Reduction formulas for higher-order indices and anomaly

Yasushi Fujimoto[a] and Robert Grigjanis
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2J1, Canada

(Received 30 December 1983; accepted for publication 23 March 1984)

Reduction formulas for $\{1^k\}$ and $\{k\}$ of SU(N) are derived which are useful for obtaining higher-order indices. We also touch on the anomaly in higher-dimensional theories.

PACS numbers: 02.20. + b, 11.15. − q

I. INTRODUCTION

Recently there have appeared several works on the higher indices[1–5] of the representations of classical groups. Indices are of importance for particle physics and other branches of physics. The second-order index which is proportional to the so-called quadratic Casimir invariant[6] appears most frequently in physics. The fourth-order index l_4 (see Refs. 5 and 7) [see Eq. (4)] has been used to obtain the correct Kronecker products and subgroup decomposition. The third-order index yields the anomaly[8] in the four-dimensional theory. To compute the indices in general is a hard task although some progress has been made over the last few years.

In the present paper we would like to present some results on the indices of SU( ). Specifically we present reduction formulas for $\{1^k\}$($=k$ times antisymmetrized representation) and $\{k\}$($=k$ times symmetrized representation). (Reduction here refers to expressing an invariant associated with a representation in terms of those associated with the fundamental representation only.) Our expectation is that the reduction formulas for $\{1^k\}$ $\{k\}$ will enable us to deal with other representations since they can be formed out of $\{1^k\}$ and $\{k\}$. Section II is devoted to the derivation of the reduction formulas. In Sec. III we will be in contact with physics and touch on the higher-order anomaly which was recently discussed by Frampton.[9,10]

II. DERIVATION OF THE REDUCTION FORMULAS

The quantity we are concerned with is[5]

$$\sum_p \mathrm{Tr}(\tilde{X}_\mu \tilde{X}_\nu \cdots \tilde{X}_\rho), \tag{1}$$

where $\tilde{X}_\mu$ is the generator of a representation of a classical group, and Σ_p is a summation over $p!$ permutations of indices $\mu, \nu, \ldots, \rho$. Various invariants can be defined through Eq. (1). For instance,

$$D^p(\rho) \equiv g^{\mu_1 \cdots \mu_p} \sum_p \mathrm{Tr}(\tilde{X}_{\mu_1} \tilde{X}_{\mu_2} \cdots \tilde{X}_{\mu_p}) \tag{2}$$

(ρ denotes a representation), is one kind of pth-order index and is related to the pth-order Casimir invariant $J_p(\rho)$ by

$$D^p(\rho) = J_p(\rho) d(\rho), \tag{3}$$

where $d(\rho)$ is the dimension of ρ. Other invariants discussed in Ref. 5 are

$$\bar{D}^{(4)}(\rho) \equiv g^{\mu_1\mu_2} g^{\mu_3\mu_4} \sum \mathrm{Tr}(\tilde{X}_{\mu_1} \tilde{X}_{\mu_2} \tilde{X}_{\mu_3} \tilde{X}_{\mu_4}), \tag{4}$$

and

$$l_4(\rho) \equiv \mathrm{Tr}(g^{ij} H_i H_j)^2,$$

where in the latter H_i are restricted to the generators of Cartan subalgebra.

So far as we know $D^p(\rho)$ is calculable in general. However as for other such as l_{2p}, an explicit compact expression has not been given. We try to obtain it for $\{1^k\}$ and $\{k\}$ as a first step toward a general expression. To deal with the problem as we follow the approach of Okubo and Patera[5] and reduce the expression of Eq. (1) to terms associated with the fundamental representation. Now to the derivation of the reduction formulas.

Let T^α be a generator of the fundamental representation of SU(N),

$$(T^\alpha)^a_b, \quad \alpha = 1,2,\ldots,N^2-1, \quad a,b = 1,2\ldots,N. \tag{5}$$

Also let δ^a_b be the $N \times N$ unit matrix. Then the generator $\tilde{T}^\alpha$ of $\{1^k\}$ can be written as

$$(\tilde{T}^\alpha)^{a_1\cdots a_k}_{b_1\cdots b_k} \equiv \frac{1}{(k-1)!} (T^\alpha \underbrace{\wedge \delta \wedge \delta \wedge}_{k-1} \cdots \wedge \delta)^{a_1\cdots a_k}_{b_1\cdots b_k}, \tag{6}$$

where the wedge products denote antisymmetrization on the lower indices only. Thus, if A, B, and C are $N \times N$ matrices,

$$\begin{aligned}(A \wedge B \wedge C)^{a_1a_2a_3}_{b_1b_2b_3} &\equiv (A^{a_1}_{b_1} B^{a_2}_{b_2} C^{a_3}_{b_3} - A^{a_1}_{b_1} B^{a_2}_{b_3} C^{a_3}_{b_2} + A^{a_1}_{b_3} B^{a_2}_{b_1} C^{a_3}_{b_2} \\ &\quad - A^{a_1}_{b_3} B^{a_2}_{b_2} C^{a_3}_{b_1} + A^{a_1}_{b_2} B^{a_2}_{b_3} C^{a_3}_{b_1} - A^{a_1}_{b_2} B^{a_2}_{b_1} C^{a_3}_{b_3}).\end{aligned} \tag{7}$$

In the above the matrix multiplication is defined by

$$(AB)^{a_1\cdots a_k}_{b_1\cdots b_k} \equiv A^{a_1\cdots a_k}_{c_1\cdots c_k} B^{c_1\cdots c_k}_{b_1\cdots b_k}. \tag{8}$$

Also trace is defined by

$$\mathrm{Tr}\, A \equiv A^{a_1\cdots a_k}_{a_1\cdots a_k}. \tag{9}$$

It is not very difficult to show that the following equation holds:

$$\begin{aligned}&\mathrm{Tr}(A_1 \wedge A_2 \wedge \cdots A_k) \\ &= (-1)^k \sum_{\{Xi\}} (-1)^{\Sigma_i X_i} \\ &\quad \times \left\{ \sum{}' (\mathrm{Tr}\, A)^{X_1} (\mathrm{Tr}\, A^2)^{X_2} \cdots (\mathrm{Tr}\, A^k)^{X_k} \right\},\end{aligned} \tag{10}$$

where $\{X_i\}$ is a partition satisfying

$$\sum_{i=1}^k i X_i = k, \tag{11}$$

[a] Address after September 1983: International Center for Theoretical Physics, Miramare, I-34100 Trieste, Italy.

 0022-2488/84/113141-07$02.50

and

$$(\mathrm{Tr}\,A^p)^{x_1} \equiv \mathrm{Tr}\,A^p_{i_1}\,\mathrm{Tr}\,A^p_{i_2}\cdots\mathrm{Tr}\,A^p_{i_{x_1}}, \tag{12}$$

where Σ' denotes a sum over permutations of $(1,2,\ldots,k)$ which yields distinct terms. That is, one does not include cyclic permutations within a trace, nor permutations which simply exchange two traces. For instance,

$$\begin{aligned}\mathrm{Tr}(A\wedge B) &= \mathrm{Tr}\,A\,\mathrm{Tr}\,B - \mathrm{Tr}(AB),\\ \mathrm{Tr}(A\wedge B\wedge C) &= \mathrm{Tr}\,A\,\mathrm{Tr}\,B\,\mathrm{Tr}\,C - \mathrm{Tr}\,A\,\mathrm{Tr}(BC)\\ &\quad - \mathrm{Tr}\,B\,\mathrm{Tr}(AC) - \mathrm{Tr}\,C\,\mathrm{Tr}\,(AB)\\ &\quad + \mathrm{Tr}(ABC) + \mathrm{Tr}(ACB).\end{aligned} \tag{13}$$

In particular note that

$$\mathrm{Tr}(A_1\wedge A_2\wedge\cdots\wedge A_k) = \mathrm{Tr}(A_{i_1}\wedge A_{i_2}\wedge\cdots\wedge A_{i_k}), \tag{14}$$

where $(i_1,\ldots,i_k)$ is any permutation of $(1,2,\ldots,k)$. It is easy to show

$$\begin{aligned}&\mathrm{Tr}(A_1\wedge\cdots\wedge A_{k-l}\wedge\underbrace{\delta\wedge\cdots\wedge\delta}_{l})\\ &\quad = [(n-k+l)!/(n-k)!]\,\mathrm{Tr}(A_1\wedge\cdots\wedge A_{k-l}).\end{aligned} \tag{15}$$

One more property we need is

$$\begin{aligned}&(A_1\wedge A_2\wedge\cdots\wedge A_k)(T\wedge\underbrace{\delta\wedge\delta\wedge\cdots\wedge\delta}_{k-1})\\ &\quad = (k-1)!\{(A_1T)\wedge A_2\wedge\cdots\wedge A_k\\ &\qquad + A_1\wedge(A_2T)\wedge\cdots\wedge A_k\\ &\qquad + \cdots + A_1\wedge A_2\wedge\cdots\wedge(A_kT)\}.\end{aligned} \tag{16}$$

Now we wish to evaluate the quantity

$$\begin{aligned}&\sum\mathrm{Tr}\,(\tilde T^{\alpha_1}\tilde T^{\alpha_2}\cdots\tilde T^{\alpha_m})\\ &\quad = \frac{1}{((k-1)!)^m}\sum\mathrm{Tr}\big[(T^{\alpha_1}\wedge\delta\wedge\cdots\wedge\delta)\\ &\qquad\times(T^{\alpha_2}\wedge\delta\wedge\cdots\wedge\delta)\cdots(T^{\alpha_m}\wedge\delta\wedge\cdots\wedge\delta)\big].\end{aligned} \tag{17}$$

Using Eq. (16) repeatedly and taking into account the symmetrization on the indices α_i we obtain

$$\begin{aligned}&\sum\mathrm{Tr}\,(\tilde T^{\alpha_1}\tilde T^{\alpha_2}\cdots\tilde T^{\alpha_m})\\ &\quad = \frac{1}{((k-1)!)^m}((k-1)!)^{m-1}\frac{1}{k}\sum_{\{n_i\}}\frac{m!}{n_1!\cdots n_k!}\\ &\qquad\times\sum\mathrm{Tr}\,(T^{n_1}\wedge T^{n_2}\wedge\cdots\wedge T^{n_k}),\end{aligned} \tag{18}$$

where $\Sigma_{\{n_i\}}$ is the sum over all partitions with n_i satisfying $\Sigma_{i=1}^k n_i = m$, and T^{n_i} is a short-hand notation for $T^{\alpha_1}T^{\alpha_2}\cdots T^{\alpha_n}$ (ordinary matrix multiplication). Using Eq. (14), we can rewrite this as a sum over partitions $\{z_i\}$, where z_i is the number of n's which are equal to i. They satisfy

$$\sum_{i=0}^m z_i = k,\qquad \sum_{i=0}^m i z_i = m. \tag{19}$$

Thus we have

$$\begin{aligned}&\sum_{\{n_i\}}\frac{m!}{n_1!\cdots n_k!}\sum\mathrm{Tr}[T^{n_1}\wedge T^{n_2}\wedge\cdots\wedge T^{n_k}]\\ &\quad = \sum_{\{z_i\}}\frac{m!k!}{z_0!z_1!\cdots z_m!(1!)^{z_1}\cdots(m!)^{z_m}}\\ &\qquad\times\sum\mathrm{Tr}\Big\{\underbrace{\delta\wedge\cdots\wedge\delta}_{z_0}\wedge\underbrace{T\wedge\cdots\wedge T}_{z_1}\\ &\qquad\wedge\underbrace{T^2\wedge\cdots\wedge T^2}_{z_2}\wedge\cdots\Big\}.\end{aligned} \tag{20}$$

Inserting this into Eq. (18) and using Eq. (15) we obtain

$$\begin{aligned}&\sum\mathrm{Tr}(\tilde T^{\alpha_1}\cdots\tilde T^{\alpha_m})\\ &\quad = m!\sum_{z_0=0}^{k-1}\binom{n-k+z_0}{z_0}\\ &\qquad\times\sum_{\{z_i\}}\frac{\sum\mathrm{Tr}\Big\{\overbrace{T\wedge\cdots\wedge T}^{z_1}\wedge\overbrace{T^2\wedge\cdots\wedge T^2}^{z_2}\overbrace{T^3\cdots\wedge T^3}^{z_3}\cdots\Big\}}{z_1!z_2!\cdots z_m!(1!)^{z_1}\cdots(m!)^{z_m}}\end{aligned} \tag{21}$$

the right-hand side contains products of traces. What we do in the following is to obtain the coefficient for each type of trace product. With that done the reduction is completed.

First we deal with the one trace case. Its coefficient is denoted as $A^m_m(n,k)$. For a particular partition $\{z_i\}$ the relevant contribution from $\Sigma\,\mathrm{Tr}\{\cdots\}$ in Eq. (21) can be pulled out by the use of Eq. (10) and it is

$$(-1)^{k-z_0+1}(k-z_0-1)!\sum\mathrm{Tr}(T^{\alpha_1}T^{\alpha_2}\cdots T^{\alpha_m}). \tag{22}$$

Thus we find,

$$\begin{aligned}A^m_m(n,k) &= m!\sum_{z_0=0}^{k-1}(-1)^{k-z_0+1}\binom{n-k+z_0}{z_0}(k-z_0-1)!\\ &\quad\times\sum_{\{z_i\}}\frac{1}{z_1!\cdots z_m!(1!)^{z_i}\cdots(m!)^{z_m}}.\end{aligned} \tag{23}$$

Then employing an identity,

$$\begin{aligned}&\sum_{\{z_i\}}\frac{1}{z_i!\cdots z_m!(1!)^{z_i}\cdots(m!)^{z_m}}\\ &\quad = \frac{1}{m!}\sum_{l=0}^{k-z_0}\frac{(-1)^l(k-z_0-l)^m}{l!(k-z_0-l)!},\end{aligned} \tag{24}$$

and defining $j = z_0 + l$ we arrive at

$$\begin{aligned}A^m_m(n,k) &= \sum_{z_0=0}^{k-1}\sum_{j=z_0}^{k}(-1)^{k-j+1}\binom{n-k+z_0}{z_0}\\ &\quad\times(k-z_0-1)!\frac{(k-j)^m}{(j-z_0)!(k-j)!}\\ &= \sum_{j=0}^{k-1}\sum_{z_0=0}^{j}(-1)^{k-j+1}\\ &\quad\times\binom{n-k+z_0}{z_0}\binom{k-z_0-1}{j-z_0}(k-j)^{m-1}\\ &= \sum_{j=0}^{k}(-1)^{k-j+1}\binom{n}{j}(k-j)^{m-1}.\end{aligned} \tag{25}$$

(This has been obtained in Ref. 5. Our result agrees with theirs.) Restricting ourselves to even $m = 2p$ we readily evaluate several coefficients in the expansion

$$\sum \mathrm{Tr}\,(\tilde{T}^{\alpha_1}\cdots\tilde{T}^{\alpha_{2p}}) = \sum_{\{n_i\}} A_{2p}^{n_1 n_2\cdots} \sum [(\mathrm{Tr}\,T^{n_1})(\mathrm{Tr}\,T^{n_2})\cdots], \tag{26}$$

where

$$\sum_i n_i = 2p, \quad n_1 \leqslant n_2 \leqslant \cdots, \quad n_i \geqslant 2.$$

Using Eq. (23) we obtain

$$A_{2p}^{22\cdots 2} = \frac{(2p)!}{2^p p!}\binom{n-2p}{k-p}, \tag{27}$$

$$A_{2p}^{22\cdots 24} = \frac{(2p)!}{2^{p-2}\cdot 4\cdot (p-2)!} \times\left[-\binom{n-2p}{k-p} + \frac{1}{6}\binom{n-2p+2}{k-p+1}\right], \tag{28}$$

$$A_{2p}^{22\cdots 233} = \frac{(2p)!}{2^{p-3}\cdot 3^2\cdot 2!(p-3)!} \times\left[-\binom{n-2p}{k-p} + \frac{1}{4}\binom{n-2p+2}{k-p+1}\right], \tag{29}$$

where $A_{2p}^{22\cdots 2}$, $A_{2p}^{22\cdots 24}$, and $A_{2p}^{22\cdots 233}$ are, respectively, the coefficients of

$$(\mathrm{Tr}\,T^2)^p, \quad (\mathrm{Tr}\,T^2)^{p-2}\cdot(\mathrm{Tr}\,T^4), \quad (\mathrm{Tr}\,T^2)^{p-3}\cdot(\mathrm{Tr}\,T^3)^2.$$

With a little more effort we can express Eq. (23) in a form similar to the above:

$$A_{2p}^{2p} = (2p-1)!\left[\alpha_1^{(2p)}\binom{n-2}{k-1} - \alpha_2^{(2p)}\binom{n-4}{k-2} + \cdots + (-1)^{p+1}\alpha_p^{(2p)}\binom{n-2p}{k-p}\right]. \tag{30}$$

The constants $\alpha_i^{(2p)}$ are tabulated in Table I for $1 \leqslant p \leqslant 7$. They can be written as

$$\alpha_k^{(2p)} = \frac{1}{(2p-1)!}\sum_{j=1}^{k} (j)^{2p-1}\beta_j^{(k)} \quad (k = 1,2,\ldots,p), \tag{31}$$

where the β's are the solutions of

$$\sum_{j=1}^{k} (j)^{2i-1}\beta_j^{(k)} = \delta_{ik}(2k-1)! \quad (i = 1,2,\ldots,k). \tag{32}$$

In particular,

$$\alpha_1^{(2p)} = \frac{1}{(2p-1)!}, \quad \alpha_p^{(2p)} = 1. \tag{33}$$

To obtain the other coefficients requires more effort. We note first that an arbitrary coefficient can be written in the following form:

$$A_{2p}^{\overbrace{22\cdots 2}^{n_2}\overbrace{33\cdots 3}^{n_3}\cdots} = \frac{(2p)!}{(2^{n_2}3^{n_3}\cdots)n_2!n_3!\cdots}\times(-1)^{p+\sum_{i=2}^{2p} n_i} \times\left\{v_1\binom{n-2p}{k-p} - v_2\binom{n-2p+2}{k-p+1} + \cdots + (-1)^{p+1}v_p\binom{n-2}{k-1}\right\}. \tag{34}$$

TABLE I. The constants of α_i^{2p}.

$2p$	α_1	α_2	α_3	α_4	α_5	α_6	α_7
2	$\frac{1}{1}$						
4	$\frac{1}{3!}$	$\frac{1}{1}$					
6	$\frac{1}{5!}$	$\frac{1}{4}$	$\frac{1}{1}$				
8	$\frac{1}{7!}$	$\frac{1}{40}$	$\frac{1}{3}$	$\frac{1}{1}$			
10	$\frac{1}{9!}$	$\frac{17}{(84)(144)}$	$\frac{7}{144}$	$\frac{5}{12}$	$\frac{1}{1}$		
12	$\frac{1}{11!}$	$\frac{31}{(144)(4200)}$	$\frac{4}{945}$	$\frac{19}{240}$	$\frac{1}{2}$	$\frac{1}{1}$	
14	$\frac{1}{13!}$	$\frac{1}{(144)(5280)}$	$\frac{13}{(144)(360)}$	$\frac{1}{108}$	$\frac{7}{60}$	$\frac{7}{12}$	$\frac{1}{1}$

$A_{2p}^{2p} = (2p-1)!\left(\alpha_1\binom{n-2}{k-1} - \alpha_2\binom{n-4}{k-2} + \alpha_3\binom{n-6}{k-3} - \cdots\right)$.

Therefore a coefficient can be represented by a vector $(v_1, v_2,\cdots,v_p, 0\cdots 0)$. For instance we have

$$\begin{aligned} A_{2p}^{22\cdots 2} &= (1, 0,\ldots,0),\\ A_{2p}^{22\cdots 233} &= (1, \tfrac{1}{4}, 0,\ldots,0),\\ A_{2p}^{22\cdots 24} &= (1, \tfrac{1}{6}, 0,\ldots,0),\\ A_{2p}^{2p} &= (\alpha_p^{(2p)}, \alpha_{p-1}^{(2p)}, \cdots, \alpha_{1,0,\cdots,0}^{(2p)}). \end{aligned} \tag{35}$$

Then by defining a vector multiplication,

$$(v_1, v_2, \ldots)(w_1, w_2), \ldots) = v_1(w_1, w_2,\ldots) + v_2(0, w_1, w_2,\ldots) + \cdots, \tag{36}$$

we find the following rules of combination and addition of coefficients.

Rule 1. If $A_{2p}^{\cdots}$ is an arbitrary coefficient in $\Sigma\,\mathrm{Tr}(\tilde{T}^{\alpha_1}\cdots\tilde{T}^{\alpha_{2p}})$ then

$$A_{2p+6}^{\cdots(3)(3)} = A_6^{33}A_{2p}^{\cdots}. \tag{37}$$

Rule 2.

$$A_{2p}^{\cdots(2n)} = A_{2n}^{2n}\,A_{2p}^{\overbrace{\cdots 22\cdots 2}^{n}}. \tag{38}$$

Rule 3.

$$A_{2p}^{\cdots(2n+1)} = \left(a_n^{(2n)}, \frac{n-1}{n}a_{n-1}^{(2n)}, \frac{n-2}{n}a_{n-2}^{(2n)},\ldots,\frac{1}{n}a_1^{(2n)}, 0,\ldots\right) \times A_{2p}^{\overbrace{\cdots 22\cdots 23}^{n-1}}. \tag{39}$$

Using these rules and Eqs. (27)–(29) one can obtain all coefficients in the reduction.

To make our procedure more understandable let us work out the $m = 2p = 8$ case:

$$\sum \mathrm{Tr}(\tilde{T}^{\alpha_1}\cdots\tilde{T}^{\alpha_8}) = A_8^{\,8}(N,k)\sum \mathrm{Tr}(T^{\alpha_1}\cdots T^{\alpha_8})$$
$$+ A_8^{\,26}(N,k)\sum [\mathrm{Tr}(T^{\alpha_1}T^{\alpha_2})\mathrm{Tr}(T^{\alpha_3}\cdots T^{\alpha_8}]$$
$$+ A_8^{\,35}(N,k)\sum [\mathrm{Tr}(T^{\alpha_1}T^{\alpha_2}T^{\alpha_3})\mathrm{Tr}(T^{\alpha_4}\cdots T^{\alpha_8})] + A_8^{\,44}(N,k)\sum [\mathrm{Tr}(T^{\alpha_1}\cdots T^{\alpha_4})\mathrm{Tr}(T^{\alpha_5}\cdots T^{\alpha_8})]$$
$$+ A_8^{\,233}(N,k)\sum [\mathrm{Tr}(T^{\alpha_1}T^{\alpha_2})\mathrm{Tr}(T^{\alpha_3}T^{\alpha_4}T^{\alpha_5})\mathrm{Tr}(T^{\alpha_6}T^{\alpha_7}T^{\alpha_8})]$$
$$+ A_8^{\,224}(N,k)\sum [\mathrm{Tr}(T^{\alpha_1}T^{\alpha_2})\mathrm{Tr}(T^{\alpha_3}T^{\alpha_4})\mathrm{Tr}T^{\alpha_5}\cdots T^{\alpha_8})]$$
$$+ A_8^{\,2222}(N,k)\sum [\mathrm{Tr}(T^{\alpha_1}T^{\alpha_2})\mathrm{Tr}(T^{\alpha_3}T^{\alpha_4})\mathrm{Tr}(T^{\alpha_5}T^{\alpha_6})\mathrm{Tr}(T^{\alpha_7}T^{\alpha_8})]. \tag{40}$$

From Eqs. (27)–(29) we have

$$A_8^{\,2222}(N,k) = \frac{8!}{2^4 4!}\binom{N-8}{k-4},$$
$$A_8^{\,233}(N,k) = \frac{8!}{2\cdot 3^2\cdot 2!}\left[-\binom{N-8}{k-4} + \frac{1}{4}\binom{N-6}{k-3}\right], \tag{41}$$
$$A_8^{\,224}(N,k) = \frac{8!}{4\cdot 2^2\cdot 2!}\left[-\binom{N-8}{k-4} + \frac{1}{6}\binom{N-6}{k-3}\right].$$

Also by solving Eqs. (31) and (32) we find

$$A_8^{\,8}(N,k) = 7!\left[-\binom{N-8}{k-4} + \frac{1}{3}\binom{N-6}{k-3} - \frac{1}{40}\binom{N-4}{k-2} + \frac{1}{7!}\binom{N-2}{k-1}\right]. \tag{42}$$

$$A_8^{\,2222} = (1, 0, 0, \ldots),$$
$$A_8^{\,233} = (1, \tfrac{1}{4}, 0, \ldots), \tag{43}$$
$$A_8^{\,224} = (1, \tfrac{1}{6}, 0, \ldots),$$
$$A_8^{\,8} = \left(1, \frac{1}{3}, \frac{1}{40}, \frac{1}{7!}, 0, \ldots\right).$$

We will also need

$$A_4^{\,4} = (1, \tfrac{1}{6}, 0, \ldots),$$
$$A_6^{\,6} = \left(1, \frac{1}{4}, \frac{1}{5!}, 0, \ldots\right). \tag{44}$$

Then $A_8^{\,26}$ can be obtained by combining three 2's in $A_8^{\,2222}$. By Rule 2 we find

$$A_8^{\,26} = A_6^{\,6}A_8^{\,2222} = (1, \tfrac{1}{4}, \tfrac{1}{5}, 0, \ldots)1, 0, \ldots) = (1, \tfrac{1}{4}, \tfrac{1}{5}, 0, \ldots). \tag{45}$$

Here, A^{35} can be obtained by combining one 2 and one 3 in A^{233}. By applying Rule 3 the generating vector is obtained from $A_4^{\,4}$ by

$$A_4^{\,4} = (1, \tfrac{1}{6}) \to (1, \tfrac{1}{2}\cdot\tfrac{1}{6}) = (1, \tfrac{1}{12}),$$
$$\therefore A_8^{\,35} = (1, \tfrac{1}{12}, 0, \ldots),$$
$$A_8^{\,233} = (1, \tfrac{1}{12}, 0, \ldots)(1, \tfrac{1}{4}, 0, \ldots) = (1, \tfrac{1}{3}, \tfrac{1}{48}, 0, \ldots). \tag{46}$$

Finally $A_8^{\,44}$ can be obtained by combining two 2's in $A_8^{\,224}$,

$$A_8^{\,44} = A_4^{\,4}A_8^{\,224} = (1, \tfrac{1}{6}, 0, \ldots)(1, \tfrac{1}{6}, 0, \ldots) = (1, \tfrac{1}{3}, \tfrac{1}{36}, 0, \ldots). \tag{47}$$

Writing these in full, according to Eqs. (27)–(29),

$$A_8^{\,26} = \frac{8!}{2\cdot 6}\left[\binom{n-8}{k-4} - \frac{1}{4}\binom{n-6}{k-3} + \frac{1}{5!}\binom{n-4}{k-2}\right],$$
$$A_8^{\,35} = \frac{8!}{3\cdot 5}\left[\binom{n-8}{k-4} - \frac{1}{3}\binom{n-6}{k-3} + \frac{1}{48}\binom{n-4}{k-2}\right], \tag{48}$$
$$A_8^{\,44} = \frac{8!}{4^2\cdot 2!}\left[\binom{n-8}{k-4} - \frac{1}{3}\binom{n-6}{k-3} + \frac{1}{36}\binom{n-4}{k-2}\right].$$

The tabulation of coefficients for $m = 2, 4, 6, 8, 10$, and 12 is presented in the Appendix.

So far we have restricted ourselves to even m. For the odd m case one has

$$A_{2p+1}^{\overbrace{22\cdots2}^{n_2}\,\overbrace{33\cdots3}^{n_3}\cdots}(N,k) = \frac{(2p+1)!}{(2^{n_2}3^{n_3}\cdots)n_2!n_3!\cdots}(-1)^{1+\Sigma_i n_i} \times \{v_1[N,k,p] + v_2[N,k,p-1] + \cdots + v_p[N,k,1]\}. \tag{49}$$

Here,

$$[N,k,p] = \frac{1}{2}(-1)^{p+1}\frac{N-2k}{N-2p}\binom{N-2p}{k-p}. \tag{50}$$

Then the coefficients are again represented by vectors and Rules 1, 2, and 3 apply without modification. The starting vector (1, 0, 0,..., 0) corresponds to

$$A_{2p+1}^{\overbrace{22\cdots23}^{p\quad 1}}(N,k) = \frac{(-1)^{p+1}(2p+1)!}{3\cdot 2^{p-1}(p-1)!}[N,k,p]. \tag{51}$$

At this point we comment on how to obtain the formulas for the symmetrized representations $\{k\}$. In this case we have to change the antisymmetrization in Eq. (6) into symmetrization. What this change amounts to is the following. Denote the coefficients for $\{1^k\}$ and $\{k\}$ as $A_{2p}^{(AS)}(N,k)$ and $A_{2p}^{(S)}(N,k)$. We know $A_{2p}^{(AS)}(N,k)$ can be expressed as follows:

$$A_{2p}^{(AS)}(N,k) = \sum_{j=0}^{k}(-1)^{k+j+1}\binom{N}{j}(k-j)^{2p-1}. \tag{52}$$

Then make a formal change $N \to -N$ to obtain

$$\binom{N}{j}\to\binom{-N}{j}=(-1)^j\binom{N+j-1}{N-1}. \tag{53}$$

The coefficient $A^{(S)}_{2p}(N,k)$ for $\{k\}$ has been found to be related to $A^{(AS)}_{2p}(-N,k)$;

$$A^{(AS)}_{2p}(-N,k)=(-1)^{k+1}A^{(S)}_{2p}(N,k). \tag{54}$$

With this we come to the end of the derivation of the reduction formulas.

We can apply our results to calculate the quantities

$$1_{2p}(\{1^k\})\equiv g^{(i_1i_2}g^{i_3i_4}\cdots g^{i_{2p-1}i_{2p})}\,\mathrm{Tr}(\mathscr{H}_{i_1}\cdots\mathscr{H}_{i_{2p}}), \tag{55}$$

where $\mathscr{H}_i$ are the elements of the Cartan subalgebra in the $\{1^k\}$ representation. We denote the elements in the fundamental representation by H_i, and $(\ )$ denotes symmetrization over indices.

By noting $\mathrm{Tr}(H_it_j)=(1/r)\delta_{ij}l_2(\{1\})$ it is straightforward to obtain the following:

$$l_2(\{1^k\})=A^{\,2}_{2}(N,k)l_2(\{1\}), \tag{56}$$

$$\begin{aligned}l_4(\{1^k\})&=A^{\,4}_{4}(N,k)l_4(\{1\})\\&\quad+\frac{r+2}{3r}A^{\,22}_{4}(N,k)[l_2(\{1\})]^2,\end{aligned} \tag{57}$$

$$\begin{aligned}l_6(\{1^k\})&=A^{\,6}_{6}(N,k)l_6(\{1\})\\&\quad+\frac{r+4}{5r}A^{\,24}_{6}(N,k)l_2(\{1\})l_4(\{1\})\\&\quad+A^{\,33}_{6}g_{(ij}g_{kl}g_{mn)}\,\mathrm{Tr}(H_iH_jH_k)\,\mathrm{Tr}(H_lH_mH_n)\\&\quad+\frac{r+4}{5r}\frac{r+2}{3r}A^{\,222}_{6}(N,k)[l_2(\{1\})]^3,\\&\quad\vdots\\l_{2p}(\{1^k\})&=A^{\,2p}_{2p}(N,k)l_{2p}(\{1\})\\&\quad+A^{\,2,2p-2}_{2p}(N,k)\frac{r+2p-2}{(2p-1)r}l_2(\{1\})l_{2p-2}(\{1\})\\&\quad+\cdots+\frac{r+2}{3r}\frac{r+4}{5r}\cdots\frac{r+2p-2}{(2p-1)r}\\&\quad\times A^{\,22\cdots2}_{2p}(N,k)[l_2(\{1\})]^p.\end{aligned} \tag{58}$$

Calculating the right-hand side is not difficult since it involves only the fundamental representation. A tabulation of the result will be given elsewhere.

So far, all the coefficients for even m have been expressed in terms of the functions

$$\binom{n-2q}{k-q}.$$

What happens when this quantity is not defined? (For example, $n-2q<k-q$.) Now, these terms arise from the expressions

$$\sum_{j=0}^{q}(-1)^j\binom{q}{j}\binom{n-2q+j}{k-2q+j},$$

each term in the summation being a contribution from a particular partition in Eq. (21). Here we assumed that $n>k\geqslant 2q$. But this need not be the case. If $k<q$, clearly none of the partitions contribute, and we simply omit the entire expression

$$\left(k<q\Rightarrow\text{“}\binom{n-2q}{k-q}\text{”}=0\right).$$

But, if $k=q+l$, $0<l<q$, the appropriate sum is

$$\begin{aligned}&\sum_{j=q-l}^{q}(-1)^j\binom{q}{j}\binom{n-2q+j}{k-2q+j}\\&\quad=\begin{cases}(-1)^q\binom{n-2q}{k-q}, & n\geqslant k+q,\\(-1)^k\binom{k+q-n-1}{k-q}, & n<2q\\0, & 2q\leqslant n<k+q.\end{cases}\end{aligned}$$

Similar considerations apply to the case of odd m.

III. CONTACT WITH PHYSICS

The index l_{2p} mentioned at the end of the last section has not yet found its use in particle physics. On the other hand a more common index has found another use in particle physics. Recently Frampton noted,[9,10] by generalizing the works of Ref. 11, that in a $d(>4)$-dimensional theory[12] the anomaly is well defined and can be separated from other divergences inherent in such a theory. To state only the result the anomaly in $2d$ dimensions is given by $A^{\,d+1}_{d+1}$, with a normalization such that the anomaly for the fundamental representation is one. If we take such a theory seriously we must take care of the anomaly, among other things. And thus the freedom from anomaly should serve as a constraint on the models in $2d$ dimensions. (Incidentally there is no anomaly in an odd dimension.)

As regards the anomaly we only note one formula which is handy to get numbers:

$$A^{\,m}_{m}[N,k]=A^{\,m}_{m}[N-1,k]+A^{\,m}_{m}[N-1,k-1]. \tag{59}$$

With the help of this formula one has to use Eq. (25) only twice to get all other numbers. Classification of the models based on freedom from anomaly will be given in a separate paper.

IV. SUMMARY

We derived a reduction formula for $\{1^k\}$ and $\{k\}$ of SU(N) which we think will serve as a basis for more general cases. Generalization of our result is currently under investigation.

ACKNOWLEDGMENTS

One of us (Y.F.) would like to thank Professor L. O'Raifeartaigh, Dr. Peccei, Dr. Y. Achiman, Dr. E. Paschos, Dr. C. H. Mo and Dr. R. J. N. Phillips, Dr. M. G. Schmidt, Professor G. Costa, Professor N. Cabbibo, Dr. G. Senjanovic, and Professor A. N. Kamal for hospitality at, respectively, D.I.A.S. (Dublin), Max Planck Institute (München), Gesamdt Hochschule Wuppteral, Universität Dortmund, Rutherford Laboratory, Universität Heidelberg, Istituto di Fisica (Padova), Istituto di Fisica (Roma), Brookhaven National Laboratory, and the Department of Physics, University of Alberta. He also thanks Professor Abdus Salam, I.A.E.A., and UNESCO for hospitality at I.C.T.P. (Trieste) and Professor J. Stern at Orsay. Further we thank Professor S. Okubo for correspondence and valuable comments.

APPENDIX. THE COEFFICIENTS IN THE REDUCTION FORMULA FOR $p = 1,2,3,4,5$

$$A^{2}_{2} = \binom{n-2}{k-1},$$

$$A^{4}_{4} = 3!\left(-\binom{n-4}{k-2} + \frac{1}{6}\binom{n-2}{k-1}\right),$$

$$A^{22}_{4} = \frac{4!}{2^2\cdot 2!}\binom{n-4}{k-2},$$

$$A^{6}_{6} = 5!\left(\binom{n-6}{k-3} - \frac{1}{4}\binom{n-4}{k-2} + \frac{1}{120}\binom{n-2}{k-1}\right),$$

$$A^{33}_{6} = \frac{6!}{3^2\cdot 2!}\left(-\binom{n-6}{k-3} + \frac{1}{4}\binom{n-4}{k-2}\right),$$

$$A^{24}_{6} = \frac{6!}{2\cdot 4}\left(-\binom{n-6}{k-3} + \frac{1}{6}\binom{n-4}{k-2}\right),$$

$$A^{222}_{8} = \frac{6!}{2^3\cdot 3!}\binom{n-6}{k-3},$$

$$A^{8}_{8} = 7!\left(-\binom{n-8}{k-4} + \frac{1}{3}\binom{n-6}{k-3} - \frac{1}{40}\binom{n-4}{k-2} + \frac{1}{7!}\binom{n-2}{k-1}\right),$$

$$A^{44}_{8} = \frac{8!}{4^2\cdot 2!}\left(\binom{n-8}{k-4} - \frac{1}{3}\binom{n-6}{k-3} + \frac{1}{36}\binom{n-4}{k-2}\right),$$

$$A^{35}_{8} = \frac{8!}{3\cdot 5}\left(\binom{n-8}{k-4} - \frac{1}{3}\binom{n-6}{k-3} + \frac{1}{48}\binom{n-4}{k-2}\right),$$

$$A^{26}_{8} = \frac{8!}{2\cdot 6}\left(\binom{n-8}{k-4} - \frac{1}{4}\binom{n-6}{k-3} + \frac{1}{120}\binom{n-4}{k-2}\right),$$

$$A^{224}_{8} = \frac{8!}{2^2\cdot 4\cdot 2!}\left(-\binom{n-8}{k-4} + \frac{1}{6}\binom{n-6}{k-3}\right),$$

$$A^{233}_{8} = \frac{8!}{2\cdot 3^2\cdot 2!}\left(-\binom{n-8}{k-4} + \frac{1}{4}\binom{n-6}{k-3}\right),$$

$$A^{2222}_{8} = \frac{8!}{2^4\cdot 4!}\binom{n-8}{k-4},$$

$$A^{10}_{10} = 9!\left(\binom{n-10}{k-5} - \frac{5}{12}\binom{n-8}{k-4} + \frac{7}{144}\binom{n-6}{k-3} - \frac{17}{(84)(144)}\binom{n-4}{k-2} + \frac{1}{9!}\binom{n-2}{k-1}\right),$$

$$A^{55}_{10} = \frac{10!}{5^2\cdot 2!}\left(-\binom{n-10}{k-5} + \frac{5}{12}\binom{n-8}{k-4} - \frac{7}{144}\binom{n-6}{k-3} + \frac{1}{576}\binom{n-4}{k-2}\right),$$

$$A^{46}_{10} = \frac{10!}{4\cdot 6}\left(-\binom{n-10}{k-5} + \frac{5}{12}\binom{n-8}{k-4} - \frac{1}{20}\binom{n-6}{k-3} + \frac{1}{720}\binom{n-4}{k-2}\right),$$

$$A^{37}_{10} = \frac{10!}{3\cdot 7}\left(-\binom{n-10}{k-5} + \frac{5}{12}\binom{n-8}{k-4} - \frac{2}{45}\binom{n-6}{k-3} + \frac{1}{1440}\binom{n-4}{k-2}\right),$$

$$A^{28}_{10} = \frac{10!}{2\cdot 8}\left(-\binom{n-10}{k-5} + \frac{1}{3}\binom{n-8}{k-4} - \frac{1}{40}\binom{n-6}{k-3} + \frac{1}{7!}\binom{n-4}{k-2}\right),$$

$$A^{334}_{10} = \frac{10!}{3^2\cdot 4\cdot 2!}\left(\binom{n-10}{k-5} - \frac{5}{12}\binom{n-8}{k-4} + \frac{1}{24}\binom{n-6}{k-3}\right),$$

$$A^{244}_{10} = \frac{10!}{2\cdot 4^2\cdot 2!}\left(\binom{n-10}{k-5} - \frac{1}{3}\binom{n-8}{k-4} + \frac{1}{36}\binom{n-6}{k-3}\right),$$

$$A^{235}_{10} = \frac{10!}{2\cdot 3\cdot 5}\left(\binom{n-10}{k-5} - \frac{1}{3}\binom{n-8}{k-4} + \frac{1}{48}\binom{n-6}{k-3}\right),$$

$$A^{226}_{10} = \frac{10!}{2^2\cdot 6\cdot 2!}\left(\binom{n-10}{k-5} - \frac{1}{4}\binom{n-8}{k-4} + \frac{1}{120}\binom{n-6}{k-3}\right),$$

$$A^{2233}_{10} = \frac{10!}{2^2\cdot 3^2\cdot (2!)^2}\left(-\binom{n-10}{k-5} + \frac{1}{4}\binom{n-8}{k-4}\right),$$

$$A^{2224}_{10} = \frac{10!}{2^3\cdot 4\cdot 3!}\left(-\binom{n-10}{k-5} + \frac{1}{6}\binom{n-8}{k-4}\right),$$

$$A^{22222}_{10} = \frac{10!}{2^5\cdot 5!}\binom{n-10}{k-5},$$

$$A^{12}_{12} = 11!\left(-\binom{n-12}{k-6} + \frac{1}{2}\binom{n-10}{k-5} - \frac{19}{240}\binom{n-8}{k-4} + \frac{4}{945}\binom{n-6}{k-3} - \frac{31}{15\cdot 8!}\binom{n-4}{k-2} + \frac{1}{11!}\binom{n-2}{k-1}\right),$$

$$A^{66}_{12} = \frac{12!}{6^2\cdot 2!}\left(\binom{n-12}{k-6} - \frac{1}{2}\binom{n-10}{k-5} + \frac{17}{240}\binom{n-8}{k-4} - \frac{1}{240}\binom{n-6}{k-3} + \frac{1}{20\cdot 6!}\binom{n-4}{k-2}\right),$$

$$A^{57}_{12} = \frac{12!}{5\cdot 7}\left(\binom{n-12}{k-6} - \frac{1}{2}\binom{n-10}{k-5} + \frac{19}{240}\binom{n-8}{k-4} - \frac{19}{6\cdot 6!}\binom{n-6}{k-3} + \frac{1}{24\cdot 6!}\binom{n-4}{k-2}\right),$$

$$A^{48}_{12} = \frac{12!}{4\cdot 8}\left(\binom{n-12}{k-6} - \frac{1}{2}\binom{n-10}{k-5} + \frac{29}{360}\binom{n-8}{k-4} - \frac{22}{7!}\binom{n-6}{k-3} + \frac{1}{6\cdot 7!}\binom{n-4}{k-2}\right),$$

$$A^{39}_{12} = \frac{12!}{3\cdot 9}\left(\binom{n-12}{k-6} - \frac{1}{2}\binom{n-10}{k-5} + \frac{3}{40}\binom{n-8}{k-4} - \frac{1}{315}\binom{n-6}{k-3} + \frac{1}{2\cdot 8!}\binom{n-4}{k-2}\right),$$

$$A^{2,10}_{12} = \frac{12!}{2\cdot 10}\left(\binom{n-12}{k-6} - \frac{5}{12}\binom{n-10}{k-5} + \frac{7}{144}\binom{n-8}{k-4} - \frac{17}{(84)(144)}\binom{n-6}{k-3} + \frac{1}{9!}\binom{n-4}{k-2}\right),$$

$$A^{444}_{12} = \frac{12!}{4^2\cdot 3!}\left(-\binom{n-12}{k-6} + \frac{1}{2}\binom{n-10}{k-5} - \frac{1}{12}\binom{n-8}{k-4} + \frac{1}{216}\binom{n-6}{k-3}\right),$$

$$A_{12}^{345}=\frac{12!}{3\cdot4\cdot5}\left(-\binom{n-12}{k-6}+\frac{1}{2}\binom{n-10}{k-5}-\frac{11}{144}\binom{n-8}{k-4}+\frac{1}{288}\binom{n-6}{k-3}\right),$$

$$A_{12}^{336}=\frac{12!}{3^2\cdot6\cdot2!}\left(-\binom{n-12}{k-6}+\frac{1}{2}\binom{n-10}{k-5}-\frac{17}{240}\binom{n-8}{k-4}+\frac{1}{480}\binom{n-6}{k-3}\right),$$

$$A_{12}^{255}=\frac{12!}{2\cdot5^2\cdot2!}\left(-\binom{n-12}{k-6}+\frac{5}{12}\binom{n-10}{k-5}-\frac{7}{144}\binom{n-8}{k-4}+\frac{1}{576}\binom{n-6}{k-3}\right),$$

$$A_{12}^{246}=\frac{12!}{2\cdot4\cdot6}\left(-\binom{n-12}{k-6}+\frac{5}{12}\binom{n-10}{k-5}-\frac{1}{20}\binom{n-8}{k-4}+\frac{1}{720}\binom{n-6}{k-3}\right),$$

$$A_{12}^{237}=\frac{12!}{2\cdot3\cdot7}\left(-\binom{n-12}{k-6}+\frac{5}{12}\binom{n-10}{k-5}-\frac{2}{45}\binom{n-8}{k-4}+\frac{1}{1440}\binom{n-6}{k-3}\right),$$

$$A_{12}^{228}=\frac{12!}{2^2\cdot8\cdot2!}\left(-\binom{n-12}{k-6}+\frac{1}{3}\binom{n-10}{k-5}-\frac{1}{40}\binom{n-8}{k-4}+\frac{1}{7!}\binom{n-6}{k-3}\right),$$

$$A_{12}^{3333}=\frac{12!}{3^4\cdot4!}\left(\binom{n-12}{k-6}-\frac{1}{2}\binom{n-10}{k-5}+\frac{1}{16}\binom{n-8}{k-4}\right),$$

$$A_{12}^{2244}=\frac{12!}{2^2\cdot4^2\cdot(2!)^2}\left(\binom{n-12}{k-6}-\frac{1}{3}\binom{n-10}{k-5}+\frac{1}{36}\binom{n-8}{k-4}\right),$$

$$A_{12}^{2235}=\frac{12!}{2^2\cdot3\cdot5\cdot2!}\left(\binom{n-12}{k-6}-\frac{1}{3}\binom{n-10}{k-5}+\frac{1}{48}\binom{n-8}{k-4}\right),$$

$$A_{12}^{2226}=\frac{12!}{2^3\cdot6\cdot3!}\left(\binom{n-12}{k-6}-\frac{1}{4}\binom{n-10}{k-5}+\frac{1}{120}\binom{n-8}{k-4}\right),$$

$$A_{12}^{22233}=\frac{12!}{2^3\cdot3^2\cdot3!\cdot2!}\left(-\binom{n-12}{k-6}+\frac{1}{4}\binom{n-10}{k-5}\right),$$

$$A_{12}^{22224}=\frac{12!}{2^4\cdot4\cdot4!}\left(-\binom{n-12}{k-6}+\frac{1}{6}\binom{n-10}{k-5}\right),$$

$$A_{12}^{222222}=\frac{12!}{2^6\cdot6!}\binom{n-12}{k-6}.$$

[1] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1972 (1976).
[2] S. Okubo, J. Math. Phys. **23**, 8 (1982).
[3] J. McKay, J. Patera, and R. T. Sharp, J. Math. Phys. **22**, 2770 (1981).
[4] S. Okubo and J. Patera, University of Rochester Report, UR-836, 1982.
[5] S. Okubo and J. Patera, University of Rochester Report, UR-845, 1983.
[6] T. P. Cheng, E. Eichten, and L. F. Li, Phys. Rev. D **9**, 2259 (1974). M. T. Vaughn, Z. Phys. C **2**, 111 (1979). P. Civitanovic, Phys. Rev. D **14**, 1536 (1976). D. R. T. Jones, Phys. Rev. D **25**, 581 (1982).
[7] W. G. McKay and J. Patera, *Tables of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Marcel Dekker, New York, 1981).
[8] J. Banks and H. Georgi, Phys. Rev. D **14**, 1159 (1976). S. Okubo, Phys. Rev. D **16**, 3528 (1977).
[9] P. H. Frampton, Phys. Lett. B **122**, 351 (1983).
[10] P. H. Frampton and T. W. Kephart, Phys. Rev. Lett. **50**, 1343 and 1347 (1983). In the former paper they obtained $A_4^4, A_4^{22}, A_6^6, A_6^{24}, A_6^{33}, A_6^{222}$. Our result agrees with theirs. It is said that they are in preparation of a paper (to appear as IFP-193-UNC) to deal with more general cases.
[11] W. A. Bardeen, Phys. Rev. **184**, 1848 (1969); J. Wess and B. Zumino, Phys. Lett. B **37**, 95 (1971).
[12] Th. Kaluza, Sitz. Preuss. Akad. Wiss. Berlin, Math. Phys. K **1**, 966 (1921); O. Klein, Z. Phys. **37**, 895 (1926); A. Salam and J. Strathdee, Ann. Phys. **141**, 316 (1982).

Space groups and their isotropy subgroups

Marko V. Jarić
Physics Department, Montana State University, Bozeman, Montana 59717 [a]

Marjorie Senechal
Department of Mathematics, Smith College, Northampton, Massachusetts 01063

(Received 14 February 1984, accepted for publication 1 June 1984)

The theory of space groups and their subgroups is related to that of their irreducible representations. Several important theorems concerning the determination of isotropy subgroups are proved. In particular, we show that the translation subgroups of isotropy subgroups are in one-to-one correspondence with certain subsets, called "substars," of the "star" characterizing the irreducible representation of the space group.

PACS numbers: 02.20. + b, 61.50.Em

I. INTRODUCTION

In this paper we show how to construct lattices of subgroups of the crystallographic space groups and explain how to obtain from them the lattice of isotropy subgroups in a given representation.

The problem of finding isotropy subgroups arises in many physical applications, notably in theories of continuous phase transitions[1] and in gauge theories.[2] In particular, in continuous structural phase transitions the symmetry of a crystal changes from one space group to a subgroup. This subgroup is the isotropy group of a vector (order parameter) in the representation space of an irreducible representation of the space group. This characterization is most explicit in the Landau theory of phase transitions[3] and in its renormalization-group generalization.[4] In the Landau theory, the isotropy groups which appear are those which correspond to the minima of the Landau free energy. The first step in identifying the minima is to determine *all* the isotropy groups of the representation.[5] An algorithm for finding the isotropy groups was recently given in Ref. 6. In this paper we present the general mathematical results on which the algorithm is based.

The direct computation of the subgroups of a space group is made possible by the algebriac characterization of an n-dimensional space group as an extension of Z^n (See Ref. 7). In Sec. II we show how these concepts can be used to construct subgroup lattices. We also describe the correspondence between this algebraic approach and the geometric point of view, in which a space group is regarded as a subgroup of $R^n \circledS O(n)$ (See Ref. 8).

Our main result is presented in Sec. III: a necessary and sufficient condition for a translation group to correspond to the translation subgroup of an isotropy subgroup is that it is reciprocal to the translation group generated by a substar of the star of the representation. This characterization greatly simplifies the computation involved in determining which of the subgroups in the subgroup lattice are in fact isotropy subgroups.

An explanation of the notation which is used in this paper is given in Table I. The Appendix contains a summary of relevant notions and results regarding a group action. Specifically, it contains several important theorems regarding centralizers and isotropy subgroups.

II. SPACE GROUPS AND THEIR SUBGROUP LATTICES

When space groups are regarded as the symmetry groups of real crystals, it is natural to think of them as groups of symmetries in three-dimensional space. This means that a space group is a subgroup of the Euclidean group, the group of all isometries of real space. This geometric viewpoint has been traditionally preferred by solid state physicists. Mathematicians and mathematical crystallographers, on the other hand, think of (n-dimensional) space groups as automorphism groups of the integral lattice. This algebraic point of view was developed by Bieberbach and others in the first half of this century to show, among other things, that the number of space group types in any finite dimension is finite,[9] and to classify them in three dimensions.[10] As we will see, the algebraic viewpoint is also very useful for characterizing and computing the subgroups of space groups.[7] We will also see that the geometric viewpoint provides complementary insights into the problem of isotropy subgroups.[11]

A. The geometric viewpoint

The Euclidean group $E(n)$ in any dimension n is the

TABLE I. The notation and symbols used in this paper.

R^n	the vector space of real n-tuples; the additive group of real n-tuples [elements: **t, f, k, r**]
Z^n	the additive group of integral n-tuples [elements: t]
$O(n)$	the group of $n \times n$ orthogonal matrices [elements: p]
$E(n)$	the group of all isometries of R^n [elements: g]
$A(n)$	the affine group of transformations of R^n
$GL(n,R)$	the group of $n \times n$ real nonsingular matrices [elements: γ]
$GL(n,Z)$	the group of $n \times n$ integral nonsingular matrices with integral inverse [elements: q, p]
$A \circledS B$	the semidirect-product-extension of A by B
A/B	the quotient group, space
$\|A\|$	number of elements in the set A
(**t**,**k**)	the orthogonal scalar product in R^n
p•**t**	the transform of **t** by p
$\oplus$	the direct sum of matrices, spaces
$\sim$	isomorphic; conjugate; reciprocal
$<$	smaller than; included; subgroup; subspace (*strict* inequality)
$\lhd$	invariant subgroup

[a] Present address, Department of Physics, North Dakota State University, Fargo, North Dakota 58105.

 0022-2488/84/113148-07$02.50

group of all isometric transformations of the real vector space R^n. It consists of linear orthogonal transformations which keep the origin fixed [the group O(n)], translations in n linearly independent directions (the additive *group* R^n) and combinations of these. Here R^n is an invariant subgroup of E(n), which is the semidirect product $R^n \circledS O(n)$. Therefore each element g of E(n) can be uniquely written as a product of an element $\mathbf{t}$ of R^n and an element p of O(n). Thus we can write $g = \mathbf{t}p$.

A space group G is the symmetry group of an n-dimensional real crystal. Therefore it is a subgroup of E(n) whose translation subgroup T_G is generated by n linearly independent vectors in R^n. G is an extension of T_G by a finite subgroup P_G of O(n) such that $P_G \sim G/T_G$ but it is not necessarily a semidirect product. (In three dimensions 73 of the 230 space group types are semidirect products; they are said to be symmorphic.) An element g of a space group G can be written as a product $\mathbf{t}\mathfrak{f}p$, where $\mathbf{t}$ is in T_G, $\mathfrak{f}$ (a "fractional translation") is in R^n and p is an element of P_G. [An alternative notation for g is $(\mathbf{t} + \mathfrak{f})p$ where additive notation is used for elements of R^n.] The point group component p and the translation component $\mathbf{t}\mathfrak{f}$ are uniquely determined by g. The term $\mathfrak{f}p$ can be regarded as a right coset representative of T_G in G; $\mathfrak{f}$ is uniquely defined modulo T_G. If G is symmorphic then $\mathfrak{f}$ can always be chosen to be $\mathbf{0}$.

The product $\mathbf{t}_3\mathfrak{f}_3p_3$ of two elements $\mathbf{t}_1\mathfrak{f}_1p_1$ and $\mathbf{t}_2\mathfrak{f}_2p_2$ is $\mathbf{t}_1\mathfrak{f}_1(p_1\cdot\mathbf{t}_2)(p_1\cdot\mathfrak{f}_2)p_1p_2$, where $p\cdot\mathbf{t}$ and $p\cdot\mathfrak{f}$ are the transforms of $\mathbf{t}$ and $\mathfrak{f}$ by p. Therefore,

$$\mathbf{t}_{1,2} \overset{\text{def}}{=} \mathbf{f}_1 + p_1\cdot\mathbf{f}_2 - \mathbf{f}_3 \in T_G. \tag{2.1}$$

Since G is an extension of T_G by P_G, a subgroup L of G must be an extension of $T_L \leqslant T_G$ by $P_L \leqslant P_G$ such that $P_L \sim L/T_L$.

In this geometric formulation, an orthonormal basis for R^n is usually assumed, and all of $\mathbf{t}$, $\mathfrak{f}$, and p are expressed in that basis.[8] However if we take the generators of T_G as a basis for R^n then T_G becomes an integral lattice. This is the starting point for the algebraic point of view.

B. The algebraic viewpoint

The above change of coordinates, which permits us to identify T_G with Z^n,[8] also effects a one-to-one mapping γ from P_G onto a subgroup $\mathfrak{P}_{\mathcal{G}}$ of GL(n,Z), the automorphism group of Z^n. We will denote $\gamma(\mathbf{t})$ by $\mathfrak{t}$, $\gamma(\mathfrak{f})$ by $\mathfrak{f}$, and $\gamma(p)$ by $\mathfrak{p}$ (see Ref. 12). Thus the image $\mathcal{G}$ of G under γ is an extension of Z^n by a finite subgroup $\mathfrak{P}_{\mathcal{G}}$ of GL(n,Z) such that $\mathfrak{P}_{\mathcal{G}} \sim \mathcal{G}/Z^n$.

Condition (2.1) now becomes

$$\mathfrak{t}_{1,2} \overset{\text{def}}{=} \mathfrak{f}_1 + \mathfrak{p}_1\cdot\mathfrak{f}_2 - \mathfrak{f}_3 \in Z^n, \tag{2.2}$$

which is known as the Frobenius congruence. The set of all fractional translations $\{\mathfrak{f}\}$ is called a vector system for $\mathcal{G}$.

Bieberbach showed that two space groups G and G' are isomorphic if and only if they are conjugate in the affine group $A(n) = R^n \circledS GL(n,R)$ (see Ref. 9). In fact, looking at the problem algebraically, we see that $\mathcal{G}$ and $\mathcal{G}'$ are conjugate in the subgroup $R^n \circledS GL(n,Z)$ of $A(n)$, since the conjugation must leave Z^n invariant. Let the space group $\mathcal{G}$ be characterized by $\mathfrak{P}_{\mathcal{G}}$ and the vector system $\{\mathfrak{f}\}$ and let the space group $\mathcal{G}'$ be characterized by $\mathfrak{P}_{\mathcal{G}'}$, and the vector system $\{\mathfrak{f}'\}$. Clearly if $\mathcal{G}$ and $\mathcal{G}'$ are conjugate in $R^n \circledS GL(n,Z)$ then $\mathfrak{P}_{\mathcal{G}'}$ and $\mathfrak{P}_{\mathcal{G}}$ must be conjugate in GL(n,Z); one says that $\mathcal{G}$ and $\mathcal{G}'$ (and also G and G') belong to the same arithmetic class. To determine the relation between $\{\mathfrak{f}\}$ and $\{\mathfrak{f}'\}$, let α be the element of $R^n \circledS GL(n,Z)$ which affects the conjugation $\mathcal{G}' = \alpha\mathcal{G}\alpha^{-1}$. Then $\alpha = \mathfrak{r}\mathfrak{q}$, where $\mathfrak{r} \in R^n$ and $\mathfrak{q} \in GL(n,Z)$. Thus $Z^n\mathfrak{f}'\mathfrak{p}' = \alpha(Z^n\mathfrak{f}\mathfrak{p})\alpha^{-1}$ implies

$$\mathfrak{p}' = \mathfrak{q}\mathfrak{p}\mathfrak{q}^{-1} \quad \text{and} \quad \mathfrak{f}' - [\mathfrak{r} + \mathfrak{q}\cdot\mathfrak{f} - (\mathfrak{q}\mathfrak{p}\mathfrak{q}^{-1})\cdot\mathfrak{r}] \in Z^n. \tag{2.3}$$

When $\mathfrak{q} = 1$, then α is just the translation $\mathfrak{r}$ and the vector systems $\{\mathfrak{f}\}$ and $\{\mathfrak{f}'\}$ are related by a shift of the origin; in this case they are said to be strongly equivalent. (If $\mathfrak{q} \neq 1$, the systems are said to be weakly equivalent.)

Within each arithmetic class, each equivalence class of solutions of (2.2) characterizes an isomorphism class of space groups. In three dimensions there are 73 arithmetic classes and 219 isomorphism classes of space groups the 230 space groups are obtained by conjugation in the subgroup of *proper* motions in $R^n \circledS GL(n,Z)$.

C. Subgroups and subgroup lattices

As we mentioned above, a subgroup of a space group is an extension of a translation subgroup by a subgroup of the point group, that is, $\mathfrak{L} \leqslant \mathcal{G}$ is an extension of $\mathfrak{T}_{\mathfrak{L}}$ by $\mathfrak{P}_{\mathfrak{L}} \sim \mathfrak{L}/\mathfrak{T}_{\mathfrak{L}}$, where $\mathfrak{T}_{\mathfrak{L}} \leqslant Z^n$ and $\mathfrak{P}_{\mathfrak{L}} \leqslant \mathfrak{P}_{\mathcal{G}}$. Therefore $\mathfrak{T}_{\mathfrak{L}}$ must be invariant under $\mathfrak{P}_{\mathfrak{L}}$. Here $\mathfrak{T}_{\mathfrak{L}}$ is isomorphic to Z^m, $0 \leqslant m \leqslant n$. We can specify $\mathfrak{T}_{\mathfrak{L}}$ by writing a set of m independent generators which are n-tuples of integers, as column vectors of an $n \times m$ matrix A. It is important to note that, alternatively, $\mathfrak{T}_{\mathfrak{L}}$ can be specified by any matrix AX, where $X \in GL(m,Z)$. Thus the condition that $\mathfrak{T}_{\mathfrak{L}}$ is invariant under $\mathfrak{P}_{\mathfrak{L}}$ becomes $\mathfrak{P}_{\mathfrak{L}}A = AX$ or, since linear independence of the m generators ensures that A has a left inverse,

$$A^{-1}\mathfrak{P}_{\mathfrak{L}}A \in GL(m,Z). \tag{2.4}$$

An element of $\mathfrak{L}$ can be written in the form $(\mathfrak{t} + \mathfrak{f}_{\mathfrak{L}} + \mathfrak{f})\mathfrak{p}$, where $\mathfrak{t} \in \mathfrak{T}_{\mathfrak{L}}$, $\mathfrak{f}_{\mathfrak{L}} \in Z^n$, $\mathfrak{p} \in \mathfrak{P}_{\mathfrak{L}}$, and $\mathfrak{f}$ is the fractional translation associated with $\mathfrak{p}$ in $\mathcal{G}$. Then $(\mathfrak{f}_{\mathfrak{L}} + \mathfrak{f})\mathfrak{p}$ can be interpreted as a right coset representative of $\mathfrak{T}_{\mathfrak{L}}$ in $\mathfrak{L}$ and therefore $\mathfrak{f}_{\mathfrak{L}}$ is uniquely determined modulo $\mathfrak{T}_{\mathfrak{L}}$. The $\mathfrak{f}_{\mathfrak{L}}$'s must satisfy the subgroup congruence[7]

$$\mathfrak{f}_{\mathfrak{L}1} + \mathfrak{p}_1\cdot\mathfrak{f}_{\mathfrak{L}2} - \mathfrak{f}_{\mathfrak{L}3} + \mathfrak{t}_{1,2} \in \mathfrak{T}_{\mathfrak{L}}, \tag{2.5}$$

which follows from Eq. (2.2).

Each solution set defines a unique subgroup of $\mathcal{G}$. We emphasize this because in order to construct the lattice of subgroups of $\mathcal{G}$ we need to specify *all* the subgroups, and not only isomorphism classes.

Now it is a simple matter to construct the lattice of subgroups of $\mathcal{G}$, that is, to determine their partial ordering. Let $\mathfrak{K}$ and $\mathfrak{L}$ be subgroups of $\mathcal{G}$, with translation subgroups $\mathfrak{T}_{\mathfrak{K}}$ and $\mathfrak{T}_{\mathfrak{L}}$ represented by matrices $A_{\mathfrak{K}}$ and $A_{\mathfrak{L}}$.

Lemma 2.1: $\mathfrak{T}_{\mathfrak{K}} \leqslant \mathfrak{T}_{\mathfrak{L}}$ if and only if $A_{\mathfrak{L}}^{-1}A_{\mathfrak{K}}$ has integer entries.

Proof: $A_{\mathfrak{L}}^{-1}A_{\mathfrak{K}} = Y$ if and only if $A_{\mathfrak{K}} = A_{\mathfrak{L}}Y$. Y is integral if and only if every element of $\mathfrak{T}_{\mathfrak{K}}$ is a linear combination of the generators of $\mathfrak{T}_{\mathfrak{L}}$. Q.E.D.

Theorem 2.1: Let $\mathfrak{K}$ and $\mathfrak{L}$ be subgroups of $\mathcal{G}$. Then $\mathfrak{K} \leqslant \mathfrak{L}$

if and only if (i) $\mathfrak{P}_{\mathfrak{R}} \leqslant \mathfrak{P}_{\mathfrak{L}}$; (ii) $\mathfrak{T}_{\mathfrak{R}} \leqslant \mathfrak{T}_{\mathfrak{L}}$; and (iii) for each element $(\mathfrak{f}_{\mathfrak{R}} + \mathfrak{f})\mathfrak{p}$ of $\mathfrak{R}$ there is an element $(\mathfrak{f}_{\mathfrak{L}} + \mathfrak{f})\mathfrak{p}$ of $\mathfrak{L}$ such that $\mathfrak{f}_{\mathfrak{R}} = \mathfrak{f}_{\mathfrak{L}} \pmod{\mathfrak{T}_{\mathfrak{L}}}$.

Proof: If $\mathfrak{R} \leqslant \mathfrak{L}$ then (i) and (ii) are immediate and also $(\mathfrak{f}_{\mathfrak{R}} + \mathfrak{f})\mathfrak{p} \in \mathfrak{L}$, which means that it can be written in the form $(\mathfrak{t} + \mathfrak{f}_{\mathfrak{L}} + \mathfrak{f})\mathfrak{p}$, where $\mathfrak{t} \in \mathfrak{T}_{\mathfrak{L}}$. Thus $\mathfrak{f}_{\mathfrak{R}}$ is congruent to $\mathfrak{f}_{\mathfrak{L}}$ mod $\mathfrak{T}_{\mathfrak{L}}$. Conversely it is obvious that (i)–(iii) ensure that $\mathfrak{R} \leqslant \mathfrak{L}$. Q.E.D.

In applications to isotropy subgroups we will be interested in the sublattice of isotropy subgroups which has as its minimal subgroup the kernel of the representation. We will also wish to identify the subgroups in this lattice by space group type. This is not a trivial problem. In the first place, if the dimension m of $\mathfrak{T}_{\mathfrak{L}}$ is less than n, then $\mathfrak{L}$ is not an n-dimensional crystallographic space group and it need not even be an m-dimensional one. Secondly, even if $m = n$, the identification of $\mathfrak{L}$ is not always obvious. First $\mathfrak{P}_{\mathfrak{L}}$ must be described as an action on Z^n and this is achieved by the conjugation $\mathfrak{P}'_{\mathfrak{L}} = A_{\mathfrak{L}}^{-1}\mathfrak{P}_{\mathfrak{L}}A_{\mathfrak{L}}$, since $A_{\mathfrak{L}}^{-1}$ maps $\mathfrak{T}_{\mathfrak{L}}$ back onto Z^n. However it is often a difficult problem to determine the conjugacy class of $\mathfrak{P}'_{\mathfrak{L}}$ in GL(n,Z), i.e., to determine the arithmetic class of $\mathfrak{L}$. Next, the equivalence class of the vector system of $\mathfrak{L}$ [see Eq. (2.3)] must be determined. We will not discuss the identification problem further here; a computer program for dealing with it has recently been written by Engel, and will be published soon.[13]

III. TRANSLATION SUBGROUPS OF ISOTROPY SUBGROUPS OF G

Throughout this section, we will adopt a geometric point of view, returning to the algebraic in the concluding section. Thus R^n will be considered to be a real vector space with an orthonormal basis, and its elements n-tuples of real numbers. The usual scalar product of $\mathbf{k} = (k_1,...,k_n)$ and $\mathbf{t} = (t_1,...,t_n)$ in R^n is $(\mathbf{k},\mathbf{t}) = k_1t_1 + \cdots + k_nt_n$.

Let T be a (translation) subgroup of R^n. Then $\widetilde{T} \overset{\text{def}}{=} \{\mathbf{k} \in R^n\colon\ (\mathbf{k},\mathbf{t}) \in Z \text{ for all } \mathbf{t} \in T\}$ is the subgroup of R^n "reciprocal" to T. Clearly $\widetilde{\widetilde{T}} \geqslant T$ and if $T < T'$ then $\widetilde{T}' \leqslant \widetilde{T}$. Consequently, although $\widetilde{\widetilde{T}}$ need not equal T it is always true that $\widetilde{\widetilde{\widetilde{T}}} = \widetilde{T}$. If T_G is the translation subgroup of an n-dimensional space group G then we can identify it with a point lattice and $\widetilde{T}_G$ with its reciprocal lattice. More generally, if $T < R^n$ is generated by m linearly independent translations, $0 \leqslant m \leqslant n$, as in the case $T \leqslant T_G$, then $\widetilde{\widetilde{T}} = T$. Furthermore, if T' is generated by m' linearly independent translations, $0 \leqslant m' \leqslant n$, then $T \leqslant T'$ if and only if $\widetilde{T} \geqslant T'$.

Each irreducible representation $\mathbf{R}$ of T_G is defined by a vector $\mathbf{k} \in R^n$ which is unique modulo $\widetilde{T}_G$; in other words $\mathbf{R}$ is labeled by a "wave" vector $\mathbf{k}$ from the first Brillouin zone $R^n/\widetilde{T}_G$ (see Ref. 14). Specifically, for each $\mathbf{t} \in T_G$, $\mathbf{R}_{\mathbf{k}}(\mathbf{t}) = e^{2\pi i(\mathbf{k},\mathbf{t})}$.

An irreducible representation $\mathbf{R}$ of a space group G is, in general, reducible when restricted to its translation subgroup T_G. Specifically,

$$\mathbf{R}(T_G) = l \overset{s}{\underset{i=1}{\oplus}} \mathbf{R}_{\mathbf{k}_i}(T_G). \tag{3.1}$$

The positive integer l and the set (star) $[\sigma_0] \overset{\text{def}}{=} \{\mathbf{k}_1,\mathbf{k}_2,...,\mathbf{k}_s\}$ are uniquely determined by $\mathbf{R}$. There is a one-to-one correspondence between the vectors $\mathbf{k}_i$ and the left cosets of the isotropy group (little point group) of $\mathbf{k}_1 \in R^n/\widetilde{T}_G$ in the group P_G: $[\sigma_0]$ is the orbit of $\mathbf{k}_1$ under the action of P_G on $R^n/\widetilde{T}_G$.

If K is the kernel of the representation $\mathbf{R}$ then an isotropy subgroup L of G is a subgroup of G which contains K. Clearly $T_K \leqslant T_L$. However the fact that L is an isotropy subgroup leads to a much stronger characterization of T_L. Establishing such a characterization is the purpose of the present section.

Let σ be any subset of $[\sigma_0]$. Then we can construct the translation group

$$T(\sigma) \overset{\text{def}}{=} \{\mathbf{t} \in T_G\colon (\mathbf{k},\mathbf{t}) \in Z \text{ for all } \mathbf{k} \in \sigma\}. \tag{3.2}$$

Clearly, $T(\sigma) \leqslant T_G$ while $T(\varnothing) = T_G$ and $T([\sigma_0]) = T_K$, where $\varnothing$ is the empty set. We now show that $T(\sigma)$ is reciprocal to the group generated by σ and $\widetilde{T}_G$.

Lemma 3.1: Let

$$Q(\sigma) \overset{\text{def}}{=} \{\mathbf{k}\colon \mathbf{k} = \mathbf{k}' + \mathbf{k}'', \text{ where } \mathbf{k}' = \Sigma n_i k_i,\ k_i \in \sigma,\ n_i \in Z, \text{ and } \mathbf{k}'' \in \widetilde{T}_G\}. \tag{3.3}$$

Then $T(\sigma) = \widetilde{Q}(\sigma)$.

Proof: Since $T(\sigma) = \{\mathbf{t} \in T_G\colon\ (\mathbf{k}_i,\mathbf{t}) \in Z \text{ for all } \mathbf{k}_i \in \sigma\}$ and $\widetilde{\widetilde{T}}_G = T_G$ we have $T(\sigma) = \{\mathbf{t} \in R^n\colon\ (\mathbf{k}'',\mathbf{t}) \in Z \text{ for all } \mathbf{k}'' \in \widetilde{T}_G$ and $(\mathbf{k}_i,\mathbf{t}) \in Z \text{ for all } \mathbf{k}_i \in \sigma\} = \{\mathbf{t} \in R^n\colon\ (\mathbf{k}' + \mathbf{k}'',\mathbf{t}) \in Z$ for all $\mathbf{k}' = \Sigma n_i \mathbf{k}_i,\ \mathbf{k}_i \in \sigma,\ n_i \in Z$, and $\mathbf{k}'' \in \widetilde{T}_G\} = \{\mathbf{t} \in R^n\colon (\mathbf{k},\mathbf{t}) \in Z$ for all $\mathbf{k} \in Q(\sigma)\} = \widetilde{Q}(\sigma)$. Q.E.D

When $Q(\sigma)$ defined by Eq. (3.3) *can be* generated by n linearly independent translations in R^n it holds that $\widetilde{\widetilde{Q}}(\sigma) = \widetilde{T}(\sigma)$ is equal to $Q(\sigma)$. Otherwise, although $\widetilde{Q}(\sigma) = T(\sigma)$, $\widetilde{\widetilde{Q}}(\sigma) = \widetilde{T}(\sigma)$ is strictly larger than $Q(\sigma)$. The first case occurs when the image $\mathbf{R}(G) \sim G/K$ is finite and corresponds to commensurate (crystal-to-crystal) phase transitions. The second case occurs when the image $\mathbf{R}(G) \sim G/K$ is infinite and corresponds to incommensurate phase transitions.

It is not necessary to consider $Q(\sigma)$ throughout this paper; it suffices to consider only $T(\sigma)$ given in Eq. (3.2). However, we introduced $Q(\sigma)$ and proved Lemma 3.1 because we found it sometimes conceptually simpler to first construct $Q(\sigma)$ using Eq. (3.3) and then to use Lemma 3.1 to find $T(\sigma) = \widetilde{Q}(\sigma)$ (see Ref. 6).

To establish our main theorem we will need several lemmas. The first of these is immediate.

Lemma 3.2: Let $\sigma' < \sigma'' \leqslant [\sigma_0]$, then $T(\sigma') \geqslant T(\sigma'')$ and, equivalently, $\widetilde{T}(\sigma') \leqslant \widetilde{T}(\sigma'')$.

Lemma 3.2 suggests that there is a maximal σ among those subsets of $[\sigma_0]$ which generate the same $T(\sigma)$. Clearly, every subset σ of $[\sigma_0]$ is contained in $\widetilde{T}(\sigma) \cap [\sigma_0]$ and so a sufficient condition for maximality is that $\sigma = \widetilde{T}(\sigma) \cap [\sigma_0]$. A σ satisfying this latter condition will be called a *substar* and denoted by $[\sigma]$. We will show in Theorem 3.1 that this is also necessary for maximality.

Lemma 3.3: The subduction frequency of a subgroup T of T_G is

$i(T) = l\,|\tilde{T} \cap [\sigma_0]|$.

Proof: By the definition of the subduction frequency (see the Appendix) and Eq. (3.1), $i(T) = l \sum_{j=1}^{s} i_{\mathbf{k}_j}(T)$. Since $i_{\mathbf{k}_j}(T) = 1$ if $\mathbf{k}_j \in \tilde{T}$ and 0 otherwise, the sum in this expression is simply the number of elements in the intersection $\tilde{T} \cap [\sigma_0]$. Q.E.D.

Notice that for a substar $[\sigma]$, $i(T([\sigma])) = l\,|[\sigma]|$.

Lemma 3.4: If $T \leqslant T_G$ then $\tilde{T} \geqslant \tilde{T}(\sigma)$, where $\sigma = \tilde{T} \cap [\sigma_0]$.

Proof: Since both T and $T(\sigma)$ are subgroups of T_G, $\tilde{T} \geqslant \tilde{T}(\sigma)$ is equivalent to $T \leqslant T(\sigma)$. Also $\tilde{\tilde{T}} = T$ so that $T = \{\mathbf{t} \in T_G\colon\ (\mathbf{k},\mathbf{t}) \in Z \text{ for every } \mathbf{k} \in \tilde{T}\}$. Since $\sigma < \tilde{T}$ it then follows that $T \leqslant T(\sigma)$. Q.E.D.

Lemma 3.5: Let $T \leqslant T_G$ and $\sigma = \tilde{T} \cap [\sigma_0]$ as in Lemma 3.4. Then σ is a substar.

Proof: We must show that $\sigma = \tilde{T}(\sigma) \cap [\sigma_0]$. As we pointed out after Lemma 3.2, $\sigma \leqslant T(\sigma) \cap [\sigma_0]$; thus we need only show the reverse inclusion. But this follows immediately from Lemma 3.4. Q.E.D.

Now we can state and prove the first theorem of this section.

Theorem 3.1: Let $[\sigma]$ be a substar of $[\sigma_0]$.

(i) $[\sigma]$ is the maximal subset among the subsets σ' of $[\sigma_0]$ such that $T(\sigma') = T([\sigma])$.

(ii) $T([\sigma])$ is an isotropy subgroup of T_G in the restriction of the representation $\mathbf{R}(G)$ to T_G.

Proof: (i) $[\sigma] = \tilde{T}([\sigma]) \cap [\sigma_0] = \tilde{T}(\sigma') \cap [\sigma_0] \geqslant \sigma'$.

(ii) Assume there is a $T \leqslant T_G$ such that $T([\sigma]) < T$. Then $\tilde{T}([\sigma]) > \tilde{T}$ and by Lemmas 3.4 and 3.5, $\tilde{T}([\sigma]) > \tilde{T}([\sigma'])$, where $[\sigma'] = \tilde{T} \cap [\sigma_0]$. Consequently, $[\sigma'] < [\sigma]$ and by Lemma 3.3, $i(T) < i(T([\sigma]))$. Therefore, $T([\sigma])$ is an isotropy subgroup of T_G by Corollary A.1 of the Appendix. Q.E.D.

Corollary 3.1: The isotropy subgroups of T_G are in one-to-one correspondence with the substars of $[\sigma_0]$.

Proof: We have shown that to each substar there corresponds a unique isotropy subgroup of T_G. Conversely, let T be an isotropy subgroup of T_G and let $[\sigma] = \tilde{T} \cap [\sigma_0]$. Then $T([\sigma]) = T$ because otherwise by Lemma 3.4 $T < T([\sigma])$ and by Lemma 3.3 $i(T) = i(T([\sigma]))$ which would imply by Corollary A.1 that T is not an isotropy subgroup of T_G. Q.E.D.

The following immediate corollary is useful in computations.

Corollary 3.2: Let $\sigma \leqslant [\sigma_0]$. Then $\tilde{T}(\sigma) \cap [\sigma_0]$ is a substar.

We need two additional lemmas in order to prove the main result of the paper: the subgroups $T([\sigma])$ are precisely the translation subgroups of isotropy subgroups of G. We denote the normalizers in P_G of $\sigma \leqslant [\sigma_0]$ and of $T \leqslant T_G$ by $N(\sigma)$ and $N(T)$, respectively. The normalizer of $\tilde{T}$ is obviously equal to $N(T)$, and in particular $N(\tilde{T}_G) = N(T_G) = P_G$.

Lemma 3.6: Let $\sigma \leqslant [\sigma_0]$. Then $N(\sigma) \leqslant N(T(\sigma))$.

Proof: Let p be in $N(\sigma)$. Then p^{-1} is in $N(\sigma)$ and for every $\mathbf{k} \in \sigma$ and every $\mathbf{t} \in T(\sigma)$ $(p\cdot\mathbf{t}, \mathbf{k}) = (\mathbf{t}, p^{-1}\cdot\mathbf{k}) \in Z$ because p is an orthogonal transformation. Therefore, $p\cdot\mathbf{t} \in T(\sigma)$ and, consequently, $p \in N(T(\sigma))$. Q.E.D.

Lemma 3.7: Let $T \leqslant T_G$ and let $[\sigma] = \tilde{T} \cap [\sigma_0]$. Then $N(T) \leqslant N([\sigma])$.

Proof: By definition, if $\mathbf{k} \in [\sigma]$ then $\mathbf{k} \in \tilde{T}$ and $\mathbf{k} \in [\sigma_0]$. If $p \in N(T)$ then $p \in N(\tilde{T}) \leqslant P_G$ and thus $\mathbf{k}'' = p\cdot\mathbf{k} \in \tilde{T}$ and $p\cdot\mathbf{k} = \mathbf{k}' + \mathbf{k}''$, $\mathbf{k}' \in [\sigma_0]$, $\mathbf{k}'' \in \tilde{T}_G \leqslant \tilde{T}$. Therefore, $\mathbf{k}' = \mathbf{k}'' - \mathbf{k}'' \in \tilde{T}$ so that $\mathbf{k}' \in \tilde{T} \cap [\sigma_0]$. Consequently, $p \in N([\sigma])$. Q.E.D.

