetry-related HF points not specified in Table V are excluded from the sum.

IV. DISCUSSION

The FSA's to BZ integrals are given in Eqs. (2), (26)–(32) with parameters and conditions summarized in Tables I–V. The results of I, Eqs. (19) and (20), correspond to the present results in the case where $F(\mathbf{k})$ is a periodic function invariant to rotations of the cubic group, $P(A++)_{11}F(\mathbf{k}) = F(\mathbf{k})$.

One objective of the present work is to determine how the special character found for Chadi–Cohen points manifests itself in the case of other symmetries besides the identity IUR. Such a comparison is made in Table VI in the case of functions $F(\mathbf{k})$ which are invariant to translations by vectors reciprocal to the fcc space lattice. From Table II under the column heading Z_N it is seen that the sum pairs $(s_N, fu_{2N}),(b_N, feu_N),(f_N, beu_N)$ have similar accuracy in the sense that independent SAPW's are formed with respect to lattice vectors inside zones S_{2N}, B_N, and F_N, respectively. These sum pairs are also complementary in the sense they may be combined to form lattice sums f_{2N}, s_{2N}, and s_{2N}, respectively. This is the reason that the sublattices feu and beu were deemed worthy of special consideration.

Chadi–Cohen points are special only in the context of approximations to certain coefficients in the finite sum expansions of functions which transform according to a particular IUR of the group. The Chadi–Cohen points correspond to sublattice fu with appropriate restrictions to an IBZ. The first six rows of Table VI compare sums over a simple cubic array of points s to sums over sublattice fu. For example, for sum type s with $N = 4$ there are 19 terms in the FSA's to the coefficients, Eq. (31). For IUR $A++$ there are 19 independent SAPW's of $A++$ symmetry and no SAPW's corresponding to surface points in the zone S_8 vanish. Similarly, for this row IUR $T++$ has 16 independent partner sets or 48 SAPW's and 10 partner sets corresponding to surface points vanish. The sum fu with $N = 8$ also corresponds to a partition of the BZ into 256 cells of equal volume, but there are only ten terms in the sum in Eq. (31). In this case there are ten independent SAPW's of $A++$ symmetry and nine SAPW's corresponding to surface points in zone S_8 vanish identically for all k-vectors in the sum fu. It is in this context that the special nature of the Chadi–Cohen points is recognized as a surface effect. However, the surface effect does not persist in any easily definable manner for symmetries other than IUR $A++$. Because the results for the simple cubic and body-centered space lattices are very similar to the results presented in Table VI, they will not be reproduced here.

Crucial to any practical calculation is an estimate of the accuracy of the method. In the present work we have simply noted that the employment of a particular type of sum corresponds to partitioning the BZ into N_0 cells of equal volume. If one knows the extremes of variation of the function under consideration within such cells, then one can made a crude estimate of the associated error. A directly related means of discussing the error was used by MacDonald.[6] To each sum in k-space there corresponds a zone in real space containing lattice vectors for which independent information can be obtained. Table II specifies the zones which correspond to each sum type, and Table IB defines the zones. MacDonald focused on the BZ average and quantified the precision of a sum type as the number of rotationally distinct lattice vectors in order of increasing magnitude which give zero contribution to a BZ average. In the present context this measure of precision corresponds to counting the number of rotationally distinct lattice vectors in the zones of Table II where N is

TABLE VI. Comparison of the number of independent SAPW partner sets for sums of similar accuracy.[a]