Theorem 3.2: (i) For every substar $[\sigma]$ there is an isotropy subgroup L of G such that $T_L = T([\sigma])$.

(ii) Conversely, if L is an isotropy subgroup then there exists a substar $[\sigma] = \tilde{T}_L \cap [\sigma_0]$ such that $T_L = T([\sigma])$.

Proof: (i) if $T([\sigma])$ is an isotropy subgroup we are done. If it is not, there is an isotropy group $L \leqslant G$ such that $L > T([\sigma])$ and $i(L) = i(T([\sigma]))$ (see Corollary A.1 of the Appendix); if $T([\sigma]) = T_L$ we are done. Thus we may assume $T([\sigma]) < T_L$. But $i(T([\sigma])) = i(L)$ implies $i(T([\sigma])) = i(T_L)$, since $T([\sigma]) < T_L \leqslant L$. Thus, by Corollary A.1 of the Appendix, $T([\sigma])$ is not an isotropy subgroup of T_G, which contradicts Theorem 3.1.

(ii) Assume that $T_L \neq T([\sigma])$; then $T_L < T([\sigma])$ by Lemma 3.4. Let L be written as a left coset decomposition with respect to T_L: $L = \cup g_i T_L$.

TABLE II. Subgroups of $p4$ such that $\mathfrak{T}_{\mathfrak{L}} = \mathfrak{T}([\sigma])$.

$\mathfrak{T}_{\mathfrak{L}}$	$\mathfrak{P}_{\mathfrak{L}}$	Vector system $\mathfrak{f}_1$	$\mathfrak{f}_p$	$\mathfrak{f}_{p^2}$	$\mathfrak{f}_{p^3}$	Subgroup type	Subgroup
$\begin{bmatrix}2&0\\0&2\end{bmatrix}$	C_2	(0,0)		(0,0)		$p2$	
		(0,0)		(1,0)		$p2$	
		(0,0)		(0,1)		$p2$	
		(0,0)		(1,1)		$p2$	$\mathfrak{R}$
	C_4	(0,0)	(0,0)	(0,0)	(0,0)	$p4$	
		(0,0)	(1,0)	(1,1)	(0,1)	$p4$	$\mathfrak{L}_1$
		(0,0)	(0,1)	(−1,1)	(1,0)	$p4$	$\mathfrak{L}_2$
		(0,0)	(1,1)	(0,0)	(1,1)	$p4$	
$\begin{bmatrix}2&0\\0&1\end{bmatrix}$	C_2	(0,0)		(0,0)		$p2$	
		(0,0)		(1,0)		$p2$	$\mathfrak{L}_3$
$\begin{bmatrix}1&0\\0&2\end{bmatrix}$	C_2	(0,0)		(0,0)		$p2$	
		(0,0)		(0,1)		$p2$	$\mathfrak{L}_4$
$\begin{bmatrix}1&0\\0&1\end{bmatrix}$	C_2	(0,0)		(0,0)		$p2$	$\mathfrak{L}_5$
	C_4	(0,0)	(0,0)	(0,0)	(0,0)	$p4$	$\mathfrak{G}$

Let $L' = \cup g_i T([\sigma])$; obviously $L < L'$. L' is a group since $N(T_L) \leqslant N(T([\sigma]))$ by Lemmas 3.6 and 3.7. By definition (see the Appendix), $i(L) = \dim \mathrm{Fix}\, L$ and $i(L') = \dim \mathrm{Fix}\, L'$. We will show that $\mathrm{Fix}\, L = \mathrm{Fix}\, L'$ and thus $i(L) = i(L')$, which contradicts our assumption that L is an isotropy subgroup (cf. Corollary A.1).

Let $v \in \mathrm{Fix}\, L$. Then v is fixed by T_L and by every $g_i \in L$. Now $i(T_L) = l\,|\widetilde{T}_L \cap [\sigma_0]| = l\,|[\sigma]|$ by Lemma 3.3 and the definition of $[\sigma]$. On the other hand, $i(T([\sigma])) = l\,|\widetilde{T}([\sigma]) \cap [\sigma_0]| = l\,|[\sigma]|$ since $[\sigma]$ is a substar. Thus $i(T_L) = i(T([\sigma]))$ and every vector fixed by T_L is fixed by $T([\sigma])$ and vice versa. Thus v is fixed by $T([\sigma])$ and by every g_i, which implies $v \in \mathrm{Fix}\, L'$, and hence $\mathrm{Fix}\, L \leqslant \mathrm{Fix}\, L'$. It follows (see Proposition A.3 of the Appendix) that $\mathrm{Fix}\, L = \mathrm{Fix}\, L'$. Q.E.D.

IV. THE LATTICE OF ISOTROPY SUBGROUPS: A TWO-DIMENSIONAL EXAMPLE

In principle, the isotropy groups are completely determined by the image $\mathbf{R}(G)$. Thus one could first determine the isotropy subgroups of $\mathbf{R}(G)$ and then "lift back" to G to find its corresponding isotropy subgroups. This procedure would be feasible if representations were classified by their kernels and images. However, such a classification does not exist even for $n = 2$; in the meantime we need an alternative method for a case-by-case study. We can avoid the explicit construction of the matrices in $\mathbf{R}(G)$ by recalling that the lattice of subgroups L of G which contain the kernel K is isomorphic to the lattice of subgroups of $\mathbf{R}(G)$. Therefore, our strategy will be to construct the lattice of subgroups $L, K \leqslant L \leqslant G$, for which T_L corresponds to a substar, and then to calculate each $i(L)$. While the construction of the subgroup lattice does not require any knowledge of the image $\mathbf{R}(G)$, the last step requires knowledge of the characters.

Let $[\sigma_0]$ be the star of the representation. It is a simple matter to calculate the substars $[\sigma]$ and the corresponding translation subgroups $\mathfrak{T}([\sigma])$. Using a standard procedure,[15] we can find a basis for each $\mathfrak{T}([\sigma])$ and form the matrix A defined in Sec. II. (In this section we return to the algebraic point of view.)

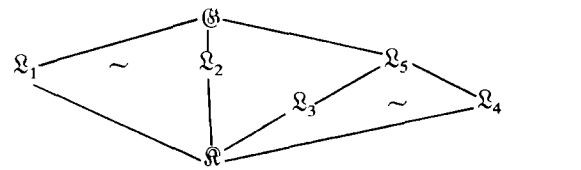
By the results of the preceeding section, we can assume that $\mathfrak{T}_{\mathfrak{L}} = \mathfrak{T}([\sigma])$ for some substar $[\sigma]$. The next step is to construct candidate subgroups $\mathfrak{L}$: for each $\mathfrak{T}_{\mathfrak{L}}$ we choose those $\mathfrak{P}_{\mathfrak{L}}$ under which it is invariant, use Eq. (2.5) to determine a vector system, and then use Theorem 2.1 to exclude those subgroups which do not include $\mathfrak{K}$. If $\mathfrak{K}$ is of finite index, the work is simplified by recalling that $|\mathfrak{P}_{\mathfrak{G}}/\mathfrak{P}_{\mathfrak{K}}| = |\mathfrak{P}_{\mathfrak{G}}/\mathfrak{P}_{\mathfrak{L}}|\,|\mathfrak{P}_{\mathfrak{L}}/\mathfrak{P}_{\mathfrak{K}}|$ and $|\mathfrak{T}_{\mathfrak{G}}/\mathfrak{T}_{\mathfrak{K}}| = |\mathfrak{T}_{\mathfrak{G}}/\mathfrak{T}_{\mathfrak{L}}|\,|\mathfrak{T}_{\mathfrak{L}}/\mathfrak{T}_{\mathfrak{K}}|$ (note: $\mathfrak{T}_{\mathfrak{G}} = Z^n$).

Finally for each $\mathfrak{L}$ in our lattice we compute the subduction frequency $i(\mathfrak{L})$. This computation is explained in detail in Ref. 6. In the following example we show only how to construct the lattice of subgroups whose subduction frequency is to be calculated.

Example: Let $\mathfrak{G}$ be the two-dimensional space group $p4$. For the purposes of this calculation, we assume $\mathbf{R}$ to be an irreducible representation of $\mathfrak{G}$ whose kernel $\mathfrak{K}$ is the invariant subgroup of $\mathfrak{G}$ specified by $A_{\mathfrak{K}} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\mathfrak{P}_{\mathfrak{K}} = C_2 = \{1, \mathfrak{p}^2\}$, and vector system $\mathfrak{f}_1 = (0,0)$, $\mathfrak{f}_{\mathfrak{p}^2} = (1,1)$, where p is the fourfold rotation. Then $|\mathfrak{G}/\mathfrak{K}| = |\mathfrak{P}_{\mathfrak{G}}/\mathfrak{P}_{\mathfrak{K}}|\,|Z^2/\mathfrak{T}_{\mathfrak{K}}| = 2 \cdot 4 = 8$. The star is $[\sigma_0] = \{(1/2,0),(0,1/2)\}$. The substars of $[\sigma_0]$ are $\varnothing$, $[\sigma_0]$, $\{(1/2,0)\}$, and $\{(0,1/2)\}$. The first two correspond to Z^2 and $\mathfrak{T}_{\mathfrak{K}}$, respectively, while if $[\sigma]$ is the third or fourth then $\mathfrak{T}([\sigma]) = \{(x,y) \in Z^2 : x \text{ is even}\}$ or $\{(x,y) \in Z^2 : y \text{ is even}\}$, respectively. We conclude that they are represented by the matrices $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, respectively.

Next we construct the subgroups $\mathfrak{L}$ with these translation lattices. We list in Table II the combinations of $\mathfrak{T}_{\mathfrak{L}}$ and $\mathfrak{P}_{\mathfrak{L}}$ which form subgroups of $\mathfrak{P}_{\mathfrak{G}}$ [see Eq. (2.5)], together with representative vector systems. Note that $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ are not invariant under C_4 and thus there are no $p4$ subgroups with these lattices.

By Theorem 2.1, $\mathfrak{L}$ does not contain $\mathfrak{K}$ unless $\mathfrak{f}_{\mathfrak{p}^2} = (1,1)$ mod $\mathfrak{T}_{\mathfrak{L}}$; this requirement eliminates seven of the subgroups in the list above. Thus our final subgroup lattice is



where $\sim$ indicates conjugate subgroups.

To determine which of these subgroups are isotropy subgroups, one would first calculate the subduction frequencies for the maximal subgroups $\mathfrak{L}_1$ and $\mathfrak{L}_5$ [note that $i(\mathfrak{L}_1) = i(\mathfrak{L}_2)$] and compare them with $i(\mathfrak{G}) = 0$: then one would calculate $i(\mathfrak{L}_3)$ and compare it with $i(\mathfrak{L}_5)$ if $\mathfrak{L}_5$ is an isotropy group or with $i(\mathfrak{G})$ if it is not. Finally, $\mathfrak{K}$ is the isotropy group by a general theorem.

ACKNOWLEDGMENTS

The authors are grateful to Professor L. Michel and the Institut des Hautes Etudes Scientifiques for hospitality during the initial stages of this work, and to Gabriele Weber-Jarić for hospitality while the paper was written.

APPENDIX: GROUP ACTION AND ISOTROPY GROUPS

In this appendix we outline the basic concepts concerning group action and isotropy subgroups which are assumed in Sec. III of the paper.

Let G be a countable group (e.g., a space group) and let $\mathbf{R}$ be a homomorphism of G into a group of unitary matrices which act on a vector space V

$$\forall g \in G,\quad \mathbf{R}: g \to \mathbf{R}(g),\quad \forall v \in V,\quad g \cdot v = \mathbf{R}(g)v, \tag{A1}$$

where $\mathbf{R}(g)$ is the unitary matrix representing g. Usually $\mathbf{R}$ is said to be a representation of G; the matrix group $\mathbf{R}(G) = \{\mathbf{R}(g): g \in G\}$, the image of G under $\mathbf{R}$, is also often called the representation. We recall that the set $K = \{g \in G: \mathbf{R}(g) = 1\}$, where 1 is the identity matrix in $\mathbf{R}(G)$, is a normal subgroup of G, $K \lhd G$, called the kernel of $\mathbf{R}$. The image $\mathbf{R}(G)$ is isomorphic to the quotient group G/K.

Throughout this appendix we will assume that $\mathbf{R}$ is a given representation.

Although the following definitions refer to the G-action defined by Eq. (A1) they may be readily extended to a more general group action.[16]

Definition A.1: $L \leqslant G$ is the *isotropy group* (*stabilizer, little group*) of $v \in V$ iff L is the largest subgroup of G such that $\forall g \in L, g \cdot v = v$. We will write $L = L(v)$. Similarly, we say that $L \leqslant G$ is an *isotropy subgroup* of G iff $\exists v \in V$ such that $L = L(v)$.

The *maximal* isotropy subgroup of G is, trivially, G itself. A reader should be warned that the term "maximal" is used in the physical literature for a maximal isotropy *strict* subgroup of G.

Definition A.2: $N \leqslant G$ is the *normalizer* of $V' \leqslant V$ in G iff N is the largest subgroup of G such that $\forall g \in N$ and $\forall v \in V'$, $g \cdot v \in V'$. We will write $N = N(V')$.

Definition A.3: $C \leqslant G$ is the *centralizer* of $V' \leqslant V$ in G iff C is the largest subgroup of G such that $\forall g \in C$ and $\forall v \in V$, $g \cdot v = v$. We will write $C = C(V')$. Similarly, we say that $C \leqslant G$ is a *centralizer in* G iff $\exists V' \leqslant V$ such that $C = C(V')$.

The following proposition is an immediate consequence of the above definitions.

Proposition A.1:

(i) Every centralizer contains K which is the minimal centralizer in G.

(ii) G is the maximal centralizer.

(iii) For every subspace $V' \leqslant V$,

$$N(V') \geqslant C(V') = \bigcap_{v \in V'} L(v), \tag{A2}$$

and, hence,

$$L(v) \geqslant C(V'), \quad \forall v \in V'. \tag{A3}$$

Definition A.4: The *G-orbit* through $v \in V$ is the set $\{g \cdot v : g \in G\} \leqslant V$. Clearly, the G-orbit through v is also the G-orbit through any of its points. Furthermore, points in the orbit are in one-to-one correspondence with the set $G/L(v)$, the set of left cosets of $L(v)$ in G. Sometimes, e.g., when $L(v) = K$, we will use the same notation for the set of corresponding coset representatives.

Since points on a G-orbit have conjugate isotropy groups, $L(g \cdot v) = gL(v)g^{-1}$, the following is clear.

Proposition A.2: Isotropy subgroups of G fall into equivalence classes under conjugation in G.

Definition A.5: Let L be any subgroup of G. Fix L is the largest vector subspace of V such that $C(\text{Fix } L) \geqslant L$.

The following proposition is an immediate consequence of the definitions.

Proposition A.3: Let L and L' be subgroups of G.

(i) If $L < L'$ then Fix $L \geqslant$ Fix L'; conversely, if Fix $L >$ Fix L' then $L < L'$.

(ii) $C(\text{Fix } L) = L$ if and only if L is a centralizer in G.

(iii) For every $V' \leqslant V$, Fix $C(V') \geqslant V'$.

(iv) Fix $L = \{v \in V : L(v) \geqslant L = \cup L'$, where $L' \in I(G), L' \geqslant L$, and $I(G)$ is the set of isotropy subgroups of $G.\}$

The dimensionality of Fix L, $i(L) = \dim$ Fix L, called the subduction frequency of L, is the number of times the trivial, identity representation occurs in the restriction of $\mathbf{R}$ to L. For example, $i(K) = \dim V$. In the case $\mathbf{R}(L)$ is finite, that is $L \cap K$ is of finite index in L, $i(L)$ can be calculated using the orthogonality relations for the characters of a finite group:

$$i(L) = \left| \frac{L}{L \cap K} \right|^{-1} \sum_{g \in |L / (L \cap K)|} \operatorname{Tr} \mathbf{R}(g). \tag{A4}$$

We emphasize that the following theorem holds for continuous as well as countable groups.

Theorem A.1: A subgroup L of G is a centralizer in G if and only if there exists no subgroup L' such that $L < L' \leqslant G$ and $i(L') = i(L)$.

Proof: Let L be a centralizer in G. Then $L = C(\text{Fix } L)$ by Proposition A.3 (ii). Let $L' \leqslant G$ be such that $L' > L$ and $i(L') = i(L)$. By Proposition A.3 (i) Fix $L' \leqslant$ Fix L and since both Fix L' and Fix L are vector subspaces of V, $i(L') = i(L)$ implies Fix $L' =$ Fix L. Furthermore, by the definition and Proposition A.3 (ii), $C(\text{Fix } L') \geqslant L'$. Consequently, $L \geqslant L'$ which is a contradiction. To prove the second part of the theorem, assume that $i(L') < i(L)$ for every subgroup L' of G such that $L' > L$. Therefore, $C(\text{Fix } L)$ is either equal to L, in which case L is a centralizer by Proposition A.3 (ii), or it is equal to such an L'. In the latter case $i(L') < i(L)$ implies Fix $L' <$ Fix L contradicting Fix $C(\text{Fix } L) \geqslant$ Fix L which follows from Proposition A.3 (iii). Q.E.D.

The following theorem, which is only true for countable groups, will facilitate the use of the previous theorem to calculate isotropy groups.

Theorem A.2: Every isotropy subgroup of G is a centralizer in G and, conversely, every centralizer in G is an isotropy subgroup of G.

Proof: Let L be an isotropy subgroup of G. It follows immediately from Proposition A.1 (iii) and the Definition A.5 that $C(\text{Fix } L) = L$, which proves that L is a centralizer. Let L be a centralizer in G such that the set of isotropy subgroups L_i of G which contain L, $L_i \geqslant L$, is countable (this is always the case for a countable group G). By Proposition A.3 (iv), we have Fix $L = \cup_i$ Fix L_i. A countable union of vector subspaces (e.g., $\cup_i$ Fix L_i) is a vector subspace (e.g., Fix L) if and only if one of the subspaces in the union (e.g., Fix L_0, $L_0 \in \{L_i\}$) contains all the others. Thus, Fix L = Fix L_0 which implies by Proposition A.3 (ii) that $L = L_0$. Hence, we proved that L is an isotropy subgroup of G. Q.E.D.

The most important result for calculating isotropy subgroups of a countable group G is the following immediate corollary to Theorems A.1 and A.2.

Corollary A.1:

(i) A subgroup L of G is an isotropy subgroup of G if and only if there exists no subgroup L' such that $L < L' \leqslant G$ and $i(L') = i(L)$.

(ii) Let $L < L' \leqslant G$. Then $i(L) = i(L')$ if and only if there exist no *isotropy* subgroups L'' of G such that $L \leqslant L'' < L'$.

A general strategy for determining all of the isotropy subgroups of G is then to systematically calculate $i(L)$ for all subgroups L of G such that $L \geqslant K$ and then to apply the above corollary.

A justification of our restriction to irreducible representations in the main text is based on the following lemma.

Lemma A.1: If $\mathbf{R}$ is reducible, $\mathbf{R} = \oplus \mathbf{R}_i$, then L is an $\mathbf{R}$-isotropy subgroup of G if and only if $L = \cap L_i$, where each L_i is an $\mathbf{R}_i$-isotropy subgroup of G.

The proof follows immediately from the observation that irreducible subspaces V_i corresponding to $\mathbf{R}_i$, $V = \oplus V_i$, are G-invariant subspaces.

Finally, although we do not use the following lemma in this paper, we prove it here since it turns out to be very useful in practical calculations of isotropy subgroups.[6]

Lemma A.2[1]: If L and L' are centralizers in G then $L \cap L'$ is also a centralizer and $i(L \cap L') \geqslant i(L) + i(L') - i(LL')$, where LL' is the group generated by L and L'.

Proof: Let $V' \leqslant V$ be the subspace generated by Fix L and Fix L'. Clearly $\forall v \in V', \forall g \in L \cap L', g \cdot v = v$. Thus $C(V') \geqslant L \cap L'$. On the other hand $C(V') = \cap_{v \in V'} L(v) \leqslant L \cap L'$. Therefore, $C(V') = L \cap L'$ and $L \cap L'$ is a centralizer in G or, equivalently, by Lemma 2.1, an isotropy subgroup of G. Furthermore, $V' = V'' \oplus V_L^{\sim} \oplus V_{L'}^{\sim}$, where $V'' = \text{Fix } L \cap \text{Fix } L'$ and $V_L^{\sim}$ and $V_{L'}^{\sim}$ are orthogonal complements of V'' in Fix L and Fix L', respectively. Clearly $\dim V' = i(L) + i(L') - \dim V''$. On the other hand $C(V'') \geqslant LL'$ and by construction, $\text{Fix}(LL') = \text{Fix } L \cap \text{Fix } L'$ implies $\dim V'' = i(LL')$. Finally, using a general relation Fix $C(V') \geqslant V'$ we arrive at $i(L \cap L') \geqslant i(L) + i(L') - i(LL')$. Q.E.D.

[1] See, for example, a review by M. V. Jarić, Physica A **114**, 550 (1982).

[2] See, for example, a review by R. Slansky, Phys. Rep. C **79**, 1 (1981).

[3] L. D. Landau, Zh. Eksp. Teor. Fiz. **7**, 19 (1937); L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, New York, 1968), 2nd ed.

[4] See A. Aharony, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 6, p. 358, and references therein.

[5] M. V. Jarić, Phys. Rev. Lett. **48**, 164 (1982).

[6] M. V. Jarić, J. Math Phys. **24**, 2865 (1983).

[7] M. Senechal, Acta Crystallogr. A **36**, 845 (1980).

[8] This distinction between the two points of view is important in concrete calculations and, clearly, for the distinction to make any sense Z^n, R^n, and O(n) must denote *concrete* realizations of these groups. Thus, throughout the paper Z^n will denote the additive group of n-tuples of all integers. R^n will denote the additive group of all n-tuples of real numbers which correspond to an orthonormal basis of R^n when R^n is considered as a vector space. Corresponding group of all $n \times n$ orthogonal matrices acting on R^n will be denoted by O(n).

[9] L. Bieberbach, Math. Ann. **72**, 400 (1911).

[10] H. Zassenhaus, Comm. Math. Helv. **21**, 117 (1948). A classification of four-dimensional space groups has been completed recently in H. Brown, R. Bülow, J. Neubüser, H. Wondratschek, and H. Zassenhaus, *Crystallographic Groups of Four-Dimensional Space* (Wiley, New York, 1978).

[11] A discussion of the geometric and algebraic viewpoints as well as the enumeration two- and three-dimensional space groups is given in M. Klemm, *Symmetrien von Ornamenten und Kristallen* (Springer, New York, 1982).

[12] With respect to the orthonormal basis of R^n, γ is an $n \times n$ matrix in GL(n,R). Explicitly, the columns of γ^{-1} are the generators of T_G in the given basis of R^n. Thus, $\gamma(\mathbf{t}) = \gamma \cdot \mathbf{t}$, $\gamma(\mathbf{f}) = \gamma \cdot \mathbf{f}$ and $\gamma(p) = \gamma p \gamma^{-1}$, where for example, the ith component of $\gamma(\mathbf{t})$ is $(\gamma \cdot \mathbf{t})_i = \Sigma_{j=1}^{n} \gamma_{ij} t_j$. Clearly, γ^{-1} is defined up to right-multiplication by an element (matrix) of GL(n,Z).

[13] P. Engel (to be published).

[14] The quotient groups like $R^n/\tilde{T}_G$ will be identified throughout this paper as the groups of cosets of $\tilde{T}_G$ in R^n. Furthermore, $\mathbf{k} \in R^n/\tilde{T}_G$ will indicate either a particular coset or any of the elements of that coset. It will be always clear which is intended. Whenever $\mathbf{k}$ stands for an element of a coset it will be irrelevant which particular element $\mathbf{k}$ is.

[15] J. W. S. Cassel, *Introduction to the Theory of Numbers* (Springer, New York, 1958).

[16] For a more general treatment see, for example, L. Michel, Rev. Mod. Phys. **52**, 617 (1980); or M. V. Jarić, *Proceedings of the XIIth International Colloquium on Group Theoretical Methods in Physics* (Trieste, 1983), and references contained therein.

Classification of systems of nonlinear ordinary differential equations with superposition principles

S. Shnider[a)]
Department of Mathematics, McGill University, Montréal, Québec H3A 2K6, Canada

P. Winternitz
Centre de recherche de mathématiques appliquées, Université de Montréal, C.P. 6128, Montréal, Québec H3C 3J7, Canada

(Received 23 September 1983; accepted for publication 30 March 1984)

A system of n first-order nonlinear ordinary differential equations $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x},t)$ is said to admit a superposition principle if its general solution can be written as a function of a finite number m of particular solutions and n constants. Such a system can be associated with the nonlinear action of a Lie group G on a space M. We show that "indecomposable" systems of ODE's with supersposition principles are obtained if and only if the Lie algebras $L_0 \subset L$, corresponding to the isotropy group H of a point and the group G, respectively, define a transitive primitive filtered Lie algebra (L,L_0). Using known results from the theory of transitive primitive Lie algebras we deduce that L_0 must be a maximal subalgebra of L and that G must be an affine group, a simple Lie group, or the direct product of two identical simple Lie groups. Affine groups lead to linear equations, the other types to nonlinear equations with polynomial or rational nonlinearities. Equations corresponding to the classical complex Lie algebras are worked out in detail.

PACS numbers: 02.20. + b, 02.40. + m, 11.10.Lm, 11.30.Na

I. INTRODUCTION

Certain nonlinear partial differential equations display the interesting property that new solutions can be obtained by a nonlinear superposition of specific known solutions (soliton superposition laws).[1] This property has been related to the matrix Riccati form of the soliton generating Bäcklund transformations.[2] A related property of certain ordinary differential equations, including all matrix Riccati equations, is the existence of nonlinear superposition principles, i.e., the possibility of expressing the general solution as a function of a finite number of particular solutions (a "fundamental set of solutions"). The question of characterizing systems of ODE's that admit nonlinear superposition principles was raised and solved by Lie.[3]

More specifically, we shall say that the vector ODE

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(x,t), \quad \mathbf{x}(t),\mathbf{f}(\mathbf{x},t) \in \mathbb{R}^n \tag{1.1}$$

admits a nonlinear superposition principle if there exist m "generic" particular solutions $\mathbf{x}_1,\ldots,\mathbf{x}_m$ and a function

$$S:\mathbb{R}^{n(m+1)} \to \mathbb{R}^n$$

such that the general solution to (1.1) can be written as

$$\mathbf{x}(t) = S(\mathbf{x}_1(t),\ldots,\mathbf{x}_m(t),\mathbf{a}), \tag{1.2}$$

where $\mathbf{a}$ is a constant vector, related to the initial conditions. The expression (1.2) will be called a superposition formula and the set $\mathbf{x}_1(t),\ldots,\mathbf{x}_m(t)$ will be called a fundamental set of solutions.

A recent series of publications[4-7] has been devoted to a study of systems of first-order ODE's with quadratic nonlinearities that satisfy Lie's criterion. Thus, explicit superposition formulas were obtained for projective and conformal Riccati equations,[4,5] as well as for the general matrix Riccati equation.[6]

The purpose of this paper is to provide a general classification of nonlinear ODE's with superposition principles and to reduce the study of these equations to that of certain "elementary" or "indecomposable" systems of equations (see below).

If $\mathbf{f}$ is linear in $\mathbf{x}$ this is just the representation of a general solution as the linear combination of n linearly independent solutions and in this case generic means linearly independent. If $\mathbf{f}$ is independent of t and nonvanishing near $\mathbf{a}_1 \in \mathbb{R}^n$ we can change coordinates on a neighborhood of $\mathbf{a}_1$ so that $\mathbf{f}$ becomes $\tilde{\mathbf{f}} = (1,0,\ldots,0)$ and the general solution to the transformed equation is $\tilde{\mathbf{x}}(t) = \tilde{\mathbf{x}}_1(t) + \mathbf{a} - \mathbf{a}_1$. Changing back to the original coordinates this becomes a nonlinear superposition with $m = 1$. Generic in this case means that the solution $\mathbf{x}_1$ has a nowhere vanishing tangent vector.

More generally for the nonlinear time-dependent system a theorem of Lie[3] gives necessary and sufficient conditions for there to exist a nonlinear superposition principle. Consider $\mathbf{f}(\mathbf{x},t)$ as defining a time-dependent vector field on $\mathbb{R}^n$

$$\xi(\mathbf{x},t) = \sum_{i=1}^{n} f^i(\mathbf{x},t)\frac{\partial}{\partial x^i}. \tag{1.3}$$

The vector fields $\xi(\mathbf{x},t_0)$ evaluated at all possible t_0 must generate a finite-dimensional subalgebra of the algebra of vector fields on $\mathbf{R}^n$, i.e., there exist k vector fields $\xi_1(\mathbf{x}),\ldots,\xi_k(\mathbf{x})$ such that

$$\xi(\mathbf{x},t) = \sum_{i=1}^{k} a_i(t)\xi_i(\mathbf{x}) \tag{1.4}$$

and

$$[\xi_i(\mathbf{x}),\xi_j(\mathbf{x})] = \sum_{l=1}^{k} c_{ij}^l \xi_l(\mathbf{x}) \tag{1.5}$$

for some constants c_{ij}^l.

[a)] Current address: Department of Mathematics, Ben Gurion University, Beersheva, Israel, 84105.

 0022-2488/84/113155-11$02.50

When these conditions hold, there is a neighborhood of the identity in a k-dimensional Lie group G which is acting as a (local) transformation group on a neighborhood U of the initial values $\mathbf{a}_0$, which we assume to be the origin of $\mathbb{R}^n$. The differential equation on $\mathbb{R}^n$ can then be solved for all initial values in U by finding the integral curve $g(t)$ through the identity in G of a time-dependent vector field on G and composing with the group action, which we write as $(g,x)\mapsto gx$ or $G\times\mathbb{R}^n\to\mathbb{R}^n$, although everything is only locally defined. The function $S(\mathbf{x}_1,\ldots,\mathbf{x}_m,\mathbf{a})$ is a composition of two maps, the first determines the solution $g(t)$ in the group from an m-tuple of solutions in $\mathbb{R}^n$ and the second is a composition of $g(t)$ with initial conditions $\mathbf{x}(0)=\mathbf{a}$

$$\{\mathbf{x}_i(t)=g(t)\mathbf{a}_i\}\to g(t)\to g(t)\mathbf{a}\,. \tag{1.6}$$

The integer m is the number of copies of $\mathbb{R}^n$ necessary to make the "generic" isotropy group of the action on $(\mathbb{R}^n)^m$ reduce to the identity.

Lie's results reduce the problem of finding all systems of n first-order ODE's with superposition principles to that of classifying the finite subalgebras of the algebra of vector fields on $\mathbb{R}^n$. They also make it possible to read off the equations directly from the expression for the fields and vice versa. Indeed, given the equations of the required form

$$\frac{dx^\mu}{dt}=\sum_{i=1}^{k}a_i(t)\xi_i^{\ \mu}(\mathbf{x})\,,\quad 1\leqslant\mu\leqslant n \tag{1.7}$$

we obtain the vector fields

$$\hat{\xi}_i(\mathbf{x})=\sum_{\nu=1}^{n}\xi_i^{\ \nu}(\mathbf{x})\frac{\partial}{\partial x^\nu}\,. \tag{1.8}$$

Given the vector fields, we can write the equations as

$$\frac{dx^\mu}{dt}=\sum_{i=1}^{k}a_i(t)\hat{\xi}_i\cdot x^\mu\,. \tag{1.9}$$

In a different context, Lie himself classified all finite subalgebras of the algebra of vector fields on $\mathbb{R}^1$ and $\mathbb{R}^2$.[8]

For $n=1$ only sl(2,$\mathbb{R}$) and its subalgebras can be realized, leading to the Riccati equation $\dot{y}=a+by+cy^2$, or linear equations, respectively. For $n=2$ infinitely many different finite-dimensional Lie algebras can be realized. Two of them, namely those that in modern terms correspond to two-dimensional symmetric spaces [quotients of sl(3,$\mathbb{R}$) and O(3,1)] lead to coupled Riccati type equations. The equations corresponding to all other subalgebras (none of them simple) partially decouple. A Riccati or linear equation is obtained in one variable; once this equation is solved the remaining equation reduces to a linear or Riccati equation for one unknown function.

In seeking a generalization of Lie's "decoupling" result and an understanding of a reduction procedure for the general equation (in $\mathbb{R}^n$) admitting a nonlinear superposition principle, one is led to the study of primitive transitive group actions. We claim that a general superposition law on $\mathbb{R}^n$ can be derived from a knowledge of superposition laws for primitive transitive group actions on $\mathbb{R}$.[9–15] Furthermore, we shall show that the "building blocks" for constructing nonlinear equations with superposition principles are obtained by constructing the vector fields, corresponding to the action of a group G on a homogeneous space G/H, where either G is a simple Lie group and H a maximal subgroup of G, or $G=K\otimes K$, $H=K_D$, where K is simple and K_D is the diagonal K-type subgroup of $K\otimes K$.

The nonlinear equations with superposition principles that we obtain are not necessarily of the Riccati type, i.e., the nonlinearities are not necessarily quadratic. The analysis does however bring out the prevalence of matrix Riccati equations in the theory.

The approach taken in this article provides both the ODE's with superposition principles and the general form of the superposition formula in each case: this is given by the group action of G on G/H.

In Sec. II we relate the problem of ODE's with superposition principles to the theory of primitive transitive Lie algebras. In Sec. III we apply known results on such Lie algebras to construct the ODE's we are interested in. In this article we restrict ourselves to the classical complex simple Lie groups. The exceptional simple Lie groups and the real Lie groups will be treated elsewhere.

II. SUPERPOSITION PRINCIPLES AND PRIMITIVE TRANSITIVE LIE ALGEBRAS

Let us consider a Lie group G acting on a manifold M. The first restriction that we make is that the orbit structure of the group is regular, i.e., admits a stratification into submanifolds of fixed orbit type. In fact we restrict our attention to the superposition of solutions, all of which take values in a stratum of one orbit type G/H. The superposition may be based on solutions $\mathbf{x}_1,\ldots,\mathbf{x}_m$ taking values in different orbits, but since they are all diffeomorphic to G/H, we will assume that all the $\mathbf{x}_i$ take values in the same orbit, which we identify with G/H and further, that H contains no normal subgroup of G. Thus, for the purpose of studying the superposition formulas we shall assume further that the group action is transitive and effective.

Assume now that the open set $U\subset\mathbb{R}^n$ is a coordinate neighborhood of the base point x_0 in G/H. We say that G acts primitively on G/H, if there is no invariant foliation. The group G acts locally primitively if there is no invariant foliation of the open set U. A group G may act primitively, but not locally primitively, when there is a foliation on a covering of G/H, or in other words, if the local foliations do not fit together to a global foliation on G/H. Golubitsky[12] deals with this problem but it will not concern us here.

We will show below that the action G fails to be locally primitive if and only if there exist coordinates $(x^1,\ldots,x^n)$ such that dx^i/dt for $i=1,\ldots,r$ is a function only of $x^1,\ldots,x^r$ and t (for some $0<r<n$). The group G acts primitively if and only if there does not exist a subgroup K such that

$$H\subset K\subset G\,,\quad \dim H<\dim K<\dim G\,. \tag{2.1}$$

The group G acts locally primitively if and only if there does not exist an algebra k between the algebras h and g: $h\subset k\subset g$, $h\neq k\neq g$.

If G does not act locally primitively, then there is an invariant foliation on U and projecting along the leaves of the foliation onto the quotient space, we can define a local action of G on a lower-dimensional manifold.

If there exists a subgroup K satisfying (2.1) we can identify the quotient space given by projecting along the leaves with a neighborhood of the base point in G/K. We can determine the curve $g(t)$ in G from solutions (that is integral curves) of the projected vector fields on G/K, if G acts effectively on G/K. This is true whenever K does not contain a normal subgroup of G. The local condition is that k does not contain an ideal of g. If K is chosen to be maximal (or k is maximal), then the action is primitive (or locally primitive).[12] If K contains no normal subgroup of G (or k contains no ideal of g) the superposition formula for G/H can be constructed from solutions on G/K.

For simplicity of notation we assume that we are dealing with the global problem, i.e., on the level of the group, as opposed to the algebra, and the group action is on G/H.

If, on the other hand, K does contain a normal subgroup of G, then denote the largest such normal subgroup N. The action of G on G/K is not effective and the superposition law on G/H cannot be derived from one on G/K without further data. We will describe how to find the curve $g(t)$ in G from the curve $g_1(t)$ determined by the action on G/K, and a second curve $n(t)$ given by an equation on a homogeneous space $N/N\cap H$. Let $J\subset L$ be the ideal in L corresponding to the normal subgroup N. The right invariant vector field $\tilde{\xi}_t\in L$ defines an element of L/J, which we denote $\tilde{\xi}_t/J$. The Lie algebra of G/N can be identified with L/J. Let $g_1(t)$ project to the solution in G/N, that is $(dg_1/dt)g_1^{-1}=\tilde{\xi}_t/J$. To find the full solution in G we must find a curve $n(t)$ in N such that $g(t)=g_1(t)n(t)$ solves $(dg/dt)g^{-1}=\tilde{\xi}_t$. This is equivalent to the equation

$$\frac{dg_1}{dt}g_1^{-1}+g_1\frac{dn}{dt}n^{-1}g_1^{-1}=\tilde{\xi}_t.$$

We have $\tilde{\xi}_t(dg_1/dt)g_1^{-1}\in J$ so we must solve the differential equation

$$\frac{dn(t)}{dt}\cdot n^{-1}(t)=g_1^{-1}(t)\left(\tilde{\xi}_t-\frac{dg_1(t)}{dt}\cdot g_1^{-1}(t)\right)g_1(t). \tag{2.2}$$

This is equivalent to studying the differential equation along the leaf L_0 of the invariant foliation through the base point given by the time-dependent vector field $g_1(t)^{-1}*(\xi_t-\xi_{1t})$, where ξ_{1t} is the vector field on G/H generated by differentiating the action of $g_1(t)$. The leaf L_0 is diffeomorphic to K/H and we are studying the action of N. As before we can reduce first to the transitive case by picking an orbit of N such as $Nx_0=N/N\cap H$ and then reducing to the primitive case if $N\cap H$ or the appropriate isotropy group is not maximal. Since $\dim K/H<n$ this process will end after at most n steps.

For the purpose of linking up with the treatment in Refs. 3–7 we will discuss this reduction at the level of vector fields. Let

$$\xi_i(\mathbf{x})=\sum_j a_{ij}(\mathbf{x})\frac{\partial}{\partial x_j} \tag{2.3}$$

generate a Lie algebra L. Consider the subalgebra

$$L_0(\mathbf{x}_0)=\{\xi\,|\,\xi\subset L,\ \xi(\mathbf{x}_0)=0\} \tag{2.4}$$

(vector fields vanishing at the origin). Transitivity of the group action on $\mathbb{R}^n$ near $\mathbf{x}_0$ is equivalent to $\dim(L/L_0(x_0))=n$. If there is a more complicated orbit structure, then $d(\mathbf{x}_0)=\dim(L/L_0(\mathbf{x}_0))$ takes values between 0 and n. Let V_k be the subset of $\mathbb{R}$, where $d(\mathbf{x}_0)=k$ and assume V_k is a manifold for each k (some V_k may be empty). On V_k the vector fields in L form an integrable distribution, foliating V_k into submanifolds of dimension k. If R is a leaf of the foliation the solutions to the differential equation with initial data on R remain on R. By restricting attention to R we are carrying out the same reduction as restricting to one orbit of the group action. If the orbit is of maximal dimension then all nearby orbits are equivalent as homogeneous spaces and the restriction to one orbit is justified. On lower-dimensional orbits whose neighboring orbits are of different orbit type, the situation is not so clear.

We will assume that we have made the reduction and that we are looking at a submanifold R of $\mathbb{R}^n$ of dimension equal to $\dim(L/L_0(\mathbf{x}))$ for all $\mathbf{x}\in R$. By changing to adapted coordinates we identify R with $\mathbb{R}^k$ canonically imbedded in $\mathbb{R}^n$; to simplify the notation set $k=n$. Thus we have an algebra of vector fields on $\mathbb{R}^n$ such that the dimension of the subspace spanned by evaluation of the vector fields at $\mathbf{x}\in\mathbb{R}^n$, $\{\xi(\mathbf{x})|\xi\in L\}$ is n for all $\mathbf{x}$.

The action is called locally primitive if there does not exist an invariant foliation defined on a neighborhood U of a base point x_0, that is a set of k functions on U $\{f_1,\ldots,f_k\}$ with $\{df_{ix}\}$ linearly independent for all $x\in U$ such that for any $\xi\in L$ $\xi f_i=h_i(f_1,\ldots,f_k)$ for some h_i depending on ξ. If such $\{f_i\}$ do exist we can assume they are the first k functions of a coordinate chart $y^1,\ldots,y^k$ and we label the remaining coordinates $z^{k+1},\ldots,z^n$. The vector fields in L have the form

$$x=(y,z),\quad \xi_i(\mathbf{x})=\sum_{j=1}^{k}a_i^j(y)\frac{\partial}{\partial y^j}+\sum b_i^j(y,z)\frac{\partial}{\partial z^j}. \tag{2.5}$$

The differential equations

$$\frac{d\mathbf{x}}{dt}=\sum a^i(t)\xi_i(\mathbf{x})\cdot\mathbf{x} \tag{2.6}$$

can be solved for y independently of z to give $y(t)=g(t)\cdot y_0$. The solution $y(t)$ can then be substituted into the full equations so that

$$\frac{dz^j}{dt}=\sum_i a^i(t)b_i^j(g(t)y_0,z). \tag{2.7}$$

These equations will only be of the form that admits a superposition principle if we have

$$\xi_i(y(t),z)\cdot z=\sum b_i^j(t)\eta_j(z)\cdot z. \tag{2.8}$$

In other words, the z components of the vector fields ξ_i evaluated at different values of y must form a finite-dimensional Lie algebra.

As an example consider the three-dimensional (decomposable) Lie algebra given by vector fields on $\mathbb{R}^+\times\mathbb{R}$:

$$\xi_1=y\frac{\partial}{\partial y}-z\frac{\partial}{\partial z},\quad \xi_2=\frac{1}{y}\frac{\partial}{\partial z},\quad \xi_3=e^{yz}\frac{\partial}{\partial z}. \tag{2.9}$$

Even though the original algebra is three-dimensional, the z components generate an ∞-dimensional algebra. Therefore, the equations for z given by substituting the solution $y(t)$ do not admit a superposition law. It is, however, possible to modify the equations for z in such a manner that a superposi-

tion formula does exist for the modified equations and that these together with the equations for y are equivalent to the original system. The vector fields $\xi_i \in L$ for which $\xi_i \cdot y = 0$ form an ideal J in L and the problem can be reduced to one which involves only this ideal. To do this, first solve for the superposition law in y-space, that is determine a curve $g_1(t)$ in the group such that for any y_0, $g_1(t)y_0 = y(t)$ solves the equation in y:

$$\frac{dy}{dt} = \sum a^i(t)\xi_i \cdot y .$$

Now let $g_1(t)$ act on (y,z) space:

$$g_1(t)x_0 = g_1(t)(y_0,z_0) = (y(t),z(t)).$$

Then

$$\frac{dg_1}{dt}\cdot x_0 = \sum b_i(t)\xi_i(g(t)\cdot x_0) ,$$

with

$$\left[\sum a_i(t)\xi_i(g(t)x_0) - \sum b_i(t)\xi_i(g(t)x_0)\right] y = 0 .$$

Therefore,

$$\sum (a_i(t) - b_i(t))\,\xi_i \in J .$$

If $J = 0$ then $\Sigma(a_i(t) - b_i(t))\xi_i = 0$ which implies $a_i(t) \equiv b_i(t)$ by linear independence of ξ_i and the solution to the superposition law on y space solves the superposition law on x-space. Otherwise we must solve the equation

$$\frac{dz}{dt} = g_1(t)^{-1}{}_* \left(\sum (a_i(t) - b_i(t))\xi_i(g_1(t)z)z\right)$$

for a function $\tilde{z}(t)$. The solution to our original system is

$$(y(t),z(t)) = g_1(t)\cdot(y_0,\tilde{z}(t))$$

(g_* denotes the derivative action of G). The solution $\tilde{z}(t)$ is described by a superposition law based on the action of the group corresponding to the ideal J.

The proposition stated below summarizes these results.

Proposition: Given a differential equation on $\mathbb{R}^n$ which has a superposition law there is an associated finite-dimensional Lie subalgebra of the algebra of vector fields $L = \{\xi_i | i = 1,\ldots, p\}$ such that the equation can be written

$$\frac{d\mathbf{x}}{dt} = \sum a_i(t)\xi_i(\mathbf{x})\cdot\mathbf{x} .$$

Let G be a (local) Lie group with algebra L acting on a neighborhood U of the origin in $\mathbb{R}^n$. If the orbit structure is regular (admits a stratification into submanifolds) and if we consider only those superposition laws based on r solutions $x_1(t),\ldots,x_r(t)$ which lie in orbits of the same type then the most general such superposition law can be derived from a knowledge of superposition laws for equations arising from transitive primitive group actions. □

For primitive group actions our primary reference is the paper by Golubitsky.[12] He studies the globally primitive group actions and notes that the maximality of the isotropy algebra L_0 is sufficient, but not necessary. For locally primitive actions the maximality of L_0 is necessary.

Proposition: There is no invariant foliation of any neighborhood U of x_0 if and only if L_0 is maximal.

Proof: Suppose there exists an invariant foliation of some neighborhood U and let F_{x_0} be the leaf through x_0. The condition that the vector field ξ on U generated by an element $\xi \in g$ be tangent to F_{x_0} defines a subalgebra of L properly containing L_0. The closure under brackets is trivial and the fact that it properly contains L_0 is a consequence of the invariance of the foliation and transitivity. Conversely suppose L_0 is not maximal. Let K be a proper subalgebra of g properly containing L_0, $L \supset K \supset L_0$, $L \neq K \neq L_0$. Define

$$D(gx_0) = \left\{\frac{d}{dt}\, g(\exp t\eta)\cdot x_0 | \eta \in K\right\} \supset T_{gx_0} M .$$

$D(gx_0)$ is well defined for all g in a sufficiently small neighborhood N of the identity in G such that the intersection of N with the isotropy group is connected. This definition provides a G invariant integrable distribution, hence an invariant foliation of codimension equal to the codimension of K in L.

III. CLASSIFICATION OF PRIMITIVE TRANSITIVE CLASSICAL LIE ALGEBRAS AND OF THE SYSTEMS OF ODE'S WITH SUPERPOSITION PRINCIPLES

We have shown in the previous section that "indecomposable" systems of ordinary differential equations with superposition principles are obtained by constructing the vector fields $\xi(\mathbf{x})$ corresponding to the infinitesimal transitive action of a group G on a homogeneous space G/H, where $H \subset G$ is a maximal subgroup of G.

Let us now look at the corresponding pair of Lie algebras (L,L_0), and reproduce some definitions from various authors.

Definition 1: The pair (L,L_0) defines a transitive primitive Lie algebra if (1) L_0 does not contain a nonzero ideal of L; (2) L_0 is maximal in L.

Definition 2: The transitive primitive Lie algebra (L,L_0) is nonlinear if there exists a nonzero subalgebra $L_1 \subset L_0$ defined by

$$L_1 = \{\xi \in L_0 | [\xi, L\,] | \supseteq L_0\}.$$

Definition 3: The transitive primitive nonlinear Lie algebra (L,L_0) is irreducible if no subspace $M \subset L$ exists such that $[L_0,M] \subseteq M$, $M \neq L$, $M \neq L_0$.

According to the arguments of Sec. II we are now in the situation of having a transitive primitive Lie algebra (L,L_0), where L is some finite-dimensional Lie algebra realized by vector fields and

$$L_0 = \{\xi(\mathbf{x}) \in L \,|\, \xi(\mathbf{x}_0) = 0\} ,$$

i.e., L_0 is the subalgebra of fields vanishing at the origin $\mathbf{x}_0$.

We can now make use of the classification of primitive transitive finite Lie algebras due to Kobayashi and Nagano,[14,15] Ochiai,[11] and Golubitsky[12] to obtain a classification of systems of ODE's with superposition principles.

Theorem 1: (See Golubitsky.[12]) Assume L is not simple. Then either

(1) L is not semisimple and there is an abelian complement V to L_0 on which L_0 acts faithfully and irreducibly. This is the case of an affine group.

(2) L is semisimple. In this case there exists a simple Lie algebra K such that $L = K \oplus K$ and L_0 is the diagonally imbedded subalgebra (isomorphic to K). □

In case (1) the Lie algebra L is either aff$(n,\mathbb{C})$ or one of its subalgebras. The vector fields in natural coordinates are given by combinations of

$$\left\{x_i \frac{\partial}{\partial x_k}\right\} \text{ and } \left\{\frac{\partial}{\partial x_i}\right\}. \tag{3.1}$$

The equations are linear, for aff$(n,\mathbb{C})$ we have $L_0 \sim \mathrm{gl}(n,\mathbb{C})$ and the superposition formula is the linear one involving n linearly independent solutions. For other affine algebras more economical nonlinear superposition formulas (involving less than n solutions) may exist.

In case (2) if we choose K as a simple classical Lie algebra, the homogeneous space G/H is a surface in a Grassmann manifold of n-planes in $\mathbb{C}^{2n}$. The semisimple group G is a subgroup of a larger simple group acting on the Grassmannian. Let us consider each of the complex classical Lie algebras separately.

(a) $K = \mathrm{sl}(n,\mathbb{C})$. Consider first the action of a larger group, SL$(2n,\mathbb{C})$ on the Grassmannian $G_n(\mathbb{C}^{2n})$ of complex n-planes in $\mathbb{C}^{2n}$. Introduce homogeneous coordinates, namely the matrix elements of two complex matrices $X \in \mathbb{C}^{n\times n}$, $Y \in \mathbb{C}^{n\times n}$. The rank of $\binom{X}{Y}$ is n and two pairs of such matrices $\tilde{X},\tilde{Y}$ and (X,Y) characterize the same point in $G_n(\mathbb{C}^{2n})$ if they satisfy

$$\xi = \begin{pmatrix}\tilde{X}\\ \tilde{Y}\end{pmatrix} = \begin{pmatrix}Xg_0\\ Yg_0\end{pmatrix}, \quad g_0 \in \mathrm{GL}(n,\mathbb{C}). \tag{3.2}$$

The action of SL$(2n,\mathbb{C})$ on the homogeneous coordinates is linear,

$$\begin{pmatrix}X'\\ Y'\end{pmatrix} = g\begin{pmatrix}X\\ Y\end{pmatrix}, \quad g = \begin{pmatrix}G_{11} & G_{12}\\ G_{21} & G_{22}\end{pmatrix}, \quad \det g = 1. \tag{3.3}$$

The isotropy group of the origin $\binom{I}{I}$ is given by the constraint $G_{11} + G_{12} = G_{21} + G_{22}$. We also introduce affine coordinates on $G_n(\mathbb{C}^{2n})$, putting

$$W \equiv XY^{-1} \quad \text{if} \quad \det Y \neq 0. \tag{3.4}$$

The action of SL$(2n,\mathbb{C})$ is now a matrix fractional linear one

$$W' = (G_{11}W + G_{12})(G_{21}W + G_{22})^{-1}. \tag{3.5}$$

Restricting to the group SL$(n,\mathbb{C}) \otimes$ SL$(n,\mathbb{C})$ under consideration, we put

$$G_{12} = 0, \quad G_{21} = 0, \quad \det G_{11} = \det G_{22} = 1 \tag{3.6}$$

and the action (3.5) reduces to

$$W' = G_{11} W G_{22}^{-1}. \tag{3.7}$$

In view of (3.7) we have

$$\det W' = \det W.$$

In order to have a transitive action we must choose a fixed value of the determinant; we put

$$\det W = 1. \tag{3.8}$$

Instead of writing out the sl$(n,\mathbb{C}) \oplus$ sl$(n,\mathbb{C})$ vector fields we shall just give the corresponding ODE's. In homogeneous coordinates they are

$$\begin{pmatrix}\dot{X}\\ \dot{Y}\end{pmatrix} = \begin{pmatrix}C & 0\\ 0 & -B\end{pmatrix}\begin{pmatrix}X\\ Y\end{pmatrix}, \quad \begin{matrix}X,Y,C,B \in \mathbb{C}^{n\times n}\\ \mathrm{Tr}\, C = \mathrm{Tr}\, B = 0,\end{matrix} \tag{3.9}$$

and in affine coordinates, using (3.4) we obtain

$$\dot{W} = WB + CW, \quad \mathrm{Tr}\, B = \mathrm{Tr}\, C = 0. \tag{3.10}$$

We thus obtain a system of n^2 linear equations, subject to the nonlinear constraint (3.8). Using (3.8) to eliminate one of the variables, say the matrix element w_{nn}, we obtain a system of $n^2 - 1$ nonlinear ODE's with rational nonlinearities. Indeed, we have

$$w_{nn} = \frac{1 + (-1)^n(w_{1n}Y_{1n} - w_{2n}Y_{2n} + \cdots + (-1)^n w_{n-1\,n}Y_{n-1\,n})}{Y_{nn}}, \tag{3.11}$$

where Y_{ik} is the subdeterminant of W corresponding to the matrix element w_{ik}. Equations (3.10) reduce to

$$\begin{aligned}
\dot{w}_{ab} &= \sum_{\alpha=1}^{n} (w_{a\alpha}B_{\alpha b} + C_{a\alpha}w_{\alpha b}, \quad 1 \leqslant a,b \leqslant n-1,\\
\dot{w}_{nb} &= \sum_{\alpha=1}^{n-1} w_{n\alpha}B_{ab} + \sum_{\alpha=1}^{n} C_{n\alpha}w_{\alpha b} + B_{nb}\frac{1 + (-1)^n(w_{1n}Y_{1n} - w_{2n}Y_{2n} + \cdots + (-1)^n w_{n-1\,n}Y_{n-1\,n})}{Y_{nn}},\\
\dot{w}_{an} &= \sum_{\alpha=1}^{n} w_{a\alpha}B_{\alpha n} + \sum_{b=1}^{n-1} C_{ab}w_{bn} + C_{an}\frac{1 + (-1)^n(w_{1n}Y_n - w_{2n}Y_{2n} + \cdots + (-1)^n w_{n-1\,n}Y_{n-1\,n})}{Y_{nn}}.
\end{aligned} \tag{3.12}$$

In the special case of $n = 2$ Eq. (3.10) can be transformed into Riccati type equations. Indeed, then (3.11) reduces to

$$w_{22} = (1 + w_{12}w_{21})/w_{11} \quad (\text{for } w_{11} \neq 0) \tag{3.13}$$

Introducing new coordinates

$$x = 1/w_{11}, \quad y = w_{12}/w_{11}, \quad z = w_{21}/w_{11}, \tag{3.14}$$

we reduce (3.12) for $n = 2$ to

$$\begin{aligned}
\dot{x} &= -(b_{11} + c_{11})x - x(b_{21}y + c_{12}z),\\
\dot{y} &= b_{12} - 2b_{11}y + c_{12}x^2 - b_{21}y^2,\\
\dot{z} &= c_{21} - 2c_{11}z + b_{21}x^2 - c_{12}z^2.
\end{aligned} \tag{3.15}$$

(b) $K = \mathrm{o}(n,\mathbb{C})$. Consider again the action of a larger group, namely O$(2n,\mathbb{C})$ on the Grassmannian $G_n^0(\mathbb{C}^{2n})$ of isotropic (with respect to the orthogonal metric) n-planes in $\mathbb{C}^{2n}$. Introduce homogeneous coordinates as in (3.2) and impose the isotropy condition

$$\xi^T K \xi = 0, \quad K = \begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix} \tag{3.16}$$

for all $\xi \in G_n^0(\mathbb{C}^{2n})$,

$$X^T X - Y^T Y = 0. \tag{3.17}$$

This implies, together with the rank condition on (X,Y), that

det $X \neq 0$, det $Y \neq 0$ and we can hence globally introduce the affine coordinates (3.4). The group $O(2n,\mathbb{C})$ acts as in (3.5) and the group element g satisfies

$$gKg^T = K. \tag{3.18}$$

The $O(n,\mathbb{C}) \otimes O(n,\mathbb{C})$ subgroup is subject to (3.6) with

$$G_{11}^T G_{11} = I, \quad G_{22}^T G_{22} = I. \tag{3.19}$$

The ODE's corresponding to the infinitesimal action of $O(n,\mathbb{C}) \otimes O(n,\mathbb{C})$ can again be written in the form (3.9) with

$$C + C^T = 0, \quad B + B^T = 0. \tag{3.20}$$

In affine coordinates we again obtain the linear equations (3.10), satisfying (3.20). The isotropy condition (3.17) implies that W satisfies the quadratic constraint

$$WW^T = W^T W = I. \tag{3.21}$$

The constraint can be solved by means of the Cayley transform

$$W = (I + V)(I - V)^{-1}, \quad V = (W - I)(W + I)^{-1}, \tag{3.22}$$

where (3.21) implies

$$V = -V^T$$

and (3.10) is tranformed into

$$\dot{V} = \frac{B+C}{2} + V\frac{B-C}{2} - \frac{B-C}{2}V - V\frac{B+C}{2}V. \tag{3.23}$$

This is a special case of the $O(2n,\mathbb{C})$ Riccati equation to be discussed below.

(c) $K = sp(2n,\mathbb{C})$. Consider the transitive action of $Sp(4n,\mathbb{C})$ on the Grassmannian of symplecticly isotropic $2n$-planes $G_{2n}^s(\mathbb{C}^{4n})$. Introduce homogeneous coordinates as in (3.2) and impose the symplectic isotropy condition

$$\xi^T \tilde{K} \xi = 0, \quad \tilde{K} = \begin{pmatrix} K_0 & 0 \\ 0 & -K_0 \end{pmatrix}, \tag{3.24}$$

where $K_0 \in \mathbb{C}^{2n \times 2n}$ is an antisymmetric nondegenerate matrix, i.e.,

$$X^T K_0 X - Y^T K_0 Y = 0, \quad X, Y \in \mathbb{C}^{2n \times 2n}. \tag{3.25}$$

Similar to the case of $K = o(n,\mathbb{C})$ we obtain, in affine coordinates (3.4), the linear equations (3.10) with a quadratic constraint:

$$WK_0 W^T = K_0 \tag{3.26}$$

and B and C satisfying

$$CK_0 + K_0 C^T = 0, \quad BK_0 + K_0 B^T = 0. \tag{3.27}$$

Removing the constraint by the transform (3.22), we obtain a special case of the $Sp(4n,\mathbb{C})$ matrix Riccati equation, namely Eq. (3.23) with B and C satisfying (3.27) and V satisfying

$$V^T K_0 + K_0 V = 0. \tag{3.28}$$

To summarize: The linear transitive primitive Lie algebras are charcterized in Theorem 1, they lead to linear ODE's in case (1), to linear ODE's with polynomial constraints in case (2). Removing the constraints we obtain nonlinear equations. For $K = o(n,\mathbb{C})$, $sp(2n,\mathbb{C})$, or $sl(2,\mathbb{C})$ the constraints are quadratic and we obtain coupled Riccati equations. For $sl(n,\mathbb{C})$ with $n \geqslant 3$ the nth-order constraint leads to equations with rational nonlinearities.

Theorem 2: If L is simple and the action of L_0 on L/L_0 is irreducible then either (1) L_0 acts faithfully and is thus a maximal reductive subalgebra; or (2) L_0 does not act faithfully, then we have the class of algebras called "nonlinear primitive irreducible transitive Lie algebras" by Ochiai.[11] They can be written as a sum

$$L = g^{-1} \dot{+} g^0 \dot{+} g^1, \quad [g^i, g^j] \subset g^{i+j},$$
$$g^i = 0, \quad i \geqslant 2, \quad i \leqslant -2, \tag{3.29}$$

and the subalgebra $L_0 = g^0 \dot{+} g^1$ is a maximal parabolic subalgebra. □

Proof: (1) If L_0 has a faithful irreducible representation it is a reductive Lie algebra.[16] It is maximal since the pair (L, L_0) defines a primitive transitive Lie algebra.

(2) If L_0 does not act faithfully on L/L_0, then there exists a nonzero element $\xi \in L_0$ such that ξ acts as zero on L/L_0. Then $x \in L_1 \neq 0$, so the algebra is nonlinear in the Kobayashi–Nagano–Ochiai sense (see Definition 2 above). By assumption L_0 acts irreducibly (so $M = L$ in Definition 3). The conclusion (3.29) and the fact that L_0 is parabolic is proven in Ref. 14.

Theorem 3: (See Veisfeiler[13] and Golubitsky.[12]) If L is simple and the action of L_0 on L/L_0 has a nontrivial invariant subspace, then L can be written as

$$L = g^{-k} \dot{+} g^{-k+1} \dot{+} \cdots \dot{+} g_0 \dot{+} g_1 \dot{+} \cdots \dot{+} g^k$$

with $g^i = 0$ for $i < -k$ or $i > k$ and $[g^i, g^j] \subset g^{i+j}$, where $L_0 = \dot{+}_{i \geqslant 0} g^i$.

Hence, L_0 does not act faithfully on L/L_0 and L is nonlinear primitive in the sense of Ochiai. Once again L_0 is parabolic. □

Let us now relate these two theorems to the problem of ODE's with superposition principles.

Case (1) of Theorem 2 can be divided into two subcases. The maximal reductive subalgebra L_0 can be imbedded into the simple algebra L reducibly (leaving a nontrivial subspace in the considered representation space invariant), or irreducibly.

Let us first consider reducibly imbedded subalgebras $L_0 \subset L$ and the corresponding ODE's with superposition principles.

(a) $L = sl(n,\mathbb{C})$. Case (1) does not occur since a maximal subalgebra $L_0 \subset sl(n,\mathbb{C})$ leaving a k-dimensional vector space invariant ($1 \leqslant k \leqslant n-1$) will be parabolic (will contain the Borel subalgebra, i.e., the maximal solvable subalgebra). A complex parabolic subalgebra will always contain a nilpotent ideal and can hence not be reductive.

(b) $L = o(n,\mathbb{C})$. A maximal reductive subalgebra is obtained if we require that L_0 leave invariant a q-dimensional nondegenerate subspace (spanned by q nonisotropic mutually orthogonal vectors). We put

$$n = p + q, \quad n - 1 \geqslant p \geqslant q \geqslant 1,$$

and obtain

$$L_0 = o(p,\mathbb{C}) \oplus o(q,\mathbb{C}). \tag{3.30}$$

Let us now make use of the imbedding

$$SO(p+q,\mathbb{C})/SO(p,\mathbb{C}) \otimes SO(q,\mathbb{C})$$
$$\sim SL(p+q,\mathbb{C})/\mathrm{Aff}(p,q,\mathbb{C}), \tag{3.31}$$

where we denote $\mathrm{Aff}(p,q,\mathbb{C})$ the group of block triangular matrices

$$\begin{pmatrix} G_{11} & 0 \\ G_{21} & G_{22} \end{pmatrix}, \quad G_{11}\in\mathbb{C}^{p\times p}, \quad G_{22}\in\mathbb{C}^{q\times q}, \quad G_{21}\in\mathbb{C}^{q\times p}, \quad \det G_{11}\cdot\det G_{22}=1. \tag{3.32}$$

We realize an element of the algebra $\mathrm{sl}(p+q,\mathbb{C})$ as

$$\xi=\begin{pmatrix} C & A \\ -D & -B \end{pmatrix}, \quad C\in\mathbb{C}^{p\times p}, \quad B\in\mathbb{C}^{q\times q}, \quad A\in\mathbb{C}^{p\times q}, \quad D\in\mathbb{C}^{q\times p}, \quad \mathrm{Tr}\,C=\mathrm{Tr}\,B, \tag{3.33}$$

and let $\mathrm{SL}(p+q,\mathbb{C})$ act on the Grassmannian of q-planes in $\mathbb{C}^{p+q}$: $G_q(\mathbb{C}^{p+q})\sim\mathrm{SL}(p+q,\mathbb{C})/\mathrm{Aff}(p,q,\mathbb{C})$. Introducing affine coordinates W as in (3.4) we can write the ODE's corresponding to the $\mathrm{SL}(p+q,C)$ infinitesimal action as

$$\dot{W}=A+WB+CW+WDW, \quad W\in\mathbb{C}^{p\times q}, \tag{3.34}$$

A,B,C,D as in (3.33).

We restrict to $\mathrm{o}(p+q,\mathbb{C})$, $p+q=n$, by imposing

$$\xi K+K\xi^{T}=0, \quad K=\begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix} \tag{3.35}$$

and obtain a special case of the rectangular matrix Riccati equation (MRE):

$$\dot{W}=A+WB+CW-WA^{T}W, \quad B=-B^{T}, \quad C=-C^{T}. \tag{3.36}$$

Note that the action of $\mathrm{O}(n,\mathbb{C})$ is not transitive. In homogeneous coordinates we have

$$\begin{pmatrix} X' \\ Y' \end{pmatrix}=\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}, \quad gKg^{T}=K. \tag{3.37}$$

The $\mathrm{O}(n,\mathbb{C})$ orbit of maximal dimension is given by the condition $\mathrm{rank}\,(X^{T}X-Y^{T}Y)=n$, which can be represented by $\binom{X}{Y}$ satisfying

$$X^{T}X-Y^{T}Y=I. \tag{3.38}$$

This imposes no symmetry condition on $W=XY^{-1}$. As the origin we can choose $(X,Y)=(0,iI)$, i.e., $W=0$. The isotropy group of the origin is $\mathrm{O}(p,\mathbb{C})\otimes\mathrm{O}(q,\mathbb{C})$, as it should be.

If we request that a maximal subalgebra $L_0\subset\mathrm{o}(n,\mathbb{C})$ leave a degenerate space invariant (containing 1 or more isotropic vectors in an orthogonal basis) then L_0 will be a maximal parabolic subalgebra and these are treated below.

(c) $L=\mathrm{sp}(2n,\mathbb{C})$. The situation is very similar to that of $\mathrm{o}(n,\mathbb{C})$. The only way to obtain a reducibly imbedded maximal reductive subalgebra is to require that $L_0\subset L$ leave invariant a $2q$-dimensional nondegenerate vector space. We then obtain

$$L_0=\mathrm{sp}(2p,\mathbb{C})\oplus\mathrm{sp}(2q,\mathbb{C}), \quad p+q=n. \tag{3.39}$$

To realize the algebra in terms of vector fields and to obtain the corresponding equations, we again imbed L_0 into $\mathrm{sl}(2p+2q,\mathbb{C})$ and use the imbedding

$$\mathrm{Sp}(2p+2q,\mathbb{C})/\mathrm{Sp}(2p,\mathbb{C})\otimes\mathrm{Sp}(2q,\mathbb{C}) \sim\mathrm{SL}(2p+2q,\mathbb{C})/\mathrm{Aff}(2p,2q,C). \tag{3.40}$$

Letting $\mathrm{SL}(2p+2q,\mathbb{C})$ act on the Grassmannian $G_{2q}(\mathbb{C}^{2p+2q})$ we again obtain the Eqs. (3.34) [replacing (p,q) by $(2p,2q)$]. Restricting to $\mathrm{sp}(2p+2q,\mathbb{C})$ we impose

$$\xi K+K\xi^{T}=0, \quad K=\begin{pmatrix} K_{2p} & \\ & -K_{2q} \end{pmatrix}, \quad K_{2j}=\begin{pmatrix} 0 & I_j \\ -I_j & 0 \end{pmatrix}, \quad j=p,q \tag{3.41}$$

and obtain a special case of the rectangular matrix Riccati equation

$$\dot{W}=A+WB+CW+WK_{2q}A^{T}K_{2p}W, \quad CK_{2p}+K_{2p}C^{T}=0, \quad BK_{2q}+K_{2q}B^{T}=0, \quad W\in\mathbb{C}^{2p\times 2q}. \tag{3.42}$$

This completes the treatment of all ODE's corresponding to reducibly imbedded maximal reductive subalgebras of the complex classical Lie algebras. We always obtain special cases of matrix Riccati equations.

The case of irreducibly imbedded reductive subalgebras of the classical Lie algebras is less uniform and more difficult to treat from the point of view of the obtained differential equations. The corresponding group–subgroup classification has been given by Dynkin.[17,18] The homogeneous spaces G/H obtained in this case include symmetric spaces, but also other classes of spaces, in particular, the isotropy irreducible homogeneous spaces studied by Wolf.[19]

We shall here restrict ourselves to two examples and postpone a detailed treatment for a future article.

(a) $\mathrm{SL}(2n,\mathbb{C})/\mathrm{Sp}(2n,\mathbb{C})$. Kobayashi and Nagano[15] have established the diffeomorphism $\mathrm{U}(2n)/\mathrm{Sp}(2n)\sim\mathrm{SO}^{*}(4n)/P$, where P is a maximal parabolic subgroup of $\mathrm{SO}^{*}(4n)$ [No. 5 in their list, which however contains a misprint: $\mathrm{SU}^{*}(4n)$ instead of $\mathrm{SO}^{*}(4n)$]. We complexify this relationship and first extend $\mathrm{SL}(2n,\mathbb{C})$ to $\mathrm{GL}(2n,\mathbb{C})$, then realize it as a subgroup of $\mathrm{SO}(4n,\mathbb{C})$.

Realize $\mathrm{O}(4n,\mathbb{C})$ as the algebra of matrices

$$\xi=\begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad \xi K+K\xi^{T}=0, \quad K=\begin{pmatrix} 0 & I_{2n} \\ I_{2n} & 0 \end{pmatrix}, \tag{3.43}$$

i.e., $D=-A^{T}$, $C=-C^{T}$, $B=-B^{T}$, and construct the Grassmannian of null planes $G^{0}_{2n}(\mathbb{C}^{4n})$. Take

$$\begin{pmatrix} X \\ Y \end{pmatrix}=\begin{pmatrix} 0 \\ I \end{pmatrix}, \quad X,Y\in\mathbb{C}^{2n\times 2n} \tag{3.44}$$

to be the homogeneous coordinates of the origin. The subalgebra of $\mathrm{O}(4n,\mathbb{C})$ leaving the origin invariant is the maximal parabolic subalgebra P. We identify $\mathrm{gl}(2n,\mathbb{C})$ with the subalgebra

$$\rho(A)=\begin{pmatrix} A & AJ+JA^{T} \\ 0 & -A^{T} \end{pmatrix}, \quad J=\begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{pmatrix}. \tag{3.45}$$

The subalgebra of $\mathrm{gl}(2n,\mathbb{C})$ leaving the origin invariant is $\mathrm{gl}(2n,\mathbb{C})\cap P$; it is represented by matrices $\rho(A)$ satisfying

$$AJ + JA^T = 0,$$

i.e., it is isomorphic to sp(2n,C), as required.

The action of the group GL(2n,C) on the origin is given by

$$\begin{pmatrix} g & gJ - J(g^T)^{-1} \\ 0 & (g^T)^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ I \end{pmatrix} \sim \begin{pmatrix} X \\ I \end{pmatrix} \tag{3.46}$$

with

$$X = gJg^T - J, \quad X^T = -X \subset \mathbb{C}^{2n\times 2n}. \tag{3.47}$$

Restricting to SL(2n,C) we have $\det g = 1$. The orbit of the origin under SL(2n,C) is thus given by

$$\begin{pmatrix} X \\ I \end{pmatrix}, \quad X + X^T = 0, \det(X + J) = 1. \tag{3.48}$$

Realizing o(4n,C) on $G^0_{2n}(\mathbb{C}^{4n})$ in affine coordinates and restricting to sl(2n,C) as in (3.45) we obtain a system of linear inhomogeneous ODE's:

$$\dot{W} = AJ + JA^T + AW + WA^T, \quad W \in \mathbb{C}^{2n\times 2n},$$
$$W + W^T = 0 \tag{3.49}$$

with the polynomial constraint

$$\det(W + J) = 1. \tag{3.50}$$

Using (3.50) to eliminate one of the matrix elements of W, say $w_{n-1\,n}$, we obtain a system of nonlinear ODE's with rational nonlinearities.

In the special case of $n = 2$ (3.50) is quadratic and in appropriate coordinates we obtain a system of coupled Riccati equations. Indeed, consider the case of SL(4,C)/SP(4,C). Put

$$W = \begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix} \tag{3.51}$$

in (3.49). Use (3.50) to eliminate f and introduce new variables

$$x = \frac{1}{1+a}, \quad y = \frac{b}{1+a}, \quad z = \frac{c}{1+a},$$
$$t = \frac{d}{1+a}, \quad u = \frac{e}{1+a}. \tag{3.52}$$

Equation (3.49) reduces to

$$\begin{aligned}
\dot{x} &= -(A_{11} + A_{22})x, \\
\dot{y} &= A_{32} + (A_{33} - A_{44})y + A_{34}z + A_{12}t - A_{14}(x^2 - zt) \\
&\quad + (-A_{23}y - A_{24}z + A_{13}t + A_{14}u)x, \\
\dot{z} &= A_{42} + A_{43}y + (-A_{22} + A_{44})z + A_{12}u + A_{13}(x^2 + yu) \\
&\quad + (A_{13}t - A_{23}y + A_{24}z)y, \\
\dot{t} &= -A_{31} + A_{21}y + (-A_{11} + A_{33})t + A_{34}u - A_{24}(x^2 + yu) \\
&\quad + (A_{13}t + A_{14}u - A_{23}y)t, \\
\dot{u} &= -A_{41} + A_{21}z + A_{43}t + (-A_{11} + A_{44})u + A_{23}(x^2 - 2t) \\
&\quad + (A_{13}t + A_{14}u - A_{24}z)u.
\end{aligned} \tag{3.53}$$

(b) SL(n,C)/SO(n,C). This case can be treated quite similarly to the previous one. SL(n,C) is extended to GL(n,C) and then treated as a subgroup of SP(2n,C) acting on SP(2n,C)/P, where P is the appropriate maximal parabolic subgroup of SP(2n,C). Again we get linear equations with a polynomial constraint, leading in general to rational nonlinearities. We shall not go into the details here.

In case (2) of Theorem 2 and in Theorem 3 the algebra L is simple, the subalgebra L_0 is parabolic (represented irreducibly in the first case reducibly in the second).

We shall now run through all classical complex Lie algebras and their maximal parabolic subalgebras and obtain the corresponding ODE's. It turns out that if L_0 acts irreducibly, we always get matrix Riccati equations. If L_0 acts reducibly we again obtain matrix Riccati equations, sometimes with additional quadratic constraints, that lead to cubic and quartic nonlinearities.

(a) *The group* SL(N,C). We partition N into $N = n + k$, $1 \leqslant n,k \leqslant N - 1$, and introduce the Grassmannian

$$G_k(\mathbb{C}^{n+k}) \sim \mathrm{SL}(n + k,\mathbb{C})/\mathrm{Aff}(n,k,\mathbb{C}), \tag{3.54}$$

where

$$H = \mathrm{Aff}(n,k,C) = \begin{pmatrix} G_{11} & 0 \\ G_{21} & G_{22} \end{pmatrix}, \quad \begin{matrix} G_{11}\in\mathbb{C}^{n\times n}, & G_{22}\in\mathbb{C}^{k\times k} \\ G_{21}\in\mathbb{C}^{k\times n}, & \end{matrix} \tag{3.55}$$

$$\det G_{11} \cdot \det G_{22} = 1.$$

The corresponding ODE's in affine coordinates are simply the most general rectangular MRE

$$\dot{W} = A + WB + CW + WDW, \quad \mathrm{Tr}\, B = \mathrm{Tr}\, C,$$
$$W,A\in C^{n\times k}, \quad B\in\mathbb{C}^{k\times k}, \quad C\in\mathbb{C}^{n\times n}, \quad D\in\mathbb{C}^{k\times n}. \tag{3.56}$$

This was treated in Ref. 6; for $n = k \geqslant 2$ the superposition formula involves precisely five particular generically chosen solutions.

Letting n and k run through all allowed values we obtain all maximal parabolic subalgebras H and MRE's of all dimensions. The action of H (or L_0) is always irreducible, so we are in the situation covered by Case (2) of Theorem 2.

(b) *The group* O(N,C). Let us realize o(N,C) as the algebra matrices $X\in\mathbb{C}^{N\times N}$ satisfying

$$XJ + JX^T = 0, \quad J = \begin{pmatrix} 0 & 0 & I_k \\ 0 & I_{N-2k} & 0 \\ I_k & 0 & 0 \end{pmatrix},$$
$$k = 1,\ldots,[N/2]. \tag{3.57}$$

The group O(N,C) will be represented by matrices G satisfying $GJG^T = J$ and a maximal parabolic subgroup H is obtained by requiring that an isotropic subspace of the representation space be left invariant. Choosing the invariant space to be

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix}, \quad f_1 f_3 \in \mathbb{C}^{k\times 1}, \quad f_2 \in \mathbb{C}^{N-2k)\times 1}$$

we find

$$H = \left\{ \begin{pmatrix} G_{11} & 0 & 0 \\ G_{21} & G_{22} & 0 \\ G_{31} & G_{32} & G_{33} \end{pmatrix} \right\}, \quad HJH^T = J, \tag{3.58}$$

where G_{ik} are complex rectangular matrices of the appropriate dimensions. Let us now introduce homogeneous coordinates on the factor space G/H, putting

$$U=\begin{pmatrix}U_1\\U_2\\U_3\end{pmatrix},\quad U_1,U_3\in\mathbb{C}^{k\times k},\quad U_2\in\mathbb{C}^{(N-2k)\times k},$$
$$U^TJU=0. \tag{3.59}$$

As usual, to get rid of the redundancy in the characterization of a point on G/H we introduce affine coordinates

$$Z=\begin{pmatrix}Z_1\\Z_2\end{pmatrix}=\begin{pmatrix}U_1U_3^{-1}\\U_2U_3^{-1}\end{pmatrix}\quad\text{for }\det U_3\neq 0. \tag{3.60}$$

The isotropy condition in (3.59) implies that the affine coordinates satisfy

$$Z_1+Z_1^T=-Z_2^TZ_2. \tag{3.61}$$

In homogeneous coordinates we can write the ODE's corresponding to the infinitesimal action of $O(N,\mathbb{C})$ on $O(N,\mathbb{C})/H$ as

$$\begin{pmatrix}\dot U_1\\\dot U_2\\\dot U_3\end{pmatrix}=\begin{pmatrix}A&B&C\\D&E&-B^T\\F&-D^T&-A^T\end{pmatrix}\begin{pmatrix}U_1\\U_2\\U_3\end{pmatrix},\quad\begin{matrix}C=-C^T\\E=-E^T\\F=-F^T\end{matrix} \tag{3.62}$$

Using (3.60) we rewrite these equations in affine coordinates as

$$\dot Z_1=C+AZ_1+Z_1A^T+BZ_2-Z_1FZ_1+Z_1D^TZ_1, \tag{3.63}$$
$$\dot Z_2=-B^T+DZ_1+EZ_2+Z_2A^T-Z_2FZ_1+Z_2D^TZ_2,$$

i.e., a system of coupled Riccati equations with the additional quadratic constraint (3.61). To get rid of the constraint we split Z_1 into its symmetric and antisymmetric part

$$Z_1=Z_{1S}+Z_{1A}=\frac{Z_1+Z_1^T}{2}+\frac{Z_1-Z_1^T}{2}$$

and eliminate Z_{1S} from the equations using (3.61). Finally we obtain the following system of ODE's:

$$\begin{aligned}\dot Z_{1A}&=C+AZ_{1A}+Z_{1A}A^T+\tfrac12(BZ_2-Z_2^TB^T)-Z_{1A}FZ_{1A}\\&\quad+\tfrac12(Z_{1A}D^TZ_2+Z_2^TDZ_{1A})\\&\quad+\tfrac14Z_2^T(DZ_2^T-Z_2D^T)Z_2-\tfrac14Z_2^TZ_2FZ_2^TZ_2,\end{aligned} \tag{3.64}$$
$$\begin{aligned}\dot Z_2&=-B^T+DZ_{1A}+EZ_2+Z_2A^T-\tfrac12DZ_2^TZ_2\\&\quad+Z_2D^TZ_2-Z_2FZ_{1A}+\tfrac12Z_2FZ_2^TZ_2.\end{aligned}$$

The matrices $A,\dots,F$ are all given functions of the independent variable t, the dependent variables are the matrix elements of Z_{1A} and Z_2. The equations obtained in general contain quartic and cubic terms, attached to the quadratic or linear ones. In special cases Eqs. (3.64) reduce to quadratic ones. Let us consider these cases.