Sum	N	N_0	N_r	$A++$	$A+-$	$A-+$	$A--$	$E'+$	$E'-$	$T++$	$T+-$	$T-+$	$T--$	# SAPW's
s	4	256	19	19	4	4	1	20	4	16	21	7	16	256
				(0)	(6)	(0)	(1)	(0)	(6)	(10)	(9)	(7)	(4)	
fu	8	256	10	10	10	2	2	10	10	20	20	12	12	256
				(9)	(0)	(2)	(0)	(10)	(0)	(6)	(10)	(2)	(8)	
s	8	2 048	85	85	40	40	19	120	56	128	155	93	128	2 048
				(0)	20)	(0)	(9)	(0)	(28)	(36)	(25)	(31)	(16)	
fu	16	2 048	60	60	60	28	28	84	84	144	144	112	112	2 048
				(25)	(0)	(12)	(0)	(36)	(0)	(20)	(36)	(12)	(32)	
s	16	16 384	489	489	336	336	231	816	560	1024	1143	889	1024	16 384
				(0)	(72)	(0)	(49)	(0)	(120)	(136)	(81)	(127)	(64)	
fu	32	16 384	408	408	408	280	280	680	680	1088	1088	960	960	16 384
				(81)	(0)	(56)	(0)	(136)	(0)	(72)	(136)	(56)	(128)	
f	4	64	8	8	2	0	0	5	1	5	5	1	3	64
				(0)	(0)	(1)	(0)	(2)	(0)	(1)	(4)	(1)	(2)	
beu	4	192	11	7	1	1	0	6	1	6	7	2	4	80
				(1)	(1)	(0)	(0)	(1)	(0)	(0)	(2)	(0)	(1)	
f	8	512	29	29	14	6	3	30	14	36	41	19	28	512
				(0)	(0)	(4)	(1)	(6)	(2)	(4)	(9)	(7)	(8)	
beu	8	1 536	56	28	13	10	4	32	15	38	44	24	31	560
				(1)	(1)	(0)	(0)	(4)	(1)	(2)	(6)	(2)	(5)	
f	16	4 096	145	145	100	68	47	204	140	272	299	205	204	4 096
				(0)	(0)	(16)	(9)	(20)	(12)	(16)	(25)	(31)	(32)	
beu	16	12 288	344	144	99	84	56	208	143	276	304	224	253	4 256
				(1)	(1)	(0)	(0)	(16)	(9)	(12)	(20)	(12)	(19)	
b	4	128	11	11	1	1	1	10	2	9	9	3	9	128
				(0)	(3)	(2)	(0)	(2)	(2)	(3)	(7)	(4)	(2)	
feu	4	128	8	8	3	3	0	10	2	7	12	4	7	128
				(3)	(1)	(0)	(1)	(2)	(2)	(5)	(4)	(3)	(4)	
b	8	1 024	45	45	18	18	11	60	28	66	75	45	66	1 024
				(0)	(7)	(4)	(1)	(4)	(7)	(9)	(14)	(12)	(6)	
feu	8	1 024	40	40	22	22	8	60	28	62	80	48	62	1 024
				(5)	(3)	(0)	(4)	(4)	(7)	(13)	(9)	(9)	(10)	
b	16	8 192	249	249	164	164	119	408	280	516	567	441	516	8 192
				(0)	(18)	(8)	(6)	(8)	(23)	(27)	(31)	(34)	(17)	
feu	16	8 192	240	240	172	172	112	408	280	508	576	448	508	8 192
				(9)	(10)	(0)	(13)	(8)	(23)	(35)	(22)	(27)	(25)	

[a] The results are for the fcc space lattice. The sum type, N, the number of k-vectors in the BZ(N_0), and the number of k-vectors in the IBZ(N_r) are listed in the first four columns. The corresponding number of independent partner sets for each IUR is specified in the next 10 columns. Immediately below these entries (enclosed by parentheses) are listed the number of partner sets which are zero because they have a zero value for their surface normalization factors. The last column lists the total number of independent SAPW's.

replaced by $2N$ and exluding the lattice vectors which are surface points of zone Z_{2N}. It appears that MacDonald also excludes any interior points which are larger in magnitude than the smallest surface lattice vector. It should be noted that MacDonald's DPC sum types are identical to our lattice f in the case of the fcc space lattice. For the bcc space lattice, MacDonald's DPC sets correspond to our sublattice feu for even q and to our lattice b for odd q. MacDonald's q is identical to our N. We have little motivation to pursue this path, since MacDonald's measure of precision is only applicable in the context of symmetry type $A++$.

For practical purposes the use of complementary sums is the most important feature that our investigation of FSA's to BZ integrals has produced. For example, suppose that the sum s with $N=4$ and the sum fu with $N=8$ are used to evaluate a coefficient for which both sums provide information. The two results tend to form an upper bound with respect to a more accurate calculation. As can be seen from the definitions of Table IA, combining these points corresponds to a sum over the lattice f with $N=8$. The mean of the two original calculations corresponds to a more accurate result, and the deviation of this mean from the original results provides an estimate of the error. The process may be continued by forming the sum beu with $N=8$. This sum combined with sum f for $N=8$ produces the sum s with $N=8$, a more accurate result. Table VI provides the information regarding the number of points involved at each step in the successive approximation scheme outlined above in the case of the fcc space lattice. For other lattices the progression is very similar.

ACKNOWLEDGMENT

One of us (J. O.) wishes to acknowledge financial support from the University of Science, Malaysia.

APPENDIX: EXAMPLES OF THE ANALYSIS FOR SURFACE POINT NORMALIZATION FACTORS

The analysis will treat points of the form $(x,y,0)$. Such points serve to demonstrate the essential features of the analysis without excessive complication.

First, observe that the invariance group of the point $(x,y,0)$ is $G_m = (E,IC_z)$ where element $IC_z = (I)(C_z)$ in an obvious extension of the notation for group elements defined in Table IIIA. Next determine the total group of the point G^{zm} for each of the zones. G^{zm} includes elements which differ from the base vector by a lattice vector for zone Z. The total group may be compactly expressed in terms of a set of coset generators C^{zm}, $G^{zm} = C^{zm}G_m$. Note that the analysis here is exactly the same as the determination of the group for a k-vector on the surface of the BZ. The group elements in C^{zm} are those needed to unite the pieces of a surface point into a whole point.Thus, the number of coset generators $^0C^{zm}$ determines the weight for a surface point, $w(k) = 1/{}^0C^{zm}$. The exception to this rule are points on the hexagonal face of the zone $f_N = B_N$, as discussed in the text.

Zone S_N contains one surface point, $(\frac{1}{2}N, j, 0)$ with $\frac{1}{2}N > j > 0$ of the type $(x, y, 0)$. This point is denoted $S(2)$ because it is contained in the combined set of surface points given in Table VA with point index (2). The coset generators are $C^{S_m} = (E,IC_x)$. Similarly, the point $F(2')$, $(i,N-i,0)$, $N > i > \frac{1}{2}N$, of zone F_N has a coset $C^{F_m} = (E, ID_{z'})$, where $ID_{z'} = (C_z)(T_1^2)(ID_x)$. Two distinct surface points of the type $(x, y, 0)$ occur for zone B_N. They are $B(4)$, $(N,j,0)$, $N > j > 0$ with $C^{B_m} = (E, IC_x)$ and $B(5)$, $(N,\frac{1}{2}N,0)$ with $C^{B_m} = (E, D_y)(E, IC_x)$, where $D_y = (I)(T_1)(ID_x)$.

The next step is to evaluate the normalization factors as defined by Eqs. (26) and (27) and to determine any additional relationships. The adjoint property of the ISO's is maintained for finite sums over sets of points in the BZ which are invariant to point group rotations. Equation (26), specialized to the diagonal terms, is rewritten as

$$\sum_{g(h_N)} w(\mathbf{k}) \exp(-\mathbf{k}\cdot\mathbf{R}_m) f_m(R,\mathbf{k})_{ii} = N(g,h,R,m)_i. \quad \text{(A1)}$$

The normalization factors are evaluated from Eqs. (A1). The finite sums are evaluated with Eq. (1). For surface terms only those group elements contained in the total group of the point, listed above for points of the form $(x,y,0)$, contribute to the normalization. The objective will be to obtain most of the results pertinent to Tables IV and V simultaneously.

The factored projection operators may be rewritten to include explicitly a factor involving group element IC_z of G_m. For example,

$$P(Tsp)_{11} = (E + sID_x)(E + pI)(E + C_x)(E - pIC_z)/16 \quad \text{(A2)}$$

is identical to the form shown in Table IIIB because $pI(E + pI) = (E + pI)$ and inversion commutes with all group elements. When this operator is applied to a plane wave as defined by Eqs. (6) and (7) the result may be expressed as

$$f_m(Tsp)_{11} = \tfrac{1}{16}(1-p)(E + pIC_x + \cdots)\exp(i\mathbf{k}\cdot\mathbf{R}_m), \quad \text{(A3)}$$

where R_m is of the form $(x,y,0)$ and only operators contained in the cosets for the expansion G^{zm} in terms of G_m are listed from the remaining factors in Eq. (A2) since only these terms contribute to the normalization. For interior points $(x, y, 0)$ the normalization found from Eqs. (1) and (A1) is

$$N(g,h,Tsp,m)_1 = N_0(g(h_N))(1-p)/16 \quad \text{(A4)}$$

and the normalization factor is $N(Tsp)_{ii} = (1-p)/16$ as defined by Eq. (27) and listed in row 6, column 7 of Table IV. For surface points the group element IC_x produces an additional contribution to the normalization. In the case of the point S (2) defined above ,the relation $IC_x(\frac{1}{2}N, j, 0(= (\frac{1}{2}N, j, 0) - (N,0,0)$. The lattice vector $(M,0,0)$ is in the set NES and from Eq. (1) can be seen to have a surface normalization factor

$$M(Tsp,g,m) = (n_0 + pn_1)/n_0. \quad \text{(A5)}$$

The corresponding entry for point $S(2)$ under the column heading $Tsp(11)$ is $(1,p,0,0)$. All entries for normalization factors and surface normalization factors were obtained by similar analysis.