(1) $k=1$. Then $C=0, F=0, Z_{1A}=0$ (because of antisymmetry). The first of Eqs. (3.64) drops out, the second reduces to

$$\dot Z_2=-B^T+EZ_2+Z_2A-\tfrac12DZ_2^TZ_2+Z_2D^TZ_2. \tag{3.65}$$

This is a complexification of the conformal Riccati equations studied earlier.[5]

(2) $k=N/2$, for N even. Then $B=E=D=Z_2=0$ and (3.64) reduces to

$$\dot Z_{1A}=C+AZ_1+Z_{1A}A^T-Z_{1A}FZ_{1A},\quad C=-C^T,\quad F=-F^T. \tag{3.66}$$

Equations (3.66) can appropriately be called "orthogonal matrix Riccati equations."

(3) $k=(N-1)/2$ for $N\gtrsim 5$ odd. In this case B is a column, D is a row, $E=0$, Z_2 is a row. Equations (3.64) again reduce to Riccati type equations:

$$\begin{aligned}\dot Z_{1A}&=C+AZ_{1A}+Z_{1A}A^T+\tfrac12(BZ_2-Z_2^TB^T)\\&\quad-Z_{1A}FZ_{1A}+\tfrac12(Z_{1A}D^TZ_2+Z_2^TDZ_{1A}),\end{aligned} \tag{3.67}$$
$$\begin{aligned}\dot Z_2&=-B^T+DZ_{1A}+Z_2A^T-\tfrac12DZ_2^TZ_2\\&\quad+Z_2D^TZ_2-Z_2FZ_{1A}.\end{aligned}$$

Returning to the question of reducibility of the primitive transitive Lie algebra (L,L_0), we have

$$L=O(N,C)=\left\{\begin{pmatrix}A&B&C\\D&E&-B^T\\F&-D^T&-A^T\end{pmatrix}\right\}\quad\begin{matrix}C=-C^T,\\E=-E^T,\\F=-F^T,\end{matrix}$$
$$L_0=\left\{\begin{pmatrix}A&0&0\\D&E&0\\F&-D^T&-A^T\end{pmatrix}\right\}, \tag{3.68}$$
$$L_1=\left\{\begin{pmatrix}0&0&0\\0&0&0\\F&0&0\end{pmatrix}\right\},$$
$$M=\left\{\begin{pmatrix}A&B&0\\D&E&-B^T\\F&-D^T&-A^T\end{pmatrix}\right\}$$

(see Definitions 2 and 3 for L_1 and M).

For $k=1$ we have $M=L$, for $k=N/2$ (N even) we have $M=L_0$. In these two cases L_0 acts irreducibly, the primitive transitive nonlinear Lie algebra is irreducible, i.e., we have case (2) of Theorem 2. For all other values of k [including $k=(N-1)/2$ for N odd] the nonlinear algebra (L,L_0) is reducible and Theorem 3 applies. The equations are in general quartic; for $k=(N-1)/2$ the cubic and quartic terms happen to drop out.

(c) The group $Sp(2N,\mathbb{C})$. Realize $sp(2N,\mathbb{C})$ as the algebra of matrices $X\in\mathbb{C}^{N\times N}$ satisfying

$$XK_{\lambda\mu}+K_{\lambda\mu}X^T=0,\quad K_{\lambda\mu}=\begin{pmatrix}0&0&I_\lambda\\0&K&0\\-I_\lambda&0&0\end{pmatrix},$$
$$K=\begin{pmatrix}0&I_\mu\\-I_\mu&0\end{pmatrix},\quad\lambda+\mu=N. \tag{3.69}$$

The maximal parabolic subalgebra L_0 will be characterized by the fact that it leaves an isotropic λ-dimensional subspace of the representation space invariant:

$$L=\left\{\begin{pmatrix}A&B&C\\D&E&KB^T\\F&-D^TK&-A^T\end{pmatrix}\right\}, \tag{3.70}$$

$$L_0=\left\{\begin{pmatrix} A & 0 & 0 \\ D & E & 0 \\ F & -D^T K & -A^T \end{pmatrix}\right\},$$

$$A,C,F\in\mathbb{C}^{\lambda\times\lambda},\quad E\in\mathbb{C}^{2\mu\times 2\mu},\quad D\in\mathbb{C}^{2\mu\times\lambda},\quad B\in\mathbb{C}^{\lambda\times 2\mu},$$

$$C=C^T,\quad F=F^T,\quad E=KE^TK. \tag{3.71}$$

The homogeneous space G/H can be identified with the Grassmannian of isotropic λ-planes; in homogeneous coordinates:

$$U=\begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix},\quad U_1,U_3\in\mathbb{C}^{\lambda\times\lambda},\quad U_2\in\mathbb{C}^{2\mu\times\lambda},\quad U^TK_{\lambda\mu}U=0. \tag{3.72}$$

Affine coordinates are introduced as in (3.60) and satisfy

$$Z_1-Z_1^T=Z_2^TKZ_2,\quad Z_1\in\mathbb{C}^{\lambda\times\lambda},\quad Z_2\in\mathbb{C}^{2\mu\times\lambda}. \tag{3.73}$$

The ODE's corresponding to the symplectic action in affine coordinates are

$$\dot{Z}_1=C+AZ_1+Z_1A^T+BZ_2-Z_1FZ_1+Z_1D^TKZ_2,$$
$$\dot{Z}_2=KB^T+DZ_1+EZ_2+Z_2A^T-Z_2FZ_1+Z_2D^TKZ_2. \tag{3.74}$$

We thus again obtain a system of coupled Riccati equations with an additional quadratic constraint (3.73). Putting

$$Z_{1S}=\tfrac{1}{2}(Z_1+Z_1^T) \tag{3.75}$$

and eliminating the antisymmetric part of Z_1 from (3.74) with the help of (3.73) we obtain a system of equations with, in general, up to quartic nonlinearities

$$\dot{Z}_{1S}=C+AZ_{1S}+Z_{1S}A^T+\tfrac{1}{2}(BZ_2+Z_2^TB^T)+\tfrac{1}{2}(Z_{1S}D^TKZ_2-Z_2^TKDZ_{1S})+\tfrac{1}{4}Z_2^TK(Z_2D^T+DZ_2^T)KZ_2-Z_{1S}FZ_{1S}-\tfrac{1}{4}Z_2^TKZ_2FZ_2^TKZ_2, \tag{3.76}$$

$$\dot{Z}_2=KB^T+DZ_{1S}+EZ_2+Z_2A^T+Z_2D^TKZ_2+\tfrac{1}{2}DZ_2^TKZ_2-Z_2FZ_{1S}-\tfrac{1}{2}Z_2FZ_2^TKZ_2,$$

The nonlinearities become quadratic in two special cases.

(1) $\mu=0$. Then $B=D=E=Z_2=0$ and we obtain the symplectic matrix Riccati equation[6] that is of special interest, e.g., in control theory[20]:

$$\dot{Z}_{1S}=C+AZ_{1S}+Z_{1S}A^T-Z_{1S}FZ_{1S},\quad C=C^T,\quad F=F^T. \tag{3.77}$$

(2) $\lambda=1$. In this case we rewrite $X\in L$ as

$$X=\begin{pmatrix} \alpha & b^T & c^T & \gamma \\ d & E_{11} & E_{12} & c \\ e & E_{21} & -E_{11}^T & -b \\ \delta & e^T & -d^T & -\alpha \end{pmatrix}, \tag{3.78}$$

$$E_{12}=E_{12}^T,\quad E_{21}=E_{21}^T,\quad E_{ab}\in\mathbb{C}^{\mu\times\mu},$$
$$b,c,d,e\in\mathbb{C}^{\mu\times 1},\quad \alpha,\delta,\gamma\in\mathbb{C}.$$

We also put

$$Z_{1S}=z,\quad Z_2=\begin{pmatrix} x \\ y \end{pmatrix},\quad z\in\mathbb{C},\quad x,y\in\mathbb{C}^{\mu\times 1}. \tag{3.79}$$

Equations (3.76) again reduce to coupled Riccati equations; in vector notation we have

$$\dot{z}=\gamma+2\alpha z+(\mathbf{b},\mathbf{x})+(\mathbf{c},\mathbf{y})+[(\mathbf{d},\mathbf{y})-(\mathbf{e},\mathbf{x})]z-\delta z^2,$$
$$\mathbf{x}=\mathbf{c}+z\mathbf{d}+E_{11}\mathbf{x}+E_{12}\mathbf{y}+\alpha\mathbf{x}+\mathbf{x}[(\mathbf{d},\mathbf{y})-(\mathbf{e},\mathbf{x})]-\delta z\mathbf{x}, \tag{3.80}$$
$$\dot{\mathbf{y}}=-\mathbf{b}+z\mathbf{e}+E_{21}\mathbf{x}-E_{11}^T\mathbf{y}+\alpha\mathbf{y}+\mathbf{y}[(\mathbf{d},\mathbf{y})-(\mathbf{e},\mathbf{x})]-\delta z\mathbf{y}.$$

The case $\mu=0$ corresponds to an irreducible nonlinear transitive primitive Lie algebra, and hence to case (2) of Theorem 2. All other cases (including $\lambda=1$ when the equations happen to be quadratic) correspond to the reducible case and hence to Theorem 3. Indeed, for $\mu\geqslant 1$ the space

$$M=\left\{\begin{pmatrix} A & B & 0 \\ D & E & KB^T \\ F & -D^TK & -A^T \end{pmatrix}\right\}$$

in invariant under L_0: $[L_0,M]\subset M$ and we have $M\neq L_0$, $M\neq L$. For $\mu=0$ we have $M=L_0$.

IV. CONCLUSIONS

It follows directly from the classical results[3] of Sophus Lie, that all systems of n ODE's of the type (1.1) with a superposition formula (1.2) are obtained from finite-dimensional subalgebras L of the algebra of vector fields on $\mathbb{R}^n$ (or $\mathbb{C}^n$). Such a direct approach is however both extremely difficult and unnecessary. We have shown that for purposes of studying superposition laws, it is reasonable to restrict the problem to a search for indecomposable systems of ODE's. These are equations from which it is not possible to decouple, by a change of dependent variables, a subset of equations involving a smaller number of variables, having a superposition formula of their own.

The restriction to the indecomposable case leads to the requirement that the finite-dimensional algebra L should correspond to the infinitesimal transitive action of a Lie group G on a homogeneous space G/H for $H\subset G$. Furthermore, it is sufficient to consider group–subgroup pairs, for which the corresponding Lie algebra–subalgebra pair (L,L_0), with $L_0\subset L$, defines a transitive primitive Lie algebra. The finite-dimensional transitive primitive Lie algebras have been classified by differential geometers.[9–15]

The subalgebra L_0 of vector fields vanishing at the origin must be maximal in L and cannot contain an ideal of L. The primitive transitive Lie algebras and the associated ODE's are listed below:

(1) L is an affine algebra acting on the abelian complement V of L_0. The equations are linear inhomogeneous ODE's.

(2) L is the direct sum of two simple Lie algebras: $L=K\oplus K$. The ODE's are coupled Riccati equations or equations with rational nonlinearities.

(3) L is simple, L_0 is a maximal reductive subalgebra. The general form of the equations has not yet been determined. However, if L_0 is imbedded reducibly in L, the equations are again coupled Riccati equations.

(4) L is simple, L_0 a maximal parabolic subalgebra. If L_0 acts irreducibly on L/L_0 we obtain various types of matrix

Riccati equations. If L_0 acts reducibly on L/L_0 the equations have polynomial nonlinearities. If L is a complex classical group, the nonlinearities are up to fourth order. In certain special cases we again obtain coupled Riccati equations (the higher-order terms drop out).

It is worth mentioning that all the obtained nonlinear ODE's with superposition principles can be interpreted as either linear equations, or coupled Riccati equations, with additional nonlinear constraints on the dependent variables.

Many questions concerning ODE's with superposition principles remain open. A detailed case-by-case treatment of the vector fields and equations corresponding to L simple and L_0 maximal reductive is forthcoming.

In this article we concentrated on the classical complex Lie algebras. The situation becomes much richer when the complex and real Cartan exceptional Lie algebras, as well as the real classical Lie algebras are also considered.

The classification of ODE's with superposition principles was performed up to arbitrary coordinate changes, i.e., up to arbitrary transformations of the dependent variables. For instance, the scalar Riccati equation $\dot{x} = a(t) + b(t)x + c(t)x^2$ represents the class of equations obtained by putting $x = \phi(y)$, where ϕ is an arbitrary single-valued differentiable function of a new dependent variable y. It is not excluded that some of the equations that we identified as having rational nonlinearities, can be transformed into equations with polynomial nonlinearities.

Equations with quadratic nonlinearities (coupled Riccati equations) are of particular interest in many applications. We have obtained many classes of such equations, but we have still not been able to characterize directly all classes of Riccati equations that admit superposition principles.

A separate problem is that of finding the actual superposition formulas and determining the number of particular solutions needed to form a "fundamental set of solutions." This has so far only been done for various types of Riccati equations.[4–7,20] The superposition formulas, in addition to providing insight into the properties of the solution space of solutions and reducing the problem of finding the general solution, to that of finding m particular solutions, also provide efficient numerical methods for solving the corresponding ODE's.[21]

ACKNOWLEDGMENTS

We thank J. Beckers and J. Harnad for helpful discussions at the early stage of this program. Work was supported in part by the Natural Sciences and Engineering Research Council of Canada and the "Fonds FCAC pour l'aide et le soutien à la recherche du Gouvernement du Québec."

[1] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[2] J. Harnad, Y. Saint-Aubin, and S. Shnider, Commun. Math. Phys. **93**, 33 (1984).
[3] S. Lie and G. Scheffers, *Vorlesungen über Continuierlichen Gruppen mit geometrischen und anderen Anwendungen* (Teubner, Leipzig, 1893).
[4] R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980).
[5] R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981); Physica D **4**, 164 (1982).
[6] J. Harnad, P. Winternitz, and R. L. Anderson, J. Math. Phys. **24**, 1062 (1983).
[7] S. Shnider and P. Winternitz, Lett. Math. Phys. **8**, 69 (1984).
[8] S. Lie and F. Engel, *Theorie der Transformationsgruppen* (Teubner, Leipzig, 1893), Vol. 3.
[9] V. W. Guillemin and S. Sternberg, Bull. Am. Math. Soc. **70**, 16 (1964).
[10] M. Demazure, in Seminaire Bourbaki, 19e année, 1966/67, 326-01.
[11] T. Ochiai, Trans. Am. Math. Soc. **124**, 313 (1966).
[12] M. Golubitsky, J. Differential Geometry **7**, 175 (1972).
[13] B. Yu. Veisfeiler, Funkts. Anal. i Prilozhen **2**, 94 (1968) [Functional Anal. Appl. **2**, 88 (1968)].
[14] S. Kobayashi and T. Nagano, J. Math. Mech. **13**, 875 (1964).
[15] S. Kobayashi and T. Nagano, J. Math. Mech. **14**, 513 (1965).
[16] N. Jacobson, *Lie Algebras* (Dover, New York, 1962).
[17] E. B. Dynkin, Trudy Moskov. Mat. Obšč. **1**, 39 (1952) [Am. Math. Soc. Transl. (Ser. 2) **6**, 245 (1957)].
[18] E. B. Dynkin, Math. Sb. **72**, (N.S. 30), 349 (1952) [Am. Math. Soc. Transl. (Ser. 2), **6**, 111 (1957)].
[19] J. A. Wolf, Acta Math. **120**, 59 (1968).
[20] M. Sorine and P. Winternitz, "Superposition Formulas for Matrix Riccati Equations Associated with Classical Lie Groups," IEEE Trans. Autom. Control (to be published).
[21] D. Rand and P. Winternitz, "Nonlinear Superposition Principles: a New Numerical Method for Solving Matrix Riccati Equations," Comp. Phys. Comm. (to be published).

A new set of Euler angles for the generalized Lorentz group

Ansaruddin Syed
Department of Mathematics, University of Karachi, Karachi, Pakistan

(Received 26 September 1983; accepted for publication 30 March 1984)

A new set of Euler angles for the generalized Lorentz group $O^+(p,n-p)$ are defined, which turn out to be much simpler than the ones defined in a couple of earlier papers, and have the useful property that each factor in the factorization of a general element itself belongs to the same group.

PACS numbers: 02.20. + b

I. INTRODUCTION

In previous papers[1,2] the author first obtained a Euler angle parametrization of the complex rotation group $O^+(n,\mathbb{C})$ and some of its subgroups $\hat{O}^+(p,n-p)$,[3] $0\leqslant p\leqslant n$, and then, using an isomorphism of $\hat{O}^+(p,n-p)$ with the physically more important generalized Lorentz (i.e., psuedo-orthogonal) group $O^+(p,n-p)$, proved that this leads to a corresponding parametrization of these latter groups also. However, in addition to being extremely cumbersome and involved, these angles have another more serious drawback also; some factors in the resulting factorization of an element of $O^+(p,n-p)$ *do not* belong to this group. The removal of this drawback is obviously very desirable if one wishes, for example, to use these angles for the explicit computation of unitary irreducible representation (UIR) matrix elements of these groups. We prove in the present paper that by using a rather simple trick, it is possible to overcome both these difficulties. To get some idea about the nature of this trick, let us recall that the Euler angles of Refs. 1 and 2 were, essentially, a collection of sets of "polar angles," this last phrase meaning the $(m-1)$ angular elements of the m spherical polar coordinates in the m-dimensional complex Euclidean space $\mathbb{C}^m$. In Refs. 1 and 2, we naturally used a "fixed" definition of these angles, i.e., the same definition for every set of polar angles used at different stages. However, apart from being quite natural and therefore assumed implicitly and unconsciously, there is no real justification for this restriction. In fact, as we show in the sequel, if we remove this condition and use suitably varied definitions at different stages, then not only does this completely remove the complexity of the definition of Euler angles of $O^+(p,n-p)$, but also leads to a factorization of elements of this group which consists entirely of simpler elements of the *same* group. We start in Sec. II below with these varied definitions and obtain two important theorems as easy consequences of them. With the help of these theorems, we define in Sec. III, the (complex) Euler angles of $\hat{O}^+(p,n-p)$ which lead in Sec. IV, in a much simpler manner than in Ref. 2, to (real) Euler angles of $O^+(p,n-p)$. In order to avoid repetition and to save space, we shall not describe here the notation already introduced in Refs. 1 and 2; thus for any unexplained notation, the reader is referred to these two papers.

II. POLAR ANGLES OF sth TYPE

Consider a Cartesian coordinate system I in the m-dimensional complex Euclidean space $\mathbb{C}^m$. Let s be any integer such that $2\leqslant s\leqslant m-1$. If the coordinates of a point $Z\in\mathbb{C}^m$ (relative to I) are

$$(z_1, z_2,\ldots, z_m),$$

with

$$t=(z_1^2+z_2^2+\cdots+z_m^2)^{1/2},\quad \text{Re } t\geqslant 0,$$

then

$$(x_2, x_3,\ldots, x_m),$$

the "polar angles of sth type relative to I" of Z are defined as follows:

$$\begin{aligned}
&x_j=\theta_j+i\varphi_j,\quad j=2,3,\ldots,m,\quad \text{all } \theta_j,\varphi_j \text{ real},\\
&z_m=t\cos x_m,\quad 0\leqslant\theta_m\leqslant\pi,\\
&z_{m-1}=t\sin x_m\cos x_{m-1}\quad 0\leqslant\theta_{m-1}\leqslant\pi,\\
&z_{s+1}=t\sin x_m\cdots\sin x_{s+2}\cos x_{s+1},\quad 0\leqslant\theta_{s+1}\leqslant\pi,\\
&z_s=t\sin x_m\cdots\sin x_{s+1}\sin x_s,\quad -\pi/2\leqslant\theta_s\leqslant\pi/2,\\
&z_{s-1}=t\sin x_m\cdots\sin x_{s+1}\cos x_s\sin x_{s-1},\\
&\quad -\pi/2\leqslant\theta_{s-1}\leqslant\pi/2,\\
&z_3=t\sin x_m\cdots\sin x_{s+1}\cos x_s\cos x_{s-1}\cdots\cos x_4\sin x_3,\\
&\quad -\pi/2\leqslant\theta_3\leqslant\pi/2,\\
&z_2=t\sin x_m\cdots\sin x_{s+1}\cos x_s\cos x_{s-1}\cdots\cos x_3\sin x_2,\\
&\quad 0\leqslant\theta_2\leqslant 2\pi,\\
&z_1=t\sin x_m\cdots\sin x_{s+1}\cos x_s\cos x_{s-1}\cdots\cos x_3\cos x_2.
\end{aligned}\tag{1}$$

These mean

$$\begin{aligned}
&\cos x_j=\frac{\pm z_j}{(z_1^2+z_2^2+\cdots+z_j^2)^{1/2}},\quad s+1\leqslant j\leqslant m,\\
&\sin x_j=\frac{\pm z_j}{(z_1^2+z_2^2+\cdots+z_j^2)^{1/2}},\quad 2\leqslant j\leqslant s,
\end{aligned}\tag{2}$$

where the sign in each right-hand side is either $(+)$ or $(-)$. Just as in Ref. 1, these angles and t determine, and are themselves determined uniquely (for $t\neq 0$), by the set of Cartesian coordinates

$$(z_1,z_2,\ldots,z_m)$$

of Z.

Let $e_j, 1\leqslant j\leqslant m$ be the unit vector

$$e_j=[0,0,\ldots,0,\underset{\substack{\downarrow\\ j\text{th place}}}{1},0,\ldots,0]^T,$$

 0022-2488/84/113166-05$02.50

and $r_{jk}(x), 1 \leqslant j,k \leqslant m, j \neq k$, be the $m \times m$ matrix of rotation through an angle x in the $(j-k)$ plane, i.e.,

$r_{jk}(x)$

$$\begin{bmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{bmatrix} \begin{matrix} \rightarrow j\text{th row} \\ \rightarrow k\text{ th row'} \end{matrix}$$

$\downarrow$ jth column $\downarrow$ k th column

where all the elements which have not been explicitly written down are supposed to be those of the unit $m \times m$ matrix.

By explicitly multiplying e_m on the left by various matrices one by one, it is now a straightforward matter to check that for $t = 1$,

$$r_{12}(x_2) r_{13}(x_3) \cdots r_{1s}(x_s) r_{1\,s+1}(-x_{s+1}) r_{s+1\,s+2}(-x_{s+2}) \cdots r_{m-1\,m}(-x_m) e_m = z,$$

where

$$z = [z_1, z_2, \ldots, z_m]^T.$$

This can be written as

$$r^T_{m-1\,m}(-x_m) \cdots r^T_{s+1\,s+2}(-x_{s+2}) r^T_{1\,s+1}(-x_{s+1}) \times r_{1s}(x_s) \cdots r_{12}(x_2) z = e_m,$$

which proves the following theorem, if we recall the well-known result that if r and r' are the column vectors representing the coordinates of the same point $P \in \mathbb{C}^m$ relative to the systems I and I', where I' is obtained from I by a rotation represented by the matrix N, then $r = N^T r$.

Theorem 1: If a point $Z \in \mathbb{C}^m$ has coordinates

$$(z_1, z_2, \ldots, z_m)$$

referred to a system I, where $z_1^2 + z_2^2 + \cdots + z_m^2 = 1$, then

$$(x_2, x_3, \ldots, x_m)$$

are the polar angles of the sth type relative to I, of Z, and I is given a sequence of rotations

$$\mathscr{J}_{12}(x_2), \ldots, \mathscr{J}_{1s}(x_s), \mathscr{J}_{1\,s+1}(-x_{s+1}),$$
$$\mathscr{J}_{s+1\,s+2}(-x_{s+2}), \ldots, \mathscr{J}_{m-1\,m}(-x_m),$$

in this order (whose resultant we denote by $\mathscr{J}$), the system I' so obtained will have its mth axis along OZ. Here

$\mathscr{J}_{jk}(x)$ = a rotation by an angle x in the $(j-k)$ plane.

Theorem 2: If (i) z of Theorem 1 is of the form

$$z = (-ix_1, \ldots, -ix_p, x_{p+1}, \ldots, x_n), \text{ all } x_i \text{ real};$$

(ii)

$$S_p = -\sum_{i=1}^{p} x_i^2,$$

$$S_j = -\sum_{i=1}^{p} x_i^2 + \sum_{i=p+1}^{j} x_i^2 \quad \text{for } j = p+1, p+2, \ldots, n;$$

(iii) S_j changes sign at $j = s$, i.e., $S_s < 0$ and $S_{s+1} > 0$;
(iv) I' is obtained from I by giving it the rotation $\mathscr{J}$ of Theorem 1,
(v) v and v' are the column vectors representing the coordinates of a point $P \in \mathbb{C}^n$ relative to I and I', respectively, then v being of the form

$$[-i\cdot\mathrm{Re}, \ldots, -i\cdot\mathrm{Re}, \mathrm{Re}, \ldots, \mathrm{Re}]^T \qquad (3)$$

$\downarrow$ pth place

implies that v' is also of this form; here Re stands for "a real number."

Note first of all, that the explicit form (i) of z shows, with the help of (ii) and relations (2), that

$$x_n = \theta_n, \ldots, \; x_{s+2} = \theta_{s+2}, \; x_{s+1} = i\varphi_{s+1}, \; x_s = i\varphi_s, \ldots, \; x_{p+1} = i\varphi_{p+1}, \; x_p = \theta_p, \ldots, \; x_2 = \theta_2,$$

where, as mentioned earlier, θ^s and φ^s are all real numbers. It therefore follows that

$$v' = r^T_{n-1\,n}(-\theta_n) \cdots r^T_{s+1\,s+2}(-\theta_{s+2}) r^T_{1\,s+1}(-i\varphi_{s+1}) \times r^T_{1s}(i\varphi_s) \cdots r^T_{1\,p+1}(i\varphi_{p+1}) r^T_{1\,p}(\theta_p) \cdots r^T_{12}(\theta_2) v.$$

Keeping in mind the form of the matrices $r^T_{jk}(x)$, we now just observe that if v is of the form (3) and we carry out the multiplication of v by the above $(n-1)$ matrices one at a time [starting with $r^T_{12}(\theta_2)$ and moving to the left] then at *each* step, we get a column vector of exactly the same form. In particular, the final vector v' is also of the same form (3), as required.

Corollary: Under the conditions of the theorem, v being of the form

$$[\mathrm{Re}, \ldots, \mathrm{Re}, i\cdot\mathrm{Re}, \ldots, i\cdot\mathrm{Re}]^T$$

$\downarrow$ pth place

implies v' is also of this form.

III. EULER ANGLES OF $\hat{O}^+(p, n-p)$

Consider now an arbitrary element $\hat{\alpha}$ of $\hat{O}^+(p, n-p)$.[4] Assuming as usual that $\hat{\alpha}$ transforms the Cartesian coordinate system I in $\mathbb{C}^n$ to another such system I', with the same origin as I, we recall that if X_j is the tip of the unit vector along the jth axis of I', the coordinates of X_j in the system I are given by the jth column of $\hat{\alpha}$. Now from the definition of $\hat{O}^+(p, n-p)$,[1] we see that the nth column of $\hat{\alpha}$, i.e., the coordinates of X_n in the system I will be

$$[-i\alpha_{1n}, \ldots, -i\alpha_{pn}, \alpha_{p+1\,n}, \ldots, \alpha_{nn}]^T,$$

where α_{ij} are all real and, of course,

$$-\sum_{i=1}^{p} \alpha_{in}^2 + \sum_{i=p+1}^{n} \alpha_{in}^2 = 1.$$

If $S_{nj}, p \leqslant j \leqslant n$ is the sum

$$S_{np} = -\sum_{i=1}^{p} \alpha_{in}^2,$$

$$S_{nj} = -\sum_{i=1}^{p} \alpha_{in}^2 + \sum_{i=p+1}^{j} \alpha_{in}^2, \quad j = p+1, \ldots, n,$$

(so that $S_{nn} = 1$), suppose that it changes sign at $j = s_n$, i.e.,

$$S_{n\,s_n} < 0 \text{ and } S_{n\,s_n+1} > 0.$$

We now take

$$x_{n2}, x_{n3}, \ldots, x_{nn}$$

as the polar angles of s_n th type relative to I, of X_n. Then (2) shows that

$$x_{nn} = \theta_{nn}, \ldots, \; x_{n\,s_n+2} = \theta_{n\,s_n+2}, \; x_{n\,s_n+1} = i\varphi_{n\,s_n+1}, \; x_{n\,s_n} = i\varphi_{n\,s_n}, \ldots, \; x_{n\,p+1} = i\varphi_{n\,p+1}, \; x_{np} = \theta_{np}, \qquad (4)$$
$$\ldots, \; x_{n2} = \theta_{n2}.$$

Let us now subject $I \equiv I^n$ to the sequence of rotations

$$\mathscr{J}_{12}(x_{n2}),\ldots,\mathscr{J}_{1\,s_n}(x_{n\,s_n}),\mathscr{J}_{1\,s_n+1}(-x_{n\,s_n+1}),$$
$$\mathscr{J}_{s_n+1\,s_n+2}(-x_{n\,s_n+2}),\ldots,\mathscr{J}_{n-1\,n}(-x_{nn}),$$

in this order (whose resultant rotation we denote by $\mathscr{J}^n$) and call the system so obtained I^{n-1}. Then if N_n is the matrix representing the rotation $\mathscr{J}^n$, the matrix

$$\hat{\alpha}^{n-1}\equiv N_n^T\hat{\alpha}$$

will transform the system I^{n-1} to I' as the coordinates of X_j in the system I^{n-1} will be represented by

N_n^T operating on the jth column of $\hat{\alpha}$

$=j$th column of $\hat{\alpha}^{n-1}$.

Now by Theorem 1,[5]

$I^{n-1}(n)=I'(n)$

$\Rightarrow$ the last column of $\hat{\alpha}^{n-1}$ is $[0,0,\ldots,0,1]^T$

$\Rightarrow$ the last row of $\hat{\alpha}^{n-1}$ is $[0,0,\ldots,0,1]$

as $\hat{\alpha}^{n-1}$ is certainly orthogonal.

Next, as any of the last $(n-p)$ columns of $\hat{\alpha}$ is of the form

$$[-i\cdot\mathrm{Re},\ldots,-i\cdot\mathrm{Re},\underset{\substack{\downarrow\\ p\text{th place}}}{\mathrm{Re}},\ldots,\mathrm{Re}]^T,$$

Theorem 2 shows that the same will be true for the last $(n-p)$ columns of $\hat{\alpha}^{n-1}$. Similarly, as any of the first p columns of $\hat{\alpha}$ is of the form

$$[\mathrm{Re},\ldots,\underset{\substack{\downarrow\\ p\text{th place}}}{\mathrm{Re}},i\cdot\mathrm{Re},\ldots,i\cdot\mathrm{Re}]^T,$$

the corollary to Theorem 2 shows that the same will be true for the first p columns of $\hat{\alpha}^{n-1}$ also. It follows that if $\hat{\alpha}^{(n-1)}$ is the matrix obtained from $\hat{\alpha}^{n-1}$ by removing the last row and column then $\hat{\alpha}^{(n-1)}\in\hat{\mathrm{O}}^+(p,n-1-p)$.

We thus have a procedure which, when applied to the collection of objects[5]

$$\left\{\begin{matrix} I\equiv I_n^n;\ I'\equiv I'_n;\ \hat{\alpha}\in\hat{\mathrm{O}}^+(p,n-p),\\ \hat{\alpha}\ \text{transforms}\ I\ \text{to}\ I'\end{matrix}\right\},$$

leads to (i) a positive integer s_n with $p\leqslant s_n\leqslant n-1$; (ii) a set of $(n-1)$ angles (4); (iii) a coordinate system I^{n-1} such that $I^{n-1}(n)=I'(n)$; (iv) a matrix $\hat{\alpha}^{n-1}=N_n^T\hat{\alpha}$ which transforms the system I^{n-1} to I' and is such that $\hat{\alpha}^{(n-1)}$ transforms the system I_{n-1}^{n-1} to I'_{n-1} (see Ref. 5) and belongs to $\hat{\mathrm{O}}^+(p,n-1-p)$, from which another such collection

$$\left\{\begin{matrix} I_{n-1}^{n-1};\ I'_{n-1};\ \hat{\alpha}^{(n-1)}\in\hat{\mathrm{O}}^+(p,n-1-p),\\ \hat{\alpha}^{(n-1)}\ \text{transforms}\ I_{n-1}^{n-1}\ \text{to}\ I'_{n-1}\end{matrix}\right\}$$

can be obtained to which the same procedure can obviously again be applied. Thus it is possible to repeatedly apply this procedure; applying it $(n-p)$ times in all, we shall end up with the collection

$$\left\{\begin{matrix} I_p^p;I'_p;\ \hat{\alpha}^{(p)}\in\hat{\mathrm{O}}^+(p,0)\equiv\mathrm{O}^+(p,\mathbb{R}),\\ \hat{\alpha}^{(p)}\ \text{transforms}\ I_p^p\ \text{to}\ I'_p\end{matrix}\right\}.$$

In the process, a large number of objects would have been defined; in order to fully identify them all, let us consider the jth step in some detail. At the end of the $(j-1)$th step, we shall have the following: (i) the set of positive integers

$$s_k,\quad k=n,n-1,\ldots,n-j+2,$$

with

$$p\leqslant s_k\leqslant k+1;$$

(ii) the collection of the set of angles

$$\{\theta_{kk},\ldots,\theta_{k\,s_k+2},i\varphi_{k\,s_k+1},\ldots,i\varphi_{k\,p+1},\theta_{kp},\ldots,\theta_{k2}\}\qquad(5)$$

obtained by giving k the values $n,n-1,\ldots,n-j+2$; (iii) the set of coordinate systems

$$I^k,\quad k=n,n-1,\ldots,n-j+1,$$

such that

$$I^k(l)=I'(l),\quad l=k+1,\ldots,n;$$

(iv) the set of matrices

$$\hat{\alpha}^k=N_{k+1}^T\cdots N_n^T\hat{\alpha},\quad k=n,n-1,\ldots,n-j+1,$$

such that $\hat{\alpha}^k$ transforms I^k to I', $\hat{\alpha}^{(k)}$ transforms I_k^k to I'_k and $\hat{\alpha}^{(k)}\in\hat{\mathrm{O}}^+(p,k-p)$.

Here

$$N_k=r_{12}(\theta_{k2})\cdots r_{1p}(\theta_{kp})r_{1\,p+1}(i\varphi_{k\,p+1})\cdots$$
$$r_{1\,s_k}(i\varphi_{k\,s_k})r_{1\,s_k+1}(-i\varphi_{k\,s_k+1})r_{s_k+1\,s_k+2}(-\theta_{k\,s_k+2})$$
$$\cdots r_{k-1\,k}(-\theta_{kk}).\qquad(6)$$

The jth step is as follows: Let the last column of $\hat{\alpha}^{(n-j+1)}$ [which, of course, gives the coordinates of $X(n-j+1)$ in the system I_{n-j+1}^{n-j+1}] be

$$[-i\alpha_{1\,n-j+1}^{n-j+1},-i\alpha_{2\,n-j+1}^{n-j+1},\ldots,-i\alpha_{p\,n-j+1}^{n-j+1},$$
$$\alpha_{p+1\,n-j+1}^{n-j+1},\ldots,\alpha_{n-j+1\,n-j+1}^{n-j+1}]^T,\qquad(7)$$

and let

$$S_{n-j+1\,p}=\sum_{i=1}^{p}-(\alpha_{i\,n-j+1}^{n-j+1})^2,$$

$$S_{n-j+1\,k}=\sum_{i=1}^{p}-(\alpha_{i\,n-j+1}^{n-j+1})^2+\sum_{p+1}^{k}(\alpha_{i\,n-j+1}^{n-j+1})^2,\quad k=p+1,\ldots,n-j+1.$$

We define s_{n-j+1} as the value of k for which $S_{n-j+1\,k}$ changes sign, i.e.,

$$S_{n-j+1\,s_{n-j+1}}\langle 0,\ S_{n-j+1\,s_{n-j+1}+1}\rangle 0.$$

We next define

$$x_{n-j+1\,n-j+1},x_{n-j+1\,n-j},\ldots,x_{n-j+1\,2},$$

as the polar angles of s_{n-j+1}th type relative to I_{n-j+1}^{n-j+1} of $X(n-j+1)$ so that (7) and relations (2) show that

$$x_{n-j+1\,n-j+1}=\theta_{n-j+1\,n-j+1},\ldots,x_{n-j+1\,s_{n-j+1}}+2$$
$$=\theta_{n-j+1\,s_{n-j+1}}+2,$$
$$x_{n-j+1\,s_{n-j+1}+1}=i\varphi_{n-j+1\,s_{n-j+1}+1},\ldots,x_{n-j+1\,p+1}$$
$$=i\varphi_{n-j+1\,p+1},$$
$$x_{n-j+1\,p}=\theta_{n-j+1\,p},\ldots,x_{n-j+1\,2}=\theta_{n-j+1\,2}.$$

We now give to I^{n-j+1} the sequence of rotations

$$\mathscr{J}_{12}(x_{n-j+1\,2}),\ldots,\mathscr{J}_{1\,s_{n-j+1}}(x_{n-j+1\,s_{n-j+1}}),$$

$$\mathscr{J}_{1\,s_{n-j+1}+1}(-x_{n-j+1\,s_{n-j+1}+1}),$$
$$\mathscr{J}_{s_{n-j+1}+1\,s_{n-j+1}+2}(-x_{n-j+1\,s_{n-j+1}+2}$$
$$\cdots\mathscr{J}_{n-j\,n-j+1}(-x_{n-j+1\,n-j+1}),$$

in this order (we denote by $\mathscr{J}^{n-j+1}$ the resultant of these rotations) and call the system so obtained I^{n-j}; it satisfies the property of (iii) above with $k=n-j$. Finally, as the last item of the jth step, $\hat{\alpha}^{n-j}$ is defined by

$$\hat{\alpha}^{n-j}=N^T_{n-j+1}\hat{\alpha}^{n-j+1}$$
$$=N^T_{n-j+1}N^T_{n-j+2}\cdots N^T_n\hat{\alpha},$$

and possesses the property of (iv) above with $k=n-j$. Here N_{n-j+1} is the matrix of the rotation $\mathscr{J}^{n-j+1}$ and is therefore given by

$$N_{n-j+1}$$
$$=r_{12}(\theta_{n-j+1\,2})\cdots r_{1\,p}(\theta_{n-j+1\,p})$$
$$\times r_{1\,p+1}(i\varphi_{n-j+1\,p+1})\cdots r_{1\,s_{n-j+1}}(i\varphi_{n-j+1\,s_{n-j+1}})$$
$$\times r_{1\,s_{n+j+1}+1}(-i\varphi_{n-j+1\,s_{n-j+1}+1})$$
$$\times r_{s_{n-j+1}+1\,s_{n-j+1}+2}(-\theta_{n-j+1\,s_{n-j+1}+2})$$
$$\cdots r_{n-j\,n-j+1}(-\theta_{n-j+1\,n-j+1}),$$

i.e., by (6) with $k=n-j+1$.

Thus we have been able to define the collection of the set of angles (5) obtained by giving k the values

$$k=n,n-1,\ldots,p+1, \tag{8}$$

and a sequence of rotations through these angles, in suitable planes and in suitable order, transforms I into I^p whose last $(n-p)$ axes are along the corresponding axes of I' and the transformation from I^p_p to I'_p is given by the determined matrix $\hat{\alpha}^{(p)}\in O^+(p,\mathbb{R})$. Taking

$$\theta_{jk},\quad 2\leqslant k\leqslant j\leqslant p \tag{9}$$

as the usual Euler angles[6] of $\hat{\alpha}^{(p)}$, we get the $\frac{1}{2}n(n-1)$ angles (5) with (8), and (9) which determine the transformation from I to I' (rotations through these angles in suitable order transforms I into I') and hence the element $\hat{\alpha}\in\hat{O}^+(p,n-p)$; these angles may therefore be taken as the Euler angles of $\hat{\alpha}$.

In the end, let us make the following remarks.

(i) If e is the $n\times n$ unit matrix, then

$$e=N^T_2N^T_3\cdots N^T_n\hat{\alpha}$$
$$\Rightarrow\hat{\alpha}=N_nN_{n-1}\cdots N_2, \tag{10}$$

where $N_k, k\geqslant p+1$ is given by (6) while

$$N_k=r_{12}(\theta_{k2})\cdots r_{1\,k-1}(\theta_{k\,k-1})r_{1k}(-\theta_{kk}) \tag{11}$$

for $2\leqslant k\leqslant p$.

(ii) Any collection of angles (5), (8), and (9) with

$$p\leqslant s_k\leqslant k-1,\quad k=n,n-1,\ldots,p+1,$$
$$0\leqslant\theta_{kn},\theta_{k\,n-1},\ldots,\theta_{k\,s_k+2}\leqslant\pi,$$
$$-\pi/2\leqslant\theta_{kp},\theta_{k\,p-1},\ldots,\theta_{k3}\leqslant\pi/2,$$
$$0\leqslant\theta_{k2}\leqslant2\pi,$$
all the φ's real numbers,

will give a (general) element of $\hat{O}^+(p,n-p)$ according to (10). [That $\hat{\alpha}$ given by (10) belongs to $\hat{O}^+(p,n-p)$ follows from the fact that each $N_k\in\hat{O}^+(p,n-p)$.]

IV. EULER ANGLES OF THE GENERALIZED LORENTZ GROUP

For the parametrization of the generalized Lorentz group $O^+(p,n-p)$, we use its isomorphism with $\hat{O}^+(p,n-p)$ considered in Refs. 1 and 2. Thus

$$\alpha\in O^+(p,n-p)\Rightarrow\hat{\alpha}\in\hat{O}^+(p,n-p),$$

where

$$(\hat{\alpha})_{jk}=\alpha_{jk}\begin{cases}\text{for } j\leqslant p \text{ and } k\leqslant p,\\ \text{for } j\geqslant p+1 \text{ and } k\geqslant p+1,\end{cases}$$
$$(\hat{\alpha})_{jk}=i\alpha_{jk}\quad\text{for } j\geqslant p+1 \text{ and } k\leqslant p,$$
$$(\hat{\alpha})_{jk}=-i\alpha_{jk}\quad\text{for } j\leqslant p \text{ and } k\geqslant p+1,$$

i.e., $\hat{\alpha}=f\alpha f^{-1}$,

f being the diagonal matrix

$$f=\operatorname{diag}(1,1,\ldots,1,i,i,\ldots,i).$$
$$\qquad\qquad\downarrow$$
$$\qquad\qquad p\text{th place}$$

We now obtain the Euler angles of $\hat{\alpha}$ as in the previous section, so as to get

$$\hat{\alpha}=N_nN_{n-1}\cdots N_2$$
$$\Rightarrow\alpha=f^{-1}\hat{\alpha}f$$
$$=f^{-1}N_nN_{n-1}\cdots N_2f,$$
$$\text{i.e., } \alpha=L_nL_{n-1}\cdots L_2, \tag{12}$$

where

$$L_k=f^{-1}N_kf,\quad k=2,3,\ldots,n.$$

Now, it is easy to verify that

$$f^{-1}r_{1k}(\theta)f=r_{1k}(\theta)\quad 2\leqslant k\leqslant p,$$
$$f^{-1}r_{1k}(i\varphi)f=l_{1k}(\varphi)\quad p+1\leqslant k\leqslant n,$$
$$f^{-1}r_{k\,k+1}(\theta)f=r_{k\,k+1}(\theta)\quad p+1\leqslant k\leqslant n-1,$$

where $l_{jk}(\varphi)$ is the matrix of the simple Lorentz transformation by an angle φ in the $(j-k)$ plane, and is therefore given by

$$\begin{bmatrix}\cosh\varphi & -\sinh\varphi\\ -\sinh\varphi & \cosh\varphi\end{bmatrix}\begin{matrix}\rightarrow j\text{th row}\\ \rightarrow k\text{ th row}\end{matrix}$$
$$\quad\downarrow\qquad\quad\downarrow$$
$$j\text{th column}\quad k\text{ th column}$$

Hence, we shall have, for $k\geqslant p+1$,

$$L_k=r_{12}(\theta_{k2})\cdots r_{1\,p}(\theta_{kp})l_{1\,p+1}(\varphi_{k\,p+1})\cdots$$
$$l_{1\,s_k}(\varphi_{k\,s_k})l_{1\,s_k+1}(-\varphi_{k\,s_k+1})r_{s_k+1\,s_k+2}(-\theta_{k\,s_k+2})$$
$$\cdots r_{k-1\,k}(-\theta_{kk}), \tag{13a}$$

while for $2\leqslant k\leqslant p$,

$$L_k=r_{12}(\theta_{k2})\cdots r_{1\,k-1}(\theta_{k\,k-1})r_{1k}(-\theta_{kk}). \tag{13b}$$

Thus we have the following set of $\frac{1}{2}n(n-1)$ Euler angles of $\alpha\in O^+(p,n-p)$:

$$\{\theta_{kk},\ldots,\theta_{k\,s_k+2},\varphi_{k\,s_k+1},\ldots,\varphi_{k\,p+1},\theta_{kp},\ldots,\theta_{k2}\} \tag{14}$$

with

$$k = n,n-1,\ldots,p+1, \tag{15}$$

and

$$\theta_{jk}, \quad 2 \leqslant k \leqslant j \leqslant p; \tag{16}$$

in terms of these, α factorizes as (12) with L_k given by (13).

[1]A. Syed, J. Math. Phys. **24**, 463 (1983).
[2]A. Syed, J. Math. Phys. **24**, 471 (1983).
[3]For the sake of conformity with the more usual notation, we replace $\hat{L}(n,r)$ and $L(n,r)$, $0<r<n$, of Refs. 1,2 by $\hat{O}^+(p,n-p)$ and $O^+(p,n-p)$, $0<p<n$, respectively, in the present paper.
[4]Recall that $\hat{O}^+(p,n-p)$ is the group $\hat{L}(n,p)$ defined in Ref. 1.
[5]See Refs. 1,2 for all unexplained notation.
[6]A. Syed, J. Natur. Sci. Math. (Pakistan) **9**, 274 (1969).

Structure and representations of the hyperoctahedral group

M. Baake
Physikalisches Institut, Universität Bonn, D-5300 Bonn 1, West Germany

(Received 3 January 1984; accepted for publication 13 April 1984)

In this paper, the main properties of the symmetry group of the n-dimensional cube are reviewed and formulated with respect to possible applications in lattice theories. The connection between the hyperoctahedral group W_n and the orthogonal group $O(n)$ is investigated by means of the canonical representation.

PACS numbers: 02.20.Qs, 02.20.Rt, 05.50. + q

I. INTRODUCTION

A. Survey

Recently the importance of finite groups grew out of the various possibilities of their application to physical problems, especially on lattices. First of all, the space symmetry of a given problem is an important factor for the calculation of possible solutions. In this context, the hyperoctahedral group W_n turns out to be a crystallographic point group in n-dimensional Euclidean space with perhaps one of the widest fields of applications of all. For this purpose, one has to know in detail the group structure and at least the canonical representation T, which is the most important one in all problems with the outer symmetry described by the hyperoctahedral group. Apart from the fact that the canonical representation, as will be shown later, is the direct link to the rotation group $O(n)$, the tensor calculus of this representation leads to all other representations of W_n.

Especially, the adjoined representation of W_n turns out to be the skew-symmetric part of $T \otimes T$, as is valid for the appropriate representations of $O(n)$ in exactly the same way.

Second, the hyperoctahedral group is correlated with discrete σ-models where the partition function on a lattice with N points is written as

$$Z = \sum_{\mathbf{s}_1 \cdots \mathbf{s}_N \in M} e^{-\beta E}, \quad \beta = \frac{1}{kT}, \tag{1.1}$$

with the energy

$$E \propto \sum_{\langle i,j \rangle} \mathbf{s}_i \cdot \mathbf{s}_j, \quad \mathbf{s}_i \in M, \quad |\mathbf{s}_i| = 1,$$

where $\langle i,j \rangle$ denotes summation over $1 \leqslant i \leqslant N$ and all j that are next neighbors. Especially, if $M \subset \mathbb{R}^n$ denotes the set of vectors with integral components, the hyperoctahedral group of dimension n describes the global symmetry of this model.

Third, the knowledge of the group characters, especially of the canonical and the adjoined ones, are required for the calculation of chiral models on lattices, where the sum over states is written as

$$Z = \sum_{g_1 \cdots g_N \in W_n} e^{-\beta E},$$

$$E = \sum_{\langle i,j \rangle} \sum_{\lambda} a_\lambda \cdot \chi_\lambda (g_i g_j^{-1}), \tag{1.2}$$

where the a_λ's denote the coupling constants of the different characters and χ_λ runs through the system of irreducible characters of W_n.

Last but not least, the same is required for calculations with lattice gauge theories where

$$E = \sum_{\text{plaquets}} \sum_{\lambda} a_\lambda \cdot \chi_\lambda (g_i g_j g_k^{-1} g_l^{-1}). \tag{1.3}$$

In many problems one would like to use $SU(n)$ and lattices with nontrivial sizes (the sums in the above formulas then turning to integrals), but such calculations, even if they are done with Monte Carlo methods, are nearly impossible from the view of available CPU time. The results of Creutz[1] concerning the Monte Carlo study of SU(2) show that an equivalent technique for $SU(n)$ with higher n is impossible. Calculations of Petcher and Weingarten[2] show that for SU(2) a very good approximation is possible if one uses finite subgroups instead of SU(2) itself. The best result was obtained with the icosahedral group, which is of order 120. Unfortunately, no family of subgroups of $SU(n)$ is known that seems to be of great promise for a generalization of this method. However, as was shown by Lovelace,[3] calculations with $SU(n)$ and $O(n)$ look similar in some cases for large n, so that results with $O(n)$ are of great interest. Obviously, direct calculations with $O(n)$ are restricted by the same reasons as mentioned above, which raises the question of approximation with the aid of nontrivial finite subgroups.

For this purpose, it is favorable that for $O(n)$ two nontrivial families of finite subgroups are known: the symmetry group of the n-dimensional simplex and the symmetry group of the n-dimensional cube. The first is identical with the symmetric group S_{n+1}, the latter with the so-called hyperoctahedral group which will now be discussed in those details that may be important for physical applications. Clearly, a lot of the results that will be stated below are well known, but can only be found in mathematical textbooks which are correlated with too general a point of view (see Refs. 4–9). Thus, the main properties of the hyperoctahedral group must be presented in a simple language, prepared for the application to physical problems. Especially, some explicit calculus with the canonical representation has to be done as well as an explicit calculation of number and order of conjugacy classes. Hence, the paper is organized as follows.

After two possible approaches to the hyperoctahedral group the structure of this group is discussed, in Sec. II and Sec. III, in some details such as the wreath product struc-

 0022-2488/84/113171-12$02.50

ture, the permutation structure, and the classification of conjugacy classes, where explicit formulas are given.

In Sec. IV, a minimal system of group generators is presented that consists of only three members without dependence on n. The following sections deal with the representations of W_n, mainly with the so-called canonical (Sec. V) and the so-called adjoined (Sec. IX) one. Perhaps the most important result of this paper is the fact that both representations appear to be a direct restriction of the appropriate O(n)-representations, i.e., the canonical and the adjoined representation of O(n) stay irreducible after restriction to the finite subgroup W_n, which is one of those properties that make W_n interesting for physics.

The canonical representation T is discussed in many details, as are matrix form, geometric interpretation, characters, and so on (Secs. V–VII).

Especially, the sum

$$A(n,k) = \frac{1}{\mathrm{ord}(W_n)} \sum_{g \in W_n} [\chi_T(g)]^{2k} \tag{1.4}$$

is investigated (Sec. VIII), which occurs in mean field approximations in chiral models (see Ref. 7).

As a nice result of these calculations we present, explicitly, the power series of the function

$$\exp[\cosh(z) - 1] \tag{1.5}$$

on the complex plane.

In Sec. IX, we use the Kronecker product $T \otimes T$ for a construction of the adjoined representation, which shows in a canonical way the relations between W_n and O(n). After a short look at the representations of dimension one we close—for the sake of completeness—with a brief description of the general classification of the irreducible W_n representations (Sec. X).

In an earlier paper we presented the structure and representations of W_4 completely and derived the connections between O(4) and W_4 in detail (Refs. 10 and 11). Recently, this group found an application in four-dimensional lattice theory (see Ref. 12).

B. Definition of the hyperoctahedral group

Let us introduce the hyperoctahedral group as the "largest" crystallographic point group of the hypercubical lattice, the latter being generated by all linear combinations of the form

$$\mathbf{x} = \sum_{k=1}^{n} x^k \mathbf{e}_k, \tag{1.6}$$

with integral coefficients x^k and the standard ON-basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ of the n-dimensional Euclidean space. This definition produces in a canonical way the wreath product structure and the permutation structure as will be shown later. Furthermore, the geometrical interpretation can be seen from the beginning.

C. An alternative approach

As is well known, the orthogonal group O(n) is the group of all $n \times n$ matrices which leave the scalar product

$$\langle \mathbf{x},\mathbf{y} \rangle := \sum_{i=1}^{n} x^i y^i \tag{1.7}$$

of the Euclidean space $\mathbb{R}^n$ invariant. Now, if we take all those matrices of O(n) that map an arbitrary vector $\mathbf{x}$ with integer components, i.e., $\mathbf{x} \in \mathbb{Z}^n$, to another vector $\mathbf{y} \in \mathbb{Z}^n$, we obtain the group O($n$,$\mathbb{Z}$), which is isomorphic to the hyperoctahedral group. A proof of this statement is given in this paper. Now we can understand why W_n occurs as the global symmetry in the special case of σ-models mentioned above where $M = \mathbb{Z}^n$. For an application the reader is referred to Ref. 13.

D. Preliminaries

The notation used is almost the same as in a preceding paper about W_4 (see Ref. 10). The symmetric group of degree n is denoted by S_n, the cyclic group of order 2 by Z_2 together with the symbol " $+_2$" for addition modulo 2, and the group

$$Z_2{}^n = \underbrace{Z_2 \otimes \cdots \otimes Z_2}_{n \text{ times}}$$

is used in the form

$$Z_2{}^n = \left\{ \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \middle| a_1,\ldots,a_n \in Z_2 \right\}, \tag{1.8}$$

together with the operation

$$\mathbf{a} +_2 \mathbf{b} := \begin{pmatrix} a_1 & +_2 & b_1 \\ & \vdots & \\ a_n & +_2 & b_n \end{pmatrix}. \tag{1.9}$$

Furthermore, isomorphic groups are connected by "$\cong$," and conjugate group elements by "$\sim_G$," where the index G denotes the group relative to which the considered elements are conjugate. Finally, the symbol " = " used for representations only denotes the equivalence, not the actual identity of the representations.

II. STRUCTURE OF W_n

A. W_n as wreath product

Let us now consider the following finite subset L_n of the n-dimensional Euclidean space $\mathbb{R}^n$,

$$L_n = \{\mathbf{e}_1; -\mathbf{e}_1; \mathbf{e}_2; -\mathbf{e}_2; \ldots; \mathbf{e}_n; -\mathbf{e}_n\}. \tag{2.1}$$

The appropriate permutation group is $S_{L_n} \cong S_{2n}$.

Now, we call two elements $\mathbf{x},\mathbf{y}$ opposite, if

$$\mathbf{y} = -\mathbf{x}. \tag{2.2}$$

In the case of L_n we therefore have n pairs of opposite elements, i.e., $(\mathbf{e}_i; -\mathbf{e}_i)$, $1 \leqslant i \leqslant n$, as shown in Fig. 1. We consider the group W_n of permutations of the elements of L_n which leave opposite elements opposite. If one regards the elements of L_n as the centers of the $2n$ faces of the n-dimensional cube, W_n turns out to be its symmetry group. In the language of Fig. 1, the elements of W_n obey the two following rules.

(1) Two elements of the same row may be interchanged.

(2) Only complete rows may be permuted.

Therefore, W_n can easily be presented as the wreath product $Z_2 \sim S_n$, i.e.,

$$W_n = \{(\mathbf{a},\pi) | \mathbf{a} \in Z_2{}^n, \pi \in S_n\}, \tag{2.3}$$

together with the multiplication rule

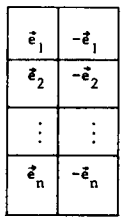


FIG. 1. Graphical presentation of the n-dimensional cube from the geometrical point of view.

$$(\mathbf{a},\pi)\cdot(\mathbf{b},\sigma) := (\mathbf{a}_\sigma + {}_2\mathbf{b}, \pi\cdot\sigma), \tag{2.4}$$

with $\mathbf{a}_\sigma$ defined by

$$(\mathbf{a}_\sigma)_k := a_{\sigma(k)} \tag{2.5}$$

(k th component of $\mathbf{a}_\sigma$). (For details about wreath products see James and Kerber[7] or Ref. 14.) Since $(\mathbf{a},\pi) = (\mathbf{b},\sigma)$ is equivalent to $\mathbf{a} = \mathbf{b}$ and $\pi = \sigma$, it is evident that W_n is a finite group of order

$$|W_n| = 2^n \cdot n!. \tag{2.6}$$

Direct calculation shows that for every n the center of W_n, i.e.,

$$C(W_n) := \{g \in W_n \mid gh = hg \quad \text{for all} \quad h \in W_n\},$$

consists of two elements,

$$C(W_n) = \{(\mathbf{0},id_{S_n});(\mathbf{1},id_{S_n})\}, \quad \mathbf{0} := \begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}, \quad \mathbf{1} := \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}. \tag{2.7}$$

B. Representations by permutations

As a consequence of $S_{L_n} \cong S_{2n}$ and $W_n \subset S_{L_n}$, W_n is isomorphic with a subgroup of S_{2n}. If we rename the squares of Fig. 1 as shown in Fig. 2, and admit again the two rules for permutations listed above, we obtain directly the following subgroup P_n of S_{2n},

$$P_n = \{\Pi \in S_{2n} \mid \bigwedge_{1\leqslant k\leqslant n} |\Pi(k+n) - \Pi(k)| = n\}. \tag{2.8}$$

With some elementary combinatorics one can see that $|P_n| = n!\cdot 2^n = |W_n|$, as it must be. Now, P_n can, together with the usual multiplication rule for permutations, be proved to be isomorphic with W_n by means of the mapping Φ defined as follows:

$$\Phi: \begin{matrix} W_n \to P_n, \\ (\mathbf{a},\pi) \mapsto \Phi(\mathbf{a},\pi), \end{matrix}$$

with

$$\Phi(\mathbf{a},\pi)(\mathbf{k}) := \begin{cases} \pi(k) + n\cdot a_k, & \text{if } 1\leqslant k\leqslant n, \\ \pi(k-n) + n(1-a_{k-n}), & \text{if } n+1\leqslant k\leqslant 2n. \end{cases} \tag{2.9}$$

Therefore, P_n can be thought of as a faithful representation of W_n by permutations.

III. CONJUGACY CLASSES

A. Classification

The group isomorphism Φ is extremely useful for the classification of the classes of conjugate elements of W_n, because we obtain the following theorem the proof of which is given in an Appendix.

Theorem: Let $(\mathbf{a},\pi), (\mathbf{b},\sigma) \in W_n$. Then $(\mathbf{a},\pi) \sim_{W_n} (\mathbf{b},\sigma)$ if and only if $\pi \sim_{S_n} \sigma$ *and* $\Phi(\mathbf{a},\pi) \sim_{S_{2n}} \Phi(\mathbf{b},\sigma)$, i.e., $(\mathbf{a},\pi)$ and $(\mathbf{b},\sigma)$ belong to the same conjugacy class of W_n, if and only if π and σ have the same cycle structure in S_n *and* $\Phi(\mathbf{a},\pi)$ and $\Phi(\mathbf{b},\sigma)$ have the same cycle structure in S_{2n}.

This classification of the conjugacy classes of W_n is of some importance because it allows the calculation of the number and the order of these classes, a parametrization of which follows later.

B. The number of conjugacy classes

The above theorem implies that for the number of conjugacy classes we have to fix a permutation $\pi \in S_n$ with the cycle structure

$$(1^{\mu_1},2^{\mu_2},\ldots,n^{\mu_n}), \quad \sum_{k=1}^{n} k\mu_k = n, \quad \mu_k \geqslant 0, \tag{3.1}$$

then to count the different cycle structures that $\Phi(\mathbf{a},\pi)$ can have in S_{2n} for the various $\mathbf{a} \in Z_2^n$, and, finally, to sum over all possible cycle structures of S_n. By the special properties of the wreath product, only two things can happen to a single cycle of an S_n-permutation π under the mapping Φ: (1) produced by the single cycle, $\Phi(\mathbf{a},\pi)$ contains two cycles of the same length, or (2) $\Phi(\mathbf{a},\pi)$ contains one cycle of twice the length.

Therefore, from μ_q S_n-cycles of length q, $\mu_q \geqslant 1$, we now get, by the action of Φ, for the various $\mathbf{a} \in Z_2^n$ $(\mu_q + 1)$ cases of appropriate cycle structures in S_{2n}:

$$q^{2\mu_q}//q^{2(\mu_q-1)}, \quad (2q)//q^{2(\mu_q-2)}, (2q)^2//$$
$$\cdots//q^2, \quad (2q)^{\mu_q-1}//(2q)^{\mu_q}. \tag{3.2}$$

Let $\pi \in S_n$ have the cycle structure $(1^{\mu_1},2^{\mu_2},\ldots,n^{\mu_n})$. The number of S_{2n}-classes in which we can find $\Phi(\mathbf{a},\pi)$ for $\mathbf{a} \in Z_2^n$ is so far given by the product

$$Q(\mu_1,\ldots,\mu_n) = \prod_{\substack{k=1 \\ \mu_k>0}}^{n} (\mu_k + 1) = \prod_{k=1}^{n} (\mu_k + 1). \tag{3.3}$$

Now the general formula for the number Q_n of conjugacy classes of W_n is easily obtained by the following summation over the partitions of n:

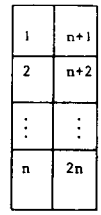


FIG. 2. Graphical presentation of the n-dimensional cube from the algebraical point of view.

TABLE I. Number of conjugacy classes of W_n for $1<n<150$.

1	2	51	130 673 928	101	2 182 704 238 810
2	5	52	164 363 280	102	2 582 113 843 795
3	10	53	206 327 710	103	3 052 274 960 840
4	20	54	258 508 230	104	3 605 324 137 485
5	36	55	323 275 512	105	4 255 414 211 990
6	65	56	403 531 208	106	5 019 039 720 949
7	110	57	502 810 130	107	5 915 411 540 970
8	185	58	625 425 005	108	6 966 891 449 865
9	300	59	776 616 430	109	8 199 492 148 220
10	481	60	962 759 294	110	9 643 456 347 436
11	752	61	1 191 580 872	111	11 333 922 675 812
12	1 165	62	1 472 454 540	112	13 311 695 737 345
13	1 770	63	1 816 715 170	113	15 624 131 048 910
14	2 665	64	2 238 075 315	114	18 326 156 951 000
15	3 956	65	2 753 078 840	115	21 481 448 170 126
16	5 822	66	3 381 689 157	116	25 163 779 232 718
17	8 470	67	4 147 937 540	117	29 458 577 635 960
18	12 230	68	5 080 752 250	118	34 464 712 775 100
19	17 490	69	6 214 880 700	119	40 296 547 446 390
20	24 842	70	7 592 053 897	120	47 086 297 888 049
21	35 002	71	9 262 292 216	121	54 986 738 253 280
22	49 010	72	11 285 536 125	122	64 174 308 234 575
23	68 150	73	13 733 486 100	123	74 852 671 601 900
24	94 235	74	16 691 879 795	124	87 256 800 664 440
25	129 512	75	20 263 074 134	125	101 657 649 931 466
26	177 087	76	24 569 214 653	126	118 367 514 773 731
27	240 840	77	29 755 845 120	127	137 746 158 532 330
28	326 015	78	35 996 306 025	128	160 207 830 443 720
29	439 190	79	43 496 760 380	129	186 229 283 983 960
30	589 128	80	52 502 280 642	130	216 358 951 877 650
31	786 814	81	63 303 821 602	131	251 227 421 256 300
32	1 046 705	82	76 246 618 325	132	291 559 408 377 040
33	1 386 930	83	91 739 827 630	133	338 187 420 114 200
34	1 831 065	84	110 268 082 280	134	392 067 356 653 475
35	2 408 658	85	132 404 776 664	135	454 296 298 907 724
36	3 157 789	86	158 827 920 009	136	526 132 805 091 285
37	4 126 070	87	190 338 386 210	137	609 020 032 390 910
38	5 374 390	88	227 881 604 535	138	704 612 097 171 965
39	6 978 730	89	272 572 552 460	139	814 804 082 539 220
40	9 035 539	90	325 725 355 088	140	941 766 219 818 916
41	11 664 896	91	388 887 409 310	141	1 087 982 771 678 576
42	15 018 300	92	463 879 670 860	142	1 256 296 287 033 025
43	19 283 830	93	552 843 114 270	143	1 449 957 907 464 780
44	24 697 480	94	658 293 423 970	144	1 672 684 577 368 315
45	31 551 450	95	783 184 076 176	145	1 928 724 031 465 432
46	40 210 481	96	930 980 399 327	146	2 222 928 642 634 633
47	51 124 970	97	1 105 744 993 420	147	2 560 839 250 666 680
48	64 854 575	98	1 312 237 775 425	148	2 948 780 346 918 950
49	82 088 400	99	1 556 031 348 120	149	3 393 968 049 615 390
50	103 679 156	100	1 843 645 820 766	150	3 904 632 614 009 852

$$Q_n = \sum_{\substack{\mu_1,\ldots,\mu_n>0 \\ \Sigma_{k=1}^n k\cdot\mu_k = n}} Q(\mu_1,\ldots,\mu_n)$$

$$= \sum_{\substack{\mu_1,\ldots,\mu_n>0 \\ \Sigma_{k=1}^n k\cdot\mu_k = n}} \left[\prod_{l=1}^{n}(\mu_l+1)\right]. \tag{3.4}$$

C. Recursive calculation of the numbers of conjugacy classes

For an explicit calculation of Q_n it is of great advantage to use the following recurrence relations. If we define for $n\geqslant 0$,

$$B(n,k) := \sum_{\substack{\mu_1,\ldots,\mu_n>0 \\ \Sigma_{j=1}^n j\cdot\mu_j = k}} \left[\prod_{l=1}^{n}(\mu_l+1)\right], \tag{3.5}$$

with $B(0,0)=1$, $B(0,k)=0$ for $k>0$, and $B(n,k) := 0$ for $k<0$, we get the formula

$$B(n+1,k) = \sum_{l=0}^{\infty}(l+1)\cdot B(n,k-l\cdot(n+1)). \tag{3.6}$$

Since $B(n,j)=0$ for $j<0$, the sum on the right side is finite and can be rewritten as

$$B(n+1,k) = \sum_{l=0}^{[k/(n+1)]}(l+1)\cdot(n,k-l\cdot(n+1)), \tag{3.7}$$

where $[\alpha]$ means the largest integer less than or equal to α (i.e., the so-called Gaussian bracket). Together with the relation

$$Q_n = B(n,n), \tag{3.8}$$

the number of conjugacy classes of W_n can easily be calculated with the aid of a computer. For $1 \leqslant n \leqslant 150$, the result is listed in Table I.

D. Parametrization and order of the conjugacy classes

To obtain a formula for the order of the conjugacy classes we first need a parametrization which, by our theorem, may be done by means of the appropriate S_n- and S_{2n}-cycles.

For this purpose, let $\pi \in S_n$ have the cycle structure $(1^{\mu_1},2^{\mu_2},\ldots,n^{\mu_n})$. Then, the above consideration concerning the action of Φ shows that $\Phi(\mathbf{a},\pi) \in S_{2n}$ can only have a cycle structure of the form

$$\begin{aligned}&(1^{2\nu_1},2^{2\nu_2+(\mu_1-\nu_1)},3^{2\nu_3},4^{2\nu_4+(\mu_2-\nu_2)},\ldots,(2n)^{2\nu_{2n}+(\mu_n-\nu_n)}),\\ &\qquad 0 \leqslant \nu_k \leqslant \mu_k, \quad \text{for } 1 \leqslant k \leqslant n,\\ &\qquad \nu_k = 0, \quad \text{for } k > n.\end{aligned} \tag{3.9}$$

Consequently, our classes have $2n$ parameters according to this description, i.e.,

$$\begin{aligned}&(\mu_1,\ldots,\mu_n;\nu_1,\ldots,\nu_n), \quad \sum_{k=1}^{n} k\cdot\mu_k = n,\\ &\qquad 0 \leqslant \nu_k \leqslant \mu_k.\end{aligned} \tag{3.10}$$

Some combinatorics on the distribution of the different $\mathbf{a} \in Z_2{}^n$ to the several classes yield the ensuing formula for the order of the conjugacy classes of W_n, which can easily be calculated with the computer:

$$\begin{aligned}&\mathrm{Ord}(\mu_1,\ldots,\mu_n;\nu_1,\ldots,\nu_n)\\ &= \frac{n!\cdot 2^n \cdot \prod_{k=1}^{n} \binom{\mu_k}{\nu_k}}{(1^{\mu_1}\mu_1!2^{\mu_2}\mu_2!\cdot\ \cdots\ \cdot n^{\mu_n}\mu_n!2^{\mu_1+\cdots+\mu_n})}.\end{aligned} \tag{3.11}$$

Using the well-known relations (compare Abramowitz and Stegun[15])

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{3.12}$$

and

$$\sum_{\substack{\mu_1,\ldots,\mu_n>0\\ \Sigma_{k=1}^{n} k\mu_k = n}} \frac{1}{1^{\mu_1}\mu_1!\cdot\ \cdots\ \cdot n^{\mu_n}\cdot\mu_n!} = 1, \tag{3.13}$$

one can easily verify the necessary relation

$$\begin{aligned}&\sum_{\substack{\mu_1,\ldots,\mu_n>0\\ \Sigma_{k=1}^{n} k\mu_k = n}} \sum_{\nu_1=0}^{\mu_1} \cdots \sum_{\nu_n=0}^{\mu_n} \mathrm{Ord}(\mu_1,\ldots,\mu_n;\nu_1,\ldots,\nu_n)\\ &= n!\cdot 2^n = |W_n|.\end{aligned} \tag{3.14}$$

IV. SOME MORE PROPERTIES OF W_n

A. Ambivalency of W_n

In this context is is interesting to remark that W_n is an ambivalent group, i.e., every element $(\mathbf{a},\pi) \in W_n$ is conjugate to its inverse. By $(\mathbf{a},\pi)^{-1} = (\mathbf{a}_{\pi^{-1}},\pi^{-1})$ and $(\mathbf{a},id)\cdot(\mathbf{a}_{\pi^{-1}},\pi^{-1}) \cdot (\mathbf{a},id)^{-1} = (\mathbf{a},\pi^{-1})$ the statement is reduced to the equivalent one for the group S_n, but $\pi \sim_{S_n} \pi^{-1}$ is a straightforward consequence of the relation

$$(i_1\cdots i_q)^{-1} = (i_q\cdots i_1), \tag{4.1}$$

where $(i_1\cdots i_q)$ is the standard notation of a q-cycle (see James and Kerber[7]). The ambivalence of W_n is of some importance, for it causes all numbers of the character table of W_n to be real.

B. On the order of the group elements

Additionally, it may be quite suitable to take a short look at the order of the several group elements $g \in W_n$, defined by the relations

$$g^{\mathrm{ord}(g)} = id_{W_n}, \quad g^k \neq id_{W_n}, \quad \text{for } 1 < k < \mathrm{ord}(g). \tag{4.2}$$

First, if $\pi \in S_n$ has the cycle structure $(1^{\mu_1},\ldots,n^{\mu_n})$ with $\Sigma_{k=1}^{n} k\cdot\mu_k = n$, we get

$$\mathrm{ord}\ \pi = \mathrm{lcm}\{k \mid \mu_k > 0\}, \tag{4.3}$$

where lcm is an abbreviation for the least common multiple. As $\mathrm{ord}(\mathbf{a},\pi) = \mathrm{ord}(\Phi(\mathbf{a},\pi))$ and conjugate elements have the same order, we can give the order of an element $(\mathbf{a},\pi) \in W_n$ as a class function in the following way.

If $(\mathbf{a},\pi) \in (\mu_1,\ldots,\mu_n;\nu_1,\ldots,\nu_n)$, then $\Phi(\mathbf{a},\pi)$ has the well-defined cycle structure in S_{2n} described above and consequently,

$$\mathrm{ord}(\mathbf{a},\pi) = \mathrm{lcm}(\{k \mid \nu_k > 0\} \cup \{2k \mid 0 < \nu_k < \mu_k\}). \tag{4.4}$$

Calculating, moreover, the highest order possible in W_n, one obtains (see Table II)

TABLE II. $\mathrm{Max}\{\mathrm{ord}(g) \mid g \in W_n\}$ for $1 < n < 60$.

N	Max. order	N	Max. order	N	Max. order
1	2	21	840	41	60 060
2	4	22	840	42	65 520
3	6	23	1 680	43	120 120
4	8	24	1 680	44	120 120
5	12	25	2 520	45	120 120
6	12	26	2 520	46	120 120
7	24	27	3 080	47	240 240
8	30	28	4 620	48	240 240
9	40	29	5 040	49	360 360
10	60	30	9 240	50	360 360
11	60	31	9 240	51	360 360
12	120	32	10 920	52	360 360
13	120	33	10 920	53	720 720
14	168	34	18 480	54	720 720
15	210	35	18 480	55	720 720
16	280	36	27 720	56	720 720
17	420	37	27 720	57	942 480
18	420	38	32 760	58	1 021 020
19	840	39	32 760	59	1 113 840
20	840	40	55 440	60	2 042 040

$$\max\{\mathrm{ord}(g)|g\in W_n\}$$
$$=\max\left\{\mathrm{lcm}(\{k\,|\nu_k>0\}\cup\{2k\,|0\leqslant\nu_k<\mu_k\})\;\middle|\;\sum_{l=1}^{n}l\cdot\mu_l=n,\quad 0\leqslant\nu_k\leqslant\mu_k,\quad 1\leqslant k\leqslant n\right\}$$
$$=\max\left\{\mathrm{lcm}\{2k\,|\mu_k>0\}\;\middle|\;\sum_{l=1}^{n}l\cdot\mu_l=n\right\}$$
$$=2\cdot\max\left\{\mathrm{lcm}\{k\,|\mu_k>0\}\;\middle|\;\sum_{l=1}^{n}l\cdot\mu_l=n\right\}$$
$$=2\cdot\max\{\mathrm{ord}(\pi)|\pi\in S_n\}. \tag{4.5}$$

C. Generators of W_n

At this point we give a minimal system of generators for the hyperoctahedral group, W_n, $n\geqslant 2$, which consists of only three elements α,β,γ, independently of the parameter n. These generators are defined by

$$\alpha:=\left(\begin{pmatrix}0\\\vdots\\0\end{pmatrix},\begin{pmatrix}123&\cdots&n\\213&\cdots&n\end{pmatrix}\right)=(\mathbf{0},(12)), \tag{4.6}$$

$$\beta:=\left(\begin{pmatrix}0\\\vdots\\0\end{pmatrix},\begin{pmatrix}123&\cdots&(n-1)n\\234&\cdots&n\ 1\end{pmatrix}\right)=(\mathbf{0},(123\ldots n)), \tag{4.7}$$

$$\gamma:=\left(\begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix},\begin{pmatrix}123&\cdots&n\\123&\cdots&n\end{pmatrix}\right)=(\mathbf{e}_n,id_{S_n}). \tag{4.8}$$

As is well known, α and β generate S_n, i.e., precisely spoken, all elements $(\mathbf{0},\pi)\in W_n$ with $\pi\in S_n$. From

$$(\mathbf{0},\pi^{-1})\cdot\gamma\cdot(\mathbf{0},\pi)=(\mathbf{e}_{\pi(n)},id)$$

and

$$(\mathbf{a},id)\cdot(\mathbf{b},id)=(\mathbf{a}+{}_2\mathbf{b},id)$$

we get all the elements

$$(\mathbf{a},id)\in W_n,\quad\text{with}\ \mathbf{a}\in {Z_2}^n.$$

Finally, $(\mathbf{0},\pi)\cdot(\mathbf{a},id)=(\mathbf{a},\pi)$ yield the rest.

For the sake of simplicity, we list the ensuing relations:

$$\alpha^2=\gamma^2=\beta^n=id_{W_n}=(\mathbf{0},id_{S_n}), \tag{4.9}$$

$$\beta^{-k}\cdot\gamma\cdot\beta^k=(\mathbf{e}_k,id_{S_n}),\quad 1\leqslant k\leqslant n, \tag{4.10}$$

$$\beta^k\cdot\alpha\cdot\beta^{-k}=(\mathbf{0},((k+1)(k+2))),\quad 0\leqslant k\leqslant n-2, \tag{4.11}$$

$$\beta^{n-1}\cdot\alpha\cdot\beta^{1-n}=(\mathbf{0},(n1)). \tag{4.12}$$

With these relations every $(\mathbf{a},\pi)\in W_n$ can easily be expressed by α,β, and γ, since every permutation $\pi\in S_n$ can be written as a product of transpositions and

$$(1k)=(12)(23)\cdot\cdots\cdot((k-1)k)$$
$$\cdot(12)(23)\cdot\cdots\cdot((k-1)k),\quad\text{for}\ k>1.$$

etc.

V. THE CANONICAL REPRESENTATION

A. Definition

Let us now proceed to the representations of the hyperoctahedral group. Quite important is a representation of degree n, by which W_n is directly connected to the n-dimensional rotation group $O(n)$. The wreath product structure of W_n implies the following definition of an n-dimensional representation, which will be called the canonical representation:

$$T:W_n\to\mathrm{aut}\,\mathbb{R}^n,$$

where

$$[T(\mathbf{a},\pi)]\mathbf{e}_i:=(-1)^{a_i}\cdot\mathbf{e}_{\pi(i)}\quad\text{for}\ 1\leqslant i\leqslant n, \tag{5.1}$$

with $\{\mathbf{e}_i\,|1\leqslant i\leqslant n\}$ being the standard ON-basis of the n-dimensional Euclidean space $\mathbb{R}^n$.

First, we have to prove the representation property

$$[T(\mathbf{a},\pi)\cdot T(\mathbf{b},\sigma)]\mathbf{e}_i$$
$$=[T(\mathbf{a},\pi)]((-1)^{b_i}\cdot\mathbf{e}_{\sigma(i)})$$
$$=(-1)^{b_i+{}_2a_{\sigma(i)}}\mathbf{e}_{\pi(\sigma(i))}=[T(\mathbf{a}_\sigma+{}_2\mathbf{b},\pi\cdot\sigma)]\mathbf{e}_i$$
$$=[T[(\mathbf{a},\pi)\cdot(\mathbf{b},\sigma)]]\mathbf{e}_i,\quad 1\leqslant i\leqslant n. \tag{5.2}$$

Furthermore, $T(\mathbf{0},id_{S_n})=Id_{\mathbb{R}^n}$ and $T(\mathbf{1},id_{S_n})=-Id_{\mathbb{R}^n}$.

Let us now calculate the representation matrices of T for the basis $\{\mathbf{e}_i\,|1\leqslant i\leqslant n\}$. From

$$[T(\mathbf{a},\pi)]\mathbf{e}_i=\sum_{j=1}^{n}\mathbf{e}_jT^j_i(\mathbf{a},\pi),$$

we obtain, by comparison with the definition of T, the formula

$$T^j_i(\mathbf{a},\pi)=(-1)^{a_i}\cdot\delta^j_{\pi(i)}. \tag{5.3}$$

Consequently, for $\mathbf{x}=\Sigma_{i=1}^n x^i\mathbf{e}_i$, we have

$$[T(\mathbf{a},\pi)]\mathbf{x}=\sum_{j=1}^{n}\left(\sum_{i=1}^{n}T^j_i(\mathbf{a},\pi)x^i\right)\mathbf{e}_j,$$

and from

$$[T(\mathbf{a},\pi)T(\mathbf{b},\sigma)]\mathbf{x}$$
$$=[T(\mathbf{a},\pi)]\sum_{i,j=1}^{n}x^i\mathbf{e}_jT^j_i(\mathbf{b},\sigma)$$
$$=\sum_{i,j,k=1}^{n}x^i\mathbf{e}_kT^k_j(\mathbf{a},\pi)T^j_i(\mathbf{b},\sigma)$$
$$=\sum_{k=1}^{n}\left(\sum_{i,j=1}^{n}T^k_j(\mathbf{a},\pi)T^j_i(\mathbf{b},\sigma)x^i\right)\mathbf{e}_k, \tag{5.4}$$

we see that the multiplication rule has now been reduced to the usual matrix multiplication.