Finally, note that no effect on the surface factors is evidenced by symmetry relations $ID_{z'}$ and D_y in the example above. These group elements provide new relations between partner sets. For example, for point $F(2')$, $ID_{z'}(i,N-i,0) = (i,N-i,0) - (N,N,0)$. Hence, $ID_{z'}\exp(i\mathbf{k}\cdot\mathbf{R}_m) = (+/-)\exp(i\mathbf{k}\cdot\mathbf{R}_m)$, where the sign is determined by the sum type. The exact sign is the sign of n_2 as given in Table II. Then, using the ISO's defined in Table IIIB,

$$\begin{aligned} P(Tsp)_{21}\exp(i\mathbf{k}\cdot\mathbf{R}_m) &= (+/-)P(Tsp)_{11}T_1ID_{z'}\exp(i\mathbf{k}\cdot\mathbf{R}_m) \\ &= -s(+/-)P(\mathrm{Tsp})_{11}\exp(i\mathbf{k}\cdot\mathbf{R}_m), \end{aligned} \quad \text{(A6)}$$

where the right member follows because $T_1ID_{z'} = C_yID_x$ and $P(Tsp)_{11}C_yID_x = -sP(Tsp)_{11}$. In general, if a group element relating zone surface points does not contribute directly to the surface factor, then an additional relationship between partner sets will occur.

[1] D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973); H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976), first recognized the Chadi–Cohen special points and displaced lattice FSA's.
[2] N. O. Folland, Phys. Rev. B. **22**, 3669 (1980).
[3] E. P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), Chap. 12.
[4] N. O. Folland, J. Math. Phys. **18**, 31 (1977).
[5] N. O. Folland, J. Math. Phys. **20**, 1274 (1979).
[6] A. H. MacDonald, Phys. Rev. B **18**, 5897 (1978).

Erratum: On Euler characteristics of compact Einstein 4-manifolds of metric signature $(+ + - -)$ [J. Math. Phys. 22, 979 (1981)]

Yasuo Matsushita
Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, Kyoto, Japan

(Received 21 April 1982; accepted for publication 29 April 1982)

PACS numbers: 02.40. – k, 99.10. + g

Matrix (16a) on page 980 should read

$$P_\sigma = \begin{bmatrix} \mu_1 + \sigma\nu_1 & 0 & 0 \\ 0 & -\mu_2 + \sigma\nu_2 & 0 \\ 0 & 0 & -\mu_3 + \sigma\nu_3 \end{bmatrix}.$$

Equation (26b) on page 981 should read

$$\bar{p}_1[M_2] = -\frac{S}{4\pi^2}\int_{M_2} \nu_1 w + \frac{2}{\pi^2}\int_{M_2} (\mu_2\nu_2) w, \quad \nu_1 \neq 0.$$

Inequality (ii) in Corollary 5 on page 981 should read $\chi[M_1] \geqslant 2$.

Equation (32) on page 982 should read

$$ds^2 = a^2(d\xi^2 + \sin^2\xi\, d\phi^2 - d\eta^2 - \sin^2\eta\, d\psi^2).$$

Erratum: Time-dependent, finite, rotating universes [J. Math. Phys. 22, 2699 (1981)]

Marcelo J. Rebouças and J. Ademir S. de Lima
Centro Brasilerio de Pesquisas Físicas/CNPq, Av. Wenceslau Braz, 71, 22.290-Rio de Janeiro-RJ-Brazil

(Received 19 March 1982; accepted for publication 26 March 1982)

PACS numbers: 98.80.Dr, 04.20.Jb, 04.20. – q, 98.80.Ft, 99.10. + g

1. Equation (3.5) should read

$$Y_1 = \frac{1}{B_0} X_1, \; Y_2 = \frac{1}{B_0} X_2, \; Y_3 = \frac{1}{(\alpha^2 - 1)^{1/2} A} X_3.$$

2. Our solutions are not spatially homogeneous: the t-constant sections (Bianchi IX) are timelike and not spacelike as we have stated in the Introduction.

We are indebted to R. T. Jantzen for calling our attention to these points.