B. Properties and character of T

At this point it is useful to prove T to be a real, faithful, orthogonal, and irreducible representation of W_n. The matrices are real by definition. The faithfulness follows from

$$T(\mathbf{a},\pi)=Id_{\mathbb{R}^n}\Leftrightarrow\bigwedge_{1\leqslant i,j\leqslant n}(-1)^{a_i}\delta^j_{\pi(i)}=\delta^j_i$$
$$\Leftrightarrow\bigwedge_{1\leqslant i\leqslant n}(a_i\equiv 0\bmod 2\ \text{and}\ \pi(i)=i)$$
$$\Leftrightarrow(\mathbf{a},\pi)=(\mathbf{0},id_{S_n})=id_{W_n}. \tag{5.5}$$

For the orthogonality it is easily checked that

$$\sum_{j=1}^{n}T^j_i(\mathbf{a},\pi)\cdot T^j_k(\mathbf{a},\pi)=(-1)^{a_i+a_k}\delta_{ik}=\delta_{ik}. \tag{5.6}$$

To prove T to be irreducible, we use the character χ_T afforded by T,

$$\chi_T(g) := \mathrm{tr}[T(g)], \quad g \in W_n. \tag{5.7}$$

Then, with

$$\chi_T(\mathbf{a},\pi) = \sum_{i=1}^{n} (-1)^{a_i} \delta^i_{\pi(i)},$$

we get

$$\begin{aligned}
&\frac{1}{|W_n|} \sum_{g \in W_n} \chi_T(g) \cdot \overline{\chi_T(g)} \\
&= \frac{1}{2^n \cdot n!} \Bigg\{ \sum_{\pi \in S_n} \sum_{\mathbf{a} \in Z_2^n} \sum_{i=1}^{n} \delta^i_{\pi(i)} \\
&\quad + \sum_{\pi \in S_n} \sum_{i \neq j} \underbrace{\sum_{\mathbf{a} \in Z_2^n} (-1)^{a_i + a_j}}_{=0} \delta^i_{\pi(i)} \delta^j_{\pi(j)} \Bigg\} \\
&= \frac{1}{n!} \sum_{i=1}^{n} \sum_{\pi \in S_n} \delta^i_{\pi(i)} = \frac{1}{n!} \sum_{i=1}^{n} (n-1)! = 1,
\end{aligned} \tag{5.8}$$

which proves the statement.

C. Useful sum rules for the characters

In Sec. IV A the ambivalency of W_n was proved. Hence, we have $\chi(g) = \chi(g^{-1})$ for every $g \in W_n$ and every irreducible character of W_n. As a consequence, we obtain the following sum rules for the canonical character χ_T:

$$\sum_{g \in W_n} \chi_T(g)\, \chi_T(gg') = \frac{n! \cdot 2^n}{n} \cdot \chi_T(g'), \tag{5.9}$$

$$\begin{aligned}
&\sum_{g_1,\ldots,g_k \in W_n} \chi_T(g_1 g_2) \cdot \chi_T(g_2 g_3) \cdot \cdots \cdot \chi_T(g_{k-1} g_k) \cdot \chi_T(g_k g_1) \\
&\quad = \begin{cases} |W_n|^k / n^{k-2}, & k \geqslant 2, \quad k \text{ even}, \\ |W_n|^k / n^{k-1}, & k \geqslant 1, \quad k \text{ odd}. \end{cases}
\end{aligned} \tag{5.10}$$

For an arbitrary irreducible character χ of W_n with dimension d these formulas read as follows:

$$\sum_{g \in W_n} \chi(g)\, \chi(gg') = (|W_n|/d) \cdot \chi(g'), \tag{5.11}$$

$$\begin{aligned}
&\sum_{g_1,\ldots,g_k \in W_n} \chi(g_1 g_2) \cdot \cdots \cdot \chi(g_k g_1) \\
&\quad = \begin{cases} |W_n|^k / d^{k-2}, & k \geqslant 2, \quad k \text{ even}, \\ |W_n|^k / d^{k-1}, & k \geqslant 1, \quad k \text{ odd}. \end{cases}
\end{aligned} \tag{5.12}$$

VI. GEOMETRICAL INTERPRETATION OF THE GROUP GENERATORS

Now, with the aid of the representation T, a geometrical interpretation of the generators α, β, and γ is possible. The representation matrices are (for $n \geqslant 2$)

$$T(\alpha) = \left(\begin{array}{cc|c} 0 & 1 & \\ 1 & 0 & 0 \\ \hline & 0 & id_{n-2} \end{array}\right), \quad T(\beta) = \left(\begin{array}{c|c} 0 & 1 \\ \hline id_{n-1} & 0 \end{array}\right),$$

$$T(\gamma) = \left(\begin{array}{c|c} -1 & 0 \\ \hline 0 & id_{n-1} \end{array}\right). \tag{6.1}$$

Thus, γ is a reflection at the hyperplane $\{\mathbf{x} \in \mathbf{R}^n | x^1 = 0\}$ and α is a reflection at the hyperplane $\{x \in \mathbf{R}^n | x^1 - x^2 = 0\}$, where $\mathbf{x} = \Sigma_{i=1}^n x^i \mathbf{e}_i$ as above.

For the generator β the two cases of even and odd space dimensions must be distinguished. First, if n is odd, i.e., $n = 2k+1$ with $k \in \mathbf{N}$, β is a pure rotation round the axis $\mathbf{u} = (1/\sqrt{n})(\mathbf{e}_1 + \mathbf{e}_2 + \cdots + \mathbf{e}_n)$ with the rotation angle $\varphi = \frac{2}{3}\pi$.

Second, if n is even, i.e., $n = 2k$ with $k \in \mathbf{N}$, β is a rotation reflection and can be written as $\beta = \gamma \cdot \beta'$, where β' is a rotation round the origin with the angle $\phi' = \pi/2$. Note that rotation round the origin means embedding of $\mathbf{R}^n$ in $\mathbf{R}^{n+1}$ and then rotation round the axis $\mathbf{u} = \mathbf{e}_{n+1}$.

This is a remarkable difference between W_{2k} and W_{2k+1} and should be kept in mind for the various calculations where the geometrical meaning is of any importance!

VII. W_n AS FINITE SUBGROUP OF O(n)

From the above considerations it is obvious that W_n is isomorphic with a finite subgroup of O(n), the isomorphism being given by the mapping

$$\begin{aligned}
T_M : \; & W_n \to \mathrm{Mat}(n,n) \\
& (\mathbf{a},\pi) \mapsto (T^j_i(\mathbf{a},\pi))_{1 \leqslant i,j \leqslant n},
\end{aligned} \tag{7.1}$$

where Mat(n,n) is the set of real $n \times n$ matrices.

Therefore, $T_M(W_n) \subset \mathrm{O}(n)$, where

$$\mathrm{O}(n) := \{A \in \mathrm{Mat}(n,n) | A^t = A^{-1}\}. \tag{7.2}$$

A short look back shows that for every $(\mathbf{a},\pi) \in W_n$ the matrix $(T^j_i(\mathbf{a},\pi))_{1 \leqslant i,j \leqslant n}$ contains only integers, so that W_n is a subgroup of

$$\mathrm{O}(n,\mathbf{Z}) := \{A \in \mathrm{O}(n) | \bigwedge_{1 \leqslant i,j \leqslant n} A^j_i \in \mathbf{Z}\}. \tag{7.3}$$

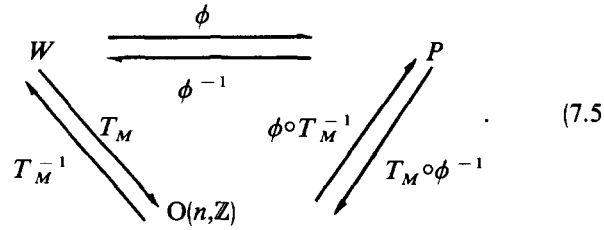
From the restrictions that all matrix elements must be integers and $A^t = A^{-1}$ we immediately obtain a special property of the matrices of O(n,$\mathbf{Z}$). In fact, they contain only the three numbers -1, 0, and $+1$, and, moreover, in every column and in every row we find $+1$ or -1 exactly once. Therefore, elementary combinatorics yield

$$\mathrm{ord}(\mathrm{O}(n,\mathbf{Z})) = n! \cdot 2^n = \mathrm{ord}(W_n)$$

with the consequence that

$$W_n \cong \mathrm{O}(n,\mathbf{Z}). \tag{7.4}$$

Thus, we have the following diagram for the isomorphisms between W_n, P_n, and O(n,$\mathbf{Z}$):



(7.5)

VIII. PROPERTIES OF THE CANONICAL CHARACTER

A. Number of elements with equal character

In chiral models as well as in gauge theories on lattices a detailed knowledge of the characters used is necessary.

One interesting problem is the calculation of the numbers of elements $g \in W_n$ with $\chi_T(g) = \alpha$, $\alpha \in \mathbb{R}$, i.e., the calculation of

$$G(n,k) := \mathrm{ord}\{g \in W_n \mid \chi_T(g) = k\}. \tag{8.1}$$

From the above considerations follows immediately $G(n,k) = 0$ if $k \notin \mathbb{Z}$ or $k > n$.

Furthermore, one can see

$$G(n,n) = 1, \quad G(n,n-1) = 0, \quad \text{and} \quad G(n,k) = G(n,-k).$$

Using the number f_n of permutations $\pi \in S_n$ without cycles of length one, which turns out to be

$$f_n = n! \cdot \sum_{k=0}^{n} \frac{(-1)^k}{k!}, \tag{8.2}$$

one can get with some combinatorics the following general formula for $G(n,k)$, $-n \leqslant k \leqslant n$,

$$G(n,k) = n! 2^{n-|k|} \sum_{l=0}^{[(n-|k|)/2]} \frac{1}{2^{2l} \cdot l! \cdot (l+|k|)!} \times \sum_{m=0}^{n-|k|-2l} \frac{(-1)^m}{m!}. \tag{8.3}$$

The relation $\Sigma_{k=-n}^{n} G(n,k) = \mathrm{ord}(W_n)$ is fulfilled, as it must be. For $1 \leqslant n \leqslant 15$ the numbers $G(n,k)$ are listed in Table III.

B. Multiplicity of the trivial representation in Kronecker products of T with itself

Let us now look at the sum

$$\frac{1}{n! 2^n} \sum_{g \in W_n} [\chi_T(g)]^m,$$

which is nothing but the multiplicity of the trivial representation in the m-fold Kronecker product of the canonical representation T with itself. Clearly, the sum equals one for $m = 0$ and vanishes for $m = 2k+1$, $k \in \mathbb{N}_0$. For $m = 2k$, $k \in \mathbb{N}$, we define

$$A(n,k) := \frac{1}{n! \cdot 2^n} \cdot \sum_{g \in W_n} [\chi_T(g)]^{2k}, \tag{8.4}$$

and obtain, after some nontrivial combinatorics, the formula

$$A(n,k) = \sum_{l=1}^{\min\{k,n\}} \frac{1}{l!} \cdot \sum_{\substack{\mu_1,\ldots,\mu_l \geqslant 1 \\ \mu_1 + \cdots + \mu_l = k}} \frac{(2k)!}{(2\mu_1)! \cdot \cdots \cdot (2\mu_l)!}. \tag{8.5}$$

Note that $A(n,k)$ is independent of n for $k \leqslant n$.

In order to simplify the above formula we look at the generating function

$$\begin{aligned} f(n,z) &:= \sum_{k>0} A(n,k) \frac{z^{2k}}{(2k)!} \\ &= \frac{1}{n! \cdot 2^n} \sum_{g \in W_n} \sum_{k>0} \frac{[\chi_T(g)]^{2k} \cdot z^{2k}}{(2k)!} \\ &= \frac{1}{2^n \cdot n!} \sum_{g \in W_n} \cosh(z \cdot \chi_T(g)) \\ &= \frac{1}{n! 2^n} \sum_{k=-n}^{n} G(n,k) \cdot \cosh(k \cdot z) \\ &= 1 + \sum_{k>1} \left\{ \frac{1}{n! 2^n} \cdot 2 \sum_{l=1}^{n} G(n,l) \cdot l^{2k} \right\} \frac{z^{2k}}{(2k)!}. \end{aligned} \tag{8.6}$$

TABLE III. Number of elements $g \in W_n$ with $\chi_T(g) = k$ for $1 \leqslant n \leqslant 15$ and $0 < k < n$.

N/K	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	1														
2	6	0	1													
3	16	15	0	1												
4	198	64	28	0	1											
5	1 728	850	160	45	0	1										
6	21 660	9 408	2 415	320	66	0	1									
7	299 808	134 715	31 808	5 481	560	91	0	1								
8	4 809 798	2 144 000	516 936	83 328	10 780	896	120	0	1							
9	86 527 744	38 635 758	9 273 024	1 518 468	185 472	19 188	1 344	153	0	1						
10	1 730 739 636	772 547 200	185 586 450	30 291 840	3 748 200	368 256	31 725	1 920	190	0	1					
11	38 075 566 560	16 996 685 046	4 082 404 480	666 740 250	82 289 856	8 176 245	671 616	49 555	2 640	231	0	1				
12	913 816 301 596	407 917 944 960	97 979 668 776	16 000 458 496	1 975 692 015	195 883 776	16 253 380	1 146 816	73 986	3 520	276	0	1			
13	23 759 213 440 896	10 605 876 226 660	2 547 463 662 016	416 017 232 631	51 364 867 840	5 094 540 451	421 930 080	30 047 446	1 857 856	106 470	4 576	325	0	1		
14	665 258 016 461 688	296 964 496 904 320	71 329 012 958 203	11 648 461 039 488	1 438 229 422 630	142 640 225 728	11 817 150 345	840 144 448	52 396 708	2 882 880	148 603	5 824	378	0	1	
15	19 957 740 338 733 120	8 908 935 052 552 275	2 139 870 268 986 240	349 453 917 677 865	43 146 828 051 600	4 279 236 816 855	354 500 203 200	25 210 186 365	1 569 865 440	87 080 175	4 315 584	202 125	7 280	435	0	1
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

With the above relation (8.3) for $G(n,l)$ a comparison of coefficients yields another formula for $A(n,k)$

$$A(n,k) = 2\sum_{l=1}^{n} \frac{l^{2k}}{2^l} \sum_{j=0}^{[(n-l)/2]} \frac{1}{2^{2j} j! \cdot (l+j)!} \times \sum_{m=0}^{n-l-2j} \frac{(-1)^m}{m!}. \tag{8.7}$$

By the special property of the numbers $A(n,k)$, one expects the generating functions $f(n,z)$ to build a sequence of functions that converges uniformly on every compact subset of the complex plane and, indeed, one finds the surprisingly simple result

$$\lim_{n\to\infty} f(n,z) = \exp(\cosh(z) - 1). \tag{8.8}$$

For a proof, look at the power series

$$\begin{aligned}
&\exp(\cosh(z) - 1) \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} (\cosh(z) - 1)^k \\
&= 1 + \sum_{k=1}^{\infty} \left\{ \sum_{l=1}^{k} \frac{(2k)!}{l!} \times \sum_{\substack{\mu_1,\ldots,\mu_l \geq 1 \\ \mu_1 + \cdots + \mu_l = k}} \frac{1}{(2\mu_1)! \cdot \cdots \cdot (2\mu_l)!} \right\} \cdot \frac{z^{2k}}{(2k)!},
\end{aligned} \tag{8.9}$$

and compare this with the results on $A(n,k)$. This may be of some importance for mean field calculations with the hypercubical group on lattices since

$$f(n,z) = \frac{1}{n! \cdot 2^n} \sum_{g \in W_n} e^{z \cdot \chi_T(g)}, \tag{8.10}$$

and

$$\exp(\cosh(z) - 1) - f(n,z) = 0(z^{2n+2}), \quad \text{for } z \to 0. \tag{8.11}$$

C. Power series of exp(cosh (z) − 1)

Additionally, the power series of $\exp(\cosh(z) - 1)$ can easily be calculated and starts with

$$\begin{aligned}
&\exp(\cosh(z) - 1) \\
&= 1 + \frac{z^2}{2!} + 4\frac{z^4}{4!} + 31\frac{z^6}{6!} + 379\frac{z^8}{8!} \\
&\quad + 6556\frac{z^{10}}{10!} + 150\,349\frac{z^{12}}{12!} + 4\,373\,461\frac{z^{14}}{14!} \\
&\quad + 156\,297\,964\frac{z^{16}}{16!} + 6\,698\,486\,371\frac{z^{18}}{18!} \\
&\quad + 337\,789\,490\,599\frac{z^{20}}{20!} + \cdots .
\end{aligned} \tag{8.12}$$

This is an explicit calculation of the first coefficients of the general formula

$$\exp(\cosh(z) - 1) = 1 + \sum_{k=1}^{\infty} A(k,k) \frac{z^{2k}}{(2k)!}. \tag{8.13}$$

Apart from the fact that even in special tables on power series one can only find the first three or four coefficients this result is a nice example of the mathematical connections between group theory and analysis.

IX. THE ADJOINED REPRESENTATION

A. Construction from $T \otimes T$

Although the canonical representation T is without any doubt the most important representation of the hyperoctahedral group, we should throw a short glance at the other ones, especially at the so-called adjoined representation, which will now be constructed from the twofold Kronecker product of the canonical representation T with itself, $T \otimes T$, which is a representation of W_n on the space $\mathbb{R}^n \otimes \mathbb{R}^n = \mathbb{R}^{n^2}$.

Using the standard ON basis $\{\mathbf{e}_i \mid 1 \leqslant i \leqslant n\}$ for $\mathbb{R}^n$ we find $\{\mathbf{e}_i \otimes \mathbf{e}_j \mid 1 \leqslant i,j \leqslant n\}$ as a new basis of $\mathbb{R}^n \otimes \mathbb{R}^n$. Furthermore, three invariant subspaces can be obtained. From

$$\begin{aligned}
&T \otimes T(\mathbf{e}_i \otimes \mathbf{e}_j - \mathbf{e}_j \otimes \mathbf{e}_i) \\
&= \sum_{k,l=1}^{n} \mathbf{e}_k \otimes \mathbf{e}_l T_i^k T_j^l - \mathbf{e}_l \otimes \mathbf{e}_k T_j^l T_i^k \\
&= \sum_{k<l} (\mathbf{e}_k \otimes \mathbf{e}_l - \mathbf{e}_l \otimes \mathbf{e}_k)(T_i^k T_j^l - T_i^l T_j^k),
\end{aligned} \tag{9.1}$$

for $i<j$, the skew-symmetric subspace $V^{(-)}$,

$$V^{(-)} := \langle\langle (1/\sqrt{2})(\mathbf{e}_i \otimes \mathbf{e}_j - \mathbf{e}_j \otimes \mathbf{e}_i) \mid 1 \leqslant i < j \leqslant n \rangle\rangle, \tag{9.2}$$

turns out to be invariant with $\dim V^{(-)} = \frac{1}{2}n(n-1)$.

In an analogous manner it follows that the symmetric subspace $\tilde{V}^{(+)}$ of dimension $\frac{1}{2}n(n+1)$, defined by

$$\tilde{V}^{(+)} := \langle\langle (1/\sqrt{2})(\mathbf{e}_i \otimes \mathbf{e}_j + \mathbf{e}_j \otimes \mathbf{e}_i) \mid 1 \leqslant i \leqslant j \leqslant n \rangle\rangle,$$

is invariant under the representation $T \otimes T$.

But $\tilde{V}^{(+)}$ cannot be irreducible because the orthogonality of the matrices T_j^i implies

$$\begin{aligned}
[T \otimes T]\left(\sum_{k=1}^{n} \mathbf{e}_k \otimes \mathbf{e}_k\right) &= \sum_{l,m} \mathbf{e}_l \otimes \mathbf{e}_m \underbrace{\sum_k T_k^l T_k^m}_{\delta^{lm}} \\
&= \sum_{m=1}^{n} \mathbf{e}_m \otimes \mathbf{e}_m.
\end{aligned} \tag{9.3}$$

Thus,

$$V^{(0)} := \langle\langle (1/\sqrt{n})(\mathbf{e}_1 \otimes \mathbf{e}_1 + \cdots + \mathbf{e}_n \otimes \mathbf{e}_n) \rangle\rangle \tag{9.4}$$

is a one-dimensional invariant subspace of $\mathbb{R}^n \otimes \mathbb{R}^n$ and of $\tilde{V}^{(+)}$ as well. Obviously, $T \otimes T$ acts trivially on $V^{(0)}$, so that $T \otimes T|_{V^{(0)}}$ is the identity representation.

Now, with $V^{(+)} := \tilde{V}^{(+)} - V^{(0)} = (V^{(-)} \oplus V^{(0)})^{\perp}$, we define two new representations

$$\begin{aligned}
&T_A : W_n \to V^{(-)} \subset \mathbb{R}^n \otimes \mathbb{R}^n, \\
&\quad g \mapsto T_A(g) := T \otimes T|_{V^{(-)}}(g),
\end{aligned} \tag{9.5}$$

$$\begin{aligned}
&T_S : W_n \to V^{(+)} \subset \mathbb{R}^n \otimes \mathbb{R}^n, \\
&\quad g \mapsto T_s(g) = T \otimes T|_{V^{(+)}}(g),
\end{aligned} \tag{9.6}$$

where the subscripts A and S stand for antisymmetric and symmectric, respectively. If T_0 denotes the trivial representation, we get the relation

$$T \otimes T = T_0 \oplus T_A \oplus T_s, \tag{9.7}$$

where nothing is said about the irreducibility of T_A or T_s. Here T_A is the so-called adjoined representation of W_n.

Since we only used the orthogonality of the representation matrices, the last relation is also valid for the appropriate representation of O(n).

B. Properties of the adjoined representation

Let us now examine the representation T_A. From

$$T_A(\mathbf{e}_i \otimes \mathbf{e}_j - \mathbf{e}_j \otimes \mathbf{e}_i) = \sum_{k<l} (\mathbf{e}_k \otimes \mathbf{e}_l - \mathbf{e}_l \otimes \mathbf{e}_k) T_{A\,ij}^{\;kl}, \quad (9.8)$$

we obtain the relation

$$T_{A\,ij}^{\;kl} = T_i^k \cdot T_j^l - T_i^l \cdot T_j^k, \quad (9.9)$$

and, for the appropriate character $\chi_A := \operatorname{tr} T_A$,

$$\chi_A(g) = \tfrac{1}{2}\{[\chi_T(g)]^2 - \chi_T(g^2)\}. \quad (9.10)$$

Obviously, T_A is real because T is real. With the last relation one can prove the irreducibility of T_A,

$$\frac{1}{\operatorname{ord}(W_n)} \sum_{g \in W_n} |\chi_A(g)|^2 = 1, \quad n \geqslant 2, \quad (9.11)$$

using the following formulas:

$$\frac{1}{\operatorname{ord}(W_n)} \sum_{g \in W_n} [\chi_T(g)]^4 = 4, \quad n \geqslant 2, \quad (9.12)$$

$$\frac{1}{\operatorname{ord}(W_n)} \sum_{g \in W_n} [\chi_T(g^2)]^2 = 4, \quad n \geqslant 2, \quad (9.13)$$

$$\frac{1}{\operatorname{ord}(W_n)} \sum_{g \in W_n} [\chi_T(g)]^2 \chi_T(g^2) = 2, \quad n \geqslant 2. \quad (9.14)$$

As a very important consequence, we find that the O(n)-representation $\tilde{T}_A$, constructed by the analogous algorithm from the fundamental O(n)-representation $\tilde{T}$ stays irreducible after restriction to the finite subgroup W_n and thus has to be irreducible itself.

In spite of this relation to the appropriate representation of O(n), the representation T_A is less important than the canonical representation T because T_A is not faithful which is an obvious consequence of its construction from the Kronecker product. Furthermore, for n even, T_A generally decomposes into two parts after restriction to the subgroup SW_n of pure rotations. Thus, a more detailed investigation of T_A is omitted here.

X. SOME REMARKS ON THE COMPLETE SYSTEM OF IRREDUCIBLE REPRESENTATIONS

A. One-dimensional representations

Before proceeding to a general classification of the W_n-representation, let us look at the one-dimensional ones. For $n \geqslant 2$, we have at least four representations of degree 1 that are clearly identical with their characters and therefore denoted by

$$\chi_k^{(1)}, \quad k = 1,\ldots,4.$$

They are

$$\chi_1^{(1)}(\mathbf{a},\pi) := 1 \quad (\text{"identity"}), \quad (10.1)$$

$$\chi_2^{(1)}(\mathbf{a},\pi) := \operatorname{sgn}(\pi) \quad (\text{"signum } S_n\text{"}), \quad (10.2)$$

$$\chi_3^{(1)}(\mathbf{a},\pi) := \operatorname{sgn}[\Phi(\mathbf{a},\pi)] \quad (\text{"signum } S_{2n}\text{"}), \quad (10.3)$$

$$\chi_4^{(1)}(\mathbf{a},\pi) := \det[T(\mathbf{a},\pi)] \quad (\text{"determinant"}). \quad (10.4)$$

The definition of T leads to the relation

$$\det[T(\mathbf{a},\pi)] = (-1)^{a_1 + \cdots + a_n} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \delta_{\pi(1)}^{\sigma(1)} \cdot \cdots \cdot \delta_{\pi(n)}^{\sigma(n)}$$

$$= (-1)^{a_1 + \cdots + a_n} \operatorname{sgn}(\pi). \quad (10.5)$$

If $\pi \in S_n$ has the cycle structure $(1^{\mu_1},\ldots,n^{\mu_n})$, where $\sum_{k=1}^n k\mu_k = n$, one gets $\operatorname{sgn}(\pi) = (-1)^{n-(\mu_1 + \cdots + \mu_n)}$. Applying this to the permutation $\Phi(\mathbf{a},\pi) \in S_{2n}$, we obtain the remarkable result

$$\operatorname{sgn}[\Phi(\mathbf{a},\pi)] = (-1)^{a_1 + \cdots + a_n}, \quad (10.6)$$

which implies a simple calculation of the "determinant" as a class function,

$$\det[T(\mathbf{a},\pi)] = \operatorname{sgn}(\pi) \cdot \operatorname{sgn}[\Phi(\mathbf{a},\pi)]. \quad (10.7)$$

Additionally, we have the following relations between our four one-dimensional representations:

$$\chi_1^{(1)} \cdot \chi_k^{(1)} = \chi_k^{(1)}, \quad \text{for any } k. \quad (10.8)$$

$$\chi_k^{(1)} \cdot \chi_k^{(1)} = \chi_1^{(1)}, \quad \text{for any } k. \quad (10.9)$$

$$\chi_2^{(1)} \cdot \chi_3^{(1)} = \chi_4^{(1)}, \quad \text{and all permutations of (2,3,4).} \quad (10.10)$$

B. Extension of S_n-representation to W_n

The representations of S_n are well known and can easily be extended to representations of W_n. If $\tilde{D}$ is a representation of S_n over V, we define

$$D: \quad W_n \to V,$$

$$(\mathbf{a},\pi) \mapsto D(\mathbf{a},\pi) := \tilde{D}(\pi). \quad (10.11)$$

The representation property is a straightforward consequence of the multiplication rule in W_n

$$D((\mathbf{a},\pi) \cdot (\mathbf{b},\sigma)) = D(\mathbf{a}_\sigma + {}_2\mathbf{b}, \pi\sigma) = \tilde{D}(\pi\sigma)$$

$$= \tilde{D}(\pi) \cdot \tilde{D}(\sigma) \quad (10.12)$$

$$= D(\mathbf{a},\pi) \cdot D(\mathbf{b},\sigma),$$

but none of these representations can be faithful.

An application of these representations without mixing up, for example, with the canonical representation reduces the symmetry to a much poorer one, especially the link to O(n) is lost.

C. A guide to the general classification of representations

Let us now proceed to a short survey of the complete system of irreducible representations of W_n. Starting with the two inequivalent, irreducible representations of Z_2, called D_0 and D_1, one obtains a complete system of pairwise inequivalent, irreducible representations of Z_2^n in the form $D_{i_1} \otimes \cdots \otimes D_{i_n}$, where $\otimes$ denotes the outer tensor product and $i_1,\ldots,i_n \in \{0,1\}$. If we combine the subscripts $i_1,\ldots,i_n$ to a vector $\mathbf{I} = (i_1,\ldots,i_n)^t$, we have 2^n different vectors (with components 0 or 1) that label the different representations of Z_2^n. Thus we use the notation $D_{\mathbf{I}} := D_{i_1} \cdot \cdots \cdot D_{i_n}$, where the tensor product is reduced to the usual multiplication since all appearing representations are of dimension 1.

We now define a subgroup of S_n

$$S_{(\mathbf{I})} := \{\pi \in S_n \mid \pi(\mathbf{I}) := (i_{\pi(1)},\ldots,i_{\pi(n)})^t = \mathbf{I}\} \subset S_n. \quad (10.13)$$

Together with

$n_1 = i_1 + i_2 + \cdots + i_n$ ($\hat{=}$ number of ones of **I**),

$n_0 = n - n_1$ ($\hat{=}$ number of zeros of **I**),

and the definitions

$$S_{(n_0)} := \{\pi \in S_n \mid \pi(k) = k, \quad \text{if } i_k = 1, \quad k = 1,\ldots,n\} \subset S_n,$$

$$S_{(n_1)} := \{\pi \in S_n \mid \pi(k) = k, \quad \text{if } i_k = 0, k = 1,\ldots,n\} \subset S_n,$$

one finds the following relations:

$$\begin{aligned} &S_{(n_0)} \cong S_{n_0}, \\ &S_{(n_1)} \cong S_{n_1}, \\ &S_{(\mathbf{I})} = S_{(n_0)} \cdot S_{(n_1)}, \\ &\operatorname{ord}(S_{(\mathbf{I})}) = n_0! \cdot n_1!. \end{aligned} \tag{10.14}$$

Furthermore, if $D^{(n_0)}$ and $D^{(n_1)}$ label irreducible representations of $S_{(n_0)}$ and $S_{(n_1)}$, respectively, we get all irreducible representations of $S_{(\mathbf{I})}$ in the form

$$D^{(n_0,n_1)} := D^{(n_0)} \otimes D^{(n_1)}. \tag{10.15}$$

The change of $D^{(n_0,n_1)}$ to a representation of the wreath product $Z_2 \sim S_{(\mathbf{I})} \subset Z_2 \sim S_n = W_n$ is the same as explained in Sec. X A,

$$\overset{\bullet}{D}{}^{(n_0,n_1)}(\mathbf{a},\pi) := D^{(n_0,n_1)}(\pi), \quad (\mathbf{a},\pi) \in Z_2 \sim S_{(\mathbf{I})}. \tag{10.16}$$

On the other hand, a change of $D_{(\mathbf{I})}$ to a representation of $Z_2 \sim S_{(\mathbf{I})}$ can be done by the definition

$$\overset{\bullet}{D}{}_{\mathbf{I}}(\mathbf{a},\pi) := D_{\mathbf{I}}(\mathbf{a}) = D_{i_1}(a_1) \cdot \cdots \cdot D_{i_n}(a_n). \tag{10.17}$$

Thus, $\overset{\bullet}{D}{}_{\mathbf{I}}(\mathbf{a},\pi) = \overset{\bullet}{D}{}_{\mathbf{I}}(\mathbf{a},id_{S_n})$, and $\overset{\bullet}{D}{}_{\mathbf{I}}(\mathbf{a},\pi)\cdot\overset{\bullet}{D}{}_{\mathbf{I}}(\mathbf{b},\sigma) = \overset{\bullet}{D}{}_{\mathbf{I}}(\mathbf{a}_\sigma + {}_2\mathbf{b},\pi\sigma)$.

Note that $\pi,\sigma \in S_{(\mathbf{I})}$ is a necessary condition for this extension which explains the definition of $S_{\mathbf{I}}$. We can now combine $\overset{\bullet}{D}{}_{\mathbf{I}}$ and $\overset{\bullet}{D}{}^{(n_0,n_1)}$ to another representation of $Z_2 \sim S_{(\mathbf{I})}$:

$$(\overset{\bullet}{D}{}_{\mathbf{I}},\overset{\bullet}{D}{}^{(n_0,n_1)})(\mathbf{a},\pi) := \overset{\bullet}{D}{}_{\mathbf{I}}(\mathbf{a},id_{S_n}) \otimes \overset{\bullet}{D}{}^{(n_0,n_1)}(\mathbf{0},\pi). \tag{10.18}$$

At this point, one has to extend these representations of $Z_2 \sim S_{(\mathbf{I})}$ to representations of $Z_2 \sim S_n$, using a complete system of representatives $\eta_1,\ldots,\eta_r$, $r = n!/n_0! \cdot n_1!$, of the group $S_{(\mathbf{I})}$ as left coset in S_n and Clifford's theory.

Since this is done in many textbooks on the representation theory of wreath products (see, for example, Kerber[3] or Osima),[9] it is omitted here. If $(\overset{\bullet}{D}{}_{\mathbf{I}},\overset{\bullet}{D}{}^{(n_0,n_1)})^*$ denote the extended representations, the following result can be stated (for a proof see Kerber[6]).

If $D_{\mathbf{I}}$ runs through a complete system of irreducible $Z_2{}^n$-representations so that every possible pair (n_0,n_1) appears exactly once, and, while **I** is kept fixed, $D^{(n_0,n_1)}$ runs through the irreducible representations of $S_{(\mathbf{I})}$, then $(\overset{\bullet}{D}{}_{\mathbf{I}},\overset{\bullet}{D}{}^{(n_0,n_1)})^*$ runs once through a complete system of pairwise inequivalent and irreducible representations of W_n.

As is well known, the number of S_m-representations is given by the number of partitions of m. Hence, with $p(0) := 1$ we have $p(n_0)\cdot p(n_1)$ representations of $S_{(\mathbf{I})}$. This yields the following formula for the number Q_n^* of W_n — representations:

$$Q_n^* = \sum_{k=0}^{n} p(k)\cdot p(n-k). \tag{10.19}$$

Since $Q_n^* = Q_n$ (number of conjugacy classes), we obtain

$$\sum_{k=0}^{n} p(k)\cdot p(n-k) = \sum_{\substack{\mu_1,\ldots,\mu_n > 0 \\ \sum_{k=1}^{n} k\cdot\mu_k = n}} \left[\prod_{l=1}^{n} (\mu_l + 1)\right]. \tag{10.20}$$

By construction, $(\overset{\bullet}{D}{}_{\mathbf{I}},\overset{\bullet}{D}{}^{(n_0,n_1)})^*$ is of dimension $(n!/n_0!n_1!)\cdot\dim(D^{(n_0,n_1)})$. As a crosscheck, one can calculate

$$\begin{aligned} &\sum_D [\dim(D)]^2 \\ &\quad = \sum_{k=0}^{n} \left(\frac{n!}{k!(n-k)!}\right)^2 \cdot \sum_R [\dim(R)]^2 \cdot \sum_T [\dim(T)]^2 \\ &\quad = \sum_{k=0}^{n} \left(\frac{n!}{k!(n-k)!}\right)^2 \cdot k! \cdot (n-k)! \\ &\quad = n! \cdot \sum_{k=0}^{n} \binom{n}{k} = n! \cdot 2^n = |W_n| \end{aligned} \tag{10.21}$$

(where D is the irreducible representation of W_n, R is the irreducible representation of S_k, and T is the irreducible representation of S_{n-k}), as is necessary.

ACKNOWLEDGMENTS

I am very grateful to Professor V. Rittenberg for helpful discussions and substantial support. Moreover, I wish to thank P. Reinicke for many useful hints.

APPENDIX: A PROOF OF THE CLASSIFICATION THEOREM FOR THE CONJUGACY CLASSES

In the main text it was stated that two elements $(\mathbf{a},\pi)$ and $(\mathbf{b},\sigma)$ of W_n are conjugate to each other if and only if π and σ have the same cycle structure in S_n *and* $\Phi(\mathbf{a},\pi)$ and $\Phi(\mathbf{b},\sigma)$ have the same cycle structure in S_{2n}. One direction of the proof is almost trivial.

If $(\mathbf{a},\pi) \sim_{W_n} (\mathbf{b},\sigma)$, i.e., $(\mathbf{b},\sigma) = (\mathbf{c},\rho)(\mathbf{a},\pi)(\mathbf{c},\rho)^{-1}$ for some $(\mathbf{c},\rho) \in W_n$, we obtain

$$\Phi(\mathbf{b},\sigma) = \Phi(\mathbf{c},\rho)\cdot\Phi(\mathbf{a},\pi)\cdot(\Phi(\mathbf{c},\rho))^{-1}$$

by the group isomorphism property of Φ and $\sigma = \rho\pi\rho^{-1}$ by application of the multiplication rule of W_n. Thus, this part of the proof is complete. In order to prove the other direction we first bear in mind that $\pi \sim_{S_n} \sigma$, i.e., $\pi = \rho\sigma\rho^{-1}$ for some $\rho \in S_n$, implies $(\mathbf{b},\sigma) \sim_{W_n} (\mathbf{0},\rho)(\mathbf{b},\sigma)(\mathbf{0},\rho^{-1}) = (\mathbf{b}_{\rho^{-1}},\pi)$. With $\mathbf{c} := \mathbf{b}_{\rho^{-1}}$ we now only have to prove that $\Phi(\mathbf{a},\pi) \sim_{S_{2n}} \Phi(\mathbf{c},\pi)$ implies $(\mathbf{a},\pi) \sim_{W_n} (\mathbf{c},\pi)$. Let π have the cycle structure $(1^{\mu_1},2^{\mu_2},\ldots,n^{\mu_n})$, $\sum_{k=1}^{n} k\cdot\mu_k = n$. Now we ask what happens to a cycle of length q under the action of the group isomorphism Φ and find that this is determined by exactly those q components of $\mathbf{a}$ the subscripts of which appear in the considered cycle. The cycle becomes doubled in number if the number of ones at the relevant positions of $\mathbf{a}$ is even, and doubled in length if this number is odd.

Now, we can make this number 0 or 1 since every pair of 1's in the range of our cycle can be cleared by an equivalence transformation of the form

$$(\mathbf{a},\pi)\sim_{W_n}(\mathbf{d},id\,)(\mathbf{a},\pi)(\mathbf{d},id\,)^{-1}=(\mathbf{d}_\pi+{}_2\mathbf{d}+{}_2\mathbf{a},\pi).$$

Explicitly, if $(m_1m_2...m_q)$ is the considered cycle and $a_{m_j}=a_{m_k}=1$, $j\neq k$, one has to choose $\mathbf{d}$ so that $d_k=0$, $k\notin\{m_1,...m_q\}$, and

$$d_{\pi(l)}+d_l=\begin{cases}1, & \text{if } l=m_j \text{ or } l=m_k,\\ 0, & \text{otherwise,}\end{cases}$$

which is always possible.

Next, if we have a single "1" in the range of our cycle $(m_1m_2\cdots m_q)$, we can move it to any other position in the range of the cycle, which is done by the same transformation as for the clearing of pairs of 1's.

Furthermore, if $\mathbf{a}$ has two cycles of length q, let us say $(m_1m_2\cdots m_q)$ and $(n_1n_2\cdots n_q)$, and $a_{m_1}=1$, $a_k=0$ for $k\in\{m_2,...m_q,n_1,...,n_q\}$, we can move the 1 from one cycle to the other so that afterwards a_{m_1} and a_{n_1} are interchanged. This is done by the transformation

$$(\mathbf{a},\pi)\sim_{W_n}(\mathbf{0},\rho)(\mathbf{a},\pi)(\mathbf{0},\rho^{-1})$$

with

$$\rho=\begin{pmatrix}m_1m_2\cdots m_{q-1}m_q n_1n_2...n_{q-1}n_q\\ n_qn_1\cdots n_{q-2}n_{q-1}m_1m_2\cdots m_{q-1}m_q\end{pmatrix},$$

where $\rho(k)=k$ if $k\notin\{m_1,...,m_q,n_1,...,n_q\}$ and $\rho\pi\rho^{-1}=\pi$, as it must be.

Last but not least let us again look at the two elements $(\mathbf{a},\pi)$ and $(\mathbf{c},\pi)$. By construction, $\Phi(\mathbf{a},\pi)$ and $\Phi(\mathbf{c},\pi)$ have the same cycle structure in S_{2n}. This guarantees that our clearing algorithm produces the same number of 1's after application to both elements.

Next, we can move the 1's to equivalent positions and to corresponding cycles with the transformations described above.

Hence, we find by construction

$$(\mathbf{a},\pi)\sim_{W_n}(\mathbf{f},\pi)\sim_{W_n}(\mathbf{c},\pi),$$

since our algorithm arrives at the same element $\mathbf{f}\in Z_2^n$ for $(\mathbf{a},\pi)$ and $(\mathbf{c},\pi)$, respectively. From the beginning we have $(\mathbf{c},\pi)\sim_{W_n}(\mathbf{b},\sigma)$ and consequently

$$(\mathbf{a},\pi)\sim_{W_n}(\mathbf{b},\sigma),$$

which completes the proof.

[1]M. Creutz, Phys. Rev. D **21**, 2308 (1980).
[2]D. Petcher and D. H. Weingarten, Phys. Rev. D**22**, 2465 (1980).
[3]C. Lovelace, "University at Large N," Rutgers University Preprint, RU-82-01.
[4]G. de B. Robinson, Proc. Cambridge Philos. Soc. **26**, 94 (1930).
[5]J. S. Frame, Nagoya Math. J. **27**, 585 (1966).
[6]A. Kerber, "Representations of Permutation Groups I," in *Lecture Notes in Mathematics* (Springer, Berlin, 1971), Vol. 240.
[7]G. James and A. Kerber, *The Representation Theory of the Symmetric Group* (Addison–Wesley, Reading, MA, 1981).
[8]S. J. Maier, J. Algebra **33**, 59 (1975).
[9]M. Osima, Math. J. Okayama Univ. **4**, 39 (1954); **6**, 81 (1956).
[10]M. Baake, B. Gemünden, and R. Oedingen, J. Math. Phys. **23**, 944 (1982).
[11]M. Baake, B. Gemünden, and R. Oedingen, J. Math. Phys. **24**, 1021 (1983).
[12]J. E. Mandula, G. Zweig, and J. Govaerts, Nucl. Phys. B **228**, 109 (1983).
[13]R. Badke, P. Reinicke, and V. Rittenberg, "The Discrete Cubic and Chiral Cubic Spin Systems in Various Dimensions and their Large N limit," preprint, Bonn-HE-84-13.
[14]M. Marcu, A. Regev, and V. Rittenberg, J. Math. Phys. **22**, 2740 (1981).
[15]M. Abramowitz and A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).

Color analysis, variational self-adjointness, and color Poisson (super)algebras

Robert Trostel
Institut für Theoretische Physik, Technische Universität Berlin, 1000 Berlin 12, West Germany and The Institute for Basic Research, Cambridge, Massachusetts 02138

(Received 19 March 1984; accepted for publication 11 May 1984)

After stating some facts concerning the calculation with "color variables" we cite some recent results of the author with respect to the color analytic extension of variational principles, self-adjointness, and Heisenberg commutation relations. As an apparent novelty, we then present the color analytic version of the Hamiltonian formalism including the construction of color Poisson brackets leading to a color (super)algebra with color derivation property.

PACS numbers: 02.30. + g, 11.10.Ef, 11.30.Pb

I. INTRODUCTION AND MAIN RESULTS

Recently, there has been great interest in generalizing the statistics in quantum field theory (e.g., Ohnuki and Kamefuchi,[1,2] Omote and Kamefuchi,[3] Levine and Tomozawa,[4] and others). This has its origin in Green's proposal[5] to use parastatistical commutation relations, being compatible with Heisenberg's equation of motion. Also in this connection, there is great interest in the recently introduced generalized superalgebras, also called color (super)algebras (e.g., Kac,[6] Rittenberg and Wyler,[7,8] Lukierski and Rittenberg,[9] and others). We refer to Agrawala,[10] Green and Jarvis,[11] Scheunert,[12] and others, for example, for the construction of Casimir invariants, tensor operators, and representations of color (super)algebras. The main idea of the color (super)algebras consists in replacing the plus or minus in the anticommutator or commutator by a complex commutation factor compatible with the grading.

In our short note, we pick up the idea of Rittenberg and Wyler,[7] for example, where they have introduced color variables in order to describe color (super)groups. But we take the color variables as the foundation of an extension of variational principles, self-adjointness, Hamiltonian formalism, Poisson brackets etc. We therefore start with an associative Γ-graded algebra $\mathbb{A}$ (over the complex numbers $\mathbb{C}$) with unit, with a finite abelian grading group Γ and equipped with a nonvanishing commutation function $\sigma: \Gamma \times \Gamma \to \mathbb{C}\{0\}$ compatible with associativity and the grading Γ. The generalization of well-known conventional supersymmetric (i.e., $\mathbb{Z}_2$-graded) theoretical physics (for instance Kostant,[13] Dell and Smolin,[14] Rogers,[15] Corwin, Ne'Eman, and Sternberg,[16] and others) consists in replacing the infinite-dimensional Banach–Grassmann algebra $\mathbb{B}_\infty$ (Rogers,[15] and Jadczyk and Pilch[17]) by $\mathbb{A}$ with an arbitrary finite abelian grading group Γ and an admissible commutation function σ. Without insisting on mathematical subtleties which can partially be found in Ref. 18, we want to collect in this note some general results. The situation if the Hessian is not regular (i.e., if one deals with constrained systems) and the globalization of the results here presented only in local coordinates will be discussed in a forthcoming paper.

We point out that the content of Secs. II and III of this article can already be found in Ref. 18. Because of mathematical homogeneity, we have decided to include Secs. II and III in this note as a preparatory step in order to familiarize the reader with color analytic calculation methods. The new results consist in the construction of a color Poisson (super)algebra and in displaying the interplay of two Γ-graded products, the ordinary associative one in $\mathbb{A}$ and the color Poisson bracket, both interrelated by the color derivation property. The color Hamilton–Jacobi differential equation is presented and also the extension to field theories is briefly indicated.

II. GENERALIZATION OF HEISENBERG'S COMMUTATION RULES

Let $\mathbb{A}$ be a Γ-graded associative Banach algebra over the complex numbers $\mathbb{C}$ and with unit[18]

$$\mathbb{A} = \bigoplus_{\gamma \in \Gamma} \mathbb{A}_\gamma, \quad \mathbb{A}_\gamma \mathbb{A}_\delta \subset \mathbb{A}_{\gamma+\delta}, \quad \forall\, \gamma,\delta \in \Gamma, \tag{1}$$

where $\mathbb{A}_\gamma$ are Banach subspaces of $\mathbb{A}$ and where $0 \in \Gamma$ denotes the neutral element in the finite abelian group Γ. Moreover $\mathbb{A}$ is characterized by a commutation function

$$\sigma: \Gamma \times \Gamma \to \mathbb{C}\setminus\{0\}, \quad \sigma(\alpha,\beta)\sigma(\beta,\alpha) = 1, \tag{2}$$

$$\sigma(\alpha,\beta)\sigma(\alpha,\gamma) = \sigma(\alpha,\beta+\gamma), \quad \forall \alpha,\beta,\gamma \in \Gamma,$$

$$q_\alpha^1 q_\beta^2 = \sigma(\alpha,\beta) q_\beta^2 q_\alpha^1, \quad \forall q_\alpha^1 \in \mathbb{A}_\alpha, q_\beta^2 \in \mathbb{A}_\beta. \tag{3}$$

Equation (3) can be found, for example, in Ref. 7. The condition (2) entails the relations

$$\sigma(\alpha,0) = \sigma(0,\alpha) = 1, \quad \sigma(\alpha,\alpha) = \pm 1, \tag{4}$$

$$\sigma(\alpha,\beta) = \sigma(-\alpha,-\beta) = \sigma(\beta,-\alpha) = \sigma(-\beta,\alpha), \quad \forall \alpha,\beta \in \Gamma,$$

which indicate that one deals with ordinary bosonic and fermionic numbers in each subspace $\mathbb{A}_\alpha$ of $\mathbb{A}$. Now let $[n_\Gamma] := (n_\gamma)_{\gamma\in\Gamma}$ denote a set of positive integers $n_\gamma \in \mathbb{N}_0$ satisfying $n_\gamma = n_{-\gamma}$. We introduce the Banach space (equipped with the product topology)

$$\mathbb{G}^{[n_\Gamma]} := \bigoplus_{\gamma \in \Gamma} (\mathbb{A}_\gamma)^{n_\gamma}, \tag{5}$$

being a $\mathbb{A}_0$-module. Elements of $\mathbb{G}^{[n_\Gamma]}$ are denoted by

 0022-2488/84/113183-07$02.50

$$x,y \in \mathbb{G}^{[n_\Gamma]},\ x = (q^i_\gamma)^{i\in\{1,\dots,n_\gamma\}}_{\gamma\in\Gamma},$$
$$y = (\eta^i_\gamma)^{i\in\{1,\dots,n_\gamma\}}_{\gamma\in\Gamma},\ q^i_\gamma,\eta^i_\gamma \in \mathbb{A}_\gamma\ . \tag{6}$$

We are interested in $\mathbb{A}$-valued functions f on $\mathbb{G}^{[n_\Gamma]}$,

$$f:\mathbb{G}^{[n_\Gamma]} \to \mathbb{A},\quad \text{i.e.} f \in \mathscr{F}(\mathbb{G}^{[n_\Gamma]};\mathbb{A}), \tag{7}$$

being sufficiently many times continuously $\mathbb{A}_0$-differentiable.[18] If $(\lambda \in \mathbb{R})$,

$$\left(\frac{\partial}{\partial q^i_\gamma} f\right)_x = \left(\frac{\partial f}{\partial q^i_\gamma}\right)_x,\quad \frac{d}{d\lambda} f_{x+\lambda y}|_{\lambda=0}$$
$$= \sum_{\gamma\in\Gamma}\sum_{i=1}^{n_\gamma} \eta^i_\gamma \cdot \left(\frac{\partial f}{\partial q^i_\gamma}\right)_x \tag{8}$$

denotes the left derivative of the function f (7) with respect to $q^i_\gamma \in \mathbb{A}_\gamma$ at $x \in \mathbb{G}^{[n_\Gamma]}$, we can derive the generalized or color Heisenberg's commutation rules[18]

$$\frac{\partial}{\partial q^i_\alpha}\frac{\partial}{\partial q^j_\beta} = \sigma(\alpha,\beta)\frac{\partial}{\partial q^j_\beta}\frac{\partial}{\partial q^i_\alpha},$$
$$\sigma(\alpha,\alpha) = +1 :\Rightarrow \left[\frac{\partial}{\partial q^i_\alpha}, q^j_\alpha\right]_- = \delta^j_i,$$
$$\sigma(\alpha,\alpha) = -1 :\Rightarrow \left[\frac{\partial}{\partial q^i_\alpha}, q^j_\alpha\right]_+ = \delta^j_i, \tag{9}$$
$$\alpha\neq\beta:\ \frac{\partial}{\partial q^i_\alpha} q^j_\beta = \sigma(\beta,\alpha) q^j_\beta \frac{\partial}{\partial q^i_\alpha},$$
$$\forall \alpha,\beta \in \Gamma,\ \forall i \in \{1,\dots,n_\alpha\},$$
$$\forall j \in \{1,\dots,n_\beta\}.$$

Note for the following, that

$$\frac{\partial}{\partial q^i_\alpha}(f_\beta g_\gamma) = \frac{\partial f_\beta}{\partial q^i_\alpha} g_\gamma + \sigma(\beta,\alpha) f_\beta \frac{\partial g_\gamma}{\partial q^i_\alpha} \tag{10}$$

is valid for arbitrary continuously $\mathbb{A}_0$-differentiable maps

$$f_\beta:\mathbb{G}^{[n_\Gamma]} \to \mathbb{A}_\beta,\quad g_\gamma:\mathbb{G}^{[n_\Gamma]} \to \mathbb{A}_\gamma.$$

We point out that

$$\left(\frac{\partial f_\beta}{\partial q^i_\alpha}\right)_x \in \mathbb{A}_{\beta-\alpha},\quad \frac{\partial f_\beta}{\partial q^i_\alpha} \in \mathscr{F}(\mathbb{G}^{[n_\Gamma]}, \mathbb{A}_{\beta-\alpha}) \tag{11}$$

holds.[18]

III. GENERALIZATION OF VARIATIONAL PRINCIPLES AND SELF-ADJOINTNESS

Let now $\mathbb{Q}$ denote the vector space of all sufficiently many times continuously differentiable maps from $[t_a,t_b]$ to $\mathbb{G}^{[n_\Gamma]}$, $t_a < t_b$; $t_a,t_b \in \mathbb{R}$. Elements of $\mathbb{Q}$ will be denoted by

$$\mathbf{x},\mathbf{y} \in \mathbb{Q},\ \mathbf{x}:t \to x(t) \in \mathbb{G}^{[n_\Gamma]},\quad \mathbf{y}:t \to y(t) \in \mathbb{G}^{[n_\Gamma]},$$
$$x(t) = (q^i_\gamma(t))^{i\in\{1,\dots,n_\gamma\}}_{\gamma\in\Gamma},\quad y(t) = (\eta^i_\gamma(t))^{i\in\{1,\dots,n_\gamma\}}_{\gamma\in\Gamma}. \tag{12}$$

Let L ("Lagrangian") be a sufficiently many times continuously $\mathbb{A}_0$-differentiable map from $[t_a,t_b]\times\mathbb{G}^{[n_\Gamma]}\times\cdots\times\mathbb{G}^{[n_\Gamma]}$ to $\mathbb{A}_0$. Using $\overset{(j)}{q}{}^i_\alpha(t) := d^j/dt^j\, q^i_\alpha(t)$, we then consider the action ω_0 being a map from $\mathbb{Q}$ to $\mathbb{A}_0$,

$$\omega_{0_{\mathbf{x}}} := \int_{t_a}^{t_b} dt\, L(t,q^i_\alpha(t),\dot{q}^i_\alpha(t),\dots,\overset{(n)}{q}{}^i_\alpha(t)), \tag{13}$$

and its variation $d\omega_0$ $(\lambda \in \mathbb{R})$,

$$(d\omega_0)_{\mathbf{x}}(\mathbf{y}) := \frac{d}{d\lambda}\omega_{o_{\mathbf{x}+\lambda\mathbf{y}}}|_{\lambda=0}$$
$$= \sum_{\gamma\in\Gamma}\sum_{j=1}^{n_\gamma}\int_{t_a}^{t_b} dt\,\eta^j_\gamma(t)\left(\sum_{l=0}^{n}(-1)^l\frac{d^l}{dt^l}\left(\frac{\partial L}{\partial \overset{(l)}{q}{}^j_\gamma(t)}\right)\right)_{\mathbf{x}}$$
$$+ \sum_{\gamma\in\Gamma}\sum_{j=1}^{n_\gamma}\sum_{l=1}^{n}\overset{(l-1)}{\eta}{}^j_\gamma(t)\cdot P_{l,j,-\gamma}(t)_{\mathbf{x}}|_{t_a}^{t_b}, \tag{14}$$

where

$$P_{l,j,-\gamma}(t)_{\mathbf{x}} := \sum_{k=l}^{n}(-1)^{k-l}\frac{d^{k-l}}{dt^{k-l}}\left(\frac{\partial L}{\partial \overset{(K)}{q}{}^j_\gamma(t)}\right)_{\mathbf{x}} \in \mathbb{A}_{-\gamma} \tag{15}$$

are the generalized canonical impulses (e.g., Ostrogradski).[19] The expression

$$\varepsilon_{j,-\gamma}(t)_{\mathbf{x}} := \varepsilon_{j,-\gamma}(t,q^i_\alpha(t),\dots,\overset{(2n)}{q}{}^i_\alpha(t))$$
$$:= \sum_{l=0}^{n}(-1)^e\frac{d^l}{dt^l}\left(\frac{\partial L}{\partial \overset{(l)}{q}{}^j_\gamma(t)}\right)_x \in \mathbb{A}_{-\gamma}, \tag{16}$$
$$\forall \mathbf{x} \in \mathbb{Q},\quad \forall t \in [t_a,t_b],\quad \forall j \in \{1,\dots,n_\gamma\},\ \forall \gamma \in \Gamma$$

is called the set of evolution terms with respect to the Lagrangian L. If on the contrary one only disposes of a set of evolution terms, one can search for a Lagrangian L so that (16) will be valid. In this case, the prescribed set of evolution terms has to satisfy

$$0 = \frac{\partial}{\partial \overset{(r)}{q}{}^j_\delta(t)}\varepsilon_{i,-\gamma}(t)_{\mathbf{x}} + \sigma(\delta,\gamma)\sum_{u=0}^{2n}(-1)^{u+1}\binom{u}{r}$$
$$\times\frac{d^{u-r}}{dt^{u-r}}\left(\frac{\partial}{\partial \overset{(u)}{q}{}^i_\gamma(t)}\varepsilon_{j,-\delta}(t)_{\mathbf{x}}\right)$$
$$\forall r = 0,1,\dots,2n\quad \forall \gamma,\delta \in \Gamma; \tag{17}$$
$$\forall i \in \{1,\dots,n_\gamma\}, j \in \{1,\dots,n_\delta\},\quad \forall \mathbf{x} \in \mathbb{Q},$$

where

$$\binom{u}{r} = \begin{cases} u!/r!(u-r)!, & \text{if } u \geqslant r, \\ 0, & \text{if } u < r. \end{cases}$$

The Lagrangian then reads up to boundary terms

$$L = \sum_{\gamma\in\Gamma}\sum_{j=1}^{n_\gamma} q^j_\gamma(t)\int_0^1 d\tau$$
$$\times\varepsilon_{j,-\gamma}(t,\tau q^i_\alpha(t),\dots,\tau\,\overset{(2n)}{q}{}^i_\alpha(t)). \tag{18}$$

We refer to Ref. 18 where the Lemma of Poincaré on Banach spaces is exposed and applied to the discussion of variational principles and self-adjointness including the consideration of boundary terms appearing during the variational proce-

dure. Because of the commutation factor $\sigma(\delta,\gamma)$ in front of the sum in (17), we call (17) the *generalized* variational self-adjointness conditions.

IV. GENERALIZATION OF THE HAMILTONIAN FORMALISM

According to, e.g., Ostrogradski,[19] we introduce the Hamiltonian

$$H_0 := \left(\sum_{\gamma\in\Gamma} \sum_{j=1}^{n_\gamma} \sum_{l=1}^{n} \overset{(l)}{q}{}^{j}_{\gamma}(t) \cdot P_{l;j,-\gamma}(t)_{\mathbf{x}} \right) - L. \tag{19}$$

We are now supposed to be able to solve (15) for the variables $\overset{(n)}{q}{}^{j}_{\gamma}$, $\overset{(n+1)}{q}{}^{j}_{\gamma}$, ..., $\overset{(2n-1)}{q}{}^{j}_{\gamma}$ so that our phase space $\mathbb{P} = (\mathbb{G}^{[n_\Gamma]})^{2n}$ is spanned by the variables R:

$$\mathbb{P} \ni R = (q^j_\gamma, \ldots, \overset{(n-1)}{q}{}^{j}_{\gamma}, P_{1,j,-\gamma}, \ldots, P_{n,j,-\gamma})^{j\in\{1,\ldots,n_\gamma\}}_{\Gamma\in\Gamma}. \tag{20}$$

The equations of motion on the configuration space

$$\sum_{l=0}^{n} (-1)^e \frac{d^l}{dt^l} \left(\frac{\partial L}{\partial \overset{(l)}{q}{}^{j}_{\gamma}(t)} \right)_{\mathbf{x}} = 0 \tag{21}$$

entail the generalized Hamilton equations on the color phase space $\mathbb{P}$

$$\begin{aligned} &\frac{d}{dt}(\overset{(l-1)}{q}{}^{j}_{\gamma}(t)) = \sigma(\gamma,\gamma) \frac{\partial}{\partial P_{e;j,-\gamma}(t)_{\mathbf{x}}} H_0(R(t),t), \\ &\frac{d}{dt}(P_{l;j,-\gamma}(t)_{\mathbf{x}}) = -\frac{\partial}{\partial \overset{(l-1)}{q}{}^{j}_{\gamma}(t)} H_0(R(t),t), \\ &l = 1,\ldots,n. \end{aligned} \tag{22}$$

Note, that $H_0(\cdot,t)$ is a map from $\mathbb{P}$ to $\mathbb{A}_0$. Let now $F(\cdot,t) \in \mathcal{F}(\mathbb{P};\mathbb{A})$ denote a map from the color phase space $\mathbb{P}$ to $\mathbb{A}$. The time evolution of F reads

$$\frac{d}{dt} F(R(t),t) = \frac{\partial F(R(t),t)}{\partial t} + \{H_0(\cdot,t), F(\cdot,t)\}_{R(t)},$$

$$\begin{aligned} &\{H_0(\cdot,t), F(\cdot,t)\}_R \\ &:= \sum_{\alpha\in\Gamma} \sum_{i=1}^{n_\alpha} \sum_{l=1}^{n} \left(\sigma(\alpha,\alpha) \frac{\partial H_0(r,t)}{\partial P_{l;i,-\alpha}} \frac{\partial F(R,t)}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} \right. \\ &\quad \left. - \frac{\partial H_0(R,t)}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} \frac{\partial F(R,t)}{\partial P_{l;i-\alpha}} \right). \end{aligned} \tag{23}$$

V. COLOR POISSON (SUPER)ALGEBRA

We now construct a color Poisson bracket on the associative Γ-graded algebra of maps

$$\mathcal{F}(\mathbb{P};\mathbb{A}) = \bigoplus_{\gamma\in\Gamma} \mathcal{F}(\mathbb{P};\mathbb{A}_\gamma)$$

such that $\mathcal{F}(\mathbb{P};\mathbb{A})$ is again a Γ-graded algebra under the color Poisson bracket $\{\{\ ,\ \}\}$:

$$\begin{aligned} &\{\{B_\beta, C_\gamma\}\} \\ &:= \sum_{\alpha\in\Gamma} \sum_{i=1}^{n_\alpha} \sum_{l=1}^{n} \left(\sigma(\alpha,\alpha)\sigma(\beta,\alpha) \frac{\partial B_\beta}{\partial P_{l;i,-\alpha}} \frac{\partial C_\gamma}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} \right. \\ &\quad \left. - \sigma(\alpha,\beta) \frac{\partial B_\beta}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} \frac{\partial C_\gamma}{\partial P_{l;i,-\alpha}} \right), \end{aligned} \tag{24}$$

$$\forall B_\beta \in \mathcal{F}(\mathbb{P};\mathbb{A}_\beta), \quad \forall C_\gamma \in \mathcal{F}(\mathbb{P};\mathbb{A}_\gamma).$$

We state the following rules: for $\{\{B_\beta,C_\gamma\}\} \in \mathcal{F}(\mathbb{P};\mathbb{A}_{\beta+\gamma})$, we have color antisymmetry,

$$\{\{B_\beta,C_\gamma\}\} = -\sigma(\beta,\gamma)\{\{C_\gamma,B_\beta\}\}; \tag{25}$$

the color derivation property,

$$\begin{aligned} \{\{B_\beta,C_\gamma D_\delta\}\} &= \{\{B_\beta,C_\gamma\}\}D_\delta \\ &\quad + \sigma(\beta,\gamma)C_\gamma\{\{B_\beta,D_\delta\}\}; \end{aligned} \tag{26}$$

and the color Jacobi identity,

$$\begin{aligned} &\sigma(\gamma,\alpha)\{\{\{\{A_\alpha,B_\beta\}\},C_\gamma\}\} + \sigma(\alpha,\beta)\{\{\{\{B_\beta,C_\gamma\}\},A_\alpha\}\} \\ &\quad + \sigma(\beta,\gamma)\{\{\{\{C_\gamma,A_\alpha\}\},B_\beta\}\} = 0; \end{aligned} \tag{27}$$

all of which can be verified after long and tedious but straightforward calculations using the nontrivial properties (2) of the commutation function σ. In order to show the technique, we, for example, prove (26):

$$\begin{aligned} \{\{B_\beta,C_\gamma D_\delta\}\} &\overset{(24)}{=} \sum_{\alpha\in\Gamma} \sum_{i=1}^{n_\alpha} \sum_{l=1}^{n} \left((\sigma(\alpha,\alpha)\sigma(\beta,\alpha) \frac{\partial B_\beta}{\partial P_{l;i,-\alpha}} \frac{\partial}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} (C_\gamma D_\delta) - \sigma(\alpha,\beta) \frac{\partial B_\beta}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} \frac{\partial}{\partial P_{l;i,-\alpha}} (C_\gamma D_\delta) \right) \\ &\overset{(10)}{=} \sum_{\alpha\in\Gamma} \sum_{i=1}^{n_\alpha} \sum_{l=1}^{n} \left(\sigma(\alpha,\alpha)\sigma(\beta,\alpha) \frac{\partial B_\beta}{\partial P_{l;i,-\alpha}} \left(\frac{\partial C_\gamma}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} D_\delta + \sigma(\gamma,\alpha) C_\gamma \frac{\partial D_\delta}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} \right) \right. \\ &\quad \left. - \sigma(\alpha,\beta) \frac{\partial B_\beta}{\partial \overset{(l-1)}{q}{}^{i}_{\alpha}} \left(\frac{\partial C_\gamma}{\partial P_{l;i,-\alpha}} D_\delta + \sigma(\gamma,-\alpha) C_\gamma \frac{\partial D_\delta}{\partial P_{l;i,-\alpha}} \right) \right) \overset{(3),(11)}{=} \{\{B_\beta,C_\gamma\}\}D_\delta \end{aligned}$$

$$+\sum_{\alpha\in\Gamma}\sum_{i=1}^{n_\alpha}\sum_{l=1}^{n}\left(\sigma(\alpha,\alpha)\sigma(\beta,\alpha)\sigma(\gamma,\alpha)\sigma(\beta+\alpha,\gamma)C_\gamma\frac{\partial B_\beta}{\partial P_{l;i,-\alpha}}\frac{\partial D_\delta}{\partial\overset{(l-1)}{q^i_\alpha}}-\sigma(\alpha,\beta)\sigma(\gamma,-\alpha)\sigma(\beta-\alpha,\gamma)C_\gamma\right.$$
$$\left.\times\frac{\partial B_\beta}{\partial\overset{(l-1)}{q^i_\alpha}}\frac{\partial D_\delta}{\partial P_{l;i,-\alpha}}\right)$$
$$\overset{(2)}{=}\{\{B_\beta C_\gamma\}\}D_\delta+\sum_{\alpha\in\Gamma}\sum_{i=1}^{n_\alpha}\sum_{l=1}^{n}\left(\sigma(\alpha,\alpha)\sigma(\beta,\alpha)\sigma(\beta,\gamma)C_\gamma\frac{\partial B_\beta}{\partial P_{l;i,-\alpha}}\frac{\partial D_\delta}{\partial\overset{(l-1)}{q^i_\alpha}}-\sigma(\alpha,\beta)\sigma(\beta,\gamma)C_\gamma\right.$$
$$\left.\times\frac{\partial B_\beta}{\partial\overset{(l-1)}{q^i_\alpha}}\frac{\partial D_\delta}{\partial P_{l;i,-\alpha}}\right)$$
$$=\{\{B_\beta,C_\gamma\}\}D_\delta+\sigma(\beta,\gamma)C_\gamma\{\{B_\beta,D_\delta\}\}.$$

■

Using (4), the definition of the color Poisson bracket (24) goes naturally over to (23) because of $H_0(\cdot,t)\in\mathcal{F}(\mathbb{P};\mathbb{A}_0)$.

Corollary I: $(\mathcal{F}(\mathbb{P};\mathbb{A}),\{\{\ ,\ \}\})$ is a *color (super)algebra* [Eq., (25) and (27)] with grading

$$\Gamma:\mathcal{F}(\mathbb{P};\mathbb{A})=\overset{\oplus}{\gamma\in\Gamma}\mathcal{F}(\mathbb{P};\mathbb{A}_\gamma)$$

and commutation function σ, the same function σ with which we have started in the associative Γ-graded algebra $\mathbb{A}$.

Corollary II: $(\mathcal{F}(\mathbb{P};\mathbb{A}),\cdot)$ is an associative Γ-graded algebra over the complex numbers with unit and with the same commutation function σ as in $\mathbb{A}$. The interplay of both the structures $(\mathcal{F}(\mathbb{P};\mathbb{A}),\{\{\ ,\ \}\})$ and $(\mathcal{F}(\mathbb{P};\mathbb{A}),\cdot)$ is determined by the color derivation property (26).

Corollary III: By virtue of the color derivation property (26), the time evolution constraint is identically satisfied; $F(\cdot,t),G(\cdot,t)\in\mathcal{F}(\mathbb{P};\mathbb{A})$:

$$\frac{d}{dt}(F\cdot G)(R(t),t)=\frac{dF(R(t),t)}{dt}G(R(t),t)+F(R(t),t)\cdot\frac{dG(R(t),t)}{dt};$$
$$\{H_0(\cdot,t),F\cdot G(\cdot,t)\}=\{H_0(\cdot,t),F(\cdot,t)\}\cdot G(\cdot,t)+F(\cdot,t)\{H_0(\cdot,t),G(\cdot,t)\},\tag{28}$$

for all Hamiltonians $H_0(\cdot,t)\in\mathcal{F}(\mathbb{P};\mathbb{A}_0)$. Also

$$\frac{d}{dt}\{\{B_\beta,C_\gamma\}\}=\left\{\left\{\frac{dB_\beta}{dt},C_\gamma\right\}\right\}+\left\{\left\{B_\beta,\frac{dC_\gamma}{dt}\right\}\right\},$$
$$\{\{H_0,\{\{B_\beta,C_\gamma\}\}\}\}=\{\{\{\{H_0,B_\beta\}\},C_\gamma\}\}+\{\{B_\beta\{\{H_0,C_\gamma\}\}\}\},$$
$$\forall B_\beta=B_\beta(\cdot,t)\in\mathcal{F}(\mathbb{P};\mathbb{A}_\beta),$$
$$C_\gamma=C_\gamma(\cdot,t)\in\mathcal{F}(\mathbb{P};\mathbb{A}_\gamma),$$
$$H_0=H_0(\cdot,t)\in\mathcal{F}(\mathbb{P};\mathbb{A}_0)\tag{29}$$

can easily be verified using the color Jacobi identity (27) and (2) and (3).

VI. COLOR HAMILTON–JACOBI DIFFERENTIAL EQUATION

In order to obtain the Hamilton–Jacobi differential equation in our color analytic framework, one proceeds as in the conventional case. One calculates the action functional ω_0 along the solution trajectories x_{Sol} (21) and obtains a function of the end points

$$\mathbb{A}_0\ni\omega_{0x_{\mathrm{Sol}}}=:S(t_a,q^i_\alpha(t_a),\ldots,\overset{(n-1)}{q^i_\alpha}(t_a);\times t_b,q^i_\alpha(t_b),\ldots,\overset{(n-1)}{q^i_\alpha}(t_b))\tag{30}$$

which satisfies the color Hamilton Jacobi differential equation:

$$-\frac{\partial S(";t_b,\overset{(j-1)}{q^i_\alpha}(t_b))}{\partial t_b}=H_0\left(t_b,\overset{(j-1)}{q^i_\alpha}(t_b),\frac{\partial S(";t_b,\overset{(j-1)}{q^i_\alpha}(t_b))}{\partial\overset{(l-1)}{q^k_\beta}(t_b)}\right)$$
$$P_{l;k,-\beta}=\frac{\partial S(";t_b,\overset{(j-1)}{q^i_\alpha}(t_b))}{\partial\overset{(l-1)}{q^k_\beta}(t_b)},\quad l=1,\ldots,n.\tag{31}$$

VII. EXTENSION TO FIELD THEORIES

Let now $\mathbb{H}^{[n_\Gamma]}$ denote the Schwartz space consisting of C^∞ maps from $\mathbb{R}^3$ to $\mathbb{G}^{[n_\Gamma]}$ vanishing with all their derivatives at infinity more rapidly then any power of $\|\mathbf{r}\|^{-1}$, $\mathbf{r}\in\mathbb{R}^3$, $\|\mathbf{r}\|:=\sqrt[+]{r_1^2+r_2^2+r_3^2}$. Elements of $\mathbb{H}^{[n_\Gamma]}$ are denoted by

$$\phi=(q^i_\gamma(\mathbf{r}))^{i\in\{1,\ldots,n_\gamma\}}_{\gamma\in\Gamma,\mathbf{r}\in\mathbb{R}^3},$$
$$\eta=(y^i_\gamma(\mathbf{r}))^{i\in\{1,\ldots,n_\gamma\}}_{\gamma\in\Gamma,\mathbf{r}\in\mathbb{R}^3}\in\mathbb{H}^{[n_\Gamma]}.\tag{32}$$

The $\mathbb{A}$-valued and in general distribution-valued functions $f\in\mathcal{F}_d(\mathbb{H}^{[n_\Gamma]};\mathbb{A})$ are supposed to be sufficiently many times functional $\mathbb{A}_0$-differentiable and ($\lambda\in\mathbb{R}$)

$$\frac{\delta}{\delta q^i_\gamma(\mathbf{r})}f[\phi]=\frac{\delta f[\phi]}{\delta q^i_\gamma(\mathbf{r})},$$
$$\frac{d}{d\lambda}f[\phi+\lambda\eta]|_{\lambda=0}=\int d^3\mathbf{r}\sum_{\gamma\in\Gamma}\sum_{i=1}^{n_\gamma}y^i_\gamma(\mathbf{r})\frac{\delta f[\phi]}{\delta q^i_\gamma(\mathbf{r})}\tag{33}$$

denotes the variational left-derivative of f with respect to $q^i_\gamma(\mathbf{r}) \in \mathbb{A}_\gamma$ at $\phi \in \mathbb{H}^{[n_\Gamma]}$. The color Heisenberg commutation rules in field theory result:

$$\frac{\delta}{\delta q^i_\alpha(\mathbf{r})}\frac{\delta}{\delta q^j_\beta(\mathbf{r}')} = \sigma(\alpha,\beta)\frac{\delta}{\delta q^j_\beta(\mathbf{r}')}\frac{\delta}{\delta q^i_\alpha(\mathbf{r})},$$

$$\sigma(\alpha,\alpha) = +1 :\Rightarrow \left[\frac{\delta}{\delta q^i_\alpha(\mathbf{r})}, q^j_\alpha(\mathbf{r}')\right]_- = \delta^j_i\delta(\mathbf{r}-\mathbf{r}'),$$

$$\sigma(\alpha,\alpha) = -1 :\Rightarrow \left[\frac{\delta}{\delta q^i_\alpha(\mathbf{r})}, q^j_\alpha(\mathbf{r}')\right]_+ = \delta^j_i\delta(\mathbf{r}-\mathbf{r}'), \tag{34}$$

$$\alpha\neq\beta :\Rightarrow \frac{\delta}{\delta q^i_\alpha(\mathbf{r})}q^j_\beta(\mathbf{r}') = \sigma(\beta,\alpha)q^j_\beta(\mathbf{r}')\frac{\delta}{\delta q^i_\alpha(\mathbf{r})},$$

$$\forall\alpha,\beta\in\Gamma,\ i\in\{1,\dots,n_\alpha\},$$

$$j\in\{1,\dots,n_\beta\};\ \mathbf{r},\mathbf{r}'\in\mathbb{R}^3.$$

In order to avoid unnecessary repetitions we remark that one only has to replace the ordinary color derivative $\partial/\partial q^i_\alpha$, the summation Σ_i over i, and the vector space $\mathbb{G}^{[n_\Gamma]}$ by the variational color derivative $\delta/\delta q^i_\alpha(\mathbf{r})$, the summation $\Sigma_i \int d^3\mathbf{r}$ over $i,\mathbf{r}$ and the vector space $\mathbb{H}^{[n_\Gamma]}$ in order to transcribe the results of the Secs. II–VI to the field theoretic case. We therefore finish the paper by at least showing the construction of the color Poisson (super)algebra in field theory

$$\{\{B_\beta,C_\gamma\}\} := \sum_{l=1}^{n}\sum_{\alpha\in\Gamma}\sum_{i=1}^{n_\alpha}\int d^3\mathbf{r}\times\left(\sigma(\alpha,\alpha)\sigma(\beta,\alpha)\frac{\delta B_\beta}{\delta P_{l;i,-\alpha}(\mathbf{r})}\frac{\delta C_\gamma}{\delta^{(l-1)} q^i_\alpha(\mathbf{r})} - \sigma(\alpha,\beta)\frac{\delta B_\beta}{\delta^{(l-1)} q^i_\alpha(\mathbf{r})}\frac{\delta C_\gamma}{\delta p_{l;i,-\alpha}(\mathbf{r})}\right), \tag{35}$$

$$B_\beta = B_\beta(\cdot,t)\in\mathscr{F}_d(\mathbb{P};\mathbb{A}_\beta),\quad C_\gamma = C_\gamma(\cdot,t)\in\mathscr{F}_d(\mathbb{P};\mathbb{A}_\gamma),$$

$$\mathbb{P} := (\mathbb{H}^{[n_\Gamma]})^{2n},$$

and by specifying the color variational self-adjointness conditions for evolution terms

$$\varepsilon_{j,-\beta}(x^\mu)\,\phi = \varepsilon_{j,-\beta}(x^\mu, q^i_\alpha(x^\mu), \partial_{\nu_1}q^i_\alpha(x^\mu),\dots, \times\ \partial_{\nu_1}\cdots\partial_{\nu_n}q^i_\alpha(x^\mu))$$

which depend (already in a symmetrized manner in the arguments, see Ref. 20, p. 395) at most on the n-jet of the color valued field functions $\phi\in\mathbb{Q}$:

$$0 = \frac{\partial}{\partial(\partial_{\nu_1}\cdots\partial_{\nu_r}q^i_\alpha)}\varepsilon_{j,-\beta}(x^\mu)_\phi + \sigma(\alpha,\beta)\sum_{u=0}^{n}(-1)^{u+1}\binom{u}{r}\sum_{\nu_{r+1},\dots,\nu_u=1}^{4}\partial_{\nu_{r+1}}\cdots\partial_{\nu_u}\left(\frac{\partial}{\partial(\partial_{\nu_1}\cdots\partial_{\nu_u}q^j_\beta)}\varepsilon_{i,-\alpha}(x^\mu)_\phi\right),$$

$$\forall r = 0,1,\dots,n,\quad \forall\alpha,\beta\in\Gamma,\quad \forall i\in\{1,\dots,n_\alpha\}\ \forall j\in\{1,\dots,n_\beta\},\quad \forall\nu_1,\dots,\nu_r\in\{1,2,3,4\},\quad \forall\phi\in\mathbb{Q},$$

$$\mathbb{Q} = \mathscr{F}([t_a,t_b],\mathbb{H}^{[n_\Gamma]}),\quad \phi: t\to\phi(t):\mathbf{r}\to(q^i_\gamma(\mathbf{r},t))^{i\in\{1,\dots,n_\gamma\}}_{\gamma\in\Gamma}\in\mathbb{G}^{[n_\Gamma]},\quad \forall x^\mu=(\mathbf{r},t)\in\mathbb{R}^3\mathbf{x}[t_a,t_b],\quad \partial_\nu := \frac{\partial}{\partial x^\nu}. \tag{36}$$

If $\varepsilon_{j,-\beta}(x^\mu)_\phi\in\mathbb{A}_{-\beta}$ fulfills (36), the Lagrangian

$$\mathscr{L}(x^\nu)_\phi := \sum_{\gamma\in\Gamma}\sum_{k=1}^{n_\gamma}q^k_\gamma(x^\mu)\int_0^1 d\tau\,\varepsilon_{k_1,-\gamma}(x^\mu,\tau q^i_\alpha(x^\mu),\dots,\tau\partial_{\nu_1}\cdots\partial_{\nu_n}q^i_\alpha(x^\mu))\in\mathbb{A}_0 \tag{37}$$

leads to

$$\varepsilon_{j,-\beta}(x^\mu)_\phi = \sum_{l=0}^{n}(-1)^l\sum_{\nu_1,\dots,\nu_l=1}^{4}\partial_{\nu_1}\cdots\partial_{\nu_l}\left(\frac{\partial\mathscr{L}}{\partial(\partial_{\nu_1}\cdots\partial_{\nu_l}q^j_\beta(x^\mu))}\right), \tag{38}$$

where

$$\sum_{\nu_0}\partial_{\nu_0}\left(\frac{\partial\mathscr{L}}{\partial\partial_{\nu_0}q^j_\beta}\right)$$

means $\partial\mathscr{L}/\partial q^j_\beta$.

Proof:

$$\sum_{l=0}^{n}(-1)^l\sum_{\nu_1,\dots,\nu_l}\partial_{\nu_1}\cdots\partial_{\nu_l}\left(\frac{\partial}{\partial(\partial_{\nu_1}\cdots\partial_{\nu_l}q^j_\beta(x^\mu))}\mathscr{L}\right)$$

$$\overset{(37)}{=}\frac{\partial}{\partial q^j_\beta}\left(\sum_\gamma\sum_k q^k_\gamma\int_0^1 d\tau\ \varepsilon_{k,-\gamma}(x^\mu,\tau q^i_\alpha,\dots,\tau\partial_{\mu_1}\cdots\partial_{\mu_n}q^i_\alpha)\right)$$

$$+\sum_{l=1}^{n}(-1)^l\sum_{\nu_1,\dots,\nu_l}\partial_{\nu_1}\cdots\partial_{\nu_l}\left(\frac{\partial}{\partial(\partial_{\nu_1}\cdots\partial_{\nu_l}q^j_\beta)}\left(\sum_\gamma\sum_k q^k_\gamma\int_0^1 d\tau\,\varepsilon_{k,-\gamma}(x^\mu,\tau q^i_\alpha,\dots,\tau\partial_{\mu_1}\cdots\partial_{\mu_n}q^i_\alpha)\right)\right)$$

$$\overset{(9)(10)}{=}\int_0^1 d\tau\,\varepsilon_{j,-\beta}(x^\mu,\tau q^i_\alpha,\dots,\tau\partial_{\mu_1}\cdots\partial_{\mu_n}q^i_\alpha)+\sum_\gamma\sum_k\sigma(\gamma,\beta)q^k_\gamma\int_0^1 d\tau\ \tau\frac{\partial}{\partial(\tau q^j_\beta)}\varepsilon_{k,-\gamma}(x^\mu,\tau q^i_\alpha,\dots,\tau\partial_{\mu_1}\cdots\partial_{\mu_n}q^i_\alpha)$$

$$+\sum_{l=1}^{n}(-1)^l\sum_{\nu_1,\dots,\nu_l}\partial_{\nu_1}\cdots\partial_{\nu_l}\left(\sum_\gamma\sum_k\sigma(\gamma,\beta)q^k_\gamma\int_0^1 d\tau\ \ \tau\frac{\partial}{\partial(\tau\partial_{\nu_1}\cdots\partial_{\nu_l}q^j_\beta)}\varepsilon_{k,-\gamma}(x^\mu,\tau q^i_\alpha,\dots,\tau\partial_{\mu_1}\cdots\partial_{\mu_n}q^i_\alpha)\right)$$

$$= \int_0^1 d\tau \frac{d}{d\tau} \left(\tau \varepsilon_{j,-\beta}(x^\mu, \tau q^i_\alpha, \ldots, \tau \partial_{\mu_1} \cdots \partial_{\mu_n} q^i_\alpha)\right) - \int_0^1 d\tau \; \tau \frac{d}{d\tau} \varepsilon_{j,-\beta}(x^\mu, \tau q^i_\alpha, \ldots, \tau \partial_{\mu_1} \cdots \partial_{\mu_n} q^i_\alpha)$$

$$+ \sum_\gamma \sum_k \sigma(\gamma,\beta) \sum_{l=0}^n (-1)^l \sum_{\nu_1,\ldots,\nu_l} \partial_{\nu_1} \cdots \partial_{\nu_l} \left(q^k_\gamma \int_0^1 d\tau \; \tau \frac{\partial}{\partial(\tau \partial_{\nu_1} \cdots \partial_{\nu_l} q^j_\beta)} \varepsilon_{k,-\gamma}(x^\mu, \tau q^i_\alpha, \ldots, \tau \partial_{\mu_1} \cdots \partial_{\mu_n} q^i_\alpha) \right)$$

$$= \varepsilon_{j,-\beta}(x^\mu, q^i_\alpha, \ldots, \partial_{\mu_1} \cdots \partial_{\mu_n} q^i_\alpha) - \int_0^1 d\tau \; \tau \sum_\gamma \sum_k \sum_{l=0} \sum_{\nu_1,\ldots,\nu_l} (\partial_{\nu_1} \cdots \partial_{\nu_l} q^k_\gamma) \cdot \frac{\partial}{\partial(\tau \partial_{\nu_1} \cdots \partial_{\nu_l} q^k_\gamma)}$$

$$\times \varepsilon_{j,-\beta}(x^\mu, \tau q^i_\alpha, \ldots, \tau \partial_{\mu_1} \cdots \partial_{\mu_n} q^i_\alpha) + \int_0^1 d\tau \; \tau \sum_\gamma \sum_k \sigma(\gamma,\beta) \sum_{l=0}^n (-1)^l \sum_{\nu_1,\ldots,\nu_l} \sum_{s=0}^l \binom{l}{s} (\partial_{\nu_1} \cdots \partial_{\nu_s} q^k_\gamma) \cdot$$

$$\times \partial_{\nu_{s+1}} \cdots \partial_{\nu_l} \left(\frac{\partial}{\partial(\tau \partial_{\nu_1} \cdots \partial_{\nu_l} q^j_\beta)} \varepsilon_{k,-\gamma}(x^\mu, \tau q^i_\alpha, \ldots, \tau \partial_{\mu_1} \cdots \partial_{\mu_n} q^i_\alpha) \right).$$

By virtue of (36), the last two terms cancel after extending the sum

$$\sum_{s=0}^l \binom{l}{s} \quad \text{to} \quad \sum_{s=0}^n \binom{l}{s} \cdots$$

(by putting $\binom{l}{s} = 0$ if $l < s$) and after substituting $l \to s$, $s \to l$ in the last double sum. ■

We emphasize that by virtue of the commutation factor $\sigma(\alpha,\beta)$ in front of the sum in (36) the color variational self-adjointness conditions (36) represent a nontrivial generalization of the old mathematical concept "variational self-adjointness" studied in the pure bosonic case by Helmholtz[21] and recently by Vainberg,[22] Tonti,[23] Santilli,[24–26] Tulzcyjew,[27] Dedecker and Tulzcyjew,[28] Abraham and Marsden,[29] Hughes and Marsden,[30] Takens,[31] Kosmann-Schwarzbach,[32] Kamo and Sugano,[33] Anderson,[34] Horndeski,[35] Vanderbauwhede,[36] Bauderon,[37] Atherton and Homsy,[38] Telega,[39] Hojman and Urrutia,[40] Trostel,[20] and others. For more references and for the technique of the modification of a prescribed set of bosonic evolution terms by virtue of integrating matrices in order to arrive at variational self-adjoint evolution terms we refer also to the books[24] of Santilli.

VIII. CONCLUDING REMARKS

We have presented some recent results concerning the calculation with numbers obeying unusual commutation relations. These numbers can be regarded as elements of an associative Γ-graded algebra A equipped with a commutation function σ satisfying nontrivial, nonlinear relations which are also known in the theory of color (super)algebras. Color Heisenberg commutation rules, generalized variational self-adjointness conditions, and a color Hamiltonian formalism result. As a novelty, we have constructed color Poisson brackets leading to an algebra called color Poisson (super)algebra satisfying the axioms of a color (super)algebra. This color Poisson (super)algebra is in addition characterized by the associative Γ-graded product structure inherited from A. The bracket and the associative product are linked by the color derivation property. We refer the reader to our paper,[18] first, concerning the concept of generalized superdifferentiability or—being more specific—A_0-differentiability which generalizes the concept of conventional (i.e., Z_2-graded) superdifferentiability according to Jadzyk and Pilch[17] to our case, if dealing with an arbitrary Γ-graded associative algebra A; second, concerning the mathematical foundation of generalized variational principles and color variational self-adjointness from the viewpoint of an exterior differential calculus for A_0-valued and A_0-differentiable p-forms ($p \in N_0$) on Banachmoduli; and third, concerning the initiation of color differential geometry characterized by unusual commutation properties of covectors within the associative wedge product.

ACKNOWLEDGMENTS

It is a sincere pleasure to thank Professor Dr. R. M. Santilli for the kind invitation to the first workshop on Hadronic Mechanics held at the Institute for Basic Research, Cambridge, Massachusetts in 1983 and for fruitful suggestions via correspondence. I also thank Professor Dr. H. Bucka, Professor Dr. K-E. Hellwig, Professor Dr. W. Muschik, Professor Dr. U. Scherz, and my father Professor Dr. Rudolf Trostel for kind help and warm encouragement.

[1] Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics* (Springer, Berlin, 1982).
[2] Y. Ohnuki and S. Kamefuchi, Physica A **114**, 416 (1982).
[3] M. Omote and S. Kamefuchi, *Lecture Notes in Physics* (Springer, Berlin, 1979), Vol. 94, p. 452.
[4] R. Y. Levine and Y. Tomozawa, Ann. Phys. **149**, 457 (1983).
[5] H. S. Green, Phys. Rev. **90**, 270 (1953).
[6] V. Kac, Adv. Math. **26**, 8 (1977).
[7] V. Rittenberg and D. Wyler, Nucl. Phys. B **139**, 189 (1978).
[8] V. Rittenberg and D. Wyler, J. Math. Phys. **19**, 2193 (1978).
[9] J. Lukierski and V. Rittenberg, Phys. Rev. D **18**, 385 (1978).
[10] V. K. Agrawala, Hadronic J. **4**, 444 (1981).
[11] H. S. Green and P. D. Jarvis, J. Math. Phys. **24**, 1681 (1983).
[12] M. Scheunert, J. Math. Phys. **24**, 2658, 2671, 2681 (1983).
[13] B. Kostant, *Lecture Notes in Mathematics* (Springer, Berlin, 1977), Vol. 570, p. 177.
[14] J. Dell and L. Smolin, Commun. Math. Phys. **66**, 197 (1979).
[15] A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[16] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[17] A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).
[18] R. Trostel, Hadronic J. **6**, 1518 (1983).
[19] M. Ostrogradski, Mem. Ac. St. Petersbourg VI **4**, 385 (1850).
[20] R. Trostel, Hadronic J. **6**, 305 (1983).
[21] H. Helmholtz, J. Reine Angew. Math. **100**, 137 (1887).
[22] M. M. Vainberg, *Variational Methods for the Study of Nonlinear Operators* (Holden–Day, San Francisco, 1964).
[23] E. Tonti, Bull. Cl. Sci. Acad. R. de Belg. **55**, 137, 262 (1969); Hadronic J. **5**, 1404 (1982).
[24] R. M. Santilli, *Foundations of Theoretical Mechanics* (Springer, New York, 1978 and 1983), Vols. I and II.
[25] R. M. Santilli, Ann. Phys. **103**, 354, 409 (1977); **105**, 227 (1977).
[26] R. M. Santilli, Hadronic J. **1**, 223, 574 (1978).

[27]W. M. Tulczyjew, *Lecture Notes in Mathematics* (Springer, Berlin, 1980), Vol. 836, p. 22.
[28]P. Dedecker and W. M. Tulczyjew, *Lecture Notes in Mathematics* (Springer, Berlin, 1980), Vol. 836, p. 498.
[29]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin Cummings, New York, 1978), p. 131 (2.5.c).
[30]T. J. R. Hughes and J. E. Marsden, Rep. Math. Phys. **12**, 35 (1977).
[31]F. Takens, *Lecture Notes in Mathematics* (Springer, Berlin, 1980), Vol. 597, p. 581.
[32]Y. Kosmann-Schwarzbach, Ann. J. H. Poincaré (to be published).
[33]H. Kamo and R. Sugano, Ann. Phys. **128**, 298 (1980).
[34]J. M. Anderson, Aequ. Math. **17**, 255 (1978).
[35]G. W. Horndeski, Tensor N. S. **28**, 303 (1974).
[36]A. L. Vanderbauwhede, Hadronic J. **1**, 1177 (1978); **2**, 620 (1979).
[37]M. Bauderon, Ann. J. H. Poincaré **36**, 159 (1982).
[38]R. W. Atherton and G. M. Homsy, Stud. Appl. Math. **LIV**, 31 (1975).
[39]J. J. Telega, J. Inst. Math. Appl. **24**, 175 (1979).
[40]S. Hojman and L. F. Urrutia, J. Math. Phys. **22**, 1896 (1981).

Necessary conditions for a unique solution to two-dimensional phase recovery

Richard Barakat[a] and Garry Newsam[b]
Division of Applied Sciences, Harvard University, Cambridge, Massachusetts 02138

(Received 31 January 1984; accepted for publication 8 June 1984)

In this paper we show that although in one dimension multiplicity of solutions to the phase reconstruction problem presents a serious problem, in two or more dimensions multiplicity is pathologically rare. We derive from a given solution pair (g,G) necessary conditions for the existence of alternative solution pairs (h,H), and a characterization of their form. The mathematical tools employed are from the theory of functions of two complex variables.

PACS numbers: 02.30.Dk

I. INTRODUCTION

The two-dimensional phase retrieval problems can be stated as: Let A and B be bounded subsets of $\mathbb{R}^2$. Given the information that $g(z_1,z_2)$ is the Fourier transform of a function $G(\omega_1,\omega_2)$ with support contained in B and the values $m(x_1,x_2) = |g(x_1,x_2)|$ on A, find the phase of g on A and reconstruct G.

The phase retrieval problem does not necessarily have a unique solution. The aim of this article is to derive from a given solution pair g and G necessary conditions for the existence (and characterization) of alternative solution pairs h and H. An intuitive start for such an investigation is the simple result that if $f(z)$ is an analytic function of the complex variable z, then so is $f^*(z^*)$ and $f(z)$ and $f^*(z^*)$ have the same modulus for real z. This suggests that if a solution g can be factored into a product of analytic functions g_1 and g_2 then the entire function of two complex variables $h(z_1,z_2) \equiv g_1(z_1,z_2)g_2^*(z_1^*,z_2^*)$ is also a possible solution. The main result of this paper is to show that all possible alternative solutions must be of this simple form.

II. ONE-DIMENSIONAL RESULTS

The following results from the theory of functions of a single complex variable are required.

Theorem 1: (Paley–Wiener Theorem[1]) Let $B \equiv [b_1,b_2]$ be a bounded interval in $\mathbb{R}$. Then for any $G \in L^2(\mathbb{R})$, $G \neq 0$ on B, the transform

$$g(z) = \int_{b_1}^{b_2} e^{izu} G(u)\,du \tag{1}$$

is an entire function and there exist constants α and β such that

$$|g(z)| < \begin{cases} \alpha e^{-b_1 \operatorname{Im} z}, & \text{if } \operatorname{Im} z > 0, \\ \beta e^{b_2 \operatorname{Im} z}, & \text{if } \operatorname{Im} z < 0. \end{cases} \tag{2}$$

The next result is the fundamental theorem providing the necessary machinery to characterize all possible solutions to the phase problem both in one and two dimensions. Although independently derived by many authors,[2–4] it appears to have been first stated by Akutowicz.[5–6] We state the result as originally presented there.

Theorem 2: Let $\mathscr{C}$ be the class of all functions $g \in L^2(\mathbb{R})$ satisfying: (a) $|g(x)| = m(x) \neq 0 \quad \forall x \in \mathbb{R}$; (b) $g = \mathscr{F}G$, where support of G is contained in a bounded interval B of $\mathbb{R}$. Then any two functions $g,h \in \mathscr{C}$ are related by equations of the form

$$h(z) = e^{i(\alpha + \beta z)} B(z) g(z), \qquad B(z) = \prod_{l=1}^{\infty} \left(\frac{z - z_l^*}{z - z_l} \right), \tag{3}$$

where the z_l form some subset of the zeros of $g(z)$. The function $(z - z_l^*)/(z - z_l)$ is termed a Blaschke factor.

Lemma 1: A necessary and sufficient condition for the infinite product $B(z)$ to converge is that[7]

$$\sum_{l=1}^{\infty} \frac{|\operatorname{Im} z_l|}{1 + |z_l|^2} < \infty. \tag{4}$$

A sufficient condition for the convergence of the infinite sum is that $G(u)$ have only a finite number of jump discontinuities over B.

It is easily shown that if G has support in an interval B and if $h(z) = B(z)g(z)$, then $H(u) = (\mathscr{F}^{-1}h)(u)$ also has support in B. Therefore, combining Theorems 1 and 2 and Lemma 1 gives the following statement on existence of multiple solutions to the one-dimensional (1-D) phase retrieval problem.

Theorem 3: Let A and B be bounded intervals in $\mathbb{R}$ with a modulus $m(x)$ specified over A and a solution pair g and G be given to the corresponding 1-D phase problem. Then if $m(x)$ has an extension to the entire real line such that $m(x) > 0$ and $G(u)$ has only finite jump discontinuities over B as well as being nonzero in neighborhoods of the endpoints of B, then all other solution pairs h,H are given by

$$h(z) = e^{i\alpha} B(z) g(z), \tag{5}$$

$$H(u) = (\mathscr{F}^{-1}h)(u), \tag{6}$$

where $B(z)$ is any finite or infinite product of Blaschke factors and $\alpha \in \mathbb{R}$.

The conditions of Theorem 3 imply that a solution $g(z)$ has a Hadamard factorization

[a] Also: RGB Associates, Box 8, Wayland, MA 01778.
[b] Present address: Centre for Mathematical Analysis, The Australian National University, Canberra, Australia.

 0022-2488/84/113190-04$02.50

$$g(z) = |g(0)|e^{i(\alpha+\beta z)}\prod_{l=1}^{\infty}\left(1-\frac{z}{z_l}\right), \tag{7}$$

where $\alpha,\beta\in\mathbb{R}$, which may be rewritten as

$$\begin{aligned} g(z) &= \left[|g(0)|e^{i(\alpha+\beta z)}\prod_{l\in\Lambda}\left(1-\frac{z}{z_l}\right)\right] \\ &\quad\times\left[\prod_{l\in(N-\Lambda)}\left(1-\frac{z}{z_l}\right)\right] \\ &\equiv g_1(z)g_2(z), \end{aligned} \tag{8}$$

where Λ is a subset of the natural numbers N. Theorem 3 thus states that any solution $h(z)$ has the form

$$h(z) = e^{i\gamma}g_1(z)g_2^*(z^*), \quad \gamma\in\mathbb{R}, \tag{9}$$

i.e., that all possible solutions are in one-to-one correspondence with all possible factorizations of $g(z)$.

III. EXTENSION TO TWO DIMENSIONS

Conditions that multiple solutions to the 2-D phase retrieval problem must satisfy can be deduced from the 1-D results. To begin, suppose that the problem as stated has a solution pair $g(z_1,z_2)$, $G(\omega_1,\omega_2)$, where $z=x+iy$ and $\omega=u+iv$ denote variables in the transform and physical domains, such that $G(u_1,u_2)$ has only a finite number of jump discontinuities over B. Then

Lemma 2: $g(z_1,z_2)$ is an entire function of the complex variables z_1,z_2 of exponential growth.

Proof: After defining the quantities

$$\begin{aligned} u_1^+ &= \max\{u_1:(u_1,u_2)\in B\}, \\ u_2^+(u_1) &= \max\{u_2:(u_1,u_2)\in B\}, \\ u_1^- &= \min\{u_1:(u_1,u_2)\in B\}, \\ u_2^-(u_1) &= \min\{u_2:(u_1,u_2)\in B\}, \end{aligned}$$

$g(z_1,z_2)$ can be expressed as

$$g(z_1,z_2) = \int_{u_1^-}^{u_1^+} du_1\, e^{iz_1u_1}\int_{u_2^-(u_1)}^{u_2^+(u_1)} du_2\, e^{iz_2u_2}G(u_1,u_2). \tag{10}$$

Writing $g(z_1,z_2)$ as $g_{z_2}(z_1)$ to indicate that $g(z_1,z_2)$ is to be considered as a function of z_1 only with z_2 fixed gives

$$g_{z_2}(z_1) = \int_{u_1^-}^{u_1^+}\tilde{G}(u_1,z_2)e^{iz_1u_1}\,du_1. \tag{11}$$

Therefore, by Theorem 1, $g_{z_2}(z_1)$ is an entire function of z_1 of exponential growth. A similar procedure shows that $g_{z_1}(z_2)$ is an entire function of z_2.

An immediate consequence of this lemma is that if a solution exists then the modulus $m(x_1,x_2)$ has an analytic extension to all of $\mathbb{R}^2$, which, under the assumption that $m(x_1,x_2)>0$, $\forall(x_1,x_2)$, is unique.

There are four points to be noted with respect to the next three paragraphs.

(1) The first point is that the zeros $\eta_k(x_2)$ may be numbered in accordance with Eq. (13).

(2) For a given k, the maps ϕ_k and ψ_k are analytic for almost all x_2. This follows from the fact that the set of singular points has dimension one less than the set of points x_2 for which the set of zeros is analytic.

(3) The next step is to note that the condition that k be fixed can be dropped because there is only a countably finite number of k's, and the set of almost all x_2 is uncountably infinite.

(4) The final step is to observe that if k is indeed an alternative solution, then its zero manifold Y must be the analytic continuation of the sets $(\rho_k(x_2),x_2)$ and therefore be continued in $X\cup X^*$. If $Y\neq X$ and $Y\neq X^*$, then any $y\in Y$ (with y also in X) is part of an analytic submanifold Y_1, with $Y_1\subset Y$ and $Y_1\subset X$ but $Y_1\neq X$. Therefore, X must be decomposable as a sum of analytic submanifolds.

Now let $h(z_1,z_2)$, $H(w_1,w_2)$ be any other solution pair to this problem; then $h_{x_2}(x_1)$ and $g_{x_2}(x_1)$ must have the same modulus $m_{x_2}(x_1)$ over the set $B_{x_2}=\{x_1:(x_1,x_2)\in B\}$; i.e., $h_{x_2}(z_1)$ and $g_{x_2}(z_1)$ are both solutions to a 1-D phase retrieval problem. Therefore, by Theorem 2, Lemma 1, and the above assumptions on g, G, and m we have that

$$h_{x_2}(z_1) = e^{i\alpha(x_2)}e^{i\beta(x_2)z_1}B_{x_2}(z_1)g_{x_2}(z_1), \tag{12}$$

where $\alpha(x_2)$ and $\beta(x_2)$ are constants dependent on x_2 only and $B(z)$ is an infinite product of Blashke factors formed from the zeros $\rho_l(x_2)$ of $g_{x_2}(z_1)$. Thus if $\eta_l(x_2)$ are the zeros of $h_{x_2}(z_1)$ we may order them so that

$$\eta_k(x_2) = \rho_k(x_2) \text{ or } \rho_k^*(x_2). \tag{13}$$

Now let $X\subset C^2$ and $Y\subset C^2$ be the sets of zeros of g and h, respectively. It is known[8] that the zeros of a function of n complex variables from an analytic set of dimension $(n-1)$, which in turn is the union of analytic manifolds of dimensions $(n-1)$ and a set of dimension at most $(n-2)$. The difference in dimensions implies that, for a fixed k, for almost all points x_2 the points $(\rho_k(x_2),x_2)$ and $(\eta_k(x_2),x_2)$ are members of analytic submanifolds of X and Y, respectively. That is, there exist maps

$$\begin{aligned} &\phi_k: C^1\to X, \quad \phi_k(x_2) = (\rho_k(x_2),x_2), \\ &\psi_k: C^1\to Y, \quad \psi_k(x_2) = (\eta_k(x_2),x_2), \end{aligned} \tag{14}$$

which are analytic on a neighborhood $N_k(x_2)$ of x_2. Since there are only countably many k's, it follows that, for almost all x_2, all of the maps $\{\phi_k,\psi_k\}_{k=1}^{\infty}$ are analytic in neighborhoods $N_k(x_2)$ of x_2. [Note the dependence of $N_k(x_2)$ on k.]

If we now suppose for simplicity that the zeros $\rho_k(x_2)$ are distinct, i.e.,

$$k\neq l \Rightarrow \rho_k(x_2)\neq\rho_l(x_2) \text{ or } \rho_l^*(x_2), \tag{15}$$

then Eq. (13), and the analyticity of ϕ_k and ψ_k imply that

$$y_k(z) = \rho_k(z) \text{ or } \rho_k^*(z^*), \quad z\in N_k(x_2). \tag{16}$$

Therefore, if $X_1\subseteq X$ and $Y_1\subset Y$ are the extensions of the neighborhoods $\{(\rho_k(z),(z),(y_k(z),z)\}_{k=1}^{\infty}$ to analytic manifolds, then

$$X_1\subseteq Y_1\cup Y_1^*, \tag{17}$$

$$Y_1\subseteq X_1\cup X_1^*. \tag{18}$$

If $Y_1\neq X_1$ then there exists a point $y\in Y_1$ and an associated neighborhood N_y such that y and N_y are contained in analytic submanifold X_2 of X, but the analytic extension of N_y is not X. Therefore, X may be decomposed into two submanifolds X_2 and X_1-X_2.

Thus, apart from alternative solutions generated by varying $\alpha(x_2)$ and $\beta(x_2)$, a necessary condition that an alternative solution h to the 2-D phase problem must satisfy is that some of its zeros be the complex conjugates of those of the original solution. Although this is the same mechanism by which an infinite number of alternative solutions to the 1-D phase problem are generated, the zeros must now satisfy the condition that they form a union of one-dimensional analytic manifolds as opposed to a union of zero-dimensional manifolds, that is a collection of connected analytic line segments as opposed to a collection of isolated points. If an alternative solution exists then either the whole manifold X has been "flipped" to its conjugate, or it has been "torn" and only partially flipped. The connected nature of X_1 implies that the existence of "dotted lines" along which tears may be made is very unlikely; this compares to the isolated points in the 1-D problem, each of which may be flipped independently of the other.

Given this condition the form of an alternative solution may be determined. Let X_1 be decomposable into submanifolds X_2, X_3 and define

$$X_{i,x_2} \equiv X_i \cap \{(z_1,z_2): x_2 \text{ fixed}, z_1 \in C^1.\} \tag{19}$$

Then the function $g_{x_2}(z_1)$ may be written as the product

$$g_{x_2}(z_1) = g_{1,x_2}(z_1) g_{2,x_2}(z_1), \tag{20}$$

where

$$g_{2,x_2}(z_1) = \prod_{(\rho_k(x_2),x_2) \in X_{2,x_2}} \left(1 - \frac{z_1}{\rho_k(x_2)}\right). \tag{21}$$

By Eq. (12) $h_{x_2}(z_1)$ may be written as

$$h_{x_2}(z_1) = g_{1,x_2}(z_1) g^*_{2,x_2}(z_1^*). \tag{22}$$

If $h(z_1,z_2)$ exists it is the analytic extension of $h_{x_2}(z_1)$, therefore, $h(z_1,z_2) = g_1(z_1,z_2) g_2^*(z_1^*,z_2^*)$.

We have not been able to show that this necessary condition for alternative solutions is also sufficient; i.e., given submanifolds X_2, X_3 and the decomposition of Eq. (20) that the $h_{x_2}(z_1)$ of Eq. (22) may be analytically continued to a function $h(z_1,z_2)$. One source of trouble is the dependence of $N_k(x_2)$ on k; it is possible that for every x_2, $\cap_{k=1}^{\infty} N_k(x_2) = \phi$. Then although each zero $(\rho_k(x_2),x_2)$ is analytic in a neighborhood of x_2 there does not exist a neighborhood over which all zeros are uniformly analytic, and therefore, a neighborhood over which the product of Eq. (21) is provably analytic.

If the cardinalities of X_{i,x_2} are finite, e.g., $g(z_1,z_2)$ is a polynomial, then sufficiency can be shown. In the polynomial case decomposability of X_1 into X_2 and X_3 is equivalent to a factorization of the polynomial. However, almost all polynomials of two variables are irreducible so that such a factorization and decomposition does not exist, therefore, alternative solutions do not exist. Irreducibility extends to general functions of two variables with infinite sets of zeros, so that exact alternative solutions are most unlikely in 2-D phase retrieval. This result on polynomials and its implications is also presented in Ref. 9.

IV. THE SUPPORT OF ALTERNATIVE SOLUTIONS

In the previous subsection conditions that alternative solutions g and h must satisfy in order that $|g(x_1,x_2)| = |h(x_1,x_2)|$ were derived. We still have to derive the necessary conditions on g and h so that (support G) = (support H). The first is that $\beta(x_2) \equiv 0$ in Eq. (12). This follows by noting that if $G(u_1,u_2)$ is nonzero in neighborhoods of points $(u_1^+,u_2),(u_1^-,u_2) \in B$ then the function $\tilde{G}(u_1,x_2)$ of Eq. (A11) will be nonzero in neighborhoods of $u_1 = u_1^+$ and $u_1 = u_1^-$ for almost all x_2. So by Theorem 3, $h_{x_2}(z_1)$ is the transform of a function $\tilde{H}(u_1,x_2)$ with support in (u_1^-,u_1^+) if and only if $\beta(x_2) \equiv 0$.

A second condition follows from noting that the boundedness of the set B implies that $h(z_1,z_2)$ is of exponential growth in z_2, so that $\alpha(x_2)$ must only be a linear function of x_2. Summarizing these results and those of the previous section gives the next theorem.

Theorem 4: Let g,G be a solution pair to the 2-D phase retrieval problem. Then any other solution pair h, H must have the form

$$h(z_1,z_2) = e^{i(\alpha_1 + \alpha_2 z_2)} g_1(z_1,z_2) g_2^*(z_1^*,z_2^*), \tag{23}$$

where $g_1 g_2$ is a factorization of g.

We have been unable to complement these necessary conditions for equality of support with sufficient conditions equivalent to those for the 1-D problem. The difficulty seems to lie in determining the role of the geometry of B; we give two examples.

1. The first example concerns convexity and is taken from Huiser and van Torn.[10] Let g,G be a solution pair, then after the change of variables to the new orthogonal coordinate systems $(s_1,s_2),(t_1,t_2)$ with

$$\begin{aligned} &s_1 = u_1 \cos\psi + u_2 \sin\psi, \quad t_1 = x_1 \cos\psi + x_2 \sin\psi, \\ &s_2 = -u_1 \sin\psi + u_2 \cos\psi, \quad t_2 = -x_1 \sin\psi + x_2 \cos\psi, \end{aligned} \tag{24}$$

and definition of the quantities

$$\begin{aligned} s_2^+(s_1) &= \max\{s_2:(s_1,s_2) \in B\}, \\ s_2^-(s_1) &= \min\{s_2:(s_1,s_2) \in B\}, \\ s_1^+(\psi) &= \max\{s_1:(s_1,s_2) \in B\}, \\ s_1^-(\psi) &= \min\{s_1:(s_1,s_2) \in B\}, \end{aligned} \tag{25}$$

the relationship $g = \mathcal{F}G$ may be rewritten as

$$g(t_1,t_2) = \int_{s_1^-(\psi)}^{s_1^+(\psi)} ds_1\, e^{is_1 t_1} \int_{s_2^-(s_1)}^{s_2^+(s_1)} ds_2\, e^{is_2 t_2} G(s_1,s_2). \tag{26}$$

For fixed t_2 the growth rate in $g_{t_2}(t_1)$ is determined by $s_1^+(\psi)$ and $s_1^-(\psi)$. Knowing these values for all ψ is equivalent to knowing all supporting hyperplanes for the set B, which by duality arguments from linear algebra is equivalent to knowing the convex hull of B. If h and H is any other solution pair then $h_{t_2}(t_1)$ must have the same growth as $g_{t_2}(t_1)$, otherwise H has support outside of the convex hull of B.

If g has a factorization $g_1 g_2$ such that the growth of g_2 is always dominated by that of g_1 (e.g., g_2 is a polynomial) then the alternative function

$$h(z_1,z_2) = g_1(z_1,z_2) g_2^*(z_1^*,z_2^*) \tag{27}$$

has the same modulus as g and support in the convex hull of B. If B is convex then h is an alternative solution, if B is not convex then it is possible that the support of H is not B even though still in the convex hull.

2. Let g,G be a solution pair, then it is trivial to show that the inverse transform of $g^*(z_1^*,z_2^*)$ is $G^*(-\omega_1,-\omega_2)$ which has support $-B$. So a sufficient condition that $g^*(z_1^*,z_2^*)$ be an alternative solution is that $B=-B$, i.e., B is invariant under rotation by 180°.

Example 1 suggests that convexity of B is necessary for existence of an alternative solution and taken with Example 2 suggests that for a factorization g into g_1g_2 and an alternative solution h of Eq. (27) then B must have symmetries linked in some fashion to those directions in which growth of g_2 dominates g_1.

V. CONCLUSIONS

Nonuniqueness in the phase retrieval problem in two dimensions appears to depend on two conditions: (1) that the zero space of g be decomposable into a union of several submanifolds, (2) that B possesses a suitable combination of convexity and symmetry. Both conditions will, in general, be difficult to satisfy compared to the 1-D phase retrieval problem. Only in the case of symmetries that effectively reduce $g(z_1,z_2)$ to a function of one variable (e.g., the possession of radial symmetry investigated in Ref. 11) will the manifold have an infinite decomposition as appears in the 1-D problem. In most cases it will be indivisible. Likewise the general two-dimensional bounded set has considerably more degrees of freedom than the one-dimensional bounded set, the interval, consequently it has far fewer symmetries. Therefore, in general the 2-D phase retrieval problem will have a unique solution if one exists.

ACKNOWLEDGMENTS

Barakat's work was supported by the Rome Air Development Center (RADC). Newsam's work was supported by the National Science Foundation (NSF). We wish to thank A. Ne'eman for valuable criticism.

[1] R. Boas, *Entire Functions* (Academic, New York, 1954).
[2] A. Walther, "The question of phase retrieval in optics," Opt. Acta **10**, 41–49 (1963).
[3] E. M. Hoffsteter, "Construction of time-limited functions with specified autocorrelation functions," IEEE Trans. Inf. Theory **IT-10**, 119–126 (1964).
[4] M. Goldberger, H. Lewis, and K. Watson, "Use of intensity correlations to determine the phase of a scattering amplitude," Phys. Rev. **132**, 2764–2787 (1963).
[5] E. J. Akutowicz, "On the determination of the phase of a Fourier integral: I," Trans. Am. Math. Soc. **83**, 179–192 (1956).
[6] E. J. Akutowicz, "On the determination of the phase of a Fourier integral: II," Proc. Am. Math. Soc. **8**, 234–238 (1957).
[7] B. Ya. Levin, *Distribution of Zeros of Entire Functions* (American Mathematical Society, Providence, RI, 1964).
[8] L. Ronkin, *Introduction to the Theory of Entire Functions of Several Complex Variables* (American Mathematical Society, Providence, RI, 1974).
[9] Yu. M. Bruck and L. G. Sodin, "On the ambiquity of the image reconstruction problem," Opt. Commun. **30**, 304–308 (1979).
[10] A. Huiser and P. van Torn, "Ambiquity of the phase reconstruction problem," Opt. Lett. **5**, 499–501 (1980).
[11] W. Lawton, "Uniqueness results for the phase retrieval problem for radial functions," J. Opt. Soc. Am. **71**, 1519–1522 (1981).

Bounds for the continuation of perturbative results to the spectral region

M. Ciulli and S. Ciulli
Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France

T. D. Spearman
School of Mathematics, Trinity College, Dublin 2, Ireland

(Received 13 December 1983; accepted for publication 9 March 1984)

The problem of analytic continuation to the boundary of the holomorphy domain from both continuous and discrete interior sets has recently been the subject of detailed analyses. This problem is important in phenomenological applications but is also of interest in theoretical calculations, e.g., in attempting to evaluate the parameters of resonances or other nonperturbative effects in QCD. Because of the inherent instability of the continuation problem it is necessary to introduce additional criteria—which should be physically based—to select the right continuation function. In this paper, the results thus obtained for continuation from a continuum are examined for stability, and bounds are derived for the errors on the boundary in terms of the uncertainty of the input data. The procedure is shown to be stable in the sense that these bounds tend to zero as the data errors go to zero.

PACS numbers: 02.30.Rz, 02.60. + y, 12.35. − i

I. INTRODUCTION

A. Summary

The problem of stabilized analytic continuation from the interior to the boundary of the domain of holomorphy has been studied in detail in some recent papers.[1,2] The idea is to use physically based hypotheses to provide the stabilization which is necessary to define a meaningful continuation process. It is possible in this way to formulate a systematic procedure for testing a hypothesis against data, data which may be either experimental or, in many interesting cases, the results of a theoretical calculation. For example, the methods which are used to determine resonance parameters from QCD fall into this category. The data set considered in Ref. 1 is a finite one, that is, the continuation is made from a finite set of discrete interior points. In Ref. 2, the problem of continuation from a continuous data set is treated. An essential aspect of both problems is that the input data, whether experimental or theoretical, are not exact, and errors must be incorporated. The significance of introducing errors is more than the simple admission of possible inaccuracy in the input; it allows a flexibility in the output without which the continuation procedure would be meaningless. This is apparent when one considers the second case above where the input forms a continuum. If the input were treated as exact, with no provision for errors, the analytic continuation would be fully determined even though it would be unstable in the sense of Hadamard, i.e., extremely sensitive in relation to minute changes of the input.

In Ref. 2, the problem of continuation from a continuum, subject to a defined error function and stabilized as indicated above by means of a supplementary physical assumption, is solved. The required analytic function is expressed in terms of the solution of a Fredholm integral equation of the second kind. Having solved this problem, it is however important to find precise error bounds for the result. Specifically, we would wish to answer the following question. Let $X^0(s)$ be the result of the continuation procedure off the data given on γ; as in Ref. 2, we will suppose that the input region, denoted by γ, is a continuous open curve inside the cut s-plane, s being the relevant variable with respect to which the true function $X^T(s)$, the actual physical amplitude or Green's function, is supposed to be analytic. If $\|X^T(s) - X^0(s)\|_\gamma < \epsilon$, where $\|\cdot\|_\gamma$ is a certain χ^2-type norm defined on γ, can we say that $|X^T(s) - X^0(s)| \to 0$, when $\epsilon \to 0$, for all points s of the holomorphy domain of $X^T(s)$, and, specifically, on the cuts? This question is answered in the affirmative. This stability analysis is the main purpose of the present paper and it is carried out in Secs. III and IV, where we obtain precise bounds on $|X^T(s) - X^0(s)|$ in the spectral region (on the cuts) in terms of ϵ. But before proceeding to this detailed analysis, it seems desirable to discuss the physical background to the problem; so, in Sec. II, we shall outline and develop the results obtained in Ref. 2.

Throughout this paper, we shall use the variable $z \equiv z(s)$ which maps the holomorphy domain of the function of interest onto the unit disk, so that the boundary cuts come on the unit circle $|z| = 1$, and the data region γ becomes a continuous curve inside the unit z-disk.

B. Background

An important problem in physics, which has attracted much attention, is to extend the results obtained from perturbative calculations to yield information which is essentially nonperturbative. The background to this is that very often the only method of calculation which is available is an iterative procedure, whereas many results of physical importance are nonperturbative in the sense that they will not be revealed by a standard perturbation calculation. One well-known method which has been applied to this problem is that of Borel summation,[3] where the information contained in the numerical values of the coefficients of a series, possibly divergent, is used to construct an integral representation with a specified domain of validity, provided certain conditions are satisfied. Unfortunately, for the problems of physical interest, it is often impossible to ascertain whether these conditions are met.[4]

 0022-2488/84/113194-10$02.50

Interest in this problem has received a particular stimulus from QCD.[5] The relevance here arises from the fact that the calculations which can be carried out in QCD (and which typically will include nonperturbative as well as perturbative components) are confined to a domain of validity which does not extend into the physical region and so, in particular, they cannot determine resonance parameters directly.

The hope of extending results calculated within one region to regions of physical interest which are outside the domain of validity of the original calculations, is based on analytic continuation. But analytic continuation off open curves—this is the problem of interest, the usual objective being to continue to the boundaries of the holomorphy domain (to the cuts)—is infinitely unstable in the Hadamard sense. This means that although the continuation is unique, errors in the input are magnified arbitrarily, so that without some stabilizing recipe, the result is meaningless. There is also the difficulty that the truncated perturbation expansion which forms the input is itself an analytic function, so that a straightforward analytic continuation of the precise input data would yield exactly the same perturbative function which is known to be incorrect in the resonance region. This highlights the importance of introducing errors in the data[2]; as we discussed above, quite apart from the recognition that the perturbative input is only approximate, errors are essential in order to allow sufficient flexibility in the analytic continuation that it may be possible to obtain the true function as a possible output, and not be uniquely restricted to the false perturbative function.

It must be clear from this, that analytic continuation alone cannot achieve the objective being sought; some supplementary information must be introduced to stabilize the problem and to remove ambiguity. The way in which this can be done, and the resolution of the above difficulties, is discussed at some length in Refs. 1 and 2. It is shown there how this supplementary physical information (which includes information about permitted behavior at singular points such as threshold and infinity, and which also includes physically based hypotheses about the types of structure permitted, such as discrete resonances) may be incorporated into the problem through a "filter"[1] acting within the function space. The operation of this filter, which acts by means of a suitably defined norm on the function space, is outlined in Sec. 2. We also show there how the discrepancy method[6] can be used to incorporate whatever hypothesis we wish to make about the type of structure to be permitted.

It should perhaps be emphasized that the need for stabilizing information arises generally in problems of this kind. Even if one tries, for example, to sum the perturbative series by means of a Bethe–Salpeter (or equivalent) integral equation, it does not mean that this difficulty has been circumvented. The terms of a perturbative series may or may not contain some latent information about the sought nonperturbative effect (see cases A and B discussed in Ref. 2). But even in case A, when the terms of the series *do* contain such latent information [as in the case with the coefficients of the series $S = 1 + z + z^2 \ldots$, which contain all the needed information about the position and residue of the pole $S = 1/(1 - z)$], it is necessary to go through a Hadamard ill-posed problem in order to make this information explicit, since indeed the results depend critically on the small terms of the above series.

Hence, to obtain a stable output, some supplementary information or assumption is required. This may be the algebraic semigroup symmetry $z \circ S = S - 1$, which is exact if $S \equiv 1 + z + z \circ z + \cdots$ but which otherwise represents a strong statement about the irrelevance of any other possible small terms of S. (If $\circ$ means ordinary algebraic multiplication, S is the geometric series referred to above; if $\circ$ means a convolution integral and $z \circ z$, $z \circ z \circ z$,... describe, for example, ladder graphs, then the integral equation $z \circ S = S - 1$ is just the Bethe–Salpeter equation.) Similar stabilizing symmetries are used when deriving linear integral equations for planar ("parquet") or other graphs configurations. For a general review of the present status of the theory, the reader may consult the recent and comprehensive paper by Jackson, Lande, and Smith.[7]

It is important to notice that it is not sufficient simply to assert that all terms which do not satisfy the semigroup translation symmetry $z \circ S = S - 1$ are small with respect to z (or to $z \circ z$, or to $z \circ z \circ z$), since the position and residue of the pole $1/(1 - z)$ are not determined by the coefficients of z or z^2 but in fact just by these small terms (the terms "..." in the expansions above!). Indeed, if one takes $10z^2$ instead of z^2 in the above sum, one simply gets $1/(1 - z) + 9z^2$ instead of $1/(1 - z)$, the parameters of the pole remaining unchanged!

So, one sees that the resonances and the other nonperturbative effects found by summing perturbative graphs by means of integral equations, depend critically on the algebraic symmetry which has been used in deriving these equations. This is a serious problem, the more so as the imprecisions of the small terms are not restricted only to some small graphs which possibly remain outside the summation scheme, but also to the imprecisions of the Lagrangian itself. This might effectively be the case with Higgs-like Lagrangian transformations and saddle point methods used with functional integrations in order to derive effective Lagrangians.

Bearing all that in mind, we attach much importance to the requirement that the stabilizing procedures should be based on physically controllable facts (or on physically controllable hypotheses) rather than on purely algebraic assumptions. Papers like that of Jackson, Lande, and Smith, where the results are checked against alternative procedures—variational schemes and hypernetted chain approximation in their case[7]—are extremely valuable for a sound founding of the theory. The methods described in this paper (and in Refs. 1 and 2) are intended to reflect this emphasis.

II. SOLUTION OF THE CONTINUATION PROBLEM—THE INTEGRAL EQUATION

A standard procedure is to map the complex E-plane, with cuts along the real axis, into the unit disk $|z| \leqslant 1$ so that the cuts map onto the unit circle[8] and the segment of the real axis which is not a cut becomes the diameter $-1 \leqslant z \leqslant 1$. The data region is a continuum, denoted by γ, which for convenience (and also because in practice this is frequently the

case) is taken to lie on the real axis γ: $z_1 \leqslant z \leqslant z_2$, where $|z_1|$, $|z_2| < 1$. The data (which we may, if we wish, suppose to be the result of a perturbation calculation) will be denoted by $a(z_\gamma)$, where $z_\gamma \in \gamma$, and we will suppose that on γ these differ from the values of the true function (the actual physical amplitude) $A(z_\gamma)$ by only a small amount $\epsilon(z_\gamma)$:

$$a(z_\gamma) = A(z_\gamma) + \epsilon(z_\gamma). \tag{1}$$

The actual deviation $\epsilon(z_\gamma)$ is of course unknown but we will require it to satisfy a χ^2-condition

$$\chi^2 \equiv \int_\gamma n(z)(\epsilon(z))^2 \, dz \leqslant 1, \tag{2}$$

where $n(z)$ gives a measure of the expected reliability of the data.

The discrepancy method,[6] which enables us to introduce a physically based hypothesis about the type of structure $A(z)$ may possess on the boundary, works in the following way.

Having decided on an appropriate hypothesis, we must express this in terms of a suitable trial function $T_k(z)$. We should stress that the trial function does not need to be, and in the spirit of our method will not normally be, an ansatz for the whole structure of the amplitude. It is only expected to describe this structure in some specific and limited range of the spectral region, and it will do this in terms of a set of variable parameters k. When testing the hypothesis T, one asks the following question: Can a set of parameters $k = k_0$ be found for which the data are compatible with the structure of the trial function $T_{k_0}(z)$? Since the parameters k typically describe such physically important quantities as resonance pole positions or residues, the determination of the values $k = k_0$ from the data is the physical problem in which we are interested and the motivation for the whole analysis. It is important to remember that any hypothesis $T_k(z)$ we may wish to test will be set against the data and rejected if it is found incompatible. For any trial function $T_k(z)$, the discrepancy function $D_k(z)$ is defined as

$$D_k(z) = A(z) - T_k(z), \tag{3}$$

and using the data $a(z_\gamma)$ for $A(z)$ we can define corresponding data $d_k(z_\gamma)$ for $D_k(z)$

$$d_k(z_\gamma) = a(z_\gamma) - T_k(z_\gamma). \tag{4}$$

A physical hypothesis which is of particular interest is that the dominant structure of the spectral function, particularly over some specified range of energy, comes from a set of discrete poles on the second Riemann sheet. This is represented by a trial function T_k which, when written as a function of the energy variable E, has the form

$$T_k(E) = \sum_{r=1}^{n} \frac{k_{1r} + ik_{2r}}{\sqrt{E} - (k_{3r} + ik_{4r})} - \frac{k_{1r} - ik_{2r}}{\sqrt{E} - (-k_{3r} + ik_{4r})}, \tag{5}$$

where k_{ir} must be real and k_{4r} negative to ensure that the poles are on the second sheet. As we have already stressed, the method is however in no way dependent on this particular form (5) of $T_k(E)$, which in what follows may be regarded as a quite general function.

The hypothesis that the structure of $A(z)$ on some specified part of the cuts can be described by $T_k(z)$ for suitable values k_0 of the parameters, may now be expressed in terms of the discrepancy function $D_k(z)$ by requiring that its structure on that part of the boundary $|z| = 1$ should be minimal. To implement this condition, one must define a measure of the structure on the boundary; this is done in terms of a norm on the function space, defined as follows:

$$\|X\| \equiv \delta[X] = \left\{ \frac{1}{2\pi} \int_0^{2\pi} (x_r(\phi))^2 \sigma(\phi) \, d\phi \right\}^{1/2}. \tag{6}$$

Here $X(z)$ are functions which are holomorphic in $|z| < 1$ and $x_r(\phi)$ are the "tangential derivatives" of their imaginary parts, defined as

$$x_r(\phi) \equiv \frac{\partial(\operatorname{Im} X(z' \equiv e^{i\phi}))}{\partial \phi} = \left. \frac{\partial(\operatorname{Re} X(z' \equiv re^{i\phi}))}{\partial r} \right|_{|z'|=1}, \tag{7}$$

and $\sigma(\phi)$ is an appropriately chosen, strictly positive weight function emphasizing that part of the boundary where we wish to test the hypothesis $T_k(z)$. We will normally consider only functions $X(z)$ which are real analytic, $X(\bar{z}) = \overline{X}(z)$, this means that $\sigma(\phi)$ may be defined as an even function $\sigma(-\phi) = \sigma(\phi)$, where $\sigma(-\phi) \equiv \sigma(2\pi - \phi)$. Other alternative norms can be defined, but this one is particularly suited to detect strong variations of the cross section [of Im $A(z)$]. Also, as will be seen in Sec. III, its stabilizing properties are particularly effective.

Strictly speaking, Eq. (6) does not define a norm over the whole space of holomorphic functions, as $x_r(\phi)$ will only determine the function $X(z)$ within the ambiguity of an arbitrary additive constant. However, if one considers the space of those functions $X(z)$ which vanish at some specified point $z = z_0$, which are holomorphic for $|z| < 1$, and have a tangential derivative $x_r(\phi)$ as defined in Eq. (7), then Eq. (6) does define a valid norm for this space. In this space $X(z)$ is uniquely determined by $x_r(\phi)$ and has the representation

$$X(z) = \frac{1}{\pi} \int_0^{2\pi} \ln \frac{e^{i\phi} - z_0}{e^{i\phi} - z} \, x_r(\phi) \, d\phi. \tag{8}$$

The kernel of Eq. (8) is the complex extension of the Neumann kernel

$$\mathcal{N}(z_0; z, z') \equiv 2 \ln |(z' - z_0)/(z' - z)| \tag{9}$$

(see the appendices of Ref. 9).

The norm $\|X\| \equiv \delta[X]$, as defined in Eq. (6), is a functional of the boundary derivative function $x_r(\phi)$. We define the functional $\mathcal{F}_1[x_r]$ as

$$\mathcal{F}_1[x_r] \equiv \delta^2[X] \equiv \frac{1}{2\pi} \int_0^{2\pi} (x_r(\phi))^2 \sigma(\phi) \, d\phi. \tag{10}$$

If we adopt the notation

$$D_k(z) = X(z) + d_0, \tag{11}$$

where

$$d_0 \equiv D_k(z_0) \tag{12}$$

we may use Eq. (8) to express the χ^2-condition (which is convenient to use in the form of an equality $\chi^2 = 1$, rather than

the inequality $\chi^2 \leqslant 1$) in terms of a second functional of the boundary derivative function $x_r(\phi)$, $\mathcal{F}_2[x_r]$, as follows:

$$\mathcal{F}_2[x_r] \equiv \chi^2[D_k] - 1 = \int_\gamma dz\, n(z)\left\{ d_k(z) - d_0 - \frac{1}{2\pi}\int_0^{2\pi} \mathcal{N}(z_0;z,e^{i\phi}) x_r(\phi)\, d\phi \right\}^2 - 1 = 0. \quad (13)$$

The Neumann kernel $\mathcal{N}(z_0;z,e^{i\phi})$ is defined in Eq. (9).

The problem to be solved is to find the extremum of the functional $\mathcal{F}_1[x_r]$ subject to the constraint $\mathcal{F}_2[x_r] = 0$. This may be done, using a Lagrange multiplier λ, by requiring that the Fréchet differential $\partial\mathcal{F}[x_r;y]$ of the functional

$$\mathcal{F}[x_r] \equiv \mathcal{F}_1[x_r] + \lambda \mathcal{F}_2[x_r] \quad (14)$$

should vanish, and also that the derivative of $\mathcal{F}[x_r]$, with respect to the subtraction constant d_0, should be zero. The Fréchet differential is a two-variable functional $\partial\mathcal{F}[x;y]$, possibly nonlinear in $x(\phi)$, but, by definition, linear in $y(\phi)$, which, when it exists, may be computed by means of the Gâteau differential formula

$$\partial\mathcal{F}[x;y] = \lim_{\alpha\to 0} \frac{\partial\mathcal{F}[x(\phi) + \alpha y(\phi)]}{\partial\alpha}, \quad (15)$$

where α is a c-number. Setting $\partial\mathcal{F}[x_r;y]$ equal to zero gives the result

$$\partial\mathcal{F}[x_r;y] \equiv \frac{1}{\pi}\int_0^{2\pi} d\phi\, y(\phi)\left\{ x_r(\phi)\sigma(\phi) - \lambda\left[\int_\gamma dz\, n(z)\mathcal{N}(z_0;z,e^{i\phi})(d_k(z) - d_0) - \frac{1}{2\pi}\int_0^{2\pi} d\phi'\, x_r(\phi') \int_\gamma dz\, n(z)\mathcal{N}(z_0;z,e^{i\phi}) \times \mathcal{N}(z_0;z,e^{i\phi'}) \right] \right\} = 0. \quad (16)$$

The requirement that $\partial\mathcal{F}/\partial d_0$ be zero yields the value of the constant d_0

$$d_0 = \frac{1}{n_\gamma}\int_\gamma n(z) d_k(z)\, dz - \frac{1}{n_\gamma}\int_\gamma dz\, n(z) \times \left\{ \frac{1}{2\pi}\int_0^{2\pi} d\phi\, \mathcal{N}(z_0;z,e^{i\phi}) x_r(\phi) \right\}, \quad (17)$$

where $n_\gamma \equiv \int_\gamma n(z)\, dz$. If we now substitute the value of d_0 given by Eq. (17) into Eq. (16), and use the condition that this equation must hold for any function $y(\phi)$, we are left with the following Fredholm integral equation for the boundary derivative function $x_r(\phi)$:

$$(x_r\sigma^{1/2})(\phi) = \lambda G_k(\phi) + \lambda \frac{1}{2\pi}\int_0^{2\pi} d\phi'\, K(\phi,\phi')(x_r\sigma^{1/2})(\phi'), \quad (18)$$

where

$$G_k(\phi) \equiv \frac{1}{\sigma^{1/2}(\phi)}\int_\gamma dz\, n(z)\mathcal{N}(z_0;z,e^{i\phi}) \times \left[d_k(z) - \frac{1}{n_\gamma}\int_\gamma dz'\, n(z') d_k(z') \right], \quad (19)$$

$$K(\phi,\phi') \equiv \frac{1}{\sigma^{1/2}(\phi)\sigma^{1/2}(\phi')} \times \left\{ \frac{1}{n_\gamma}\int_\gamma dz\, n(z)\mathcal{N}(z_0;z,e^{i\phi}) \cdot \int dz'\, n(z')\mathcal{N}(z_0;z',e^{i\phi'}) - \int_\gamma dz\, n(z)\mathcal{N}(z_0;z,e^{i\phi})\mathcal{N}(z_0;z,e^{i\phi'}) \right\}. \quad (20)$$

Notice that the kernel $K(\phi,\phi')$ is Hilbert–Schmidt (symmetric, and, in this case, also uniformly continuous).

Having solved Eq. (18) for $x_r(\phi)$ in terms of λ,[10] the solution must be inserted into Eq. (13) [after d_0 has been replaced by the right-hand side of Eq. (17)] in order to determine the value of λ. The optimal boundary derivative function $x_r^0(\phi)$ thus obtained can be inserted into Eq. (8) to yield the function $X^0(z)$ for $|z| < 1$. Using the value of d_0 from Eq. (17) in Eq. (11), we obtain the optimal discrepancy function $D_k^0(z)$ for the parameter set k. The minimum value $\delta_0(k)$ obtained for the functional $\delta[D_k^0]$ is given by

$$\delta_0(k) \equiv \delta[D_k^0] = \left\{ \frac{1}{2\pi}\int_0^{2\pi} d\phi\, (x_r^0(\phi))^2\sigma(\phi) \right\}^{1/2}. \quad (21)$$

The above calculation allows one to evaluate $\delta_0(k)$ for any parameter set k. If $\delta_0(k)$ has a distinct minimum at $k = k_0$ then one can say that the data favor the particular trial function $T_{k_0}(z)$, and, within the context of the hypothesis represented by T_k, one can say that the data have selected the values $k = k_0$ of the parameters. As well as selecting a set of parameters within one hypothesis, the method also allows the data to discriminate between hypotheses. Thus, for an inadequate hypothesis $T_k(z)$, $\delta_0(k)$ would be expected to be consistently large, with no pronounced minimum. If, on the other hand $\delta_0(k)$ is small with no significant minimum, the conclusion must be that the data are not sufficiently accurate to provide an adequate evaluation of the hypothesis.

When a set of parameters k_0 has been selected, the function $A_{k_0}^0(z)$, which defines the analytic continuation from the data according to the criteria and procedures described above, is now completely determined: It is the following function:

$$A_{k_0}^0(z) = d_0 + \frac{1}{\pi}\int_0^{2\pi} \ln\left(\frac{e^{i\phi} - z_0}{e^{i\phi} - z}\right) x_r^0(\phi)\, d\phi + T_{k_0}(z), \quad (22)$$

where the constant d_0 is given by Eq. (17), and where $x_r^0(\phi)$ now stands for the solution of the Fredholm integral equation (18) when $k = k_0$.

III. AN EXPLICIT ERROR BOUND

Let us suppose that the integral equation (18) corresponding to a specific physical problem has been solved and that the actual form of the function $X^0(z) \equiv D_k^0(z) - d_0$ having the least norm δ_0 [see Eq. (21)] is known. If the assumptions [i.e., the functions $T_k(z)$, $\sigma(z)$, $n(z)$] used to stabilize the continuation process were correct, the function $X^T \equiv A^T(z) - T_k(z) - d_0$ corresponding to the true but unknown amplitude $A^T(z)$ should have a norm δ_T (also unknown) not much larger than δ_0 itself. It is the aim of this and the next sections to derive bounds for the deviation function

$$F(z) \equiv X^T(z) - X^0(z) \equiv A^T(z) - A^0(z), \quad (23)$$

depending on δ_T, on δ_0, and on the precision $n(z)$ of the data in the input region γ. In Sec. IV, we shall derive the least (the best possible) upper bound $E^0(z)$, for given weights $\sigma(z)$ and $n(z)$, as the solution of an extremum problem (a Fredholm integral equation) similar to Eq. (18). Although the knowledge of the precise value of this optimal bound is very important when dealing with a specific physical problem, the solution of the corresponding Fredholm integral equation can be found only numerically. Therefore, we shall first derive here an approximate bound, $E(z)$, for the deviation function $|F(z)|$, which has the advantage of being explicitly computable and hence gives us direct insight into the contribution of the various factors. One might use this information to derive an optimal strategy, taking for instance a suitable balance between the values of the weight function $\sigma(z)$ inside and outside the energy "window," and so on.

Pursuing this objective, we make the following modification in the initial problem:

Instead of the inequality

$$\int_\gamma dz(X^0(z) - X^T(z))^2 n(z) \leqslant 4 \tag{24}$$

[which one obtains from the triangle inequality for the n-weighted $\mathscr{L}^2$-norms on γ of $X^0(z) - d_k(z) + d_0$ and $X^T(z) - d_k(z) + d_0$], we shall use the $\mathscr{L}^\infty$-condition

$$|F(z)| \equiv |X^0(z) - X^T(z)| < \epsilon \quad \text{for} \quad z \epsilon \gamma'. \tag{25}$$

Here γ' is some subset of γ. Indeed, in order that an inequality of the type

$$\int_\gamma dz |F(z)|^2 < \eta^2 l_\gamma \tag{26}$$

should hold [for simplicity we have taken here the case in which $n(z)$ is constant and equal to $1/\eta^2$, l_γ being the length of γ], it is necessary that

$$|F(z)|_\gamma < \epsilon \tag{27}$$

on some subset γ' of the initial interval γ; if $\epsilon = \alpha\eta$, with $\alpha > 1$, then the measure l_γ of γ' must be larger than $(1 - 1/\alpha^2) l_\gamma$.

One could have specified the accuracy of the data from the beginning by means of a point-wise inequality $|a^{\text{pert}}(z) - A(z)| < \epsilon(z)$ rather than by a $\mathscr{L}^2$ one. In this case the inequality (27) would have been the natural formulation of the problem, but we wanted to stress here that, with some loss of information, the inequality (27) also follows from our original $\mathscr{L}^2\chi^2$-condition.

The sole remaining information about $F(z)$ [the difference between the (known) optimal continuation function $X^0(z)$ and the true but unknown one, $X^T(z)$], is that it is holomorphic in the unit disk and that the σ-weighted $\mathscr{L}^2$-norm of the boundary values of the (tangential) derivative of its imaginary part

$$f_r(\phi) \equiv \frac{\partial \operatorname{Im} F(e^{i\phi})}{\partial \phi} = x_r^0(\phi) - x_r^T(\phi) \tag{28}$$

is bounded

$$\frac{1}{2\pi} \int_0^{2\pi} d\phi f_r^2(\phi)\sigma(\phi) \leqslant \Delta^2 \equiv (\delta_0 + \delta_T)^2 \leqslant (2\delta_T)^2. \tag{29}$$

Here δ_T is the δ-norm [Eq. (6)] of the discrepancy associated with the true (but unknown) amplitude $A^T(z)$. If the definition (6) of norm (which is required to be small) and the choice of the trial function $T_{k_0}(z)$ are physically reasonable, then δ_T should be small, certainly finite. From (29) and from the integral representation

$$F(z) = F(z_0) + \frac{1}{\pi} \int_0^{2\pi} d\phi' \ln\left(\frac{e^{i\phi'} - z_0}{e^{i\phi'} - z}\right) f_r(\phi') \tag{30}$$

of $F(z)$ for $|z| \leqslant 1$, we may show that $F(z)$ is bounded in the whole (closed) unit disk. Indeed, using the Schwartz inequality, we get

$$\begin{aligned} |F(z) - F(z_0)|^2 &\equiv \left| \frac{1}{2\pi} \int_0^{2\pi} d\phi' \, 2 \ln \frac{e^{i\phi'} - z_0}{e^{i\phi'} - z} \frac{\sigma^{1/2}(\phi')}{\sigma^{1/2}(\phi')} f_r(\phi') \right|^2 \\ &\leqslant \frac{1}{2\pi} \int_0^{2\pi} d\phi' \frac{4}{\sigma(\phi')} \left| \ln \frac{e^{i\phi'} - z_0}{e^{i\phi'} - z} \right|^2 \\ &\quad \times \frac{1}{2\pi} \int_0^{2\pi} d\phi' f_r^2(\phi')\sigma(\phi'). \end{aligned} \tag{31}$$

The second integral satisfies the bound (29), while the first is bounded even if z is on the unit circle. (If $z = e^{i\phi}$, the integral is finite, in spite of the logarithmic divergence of the integrand at $\phi' = \phi$.) Taking z_0 on γ', where $|F(z_0)| < \epsilon$, we thus obtain the result

$$|F(e^{i\phi})| \leqslant M. \tag{32}$$

Our aim is to show that $M \to 0$, when $\epsilon \to 0$. The proof which follows will use a combination of the Nevanlinna bound (see below, Sec. A), and of a limitation on the growth of $F(z)$ over any interval (z_1, z_2) contained in the closed unit disk $|z| \leqslant 1$ because of the inequality (31) (see Sec. B below).

A. The Nevanlinna bound

Define $\omega(z)$ to be that (real) harmonic function which is equal to one on γ' and zero on the unit circle Γ,

$$\begin{aligned} &\nabla^2 \omega(z) = 0 \quad \text{for} \quad z: |z| < 1, z \notin \gamma', \\ &\omega(z \in \gamma') = 1, \quad \omega(z \in \Gamma) = 0. \end{aligned} \tag{33}$$

Now let D' be the domain obtained if one removes from the disk $|z| < 1$ the points of γ' as well as small circular neighborhoods around the zeros of the function $F(z)$, each neighborhood being bounded by sufficiently small circles γ_i so that inside any of them $\ln |F(z)|$ will be less than both $\ln \epsilon$ and $\ln M$. The function $\ln F(z)$ being now holomorphic around each of the points of D', $\operatorname{Re} \ln F(z) \equiv \ln |F(z)|$ will be harmonic in D'. On the other hand, on the boundary $\partial D'$ of D' we have

$$\ln |F(z)| \leqslant \omega(z) \cdot \ln \epsilon + (1 - \omega(z)) \cdot \ln M, \quad z \in \partial D' \equiv \gamma' \cup \Gamma \cup \gamma_i. \tag{34}$$

[This follows on γ' from (27) and from $(1 - \omega(z))|_{\gamma'} = 0$, on Γ from (32) and $\omega(z)|_\Gamma = 0$, and on the circles γ_i from the fact that $|F(z)|$ is (there) less than $\min(\epsilon, M)$ and that $\omega(z) \leqslant 1$ throughout the whole unit disk.]

But both sides of the inequality (34) being harmonic functions, this property, which is valid initially for the boundary $\partial D'$, is automatically extended throughout the whole domain D', and since it is also satisfied inside the small

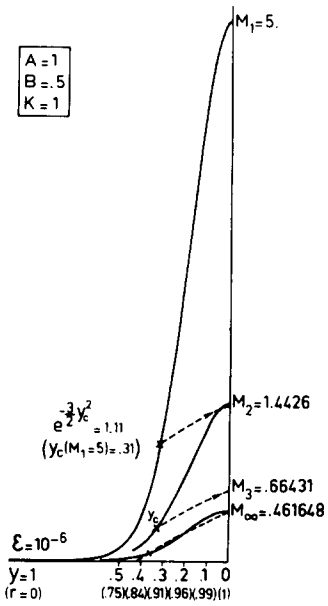


FIG. 1. Interplay between the Nevanlinna and the derivative bound near $r = 1$, in y coordinates ($r = 1 - y^2$).

neighborhoods of the zeros of $F(z)$, it is valid throughout the whole (closed) unit disk. Hence,

$$|F(z)| \leqslant M\left(\frac{M}{\epsilon}\right)^{-\omega(z)} \quad \text{for all} \quad |z| \leqslant 1. \tag{35}$$

The Nevanlinna bound (35) is not expected to be the best bound which one could find for $|F(z)|$ from (27) and (32), as the inequality (35) would have been valid even if γ' had been a part of the boundary of the holomorphy domain of $F(z)$, whereas in fact, γ' is completely embedded into the latter. However, (35) is sufficient to provide, in combination with (31), a mechanism which makes M vanish with ϵ.

Although (35) predicts that for every interior point $|z| < 1$, $|F(z)|$ tends to zero when ϵ tends to zero [for such points $\omega(z)$ is strictly positive], the condition (35) has no predictive value at all for boundary points, since the inequality (35) merely repeats there the already known fact that $|F(z)|$ is bounded on Γ by $M(\omega(z = e^{i\phi}) \equiv 0!)$ However, even if no better condition could be obtained, this would mean that, at least for small ϵ's, a huge increase of the modulus of $F(z)$ would be allowed to take place in the immediate neighborhood of the boundary (see Fig. 1). But considerations similar to those which lead to Eq. (31) limit the rate of growth of $|F(z)|$ near $|z| = 1$ (see Sec. B below) and this in turn imposes strong constraints on M, linking its value to that of ϵ.

The above discussion emphasizes the importance of the way in which the Nevanlinna bound behaves at points $z = (1 - x)e^{i\phi}$, for x small. As it is well known from elementary potential theory, $\omega(z)$ should be harmonic in the entire complex plane cut along γ' [where $\omega(z \in \gamma') = +1$] and along the image γ'' of γ' through the unit circle Γ, where $\omega(z \in \gamma'') = -1$. This means that the points of Γ are all within the domain in which $\omega(z)$ is harmonic, with the result that $\omega(z)$ has a well-defined gradient at points on the circle, so that we may write

$$\omega(z) = (1 - x)e^{i\phi}) = K(\phi)x + \cdots, \tag{36}$$

where

$$-K(\phi) \equiv \left.\frac{\partial \omega(re^{i\phi})}{\partial r}\right|_{r=1}$$

is positive and bounded. Hence, for points close to Γ, the Nevanlinna bound (35) may be rewritten as

$$|F(z = (1 - x)e^{i\phi})| \leqslant M \exp[-K(\phi)\ln(M/\epsilon)x]. \tag{37}$$

B. A limit for the growth of $F(z)$

For large M and small ϵ, the right-hand side of the inequality (37) has a very sharp growth when x tends to zero ($|z| \to 1$). However, the other inequality (31) severely limits the rate of growth of $F(z)$ near the unit circle. Indeed, writing $1/\sigma(\phi)$ as the real part of the boundary values of the holomorphic weight function $S(z)$ (see Ref. 9), where

$$S(z) = \frac{1}{2\pi}\int_0^{2\pi} d\phi' \frac{e^{i\phi'} + z}{e^{i\phi'} - z}\frac{1}{\sigma(\phi')}, \quad \frac{1}{\sigma(\phi)} \equiv \mathrm{Re}\, S(e^{i\phi}), \tag{38}$$

the bound (31) reads

$$|F(z_0) - F(z_x)| \leqslant (\delta_T + \delta_0)\left\{\mathrm{Re}\frac{4}{2\pi}\int_0^{2\pi} d\phi' \times \left|\ln\frac{e^{i\phi'} - z_x}{e^{i\phi'} - z_0}\right|^2 S(e^{i\phi'})\right\}^{1/2} \equiv (\delta_T + \delta_0)(\mathrm{Re}\, I)^{1/2}. \tag{39}$$

To see how (39) limits the growth of $F(z)$ near the boundary we take $z_x = (1 - x)e^{i\phi}$ and $z_0 = e^{i\phi}$. The integral

$$I = \frac{4}{2\pi i}\oint \frac{d(e^{i\phi'})}{e^{i\phi'}} S(e^{i\phi'})\ln\left(\frac{e^{i\phi'} - z_x}{e^{i\phi} - z_0}\right)\ln\left(\frac{1 - e^{i\phi'}\bar{z}_x}{1 - e^{i\phi'}\bar{z}_0}\right) \tag{40}$$

may then be computed by moving the integration contour around the singularities of the integrand. Observing that the residue of the pole at $z' = 0$ is zero so that the sole contribution to the integral is due to the cut of the first of the two logarithms of (40) along the radial segment $z' = r'e^{i\phi}$, $1 - x \leqslant r' \leqslant 1$, we get the result

$$\mathrm{Re}\, I = \mathrm{Re}\, 4\int_{z_x}^{z_0} dz' \frac{S(z')}{z'}\ln\frac{1 - z'\bar{z}_x}{1 - z'\bar{z}_0} \equiv 4\int_{1-x}^{1} dr' \frac{\mathrm{Re}\, S(r'e^{i\phi})}{r'}\ln\frac{1 - r'(1 - x)}{(1 - r')}. \tag{41}$$

If x is small, $\mathrm{Re}(S(r'e^{i\phi}))/r'$ can be expanded around $z' = z_0 (\equiv e^{i\phi})$,

$$\frac{\mathrm{Re}(S(z'))}{r'} \equiv \mathrm{Re}\left(z_0\frac{S(z')}{z'}\right) = \mathrm{Re}\left\{z_0\left[\frac{S(z_0)}{z_0} + \frac{1}{1!}\left(\frac{S'(z_0)}{z_0} - \frac{S(z_0)}{z_0^2}\right)(z' - z_0) + \cdots\right]\right\}$$

$$= \mathrm{Re}\, S(z_0) + (r' - 1)\left\{\frac{\partial\, \mathrm{Re}\, S(z')}{\partial r'}\Big|_{r'=1} - \mathrm{Re}\, S(z_0)\right\} + \cdots, \tag{42}$$

where we have used the fact that $(z' - z_0) \equiv z_0(r' - 1)$ and that

$$\mathrm{Re}(S'(z_0)\cdot z_0) = \left.\frac{\partial(\mathrm{Re}\, S(re^{i\phi}))}{\partial r}\right|_{r=1}.$$

The bound (39) on the rate of growth of $F(z)$ in the neighborhood of the boundary, in terms of the value of the weight function $S(z)$ and its radial derivative at $z = z_0$ now becomes

$$|F(e^{i\phi}) - F((1-x)e^{i\phi})|^2 \leqslant (\Delta F(\phi))^2 \equiv (\delta_0 + \delta_T)^2 \left\{ x 8 \ln 2\, \mathrm{Re}\, S(z_0) + 2x^2 \left[(2\ln 2 - 1)\mathrm{Re}\, S(z_0) - \left.\frac{\partial\, \mathrm{Re}\, S(z')}{\partial r'}\right|_{r'=1} \right] \right\} + \cdots . \tag{43}$$

This may be rephrased as

$$|(Fe^{i\phi}) - F((1-y^2)e^{i\phi})| \leqslant (\Delta F(\phi))^2 = yA(\phi) + y^3 B(\phi) + O(y^5), \tag{44a}$$

where we have introduced the notation $y = \sqrt{x}$ and the coefficient functions $A(\phi)$, $B(\phi)$:

$$A(\phi) \equiv 2\sqrt{2\ln 2}\,\sqrt{\mathrm{Re}\, S(e^{i\phi})}(\delta_0 + \delta_T), \tag{44b}$$

$$B(\phi) = \frac{1}{2(2\ln 2)^{1/2}}\left[(2\ln 2 - 1)\sqrt{\mathrm{Re}\, S(e^{i\phi})} - 2\left.\frac{\partial\sqrt{\mathrm{Re}\, S(r'e^{i\phi})}}{\partial r'}\right|_{r'=1}\right](\delta_0 + \delta_T).$$

If one is interested to have a bound for $|F(e^{i\phi}) - F((1-y^2)e^{i\phi})|$ independent of the magnitude of y, this may easily be written in terms of the upper bound $\overline{\mathrm{Re}\, S(e^{i\phi})}$ of $\mathrm{Re}\, S(re^{i\phi})$ in the interval $1 - y^2 < r < 1$. It reads

$$|F(e^{i\phi}) - F((1-y^2)e^{i\phi})| \leqslant y2\sqrt{(2-y^2)\ln 2}/(1-y^2)\cdot(\overline{\mathrm{Re}\, S(e^{i\phi})})^{1/2}(\delta_0 + \delta_T). \tag{44c}$$

Equations (44a) and (44b) make explicit the relative contributions of $\mathrm{Re}\, S(e^{i\phi}) \equiv 1/\sigma(\phi)$ and the derivative term $\partial\, \mathrm{Re}\, S(e^{i\phi})/\partial r$. This is of practical importance in devising an optimum strategy with regard to the choice of the weight function $\sigma(\phi)$. On the one hand, we may wish to increase the accuracy within a particular range $[\phi_1, \phi_2]$ by making $\sigma(\phi)$ relatively large (and $\mathrm{Re}\, S$ small) within that interval, but this benefit is offset by the contribution introduced through the derivative term $\partial\, \mathrm{Re}\, S(e^{i\phi})/\partial r$ which can become large if $\sigma(\phi)$ has a large variation.[11]

C. Vanishing of the bound M of $F(z)$ on the boundary

If M is large with respect to ϵ, there will be a sharp drop of the Nevanlinna bound near $r = 1$, as one may see in Fig. 1 (where the variable $y = \sqrt{x} = \sqrt{1-r}$ has been used for convenience),

$$|F(z = (1-y^2)e^{i\phi})| \leqslant Me^{-Qy^2}, \tag{45}$$

where [see Eq. (37)] $Q \equiv K(\phi)\ln(M/\epsilon)$. But, as has been shown in Sec. B, such a violent variation of the modulus of F is ruled out. So, starting with some $M = M_0$ as in Fig. 1, we will end up with a smaller $M = M_1$, if we take account of the fact that the modulus of the increase of F between $z_x = (1-y^2)e^{i\phi}$ and $z_0 = e^{i\phi}$ cannot exceed $\Delta F(\phi)$ given by Eq. (44).

To simplify the discussion, we shall suppose for the moment that the coefficients entering the right-hand side of Eq. (44) do not depend on ϕ, or, more correctly, that by A, B, and K we understand $\sup_\phi(A(\phi))$, $\sup_\phi(B(\phi))$ and $\inf_\phi K(\phi)$, respectively. At the end of this subsection we shall drop this condition.

The new value M_1 which replaces M depends on the point y beyond which the rule (44) is applied. Obviously, the lowest M_1 is obtained if y_1 is chosen to be the tangential point y_t where the curves Me^{-Qy^2} and const $- Ay - By^3$ touch, the value of the constant being chosen to ensure tangency. This corresponds to M given by the implicit equation

$$M = \inf_{0<y<1}(Me^{-Qy^2} + Ay + By^3). \tag{46}$$

However, in order to simplify the calculation, instead of y_t, we shall use the (nearby) point y_c, where the second derivative is maximal, given by the condition

$$\frac{d^3 e^{-Qy^2}}{dy^3} \equiv -8Q^3 e^{-Qy_c^2} y_c \left\{ y_c^2 - \frac{3}{2Q} \right\} = 0, \tag{47}$$

the relevant solution of which is $y_c = \sqrt{3/2Q}$.

Substituting this value of y as an approximate solution in Eq. (46) above, we get the following fixed-point equation for M:

$$M = Me^{-3/2} + \sqrt{\frac{3}{2Q}}\left(A + \frac{3B}{2Q}\right), \tag{48}$$

which may be rewritten in the following form, which, if ϵ is small, is convenient for iterations:

$$M = \frac{e^{3/2}}{(e^{3/2}-1)}\sqrt{\frac{3}{2K\ln(M/\epsilon)}}\left(A + \frac{3B}{2K\ln(M/\epsilon)}\right). \tag{49}$$

(See Fig. 2.) An asymptotic solution M_0, valid for vanishingly small ϵ's, which is necessary if we want to prove that $M_0 \to 0$ when $\epsilon \to 0$, can readily be found by observing that for large values of the ratio M/ϵ the B term is negligible with respect to the A term. Hence, for very small values of ϵ, we have

$$\frac{M^2}{\epsilon^2}\ln\left(\frac{M}{\epsilon}\right)^2 = \frac{3e^3A^2}{(e^{3/2}-1)^2K}\frac{1}{\epsilon^2}. \tag{50}$$

But the asymptotic form of the solution of the transcendent equation $X\ln X = a$, for $a \to \infty$, $(a \sim 1/\epsilon^2)$, is $X_0 = a/\ln a$; indeed $\lim_{a\to\infty}(1/a)X_0 \ln X_0 = \lim_{a\to\infty}\{(1/\ln a)(\ln a - \ln\ln a)\} = 1$. Hence the corresponding form for M is

$$M_0 = \frac{e^{3/2}}{(e^{3/2}-1)}\frac{\sqrt{3}A}{\sqrt{2}\sqrt{K}}\left[\ln\left(\frac{e^{3/2}}{(e^{3/2}-1)}\frac{\sqrt{3}A}{\sqrt{K}}\frac{1}{\epsilon}\right)\right]^{-1/2}. \tag{51}$$

So, we have shown that the modulus of the difference $F(z)$ between the unknown amplitude $A^T(z)$ and the comput-

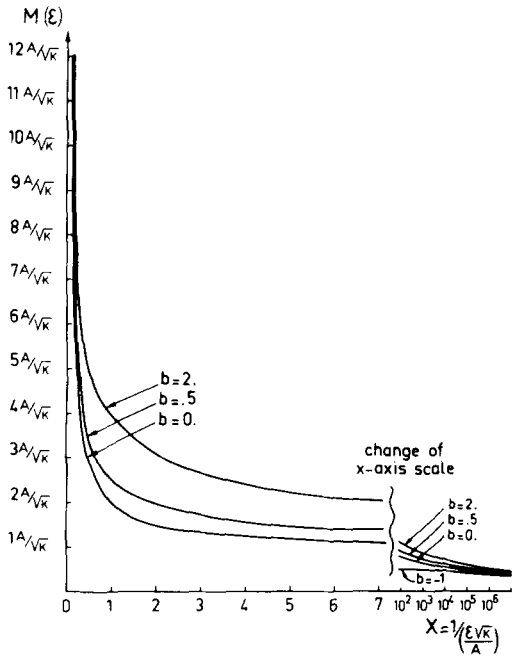


FIG. 2. Because of its internal symmetries, it is convenient to rewrite Eq. (49) in terms of $m \equiv M/(A/\sqrt{K})$ and $\epsilon' \equiv \epsilon/(A/\sqrt{K})$. Then, if $b = B/(AK)$, we have

$$m = \frac{e^{1.5}}{e^{1.5}-\epsilon'}\frac{3^{1/2}}{2^{1/2}}\frac{1}{(\ln(m/\epsilon'))^{1/2}}\left\{1+\frac{3}{2}\frac{b}{\ln(m/\epsilon')}\right\}.$$

In Fig. 2 the dependence on m versus $x = 1/\epsilon'$ is plotted for three different values of b. One should notice the initial sharp drop of $M = m(A/\sqrt{K})$ with $1/\epsilon$, which slows down considerably for x greater than that corresponding to the maximum curvature point. Nevertheless, $M \to 0$, when $x \to \infty$ ($\epsilon \to 0$).

ed one $A^{0}(z)$ does tend to zero when the errors of the data tend to zero, even on the boundary $|z| = 1$ of the domain of holomorphy. This is not a trivial consequence of the Nevanlinna principle alone, but follows from combining this with the stabilizing condition used on the cuts. To compute M for a value of ϵ different from zero, one may start with the asymptotic form (51) and iterate it using Eq. (49); since the right-hand side of (51) contains only terms which vary slowly with M, the whole procedure converges rapidly.

In the above calculation, we took no account of the ϕ-dependence of either $K(\phi)$ or of the coefficients $A(\phi)$ and $B(\phi)$. We shall now indicate an approximate procedure for introducing this ϕ-dependence, which will yield a ϕ-dependent bound on the unit circle

$$|F(e^{i\phi})| < \mathcal{M}(\phi). \tag{52}$$

In the next section, we shall describe the calculation which leads to the best possible bounds for $F(z)$ at each point z; here we are concerned to obtain a result which is explicit, thus displaying the individual effects of the various terms, even though in deriving this, we make approximations so that the bound obtained is not the optimal one. The iterative procedure outlined below is a reasonable approximation for small ϵ and provided that γ does not extend too close to the unit circle. We define a weight function $C(z)$, holomorphic and without zeroes in the unit disk, in terms of a real function $\Omega(\phi)$, as follows:

$$C(z) = \exp\left\{\frac{1}{2\pi}\int_0^{2\pi} d\phi\,\frac{e^{i\phi}+z}{e^{i\phi}-z}(\Omega(\phi)-\overline{\Omega})\right\}, \tag{53a}$$

where

$$\overline{\Omega} \equiv \frac{1}{2\pi}\int_0^{2\pi} d\phi\,\Omega(\phi). \tag{53b}$$

The above construction ensures that $|C(0)| = 1$, since

$$\ln|C(z)| = \frac{1}{2\pi}\int_0^{2\pi} d\phi\,\mathrm{Re}\left(\frac{e^{i\phi}+z}{e^{i\phi}-z}\right)(\Omega(\phi)-\overline{\Omega}). \tag{54}$$

Being a harmonic function, the value of $\ln|C(z)|$ at $z = 0$ is the mean of its value $(\Omega(\phi)-\overline{\Omega})$ on the boundary (which is zero by construction). We now use this weight function $C(z)$, which is allowed to have an arbitrary ϕ-dependence, to replace $F(z)$ by $\tilde{F}(z)$

$$F(z) = \tilde{F}(z)C(z). \tag{55}$$

Having defined $C(z)$ so that $|C(0)| = 1$, we shall make the approximation of supposing that the error ϵ applies unchanged to $\tilde{F}(z)$. This is reasonable provided that γ does not extend too far from the origin and also provided that the ϕ-dependence of Ω (which will be determined below, and which reflects the ϕ-dependence of A and B) is not too marked. If we now start with a bound M for $|\tilde{F}(e^{i\phi})|$ and proceed as before, since

$$|\tilde{F}(z)| \leqslant M(M/\epsilon)^{-\omega(z)}, \tag{56}$$

then for z close to (and within) the unit circle, [putting $z = (1-x)z_0$, $z_0 \equiv e^{i\phi}$], we find that

$$\begin{aligned}|F(z=(1-x)e^{i\phi})| &\leqslant M\exp[-K(\phi)(\ln(M/\epsilon))x]\,|C((1-x)e^{i\phi})| \\ &= M|C(z_0)|\exp\left[-K(\phi)\ln\left(\frac{M}{\epsilon}\right)x\right] \\ &\quad\times\left(1-x\left.\frac{\partial\ln|C(re^{i\phi})|}{\partial r}\right|_{r=1}+\cdots\right).\end{aligned} \tag{57}$$

Now define $\mathcal{M}(\phi)$ as

$$\mathcal{M}(\phi) \equiv M|C(z_0 \equiv e^{i\phi})|, \tag{58}$$

put $y^2 \equiv x$ as before, and combine Eq. (58) with the bound for ΔF [Eq. (44)], taking $y = y_c \equiv \sqrt{3/(2K(\phi)\ln(M/\epsilon))}$, to obtain the equation for $\mathcal{M}(\phi)$,

$$\begin{aligned}\mathcal{M}(\phi) &\equiv M|C(z_0)| \\ &= \frac{e^{3/2}}{\left(e^{3/2}-1-\left.\frac{\partial\ln|C(re^{i\phi})|}{\partial r}\right|_{r=1}\cdot\frac{3}{2K(\phi)\ln(M/\epsilon)}\right)} \\ &\quad\times\frac{\sqrt{3}}{[2K(\phi)\ln(M/\epsilon)]^{1/2}}\left\{A(\phi)+\frac{3B(\phi)}{2K(\phi)\ln(M/\epsilon)}\right\}.\end{aligned} \tag{59}$$

To solve this equation by iteration, one first neglects the term containing $\partial\ln|C(re^{i\phi})|/\partial r$ in the first denominator (this will be small relative to $e^{3/2}-1$ if ϵ is sufficiently small), and one gives M the asymptotic value M_0 from Eq. (51). This yields a solution $\mathcal{M}_1(\phi)$ whose ϕ-dependence comes from $K(\phi)$, $A(\phi)$, $B(\phi)$ and which determines the normalized function $\Omega(\phi)-\overline{\Omega}$ appropriate to this state of the iteration,

$$\Omega_1(\phi)-\overline{\Omega}_1 = \ln|C_1(e^{i\phi})| \equiv \ln(\mathcal{M}_1(\phi)/M_1),$$

where

$$\ln M_1 \equiv \frac{1}{2\pi}\int_0^{2\pi} d\phi \ln \mathscr{M}_1(\phi). \tag{60}$$

Equation (60) follows from Eq. (53a) which may now be used to construct $C(z)$ and thus $\partial \ln|C(re^{i\phi})|/\partial r|_{r=1}$. For small ϵ, the iteration will converge rapidly to yield the desired bound $\mathscr{M}(\phi)$ for the difference between the true (unknown) function $A^T(e^{i\phi})$ and the computed one $A^0(e^{i\phi})$ for each point $e^{i\phi}$ on the unit circle.

Although the stability of the whole procedure has been proven (i.e., the vanishing of M for $\epsilon \to 0$), the dependence of M on ϵ in Eq. (51) is, of course, a weak one [$1/\sqrt{\ln(1/\epsilon)}$, as $\epsilon \to 0$]. It is interesting to look at the graph of M with respect to $1/\epsilon$: At the beginning the variation of M with $1/\epsilon$ is large, and so, the value of $1/\epsilon_c$ for which this curve has a maximum curvature gives a useful target value for the required precision of the perturbative calculus. [For example, for the function $1/\sqrt{\ln X}$ this point is $X_c = 1.591$, which, if we approximate M by its asymptotic form (51), corresponds to an $\epsilon_c = 1.40A/\sqrt{K}$. Since this point is near that where the derivative of the right-hand side of Eq. (49) with respect to M equals -1, ϵ_c may easily be recognized numerically by the slowing down/failure of the iteration (49) for $\epsilon > \epsilon_c$.] The benefit from reducing ϵ much below this level is not great.

Once again, we emphasize that the method of this section, and in particular the way in which the Nevanlinna bound was applied [bearing in mind that $F(z)$ is analytic on γ], involved approximations which were not required in principle. Thus, the bounds derived above are not the best possible,[12] although their explicit form makes them interesting and valuable. In the next section, we consider the problem of determining the best possible bound as a function of ϕ and show that this satisfies a Fredholm equation.

IV. DERIVATION OF THE OPTIMAL BOUND $E^0(z)$

The problem to be solved may be stated simply. The function $F(z) \equiv A^T(z) - A^0(z)$ is required (a) to be holomorphic in the unit disk, (b) to satisfy a χ^2-condition on γ of the form

$$\chi^2[F] \equiv \int_\gamma dz\, n(z)|F(z)|^2 = 4, \tag{61}$$

and (c) to have a norm less than some constraint Δ^2,

$$\delta^2[F] \equiv \frac{1}{2\pi}\int_0^{2\pi} (f_r(\phi))^2 \sigma(\phi) d\phi < \Delta^2.$$

Subject only to these conditions, we want to know the maximum value $E^0(z)$ which $|F(z)|$ could have at any point in the unit disk and more particularly on the unit circle.

Observe that the value of $|F(z)|$ at a specified point z is a functional of $F(z)$. But χ^2 and δ^2 are also functionals of $F(z)$, so that the problem may be completely expressed in terms of these three functionals. Instead of being expressed in terms of F, they may equally be regarded as functionals of the boundary derivative function f_r, since $f_r(\phi)$ [together with the subtraction constant $d_0 \equiv F(z_0)$] determines $F(z)$. So, we define

$$\mathscr{F}_1[f_r] \equiv \int_\gamma dz\, n(z) \times \left\{ d_0 + \frac{1}{\pi}\int_0^{2\pi} \ln\left(\frac{e^{i\phi} - z_0}{e^{i\phi} - z}\right) f_r(\phi) d\phi \right\}^2 - 4, \tag{62}$$

$$\mathscr{F}_2[f_r] \equiv \frac{1}{2\pi}\int_0^{2\pi} (f_r(\phi))^2 \sigma(\phi) d\phi - \Delta^2, \tag{63}$$

$$\mathscr{F}_3^{(z)}[f_r] = \left| d_0 + \frac{1}{\pi}\int_0^{2\pi} \ln\left(\frac{e^{i\phi} - z_0}{e^{i\phi} - z}\right) f_r(\phi) d\phi \right|^2 \quad (\equiv (E^0(z))^2). \tag{64}$$

The conditions $\chi^2 = 4$ and $\delta^2 = \Delta^2$ thus become $\mathscr{F}_1[f_r] = 0$ and $\mathscr{F}_2[f_r] = 0$, respectively; these are the two constraints subject to which we must maximize $\mathscr{F}_3^{(z)}[f_r]$. To incorporate the constraints, we introduce two Lagrange multipliers and define the combined functional

$$\mathscr{F}[f_r] = \lambda_1 \mathscr{F}_1[f_r] + \mathscr{F}_2[f_r] + \lambda_3 \mathscr{F}_3^{(z)}[f_r]. \tag{65}$$

(We have written $\mathscr{F}$ as $\lambda_1\mathscr{F}_1 + \mathscr{F}_2 + \lambda_3\mathscr{F}_3$ rather than $\lambda_1\mathscr{F}_1 + \lambda_2\mathscr{F}_2 + \mathscr{F}_3$ for notational convenience.) The extremum condition is that the Fréchet differential of $\mathscr{F}$ and the derivative of $\mathscr{F}$ with respect to d_0 should both vanish,

$$\partial\mathscr{F}[f_r; y] = 0, \tag{66}$$

$$\frac{\partial\mathscr{F}}{\partial d_0} = 0. \tag{67}$$

As an alternative problem, we can look for bounds on the real or imaginary parts. To do this the functional $\mathscr{F}_3^{(z)}[f_r]$ of Eq. (64) has simply to be replaced by the appropriate forms. Thus the value of Re $F(z)$ at a specified point z is given by

$$\mathscr{F}_3^{(z,re)}[f_r] = d_0 + \frac{1}{2\pi}\int_0^{2\pi} \mathscr{N}(z_0;z,e^{i\phi}) f_r(\phi) d\phi \quad (\equiv E^{0re}(z)), \tag{64'}$$

and Im $F(z)$ by

$$\mathscr{F}_3^{(z,im)}[f_r] = \frac{1}{2\pi}\int_0^{2\pi} \mathscr{M}(z_0;z,e^{i\phi}) f_r(\phi) d\phi \quad (\equiv E^{0im}(z)), \tag{64''}$$

where $\mathscr{N}(z_0;z,z')$ and $\mathscr{M}(z_0;z,z')$ are the real and imaginary parts, respectively, of $2\ln((z'-z_0)/(z'-z))$. The two functionals for Re $F(z)$ and Im $F(z)$ are simpler than that for $|F(z)|$, which is quadratic in f_r; so, in order to keep the derivation as simple as possible, and because the bounds for Re $F(z)$ and Im $F(z)$ are of comparable physical interest, we choose the Re $F(z)$ case for the derivation which follows. In this case Eq. (66) gives

$$f_r(\phi)\sigma(\phi) + \frac{\lambda_3}{2}\mathscr{N}(z_0;z,e^{i\phi}) + \lambda_1 d_0 \int_\gamma dz'\, n(z')\mathscr{N}(z_0;z',e^{i\phi}) + \frac{\lambda_1}{2\pi}\int_0^{2\pi} d\phi' \int_\gamma dz'\, n(z') \times \mathscr{N}(z_0;z',e^{i\phi})\mathscr{N}(z_0;z',e^{i\phi'}) f_r(\phi') = 0. \tag{68}$$

As for the derivation of the integral equations (18) [see Eqs. (15) and (16)], we have required that $\partial\mathscr{F}[f_r; y]$ should vanish for arbitrary $y(\phi)$ and so have put the factor multiplying $y(\phi)$ equal to zero to obtain Eq. (68). The condition $\partial\mathscr{F}/\partial d_0 = 0$

[Eq. (67)] gives the value of d_0 ($n_\gamma \equiv \int_\gamma dz'\, n(z')$),

$$d_0 = -\frac{1}{n_\gamma}\left\{\frac{\lambda_3}{2\lambda_1} + \frac{1}{2\pi}\int_0^{2\pi} d\phi' \int_\gamma dz'\, n(z') \times \mathcal{N}(z_0;z',e^{i\phi'}) f_r(\phi')\right\}. \tag{69}$$

Substituting this in Eq. (68) yields the integral equation

$$\sigma^{1/2}(\phi) f_r(\phi) = \lambda_3 G^z(\phi) + \frac{\lambda_1}{2\pi}\int_0^{2\pi} K(\phi,\phi')\sigma^{1/2}(\phi') f_r(\phi') d\phi', \tag{70}$$

where

$$G^z(\phi) = \frac{-1}{2\sigma^{1/2}(\phi)}\left\{\mathcal{N}(z_0;z,e^{i\phi}) - \frac{1}{n_\gamma}\int_\gamma dz'\, n(z')\mathcal{N}(z_0;z',e^{i\phi})\right\}, \tag{71}$$

$$K(\phi,\phi') = \frac{-1}{\sigma^{1/2}(\phi)\sigma^{1/2}(\phi')}\int_\gamma dz'\, n(z')\mathcal{N}(z_0;z',e^{i\phi}) \times\left\{\mathcal{N}(z_0;z',e^{i\phi'}) - \frac{1}{n_\gamma}\int_\gamma dz''\, n(z'')\mathcal{N}(z_0;z'',e^{i\phi'})\right\}. \tag{72}$$

We see at once that the kernel $K(\phi,\phi')$ is identical with that of Eq. (20). This means that the program which was used to obtain the numerical solution to the main problem, as set out in Sec. II, may be applied directly to the integral equation (70). The desired optimal error bound $E^{0re}(z)$ for $\mathrm{Re}\, A^T(z) - \mathrm{Re}\, A^0(z)$, [see Eq. (64')], given in numerical form by means of the solution of the integral equation (70), may thus be obtained with little extra effort.

[1] S. Ciulli and T. D. Spearman, Phys. Rev. D **27**, 1580 (1983).

[2] S. Ciulli and T. D. Spearman, "Non-Perturbative Results by Function-Analytic Methods," preprint IHES/P/83/5; also, J. Phys. Lett. to be published (Sept. 1984). See also S. Ciulli and T. D. Spearman, "An Integral Equation for the Continuation of Perturbative Expansions," to be published in Nuovo Cimento.

[3] See for example Barry Simon, "Large orders and Summability of Eigenvalue Problems," Int. J. Quantum Chem. Proc. Sanibel Workshop (1981); or G. Auberson and G. Mennessier, J. Math. Phys. **22**, 2472 (1981).

[4] N. N. Khuri, Phys. Rev. D **23**, 2285 (1981).

[5] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B **147**, 385, 448, 519 (1979); see also the review of H. R. Rubinstein, Proc. Paris Conf. 1982, J. Phys. (Paris) Colloq. C-3, Suppl. **43** (12), C3-249 (1982), and the references given there.

[6] J. Hamilton, P. Menotti, T. D. Spearman, and W. S. Woolcock, Nuovo Cimento **20**, 519 (1961).

[7] A. D. Jackson, A. Lande, and R. A. Smith, Phys. Rep. **86**, 55 (1982).

[8] The precise form of this mapping is given in Ref. 1 (following Fig. 2).

[9] S. Ciulli and T. D. Spearman, J. Math. Phys. **23**, 1752 (1982).

[10] Equation (18) can be solved numerically using standard procedures for Fredholm equations of the second kind, since the kernel is Hilbert–Schmidt and thus well behaved.

[11] Notice that $\partial\, \mathrm{Re}\, S(re^{i\phi})/\partial r|_{r=1} = \partial\, \mathrm{Im}\, S(e^{i\phi})/\partial\phi$. If $\sigma(\phi) = \text{const}$ for $0 \leqslant \phi \leqslant 2\pi$, then $\partial S/\partial\phi = 0$, but any variation in $\sigma(\phi)$ will introduce a contribution to $\partial\, \mathrm{Im}\, S(e^{i\phi})/\partial\phi$. $\mathrm{Im}\, S(e^{i\phi})$ is given in terms of $\sigma(\phi)$ by a principal value integral.

[12] For example if δ_0 is a clear minimum, so that A^0 represents the nearest point to the origin of the convex region of the function space within which A^T also lies, then, if we had reason to know that δ_T was close to δ_0 (it could not be less than δ_0!), A^T would have to be very close to A^0, much closer than the above bound would indicate. [If δ_T were equal to δ_0 then A^T and A^0 would have to coincide because of the proper convexity of the set of functions satisfying the condition (13); this (proper convexity property) might not be true if other norms were used instead of the L^2 one.]

Commutants of a family of operators on a partial inner product space

J. Shabani[a)]
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

(Received 3 January 1984; accepted for publication 30 March 1984)

We consider all different possible definitions of commutants and bicommutants for an x-invariant family of operators on a partial inner product space. We investigate their behavior with respect to the weak topology and we describe the situation when all commutants (resp. all bicommutants) coincide.

PACS numbers: 02.30.Tb

I. INTRODUCTION

Unbounded commutants of unbounded operator families have been studied in recent years by many authors.[1-7]

In this paper we start with a x-invariant family $\mathscr{R}$ of the space Op V of all operators on a partial inner product (PIP) space V,[8,9] and we define the commutant of $\mathscr{R}$ within Op V. The study of this commutant shows that it is too pathological and this suggests another more suitable concept of commutant, namely one starts with $\mathscr{R}$ and defines the commutant within the *-algebra Reg V of all regular operators on the PIP space V[10] and the bicommutant within Op V.

We assume that V is quasicomplete in its canonical Mackey topology $\tau(V, V^\#)$. This implies that Reg V is isomorphic to the *-algebra $L^+(V^\#)$ of all operators $A \in \text{Op } V$ such that A and its adjoint leave $V^\#$ invariant. The space Op V is equipped with the weak topology defined by the following system of seminorms: $A \to |<A\phi, \psi>|$; $\phi, \psi \in V^\#$. On Reg $V \simeq L^+(V^\#)$ we will consider the weak topology inherited from Op V.

The paper is organized as follows. In Sec. II we recall briefly some basic facts about PIP spaces and operators on them. In Sec. III we introduce the different possible definitions of commutant for an x-invariant family $\mathscr{R}$ of Op V and we study the relationships between our commutants and bicommutants with the ones considered in Refs. 2, 4, 5, and 6. In Sec. IV we study the weak closedness of the commutants and bicommutants. In Sec. V we give some criteria in order that the bicommutant coincide with the weak closure of $\mathscr{R}$. In Sec. VI we compare our commutants with the one introduced in Ref. 1 and we describe the situation when all commutants (resp. all bicommutants) coincide.

II. PIP-SPACES AND OPERATORS ON THEM[8,9]

A PIP-space V is a complex vector space with the following structure.

(i) $\mathscr{T} = \{V_r, r \in I\}$ is a collection of vector subspaces of V which covers V and is an involutive lattice with respect to set interesection, vector sum, and lattice involution: $V_r \leftrightarrow V_{\bar{r}}$. Besides elements of $\mathscr{T}$, we consider also the extreme spaces

$$V^\# = \bigcap_{r \in I} V_r \quad \text{and} \quad V = \bigcup_{r \in I} V_r.$$

(ii) A nondegenerate Hermitian form $\langle \cdot | \cdot \rangle$ (the partial inner product) is defined on $\cup_r V_r \times V_{\bar{r}}$.

(iii) There exists a unique element $O = \bar{O}$ in I such that $V_o = V_{\bar{o}} \equiv \mathfrak{H}$ is a Hilbert space with respect to $\langle \cdot | \cdot \rangle$.

The nondegeneracy assumption $(V^\#)^\perp = \{O\}$ implies that every pair $\langle V_r, V_{\bar{r}} \rangle$, as well as $\langle V^\#, V \rangle$ is a dual pair with respect to the form $\langle \cdot | \cdot \rangle$. We may therefore equip each V_r with its Mackey topology $\tau(V_r, V_{\bar{r}})$ and similarly for $V^\#, V$.

An operator A on a PIP space V is a map $D(A) \to V$, where $D(A)$ is the largest union of V_r's such that the restriction of A to any of them is linear and continuous into V. The set of all operators on V, denoted by Op V is isomorphic to $\mathscr{L}(V^\#, V) = \{\text{linear continuous maps } V^\# \to V\}$. Equivalently Op $V \simeq B(V^\#, V^\#) = \{\text{separately continuous sesquilinear forms on } V^\# \times V^\#\}$. Thus, Op V is a vector space. Moreover Op V carries an involution $A \leftrightarrow A^x$ (adjoint of A), but it is not an algebra for the multiplication is not always defined. Such sets are called partial-x-algebras.[11] In order to avoid this undesirable feature, one has to consider a smaller class of operators; the so-called regular operators.[10]

An operator A on a PIP space V is called regular if $D(A) = D(A^x) = V$. Equivalently, a regular operator is a linear continuous map of V into itself, which maps $V^\#$ into itself continuously. The set of all regular operators on V, denoted by Reg V is a *-algebra.

The space Op V contains another remarkable subset, namely

$$C(V^\#, \mathfrak{H}) = \{A\text{-closable in } \mathfrak{H} \mid V^\# \subset D(A) \cap D(A^*)\}$$
$$= \{A \in \text{Op } V \mid A^{(*)}: V^\# \to \mathfrak{H}\}.$$

We have Reg $V \subseteq C(V^\#, \mathfrak{H}) \subseteq$ Op V.

From now on, we will assume that V is quasicomplete in its Mackey topology. This implies in particular that we can simply identify Reg V with the *-algebra $L^+(V^\#)$ of operators A such that $V^\# \subset D(A) \cap D(A^*)$ and $A^{(*)} V^\# \subset V^\#$ (see Ref. 10, Proposition 2.5). An Op*-algebra on V is a *-subalgebra with unit of $L^+(V^\#)$. The condition of Mackey quasicompleteness of V is actually satisfied in almost all examples; the only known exceptions are quite pathological.[12]

We will endow Op V with the weak topology defined by the following family of seminorms:

$$A \mapsto |<A\varphi, \psi>|; \quad \varphi, \psi \in V^\#.$$

On Reg $V \simeq L^+(V^\#)$ we will consider the weak topology inherited from Op V.

[a)] Boursier Administration générale de la Coopération au développement, Belgium.

 0022-2488/84/113204-05$02.50

III. COMMUTANTS AND BICOMMUTANTS

Let $A, B \in \text{Op } V$. Following Ref. 11 the operator A is called the left multiplier of B (resp. B is the right multiplier of A) and we note $A \in L(B)$ [resp. $B \in R(A)$] if the product AB is defined. Similarly if $\mathscr{R}$ is a subset of Op V, we may define

$$L\mathscr{R} \equiv \bigcap_{B \in \mathscr{R}} L(B) = \{C \in \text{Op } V \mid CB \text{ exists } \forall B \in \mathscr{R}\}.$$

Definition 3.1: A subset $\mathscr{R}$ of Op V is said to be a partial x-algebra if the following conditions are satisifed.

(i) $\mathscr{R}$ contains the identity.

(ii) $\mathscr{R}$ is x-invariant, i.e., $A \in \mathscr{R}$ implies $A^x \in \mathscr{R}$.

(iii) If $A, B \in \mathscr{R}$ and $A \in L(B)$, then $AB \in \mathscr{R}$.

It is easily verified that Op V satisfies these conditions.

Let $\mathscr{R}$ be an x-invariant subset of Op V. It is natural to define the commutant of $\mathscr{R}$ in the following way:

$$\mathscr{R}' = \{X \in \text{Op } V \mid X \in L\mathscr{R} \cap R\mathscr{R} \text{ and } AX = XA, \quad \forall A \in \mathscr{R}\}.$$

This set is a vector subspace of $L\mathscr{R} \cap R\mathscr{R}$. Furthermore it is x-invariant and contains the identity, but it is neither a partial-x-algebra nor a weakly closed subspace of Op V. Now we may define the following bicommutant:

$$\mathscr{R}'' = (\mathscr{R}')' = \{Y \in \text{Op } V \mid Y \in L\mathscr{R}' \cap R\mathscr{R}' \text{ and } YX = XY, \quad \forall X \in \mathscr{R}'\}.$$

This set is an x-invariant subspace of $L\mathscr{R}' \cap R\mathscr{R}'$. Moreover it contains the identity and $\mathscr{R}$ itself. From this fact and the relation $\mathscr{R}_1 \subset \mathscr{R}_2$ implies $\mathscr{R}_2' \subset \mathscr{R}_1'$ we get that $\mathscr{R}''' = \mathscr{R}'$ and therefore $\mathscr{R}'''' = \mathscr{R}''$. But in general $\mathscr{R}''$ is not a partial-x-algebra and furthermore it is not closed with respect to the weak topology.

The lesson of these considerations is that the concept of commutant introduced above is too pathological and thus not interesting for applications. One can improve the situation by considering the case when the product of two operators is always defined. This is true in particular whenever one of the two operators is regular.[10] Thus another (well-behaved) concept of commutant suggests itself: let $\mathscr{R}$ be an x-invariant subset of Op V. We may define the following commutant:

$$\mathscr{R}_c' = \{X \in \text{Reg } V \simeq L^+(V^\#) \mid AX = XA, \quad \forall A \in \mathscr{R}\}.$$

As Reg $V \subset L\mathscr{R} \cap R\mathscr{R}$, the commutant $\mathscr{R}_c'$ is contained in $\mathscr{R}'$. The set $\mathscr{R}_c'$ is an Op*-algebra on $V^\#$. To prove this, it suffices to show that if $X, Y \in \mathscr{R}_c'$ then their product XY also belongs to $\mathscr{R}_c'$.

We have $XA = AX$ and $YA = AY$, $\forall A \in \mathscr{R}$. From these relations, it follows that $\forall f, g \in V^\#$ we have

$$AXYf = AXg = XAg = XAYf = XYAf, \quad \text{i.e., } XY \in \mathscr{R}_c'.$$

One can also start with a subset of regular operators and define the commutant within Op V. Namely: let $\mathscr{R}$ be an Op*-algebra on $V^\#$. We may define the following commutant:

$$\mathscr{R}_0' = \{X \in \text{Op } V \mid AX = XA, \quad \forall A \in \mathscr{R}\}$$

This set is an x-invariant subset of Op V. Clearly $\mathscr{R}_c' = \mathscr{R}_0' \cap L^+(V^\#)$ coincides with Inoue's commutant[4] and we have the following relation between the different types of commutants:

$$\mathscr{R}_c' \subseteq \mathscr{R}_\sigma' \subseteq \mathscr{R}_0',$$

where $\mathscr{R}_\sigma'$ is the "weak unbounded commutant" introduced in Ref. 2.

Now we may define two different bicommutants.

(i) $\mathscr{R}_{c0}'' = \{Y \in \text{Op } V \mid YX = XY, \quad \forall X \in \mathscr{R}_c'\}$,

which is an x-invariant subset of Op V containing $\mathscr{R}$. From the above definition, it follows that if $\mathscr{R}$ is a x-invariant subset of Op V, then $\mathscr{R}_{c0c}''' = \mathscr{R}_c'$ and $\mathscr{R}_{c0c0}''' = \mathscr{R}_{c0}''$.
This bicommutant is related to $\mathscr{R}''$ in the following way:

$$\mathscr{R}'' \subset \mathscr{R}_{c0}'' = (\mathscr{R}_c')' \subset \text{Op } V.$$

(ii) $\mathscr{R}_{0c}'' = \{Y \in L^+(V^\#) \mid YX = XY, \quad \forall X \in \mathscr{R}_0'\}$.

This set is an Op*-algebra.

The two bicommutants $\mathscr{R}_{0c}''$ and $\mathscr{R}_{c0}''$ are related to the four ones considered in Ref. 5 in the following way:

$$\begin{array}{ccccc} \mathscr{R} & & & & \\ & & & \subseteq & \mathscr{R}_{\sigma\sigma}'' \\ & & & & \cup| \\ \mathscr{R} & \subseteq & \mathscr{R}_{0c}'' & \subseteq & \mathscr{R}_{\sigma c}'' \end{array} \qquad \begin{array}{ccccc} \subseteq & \mathscr{R}_{c0}'' & \subseteq & \text{Op } V \\ & \cup| & & \cup| \\ \subseteq & \mathscr{R}_{c\sigma}'' & \subseteq & C(V^\#, \mathfrak{H}) \\ & \cup| & & \cup| \\ \subseteq & \mathscr{R}_{cc}'' & \subseteq & L^+(V^\#), \end{array}$$

IV. CLOSEDNESS OF THE COMMUTANTS AND BICOMMUTANTS

One of the main properties of the algebra $B(\mathfrak{H})$ of bounded operators in $\mathfrak{H}$ is that the commutant and bicommutant (in the usual sense) are weakly closed. If we try to extend this property to the commutants and bicommutants introduced in this paper, it turns out that this property remains valid whenever one starts with an Op*-algebra on $V^\#$, i.e., $\mathscr{R}_0'$ and $\mathscr{R}_{c0}''$ are weakly closed in Op V. But if one starts with an x-invariant subset of Op V, then the commutants [here $\mathscr{R}_c'$ and $(\mathscr{R}_0')_c'$] need not be closed in $L^+(V^\#)$ with respect to the weak topology.

Proposition 4.1: Let $\mathscr{R}$ be an x-invariant subset of Op V. If $\mathscr{R}V^\# \subset V^\#$, then $\mathscr{R}_c'$ is weakly closed in $L^+(V^\#)$.

Proof: Let $\{A_\alpha\}$ be a net in $\mathscr{R}_c'$ such that $A_\alpha \xrightarrow{w} A$ with $A \in L^+(V^\#)$, i.e., $(A_\alpha\varphi, \psi) \to (A\varphi, \psi)$, $\forall \varphi, \psi \in V^\#$. We have

to prove that $A \in \mathscr{R}'_c$. Let $B \in \mathscr{R}$. This means that $BA_\alpha = A_\alpha B$ and we get the following relation:

$$(A\varphi, B^*\psi) = \lim_\alpha (A_\alpha \varphi, B^*\psi) = \lim_\alpha (B\varphi, A^*_\alpha \psi) = (B\varphi, A^*\psi).$$

This implies that $A \in \mathscr{R}'_\sigma$, and since $A \in L^+(V^\#)$, we have that $A \in \mathscr{R}'_c$.

Remark 4.2: The condition $\mathscr{R}V^\# \subset V^\#$ implies that $\mathscr{R}$ is an Op*-algebra on $V^\#$ and therefore $\mathscr{R}'_c$ is Inoue's commutant, which is indeed weakly closed in $L^+(V^\#)$ (see Ref. 4).

Proposition 4.3: If $\mathscr{R}$ is an Op*-algebra on $V^\#$, then we have the following.

(i) $\mathscr{R}'_0$ is weakly closed in Op V.

(ii) If moreover $\mathscr{R}'_0 V^\# \subset V^\#$, then $\mathscr{R}''_{0c}$ is weakly closed in $L^+(V^\#)$.

The proof is similar to that of Proposition 4.1.

Remark 4.4: The condition $\mathscr{R}'_0 V^\# \subset V^\#$ means that $\mathscr{R}'_0$ is an Op*-algebra on $V^\#$, i.e., $\mathscr{R}'_0 = \mathscr{R}'_c$. Thus $\mathscr{R}''_{0c}$ is not weakly closed unless it is equal to $\mathscr{R}''_{cc}$.

Corollary 4.5: If $\mathscr{R}$ is an x-invariant subset of Op V, then $\mathscr{R}''_{c0}$ is weakly closed in Op V.

Proposition 4.6: If $\mathscr{R}$ is an x-invariant subset of Op V, then the commutant of $\mathscr{R}$ is equal to the commutant of its weak closure, i.e., $\mathscr{R}'_c = (\overline{\mathscr{R}}^w)'_c$.

Proof: The inclusion $(\overline{\mathscr{R}}^w)'_c \subset \mathscr{R}'_c$ follows clearly from the fact that $\mathscr{R} \subset \overline{\mathscr{R}}^w$. Let us now prove that $\mathscr{R}'_c \subset (\overline{\mathscr{R}}^w)'_c$. Let $B \in \overline{\mathscr{R}}^w$. Then there exists a net $\{B_\alpha\} \subset \mathscr{R}$ such that $B_\alpha \overset{w}{\to} B$. Let $X \in \mathscr{R}'_c$, i.e., $XB_\alpha = B_\alpha X$. We have

$$(B\varphi, X^*\psi) = \lim_\alpha (B_\alpha \varphi, X^*\psi)$$

$$= \lim_\alpha (X\varphi, B^*_\alpha \psi) = (X\varphi, B^*\psi),$$

i.e., $X \in (\overline{\mathscr{R}}^w)'_c$.

Corollary 4.7: Let $\mathscr{R}$ and $\mathscr{R}_1$ be two x-invariant subsets of Op V such that $\mathscr{R}_1 \subset \mathscr{R}$ and $\mathscr{R}_1$ is weakly dense in $\mathscr{R}$. then $(\mathscr{R}_1)'_c = \mathscr{R}'_c$.

Proof: The inclusion $\mathscr{R}'_c \subset (\mathscr{R}_1)'_c$ is obvious. Now since $\mathscr{R}_1$ is weakly dense in $\mathscr{R}$, we have $\overline{\mathscr{R}}^w_1 \supset \mathscr{R}$ and therefore $(\overline{\mathscr{R}}^w_1)'_c \subset \mathscr{R}'_c$. From Proposition 4.6 we know that $(\overline{\mathscr{R}}^w_1)'_c = (\mathscr{R}_1)'_c$. This implies that $(\mathscr{R}_1)'_c \subset \mathscr{R}'_c$ and thus we obtain the equality $(\mathscr{R}_1)'_c = \mathscr{R}'_c$.

Remark 4.8: In Proposition 4.6 one can start with an Op*-algebra and consider the commutant $\mathscr{R}'_0$. In this case the equality $\mathscr{R}'_0 = (\overline{\mathscr{R}}^w)'_0$ does not hold since $\mathscr{R}'_0$ does not leave $V^\#$ invariant. We have only the inclusion $\mathscr{R}'_0 \supset (\overline{\mathscr{R}}^w)'_0$.

V. BICOMMUTANTS AND THE WEAK CLOSURE OF $\mathscr{R}$

For algebras of bounded operators, it is well known that the usual bicommutant of a nondegenerate *-algebra $\mathscr{M}$ of $B(\mathfrak{H})$ coincides with the weak closure of $\mathscr{M}$, i.e., $\mathscr{M}'' = \overline{\mathscr{M}}^w$. When one starts with an x-invariant subset $\mathscr{R}$ of Op V (or an Op*-algebra on $V^\#$), this property need not be satisfied and our aim in this section is to find sufficient conditions on $\mathscr{R}$ in order that the above property remains true.

Let $\mathscr{R}$ be an x-invariant subset of Op V and consider $\mathscr{R}_b \equiv \mathscr{R} \cap B(\mathfrak{H})$. Following Ref. 8 we will say that a subspace W of a PIP space V is orthocomplemented in V, if W is the range of an orthogonal projection P, i.e., $W = PV$. We will say that $\mathscr{R}_b$ satisfies the condition (α) if $\forall\, \varphi \in V^\#$ the norm closure of $\mathscr{R}_b\varphi$ is orthocomplemented in V.

Proposition 5.1: Let $\mathscr{R}$ be an x-invariant subset of Op V. If $\mathscr{R}_b$ satisfies (α), then $(\mathscr{R}_b)''_{c0} = \overline{\mathscr{R}}^w_b$.

Proof: First, we know from Corollary 4.5 that $(\mathscr{R}_b)''_{c0}$ is weakly closed and hence $\overline{\mathscr{R}}^w_b \subset (\mathscr{R}_b)''_{c0}$. Now let $\varphi \in V^\# \subset \mathfrak{H}$ and consider the closed subspace $\overline{\mathscr{R}_b\varphi}^{\|\cdot\|}$ of $\mathfrak{H}$. Let P be the projection on $\overline{\mathscr{R}_b\varphi}^{\|\cdot\|}$. Since all elements $A \in \mathscr{R}_b$ are bounded and $\mathscr{R}_b$ is an algebra, $\mathscr{R}_b$ leaves $\overline{\mathscr{R}_b\varphi}^{\|\cdot\|}$ invariant and this implies that $AP = PA$, i.e., $P \in (\mathscr{R}_b)'_w \equiv (\mathscr{R}_b)'_\sigma \cap B(\mathfrak{H})$. On the other hand P leaves the space $\overline{\mathscr{R}_b\varphi}^{\|\cdot\|}$ invariant and therefore $P \in (\mathscr{R}_b)'_s \equiv (\mathscr{R}_b)'_c \cap B(\mathfrak{H})$, i.e., $P \in \mathscr{R}'_c$. Let now $B \in (\mathscr{R}_b)''_{c0}$ and $\psi \in V^\#$. Then

$$((1-P)B\varphi, \psi) = (B\varphi, (1-P)\psi) = ((1-P)\varphi, B^*\psi) = 0,$$

i.e., $B\varphi = PB\varphi$, which in turn implies that $B\varphi \in \overline{\mathscr{R}_b\varphi}^{\|\cdot\|}$ and hence $B \in (\overline{\mathscr{R}_b})^s \subset (\overline{\mathscr{R}_b})^w$.

Proposition 5.2: Let $\mathscr{R}$ be an x-invariant subset of Op V. If $\mathscr{R}_b$ satisfies (α) and $(\mathscr{R}_b)'_c = \mathscr{R}'_c$, then $\mathscr{R}''_{c0} = \overline{\mathscr{R}}^w$.

Proof: The inclusion $\mathscr{R}''_{c0} \supset \overline{\mathscr{R}}^w$ follows from the fact that $\mathscr{R}''_{c0}$ is weakly closed. Now, the equality $\mathscr{R}'_c = (\mathscr{R}_b)'_c$ together with Proposition 5.1. imply that $\mathscr{R}''_{c0} = (\mathscr{R}_b)''_{c0} = \overline{\mathscr{R}}^w_b$. Taking into account the relation $\overline{\mathscr{R}}^w_b \subset \overline{\mathscr{R}}^w$, we obtain the inclusion $\mathscr{R}''_{c0} \subset \overline{\mathscr{R}}^w$.

Corollary 5.3: Let $\mathscr{R}$ be an x-invariant subset of Op V. If $\mathscr{R}_b$ satisfies (α) and $\mathscr{R}_b$ is weakly dense in $\mathscr{R}$, then $\mathscr{R}''_{c0} = \overline{\mathscr{R}}^w$.

Proof: Since $\mathscr{R}_b$ is weakly dense in $\mathscr{R}$, it follows from Corollary 4.7. that $(\mathscr{R}_b)'_c = \mathscr{R}'_c$. Now using Proposition 5.2 we obtain the needed equality.

Proposition 5.4: Let $\mathscr{R}$ be an x-invariant subset with unit of Op V and assume that $\forall\, \varphi \in V^\#$, the $\sigma(V, V^\#)$-closure of $\mathscr{R}\varphi$ is orthocomplemented in V. If $\forall\, \varphi \in V^\#$, $P_\varphi \in \mathscr{R}'_c$ [where $P_\varphi \in L^+(V^\#)$] is the projection on $\overline{\mathscr{R}\varphi}^{\sigma(V, V^\#)}$ then $\mathscr{R}''_{c0} = \overline{\mathscr{R}}^w$.

Proof: The inclusion $\overline{\mathscr{R}}^w \subset \mathscr{R}''_{c0}$ is obvious since $\mathscr{R}''_{c0}$ is weakly closed. Now let $\varphi, \psi \in V^\#$ and $B \in \mathscr{R}''_{c0}$. Since $P_\varphi \in \mathscr{R}'_c \subset L^+(V^\#) \simeq \mathrm{Reg}\, V$, P_φ is defined on the whole space V and $PB = BP$. On the other hand, we have the following relation:

$$((1-P_\varphi)B\varphi, \psi) = (B\varphi, (1-P_\varphi)\psi) = ((1-P_\varphi)\varphi, B^*\psi) = 0,$$

i.e., $B\varphi = P_\varphi B\varphi$ which implies that $B\varphi \in \overline{\mathscr{R}\varphi}^{\sigma(V, V^\#)}$ and therefore $B \in \overline{\mathscr{R}}^w$.

Let now $\mathscr{R}$ be an Op*-algebra on $V^\#$ and P the projection on the space $\overline{\mathscr{R}\varphi}^{\sigma(V^\#, V^\#)}$: $\varphi \in V^\#$. Then on one hand every element $A \in \mathscr{R}$ is $\sigma(V^\#, V^\#)$ continuous and hence the space $\overline{\mathscr{R}\varphi}^{\sigma(V^\#, V^\#)}$ is invariant under $\mathscr{R}$, which means that $PA = AP$, i.e., $P \in \mathscr{R}'_0$. But on the other hand, by definition P belongs to $L^+(V^\#)$ so that finally $P \in \mathscr{R}'_0 \cap L^+(V^\#) \equiv \mathscr{R}'_c$. So, if in Proposition 5.4 one starts with an Op*-algebra on $V^\#$, the assumption $P \in \mathscr{R}'_c$ is automatically fulfilled. But in this case we have to consider the bicommutant $\mathscr{R}''_{0c}$ which is not weakly closed in $L^+(V^\#)$ unless it coincides with $\mathscr{R}''_{cc}$. We summarize these considerations in the following.

Proposition 5.5: Let $\mathscr{R}$ be an Op*-algebra on $V^{\#}$. If $\forall\, \varphi \in V^{\#}$ the $\sigma(V^{\#}, V^{\#})$ closure of $\mathscr{R}\varphi$ is orthocomplemented in $V^{\#}$, then $\mathscr{R}''_{cc} = \overline{\mathscr{R}}^{w}$.

VI. COINCIDENCE OF DIFFERENT TYPES OF COMMUTANTS AND BICOMMUTANTS

If we start with an Op*-algebra $\mathscr{R}$ on $V^{\#}$, we may define three different commutants, namely $\mathscr{R}'_c$, $\mathscr{R}'_\sigma$, and $\mathscr{R}'_0$, and six different bicommutants: $\mathscr{R}''_{c0}$, $\mathscr{R}''_{c\sigma}$, $\mathscr{R}''_{cc}$, $\mathscr{R}''_{0c}$, $\mathscr{R}''_{\sigma\sigma}$, and $\mathscr{R}''_{\sigma c}$ (see Refs. 2, 5, and 6). We should like to have only one concept of unbounded commutant and unbounded bicommutant for unbounded algebras. Thus our aim in this section is to find sufficient conditions in order that all the commutants (resp. all bicommutants) coincide.

Let $\mathscr{R}$ be an Op*-algebra on $V^{\#}$. We can define $\mathscr{R}'_c$, $\mathscr{R}'_\sigma$, and $\mathscr{R}'_0$. In general these commutants are related in the following way: $\mathscr{R}'_c \subseteq \mathscr{R}'_\sigma \subseteq \mathscr{R}'_0$. At this stage one question arises immediately: when do they coincide? Obviously a necessary condition for the coincidence of these commutants is that they all belong to $L^{+}(V^{\#})$.

To answer the above question we will compare $\mathscr{R}'_0$ with a commutant introduced in Ref. 1 for closed Op*-algebras.[13] This commutant, to be denoted by $\mathscr{R}'_A$ will be described below. We will say that a closed Op*-algebra $\mathscr{R}$ satisfies the condition I_0 (see Ref. 1) if $\mathscr{R}$ contains a generating monotone increasing sequence $A_n \geqslant 1$ such that $A_n V^{\#} = V^{\#}$.

Let $\mathscr{R}$ be an Op*-algebra on $V^{\#}$. Then $\mathscr{R}$ defines on $V^{\#}$ a locally convex topology $t_{\mathscr{R}}$ (see Ref. 13) (the so-called $\mathscr{R}$-topology) by the following seminorms: $f \mapsto \|Af\|$; $f \in V^{\#}$, $A \in \mathscr{R}$. This topology is the coarsest locally convex topology on $V^{\#}$ such that every $A \in \mathscr{R}$ is continuous from $V^{\#}[t_{\mathscr{R}}]$ into $\mathfrak{H}$ endowed with the usual Hilbert space norm topology.

Definition 6.1: The Op*-algebra $\mathscr{R}$ is called closed if $V^{\#}[t_{\mathscr{R}}]$ is complete.

Definition 6.2: Let $\mathscr{R}$ be a closed Op*-algebra on $V^{\#}$ and $B(V^{\#}[t_{\mathscr{R}}], V^{\#}[t_{\mathscr{R}}])$ the set of all sesquilinear forms which are jointly continuous in the $\mathscr{R}$-topology, i.e., for $\beta \in B(V^{\#}[t_{\mathscr{R}}], V^{\#}[t_{\mathscr{R}}])$ there is an $A \in \mathscr{R}$ such that for some constant M and all $x, y \in V^{\#}$,

$$|\beta(x, y)| \leqslant M \|Ax\| \, \|Ay\|.$$

One can define the following commutant[1]:

$$\mathscr{R}'_A = \{\beta \in B(V^{\#}[t_{\mathscr{R}}], V^{\#}[t_{\mathscr{R}}]) | \beta(Af, g) = \beta(f, A^{*}g);$$
$$\forall\, f, g \in V^{\#} \text{ and } A \in \mathscr{R}\}.$$

Proposition 6.3: (See Ref. 1). If $\mathscr{R}$ is an Op*-algebra on $V^{\#}$ satisfying I_0, then (i) $\mathscr{R}'_A$ is an Op*-algebra on $V^{\#}$ satisfying I_0, but $\mathscr{R}'_A$ is not closed; and (ii) the $\mathscr{R}$ – topology is metrizable; it is defined by the seminorms

$$f \mapsto \|A_n f\|; \quad n \in \mathbb{N}, \quad f \in V^{\#}.$$

From now on we assume that the closed Op*-algebra $\mathscr{R}$ satisfies I_0. Let us compare $\mathscr{R}'_A$ with $\mathscr{R}'_0$. First, we know that Op V is isomorphic to the space $B(V^{\#}_\tau, V^{\#}_\tau)$ of all Mackey separately continuous sesquilinear forms on $V^{\#} \times V^{\#}$ (see Ref. 8). Thus $\mathscr{R}'_A$ will coincide with $\mathscr{R}'_0$ if in particular

$$B(V^{\#}[t_{\mathscr{R}}], V^{\#}[t_{\mathscr{R}}]) \equiv B(V^{\#}_\tau, V^{\#}_\tau).$$

Since the PIP space V possesses a central Hilbert space $\mathfrak{H}$, the topologies $t_{\mathscr{R}}$ and $\tau(V^{\#}, V)$ are comparable. In general the Mackey topology is strictly finer than the $\mathscr{R}$-topology and we have the following situation (where $V_{\mathscr{R}}$, the dual of $V^{\#}[t_{\mathscr{R}}]$, and V_τ, the dual of $V^{\#}[\tau]$, are endowed with their Mackey topologies, and each arrow denotes a continuous embedding with dense range):

$$V^{\#}[\tau] \to V^{\#}[t_{\mathscr{R}}] \to \mathfrak{H} \to V_{\mathscr{R}} \to V_\tau.$$

Now the $\mathscr{R}$-topology is metrizable and this implies that[14] on one hand $t_{\mathscr{R}} = \tau(V^{\#}[t_{\mathscr{R}}], V_{\mathscr{R}})$. On the other hand, $V^{\#}[t_{\mathscr{R}}]$ is a Fréchet space; it is therefore barreled. Thus, $V^{\#}[t_{\mathscr{R}}] = V^{\#}[\tau]$ and the dual of $V^{\#}[t_{\mathscr{R}}]$ is τ-quasicomplete, i.e., $V_{\mathscr{R}} = V_\tau = V^{\frown}$. Furthermore, since $V^{\#}[t_{\mathscr{R}}] = V^{\#}[\tau]$ is a Fréchet space, every separately continuous sesquilinear form on $V^{\#} \times V^{\#}$ is jointly continuous and we get that $B(V^{\#}[t_{\mathscr{R}}], V^{\#}[t_{\mathscr{R}}]) \equiv B(V^{\#}_\tau, V^{\#}_\tau)$, which in turn implies that $\mathscr{R}'_A = \mathscr{R}'_0$. Since $\mathscr{R}$ satisfies I_0, $\mathscr{R}'_0$ is an Op*-algebra on $V^{\#}$ and hence $\mathscr{R}'_0 = \mathscr{R}'_c$. Now, taking into account the relation $\mathscr{R}'_c \subseteq \mathscr{R}'_\sigma \subseteq \mathscr{R}'_0$ we obtain the equality $\mathscr{R}'_c = \mathscr{R}'_\sigma = \mathscr{R}'_0 \subset L^{+}(V^{\#})$. Summarizing this analysis we get the following.

Proposition 6.4: Let V be an arbitrary PIP space and $\mathscr{R}$ a closed Op*-algebra on $V^{\#}$ satisfying I_0. Then $\mathscr{R}'_c = \mathscr{R}'_\sigma = \mathscr{R}'_0 = \mathscr{R}'_A \subseteq L^{+}(V^{\#})$.

In practice, it is much easier to start with the closed Op*-algebra $\mathscr{R}$ on $V^{\#}$ satisfying I_0, and build the canonical PIP space V associated to it, following the construction of Ref. 15. In this way we get another PIP space structure around $\mathfrak{H}$, namely the lattice generated by all the Hilbert spaces $D(\overline{A})$ (with graph norm); $A \in \mathscr{R}$.

Proposition 6.5: Let $\mathscr{R}$ be a closed Op*-algebra on $V^{\#}$ satisfying I_0. If V is the PIP space generated by $\mathscr{R}$, then $\mathscr{R}'_c = \mathscr{R}'_\sigma = \mathscr{R}'_0 = \mathscr{R}'_A \subset L^{+}(V^{\#})$.

Now, let us look at the bicommutants. If $\mathscr{R}$ is an Op*-algebra on $V^{\#}$, then we may define six different bicommutants, namely: $\mathscr{R}''_{\sigma\sigma}$, $\mathscr{R}''_{c0}$, $\mathscr{R}''_{c\sigma}$, $\mathscr{R}''_{cc}$, $\mathscr{R}''_{0c}$, and $\mathscr{R}''_{\sigma c}$. Obviously in order that these bicommutants coincide it is necessary that they all belong to $L^{+}(V^{\#})$.

Assume that $\mathscr{R}$ is closed and satisfies the condition I_0. Then, $\mathscr{R}'_A$ is an Op*-algebra on $V^{\#}$, but it is not closed. So in general

$$\mathscr{R}''_{AA} = \{\gamma \in B(V^{\#}[t_{\mathscr{R}'_A}], V^{\#}[t_{\mathscr{R}'_A}]) | \gamma(Cf, g)$$
$$= \gamma(f, C^{*}g); \quad \forall\, f, g \in V^{\#}, C \in \mathscr{R}'_A\}$$

is not an Op*-algebra and in this case no general connection can be found between $\mathscr{R}''_{AA}$ and, e.g., $\mathscr{R}''_{cc}$. But $\mathscr{R}'_A$ is an Op*-algebra and we know that it can always be extended by continuity to the Op*-algebra (its closure) $\overline{\mathscr{R}'_A}$ on $\overline{V^{\#}[t_{\mathscr{R}'_A}]} \equiv V^{\#}[t_{\mathscr{R}'_A}]$. We can then consider the bicommutant

$$\left(\overline{\mathscr{R}'_A}, V^{\#}\left[t_{\overline{\mathscr{R}'_A}}\right]\right)'_A \equiv \left(\overline{\mathscr{R}'_A}\right)'_A$$
$$= \left\{\gamma \in B\left(V^{\#}\left[t_{\overline{\mathscr{R}'_A}}\right], V^{\#}\left[t_{\overline{\mathscr{R}'_A}}\right]\right) | \gamma(Cf, g)\right.$$
$$\left. = \gamma(f, C^{*}g); \quad \forall\, f, g \in V^{\#}, C \in \overline{\mathscr{R}'_A}\right\},$$

which is an Op*-algebra on $V^{\#}\left[t_{\overline{\mathscr{R}'_A}}\right]$. In general this com-

mutant is contained in $\mathscr{R}''_{AA}$. Since $\mathscr{R}'_A$ satisfies I_0, $\overline{\mathscr{R}'_A}$ also satisfies I_0 (see Ref. 1). Thus Proposition 5.4 gives us the following result for $\overline{\mathscr{R}'_A}$: $(\overline{\mathscr{R}'_A})'_c = (\overline{\mathscr{R}'_A})'_\sigma = (\overline{\mathscr{R}'_A})'_0 = (\overline{\mathscr{R}'_A})'_A \subset L^+(V^\#)$. On the other hand the Op*-algebra $\mathscr{R}$ itself satisfies the conditions of Proposition 6.4 so $\mathscr{R}'_c = \mathscr{R}'_\sigma = \mathscr{R}'_0 = \mathscr{R}'_A$. These two relations together give the equality between the six different bicommutants for $\overline{\mathscr{R}'_A}$.

Proposition 6.6: Let V be an arbitrary PIP space and $\mathscr{R}$ a closed Op*-algebra on $V^\#$ satisfying I_0. Then

$$(\overline{\mathscr{R}'_0})'_c = (\overline{\mathscr{R}'_\sigma})'_c = (\overline{\mathscr{R}'_c})'_c = (\overline{\mathscr{R}'_\sigma})'_\sigma = (\overline{\mathscr{R}'_c})'_\sigma = (\overline{\mathscr{R}'_c})'_0 = (\overline{\mathscr{R}'_A})'_A \subset L^+(V^\#). \quad (*)$$

Proposition 6.7: Let $\mathscr{R}$ be a closed Op*-algebra on $V^\#$ satisfying I_0. If V is the PIP space generated by $\overline{\mathscr{R}'_A}$, then we obtain (*).

VII. THE BOUNDED PART OF THE COMMUTANT $\mathscr{R}'_0$

Definition 7.1: An Op*-algebra $\mathscr{R}$ is called symmetric if for every $A \in \mathscr{R}$, $(1 + A^*A)^{-1}$ exists and lies in $\mathscr{R}_b \equiv \mathscr{R} \cap B(\mathfrak{H})$.

Proposition 7.2: If V is a PIP-space and $\mathscr{R}$ a symmetric Op*-algebra on $V^\#$, then $\mathscr{R}'_0 = (\mathscr{R}_b)'_0$.

Proof: The inclusion $(\mathscr{R}_b)'_0 \supset \mathscr{R}'_0$ follows from the fact that $\mathscr{R}_b$ is contained in $\mathscr{R}$. Now assume that $C \in (\mathscr{R}_b)'_0$ and let $A = A^+ \equiv A^*|_{V^\#}$. Since $\mathscr{R}$ is symmetric we know that $(1 + A^2)^{-1}$ and $A(1 + A^2)^{-1}$ belong to $\mathscr{R}_b$. For all $f, g \in V^\#$ we have

$$\langle CA(1 + A^2)^{-1} f | g \rangle = \langle A(1 + A^2)^{-1} Cf | g \rangle = \langle AC(1 + A^2)^{-1} f | g \rangle.$$

Since $(1 + A^2)^{-1} V^\# = V^\#$, we have $C \in \mathscr{R}'_0$ and this implies the equality $(\mathscr{R}_b)'_0 = \mathscr{R}'_0$.

Remark 7.3: A similar proof in Ref. 5 shows that $\mathscr{R}'_\sigma = (\mathscr{R}_b)'_\sigma$ so that finally for symmetric Op*-algebras each commutant is equal to the analogous commutant of the bounded part.

Proposition 7.4: $(\mathscr{R}'_0)_b$ is a von Neumann algebra.

Proof: Since $\mathscr{R}'_0$ is weakly closed in Op V, $(\mathscr{R}'_0)_b$ is closed in $B(\mathfrak{H}) = \{(0, 0)\text{-representatives of } \mathscr{R}'_0\}$ (see Ref. 8) with respect to the weak topology inherited from Op V and therefore $(\mathscr{R}'_0)_b$ is closed in $B(\mathfrak{H})$ with respect to the usual weak topology of $B(\mathfrak{H})$.

Assume now that the Op*-algebra $\mathscr{R}$ is closed and satisfies the condition I_0 (see Ref. 1). Let $\tilde{\mathscr{R}}$ be the Op*-algebra generated by $\mathscr{R}$ and all A_n^{-1}, and N the von Neumann algebra generated by all bounded operators in $\tilde{\mathscr{R}}$. Then we have the following.

Proposition 7.5: Let V be a PIP-space and $\mathscr{R}$ a closed Op*-algebra on $V^\#$ satisfying I_0. Then we have the following.

(i) $(\mathscr{R}'_0)_b = N'$ [commutant in the usual sense in $B(\mathfrak{H})$].

(ii) If moreover $\mathscr{R}'_0$ is symmetric then it is an EW^*-algebra in the sense of Dixon[16] (i.e., a symmetric Op*-algebra whose bounded part is a von Neumann algebra).

Proof: (i) In Ref. 1 it is proved that if $\mathscr{R}$ is closed and satisfies I_0, then $(\mathscr{R}'_A)_b = N'$. On the other hand $\mathscr{R}'_A = \mathscr{R}'_0$ by Proposition 6.4, so that finally $(\mathscr{R}'_0)_b = N'$.

ACKNOWLEDGMENTS

I want to thank Professor J. P. Antoine and Dr. F. Mathot, Dr. C. Trapani, and Dr. W. Timmermann for many fruitful discussions.

[1] H. Araki and J-P. Jurzak, Publ. RIMS, Kyoto Univ. **18**, 1013 (1982).
[2] G. Epifanio and C. Trapani, J. Math. Phys. **25**, 2633(1984).
[3] S. Gudder and W. Scruggs, Pac. J. Math. **70**, 369 (1977).
[4] A. Inoue, Proc. Am. Math. Soc. **69**, 97 (1978).
[5] F. Mathot, "Topological properties of unbounded bicommutants," Université Catholique de Louvain, Preprint UCL-IPT-82-22.
[6] F. Mathot and J. Shabani, "Generalized unbounded commutants," Université Catholique de Louvain, to appear in *Proceedings of the II International Conference* on *Operator Algebras, Ideals and their Applications in Theoretical Physics 1983* (Teubner, Leipzig).
[7] R. T. Powers, Commun. Math. Phys. **21**, 85 (1971).
[8] J. P. Antoine and A. Grossmann, J. Funct. Anal. **23**, 369, 379 (1976); J. Math. Phys. **19**, 329 (1978).
[9] J. P. Antoine, J. Math. Phys. **21**, 268, 2067 (1980).
[10] J. P. Antoine and F. Mathot, Ann. Inst. H. Poincaré **37**, 29 (1982).
[11] J. P. Antoine and W. Karwowski, in *Quantum Theory of Particles and Fields*," edited by J. Lukierski and B. Jancewicz, (World Scientific, Singapore, 1983), pp. 13–30; and *Proceedings of the II International Conference on Operator Algebras, Ideals and their Applications in Theoretical Physics 1983* (Teubner, Leipzig, to be published).
[12] M. Friedrich and G. Lassner, Wiss. Z. Karl-Marx Univ. Leipzig, Math. Naturwiss. R **27**, 245 (1978).
[13] G. Lassner, Rep. Math. Phys. **3**, 279 (1972).
[14] G. Köthe, *Topological Vector Spaces I* (Springer Verlag, Berlin, 1969).
[15] J. P. Antoine and F. Mathot, "Partial inner product spaces generated by unbounded operators," Universite Catholique de Louvain Preprint UCL, in preparation.
[16] P. G. Dixon, Proc. London Math. Soc. **23**, 53 (1971).

Holonomy groups, sesquidual torsion fields, and SU(8) in $d = 11$ supergravity

L. Castellani, R. D'Auria, and P. Fré
Istituto di Fisica dell' Universitá, Torino and INFN Sezione di Torino, Torino, Italy

P. van Nieuwenhuizen[a)]
Institute for Theoretical Physics, Princetonplein 5, P. O. Box 80.006, 3508 TA Utrecht, The Netherlands

(Received 1 June 1983; accepted for publication 25 May 1984)

The torsion and its curl form an anti-self-dual SO(8) tensor field $F_{ABCD}(y)$. The Maxwell equations are solved if this tensor is covariantly constant, while the Einstein equations are solved if it satisfies an algebraic relation at the origin. Such F_{ABCD} are found as the kernel of the holonomy group in the 35 representation of SO(8) and they extend the SO(8) of the round S_7 to SU(8).

PACS numbers: 02.40. + m, 04.50. + h, 11.30.Pb

Consider a general seven-dimensional compact but not necessarily connected Riemannian manifold M_7 with spin connection B_α^{ab} and vielbein B_α^a. The structure equations of this M_7 read

$$T^a \equiv dB^a + B_b^a \wedge B^b = 0\,, \tag{1a}$$

$$\mathscr{R}^{ab} \equiv dB^{ab} + B_c^a \wedge B_b^c = \mathscr{R}_{cd}^{ab} B^c \wedge B^d\,. \tag{1b}$$

We recall that a Riemannian manifold satisfies (1a) with $B^{ab} = B_c^a \eta^{cb} = -B^{ba}$. We will assume that B^a and B^{ab} admit an isometry group G which acts transitively on M_7 with subgroup H. Hence, we shall have $H \subset \mathrm{SO}(7)$, M_7 will be identified with the coset space G/H, and the intrinsic components of the curvature R_{cd}^{ab} will be constants.

A solution of supergravity must satisfy (1a) and the Ricci tensors obtained from (1b) must satisfy the Einstein equation on M_7

$$\mathscr{R}_{ac}{}^{cb} = -\tfrac{1}{6}(F_{acde} F_b^{cde} - \tfrac{1}{12} g_{ab}\{F_{cdef}^2 - 24e^2\})\,, \tag{2}$$

which follows from the Freund–Rubin ansatz,[1] where F_{MNPQ} splits into a 7-part $F_{abcd}(y)$ and a 4-part $F_{mnrs} = ie\epsilon_{mnrs}$. (Our conventions are the same as in Ref. 10.) With this ansatz the Maxwell equations in 4-space-time are automatically satisfied, while in 7-space they read

$$D_a^{\mathrm{SO}(7)} F^{abcd} + e\sqrt{2}\epsilon^{bcdefgh} F_{efgh} = 0\,. \tag{3}$$

Indices $M,N,\ldots$ run from 1,11 while $m,n,\ldots = 1,4$ and $a,b,\ldots = 1,7$. All indices are flat and F_{MNPQ} has unit strength.

The spin connection and vielbein in (1a) and (1b) can be combined into an SO(8) connection Ω_A^B $(A,B = 1,8)$

$$\Omega^{ab} = B^{ab}\,, \quad \Omega^{b8} = -\Omega^{8b} = \rho B^b\,, \tag{4}$$

where ρ is at this point an arbitrary constant. With this SO(8) connection we will define SO(8) covariant derivatives.

We now show that the Maxwell equation in (3) is automatically solved by introducing an SO(8) covariantly constant anti-self-dual tensor $F_{ABCD}(y)$

$$D^{\mathrm{SO}(8)} F_{ABCD} = dF_{ABCD} + 4\Omega_{[D}{}^{D'} F_{ABC]D'} = 0\,, \tag{5}$$

$$F_{ABCD} = -(1/4!)\epsilon_{ABCDEFG} F^{DEFG}\,. \tag{6}$$

Decomposing the SO(8) indices into SO(7) indices, one finds

$$D_e^{\mathrm{SO}(7)} F_{abcd} + 4\rho\delta_{e[d} F_{abc]8} = 0\,, \tag{7}$$

$$D_e^{\mathrm{SO}(7)} F_{abc8} - \rho F_{abce} = 0\,, \tag{8}$$

$$F_{abcd} = -(1/3!)\epsilon_{abcd}^{efg} F_{efg8}\,. \tag{9}$$

Equation (8) is the definition of the photon curl and tells us that F_{abc8} is proportional to the photon field A_{abc}. Recalling that F_{abcd} has unit strength we obtain

$$F_{abc8} = -\rho A_{abc}\,. \tag{10}$$

Equation (9) defines what we call sesquiduality, namely the fact that the potential A_{abc} is dual to its curl. The anti-self-duality in (6) makes (7) and (8) equivalent. Contracting the indices e and a in (7) and using (9) we retrieve the Maxwell equation (3), provided we fix ρ appropriately, namely,

$$\rho = -6e\sqrt{2}\,. \tag{11}$$

The integrability condition of (5) reads

$$R_{[D}{}^{D'} F_{ABC]D'} = 0\,, \tag{12}$$

where $R^{ab} = \mathscr{R}^{ab} - \rho^2 B^a \wedge B^b$ and $R^{a8} = \rho T^a$ are the components of the SO(8) curvature 2-form defined on M_7. Since M_7 is by assumption Riemannian, T^a in (1a) vanishes and (12) is equivalent to

$$\mathscr{R}_{[d}{}^{d'} F_{abc]d'} = 0\,, \quad \mathscr{R}_{[c}{}^{c'} A_{ab]c'} = 0\,. \tag{13}$$

For the round S_7,[2] where M_7 is maximally symmetric, $\mathscr{R}_{cd}^{ab}$ is a product of Kronecker delta functions. Choosing the normalization of $\mathscr{R}_{cd}^{ab}$ such that R_{cd}^{ab} vanishes, the integrability condition (13) is satisfied. In general the R_{cd}^{ab} do not vanish although they are constant, and the subalgebra of SO(7) generated by the following linear combination of SO(7) generators:

$$C_{cd} = (\mathscr{R}_{cd}^{ab} - \rho^2 \delta_{cd}^{ab}) J_{ab}\,, \tag{14}$$

must annihilate all SO(7) irreducible representations contained in the SO(8) representation F_{ABCD}. Because of sesquiduality there is only one independent SO(7) irreducible part in this case, namely F_{abcd}. We see thus that a covariantly constant anti-self-dual representation F_{ABCD} of SO(8) exists if and only if there exists an F_{abcd} which is a singlet under the group generated by the C_{cd} in (14).

The group generated by SO(7) generators in (14) is called the holonomy group. Its relevance for supergravity was first noted in the case of covariantly constant spinors on the squashed S_7.[3] Our general discussion applies to any re-

[a)] On leave from The State University of New York at Stony Brook, Stony Brook, New York.

 0022-2488/84/113209-05$02.50

presentation and any manifold. The holonomy group is representation independent on a given manifold. For example, the holonomy group G_2 for covariantly constant spinors on the squashed S_7 is also the holonomy group for our F_{abcd} on the same manifold. It was recently shown that the squashed S_7 is the coset $\mathrm{Sp}(4)\times\mathrm{Sp}(2)/\mathrm{Sp}(2)\times\mathrm{Sp}(2)$.[4]

On a different manifold, the holonomy group will be different, examples being the round sphere whose holonomy group is the unit group, or the coset spaces $\mathrm{SU}_3\times\mathrm{SU}_2\times\mathrm{U}_1/\mathrm{SU}_2\times\mathrm{U}_1\times\mathrm{U}_1$,[5] whose holonomy group is SO(7) or (in one case) SU_3 as we shall discuss.

Let us briefly discuss the relation between our definition of holonomy group and the holonomy group one is familiar with in general relativity, that is, the holonomy group of Riemannian geometry. In general relativity one parallel transports a vector v^a around a closed curve and finds a holonomy group generated by $\mathscr{R}^{ab}_{cd}J_{ab}$ while in our case it is generated by $R^{ab}_{cd}J_{ab}$. Thus, the maximally symmetric n-sphere has in general relativity the holonomy group SO(n) while in our case S_n has as holonomy group the unit group. Said differently, in general relativity one considers the holonomy group of the metric connection (= spin connection) while we consider here the holonomy group of the de Sitter connection [SO(8) in our case] which is the sum of the spin connection plus the vielbein term. The vector representation of the de Sitter group splits into a vector plus a scalar under the Lorentz subgroup, and since one restricts one's attention in general relativity to Lorentz vectors, one cannot consider the de Sitter holonomy group. However, for such representations as spinors and anti-self-dual F_{ABCD} which remain irreducible under the Lorentz subgroup, one can define both holonomy groups.

Returning to the main theme, we note that we have solved the Maxwell equations provided we can find an SO(7) tensor F_{abcd} which is invariant under the C_{cd} in (14). The $\mathscr{R}^{ab}_{cd}$ which appear in (14) must also satisfy the Einstein equation in (2). The Ricci tensors $\mathscr{R}^{bc}_{ab}$ in (2) are constant, but the Maxwell curls F_{abcd} depend on the coordinates y of M_7. Because of the covariant constancy, the dependence of $F_{ABCD}(y)$ on y can be obtained by sweeping out $F_{ABCD}(0)$ from the origin in a way we will now first describe. Afterwards, we will come back to the Einstein equations.

Consider any SO(8) covariantly constant SO(8) representation, denoted generically by $F(y)$. Consistency of the covariant constancy requires that the SO(8) Lie algebra valued curvatures $R^{AB}J_{AB}$ in (14) annihilate $F(y)$ at all y. Hence, $F(y)$ must belong to the null space N of the matrix $R^{AB}J_{AB}$

$$R^{AB}J_{AB}v=0\Leftrightarrow v\in N,\quad F(y)\in N\quad\text{for all } y. \tag{15}$$

The 21 SO(8) Lie algebra valued matrices $R^{AB}_{cd}J_{AB}\equiv C_{cd}$ lie actually in SO(7) since the torsion T^a vanishes as explained above. [Our connection is the Riemann (= metric) connection which has per definition vanishing torsion, and not the connection which appears in the Maurer–Cartan equation of the coset manifold G/H. This Riemannian connection is in general not equal to the H-connection, but we need it since it appears in the field equations of supergravity.] In general the C_{cd} do not already span a Lie algebra; however their commutators generate the holonomy algebra C of the given manifold, which may coincide with SO(7) or be a proper subgroup thereof. Given the null space N of C, we denote by Σ the set of generators of SO(8) which annihilate N. Clearly Σ is a (proper or inproper) subalgebra of SO(8) and contains C:

$$C=\mathrm{SO}(7)\cap\Sigma. \tag{16}$$

Since $CF(y)=0$ for all y, and $F(y)\in N$, also $\Sigma F(y)=0$ for all y. Let us denote by $g(y)$ the SO(8) matrix which sweeps $F(y)$ out from the origin and the group generated by Σ by S. Then $gS\in\mathrm{SO}(8)$ and $(gS-Sg)v=0$ for $v\in N$. Hence,

$$F(y)=g(y)F(0),\quad g(y)\in\mathrm{SO}(8)\quad\text{for all } y. \tag{17}$$

Clearly, $g(y)$ maps nulvectors of Σ into null vectors of Σ. Let us denote the generators of $g(y)$ by Z. Then we must have

$$[Z,\Sigma]\subset\Sigma. \tag{18}$$

These generators Z define a new subalgebra of SO(8) because of the Jacobi identities. Z is the normalizer of Σ in SO(8), i.e., Z is the largest subalgebra of SO(8) in which Σ is an invariant subalgebra (an ideal). The matrix $g(y)$ can now be written as $\exp\epsilon^I(y)Z_I$, where I labels the generators of Z. The relation between $\epsilon^I(y)$ and the SO(8) connection $\Omega_A{}^B(y)$ is dictated by the requirement that F be covariantly constant, namely by

$$(dg(y)+\Omega^{AB}(y)J_{AB})F(0)=0. \tag{19}$$

Example: Suppose $\Sigma=\mathrm{SO}(p)$. Writing SO(8) as $\mathrm{SO}(p+q)$, clearly $[\mathrm{SO}(p),\mathrm{SO}(q)]=0$, hence Z contains $\mathrm{SO}(p)\times\mathrm{SO}(q)$. Since the Grassmann manifold $\mathrm{SO}(p+q)/\mathrm{SO}(p)\times\mathrm{SO}(q)$ is symmetric, no other generators of $\mathrm{SO}(p+q)$, when commuted with $\mathrm{SO}(p)$, produce $\mathrm{SO}(p)$, hence, $Z=\mathrm{SO}(p)\times\mathrm{SO}(q)$.

Having discussed how $F_{ABCD}(y)$ depends on y, let us now come back to the Einstein equations in (2). Since

$$F_{abcd}(y)=Y^A_a(y)Y^B_b(y)Y^C_c(y)Y^D_d(y)F_{ABCD}(0), \tag{20}$$

where $Y^B_A(z)$ are SO(8) matrices in the vector (= defining) representation [the product of four Y's is the matrix $g(y)$ in (17)], we can write

$$\begin{aligned}F_{acde}(y)F^{cde}_b(y)&=Y^A_a(y)Y^B_b(y)F_{ACDE}(0)F_B^{C'D'E'}(0)\\&\quad\times(\delta^C_{C'}-Y^C_8Y^8_{C'})(\delta^D_{D'}-Y^D_8Y^8_{D'})(\delta^E_{E'}-Y^E_8Y^8_{E'}).\end{aligned} \tag{21}$$

Due to the antisymmetry of the $F_{ACDE}(0)$, only terms with three and two Kronecker deltas contribute

$$\begin{aligned}F_{acde}(y)F^{cde}_b(y)&=Y^A_a(y)Y^B_b(y)[F_{ACDE}(0)F_B^{CDE}(0)\\&\quad-3F_{ACDE}(0)F_{BC'}{}^{DE}(0)Y^C_8(y)Y^{C'}_8(y)].\end{aligned} \tag{22}$$

For the completely contracted term $F^2_{abcd}(y)$ in (2) one finds

$$F^2_{abcd}(y)=F^2_{ABCD}(0)-4F_{ACDE}(0)F^{ACDE'}Y^E_8(y)Y^8_{E'}(y). \tag{23}$$

Since $\mathscr{R}^{cd}_{ab}$ in (2) is constant, so must be (22) and (23). Requiring that (22) and (23) must each separately be constant (clearly a sufficient as well as a necessary condition) we can fulfill this requirement by imposing the following condition on

TABLE I. Table of holonomy groups of solutions.[a]

G/H Internal coset space	C Holonomy group in SO(7)	Σ Extension of the holonomy group to SO(8)	Z Centralizer of the Σ
$S_7 = \frac{SO(8)}{SO(7)}$	1	1	SO(8)
$S_7^{sq} = \frac{SO(5) \otimes SO(3)}{SO(3) \otimes SO(3)}$	G_2	SO(7)	1
$\frac{SU(3) \times SU(2) \times SU(1)}{SU_2 \times U_1 \times U_1}$ $p/q \neq 1$	SO(7)	SO(8)	1
$\frac{SU_3 \times SU_2 \times U_1}{SU_2 \times U_1 \times U_1}$ $p/q = 1$	SU(3)	SO(6)	SO(2) $\otimes$ SO(6)
$\frac{SU_2 \times SU_2 \times SU_2}{U_1 \times U_1}$ $p \neq q \neq r$	SO(7)	SO(8)	1
$\frac{SU_2 \times SU_2 \times SU_2}{U_1 \times U_1}$ $p = q = r = 1$	SU(3)	SO(6)	SO(2)$\times$SO(6)

[a] The last two cases are treated in Ref. 9.

$F_{ABCD}(0)$:

$$F_{ACDE}(0)F^{BC'DE}(0) = \alpha(\delta_A^B \delta_C^{C'} - \delta_A^{C'} \delta_C^B) + h_{AC}^{BC'}, \quad (24)$$

where α is a constant and where h is antisymmetric in AB and CC'. In that case the y-dependence drops out of the Einstein equations and one is led to the following two algebraic equations:

$$\mathcal{R}_{ac}^{cb} = -\tfrac{1}{6}(7\alpha - 3\alpha - \tfrac{1}{12}\{28\alpha - 24e^2\})\delta_{ab}, \quad (25)$$

$$\mathcal{R}_{mk}^{kn} = -\tfrac{1}{6}(-6e^2 - \tfrac{1}{12}\{28\alpha - 24e^2\})\delta_{mn}. \quad (26)$$

Due to the condition in (24), M_7 is an Einstein space and M_4 an anti-de Sitter space-time. The constant α is not arbitrary, but must be fixed such that the $\mathcal{R}_{ab}$ satisfy the holonomy condition, namely such that C_{cd} in (14) have a nontrivial null space. Hence, the radii of internal and external space are fixed and universal.

To appease the anxieties of the reader that a tensor $F_{ABCD}(0)$ with the properties in (24) may not exist, we merely give an example. Consider $F_{ABCD}(0) = a\bar{\eta}\Gamma_{ABCD}\eta$, where η is any constant Majorana–Weyl spinor in 8 dimensions with $\bar{\eta}\eta = 1$. Then, by Fierzing, one finds (24) with h equal to zero and $\alpha = 12a^2$. (The 8×8 matrix Γ_8 is here equal to $-i$.)

This concludes our general treatment. We now give three examples, see also Table I.

(*i*) *Round S_7 with torsion:* Since $\mathcal{R}_{cd}^{ab}$ is maximally symmetric, the holonomy group C is either SO(7) or the unit group, depending on the value of the cosmological constant Λ. By choosing Λ such that the SO(8) curvature vanishes, $C = 0$. This value of Λ is indeed the one which follows from the Maxwell equations. Since the null space is the whole space, there are no SO(8) generators which leave the null space invariant. Hence also $\Sigma = 0$. It follows that the normalizer of Σ in SO(8) coincides with SO(8): $Z = SO(8)$. Hence, in this case a general representation F depends on y as $F(y) = g(y)F(0)$, where $g(y)$ is any SO(8) matrix whose coefficients depend on y.[6]

(*ii*) *The squashed S_7 with torsion:* In this case C equals G_2 (recall, C is a subgroup of SO(7) and representation independent). From the explicit form of the Riemann curvatures, C was identified as G_2.[3] Hence, the subalgebra of SO(8) which leaves, for example (recall that Σ is representation independent), an SO(8) spinor invariant, is SO(7). The normalizer Z of SO(7) in SO(8) is this SO(7) itself (see the example). Hence, the spinor is invariant under Z, and thus the spinor is actually constant. Thus the solution is $F_{ABCD}(y) = F_{ABCD}(0) = \bar{\eta}\Gamma_{ABCD}\eta$, which is the dual of the associator of the octonions.

(*iii*) *$SU_3 \times SU_2 \times U_1 / SU_2 \times U_1 \times U_1$ with torsion:* There is actually an infinite class of these coset spaces, depending on the choice of the U(1) subgroups.[5] The topology of these coset spaces depends only on the ration p/q of two integers p and q. If $p \neq q$, the holonomy group is SO(7) in which case there is no sesquidual torsion and at least our method does not provide a solution. If $p = q$, the holonomy group is SU_3 [G_2 is the subgroup of spin (7) which leaves one spinor invariant ($\underline{8}$ spin $= \underline{7} + \underline{1}$). The subgroup of G_2 which leaves an element of this $\underline{7}$ invariant is SU_3.] The subgroup of spin (8) which leaves two spinors invariant is SO(6), thus $\Sigma = SO(6)$. [Indeed, SO(6)$\cap$spin(7) $= SU_3$. Note also that spin (7) $\cap$SO(7) $= G_2$.] The normalizer of SO(6) in SO(8) equals SO(6)$\times$SO(2); see the example. Hence, $g(y)$ lies in

$SO(6)\times SO(2)$. Since SO(6) lies in Σ, it acts trivially on $F(0)$, hence $g(z)$ depends on only one coordinate φ: $g(y) = \exp \epsilon(\varphi) T(SO_2)$. One may first compute all the components of Ω^{AB} and then explicitly solve $(dy + \Omega g)\, F = 0$. The Ω^{AB} connection is an algebraic expression in terms of the vielbein and H-connection of the coset manifold. The only O(2) generator which does not act trivially on $F(y)$ is multiplied by a linear combination of coset vielbeins and H-connections. Hence, calling this combination V, we must solve $d \exp \epsilon(\varphi) T(SO_2) + V \exp \epsilon(\varphi) T(SO_2) = 0$.

Now we will argue that we can choose coordinates on the coset manifold $M_7 = G/H$ such that V only depends on φ. The argument goes as follows: Since there are two covariantly constant spinors (SU_3 is the holonomy group), the supergravity model will have an $N = 2$ supersymmetry. The supergroup will be $Osp(2/4)\times S'$, where S' is a purely bosonic group. Hence, in the bosonic sector, there will be an SO(2) generator which commutes with all other bosonic generators [namely Sp(4) and S']. Choosing the coordinate φ along the direction of this generator, $V = V(\varphi)$ is a well-determined function of φ, and one can solve the different equation for $\epsilon(\varphi)$.

We now change gears and come back to a property of the torsion tensor on the round seven sphere. As we will now derive using the formalism outlined above, its symmetry group is SO(7) as first conjectured by Warner, and proven by Castellani and Warner, and Englert *et al.*[8] and not G_2, as initially advocated by several authors, the present authors included.[6] We recall that

$$F_{ABCD}(y) = Y_A^{A'}(y) \cdots Y_D^{D'}(y) F_{A'B'C'D'}(0)\,, \tag{27}$$

where $Y_A^{A'}(y)$ defines the coset elements at y. We recall the definition of a Killing vector $k_{AB}^{\mu}(y)$. Let g be an arbitrary group element of SO(8) near the identity, $g = I + \epsilon^{AB} J_{AB}$ with constant ϵ^{AB}. Then [we use here the right cosets $HY(z)$ of SO(8)/SO(7)]

$$Y(y)g = H(y,g) Y(y')\,, \tag{28}$$

where $y'^{\mu} = y^{\mu} + \epsilon^{AB} k_{AB}^{\mu}(y)$, and $H(y,g)$ is an element of SO(7) subgroup given by $H(y,g) = 1 + \epsilon^{AB} W_{AB}^{ab}(y) J_{ab}$. Thus we find for the Lie derivative of the spherical harmonic $Y_A^B(y)$ the following result:

$$\epsilon^{AB}\left[l_{\mathbf{k}_{AB}} Y_c^D(y) + W_{AB,C}{}^{c'}(y) Y_{c'}^D(y) \right] = Y_c^{D'}(y) \epsilon_{D'}{}^{D}\,, \tag{29}$$

$$\epsilon^{AB}\left[l_{\mathbf{k}_{AB}} Y_8^D(y) \right] = Y_8^{D'} \epsilon_{D'}{}^{D}\,. \tag{30}$$

In particular, if we use the sweeping out matrix $Y_A^B(y)$ to define vielbeins and connections by dYY^{-1}, the rigid SO(8) transformations cancel and these vielbeins and H-connections are invariant under the Lie derivative up to H-gauge transformation with parameter $W_{AB,c}{}^{c'}(y)$.

Returning to the problem of the invariance group of $F_{ABCD}(y)$, we must find which Killing vectors $\epsilon^{AB}\mathbf{k}_{AB}$ leave $F_{ABCD}(y)$ invariant up to a W_{AB} SO(7) transformation. Clearly, the sum of the action of a Killing vector and a $(W_{AB})_a^b$ rotation of the index a of the harmonic $Y_a^B(y)$ is equivalent to an SO(8) rotation of the index B. Hence, the symmetry group of $F_{abcd}(y)$ is the same as that of $F_{abc8}(y)$ and is equal to the SO(8) subgroup which leaves $F_{ABCD}(0)$ invariant. Since $F_{ABCD}(0)$ is equal to $\bar{\eta}\Gamma_{ABCD}\eta$ and the Dirac matrices are SO(8) invariant tensors, the answer is the following: The invariance group of $F_{abcd}(y)$ is the stability group of a spinor in spin (8). Now spin (8) are the same matrices as SO(8) in the vector (= defining) representation, and the subgroup of the latter which leaves a given 8-component vector invariant, is SO(7). Thus,

$$F_{abcd}(y) = Y_a^{a'}(y) Y_b^{b'}(y) Y_c^{c'}(y) \left[Y_d^{d'}(y) F_{a'b'c'd'}(0) + Y_d^8(y) F_{abc8}(0) \right] \tag{31}$$

is physically invariant under a full SO(7). The G_2 subgroup leaves each term separately invariant, but the remaining seven generators leave only the sum invariant.

In (6) we introduced the notion of Killing vectors and saw that there were as many of them as there are generators in the group G. For the round seven sphere with $G = SO(8)$ there are thus 28 Killing vectors. One can use covariantly constant spinors to give an explicit representation.[2] Actually, there are two sets of covariantly constant spinors which can be written as[6]

$$\eta_{\pm}(y) = (1 + y^2)^{-1/2}(1 \pm \not{y})\xi\,, \quad \not{y} \equiv i y^{\alpha} \Gamma_{\alpha} \quad (\alpha = 1,7)\,. \tag{32}$$

There must be a relation between the two sets of Killing vectors

$$\begin{aligned} K_{IJ}^{a,+} &= \eta_{I,+}^T(y) \Gamma^a \eta_{J,+}(y) \\ \text{and} \qquad & \qquad (33) \\ K_{IJ}^{a,-} &= \eta_{I,-}^T(y) \Gamma^a \eta_{J,-}(y)\,. \end{aligned}$$

Although it was clear to most that such a relation should exist, its form was unknown. We now present it here.

Define $\xi_J = (0,0,\ldots,1,0,\ldots,0)$. Then,

$$K_{IJ}^{a,+}(y) = \left[(1 - \not{y})(1 + y^2)^{-1} \Gamma^a (1 + \not{y}) \right]_{IJ}\,. \tag{34}$$

Contract now with 8×8 matrices Γ_{JI}^{AB} where the 16×16 SO(8) matrices Γ^A are related to the SO(7) matrices Γ^a by $\Gamma^A = \Gamma^a \times \tau^2$, $\Gamma^8 = I \times \tau'$, so that $\Gamma^{8a} = i\Gamma^a$. Since in a trace one gets the same answers if one replaces Γ^a by $-\Gamma^a$ we have the identity

$$K_{IJ}^{a,+} \begin{pmatrix} \Gamma^{ab} \\ i\Gamma^a \end{pmatrix}_{JI} = -K_{IJ}^{a,-} \begin{pmatrix} \Gamma^{ab} \\ -i\Gamma^a \end{pmatrix}_{JI}\,. \tag{35}$$

Using the completeness relation

$$\Gamma_{JI}^{AB} \Gamma_{KL}^{AB} = 8(\delta_{JK}\delta_{IL} - \delta_{JL}\delta_{IK})\,, \tag{36}$$

we arrive at the desired relation

$$K_{LK}^{a,+}(y) = \tfrac{1}{16} K_{IJ}^{a,-}(y) (\Gamma_{JI}^{ab} \Gamma_{KL}^{ab} + \Gamma_{JI}^{a} \Gamma_{KL}^{a})\,. \tag{37}$$

The last issue we want to address ourselves to is a group theoretical aspect of the parallelizing torsion [i.e., of the internal photon $A_{\alpha\beta\gamma}(y)$ on the round S_7 (Ref. 7)]. Consider the SU(8) Cartan–Maurer equation instead of the SO(8) Maurer–Cartan equation. Decomposed with respect to SO(7) they read

$$dB^a + B_b^a \wedge B^b + (2/\rho)\epsilon^{ab_1b_2b_3c_1c_2c_3} B_{b_1b_2b_3} \wedge B_{c_1c_2c_3} = 0\,, \tag{38}$$

$$dB^{ab} + B_c^a \wedge B^{cb} - \rho^2 B^a \wedge B^b - 72 B^{acd} \wedge B_{cd}^b = 0\,, \tag{39}$$

$$dB^{abc} + 3B^{[c}{}_{c'} \wedge B^{ab]c'} + (\rho/6)\epsilon^{abcde_1e_2e_3} B_d \wedge B_{e_1e_2e_3} = 0\,. \tag{40}$$

We obtained these results in Ref. 6 by evaluating $dM + M \wedge M = 0$ with $M = g^{-1}\, dg = \frac{1}{4} B^{ab} \Gamma_{ab} - (i/2)\rho B^{a} \Gamma_{a} + B^{abc} \Gamma_{abc}$. Suppose that we express these SU(8) connections in terms of the SO(8) connections $B^{ab}_{(0)}$ and $B^{a}_{(0)}$ as follows:

$$B^{a} = \lambda B^{a}_{(0)}\,, \quad B^{ab} = B^{ab}_{(0)} + \alpha A^{abc} B^{(0)}_{c}\,,$$
$$B^{abc} = \beta B^{abcd} B^{(0)}_{d}\,, \quad F^{abcd} = \pm (e/2) \bar{\eta} \Gamma^{abcd} \eta\,, \tag{41}$$

where α, β, and λ are constants. Let us further assume that $B^{ab}_{(0)}$, $B^{a}_{(0)}$, and A^{abc} constitute the solution of the round S_7 with torsion. It is remarkable that in that case (41) satisfies (38)–(40). One could rewrite (38)–(40) in a suggestive manner as

$$R^{AB} + kB^{APQR} B^{B}{}_{PQR} = 0 \quad [k = \text{constant}]\,,$$
$$D^{\mathrm{SO}(8)} B_{PQRS} = 0\,, \tag{42}$$

where B_{PQRS} is anti-self-dual and $B_{abc8} = \beta F_{abcd} B^{d}_{(0)}$. Clearly one can split B^{ab} into the parallelizing connection B^{ab}_{E} of Englert plus a remainder ΔB^{ab}_{E}. In that case the structure equations of SU(8) in (38) and (39) reduce to the structure equations of the parallelized S_7. Equation (40) is a differential equation for the torsion field which, upon contraction of a pair of indices, yields the Maxwell equations. Thus, the torsion which flattens the seven sphere is one of the components of a particular flat connection of SU(8). We conjecture that the torsion of other solutions of $d = 11$ supergravity is also one of the components of the flat connection of a group which would be a "hidden symmetry" in the four-dimensional theory.

ACKNOWLEDGMENT

We thank Tullio Regge for stressing the usefulness of the concept of sesquiduality.

Research was supported by NATO grant 23681.

[1] P. G. O. Freund and M. A. Rubin, Phys. Lett. B **97**, 233 (1980).
[2] M. J. Duff and C. N. Pope, as reported in M. J. Duff and D. J. Toms, *Proceedings Erice Europhysics Conference, 1981* (Plenum, New York, 1982).
[3] M. A. Awada, M. J. Duff, and C. N. Pope, Phys. Rev. Lett. **50**, 294 (1983).
[4] F. A. Bais, H. Nicolai, and P. van Nieuwenhuizen, Nucl. Phys. B **228**, 333 (1983).
[5] L. Castellani, R. D'Auria, and P. Fré, Nucl. Phys. B **239**, 610 (1984).
[6] R. D'Auria, P. Fré, and P. van Nieuwenhuizen, Phys. Lett. B **122**, 225 (1983).
[7] F. Englert, Phys. Lett. B **119**, 339 (1982).
[8] N. P. Warner, Phys. Lett. B **128**, 169 (1983); L. Castellani and N. P. Warner, Phys. Lett. B **130**, 47 (1983); F. Englert, M. Rooman, and P. Spindel, Phys. Lett. B **130**, 50 (1983).
[9] R. D'Auria, P. Fré, and P. van Nieuwenhuizen, Phys. Lett. B **136**, 347 (1984).
[10] E. Bergshoeff, M. de Roo, B. de Wit, and P. van Nieuwenhuizen, Nucl. Phys. B **195**, 97 (1982).

Liouville and Painlevé equations and Yang–Mills strings

Cihan K. Saçlioğlu
Physics Department, Boğaziçi University, Bebek, Istanbul, Turkey

(Received 23 February 1984; accepted for publication 15 June 1984)

Stringlike solutions of the self-dual Yang–Mills equations (dimensionally reduced to R^2) are sought. A multistring *Ansatz* results in the sinh–Gordon and Liouville equations. According to a general theorem, the solutions must be either real and singular and have infinite action, or complex and nonsingular, with zero action. In the Liouville case, explicit arbitrarily separated n-string solutions of both classes are given. The magnetic flux for these solutions is found to be the Chern class of a Kaehler manifold, and it consequently assumes quantized values $4\pi n/e$. The axisymmetric version of the sinh–Gordon is solved by the third Painlevé transcendent P_3, using the results on P_3 by Wu *et al.* [Phys. Rev. B **13**, 316 (1976)] and McCoy *et al.* [J. Math. Phys. **18**, 10 (1977)]. The axisymmetric case can be cast into the Ernst equation framework for the generation of further solutions. In the Appendix, the Euclideanized Ernst equation is shown to give self-dual Gibbons–Hawking gravitational instantons.

PACS numbers: 02.40.Ky, 11.15.Kc

I. INTRODUCTION

Self-dual Yang–Mills (SDYM) equations in four Euclidean dimensions share many properties with totally integrable systems in two dimensions. For example, Bianchi–Backlund transformations, nonlocal conservation laws, an associated linear problem, and a Kac–Moody algebra have been constructed[1] for SDYM fields.

Our aim in this paper is twofold: The first is to point out that another feature of totally integrable systems, namely that of reducibility of the field equations[2] under certain restrictions to ordinary differential equations of the Painlevé type, also follows naturally from the SDYM equations when stringlike solutions of the SDYM equations are sought.[3] This, together with the fact that self-dual monopoles, i.e., the once dimensionally reduced form of the theory, have also been related to integrable systems,[4] further strengthens the possibility that integrability is an inherent property of the SDYM equations, rather than a property limited to specific *Ansätze*.

Our second purpose is to present and examine these stringlike solutions. The search for such solutions constitutes a natural step in dimensional reduction: Just as monopoles solve the static SDYM equations, one may expect to find Nielsen–Olesen strings[5] when the theory is further reduced to R^2. Indeed, this reduction results in two "Higgs fields" (say A_3^a and A_4^a), which is just what is needed to break SU(2) completely. However, unlike with monopoles, here there is a price to be paid: A general theorem dictates[6] that these solutions have to be either real and singular, or complex and nonsingular. The action per unit x_3x_4 is then infinite in the former case and zero in the latter. Nevertheless, we believe there are good reasons to warrant an examination of both classes of solutions, in addition to the obvious one that they are there.

(i) Complex Yang–Mills solutions may play a physical role, just as complex solutions of the anharmonic oscillator represent the WKB approximation.[7]

(ii) The complex solutions, owing to their vanishing action, will have the same Boltzmann factor as the vacuum in the functional integral.

(iii) The solutions exhibit flux quantization in multiples of $(2\pi/e)$ as in the original Nielsen–Olesen model. The topological origins of the quantization is clearly seen in the Liouville case, which also admits an explicit, arbitrarily separated n-string solution with flux equal to $(4\pi/e)n$.

(iv) General approaches to the SDYM equations such as Yang's equations,[8] the Atiyah–Ward method,[9] and Backlund transformations[1] all involve a complexification of coordinates and/or field quantities. Thus complex solutions are a natural part of this general framework. Indeed, Burns[10] has recently given a class of complex nonsingular solutions based on the Atiyah–Ward *Ansatz*. In the following, we will present a new set of explicit solutions, both in real and complex forms.

Finally, it should be noted that real, nonsingular strings also do not inherit all the attractive features of SDYM monopoles: First of all, they can only be obtained by going outside the pure Yang–Mills system; second, their Bogomolny equations[11] reduce to the Poisson–Boltzmann equation

$$(\partial_1^2 + \partial_2^2)\rho = e^{2\rho} - 1, \tag{1}$$

which is believed[12] to be nonintegrable. The pure Yang–Mills strings, on the other hand, are described by the sinh–Gordon or the Liouville equations, both of which are integrable. The previously mentioned Painlevé equations are in fact just the axisymmetric special cases of these partial differential equations. Notice that (1), in contrast, does not reduce to any of the Painlevé equations.

The contents of the paper, in more detail, are as follows: In Sec. II, we first reproduce the aforementioned theorem due to Lohe.[6] We then present a general *Ansatz* depending on two coordinates and show that self-duality results in a single function satisfying either the sinh–Gordon or the Liouville equation. In Sec. III, we consider the axisymmetric cases of these equations and thus obtain special Painlevé transcendents of the third and fifth kinds, P_3 and P_5. We then summarize the results found in a previous paper on a parti-

 0022-2488/84/113214-07$02.50

cular P_3 solution. The same P_3 has been studied (see Ref. 13) in the context of the Ising model. Some of these mathematical results are directly applicable here. We treat the Liouville equation in Sec. IV and present its multistring solutions in both real singular and complex nonsingular forms. The flux is seen to be the integral of a Kaehler form and given by $(4\pi/e)$ times the number of strings. The axisymmetric solutions are cast into the Ernst equation[14] formalism in Sec. V, with a view to generating axisymmetric n-string solutions of P_3 using the solution generating techniques associated with the Ernst equation. After some concluding remarks in Sec. VI, we return in the Appendix to the Ernst equation, this time in Euclidean space-time, and point out that it, too, is related to self-duality, but in a different context: Solutions of this system are Gibbons–Hawking metrics[15] with self-dual Riemann tensors.

II. GENERAL PROPERTIES OF SDYM STRINGS

A. A no-go theorem

An appropriate starting point for seeing what can be done with SDYM fields reduced to the plane is the following argument of Lohe[6]: The action per unit time per unit length, i.e., the tension, can be written as

$$T=\int dx_1\,dx_2\{\tfrac{1}{4}(F^a_{ij}\pm e\epsilon_{ij}\epsilon^{abc}\psi^b\varphi^c)^2 + \tfrac{1}{2}((D_i\psi)^a\mp\epsilon_{ij}(D_j\varphi)^a)^2 \mp((e/2)F^a_{ij}\epsilon_{ij}\epsilon^{abc}\psi^b\varphi^c-\epsilon_{ij}(D_j\varphi)^a(D_i\psi)^a)\}, \tag{2}$$

where $i,j=1,2$; $\psi^a=A^a_3$ and $\varphi^a=A^a_4$. For self-dual solutions one is left with the last term. Using $[D_i,D_j]=eF_{ij}$ on it, one obtains

$$T=\int dx_1\,dx_2\{\partial_1(\psi^a(D_2\varphi)^a)-\partial_2(\psi^a(D_1\varphi)^a)\}, \tag{3}$$

which, if converted to a line integral along a large closed curve, would appear to go to zero, being proportional to the covariant derivatives at infinity. As the original expression is positive definite for real Yang–Mills fields, one is forced to conclude that real, nonsingular self-dual solutions with asymptotically vanishing covariant derivatives are pure gauge.

As mentioned earlier, one possible remedy is to add a quartic Higgs potential by hand and thus abandon the pure Yang–Mills case. The Bogomolny equations then result in (1). However, if one decides to stay within the Yang–Mills system, the price to be paid is either a singularity or a complex solution: Singularities prohibit the conversion of (1) to a line integral, while complex potentials can give zero action without vanishing field strengths. A third possibility of a doubly periodic solution, where the covariant derivatives do not vanish at infinity, will be mentioned in Sec. VI.

B. A general *Ansatz*

We restrict ourselves to the group SU(2) throughout. Let us first note an important difference between monopole and vortex solutions: In the former, a massless U(1) field $A_\mu=\hat{\varphi}\cdot\mathbf{A}_\mu$ survives, while in the latter the symmetry is completely broken, so that the magnetic field exponentially decreases as we leave the center of the vortex. This broken magnetic field is projected out by $\mathcal{F}_{12}=\mathbf{F}_{12}\cdot\hat{\varphi}\times\hat{\psi}$. Thus we should seek an *Ansatz* where the "massive photon" $\mathbf{A}_{1,2}$ and the "Higgs fields" $\mathbf{A}_3\equiv\psi$, $\mathbf{A}_4=\varphi$ are all nonparallel to each other. We may therefore try

$$A_1=i\sigma_3 f,\quad A_2=i\sigma_3 g,\quad A_3=i\sigma_2 k,\quad A_4=i\sigma_1 h, \tag{4}$$

where the $A_\mu=A^a_\mu e\sigma^a/2i$ depend on (x_1,x_2) only. Self-duality gives

$$g_1-f_2=2hk, \tag{5a}$$

$$k_1/h=h_1/k=2g, \tag{5b}$$

$$k_2/2h=h_2/2k=-2f, \tag{5c}$$

where $f_1=\partial f/\partial x_1$, etc. Rearranging, one gets

$$hh_1-kk_1=hh_2-kk_2=0, \tag{6}$$

which of course implies

$$h^2-k^2=\text{const.} \tag{7}$$

The *Ansatz* thus corresponds to a hyperbola in (h,k) space, admitting equivalent parametrizations for (h,k) for a given solution. If, however, the constant is set equal to zero, the hyperbola degenerates to its asymptotes, and one obtains a new and distinct solution. With a nonzero constant, we may use the two equivalent parametrizations

$$h=a\cosh(\omega/2),\quad k=a\sinh(\omega/2); \tag{8a}$$

$$h=a\sec\chi,\quad k=a\tan\chi, \tag{8b}$$

while for the degenerate case we may set

$$h=k=a\exp\zeta, \tag{9}$$

where a is an integration constant with the dimensions of an inverse length. Substituting (8) and (9) in (5) and using new dimensionless coordinates $(x,y)=2a(x_1,x_2)$, the self-duality equations reduce to

$$\nabla^2\omega=\sinh\omega, \tag{10a}$$

$$\nabla^2\chi+(\nabla\chi)\cdot(\nabla\chi)\tan\chi=\tan\chi, \tag{10b}$$

and

$$\nabla^2\zeta=e^{2\zeta}. \tag{11}$$

C. Field strengths and invariants

We shall not refer to (10b) any further, except for a brief discussion showing the equivalence of a particular P_3 to a P_5 in the next section. Below we give the potentials and field strengths for (10a) and (11); the former in Eq. (12) and (13) and the latter in (14) and (15):

$$A_1=-i\sigma_3(a/2)\omega_y,\quad A_2=i\sigma_3(a/2)\omega_x,$$
$$A_3=i\sigma_2 a\sinh(\omega/2),\quad A_4=i\sigma_1 a\cosh(\omega/2), \tag{12}$$

$$F_{12}=F_{34}=ia^2\sigma_3\sinh\omega, \tag{13a}$$

$$F_{13}=F_{42}=ia^2(\sigma_2\omega_x\cosh(\omega/2)-\sigma_1\omega_y\sinh(\omega/2)), \tag{13b}$$

$$F_{14}=F_{23}=ia^2(\sigma_1\omega_x\sinh(\omega/2)+\sigma_2\omega_y\cosh(\omega/2)), \tag{13c}$$

and

$$A_1 = -i\sigma_3 a\zeta_y,\ A_2 = i\sigma_3 a\zeta_x,\ A_3 = i\sigma_2 a e^{\zeta},\ A_4 = i\sigma_1 a e^{\zeta}; \tag{14}$$

$$F_{12} = F_{34} = 2i\sigma_3 a^2 e^{2\zeta}, \tag{15a}$$

$$F_{13} = F_{42} = 2ia^2(\sigma_2\zeta_x - \sigma_1\zeta_y)e^{\zeta}, \tag{15b}$$

$$F_{14} = F_{23} = 2ia^2(\sigma_1\zeta_x + \sigma_2\zeta_y)e^{\zeta}. \tag{15c}$$

The tension T and the magnetic flux Φ are also easily calculated. From

$$T = \frac{1}{4a^2}\int\int dx\,dy\{\mathbf{F}_{12}\cdot\mathbf{F}_{12} + \mathbf{F}_{13}\cdot\mathbf{F}_{13} + \mathbf{F}_{14}\cdot\mathbf{F}_{14}\} \tag{16}$$

and

$$\Phi = \frac{1}{4a^2}\int\int dx\,dy\,\mathbf{F}_{12}\cdot\hat{\varphi}\times\hat{\psi}, \tag{17}$$

we find for the sinh–Gordon case

$$T = \frac{a^2}{e^2}\int\int dx\,dy\,\nabla\cdot(\sinh\omega\nabla\omega)$$

$$= \frac{a^2}{e^2}\int\int dx\,dy\,\nabla^2\cosh\omega, \tag{18}$$

$$\Phi = \frac{1}{2e}\int\int dx\,dy\nabla^2\omega = \frac{1}{2e}\int\int dx\,dy\sinh\omega, \tag{19}$$

while for the Liouville we get

$$T = \frac{4a^2}{e^2}\int\int dx\,dy\,\nabla\cdot(e^{2\zeta}\nabla\zeta)$$

$$= \frac{2a^2}{e^2}\int\int dx\,dy\,\nabla^2 e^{2\zeta}, \tag{20}$$

$$\Phi = \frac{1}{e}\int\int dx\,dy\,\nabla^2\zeta = \frac{1}{e}\int\int dx\,dy\,e^{2\zeta}. \tag{21}$$

Since all quantities have now been expressed in terms of ω and ζ, it only remains to find explicit solutions of (10a) and (11). In the following sections, an axisymmetric solution for P_3 and a more general one for the Liouville equation will be given.

III. PAINLEVÉ TRANSCENDENTS AND A SOLUTION OF P_3

A. Painlevé equations

It has been shown that there are only 50 ordinary differential equations of the form

$$\frac{d^2u}{dz^2} = F\left(u, \frac{du}{dz}, z\right), \tag{22}$$

where F is a rational function of its arguments, and the structure of the equation fixes the positions of the essential singularities and branch points regardless of the values of the integration constants.[16] Forty-four of these define familiar elementary functions including the elliptic ones, while six give rise to new transcendental functions called the Painlevé transcendents $P_1 - P_6$ (see Ref. 17). The interest of physicists in these functions stems from the experience that they always appear in connection with integrable systems: P_2 and the modified KdV equation,[18] P_5 and the Regge–Lund string model,[19] P_3 and the Ernst equation,[19] to name just a few. Furthermore, in integrable quantum theories, the P's parametrize correlation functions.[13,20]

In our problem, Painlevé equations follow directly from (10) and (11) when axisymmetric special solutions are sought and the transformations $\exp(\omega/2) = u$, $\exp(-2i\chi) = -w$, and $\exp\zeta = v$ are made. Then, with $r = \sqrt{x^2 + y^2}$, (10a) and (10b) lead to, respectively,

$$u'' + (1/r)u' = (1/u)(u')^2 + \tfrac{1}{4}(u^3 - u^{-1}), \tag{23a}$$

$$w'' + \frac{1}{r}w' = (w')^2\left\{\frac{1}{2w} + \frac{1}{w-1}\right\} + \frac{2w(w+1)}{w-1}, \tag{23b}$$

while the Liouville equation gives

$$v'' + (1/r)v' = (1/v)(v')^2 + v^3. \tag{24}$$

In the notation of Ince,[17] (23a) is a P_3 with $\alpha = \beta = 0$, $\gamma = -\delta = \frac{1}{4}$; (23b) a P_5 with $\alpha = \beta = \gamma = 0$, $\delta = 2$, and (24) a special case of (23a) with $\delta = 0$. From the fact that (23a) and (23b) come from the same constraint, it is clear that they must represent the same equation; and one can indeed be obtained from the other via the substitution $\cosh\omega/2 = \sec\chi$. This shows that for certain values of the parameters $\alpha, \beta, \gamma, \delta$, Painlevé transcendents of different kinds can become identical. Similarly, (24) with $\gamma = 1$, $\delta = 0$ can be converted into a form with $\gamma = 0$, $\delta = -1$ by inverting v.

B. A solution of P_3

Equation (24) is known to be soluble in terms of rational functions because of its special choice of coefficients. This, of course, is a consequence of the general solution to (11) given by Liouville.[21] We shall therefore treat (24) in the next section as a special case of the general equation. The P_3 of (23a), on the other hand, has been studied in Ref. 13 by their derivation of the correlation functions of the Ising model in the scaling limit. These authors give a simple one-parameter family of approximate solutions for $r \gg 1$ and $r \ll 1$, and a numerical solution in between for a value of the parameter corresponding to the critical point. We summarize below some of their pertinent results and those of an earlier paper.[3]

First, noting that $\omega = 0$ is a solution of (10a), we can look for an axisymmetric solution for which ω and $\omega' \to 0$ as $r \to \infty$. This is needed to make the fields pure gauge at infinity. Thus, neglecting nonlinear terms in (10a), we get

$$\omega'' + (1/r)\omega' - \omega = 0, \tag{25}$$

which gives $\omega \to cK_0(r)$ as $r \to \infty$. The corresponding solution of (23a) in Ref. 13 is

$$u(\lambda, r) \to 1 - 2\lambda K_0(r). \tag{26}$$

Thus $4\lambda = -c$.

In Ref. 13 the small-r behavior of $u(\lambda, r)$ is shown to depend crucially on the value of the parameter λ, and explicit approximate solutions in terms of elementary functions are given. The case studied in most detail, $\lambda = 1/\pi$, which represents the critical point in the Ising model, has also there been numerically solved for intermediate values of r, revealing a rapid monotonic increase to unity from zero at the origin. The role of the parameter λ is not clear in the Yang–Mills problem, but the case $\lambda = 1/\pi$, being the least-singular and best-examined one, will also be the only one treated here.

The solutions for other values of λ can, in principle, be brought to the same degree of completion when numerical work for intermediate r is extended to them. The $\lambda = 1/\pi$ solution behaves at the origin as

$$u(r) \to -(r/2)(\ln(r/8) + \gamma_E) + O(r^5 \ln^3 r), \tag{27}$$

where $\gamma_E = 0.577\,215\ldots$ is the Euler–Mascheroni constant.

To see what this behavior implies for the Yang–Mills field, we first apply the gauge transformation $U = \exp(-i\theta\sigma_3/2)$, where $\theta = \tan^{-1} y/x$ as usual, on the axisymmetric case of (12). The result is

$$A_r = 0,\ A_\theta = i\sigma_3 a(\omega'/2 + 1/r),$$
$$A_3 = i\sigma_\theta a \sinh(\omega/2),\ A_4 = i\sigma_r a \cosh(\omega/2), \tag{28}$$

with $\sigma_r = (\sigma_1 \cos\theta + \sigma_2 \sin\theta)$ and $\sigma_\theta = (\sigma_2 \cos\theta - \sigma_1 \sin\theta)$. This form of the potential with only an azimuthal component is clearly suited for describing strings along the x_3 axis. The asymptotic behavior $\omega \underset{r\to\infty}{\to} -4K_0(r)/\pi$ is now seen to give for the fields

$$|A^3_\theta| \underset{r\to\infty}{\to} 1/eR - (4a/e\pi)K_1(2aR), \tag{29a}$$

$$|\psi| = |A^\theta_3| \underset{r\to\infty}{\to} (4a/e\pi)K_0(2aR), \tag{29b}$$

$$|\varphi| = |A^r_4| \underset{r\to\infty}{\to} 2a/e, \tag{29c}$$

$$|F^3_{12}| \underset{r\to\infty}{\to} (8a^2/e\pi)K_0(2aR), \tag{29d}$$

where clearly $2aR = r$. Interpreting A_4 as a scalar field, (29a) and (29c) are precisely the same asymptotic expressions found in the original Nielsen–Olesen solution. On the other hand, $|\psi|$ vanishes at large r, and this results in the breakdown of the usual Bogomolny proportionality between topological charge and the tension. At the origin, furthermore, some field components diverge like $1/r \ln r$ unlike in the Nielsen–Olesen case. This of course is the singularity needed to circumvent the no-go theorem.

The flux can be found using (19), (26), and (27):

$$\Phi = \frac{\pi}{e}\int_0^\infty r\,dr \frac{1}{r}\frac{d}{dr}\left(r\frac{d\omega}{dr}\right)$$

$$= \frac{\pi}{e} r\omega'\Big|_0^\infty = \frac{2\pi}{e}\frac{ru'}{u}\Big|_0^\infty = -\frac{2\pi}{e}. \tag{30}$$

This is again exactly the Nielsen–Olesen value. The tension, on the other hand, diverges as expected:

$$T = \frac{2\pi a^2}{e^2}\int_0^\infty r\,dr \frac{1}{r}\frac{d}{dr}(r\cosh\omega) = \frac{2\pi a^2}{e^2}[r\omega' \sinh\omega]_0^\infty$$
$$= \frac{8\pi a^2}{e^2}\lim_{r\to 0}\frac{1}{r^2 \ln^2(r/2)}. \tag{31}$$

IV. REAL AND COMPLEX SOLUTIONS BASED ON THE LIOUVILLE EQUATION

A. The general solution

Liouville[21] in 1853 expressed the most general solution to the Minkowski version of (11) in terms of two arbitrary functions $F(x-y)$ and $G(x+y)$. In Euclidean space, one of these becomes an analytic function of the variable $z = x + iy$ and the other a function of the variable z^*. If we also require that the solution be real (this will hold even for our complex Yang–Mills solutions), then ζ must be the following symmetric combination of an analytic function $g(z)$ and its conjugate g^*:

$$\zeta = \frac{1}{2}\ln\left[4\left|\frac{dg}{dz}\right|^2 \frac{1}{(1-g^*g)^2}\right]. \tag{32}$$

It is immediately obvious that the solution is singular along a curve defined by $g^*g = 1$, in contrast to the singularity of P_3 which was limited to the origin. A famous theorem, also by Liouville, states that the singularity cannot be avoided unless g is trivial, i.e., a constant. We can, however, trade a real, singular Yang–Mills solution for a complex, nonsingular one simply by $(h,k) \to (ih, ik)$. Interestingly, gauge-invariant quantities such as the flux and the tension still remain real under this transformation. The Liouville equation now takes the form

$$\nabla^2\zeta = -e^{2\zeta}, \tag{33}$$

to which the solution is

$$\zeta = \frac{1}{2}\ln\left[4\left|\frac{dg}{dz}\right|^2 \frac{1}{(1+g^*g)^2}\right], \tag{34}$$

showing that the singularity has indeed disappeared. Note $g(z)$ and $1/g(z)$ correspond to the same solution ζ. The "magnetic field" $\mathcal{F}_{12}$ is given by

$$\mathcal{F}_{12} = \mathbf{F}_{12}\cdot\hat{\varphi}\times\hat{\psi} = \frac{16a^2}{e^2}\left|\frac{dg}{dz}\right|^2 \frac{1}{(1+g^*g)^2}, \tag{35a}$$

$$\mathcal{F}_{12} = (16a^2/e^2)\partial_z\,\partial_{z^*} \ln(1+g^*g). \tag{35b}$$

Thus $\mathcal{F}_{12}$ is a Kaehler form,[22] obtainable from a Kaehler potential $\ln(1+g^*g)$. The flux Φ is the integral of this form:

$$\Phi = \frac{1}{4a^2}\int dx\,dy\,\mathcal{F}_{12}$$
$$= \frac{16ia^2}{8a^2e}\int dz\wedge dz^*\left|\frac{dg}{dz}\right|^2 \frac{1}{(1+g^*g)^2}$$
$$= \frac{2i}{e}\int \frac{dg\wedge dg^*}{(1+g^*g)^2}. \tag{36}$$

This is recognized to be proportional to the surface area of a sphere of unit diameter projected stereographically onto the complex g-plane.

B. Particular solutions

1. Nonsingular complex axisymmetric solutions

To obtain these, we must clearly set $g = z^n = r^n e^{in\theta}$ so that there is no angular dependence in (34). Furthermore, n must be an integer since otherwise the analyticity requirement fails along the positive x axis of our xy plane and (34) ceases to be a solution there. With this choice of g and the gauge used in (28), we find from (14) and (34)

$$A_r = 0,\ A_\theta = i\sigma_3 a\left(\frac{-2nr^{2n-1}}{(1+r^{2n})} + \frac{n}{r}\right),$$
$$A_3 = i\sigma_\theta a\frac{2nr^{n-1}}{1+r^{2n}},\ A_4 = i\sigma_r a\frac{2nr^{n-1}}{1+r^{2n}}, \tag{37}$$

and

$$F_{12} = -i\sigma_3(8a^2n^2r^{2n-2}/(1+r^{2n})^2). \tag{38}$$

One can now check explicitly from (36), (38), or (21) that the flux Φ is quantized in units of $4\pi/e$:

$$\Phi = -\frac{4\pi n}{e}\int_0^\infty \frac{2nr^{2n-1}\,dr}{(1+r^{2n})^2} = -\frac{4\pi n}{e}. \tag{39}$$

This, of course, agrees with the result found from the circuital integral of A_θ in (37), as long as one is careful to cancel the contribution of the pure gauge n/r term along the circle at infinity by that along a small circle around the origin. The quantization evidently comes from the fact that we cover the $g(z)$ plane (and the spherical area it is projected into) n times as we cover the xy plane once. Note that the basic value of flux for the Liouville case is twice that in the sinh–Gordon one.

The tension density is proportional to $\nabla^2 e^{2\zeta}$ by (20). This latter quantity is easily found to be

$$\nabla^2 e^{2\zeta} = 4n^2\{[(2n-2)^2r^{2n-4} + (8-16n^2)r^{4n-4} + (2n+2)^2r^{6n-4}]/(1+r^{2n})^4\}. \tag{40}$$

This is invariant under $n \to -n$ as expected. Here $\mathcal{F}_{12}$ and the energy density (40) display increasingly sharp extrema approaching $r=1$ with increasing n, and then rapidly diminish to vanishingly small positive values for $r \gtrsim 2$. This accumulation of the tension density and the fields near $r=1$ for large n is reminiscent of the situation with axisymmetric multimonopoles.[23]

2. Real singular axisymmetric solutions

The real versions of the above set of solutions can be obtained by substituting $g = z^n$, this time in (32). This results in solutions which are singular on a ring $r^{2n}=1$ and the flux integral becomes undefined. However, it is interesting that if one wishes to give meaning to the divergent integral

$$I_2 = \int_0^\infty \frac{2nr^{2n-1}}{(-1+r^{2n})^2}\,dr = \int_{-1}^\infty \frac{du}{u^2}, \tag{41}$$

via analytic regularization,[24] one finds

$$I_2 = \left[\int_{-1}^\infty \frac{du}{u^\lambda}\right]_{\lambda=2} = 1, \tag{42}$$

which is again the result (39). More generally, this could be a way of defining singular variants of integrals such as (36) for noncompact Kaehlerian manifolds.

3. Nonaxisymmetric solutions

We have seen that $g = z^n$ gives a solution of flux $4\pi n/e$ centered around the origin. The flux value clearly does not change if one instead chooses

$$g = (z-z_1)(z-z_2)\cdots(z-z_n). \tag{43}$$

This is the case since $g(z)$ still approaches z^n at infinity for finite separations between the string centers $z_k = x_k + iy_k$. Hence at infinity one reobtains the A_θ in (37), whose circuital integral along a very large circle yields $4\pi n/e$.

V. THE ERNST EQUATION AND AXISYMMETRIC STRINGS

We have so far obtained a general n-string solution of (11), but only a single axisymmetric solution of flux $2\pi/e$ of (10a). It would also clearly be desirable to find more general solutions of this latter equation. Leaving aside for the moment the question of how separated n-string configurations satisfying (10a) may be found, we address that of constructing axisymmetric P_3 solutions with a multiple of the basic flux. A successful strategy in the problem of axisymmetric multimonopole solutions employs[23] the solution-generating techniques[25] of the Ernst equation of general relativity by first casting known simple solutions into the Yang–Ernst formalism. Our aim here is to point out the applicability of the method to the string problem by showing how the given solutions fit into the Ernst framework, leaving the actual generation of new ones to a future note.

In the Yang–Ernst formalism we must parametrize our solutions in terms of two functions $F(R,X_3)$ and $G(R,X_3)$ (notice $2aR = r$ and $2aX_3 = z$) obeying the Ernst equations

$$F\nabla^2 F = \nabla F\cdot\nabla F - \nabla G\cdot\nabla G, \tag{44a}$$

$$F\nabla^2 G = 2\nabla F\cdot\nabla G, \tag{44b}$$

which guarantee self-duality. Here, unlike in (10), $\nabla^2 \equiv \partial_1^2 + \partial_2^2 + \partial_3^2$. Potentials with self-dual field tensors are then expressed in terms of solutions of (44) by

$$A_1 = (i/2F)(F_2\sigma_3 + G_1\sigma_2 + G_2\sigma_1), \tag{45a}$$

$$A_2 = (i/2F)(-F_1\sigma_3 + G_2\sigma_2 - G_1\sigma_1), \tag{45b}$$

$$A_3 = (i/2F)G_3\sigma_2, \tag{45c}$$

$$A_4 = (i/2F)(F_3\sigma_3 + G_3\sigma_1). \tag{45d}$$

The question here is what F and G should be in order to recover the solutions (28) and (37), and which, if any, gauge transformations are needed. This has been dealt with explicitly in Ref. 19 regarding P_5; and it is not hard to find the P_3 and the Liouville F's and G's by following this example. For P_3, the result is

$$F = \exp(2aX_3)\operatorname{sech}(\omega(2aR)/2),$$

$$G = \exp(2aX_3)\tanh(\omega(2aR)/2). \tag{46}$$

This is equivalent to (28) up to two gauge transformations: first, one of the form

$$U_1 = \exp(i\sigma_2\tan^{-1}(e^{\omega/2})) \tag{47}$$

followed by $U_2 = \exp(-i\theta\sigma_3/2)$.

Although the general Liouville solution has already been given, we also write down the F and G for the axisymmetric solution (37) for completeness: One simply takes

$$F = \exp\zeta,\quad G = 2aX_3, \tag{48}$$

and no gauge transformation other than U_2 is needed.

VI. CONCLUDING REMARKS

We have so far examined two ways of avoiding the no-go theorem of Sec. II A: allowing singular or complex solutions. Coupled with the two general multistring *Ansätze* based on the sinh–Gordon and Liouville equations, there are thus four classes of solutions to be studied. While both the complex and the real string solutions of the Liouville case

were presented, the only explicit solution of the sinh–Gordon case so far is the real, singular, axisymmetric P_3 one. It would thus be of interest to (i) find a complex axisymmetric and nonsingular version of P_3, (ii) search for axisymmetric solutions of higher flux by generating new solutions of the Ernst system set up in Sec. V, and (iii) look for separated multistrings. We intend to return to these problems in a future publication.

Should solutions of type (iii) exist, the no-go theorem might be beaten in yet another way: Finite tension, real and nonsingular Yang–Mills strings are not forbidden if the covariant derivatives do not go to zero at infinity. This would happen if one had a doubly periodic arrangement of vortices over the whole xy plane. The presence of a length scale (in the form of a lattice spacing) would presumably set a scale for the tension per string (the P_3 vortex might be reobtained by putting the origin at one vortex and letting the spacing as well as the tension go to infinity). Interestingly, this is exactly the picture of the Copenhagen "spaghetti vacuum,"[26] before the spaghetti is "cooked" by quantum fluctuations and Lorentz invariance recovered. Mathematically, such a vacuum configuration would also be described by a doubly periodic function. On the other hand, Painlevé transcendents are known to be asymptotically related to elliptic functions.[17] The appearance of exactly the same P_3 in an intrinsically doubly periodic problem, i.e., the Ising model, is another hint in this direction. In principle, one only needs to find a doubly periodic solution of (10a) and check whether these expectations are borne out.

Regarding the question of integrability the emergence of integrable equations and Painlevé transcendents in both the twice dimensionally reduced SDYM equations considered here, and in the radial equations for a self-dual Yang–Mills monopole,[4] strengthens the likelihood that integrability is an inherent property of the SDYM system. In fact, the particular equations found, i.e., the sinh–Gordon and the Liouville, have to do with a specific integrable system, the so-called Toda field theory, as do the radial equations for a self-dual monopole. As the Toda system of equations also naturally gives rise to the Dodd–Bullough equation,[27] it is very likely that a Yang–Mills *Ansatz* leading to this equation also exists. It is possible that the class of solutions found by Burns[10] are related to the Dodd–Bullough equation.

Finally, there remains an interesting mathematical problem to be investigated. Nahm[28] has recently shown how to adapt the ADHM formalism[29] to the multimonopole problem. It should similarly be possible to incorporate the Painlevé solutions found here into the ADHM framework. This would not only reveal an unexpected connection between the theory of differential equations and the ADHM construction, but could possibly lead to a generalization of the Painlevé transcendents by using groups larger than SU(2).

ACKNOWLEDGMENTS

The greater part of this work was done while in Bonn on an Alexander von Humboldt Stiftung fellowship. I gratefully acknowledge useful conversations with A. Comtet, A. Chakrabarti, T. Eguchi, R. Güven, B. Julia, W. Nahm, R. Stora, and my colleagues in Bonn, whom I thank for their hospitality.

This work is supported in part by the Scientific and Technological Research Council of Turkey (TBTAK).

APPENDIX: THE EUCLIDEAN ERNST EQUATION AND SELF-DUAL RIEMANN TENSOR

The functions F and G in Eq. (44) parametrize a stationary axially symmetric metric[30] of Minkowskian signature:

$$ds^2 = -F(dt-\Omega\, d\theta)^2 + (1/F)[e^{2\gamma}(dR^2+dZ^2)+R^2\,d\theta^2]. \tag{A1}$$

Here G is hidden in the function Ω, related to it through

$$G_R = (F^2/R)\Omega_Z, \quad G_Z = -(F^2/R)\Omega_R. \tag{A2}$$

In going to Euclidean space, the metric (A1) must remain real after $t\to it$; thus we must also have $(\Omega,G)\to(i\Omega,iG)$. The vacuum Einstein equations, to which (44a) and (44b) also belong, now become

$$F\nabla^2 F = \nabla F\cdot\nabla F + \nabla G\cdot\nabla G, \tag{A3a}$$

$$F\nabla^2 G = 2\nabla F\cdot\nabla G, \tag{A3b}$$

$$\nabla^2 2\gamma = (1/F^2)(G_z^2 - F_Z^2), \tag{A4a}$$

$$2\gamma_z = (R/F^2)(F_R F_Z - G_R G_Z), \tag{A4b}$$

$$2\gamma_R = (R/2F^2)(F_R^2 - F_Z^2 - G_R^2 + G_Z^2), \tag{A4c}$$

where (A3) is obviously the Euclideanized Ernst equation.

A remarkable simplification occurs if one looks for solutions with

$$F = G. \tag{A5}$$

Equations (A4b) and (A4c) then imply γ can at most be a constant, which we set to zero for convenience. Furthermore, the single resulting equation

$$F\nabla^2 F = 2\nabla F\cdot\nabla F \tag{A6}$$

simply means that $V = F^{-1}$ is harmonic:

$$\nabla^2 V = 0. \tag{A7}$$

Introducing the vector $\mathbf{\Omega}_H = \Omega\hat{e}_\theta/R$ we can now write the Euclideanized metric with $V^{-1} = F = G$ and $\gamma = 0$ as

$$ds^2 = V^{-1}(dt - \mathbf{\Omega}_H\cdot d\mathbf{x})^2 + V\, d\mathbf{x}\cdot d\mathbf{x}. \tag{A8}$$

On the other hand, using the freedom in the sign of G turns (A2) into

$$\nabla V = \pm\nabla\times\mathbf{\Omega}_H, \tag{A9}$$

which is also consistent with (A7).

The metric (A8), subject to the condition (A9), was first written down by Gibbons and Hawking,[15] who gave a class of metrics with self-dual Riemann tensors. In their work, self-duality is guaranteed by (A9). Among all possible solutions of (A7) and (A9), the special case, where V is a superposition of "Newtonian mass point potentials" and $\mathbf{\Omega}_H$ a superposition of corresponding "magnetic mass vector potentials," gives Taub–NUT or gravitational instanton metrics. But this special restriction is built into our metric: $\mathbf{\Omega}_H$ is already in the azimuthal direction, so that with an axisym-

metric V of the form

$$V = \epsilon + \sum_{i=1}^{N} \frac{m}{[R^2 + (Z - Z_i)^2]^{1/2}}, \tag{A10}$$

Ω_H would indeed be a sum of Dirac monopole potentials. Here $\epsilon = 0$ and $\epsilon = 1$ represent collinear $N - 1$ gravitational instanton and N-center Taub–NUT metrics, respectively. A single m is necessary in order to avoid string singularities.

One might wonder at what point self-duality has been introduced into the standard Ernst framework defined by (A1)–(A4). The answer is in Eq. (A5), which is an extra condition singling out the self-dual solutions to (A3) and (A4). Thus the Minkowski form of the Ernst equations corresponds to self-duality for Yang–Mills fields, while a particularly simple special case of the Euclidean version produces metrics with self-dual Riemann tensors.

[1]L. L. Chau, in *Group Theoretical Methods in Physics*, Istanbul, 1982, edited by M. Serdaroğlu and E. İnönü (Springer-Verlag, Berlin, 1983).
[2]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[3]C. Saçlioğlu. Nucl. Phys. B **178**, 361 (1981).
[4]D. Olive, ICTP/81/82-8; A. Lenzov and M. Saveliev, Lett. Math. Phys. **3**, 207 (1979).
[5]H. Nielsen and P. Olesen, Nucl. Phys. B **61**, 45 (1973).
[6]M. Lohe, Phys. Lett. B **70**, 325 (1977).
[7]J. Richard and A. Rouet, Nucl. Phys. B **183**, 251 (1981) and **185**, 47 (1981); R. Abbott, Durham preprint DTP-82/1, 1982; R. Abbott and W. Zakrzewski, Durham preprint DTP-83/1, 1983; A. Lapedes and E. Mottala, "Complex path integrals and finite temperature"; R. Stora (private communication).
[8]C. N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[9]M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[10]A. D. Burns, Durham preprint DTP-83/2, 1983.
[11]E. Bogomolny, Sov. J. Nucl. Phys. **24**, 449 (1976).
[12]B. Julia (private communication).
[13]T. T. Wu, B. M. McCoy, C. A. Tracy, and E. Barouch, Phys. Rev. B **13**, 316 (1976); B. M. McCoy, C. A. Tracy, and T. T. Wu, J. Math. Phys. **18**, 10 (1977).
[14]F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[15]G. W. Gibbons and S. W. Hawking, Phys. Lett. B **78**, 430 (1978).
[16]P. Painleve, Acta Math. **25**, 1 (1902); B. Gambier, Acta Math. **33**, (1910).
[17]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), p. 345.
[18]B. M. McCoy, C. A. Tracy, and T. T. Wu, Phys. Lett. A **61**, 283 (1977).
[19]P. Lund and T. Regge, Phys. Rev. D **14**, 1524 (1976); H. Morris and R. Dodd, Phys. Lett. A **75**, 249 (1980).
[20]M. Sato, T. Miwa, and M. Jimbo, RIMS (Kyoto), 207 (1976); E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, RIMS (Kyoto), 362 (1981).
[21]J. Liouville, J. Math. Appl. **18**, 71 (1853).
[22]T. Eguchi, P. Gilkey, and A. Hanson, Phys. Rep. **66**, 213 (1980).
[23]P. Forgacs, Z. Horvath, and L. Palla, Phys. Lett. B **109**, 200 (1981).
[24]E. Speer, Ann. Math. Stud. **62**, (1969).
[25]W. Kinnersly and D. Chitre, J. Math. Phys. **18**, 1538 (1977).
[26]P. Olesen, talk in Karpatz, Poland, 1981, NBI-HE-81-5, and references therein.
[27]R. Dodd and R. Bullough, Proc. R. Soc. London, Ser. A **352**, 481 (1977).
[28]W. Nahm, in Ref. 1.
[29]M. Atiyah, N. Hitchin, V. Drinfeld, and Yu. Manin, Phys. Lett. A **65**, 185 (1978).
[30]A. Papapetrou, Ann. Phys. **12**, 309 (1953).

First integrals via polynomial canonical transformations

P. G. L. Leach
Department of Applied Mathematics, University of the Witwatersrand, Johannesburg, 2001, South Africa

(Received 21 July 1983; accepted for publication 6 April 1984)

Maharatna, Dutt, and Chattarji [J. Math. Phys. **20**, 2221 (1979)] discussed the use of time-dependent canonical transformations for the determination of first integrals for time-dependent Hamiltonian systems. One particular proposal that successive time-dependent polynomial canonical transformations will enable first integrals to be found for a wider variety of time-dependent polynomial Hamiltonians than can be obtained using time-dependent linear canonical transformations is shown to be not true for the paradigm which they selected. It is suggested that their ansatz is ill-founded in general.

PACS numbers: 03.20. + i

I. INTRODUCTION

In a paper, which appeared in 1979, Maharatna, Dutt, and Chattarji[1] (hereinafter referred to as MDC) discussed some applications of time-dependent canonical transformations to nonlinear nonconservative systems. Their aim was to determine first integrals for such systems.

In the first instance they applied the method of time-dependent linear canonical transformations[2] to a one-dimensional time-dependent nonlinear Hamiltonian. To facilitate the discussion a Hamiltonian system with the lowest order of nonlinearity (i.e., terms cubic in the canonical variables q and p were added to a quadratic Hamiltonian) was treated. It was shown that only when the time-dependent coefficient in the Hamiltonian had specific forms of time dependence could a first integral be found by this method. Even with this limitation the result was a useful addition to the then state of the art. The same method was applied to a linear damped system and a first integral was obtained without restriction on the nature of the time dependence.

Taking as a paradigm the Hamiltonian of a damped Duffing oscillator, they showed that it was possible to reduce the degree of nonlinearity by means of a time-dependent quadratic canonical transformation. This resulted in the degree of the Hamiltonian being reduced from 4 to 3. The coefficients of the transformed Hamiltonian were still time-dependent.

They then proposed that it may be possible to remove the explicit time dependence of a nonlinear Hamiltonian system by means of a succession of nonlinear canonical transformations. In the particular instance of the Hamiltonian of the Duffing oscillator which is a polynomial in the canonically conjugate variables q and p, the nonlinear canonical transformations were to be quadratic polynomials in q and p. The suggested procedure was to apply a canonical transformation of the form

$$q' = \sum_{j=0}^{2} \sum_{i=0}^{2-j} a_{ij}(t) q^i p^j,$$
$$p' = \sum_{j=0}^{2} \sum_{i=0}^{2-j} b_{ij}(t) q^i p^j, \tag{1.1}$$

where q' and p' are the new variables and q and p the old variables, to a time-dependent polynomial Hamiltonian so that the degree of the transformed Hamiltonian was one less than that of the original transformation. This process was to be repeated until a quadratic Hamiltonian with time-dependent coefficients was obtained. The time-dependence of this Hamiltonian would then be removed by means of a time-dependent linear transformation as has been described elsewhere (Leach[2]).

The calculations for this procedure published in MDC were not complete and further work along their lines does not seem to have appeared in the literature. The aim of this note is to examine two questions raised by this part of the work of MDC. The first is whether a class of time-dependent systems wider than that which can be treated by time-dependent linear transformations can be treated successfully by means of time-dependent polynomial transformations. The second is whether in fact such polynomial transformations exist in a canonical context.

II. STRUCTURE OF THE FIRST INTEGRAL OF MDC

As a paradigm for their method, MDC used the Hamiltonian of the damped Duffing oscillator

$$H(q,p,t) = \tfrac{1}{2}p^2 e^{-2\gamma t} + \tfrac{1}{2}\omega^2 q^2 e^{2\gamma t} - \tfrac{1}{4}\epsilon q^4 e^{2\gamma t}. \tag{2.1}$$

We propose to use a slightly more general paradigm, viz.

$$H(q,p,t) = \tfrac{1}{2}p^2 + \tfrac{1}{2}q^2 + \tfrac{1}{4}\lambda(t) q^4. \tag{2.2}$$

We explain how the Hamiltonian (2.2) is more general below. Given a Hamiltonian

$$H(q,p,t) = \tfrac{1}{2}a(t)p^2 + \tfrac{1}{2}b(t)q^2 + \tfrac{1}{4}c(t)q^4, \tag{2.3}$$

the function $a(t)$ may be removed by means of the change of time variable

$$t \to t' : t' = \int^t a(u)\, du, \tag{2.4}$$

giving an equivalent description of the same system by the Hamiltonian

$$H'(q,p,t') = \tfrac{1}{2}p^2 + \tfrac{1}{2}b'(t')q^2 + \tfrac{1}{4}c'(t')q^4, \tag{2.5}$$

where

$$b'[t'(t)] = b(t)/a(t), \quad c'[t'(t)] = c(t)/a(t). \tag{2.6}$$

The function $b'(t')$ is then removed by the time-dependent linear canonical transformation

 0022-2488/84/113221-06$02.50

$$(q,p)\rightarrow(Q,P: Q=q/\rho,\quad P=\rho p-\dot{\rho}q), \tag{2.7}$$

where $\rho(t')$ (the overdot denotes d/dt') is a solution of the differential equation

$$\ddot{\rho}+b'\rho=\rho^{-3}. \tag{2.8}$$

The transformed Hamiltonian is

$$\overline{H}'(Q,P,t')=\rho^{-2}\{\tfrac{1}{2}P^2+\tfrac{1}{2}Q^2+\tfrac{1}{4}c'\rho^6Q^4\}. \tag{2.9}$$

The further change of time variable

$$t'=\rightarrow T: T=\int^{t'}\rho^{-2}(u)\,du \tag{2.10}$$

gives the equivalent Hamiltonian

$$\overline{H}(Q,P,T)=\tfrac{1}{2}P^2+\tfrac{1}{2}Q^2+\tfrac{1}{4}\Lambda(T)Q^4, \tag{2.11}$$

where

$$\Lambda[T(t')]=c'(t')\rho^6(t'). \tag{2.12}$$

The Hamiltonian (2.11) has the form given in (2.2). The reduction of (2.3) to (2.2) by the method outlined above enables us to use the point transformation of (2.7) rather than a more general linear transformation which would introduce higher powers of the momentum.

The method proposed in MDC was to apply (i) a time-dependent quadratic canonical transformation of the type (1.1) to reduce (2.1) to a time-dependent Hamiltonian cubic in q' and p'; (ii) a second quadratic canonical transformation of the same form, viz.

$$\begin{aligned}Q'&=\sum_{j=0}^{2}\sum_{i=0}^{2-j}c_{ij}(t)q'^ip'^j,\\ P'&=\sum_{j=0}^{2}\sum_{i=0}^{2-j}d_{ij}(t)q'^ip'^j,\end{aligned} \tag{2.13}$$

to reduce the cubic Hamiltonian to a time-dependent quadratic Hamiltonian; and finally (iii) a time-dependent linear canonical transformation

$$\begin{aligned}Q&=\sum_{j=0}^{1}\sum_{i=0}^{1-j}e_{ij}(t)Q''^iP''^j,\\ P&=\sum_{j=0}^{1}\sum_{i=0}^{1-j}f_{ij}(t)Q''^iP''^i,\end{aligned} \tag{2.14}$$

to convert the time-dependent quadratic Hamiltonian to a time-independent form. The resulting Hamiltonian is a first integral of the system it describes and so of the original Hamiltonian system.

The suggestion in MDC is that, by the method outlined above, first integrals may be found for a wider class of time-dependent systems than the class which may be treated by time-dependent linear canonical transformations. Their posited first integral is quadratic in the final set of canonical variables, Q and P, hence quadratic in Q' and P', quartic in q' and p', and so octic in q and p. Before we examine the existence of an octic integral for a Hamiltonain of the type (2.2), we shall determine the conditions under which (2.2) has a first integral in terms of a linear canonical transformation.

III. A FIRST INTEGRAL FOR A HAMILTONIAN OF TYPE (2.2)

First integrals for Hamiltonians of the form

$$H(q,p,t)=\tfrac{1}{2}p^2+V(q,t), \tag{3.1}$$

which are polynomials in p have been treated recently by Gascon, Ramos, and Aguirre-Dabon[3] and Lewis and Leach.[4] The approach was quite direct in that a first integral of the form

$$I(q,p,t)=\sum_{i=0}^{n}x_i(q,t)\,p^i \tag{3.2}$$

was assumed and the Liouville equation

$$\frac{\partial I}{\partial t}+[I,H]_{qp}=0 \tag{3.3}$$

solved.

For the case $n=2$ it was found that when

$$\begin{aligned}V(q,t)=\rho^{-2}\Big(&\frac{1}{2}[(\dot{\rho}q-b\rho^{-1})^2-(\dot{\rho}^2+\rho\ddot{\rho})q^2+2\dot{b}q]\\&+G(q\rho^{-1}+\int b\rho^{-3}\,dt)\Big)+g(t),\end{aligned} \tag{3.4}$$

the first integral was

$$I(q,p,t)=\frac{1}{2}(\rho p-\dot{\rho}q)^2+G\Big(q\rho^{-1}+\int b\rho^{-3}\,dt\Big), \tag{3.5}$$

where $\rho(t)$, $b(t)$, and $g(t)$ are arbitrary functions of time and G is an arbitrary function of its argument subject to obvious restrictions of differentiability.[5] Matching (3.4) to the Hamiltonian (2.2) for which

$$V(q,t)=\tfrac{1}{2}q^2+\tfrac{1}{4}\lambda(t)q^4, \tag{3.6}$$

we find that

$$I(q,p,t)=\tfrac{1}{2}(\rho p-\dot{\rho}q)^2+\tfrac{1}{2}(q\rho^{-1})^2+\tfrac{1}{2}K(q\rho^{-1})^4, \tag{3.7}$$

$$\lambda(t)=K\rho^{-6}(t), \tag{3.8}$$

and that $\rho(t)$ is a solution of the differential equation

$$\ddot{\rho}+\rho=\rho^{-3}, \tag{3.9}$$

i.e,

$$\rho^2(t)=A_0+A_1\sin 2t+A_2\cos 2t, \tag{3.10}$$

where

$$A_0^2-A_1^2-A_2^2=1. \tag{3.11}$$

[Note: It is conventional to make the coefficient of ρ^{-3} in (3.9) as unity. Any nonzero constant will suffice and this could be used to subsume the K of (3.7) and (3.8), but there is no gain in generality. A zero coefficient of ρ^{-3} will cause $\rho(t)$ to have periodic zeros and so limit the domain in time of the problem.]

The first integral (3.7) may obtained from the Hamiltonian (2.2) by means of the linear canonical transformation

$$(q,p)\rightarrow(Q,P: Q=q\rho^{-1},\quad P=\rho p-\dot{\rho}q), \tag{3.12}$$

which produces the transformed Hamiltonian

$$\overline{H}[Q(q,t),P(q,p,t),t]=\rho^{-2}I(q,p,t). \tag{3.13}$$

IV. AN OCTIC FIRST INTEGRAL FOR A HAMILTONIAN OF TYPE (2.2)

For the case $n>3$, the solution of (3.3) cannot be obtained explicitly for a general $V(q,t)$ (see both Refs. 3 and 4). However, in this case for which the form of the potential is

given as a simple polynomial in q [cf. (3.6)], it is feasible to assume the form of the first integral in (3.2) for general n. In particular we may take the first integral to have the form implied in MDC, i.e., one which is octic in q and p by assuming a first integral for the Hamiltonian (2.2) to be

$$I(q,p,t)=\sum_{i=0}^{8}x_i(q,t)\,p^i, \tag{4.1}$$

where $x_i(q,t)$ is a polynomial in q of degree $8-i$ with time-dependent coefficients.

Substitution of (2.2) and (4.1) into the Liouville equation (3.3) and separation by equating coefficients of powers of p to zero gives rise to the ten equations

$$\begin{aligned}&\frac{\partial x_8}{\partial q}=0,\\&\frac{\partial x_8}{\partial t}+\frac{\partial x_7}{\partial q}=0,\\&\frac{\partial x_i}{\partial t}+\frac{\partial x_{i-1}}{\partial q}=(i+1)x_{i+1}(q-\lambda q^3),\quad i=7,\ldots,1,\\&\frac{\partial x_0}{\partial t}=x_1(q+\lambda q^3).\end{aligned} \tag{4.2}$$

Given that each $x_i(q,t)$ is a polynomial in q of degree $8-i$ we define

$$\begin{aligned}&x_0(q,t)=\sum_{j=0}^{8}a_j(t)q^j,\quad x_1(q,t)=\sum_{j=0}^{7}b_j(t)q^j,\\&x_2(q,t)=\sum_{j=0}^{6}c_j(t)q^j,\quad x_3(q,t)=\sum_{j=0}^{5}d_j(t)q^j,\\&x_4(q,t)=\sum_{j=0}^{4}e_j(t)q^j,\quad x_5(q,t)=\sum_{j=0}^{3}f_j(t)q^j,\\&x_6(q,t)=\sum_{j=0}^{2}g_j(t)q^j,\quad x_7(q,t)=\sum_{j=0}^{1}h_j(t)q^j,\\&x_8(q,t)=k_0(t).\end{aligned} \tag{4.3}$$

The expressions in (4.3) are then substituted in Eqs. (4.2) and separated by equating coefficients of powers of q to zero. the resulting equations simplify considerably. For those functions of time which are not identically zero (with one exception) the following set of equations remains to be solved:

$$\dot e_0+d_1=0, \tag{4.4a}$$
$$\dot d_1+2c_2=4e_0, \tag{4.4b}$$
$$\dot c_0+b_1=0, \tag{4.4c}$$
$$\dot c_2+3b_3=3d_1, \tag{4.4d}$$
$$\dot c_4+5b_5+3\lambda d_1, \tag{4.4e}$$
$$\dot b_1+2a_2=2c_0, \tag{4.4f}$$
$$\dot b_3+4a_4=2c_2+2\lambda c_0, \tag{4.4g}$$
$$\dot b_5+6a_6=2c_4+2\lambda c_2, \tag{4.4h}$$
$$\dot a_2=b_1, \tag{4.4i}$$
$$\dot a_4=b_3+\lambda b_1, \tag{4.4j}$$
$$\dot a_6=b_5+\lambda b_3, \tag{4.4k}$$
$$\dot a_8=\lambda b_5, \tag{4.4l}$$
$$c_4=\lambda e_0, \tag{4.4m}$$
$$4a_8=\lambda c_4. \tag{4.4n}$$

The only function which is identically zero not included in (4.4) is $a_0(t)$ which is an arbitrary constant A. As I and $I+A$ have the same meaning, it is of no import.

Using Eqs. (4.4a), (4.4b), (4.4d), (4.4e), (4.4m), and (4.4n) we may express d_1, c_2, b_3, c_4, and a_8 in terms of e_0 and λ and their derivatives. Equation (4.4l) becomes a consistency condition on λ and e_0 and reduces to

$$2\dot\lambda e_0+3\lambda\dot e_0=0. \tag{4.5}$$

Integrating (4.5),

$$\lambda^2 e_0^3=\text{const.} \tag{4.6}$$

If we let

$$\lambda(t)=\rho^{-6}(t), \tag{4.7}$$

we then have by back substitution

$$e_0=\alpha\rho^4, \tag{4.8a}$$
$$d_1=-4\alpha\rho^3\dot\rho, \tag{4.8b}$$
$$c_2=2\alpha(\rho^4+2\rho^2\dot\rho^2+\rho^3\ddot\rho), \tag{4.8c}$$
$$b_3=-2\alpha/3(10\rho^3\dot\rho+6\rho\dot\rho^3+9\rho^2\dot\rho\ddot\rho+\rho^3\dddot\rho), \tag{4.8d}$$
$$b_5=-2\alpha\rho^{-3}\dot\rho, \tag{4.8e}$$
$$c_4=\alpha\rho^{-2}, \tag{4.8f}$$
$$a_8=\tfrac14\alpha\rho^{-8}, \tag{4.8g}$$

where α is an arbitrary constant. Returning to Eq. (4.4), (4.4h) yields

$$a_6=\alpha(\rho^{-2}+\rho^{-4}\dot\rho^2+\rho^{-3}\ddot\rho), \tag{4.9}$$

and (4.4k) becomes a consistency condition on $\rho(t)$, viz.,

$$\rho^3\dddot\rho+3\rho^2\dot\rho\ddot\rho+4\rho^3\dot\rho=0. \tag{4.10}$$

Integration of (4.10) gives

$$\ddot\rho+\rho=\rho^{-3}, \tag{4.11}$$

in which the constant of integration has been taken as unity.

Equations (4.4c), (4.4f), and (4.4i) may be solved independently of the others to yield

$$\begin{aligned}&c_0=B_0-B_1\sin 2t+B_2\cos 2t,\\&b_1=-2B_1\cos 2t+2B_2\sin 2t,\\&a_2=B_0-B_1\sin 2t-B_2\cos 2t.\end{aligned} \tag{4.12}$$

From (4.4g) and using (4.8d) and (4.11) we find that

$$\begin{aligned}a_4=&\tfrac12\rho^{-6}(B_0+B_1\sin 2t+B_2\cos 2t)\\&+\alpha(\dot\rho^2+\rho^{-2})^2.\end{aligned} \tag{4.13}$$

Finally (4.4) produces the consistency requirement

$$\begin{aligned}&\rho^2(-2B_1\cos 2t+2B_2\sin 2t)\\&+2\rho\dot\rho(B_0+B_1\sin 2t+B_2\cos 2t)=0,\end{aligned} \tag{4.14}$$

from which it follows that

$$B_0+B_1\sin 2t+B_2\cos 2t=\beta\rho^2, \tag{4.15}$$

where β is an arbitrary constant. This is consistent with (4.11) which has the solution

$$\rho^2(t)=A_0+A_1\sin 2t+A_2\cos 2t, \tag{4.16}$$

where

$$A_0^2 - A_1^2 - A_2^2 = 1 . \tag{4.17}$$

We may now write down the first integral by substituting the various coefficients which have been determined above into (4.3) and thence (4.2). In doing this any second and third derivatives of ρ are replaced by expressions in ρ and $\dot\rho$ by using (4.11). We find that

$$I(q,p,t) = \beta I_1 + \alpha I_1^2 , \tag{4.18}$$

where

$$I_1(q,p,t) = (\rho p - \dot\rho q)^2 + (q\rho^{-1})^2 + \tfrac{1}{2}(q\rho^{-1})^4, \tag{4.19}$$

is a first integral for

$$H(q,p,t) = \tfrac{1}{2}p^2 + \tfrac{1}{2}q^2 + \tfrac{1}{4}q^4\rho^{-6}. \tag{4.20}$$

This is essentially the first integral given by (3.7), i.e., the extension of the p dependence of the first integral to higher powers has not given rise to a first integral other than the one already calculated nor has it extended the nature of the permitted time dependence in the potential beyond that which was already permitted.

V. THE EXISTENCE OF A QUARTIC CANONICAL TRANSFORMATION

For a paradigm of the type in MDC we have seen that there has been no advance over the result which would have been obtained using only a time-dependent linear canoncial transformation. In this case, the first question raised in the Introduction must be answered in the negative. However, it has been possible to obtain a first integral which is quadratic in a variable I_1, which is quartic polynomial in q and p.

It is possible then that the second question can be answered in the affirmative? Let us recall our argument of Sec. II. The combination of two successive quadratic transformations followed by a linear transformation is to produce a first integral quadratic in the new variables and so octic in the original variables. The composition of the series of three transformations is a transformation quartic in the original variables. So the argument was that a quartic transformation implies an octic first integral. However, the existence of an octic first integral, which was demonstrated in Sec. III for a limited type of time dependence in the potential, does not necessarily imply the existence of the proposed quartic transformation. We now turn to an examination of the second question.

It has been suggested in MDC that a Hamiltonian of the form

$$H(q,p,t) = \tfrac{1}{2}p^2 + \tfrac{1}{2}q^2 + \tfrac{1}{4}\lambda q^4 \tag{5.1}$$

can be transformed by a canonical transformation of the type

$$\begin{aligned} Q(q,p,t) &= a + bp + cp^2 + dp^3 + ep^4, \\ P(q,p,t) &= f + gp + hp^2 + kp^3 + mp^4, \end{aligned} \tag{5.2}$$

in which the coefficients a to m are polynomials in q of degree $4-i$, where i is the power of p multiplying the coefficient, to a Hamiltonian which is quadratic in Q and P. We are at liberty to set the form of the quadratic. The reason for this is as follows. Suppose that a quartic transformation to some quadratic Hamiltonian exists. This quadratic Hamiltonian may then be transformed by a linear canonical transformation to any other quadratic Hamiltonian as was shown in Ref. 2. As the composition of a quartic and a linear transformation is itself quartic, it follows that we can select the final quadratic Hamiltonian at will. We select

$$\bar H = \tfrac{1}{2}P^2. \tag{5.3}$$

We could attempt to match these Hamiltonians [(5.1) and (5.3)] by using the generating function approach in which for a generating function of the form $F_0(q,p,t)$,[6]

$$\begin{aligned} \frac{\partial F_0}{\partial q} &= p - P\frac{\partial Q}{\partial q}, \\ \frac{\partial F_0}{\partial p} &= -P\frac{\partial Q}{\partial p}, \\ \bar H &= H + P\frac{\partial Q}{\partial t} + \frac{\partial F_0}{\partial t}, \end{aligned} \tag{5.4}$$

where all functions are expressed in terms of q, p, and t. However, this produces equations for the coefficients of p in $Q(q,p,t)$ and $P(q,p,t)$ which are nonlinear. To avoid such equations we adopt a different procedure here. We require (cf. Refs. 1 and 2) that the equations of motion derived from H and $\bar H$ be the same when expressed in terms of q and p and then impose the canonical requirement

$$[Q,P]_{qp} = 1. \tag{5.5}$$

Hamilton's equations for H and $\bar H$ are

$$\dot Q = P, \quad \dot q = p, \quad \dot P = 0, \quad \dot p = -(q + \lambda q^3). \tag{5.6}$$

Substituting for Q and P from (5.2) and using the expressions for $\dot q$ and $\dot p$ in (5.6), we separate the equations $\dot Q = P$ and $\dot P = 0$ by equating coefficients of powers of p to zero to obtain

$$0 = \frac{\partial e}{\partial q}, \tag{5.7a}$$

$$m = \frac{\partial d}{\partial q} + \frac{\partial e}{\partial t}, \tag{5.7b}$$

$$k = \frac{\partial c}{\partial q} + \frac{\partial d}{\partial t} - 4e(q + \lambda q^3), \tag{5.7c}$$

$$h = \frac{\partial b}{\partial q} + \frac{\partial c}{\partial t} - 3d(q + \lambda q^3), \tag{5.7d}$$

$$g = \frac{\partial a}{\partial q} + \frac{\partial b}{\partial t} - 2c(q + \lambda q^3), \tag{5.7e}$$

$$f = \frac{\partial a}{\partial t} - b(q + \lambda q^3), \tag{5.7f}$$

and

$$0 = \frac{\partial m}{\partial q}, \tag{5.8a}$$

$$0 = \frac{\partial k}{\partial q} + \frac{\partial m}{\partial t}, \tag{5.8b}$$

$$0 = \frac{\partial h}{\partial q} + \frac{\partial k}{\partial t} - 4m(q + \lambda q^3), \tag{5.8c}$$

$$0 = \frac{\partial g}{\partial q} + \frac{\partial h}{\partial t} - 3k(q + \lambda q^3), \tag{5.8d}$$

$$0 = \frac{\partial f}{\partial q} + \frac{\partial g}{\partial t} - 2h(q + \lambda q^3), \tag{5.8e}$$

$$0 = \frac{\partial f}{\partial t} - g(q + \lambda q^3), \tag{5.8f}$$

respectively. [Note: The equations (5.7a) and (5.8a) were anticipated from the definition of the transformation in (5.2).] The first of (5.7) and (5.8) give

$$e = \alpha, \quad m = \beta, \tag{5.9}$$

where α and β are arbitrary functions of time, subject to any later differentiability requirements, as will be any other Greek letter below. From (5.7b), (5.8b), and (5.9),

$$d = (\beta - \dot{\alpha})\, q + \gamma, \quad k = -\dot{\beta} q + \delta. \tag{5.10}$$

The next pair of equations becomes

$$\frac{\partial c}{\partial q} = 4\alpha(q + \lambda q^3) - (\dot{\beta} - \ddot{\alpha})\, q - \dot{\gamma} - \dot{\beta} q + \delta,$$
$$\frac{\partial h}{\partial q} = 4\beta(q + \lambda q^3) + \ddot{\beta} q - \dot{\delta}. \tag{5.11}$$

Since c and h are to be at most quadratic in q and $\lambda \neq 0$,

$$\alpha = 0, \quad \beta = 0, \tag{5.12}$$

so that

$$c = (\delta - \dot{\gamma}) q + \epsilon, \quad h = -\dot{\delta} q + \sigma. \tag{5.13}$$

The next pair [(5.7d) and (5.8d)] becomes

$$\frac{\partial b}{\partial q} = 3\gamma(q + \lambda q^3) - (\dot{\delta} - \ddot{\gamma})\, q - \dot{\epsilon} - \dot{\delta} q + \sigma,$$
$$\frac{\partial g}{\partial q} = 3\delta(q + \lambda q^3) + \ddot{\delta} q - \dot{\sigma}. \tag{5.14}$$

As b and g are at most cubic in q,

$$\gamma = 0, \quad \delta = 0, \tag{5.15}$$

and so

$$b = (\sigma - \dot{\epsilon})\, q + \mu, \quad g = -\dot{\sigma} q + \eta. \tag{5.16}$$

Integrating (5.7e) and (5.8e) we obtain

$$a = \tfrac{1}{2}\epsilon\lambda q^4 + \tfrac{1}{2}(2\epsilon - 2\dot{\sigma} + \ddot{\epsilon})\, q^2 + (\eta - \dot{\mu})\, q + \nu,$$
$$f = \tfrac{1}{2}\sigma\lambda q^4 + \tfrac{1}{2}(2\sigma - \ddot{\sigma})\, q^2 - \dot{\eta} q + \xi. \tag{5.17}$$

The remaining equations (5.7f) and (5.8f) impose consistency requirements. Separating them by equating coefficients of powers of q to zero we have from (5.7f)

$$3\dot{\epsilon}\lambda + \epsilon\dot{\lambda} = 3\sigma\lambda, \tag{5.18a}$$
$$\mu\lambda = 0, \tag{5.18b}$$
$$\dddot{\epsilon} + 4\dot{\epsilon} = 4\sigma + 3\ddot{\sigma}, \tag{5.18c}$$
$$\ddot{\mu} + \mu = 2\dot{\eta}, \tag{5.18d}$$
$$\xi = \dot{\nu}, \tag{5.18e}$$

and from (5.8f)

$$3\dot{\sigma}\lambda + \sigma\dot{\lambda} = 0, \tag{5.19a}$$
$$\eta\lambda = 0, \tag{5.19b}$$
$$\ddot{\sigma} + 4\dot{\sigma} = 0, \tag{5.19c}$$
$$\ddot{\eta} + \eta = 0, \tag{5.19d}$$
$$\dot{\xi} = 0. \tag{5.19e}$$

Since $\lambda \neq 0$, (5.18b) and (5.19b) give

$$\eta = 0, \quad \mu = 0, \tag{5.20}$$

and (5.18d) and (5.19d) are satisfied identically. From (5.19) we obtain

$$\sigma(t) = A_0 + A_1 \sin 2t + A_2 \cos 2t. \tag{5.21}$$

From (5.19a) and (5.18a)

$$\lambda = B\sigma^{-3}, \quad \epsilon = (t + D)\sigma, \tag{5.22}$$

respectively, and from (5.19e) and (5.18e)

$$\xi = C, \quad \nu = Ct + E, \tag{5.23}$$

where A through E are constants. The one remaining equation (5.18c) is satisfied identically.

The coefficients in (5.2) are

$$e = 0, \quad d = 0, \quad c = (t + D)\sigma, \quad b = -(t + D)\dot{\sigma} q,$$
$$a = (t + D)\{\tfrac{1}{2}\sigma\lambda q^4 + \tfrac{1}{2}(2\sigma + \ddot{\sigma})\, q^2 + C\} + E - CD,$$
$$m = 0, \quad k = 0, \quad h = \sigma, \quad g = -\dot{\sigma} q,$$
$$f = \tfrac{1}{2}\sigma\lambda q^4 + \tfrac{1}{2}(2\sigma + \ddot{\sigma}) q^2 + C, \tag{5.24}$$

from which it is evident that

$$Q(q, p, t) = (t + D)\, P(q, p, t) + E - CD. \tag{5.25}$$

Thus Q and P are not canonically conjugate and so the second question raised in the context of the work in MDC also must be answered in the negative.

VI. DISCUSSION

In one sense the results obtained above are negative. For a polynomial Hamltonian of type (2.2) there has been no advance in the nature of the permitted time dependence in the potential over and above that which can be treated using linear canonical transformations. Increasing the degree in p and q of the polynomial first integral in the context of the proposal in MDC has not yet yielded any further results, just a lot of hard work. The assumption that polynomial canonical transformations, which are of equal degree (greater than 1), exist also has been shown to be false in this context.

Indeed it was more from a wish to follow the spirit of the work of Maharatna, Dutt, and Chattarji that the calculations in Sec. V were made. We would not expect a result such as they proposed on more fundamental grounds. We may illustrate this point easily. We may without loss of generality [as discussed in Sec. V in the lines preceeding (5.3)] select as P the invariant I_1 (4.19) obtained in Sec. IV, i.e., take

$$P(q, p, t) = (\rho p - \dot{\rho} q)^2 + (q\rho^{-1}) + \tfrac{1}{2}(q\rho^{-1})^4. \tag{6.1}$$

To simplify the following discussion we interpose the linear canonical transformation

$$q' = q\rho^{-1}, \quad p' = \rho p - \dot{\rho} q, \tag{6.2}$$

so that

$$P(q', p', t) = p'^2 + q'^2 + \tfrac{1}{2} q'^4. \tag{6.3}$$

A $Q(q', p', t)$ canonically conjugate to $P(q', p', t)$ satisfies the linear partial differential equation

$$\frac{\partial Q}{\partial q'}\frac{\partial P}{\partial p'} - \frac{\partial Q}{\partial p'}\frac{\partial P}{\partial q'} = 1, \tag{6.4}$$

the characteristics for which are found from

$$\frac{dq'}{\partial P/\partial p'} = \frac{-dp'}{\partial P/\partial q'} = \frac{dQ}{1}. \tag{6.5}$$

The two characteristics are

$$u = P, \quad v = Q - \int^{q'} \frac{\partial P}{\partial p'}(u,\eta,t)^{-1}\, d\eta, \tag{6.6}$$

where in the integrand p' has been replaced by a function of u, q', and t by inversion of the first characteristic. With P as in (6.3), to within an additive function of P,

$$Q = \int^{q'} \tfrac{1}{2}(u - \eta^2 - \tfrac{1}{2}\eta^4)^{-1}\, d\eta, \tag{6.7}$$

which is an elliptic integral of the first kind. [Alternatively we could take as characteristics

$$u = P, \quad v = Q + \int^{P'} \frac{\partial P}{\partial q'}(u,\eta,t)^{-1}\, d\eta, \tag{6.6'}$$

for which the analog to (6.7) is

$$Q = -\int^{p'} \tfrac{1}{2}\{-1 + \sqrt{[1 + 2P - 2\eta^2]}\}^{-1/2} \times \{1 + 2P - 2\eta^2\}^{-1/2}\, d\eta. \tag{6.7'}$$

This also is an elliptic integral of the first kind.] Consequently it is not surprising tht $Q(q', p', t)$ is not a quartic polynomial in q' and p' and hence $Q(q, p, t)$ not a quartic polynomial in q and p.

We must emphasize the restriction that both P and Q be polynomials. Sense can be made of the proposal in MDC if we take P, say, to be a polynomial and allow Q to be an infinite series. From (6.7), for example, we see that

$$Q(q', p', t) = \frac{1}{2\sqrt{P}} \sum_{i=0}^{\infty} \frac{(-1)^i}{P^i} \binom{-\frac{1}{2}}{i} \times \int^{q'} (\eta^2 + \tfrac{1}{2}\eta^4)^i\, d\eta, \tag{6.8}$$

where $\binom{-\frac{1}{2}}{i}$ is the binomial coefficient. However, this seems to be a cumbersome approach to a problem which can be solved more easily by other methods. When applied to a Hamiltonian with a general potential $V(q,t)$ it is difficult to see how satisfactory progress could be made.

We may conclude on a more positive note. It would appear that the results pertaining to the permitted nature of the time dependence in a potential $V(q,t)$ are going to be the same whether a linear canonical transformation is used or something more complicated as suggested above in the combination (6.3) and (6.8). If we wish to expand the nature of the permitted time dependence, some other approach is necessary. Two such approaches are found in Refs. 7 and 8.

[1] A. Maharatna, R. Dutt, and D. Chattarji, J. Math. Phys. **20**, 2221 (1979).
[2] P. G. L. Leach, J. Math. Phys. **18**, 1608 (1977).
[3] F. G. Gascon, F. B. Ramos, and E. Aguirre-Dabon, J. Math. Phys. **23**, 2281 (1982).
[4] H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982).
[5] The result may be quoted in a variety of ways. This form is given in P. G. L. Leach, Phys. Lett. A **84**, 161 (1981).
[6] H. R. Lewis and P.G.L. Leach, J. Math. Phys. **23**, 165 (1982).
[7] H. R. Lewis and P. G. L. Leach, Los Alamos National Laboratory preprint LA-UR-82-1417.
[8] P. G. L. Leach, H. R. Lewis, and W. Sarlet, J. Math. Phys. **25**, 486 (1984).

Factorization of the 2×2 matrix recursion operator of the coupled KdV equation

Raju N. Aiyer
Laser Division, Bhabha Atomic Research Centre, Bombay 400 085, India

(Received 23 February 1984; accepted for publication 20 April 1984)

The recursion operator for the infinitesimal transformations about solutions of the coupled KdV equation is a 2×2 matrix whose elements are of the fourth order. This formidable looking operator is written as the product of four 2×2 matrix operators whose elements are of the first order. Auxiliary functions introduced to factorize the recursion operator lead to the scattering equation for the equation. The factorization of the recursion operator for the sine–Gordon equation is also presented.

PACS numbers: 03.40.Kf, 02.90. + p

I. INTRODUCTION

The existence of a recursion operator which generates an infinity of infinitesimal transformations (IT) about an arbitrary solution of a nonlinear evolution equation (NLEE)

$$u_t(x,t) = K(u) \tag{1.1}$$

is a necessary condition for the integrability of (1.1).[1,2] A simple derivation of the recursion operator is also presented in these papers. Many properties of these operators have been studied in Refs. 3–5. The eigenfunctions of these operators have been shown, in many cases, to form a complete set and used to study near-integrable systems by perturbation.[6–9] The soliton solutions can be written as a linear combination of the discrete eigenfunctions of the recursion operators.[3,8] Finally the recursion operator and the operator determining the time evolution of the IT form a Lax pair.[1–3] These properties of the recursion operator suggests their study in their own right. However, more often, the recursion operators are more complicated than the scattering operator which, on the other hand, might be more difficult to obtain. Simplification of the recursion operators might therefore prove useful.

In this paper we simplify the 2×2 matrix recursion operators for the coupled KdV equation and the sine–Gordon (SG) equation by factorizing them. Factorization of operators associated with integrable NLEE was first done by Fordy and Gibbons.[10,11] They factorized the third-order scattering operators associated with two fifth-order KdV-like equations and were able to relate the equations to a single modified equation.

The recursion operator for the coupled-KdV equation[12–14]

$$\begin{aligned} u_t &= \tfrac{1}{2}u_{3x} + 3uu_x - 6\phi\phi_x, \\ \phi_t &= -\phi_{3x} - 3u\phi_x \end{aligned} \tag{1.2}$$

is a 2×2 matrix whose elements are fourth-order operators.[15] We show that this operator can be written as a product of four 2×2 matrices whose elements are of the first order. New dependent variables introduced to factorize the recursion operator lead to the scattering operator for (1.2).[13,14] We also show that the inverse of the fourth-order recursion operator cannot be obtained in a closed form.

In Sec. II we obtain the factors. In Sec. III we discuss the inverse. In Sec. IV the scattering operators are derived. In Sec. V the factors of the scalar recursion operator of the SG equation are derived and the inverse of the recursion operator obtained.

II. FACTORIZATION OF THE RECURSION OPERATOR OF THE COUPLED KdV EQUATION

The recursion operator $T(u,\phi)$ is of the form[15]

$$T(u,\phi) = (T_{ij}(u,\phi)), \quad i,j = 1,2, \tag{2.1}$$

where

$$\begin{aligned} T_{11}(u,\phi) &= \tfrac{1}{4}D^4 + 2uD^2 + 3u_xD + 2u_{xx} \\ &\quad + \tfrac{1}{2}u_{3x}D^{-1} + 4u^2 + 3uu_xD^{-1} \\ &\quad + u_xD^{-1}u - 4\phi^2 - 6\phi\phi_xD^{-1}, \\ T_{12}(u,\phi) &= -4\phi_xD - \phi_{xx} - 5\phi D^2 - 2u_xD^{-1}\phi - 4u\phi, \\ T_{21}(u,\phi) &= -\phi_{3x}D^{-1} - 3\phi_{xx} - \tfrac{5}{2}\phi_xD \\ &\quad - 3\phi_xuD^{-1} + \phi_xD^{-1}u, \\ T_{22}(u,\phi) &= -D^4 - 4\phi^2 - 2\phi_xD^{-1}\phi - 4uD^2 - 2u_xD, \end{aligned} \tag{2.2}$$

where $D \equiv \partial/\partial x$ and D^{-1} is the inverse operator $\int_{-\infty}^{x} dx_1$.

We first write $T(u,\phi)$ as a product of two second-order operators because one could make a reasonable guess for the form of these second-operator operators:

$$T(u,\phi) = T_1(u,\phi)\,T_2(u,\phi),$$

where

$$T_1(u,\phi) = (t_{ij}(u,\phi)), \quad i,j = 1,2,$$

and

$$\begin{aligned} t_{ij}(u,\phi) &= a_{ij}D^2 + b_{ij}u + c_{ij}u_xD^{-1} \\ &\quad + d_{ij}\phi_xD^{-1} + e_{ij}\phi, \quad i,j = 1,2. \end{aligned} \tag{2.3}$$

The a_{ij}, b_{ij}, etc., are constants to be determined, and $T_2(u,\phi)$ is a matrix of the same form with different constants. The forms for $T_1(u,\phi)$ and $T_2(u,\phi)$ are chosen so that acting on $(u_x,\phi_x)^t$ they will give the most general third-order terms of the form that appear on the rhs of (1.2). Next we find the product $T_1(u,\phi)T_2(u,\phi)$ and compare the coefficients with (2.2) to find the constants in (2.3) and in $T_2(u,\phi)$. The large number of equations are not difficult to solve and we get

 0022-2488/84/113227-04$02.50

$$T_1(u,\phi)$$
$$= \begin{pmatrix} mD^2 + 4mu + 2mu_x D^{-1} & 4n\phi + 2n\phi_x D^{-1} \\ 4m\phi + 2m\phi_x D^{-1} & nD^2 + 4nu + 2nu_x D^{-1} \end{pmatrix}, \tag{2.4a}$$

$$T_2(u,\phi) = \begin{pmatrix} \frac{1}{4m} D^2 + \frac{1}{m} u + \frac{1}{2m} u_x D^{-1} & -\frac{1}{m}\phi \\ -\frac{1}{n}\phi - \frac{1}{n}\phi_x D^{-1} & -\frac{1}{n} D^2 \end{pmatrix}, \tag{2.4b}$$

where m and n are arbitrary nonzero constants. The product $T_1 T_2$ is independent of m,n. It is interesting to note that with $m = \frac{1}{2}$, $n = -1$,

$$T_1(u,\phi)\begin{pmatrix} u_x \\ \phi_x \end{pmatrix}$$

is just the rhs of (1.2). This means that, though there is no recursion operator connecting (u_x,ϕ_x) and (u_t,ϕ_t) [(u_t,ϕ_t) is the rhs of (1.2)], the operator connecting these two is a factor of the recursion operator. We have observed a similar result for the fifth-order Sawada–Kotera equation

$$u_t = u_{5x} + 5uu_{3x} + 5u_x u_{xx} + 5u^2 u_x$$

whose sixth-order recursion operator[16] does not connect u_x and u_t but the fourth-order operator connecting them is a factor of the recursion operator.

Now $T_1(u,\phi)$ and $T_2(u,\phi)$ can be written as products of first-order operators. For $T_1(u,\phi)$ the factors are easily obtained. Consider the following two operators:

$$T_\pm = D^2 + 4(u \pm \phi) + 2(u_x \pm \phi_x)D^{-1}. \tag{2.5}$$

Now $T_1(u,\phi)$ with $m = n = 1$ can be written as (this choice of m,n is not necessary, i.e., one can proceed with arbitrary m,n)

$$T_1(u,\phi) = \frac{1}{2}\begin{pmatrix} T_+ + T_- & T_+ - T_- \\ T_+ - T_- & T_+ + T_- \end{pmatrix}$$
$$= \frac{1}{2}\begin{pmatrix} D + 2f_x/f & D + 2g_x/g \\ D + 2f_x/f & -(D + 2g_x/g) \end{pmatrix}$$
$$\times \begin{pmatrix} D(D - 2f_x/f)D^{-1} & D(D - 2f_x/f)D^{-1} \\ D(D - 2g_x/g)D^{-1} & -D(D - 2g_x/g)D^{-1} \end{pmatrix}. \tag{2.6}$$

In writing (2.6) we have made use of the result

$$T_+ = (D + 2f_x/f)\,D\,(D - 2f_x/f)\,D^{-1},$$
$$T_- = (D + 2g_x/g)\,D\,(D - 2g_x/g)\,D^{-1}, \tag{2.7}$$

where

$$u + \phi = -f_{xx}/f, \tag{2.8a}$$
$$u - \phi = -g_{xx}/g. \tag{2.8b}$$

To factorize $T_2(u,\phi)$ (we put $m = n = 1$), let

$$T_2(u,\phi) = \begin{pmatrix} a_{11}D + f_{11} & a_{12}D + f_{12} \\ D(a_{21}D + f_{21})D^{-1} & D(a_{22}D + f_{22})D^{-1} \end{pmatrix}$$
$$\times \begin{pmatrix} D(b_{11}D + g_{11})D^{-1} & b_{12}D + g_{12} \\ D(b_{21}D + g_{21})D^{-1} & b_{22}D + g_{22} \end{pmatrix}. \tag{2.9}$$

Constants a_{ij} and b_{ij} and functions $f_{ij}(x,t)$ and $g_{ij}(x,t)$ are to be determined. This form for the product matrices was suggested by the following observations. The (1,2) element of $T_2(u,\phi)$ is $-\phi(x,t)$. This can be written as

$$-\phi(x,t) = -\tfrac{1}{2}\{D^2 + u + \phi - D^2 - (u - \phi)\}$$
$$= -\tfrac{1}{2}\{(D + f_x/f)(D - f_x/f)$$
$$-(D + g_x/g)(D - g_x/g)\}. \tag{2.10}$$

The (2,1) element of $T_2(u,\phi)$ is

$$-(\phi(x,t) + \phi_x(x,t)\,D^{-1}) = -D\phi(x,t)\,D^{-1}, \tag{2.11}$$

that is, obtained by a similarity transformation of (2.10). The (1,1) element of $T_2(u,\phi)$ is the same as the (1,1) element of $T_1(u,\phi)$ and can be written as the sum of T_+ and T_-. The factors of these are given in (2.7). Thus from (2.7), (2.10), and (2.11) the forms of the products are fixed and these lead to the form of the elements in (2.9). Taking the product (2.9) and comparing it with (2.4b) we got two sets of factors for $T_2(u,\phi)$. The first factor is

$$T_2(u,\phi)$$
$$= \frac{1}{4}\begin{pmatrix} D - \frac{3}{2}F - \frac{1}{2}G & D - \frac{1}{2}F - \frac{3}{2}G \\ D(D + 3F - 3G)D^{-1} & -D(D - F + G)D^{-1} \end{pmatrix}$$
$$\times \begin{pmatrix} \frac{1}{2}D(D - F + 3G)D^{-1} & -2D + F - G \\ \frac{1}{2}D(D + 3F - G)D^{-1} & 2D - 3F + 3G \end{pmatrix}, \tag{2.12}$$

where

$$F_x - (2F - G)G = u + \phi = -f_{xx}/f, \tag{2.13a}$$
$$G_x - G^2 = u - \phi = -g_{xx}/g. \tag{2.13b}$$

This gives $G = -g_x/g$. We obtain F by solving (2.13a):

$$F = -\left(\frac{1}{g^2}\right)\int_{-\infty}^{x}\left\{\left(\frac{g^2 f_{x_1 x_1}}{f}\right) + g_{x_1}^2\right\} dx_1, \tag{2.14}$$

where $F(x,t)$ has a rather complex form.

The other factor is

$$T_2(u,\phi)$$
$$= \frac{1}{4}\begin{pmatrix} D - \frac{3}{2}\overline{G} - \frac{1}{2}\overline{F} & D - \frac{1}{2}\overline{G} - \frac{3}{2}\overline{F} \\ D(D + \overline{G} - \overline{F})D^{-1} & -D(D - 3\overline{G} + 3\overline{F})D^{-1} \end{pmatrix}$$
$$\times \begin{pmatrix} \frac{1}{2}D(D - \overline{G} + 3\overline{F})D^{-1} & -2D - 3\overline{G} + 3\overline{F} \\ \frac{1}{2}D(D + 3\overline{G} - \overline{F})D^{-1} & 2D + \overline{G} - \overline{F} \end{pmatrix}, \tag{2.15}$$

where

$$\overline{G}_x - \overline{G}^2 = u + \phi = -f_{xx}/f, \tag{2.16a}$$

$$\overline{F}_x - (2\overline{F} - \overline{G})\,\overline{G} = u - \phi = -g_{xx}/g. \tag{2.16b}$$

The forms are very similar to (2.13a) and (2.13b) with $(u + \phi)$ and $(u - \phi)$ interchanged. In (2.16) it is $u + \phi$ that obeys a simple Riccati equation.

Combining (2.6) and (2.12) or (2.15) we have the recursion operator $T(u,\phi)$ as a product of four 2×2 matrices whose elements are of first order involving new dependent variables $f(x,t)\,g(x,t)$ defined by (2.8).

We wish to remark that all the elements of $T_1(u,\phi)$ and $T_2(u,\phi)$ given by (2.4) except the (2,2) element of $T_2(u,\phi)$ can

be written as a sum or difference of the following operators as can be seen from (2.6), (2.7), (2.10), and (2.11):

$$D^2 + 4(u \pm \phi) + 2(u_x \pm \phi_x)D^{-1}, \tag{2.17a}$$

$$D^2 + (u \pm \phi), \tag{2.17b}$$

$$D^2 + (u \pm \phi) + (u_x \pm \phi_x) D^{-1} = D(D^2 + (u \pm \phi)) D^{-1}. \tag{2.17c}$$

These are the only operators, of all the operators of the form

$$D^2 + \alpha(u \pm \phi) + \beta(u_x \pm \phi_x) D^{-1}$$

that can be factorized using the substitutions (2.8).

The scattering operator $L(u,\phi)$ for (1.2)[14,15] can also be simply factorized using (2.8). Thus

$$L(u,\phi) = (D^2 + u + \phi)(D^2 + u - \phi)$$
$$= (D + 2f_x/f) D (D - 2f_x/f) D^{-1}$$
$$\times (D + 2g_x/g) D (D - 2g_x/g) D^{-1}. \tag{2.18}$$

III. INVERSE OF THE RECURSION OPERATOR $T(u,\phi)$

The inverse of $T_1(u,\phi)$ is simply obtained because of the symmetry of the matrix. One needs to know only the inverse of each of the elements. Since

$$(D \pm 2f_x/f)^{-1} = f^{\mp 2} \int_{-\infty}^{x} dx_1 f^{\pm 2}(x_1,t), \tag{3.1}$$

we get

$$T_1^{-1}(u,\phi) = \frac{1}{2}\begin{pmatrix} D\left(f^2 \int_{-\infty}^{x} dx_1 f^{-2}\right) D^{-1} & D\left(g^2 \int_{-\infty}^{x} dx_1 g^{-2}\right) D^{-1} \\ D\left(f^2 \int_{-\infty}^{x} dx_1 f^{-2}\right) D^{-1} & -D\left(g^2 \int_{-\infty}^{x} dx_1 g^{-2}\right) D^{-1} \end{pmatrix}$$
$$\times \begin{pmatrix} f^{-2} \int_{-\infty}^{x} dx_1 f^2 & f^{-2} \int_{-\infty}^{x} dx_1 f^2 \\ g^{-2} \int_{-\infty}^{x} dx_1 g^2 & -g^{-2} \int_{-\infty}^{x} dx_1 g^2 \end{pmatrix}. \tag{3.2}$$

The inverse of $T_2(u,\phi)$ could not be obtained in a closed form. To obtain the inverse of a matrix operator of the form.

$$\begin{pmatrix} a_{11}D + f_{11} & a_{12}D + f_{12} \\ D(a_{21}D + f_{21}) D^{-1} & D(a_{22}D + f_{22}) D^{-1} \end{pmatrix} \tag{3.3}$$

appearing in (2.9), in a simple closed form we found that the condition

$$f_{11}/a_{11} = f_{12}/a_{12} \tag{3.4}$$

has to be satisfied. Similar conditions should hold for other f_{ij} and g_{ij}. It is seen from (2.12) or from (2.15) that the elements in (2.12) and (2.15) do not satisfy (3.4). In deriving (2.12) we had fixed a_{ij} and b_{ij} in (2.9) early in the calculations. Leaving these constants undetermined until the end to see if the constants could be adjusted to give the desired relation for f_{ij} and g_{ij} did not help. Another alternative was to see if the factors could have a different form from the one assumed in (2.9). We wrote

$$T_2(u,\phi) = r \cdot s,$$

where r and s are 2×2 matrices whose elements are of the form

$$r_{ij} = a_{ij}D + f_{ij}(x,t),$$
$$s_{ij} = D(b_{ij}D + g_{ij}(x,t))D^{-1}, \quad i,j = 1,2,$$

where the constants a_{ij} and b_{ij} and the functions $f_{ij}(x,t)$ and $g_{ij}(x,t)$ are to be determined. This did not give consistent solutions for f_{ij} and g_{ij}. However one can write each of the factors in (2.12) or in (2.15) as a sum of matrices which have simple inverses. But this would lead to an infinite series for the inverses of the matrices in (2.12) or in (2.15) and hence not of interest.

IV. SCATTERING EQUATIONS

By differentiating (2.8b) twice with respect to x we get

$$L(u,\phi)\, g(x,t) \equiv \{D^4 + 2uD^2 + 2(u_x - \phi_x)D + (u_{xx} - \phi_{xx}) + (u^2 - \phi^2)\}\, g(x,t) = 0. \tag{4.1}$$

It is easy to see by expanding $L(u,\phi) = (D^2 + u + \phi)(D^2 + u - \phi)$ in (2.18) that it is the same as the rhs of (4.1).

Differentiating (2.8a) twice with respect to x gives the adjoint of $L(u,\phi)$.

V. SINE-GORDON EQUATION

The polynomial recursion operator for the SG equation

$$\phi_{xt} = \sin\phi \tag{5.1}$$

is[17]

$$T_s(\phi) = D^2 + \phi_x \int_{-\infty}^{x} dx_1\, \phi_{x_1} \frac{\partial}{\partial x_1}. \tag{5.2}$$

We write

$$T_s(\phi) = (D + p) D^{-1}(D + q)D. \tag{5.3}$$

Equating this to (5.2) we get

$$p = -q = c\phi_x, \quad c^2 = -1, \tag{5.4}$$

and

$$T_s(\phi) = (D + i\phi_x) D^{-1}(D - i\phi_x)D$$
$$= (D - i\phi_x)D^{-1}(D + i\phi_x)D. \tag{5.5}$$

The ordering of $D \pm i\phi_x$ is of no consequence as $c = \pm i$. No new dependent variables are needed to factorize $T_s(\phi)$.

Since the inverse $(D \pm i\phi_x)^{-1}$ of $D \pm i\phi_x$ is

$$(D \pm i\phi_x)^{-1} = e^{\mp i\phi(x,t)} \int_{-\infty}^{x} dx_1\, e^{\pm i\phi(x_1,t)}, \tag{5.6}$$

one immediately obtains

$$T_s^{-1}(\phi) = \frac{1}{2}\left\{ \int_{-\infty}^{x} dx_1\, e^{i\phi(x_1,t)} \int_{-\infty}^{x_1} dx_2\, e^{-i\phi(x_2,t)} + \int_{-\infty}^{x} dx_1\, e^{-i\phi(x_1,t)} \int_{-\infty}^{x_1} dx_2\, e^{i\phi(x_2,t)} \right\},$$

the result derived in Ref. 17, after considerable algebra.

VI. CONCLUSION

Recursion operators for the IT about solutions of integrable NLEE can be found directly from the given NLEE. The eigenfunctions of these operators have interesting properties and it will be useful to study the recursion operators directly. Hoping that factorization of these operators would be a step in this direction we have shown that the 2×2 recursion operator of the coupled KdV equation with fourth-order elements can be written as a product of four 2×2 matrices whose elements are of the first order. We have been able to show from these factors that the inverse of these recursion operators cannot be written in a closed form. The auxiliary functions introduced to factorize the recursion operator lead to the scattering equations for the NLEE. We have also factorized the polynomial recursion operator of the SG equation and very simply obtained its inverse.

ACKNOWLEDGMENT

We thank Dr. R. Subramanian for helpful discussions.

[1] A. S. Fokas, J. Math. Phys. **21**, 1318 (1980).
[2] R. N. Aiyer, J. Phys. A; Math. Gen. **16**, 913 (1983).
[3] B. Fuchssteiner, Nonlinear Anal. TMA **3**, 849 (1979).
[4] B. Fuchssteiner, Prog. Theor. Phys. **65**, 861 (1981).
[5] B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[6] D. J. Kaup, J. Math. Anal. Appl. **54**, 849 (1976).
[7] M. B. Fogel, S. E.Trullinger, A. R. Bishop, and J. A. Krumhansl, Phys. Rev. B **15**, 1578 (1977).
[8] A. C. Newell, in *Solitons, Topics in Current Physics*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980), Vol. 17.
[9] J. P. Keener and D. W. McLaughlin, J. Math. Phys. **18**, 2008 (1977).
[10] A. P. Fordy and J. Gibbons, J. Math. Phys. **21**, 2508 (1980).
[11] A. P. Fordy and J. Gibbons, J. Math. Phys. **22**, 1170 (1981).
[12] R. Hirota and J. Satsuma, Phys. Lett. A **85**, 407 (1981).
[13] R. K. Dodd and A. P. Fordy, Phys. Lett. A **89**, 168 (1982).
[14] G. Wilson, Phys. Lett. A **89**, 332 (1982).
[15] R. N. Aiyer, Phys. Lett. A **93**, 368 (1983).
[16] B. Fuchssteiner and W. Oevel, J. Math. Phys. **23**, 358 (1982).
[17] R. N. Aiyer, J. Phys. A; Math. Gen. **16**, 255 (1983).

Inverse scattering for geophysical problems. II. Inversion of acoustical data

A. G. Ramm
Mathematics Department, Kansas State University, Manhattan, Kansas 66506

A. B. Weglein
Geophysical Research, Sohio Petroleum Company, One Lincoln Center, 5400 LBJ Freeway, Dallas, Texas 75240

(Received 29 December 1983; accepted for publication 30 March 1984)

An algorithm is given for finding the density and bulk modulus (refraction coefficient) of an inhomogeneity from the knowledge of the scattered field on the surface of the earth for all positions of the source and receiver on this surface and for two arbitrary fixed frequencies in the Born approximation. An alternative inversion method using the low-frequency data is also given.

PACS numbers: 03.40.Kf, 91.90. + p

I. INTRODUCTION AND BASIC FORMULAS

Consider the reduced wave equation describing the acoustic wave propagation in the three-dimensional space

$$\frac{\omega^2 G}{K(x)} + \nabla\cdot\frac{1}{\rho}\nabla G = -\delta(x-y), \tag{1}$$

where $x, y \in \mathbb{R}$, $K(x)$ is the bulk modulus and $\rho(x)$ is the density. Let us assume that

$$\frac{1}{K(x)} = \frac{1}{K_r} - \frac{a_1(x)}{\rho_r}, \quad \frac{1}{\rho} = \frac{1}{\rho_r} - \frac{a_2(x)}{\rho_r}, \tag{2}$$

where K_r and ρ_r are positive constants, $a_1(x)$ and $a_2(x)$ are smooth functions with compact support, i.e., $a_1 = a_2 = 0$ for $|x| > R$, where $R > 0$ is an arbitrary large fixed number.

Equation (1) can be written as

$$(\nabla^2 + \omega^2/c^2)G - \omega^2 a_1(x)G - \nabla\cdot(a_2(x)\nabla G) = -\rho_r\delta(x-y), \tag{3}$$

where $c^2 \equiv K_r\,\rho_r^{-1}$. For simplicity let us choose the units in which

$$\rho_r = 1, \quad c = 1. \tag{4}$$

This is possible: if one sets $x = \alpha x'$, $\omega = \beta\omega'$, $y = \alpha y'$, where α and β are constants then Eq. (3) takes the form

$$(\alpha^{-2}\nabla'^2 + (\beta^2/c^2)\omega'^2)G - \beta^2\omega'^2 a_1 G - \alpha^{-2}\nabla'\cdot(a_2\nabla' G) = -\rho_r\alpha^{-3}\delta(x'-y'). \tag{5}$$

One can choose $\alpha = \rho_r$, $\beta = c/\rho_r$, and define $a_1' = c^2 a_1$, $a_2' = a_2$. Then (3) takes the form

$$(\nabla'^2 + \omega'^2)G - \omega'^2 a_1' G - \nabla'\cdot a_2'\nabla G = -\delta(x'-y'). \tag{6}$$

Therefore, one can study Eq. (3) under assumption (4):

$$(\nabla^2 + \omega^2)G - \omega^2 a_1(x)G - \nabla\cdot(a_2(x)\nabla G) = -\delta(x-y). \tag{7}$$

This equation can be written as ($\int = \int_{\mathbb{R}^3}$, $dz = dz_1\,dz_2\,dz_3$)

$$G(x,y) = g(x,y) - \omega^2\int g(x,z)a_1(z)G(z,y)dz - \int g(x,z)\nabla_z\cdot(a_2(z)\nabla_z G(z,y))dz, \tag{8}$$

$$g = \frac{\exp(i\omega|x-y|)}{4\pi|x-y|}. \tag{9}$$

In the operator form (8) can be written as

$$G = g - \omega^2 g a_1 G - g\nabla\cdot a_2\nabla G \equiv g - gVG, \tag{10}$$

where

$$V = \omega^2 a_1 + \nabla\cdot a_2\nabla. \tag{11}$$

Notice that $G = g - GVg$, so that $GVg = gVG$. Let us define the T-matrix:

$$T = V - VGV. \tag{12}$$

It is obvious (and well known) that T satisfies the equation

$$T = V - VgT \tag{13}$$

and

$$Tg = VG. \tag{14}$$

An equivalent definition of T is $T = VGg^{-1}$. The operator (11) in the wavenumber representation is an integral operator with the kernel $(2\pi)^{-3}\tilde{V}(k,p)$, where

$$\tilde{V}(k,p) = \iint dx\,dy\,V(x,y)\exp[i(k\cdot x - p\cdot y)],$$

so that

$$\tilde{V}(k,p) \equiv \omega^2\tilde{a}_1(k-p) - k\cdot p\tilde{a}_2(k-p), \quad \tilde{a}(k) = \int \exp(ik\cdot x)a(x)dx. \tag{15}$$

The quantity $G - g = G_s$ is the scattered field,

$$G_s = -gVG = -gTg. \tag{16}$$

The last equality follows from (14). Let us keep the x and y variables (the position of the geophone and the source) on the plane $x_3 = 0$ and Fourier transform (16) in $\hat{x} = (x_1, x_2)$ and $\hat{y} = (y_1, y_2)$. Then ($\hat{k}\cdot\hat{x} \equiv k_1x_1 + k_2x_2$)

$$\begin{aligned}\tilde{G}_s(\hat{k},\hat{p},\omega) &= \int \exp(i\hat{k}\cdot\hat{x} - i\hat{p}\cdot\hat{y})G_s(\hat{x},\hat{y},\omega)d\hat{x}\,d\hat{y} \\ &= -\int \exp[i(\hat{k}\cdot\hat{x} - \hat{p}\cdot\hat{y})]gTg\,d\hat{x}\,d\hat{y} \\ &= -\iint dz'\,dz''\,T(z',z'') \\ &\quad\times\int \frac{\exp(i\omega|z'-\hat{x}| + i\hat{k}\cdot\hat{x})}{4\pi|z'-\hat{x}|}d\hat{x} \\ &\quad\times\int \frac{\exp(i\omega|z''-\hat{y}| - i\hat{p}\cdot\hat{y})}{4\pi|z''-\hat{y}|}d\hat{y} \\ &= -\iint dz'\,dz''\,T(z',z'') \\ &\quad\times e^{i\hat{k}\cdot\hat{z}' - i\hat{p}\cdot\hat{z}''}h(|z_3'|,|\hat{k}|)h(|z_3''|,|\hat{p}|),\end{aligned} \tag{17}$$

 0022-2488/84/113231-04$02.50

where [see Appendix A and Ref. 1, formula (8.6.21)]

$$h(|z_3'|,|\hat{k}|)=\begin{cases}\frac{i}{2}\dfrac{\exp[i|z_3'|(\omega^2-|\hat{k}|^2)^{1/2}]}{(\omega^2-|\hat{k}|^2)^{1/2}}, & \omega>|\hat{k}|,\\ \frac{1}{2}\dfrac{\exp[-|z_3'|(|\hat{k}|^2-\omega^2)^{1/2}]}{(|\hat{k}|^2-\omega^2)^{1/2}}, & \omega<|\hat{k}|,\\ \infty, & \omega=|\hat{k}|.\end{cases} \tag{18}$$

Let us define

$$k_3=\begin{cases}(\omega^2-|\hat{k}|^2)^{1/2}, & |\hat{k}|<\omega,\\ i(|\hat{k}|^2-\omega^2)^{1/2}, & |\hat{k}|>\omega,\end{cases} \tag{19}$$

and p_3 in the same way. Then

$$|\hat{k}|^2+k_3^2=\omega^2,\quad |\hat{p}|^2+p_3^2=\omega^2. \tag{20}$$

Vectors $\hat{k},\hat{p}$ are real valued but k_3 and p_3 may be complex numbers. If V has compact support (and this was our assumption) then T has compact support as follows from (12). Therefore, the Fourier transform of T makes sense for complex k_3 and p_3. From (17) and (19) one obtains

$$\tilde{G}_s=\frac{1}{4}\iint dz'\,dz''\;T(z',z'')\times\frac{\exp[i(\hat{k}\cdot\hat{z}'+k_3|z_3'|)]}{k_3}\cdot\frac{\exp[-i(\hat{p}\cdot\hat{z}''-p_3|z_3''|)]}{p_3}. \tag{21}$$

Let us assume that the support of V belongs to the half-space $z_3<0$. Then $|z_3|=-z_3$ and (21) takes the form

$$\tilde{G}_s=(1/4k_3p_3)\tilde{T}(k',p),\quad |k|=|p|=\omega,$$
$$k'\equiv(\hat{k},-k_3),\quad |k'|=|k|, \tag{22}$$

where

$$\tilde{T}(k,p)\equiv\iint dz'\,dz''\;T(z',z'')e^{i(k\cdot z'-p\cdot z'')}. \tag{23}$$

The basic problem can be formulated as follows. Given the scattering data $G(\hat{x},x_3=0,\hat{y},y_3=0,\omega)$ for all $\hat{x},\hat{y}$ and small ω find $a_1(x)$ and $a_2(x)$. This problem is discussed in the next section. In Ref. 2 the case $a_2=0$ was treated. Our method uses some ideas from Refs. 2 and 3 but the presentation is self-contained. Formulas (12), (15), and (22) are basic for our first inversion scheme presented in Sec. II. In Ref. 2 the inversion scheme was considered for $\omega\to 0$ in which case the Born approximation reduces to the exact solution. In the present paper the inversion scheme is given in the Born approximation for arbitrary $\omega>0$. The scheme can be generalized to include the dissipative terms corresponding to first derivative in time with a coefficient depending on x but not on t. The reason why we cannot handle the inversion within the exact theory (as in Ref. 2) rather than in the Born approximation is that the perturbation we consider does not become small as $\omega\to 0$ (unlike in the case $a_2=0$ considered in Ref. 2). An alternative inversion scheme similar to the one given in Ref. 2 is described in Sec. III. The scheme in Ref. 2 can be carried through for any $\omega>0$ in the Born approximation.

II. BASIC INVERSION SCHEME

Formulas (12) and (15) show that under the assumption

$$|k|=|p|=\omega \tag{24}$$

[which means that the T-matrix in (22) is known "on shell"] the potential V is

$$\tilde{V}(k,p)=\omega^2W,\quad W=\tilde{a}_1(k-p)-k^0\cdot p^0\tilde{a}_2(k-p), \tag{25}$$

where $k^0=k/|k|, p^0=p/|p|$ are unit vectors which do not depend on ω. Since the functions a_1 and a_2 have compact support, the functions $\tilde{a}_1$ and $\tilde{a}_2$ are entire functions of the three complex variables k_1,k_2,k_3. These functions decay at infinity in the real space $\mathbb{R}^3$.

In the Born approximation $T=V$ [see (14)] and (22) becomes in this approximation

$$\tilde{G}_s=(1/4k_3p_3)\tilde{V}(k',p)=(\omega^2/4k_3p_3)W(k',p). \tag{26}$$

Thus

$$4\tilde{G}_sk_3p_3/\omega^2=W(k',p). \tag{27}$$

Let

$$k'-p=q,\quad k'+p=s, \tag{28}$$
$$k'=(s+q)/2,\quad p=(s-q)/2. \tag{29}$$

Then

$$k'\cdot p=(|s|^2-|q|^2)/4,\quad s\cdot q=0, \tag{30}$$
$$k'^0\cdot p^0=(|s|^2-|q|^2)/4\omega^2, \tag{31}$$
$$|q|^2=2\omega^2(1-k'^0\cdot p^0), \tag{32}$$
$$|s|^2=|k'|^2+|p|^2+2k'\cdot p=2\omega^2(1+k'^0\cdot p^0),\quad \omega>0,$$
$$|s|^2-|q|^2=4\omega^2-2|q|^2, \tag{33}$$

thus

$$W(k',p)=W=\tilde{a}_1(q)-\tilde{a}_2(q)(|s|^2-|q|^2)/4\omega^2=\tilde{a}_1(q)-\tilde{a}_2(1-|q|^2/2\omega^2),\quad \omega>0. \tag{34}$$

Let us take two arbitrary frequencies ω_1 and $\omega_2\neq\omega_1$ and solve the two equations

$$W_1=\tilde{a}_1(q)-\tilde{a}_2(q)(1-|q|^2/2\omega_1^2), \tag{35}$$
$$W_2=\tilde{a}_1(q)-\tilde{a}_2(q)(1-|q|^2/2\omega_2^2), \tag{36}$$

for $\tilde{a}_1$ and $\tilde{a}_2$. The result is

$$\tilde{a}_2=\frac{2(W_1-W_2)}{|q|^2}\frac{\omega_1^2\omega_2^2}{\omega_2^2-\omega_1^2}, \tag{37}$$

$$\tilde{a}_1=W_1+(W_1-W_2)\frac{2\omega_1^2\omega_2^2}{(\omega_2^2-\omega_1^2)|q|^2}\left(1-\frac{|q|^2}{2\omega_1^2}\right). \tag{38}$$

Taking the inverse Fourier transform one obtains $a_1(x)$ and $a_2(x)$. An important point is that one cannot find both functions a_1 and a_2 if the data are given at a fixed frequency. Indeed, from (34) it follows that for a fixed frequency the data depends on q only and there is no parameter to vary in order to find both $\tilde{a}_1$ and $\tilde{a}_2$. This conclusion is not at all obvious: at first glance one can think that the two conditions $|k|=|p|=\omega$ leave four degrees of freedom in the six-dimensional space $\mathbb{R}^3_k\times R^3_p$, which should be enough to determine two functions $\tilde{a}_1$ and $\tilde{a}_2$ of three variables. This argument however does not work as one saw above. The reason is that the function $W(k',p)$ has a very special structure as a function of two vectors k' and p. One more remark: Eqs. (37) and (38) determine the Fourier transforms of a_1 and a_2 for real-valued vectors q in the ball $|q|<2\omega$ only. Since a_1 and a_2 vanish outside of a compact domain by assumption, their Fourier

transforms are entire analytic functions. Therefore, if their Fourier transforms are known in the ball $|q| < 2\omega$ they are uniquely defined everywhere. From the numerical point of view, if ω is large one can Fourier invert the functions $\tilde{a}_j(q)$ which are set to be zero for $|q| > 2\omega$. A more elaborate numerical inversion and a study of the stability of the Fourier inversion of entire functions measured in a ball of finite radius is given in Ref. 2b.

III. AN ALTERNATIVE INVERSION SCHEME

If $\omega \to 0$ the limit of Eq. (8) is

$$G(x,y) = g_0(x,y) - \int g_0(x,z)\nabla_z \cdot a_2(z)\nabla_z G(z,y)dz,$$

$$g_0 = 1/(4\pi|x-y|). \tag{39}$$

Consider the scattered field in the Born approximation on the plane $x_3 = 0$:

$$G_s(\hat{x},\hat{y}) \equiv G_s(\hat{x},\hat{y},\omega = 0)$$
$$= -\frac{1}{16\pi^2}\int \frac{1}{|\hat{x}-z|}\nabla_z \cdot a_2(z)\nabla_z \frac{1}{|z-\hat{y}|}dz. \tag{40}$$

This equation can be solved analytically by the method given in Ref. 2. Let us take the Fourier transform in $\hat{x}$ and $\hat{y}$:

$$f(\lambda,\mu) = -\frac{16\pi^2}{(2\pi)^4}\int d\hat{x}\, d\hat{y} \exp\{i(\lambda\cdot\hat{x} + \mu\cdot\hat{y})\} G_s(\hat{x},\hat{y})$$
$$= \frac{1}{(2\pi)^2}\int \frac{\exp(i\lambda\cdot\hat{z} - |\lambda|\,|z_3|)}{|\lambda|}$$
$$\times \nabla_z \cdot a_2(z)\nabla_z e^{i\mu\cdot\hat{z}} \cdot \frac{e^{-|\mu|\,|z_3|}}{|\mu|}dz$$
$$= \frac{-1}{(2\pi)^2|\lambda|\,|\mu|}\int (i\lambda + |\lambda|e_3)$$
$$\times \exp(i\lambda\cdot\hat{z} + |\lambda|z_3)a_2(z)(i\mu + |\mu|e_3)$$
$$\times \exp(i\mu\cdot\hat{z} + |\mu|z_3)dz$$
$$= -\frac{1}{(2\pi)^2|\lambda|\,|\mu|}\int dz\, a_2(z)\exp[i(\lambda+\mu)\cdot\hat{z}$$
$$+ (|\mu| + |\lambda|)z_3]\cdot(-\mu\cdot\lambda + |\lambda|\cdot|\mu|). \tag{41}$$

Here we used the formulas

$$\frac{1}{(2\pi)^2}\int \frac{\exp(i\lambda\cdot\hat{x})}{|\hat{x}-z|}d\hat{x} = \frac{\exp(i\lambda\cdot z - |\lambda|\,|z_3|)}{2\pi|\lambda|},$$
$$-|z_3| = z_3. \tag{42}$$

The second formula (42) holds since we assume that $a_2(z) = 0$ if $z_3 > 0$. From (41) it follows that the scattering data $G_0(\hat{x},\hat{y})$ determines the function

$$\psi(\lambda,\mu) \equiv -\frac{(2\pi)^2 f(\lambda,\mu)|\lambda|\,|\mu|}{|\lambda|\,|\mu| - \lambda\cdot\mu}$$
$$= \int dz\, a_2(z)\exp[i(\lambda+\mu)\cdot\hat{z} + (|\mu| + |\lambda|)z_3]. \tag{43}$$

Let $\lambda + \mu = p = (p_1,p_2)$, $|\lambda| = p_3$, $|\mu| = p_4$, $q = p_3 + p_4$. (Do not confuse these p and q with p and q in Sec. II.) Then ψ, defined in these variables, can be written as

$$\psi = F(p_1,p_2,p_3,p_4),$$
$$F(p_1,p_2,p_3,p_4) = \int dz\, a_2(z)\exp(ip\cdot\hat{z} + qz_3).$$

Let us set $p_3 = p_4 = q/2$. Then

$$F(p_1,p_2,q/2,q/2)$$
$$\equiv h_2(p_1,p_2,q) = h(p,q)$$
$$= \int_{-\infty}^{\infty}\int d\hat{z}\exp(ip\cdot\hat{z})\int_0^\infty d\zeta \exp(-q\zeta)a_2(\hat{z},-\zeta). \tag{44}$$

From the data $h(p,q)$ one finds $a_2(z) = a_2(\hat{z},-\zeta)$ by taking the two-dimensional inverse Fourier transform and then the inverse Laplace transform. If $a_2(z)$ is found then Eq. (8) in the Born approximation can be rewritten as ($x = \hat{x}, y = \hat{y}$)

$$-16\pi^2 \lim_{\omega\to 0}\frac{G_s - Q}{\omega^2} = \int \frac{a_1(z)dz}{|\hat{x}-z|\,|\hat{y}-z|}, \tag{45}$$

where Q in (45) is defined to be the third term in the right-hand side of (8) with g substituted in place of G. If a_2 is found then Q is known so that the left-hand side of (45) is known. Let us denote this known function by $h_1(\hat{x},\hat{y})$.

Equation (45) which can be written as

$$\int \frac{a_1(z)dz}{|\hat{x}-z|\,|\hat{y}-z|} = h_1(\hat{x},\hat{y}) \tag{46}$$

was solved in Ref. 2. Thus, the alternative inversion scheme which requires the knowledge of the scattered field $G_s = G - g$ on the plane $x_3 = 0$ for all positions $\hat{x}$ and $\hat{y}$ of the receiver and source and for small ω is as follows. First, find a_2 from Eq. (44). Secondly, find a_1 from Eq. (46).

IV. BIBLIOGRAPHICAL SKETCH

Born (1926)[4] first used the approximate linear data-perturbation relationship in atomic scattering calculations. This approximation was applied to an inverse acoustic scattering problem, with T-matrix data, by Wolf (1969).[5] Cohen and Bleistein (1979)[6] applied this approximation to the constant density acoustic equation for coincident source receiver data on the $z_3 = 0$ plane. The Born inversion of the variable density acoustic equation was addressed in Raz (1981),[7] Clayton and Stolt (1981),[3] and Wilcox (1983).[8] These works require, for a single determination of bulk modulus and density variations, sources and receivers everywhere on the surface of the earth and *all* temporal frequencies. In contrast, the techniques presented in this paper require reflection data at either a single, very small temporal frequency or two arbitrary frequencies. Coen, Cheney, and Weglein (1984)[9] present an exact, two-dimensional, two-temporal-frequency, acoustic inversion method which requires transmission as well as reflection data. A two-temporal-frequency Born inversion method which also requires transmission and reflection data is given in Devaney (1983).[10]

ACKNOWLEDGMENT

A. B. Weglein would like to thank Sohio Petroleum Company for permission to publish this work.

APPENDIX A: DERIVATION OF EQ. (18)

Here the details of the computation of integral (18) are given. One has

$$\int \frac{\exp(-i\hat{k}\cdot\hat{x}+i\omega|\hat{x}-z)}{4\pi|\hat{x}-z|}\,d\hat{x}\Big|_{\hat{x}-\hat{z}=\hat{u}} = e^{-i\hat{k}\cdot\hat{z}}\int \frac{\exp[-i|\hat{k}|\,|\hat{u}|\cos\phi+i\omega(|\hat{u}|^2+z_3^2)^{1/2}]}{4\pi(|\hat{u}|^2+z_3^2)^{1/2}}\,d\hat{u}$$

$$= e^{-i\hat{k}\cdot\hat{z}}\frac{1}{4\pi}\int_0^\infty \frac{dr\,r\exp[i\omega(r^2+z_3^2)^{1/2}]}{(r^2+z_3^2)^{1/2}}\int_0^{2\pi} e^{-i|\hat{k}|r\cos\phi}\,d\phi$$

$$= \frac{e^{-i\hat{k}\cdot\hat{z}}}{|\hat{k}|^{1/2}}\frac{1}{4\pi}\int_0^\infty \frac{dr\sqrt{r}\exp[i\omega(r^2+z_3^2)^{1/2}]\sqrt{r|\hat{k}|}}{(r^2+z_3^2)^{1/2}}\,2\pi J_0(|\hat{k}|r) = \frac{e^{-i\hat{k}\cdot\hat{z}}}{2|\hat{k}|^{1/2}}\chi, \tag{A1}$$

where [Ref. 1, formula (8.6.21)]

$$\chi = \begin{cases} i\left(\dfrac{\pi|z_3|}{2}\right)^{1/2}|\hat{k}|^{1/2}\dfrac{H^{(1)}_{-1/2}(|z_3|\sqrt{\omega^2-|\hat{k}|^2})}{(\omega^2-|\hat{k}|^2)^{1/4}}, & \omega>|\hat{k}|, \\ \left(\dfrac{2|z_3|}{\pi}\right)^{1/2}|\hat{k}|^{1/2}\dfrac{K_{1/2}(|z_3|\sqrt{|\hat{k}|^2-\omega^2})}{(|\hat{k}|^2-\omega^2)^{1/4}}, & \omega<|\hat{k}|, \end{cases} \tag{A2}$$

where

$$H^{(1)}_{-1/2}(z) = (2/\pi z)^{1/2}e^{iz},\quad K_{1/2}(z) = (\pi/2z)^{1/2}e^{-z}. \tag{A3}$$

From (A1)–(A3) formula (18) follows.

APPENDIX B: SUFFICIENT CONDITIONS FOR CONVERGENCE OF THE BORN SERIES IN ACOUSTICAL PROBLEMS

Let

$$u = g+\omega^2\int ga_1u\,dz+\int g\nabla\cdot a_2\nabla u\,dz$$

$$\equiv g+\omega^2T_1u+T_2u,\quad \int=\int_{R^3}.$$

Let $a_1=a_2=0$ if $|x|>R$. By c we denote various constants. The Born series $\Sigma_{n=0}^{\infty}(\omega^2T_1+T_2)^nf$ converges in some space H if $\|\omega^2T_1\|+\|T_2\|<1$, where $\|T\|$ is the norm of a linear operator T acting on this space. Let us take as the space H the Sobolev space H_R^2 of the functions defined in the ball $B_R=\{x:|x|\leqslant R\}$. Notice that since a_1 and a_2 have support in this ball the values of f outside this ball do not influence the values of $T_jf, j=1,2$. The following estimate is well known:

$$\|gf\|_{2,R}\leqslant c_R\|f\|_{0,R}, \tag{B1}$$

where

$$\|f\|^2_{m,R}=\int_{|x|<R}\sum_{j=0}^{m}|D^jf|^2\,dx,$$

D^jf denotes an arbitrary derivative of order j. Since we are going to let $\omega\to 0$ it is sufficient to give conditions under which $\|T_1\|\leqslant c$, $\|T_2\|<1$. We have [using the inequality (B1)]

$$\|T_1f\|_{2,R}\leqslant c_R\|a_1f\|_{o,R}$$
$$\leqslant c_R\max|a_1|\cdot\|f\|_{0,R} \tag{B2}$$
$$\leqslant c_R\max|a_1|\cdot\|f\|_{2,R}.$$

Thus, $\|T_1\|\leqslant$ const if $\max|a_1|<\infty$. Furthermore,

$$\|T_2f\|_{2,R}\leqslant c_R\|\nabla\cdot a_2\nabla f\|_{o,R}$$
$$\leqslant c_R\max|a_2|\,\|\nabla^2f\|_{0,R}+c_R\max|\nabla a_2|\,\|\nabla f\|_{0,R}$$
$$\leqslant c_R|a_2|_1\cdot\|f\|_{2,R},\quad |a_2|_1=\max|a_2|+\max|\nabla a_2|. \tag{B3}$$

Therefore, $\|T_2\|<1$ if $c_R|a_2|<1$. The following result is proved.

Theorem: Let a_1 and a_2 have support in the ball B_R. If $\max|a_1|<c_1$, $\max|a_2|+\max|\nabla a_2|<c_2$, and $q\equiv\omega^2c_1c_R+c_2c_R<1$ then the Born series converges in H_R^2 and is majorized by the geometrical series with ratio q, i.e., $\|(\omega^2T_1+T_2)^n\|\leqslant q^n$, $0<q<1$.

[1] H. Bateman and A. Erdelyi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954).
[2] A. G. Ramm, (a) "Inverse scattering for geophysical problems I," Phys. Lett. A **99**, 258 (1983). (b) "An inversion formula in scattering theory," Phys. Lett. A **99**, 201 (1983).
[3] R. Clayton and R. Stolt, "A Born–WKBJ Inversion method for acoustic reflection data," Geophysics **46**, N 11, 1559–1567 (1981).
[4] M. Born, Z. Phys. **37**, 863 (1926); **38**, 803 (1926).
[5] E. Wolf, "Three-dimensional structure determination of semi-transparent objects from holographic data," Opt. Commun. **1**, 153–156 (1969).
[6] J. Cohen and N. Bleistein, "Velocity inversion for acoustic waves," Geophys. **44**, 1077–1087 (1979).
[7] S. Raz, "Three-dimensional velocity profile inversion from finite offset scattering data," Geophysics **46**, (6), 837–842 (1981).
[8] C. Wilcox, Math. Methods Appl. Sci. **5**, 276 (1983).
[9] S. Coen, M. Cheney, and A. Weglein, J. Math. Phys. **25**, 1857 (1984).
[10] A. Devaney, IEEE Trans. Sonics Ultrason. **SU-30**, 355 (1983).

Electrodynamics of memory-dependent nonlocal elastic continua

A. Cemal Eringen
Princeton University, Princeton, New Jersey 08544

(Received 8 June 1983; accepted for publication 13 April 1984)

Balance laws and constitutive equations are given for elastic continua with memory of past motions and electromagnetic fields. Nonlinear, finite-linear, and linear constitutive equations are obtained and restricted by the second law of thermodynamics. Memory-dependent nonlocal piezoelectricity, piezomagnetism, heat and electric conduction, viscoelasticity, and other allied physical phenomena are in the domain of the general theory. The theory is applied to discuss infrared dispersion and lattice vibrations, natural optical activity, anomalous skin effect, and superconductivity, indicating the power and the potential of the nonlocal theory.

PACS numbers: 03.50.De

I. INTRODUCTION

In a previous paper,[1] I presented a nonlocal continuum theory of elastic solids subject to electromagnetic (EM) interactions. The electric and heat conductions were not included in the theory and the memory dependence was not considered. In order to discuss the absorption and damping of waves at all frequencies, one needs to take into account the effect of strain and EM field histories in the constitutive equations. Moreover, the heat and electric conduction cannot be discussed without the development of constitutive equations for the heat and current vectors generalizing classical Fourier and Ohm's laws. The *raison d'être* of the present paper is the development of a rather general continuum theory which includes these effects. The power and potential of the theory is then demonstrated by treating certain problems which fall clearly outside of the domain of classical (local) continuum theory.

There exist many interesting physical phenomena for which classical field theories are not applicable without proper modifications. The failure of classical theories stems from the fact that they do not possess a natural internal characteristic length and a characteristic time. Yet all physical phenomena depend, to some extent, on such characteristic scales because of the discrete (atomic) nature of materials and the relaxation time. In order to explain such physical phenomena, classical theories are modified, often in an *ad hoc* fashion, without reference to fundamental laws. For example, anomalous skin effects and superconductivity require approaches not entirely within the realm of Maxwell's theory of electromagnetism. Yet, these can be shown to be part of the nonlocal continuum theory (Secs. XI , XII), which also include Maxwell's theory as a special case. Much of the nonquantum aspects of the electron theory can be explained by means of the nonlocal theory, e.g., interaction of electrons with lattice vibrations, scattering of excitons near the boundaries of the Brillouin zone, infrared dispersion, absorption of waves, phase transition, nonlinear optics, streaming birefringence, piezoelectricity, etc. Some of these problems are treated here (cf. Secs. VIII–XII).

There exist a large number of references and texts on each of these topics with approaches based on semiclassical, atomic, and quantum mechanical ideas (to cite a few, cf. Refs. 2–7). Wave-number-dependent dielectrics are examples of nonlocality[8] which are relevant to semiconductor devices. However, quantum and statistical mechanical approaches are difficult and cannot be carried out, except in simple situations. Nonlinear, nonlocal problems which are especially relevant to phase transition pose insurmountable difficulties on microscopic grounds.

A macroscopic field theory of rigid-body electrodynamics, based on Fourier formalism, exists as surveyed by Rukhadze and Silin.[9] However, this approach is entirely formal treating only the linear theory and thermodynamical restrictions are not considered. In other fields, e.g., nonlocal elasticity,[10,11] fluid dynamics,[12,13] liquid crystals,[14] and electromagnetic theory,[15–17] nonlocal theory, have registered significant advances and resolved some long standing controversy.[18,19]

Motivated with this progress, I develop here a general theory of memory-dependent nonlocal electromagnetic elastic solids. Balance laws and the second law of thermodynamics are given in Secs. II and III. In Secs. IV and V, I develop general constitutive equations and thermodynamics for the nonlinear theory.

Section VI contains an account of the *finite-linear* constitutive equations which are useful for materials with weak memory, but large fields. In Sec. VII, I give the linear constitutive equations for anisotropic heat and electric conducting electro-magneto-elastic solids. Section VIII presents a discussion of elastic dielectrics. With Sec. IX, applications of the theory begin to infrared dispersion and lattice vibrations. Natural optical activity is discussed in Sec. X; anomalous skin effect in Sec. XI and superconductivity in Sec. XII. Results obtained are in conformity with other approaches based on semiclassical formalisms or electron theory. They are unified under one theory and contain other possibilities for the treatment of more general and nonlinear problems.

II. BALANCE LAWS

The body at the natural state occupies a region $V-\Sigma$, the volume V excluding a discontinuity surface Σ. The motion carries a material point $\mathbf{X}\in V-\Sigma$ to a spatial place

 0022-2488/84/113235-15$02.50

$\mathbf{x} \in \mathscr{V} - \sigma$, where $\mathscr{V} - \sigma$ is the image of $V - \Sigma$ at time t. The motion is a bijective mapping expressed by

$$\mathbf{x} = \mathbf{x}(\mathbf{X}, t) \leftrightarrow \mathbf{X} = \mathbf{X}(\mathbf{x}, t). \tag{2.1}$$

We employ a rectangular frame of reference so that rectangular coordinates of $\mathbf{x}$ and $\mathbf{X}$ are denoted by x_k and X_K, respectively ($k, K = 1,2,3$). Since (2.1) is bijective, the Jacobian must be positive

$$J \equiv \det(x_{k,K}) > 0. \tag{2.2}$$

Henceforth, we employ a comma to denote partial derivative and a dot to express the material derivative. The usual summation convention on repeated indices is also assumed, e.g.,

$$x_{k,K} = \frac{\partial x_k}{\partial X_K}, \quad \dot{x}_k = \left.\frac{\partial x_k}{\partial t}\right|_{\mathbf{X}} = v_k(\mathbf{x}, t),$$

$$a_k = \dot{v}_k = \frac{\partial v_k}{\partial t} + v_{k,l} v_l. \tag{2.3}$$

Balance laws of nonlocal electromechanical continua were given in Ref. 15. Here we are concerned with inert bodies consisting of single substance, i.e., mixtures and chemical reactions are excluded. Moreover, we assume that the balance laws are valid for a macroscopic volume element large enough to contain large number material points (atoms, molecules), but small enough so that the body contains a large number of these elements. This situation is reminiscent of statistical mechanical ensembles. Under these conditions, nonlocal residuals (action at a distance) appearing in the balance laws can be neglected as compared to fields at a reference point $\mathbf{X}$ in the body. Of course, the fields at a reference point are still influenced by the nonlocal intermolecular attractions. For example, the stress and electric polarization at $\mathbf{X}$ depend on the strain and the electric field at *all* other points of the body, but gravitational variations with distance is unimportant. Under these conditions, Maxwell's equations remain valid in $\mathscr{V} - \sigma$.

$$\nabla \cdot \mathbf{D} = q, \tag{2.4}$$

$$\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = \mathbf{0}, \tag{2.5}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{2.6}$$

$$\nabla \times \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} = \frac{1}{c}\mathbf{J}, \tag{2.7}$$

$$\frac{\partial q}{\partial t} + \nabla \cdot \mathbf{J} = 0, \tag{2.8}$$

where $\mathbf{D}$, $\mathbf{E}$, $\mathbf{B}$, $\mathbf{H}$, $\mathbf{J}$, and q are, respectively, the electric displacement vector, electric vector, magnetic induction vector, magnetic field vector, current density vector, and the charge density. c is the speed of light in vacuum.

Maxwell's equations are supplemented with the mechanical balance laws, valid in $\mathscr{V} - \sigma$ (cf. Ref. 20, Sec. 10.9)

$$\rho_0/\rho = \det x_{k,K}, \tag{2.9}$$

$$t_{kl,k} + \rho(f_l - \dot{v}_l) + {}_M f_l = 0, \tag{2.10}$$

$$t_{kl} + P_k \mathscr{E}_l + \mathscr{M}_k B_l = t_{lk} + P_l \mathscr{E}_k + \mathscr{M}_l B_k \equiv {}_E t_{kl}, \tag{2.11}$$

$$\rho\dot{\epsilon} - t_{kl} v_{l,k} - q_{k,k} - \rho h - \rho \mathscr{E} \cdot (\mathbf{P}/\rho)^{\cdot} + \mathscr{M} \cdot \dot{\mathbf{B}} - \mathscr{J} \cdot \mathscr{E} = 0, \tag{2.12}$$

where ρ_0 is the mass density in $V - \Sigma$ and ρ, $t_{kl}, f_l, v_l, \epsilon$, q_k, h are, respectively, the mass density, stress tensor, body force density, velocity vector, internal energy density, heat vector, and the heat source in $\mathscr{V} - \sigma$. $\mathscr{E}$, $\mathscr{M}$, and $\mathscr{J}$ are the electric vector, magnetization vector, and current vector in the *proper* (comoving) frame, as defined by

$$\mathscr{E} = \mathbf{E} + (1/c)\mathbf{v} \times \mathbf{B}, \quad \mathscr{M} = \mathbf{M} + (1/c)\mathbf{v} \times \mathbf{P}, \quad \mathscr{J} = \mathbf{J} - q\mathbf{v}. \tag{2.13}$$

Here $\mathbf{M}$ and $\mathbf{P}$ are, respectively, the magnetization and polarization vectors in the fixed frame so that

$$\mathbf{D} = \mathbf{E} + \mathbf{P}, \quad \mathbf{B} = \mathbf{H} + \mathbf{M}. \tag{2.14}$$

The EM body force ${}_M\mathbf{f}$ is given by (Ref. 19, Sec. 10.6)

$${}_M\mathbf{f} = q\mathbf{E} + \frac{1}{c}\mathbf{J} \times \mathbf{B} + (\nabla \mathbf{E}) \cdot \mathbf{P} + (\nabla \mathbf{B}) \cdot \mathbf{M} + \frac{1}{c}[(\mathbf{P} \times \mathbf{B}) v_k]_{,k} + \frac{1}{c}\frac{\partial}{\partial t}(\mathbf{P} \times \mathbf{B}). \tag{2.15}$$

Accompanying Maxwell's equations and mechanical balance laws, we have the *jump* conditions across σ. These conditions give boundary conditions when σ is made to coincide with the surface of the body. For brevity, we do not list these conditions here. They can be found in Ref. 20, Sec. 10.17.

III. SECOND LAW OF THERMODYNAMICS

The second law of thermodynamics is a statement about the dissipative process expressing the physical fact that the total dissipation in a body is non-negative. The localized form of the second law used in classical field theories places severe restrictions on thermodynamic behavior of materials. For the nonlocal theory, the local form of the entropy inequality is given by[10,11,20]

$$\rho\dot{\eta} - \nabla \cdot (\mathbf{q}/\theta) - (\rho h/\theta) - \rho\hat{s} \geqslant 0, \tag{3.1}$$

where η is the entropy density, $\theta > 0$ is the absolute temperature, and $\hat{s}$ is the nonlocal entropy residual resulting from entropy exchanges between the reference point and the rest of the body. It is subject to the restriction

$$\int_{\mathscr{V}-\sigma} \rho\hat{s}\, dv = 0. \tag{3.2}$$

If we eliminate h between (3.1) and (2.12) and employ the following expressions:

$$J = \rho_0/\rho, \quad \dot{C}_{KL} = (v_{k,l} + v_{l,k}) x_{k,K} x_{l,L}, \tag{3.3}$$

$$\Psi = \psi - \rho_0^{-1} \Pi_K \mathscr{E}_K = \epsilon - \theta\eta - \rho_0^{-1} \Pi_K \mathscr{E}_K, \tag{3.4}$$

$$\begin{aligned} T_{KL} &= J X_{K,k} X_{L,l} t_{kl}, & Q_K &= J X_{K,k} q_k, \\ \Pi_K &= J X_{K,k} P_k, & M_K &= J X_{K,k} \mathscr{M}_K, \\ C_{KL} &= x_{k,K} x_{k,L}, & \mathscr{E}_K &= \mathscr{E}_k x_{k,K}, \\ \theta_{,K} &= \theta_{,k} x_{k,K}, & B_k &= B_k x_{k,K} \\ & & \mathscr{J}_k &= J X_{K,k} \mathscr{J}_k, \end{aligned} \tag{3.5}$$

we obtain

$$-\rho_0(\dot{\Psi}+\eta\dot{\theta})+\tfrac{1}{2}{}_E T_{KL}\dot{C}_{KL}+(1/\theta)Q_K\theta_{,K}-\Pi_K\dot{\mathscr{E}}_K$$
$$-M_K\dot{B}_K+\mathscr{J}_K\mathscr{E}_K-\rho_0\theta\hat{s}\geqslant 0. \tag{3.6}$$

By means of (3.3) to (3.5), the equation of energy (2.12) can be written in the material form

$$\rho_0(\dot{\Psi}+\theta\dot{\eta}+\dot{\theta}\eta)-\tfrac{1}{2}{}_E T_{KL}\dot{C}_{KL}-Q_{K,K}-\rho_0 h$$
$$+\Pi_K\dot{\mathscr{E}}_K+M_K\dot{B}_K-\mathscr{J}_K\mathscr{E}_K=0, \tag{3.7}$$

Eq. (3.7) serves as the equation of heat conduction. The inequality (3.6) will be employed to place restrictions on the constitutive equations.

After dividing (3.6) with the positive quantity θ, we integrate it over the volume $V-\Sigma$, to obtain

$$\int_{V-\Sigma}\frac{1}{\theta}\Big[-\rho_0(\dot{\Psi}+\eta\dot{\theta})+\frac{1}{2}{}_E T_{KL}\dot{C}_{KL}+\frac{1}{\theta}Q_K\theta_{,K}$$
$$-\Pi_K\dot{\mathscr{E}}_K-M_K\dot{B}_K+\mathscr{J}_K\mathscr{E}_K\Big]\,dV\geqslant 0. \tag{3.8}$$

It is posited that Eq. (3.8) must not be violated for all thermodynamic processes that are physically admissible.

IV. CONSTITUTIVE EQUATIONS

According to the axiom of causality,[20,21] all physical processes that take place in a body are the result of motions (deformations) in the past up to and including the present time. When the intrinsic deformations of sub-bodies in a volume element are considered, this implies the history of centroidal motions of the volume element and the memory of temperature, temperature gradients, polarizations, and magnetizations. This is equivalent to the selection of the independent constitutive variables:

$$\mathscr{Y}'\equiv\{\mathbf{x}',\theta',\theta'_{,K},\mathscr{E}'_K,B'_K\}, \tag{4.1}$$

where a prime is used to denote the values of functions at $(\mathbf{X}',t-\tau')$, e.g.,

$$\mathbf{x}'=\mathbf{x}(\mathbf{X}',t-\tau'),\quad B'_K=B(\mathbf{X}',t-\tau'),$$
$$\mathbf{X}'\in V-\Sigma,\quad 0\leqslant\tau'<\infty. \tag{4.2}$$

Values of these functions at $(\mathbf{X},t)$ will be denoted without a prime, e.g., $\mathbf{x}=\mathbf{x}(\mathbf{X},t)$, $B_K=B_K(\mathbf{X},t)$.

Constitutive equations express the functional dependence of the set

$$Z\equiv\{\Psi,\eta,{}_E T_{KL},\Pi_K,M_K,Q_K,\mathscr{J}_K\} \tag{4.3}$$

at $(\mathbf{X},t)$ on the set (4.1), e.g.,

$$\Psi(\mathbf{X},t)=\mathscr{F}[\mathbf{x}',\theta',\theta'_{,K},\mathscr{E}'_K,B'_K]. \tag{4.4}$$

They also depend on $\mathbf{X}$ for inhomogeneous materials.

The response functionals, such as $\mathscr{F}$, must be form invariant under arbitrary spatial translations and rotations. This implies that Ψ will depend on $\mathbf{x}'$ and $\mathbf{x}$ only through the distance $|\mathbf{x}'-\mathbf{x}|$. Since the distance can be expressed as a functional[22] of C_{KL}, it proves to be convenient to replace $\mathbf{x}'$ in (4.4) by $C'_{KL}\equiv C_{KL}(\mathbf{X}',t-\tau')$.

For the discussion of thermodynamic restrictions, we distinguish each of the set of dependent variables into two categories, e.g.,

$$\mathbf{C}^\tau\equiv\mathbf{C}(\mathbf{X}',t-\tau')\quad\begin{cases}\text{when } \mathbf{X}'\neq\mathbf{X}, & 0\leqslant\tau'<\infty\\ \text{when } \mathbf{X}'=\mathbf{X}, & 0<\tau<\infty\end{cases} \tag{4.5}$$
$$\mathbf{C}\equiv\mathbf{C}(X,t).$$

For brevity, we abbreviate the following collections by

$$G^\tau\equiv\{\mathbf{C}^\tau,\theta^\tau,\theta^\tau_{,K},\mathscr{E}^\tau_K,B^\tau_K\},$$
$$G\equiv\{\mathbf{C},\theta,\theta_{,K},\mathscr{E}_K,B_K\}, \tag{4.6}$$
$$G'\equiv\{G^\tau,G\}=\{\mathbf{C}',\theta',\theta'_{,K},\mathscr{E}'_K,B'_K\}.$$

We assume that members of G' are continuously differentiable with respect to their arguments. In order to introduce a topology to the space of function G^τ, we define the inner product of two such sets by

$$(G^\tau_1,G^\tau_2)_H=\int_0^\infty d\tau'\int_{V-\Sigma}H(\mathbf{X}'-\mathbf{X},\tau')G^\tau_1\cdot G^\tau_2\,dV', \tag{4.7}$$

where

$$G^\tau_1\cdot G^\tau_2\equiv\mathrm{tr}(\mathbf{C}^\tau_1\mathbf{C}^\tau_2)+\theta^\tau_1\theta^\tau_2$$
$$+\theta^\tau_{1,K}\theta^\tau_{2,K}+\mathscr{E}^\tau_{1K}\mathscr{E}^\tau_{2K}+B^\tau_{1K}B^\tau_{2K}. \tag{4.8}$$

The *influence function* H is a positive decreasing function of its argument such that the integrals in (4.7) converge and

$$H(\mathbf{0},0)=1. \tag{4.9}$$

This function emphasizes the dependence on deformations, temperature, temperature gradients, and EM field near the reference point $\mathbf{X}$ at the present time t over their past histories and distant points $\mathbf{X}'$ from $\mathbf{X}$. This is in accordance with the *attenuating neighborhood* and *fading memory* hypothesis,[20,21] based on the nature of intermolecular forces. There exist many choices for the influence functions. As an example, we mention

$$H(\mathbf{X}',\tau')=\exp(-\alpha|\mathbf{X}'|-\beta\tau'),\quad \alpha,\beta>0. \tag{4.10}$$

Physically, more realistic forms of these functions can be selected approximating the interatomic force potential.[23–25] The space of functions G^τ is a Hilbert space $\mathscr{H}$ with a finite norm defined by

$$\|G^\tau\|=(G^\tau,G^\tau)^{1/2}_H. \tag{4.11}$$

In a Hilbert space, any continuous, linear, real-valued function $f(F)$ has the unique Riez–Fréchet representation (cf. Ref. 26, p. 421)

$$f(F)=(F,G^\tau)_H \tag{4.12}$$

valid for all G^τ.

It is now possible to calculate $\dot{\Psi}$. Let

$$\rho_0\Psi=F, \tag{4.13}$$

then

$$\rho_0\dot{\Psi}=\frac{\partial F}{\partial G}\dot{G}+\delta F(G^\tau|\dot{G}^\tau,G), \tag{4.14}$$

where δF is the Fréchet derivative of F with respect to G^τ. It is a linear functional of $\dot{G}^\tau$. Consequently, it can be expressed in the form

$$\delta F=\int_0^\infty d\tau'\int_{V-\Sigma}\frac{\delta F}{\delta G^\tau}\dot{G}^\tau\,dV'. \tag{4.15}$$

The operator $\delta(\ )/\delta(\ )$ represents the Fréchet partial derivative.

Substituting (4.14) into the entropy inequality (3.8), we obtain

$$\int_{V-\Sigma}\frac{1}{\theta}\Big[-\Big(\rho_0\eta+\frac{\partial F}{\partial\theta}\Big)\dot\theta + \frac{1}{2}\Big({}_E T_{KL}-2\frac{\partial F}{\partial C_{KL}}\Big)\dot C_{KL} - \Big(\Pi_K+\frac{\partial F}{\partial \mathscr{E}_K}\Big)\dot{\mathscr{E}}_K-\Big(M_K+\frac{\partial F}{\partial B_K}\Big)\dot B_K-\frac{\partial F}{\partial\theta_{,K}}\dot\theta_{,K} + \frac{1}{\theta}Q_K\theta_{,K}+\mathscr{J}_K\mathscr{E}_K-\delta F\Big]dV\geqslant 0. \tag{4.16}$$

This inequality is linear in the rates $\dot\theta$, $\dot C_{KL}$, $\dot{\mathscr{E}}_K$, $\dot B_K$, and $\dot\theta_{,K}$. For arbitrary and independent variations of these quantities throughout $V-\Sigma$, the inequality cannot be maintained unless

$$\eta=-\frac{1}{\rho_0}\frac{\partial F}{\partial\theta}=-\frac{1}{\rho_0}\frac{\partial\dot F}{\partial\dot\theta}, \tag{4.17}$$

$${}_E T_{KL}=2\frac{\partial F}{\partial C_{KL}}=2\frac{\partial\dot F}{\partial\dot C_{KL}}, \tag{4.18}$$

$$\Pi_K=-\frac{\partial F}{\partial\mathscr{E}_K}=-\frac{\partial\dot F}{\partial\dot{\mathscr{E}}_K}, \tag{4.19}$$

$$M_K=-\frac{\partial F}{\partial B_K}=-\frac{\partial\dot F}{\partial\dot B_K}, \tag{4.20}$$

$$0=\frac{\partial F}{\partial\theta_{,K}}=\frac{\partial\dot F}{\partial\dot\theta_{,K}}, \tag{4.21}$$

$$\int_{V-\Sigma}\frac{1}{\theta}\Big(\frac{1}{\theta}Q_K\theta_{,K}+\mathscr{J}_K\mathscr{E}_K-\delta F\Big)dV\geqslant 0. \tag{4.22}$$

On the right-hand sides of (4.17)–(4.21), we also give alternative forms of constitutive equations in terms of gradients of $\dot F$. These forms are useful in the construction of special constitutive equations. From these equations, it is clear that η, ${}_E T_{KL}$, Π_K, and M_k are determined in terms of the free-energy functional alone.[27] Equation (4.21) indicates that F cannot depend on the present value of $\theta_{,K}$ at $\mathbf{X}$. In general, nothing can be said on the dependence of F on $\theta^r_{,K}\equiv\theta_{,K}(\mathbf{X}',t-\tau'),\mathbf{X}'\neq\mathbf{X},\tau'\neq t$. For the present treatment, we shall assume that F is independent of the history of the temperature gradients of other points as well.

To complete the theory, separate constitutive equations will have to be written for Q_K and $\mathscr{J}_K$.

Spatial forms of constitutive equations follow from (3.5):

$$\eta=-\frac{1}{\rho_0}\frac{\partial F}{\partial\theta}, \tag{4.23}$$

$${}_E t_{k,l}=(\rho/\rho_0){}_E T_{KL}x_{k,K}x_{l,L}, \tag{4.24}$$

$$P_k=(\rho/\rho_0)\Pi_K x_{k,K}, \tag{4.25}$$

$$\mathscr{M}_k=(\rho/\rho_0)M_K x_{k,K}, \tag{4.26}$$

$$q_k=(\rho/\rho_0)Q_K x_{k,K}, \tag{4.27}$$

$$\mathscr{J}_k=(\rho/\rho_0)\mathscr{J}_K x_{k,K}. \tag{4.28}$$

We conclude this section by representation of the free-energy functional by means of Stone–Weirstrass theorem. According to this theorem, a real, continuous, scalar-valued functional of G' may be represented uniformly by a polynomial in a set of real continuous, linear, scalar-valued functionals of G'. Accordingly, we may write

$$\rho_0\Psi=F=\sum_{\alpha=1}^{N}F_\alpha+\sum_{\alpha=1}^{N}\sum_{\beta=1}^{N}F_\alpha F_\beta, \tag{4.29}$$

where F is a linear functional in G', which, in accordance with the Reisz representation theorem, may be expressed as

$$F_\alpha=f^0_\alpha(\mathbf{X})G(\mathbf{X},t)+\int_{V-\Sigma}f_\alpha(\mathbf{X}',0;\mathbf{X})G(\mathbf{X}',t)dV' + \int_0^\infty d\tau'\int_{V-\Sigma}\frac{\partial f_\alpha(\mathbf{X}',\tau';\mathbf{X})}{\partial\tau'} \times G(\mathbf{X}',t-\tau')dV'. \tag{4.30}$$

Note that in (4.30), the first terms give an explicit display of the dependence upon the current state at the present time. Alternatively, through integration by parts, (4.30) may be expressed as

$$F_\alpha=f^0_\alpha(\mathbf{X})G(\mathbf{X},t) + \int_{-\infty}^{t}d\tau'\int_{V-\Sigma}f_\alpha(\mathbf{X}',t-\tau';\mathbf{X}) \times\frac{\partial G(\mathbf{X}',\tau')}{\partial\tau'}dV'. \tag{4.31}$$

Carrying (4.31) into (4.29), we obtain a formal representation of the free energy in terms of the members of G' listed in the third equation of (4.6). In this way, nonlinear, nonlocal constitutive equations are constructed. Explicit expressions are too lengthy to list here.

V. ADDITIVE FUNCTIONALS

For the additive functionals, Friedman and Katz[28] gave a representation theorem according to which

$$\rho_0\psi=F=\int_0^\infty d\tau'\int_{V-\Sigma}S(\underline{G}^r,\underline{G},\mathbf{X}',\mathbf{X},\tau')dV', \tag{5.1}$$

where an underline is used to indicate that the lists of functions $\underline{G}^r$ and $\underline{G}$ exclude the temperature gradients. Since only the symmetric part of S in $\mathbf{X}$ and $\mathbf{X}'$ contributes to the total free energy of the body, we may select S as a symmetric function of its argument functions at $\mathbf{X}$ and $\mathbf{X}'$. This can be done by decomposing G^r as

$$G^r=\{G^s,G^t,G^{st}\}, \tag{5.2}$$

where

$$\begin{aligned} &G^s\equiv G(\mathbf{X}',t), && \mathbf{X}'\neq\mathbf{X},\\ &G^t\equiv G(\mathbf{X},t-\tau'), && 0<\tau'<\infty,\\ &G^{st}\equiv G(\mathbf{X}',t-\tau'), && \mathbf{X}'\neq\mathbf{X},\ 0<\tau'<\infty. \end{aligned} \tag{5.3}$$

If we let a superposed asterisk represent interchange of $\mathbf{X}$ and $\mathbf{X}'$, i.e.,

$$\overset{*}{A}(\mathbf{X}'\mathbf{X})=A(\mathbf{X},\mathbf{X}'), \tag{5.4}$$

then clearly,

$$\overset{*}{G}=G^s,\quad \overset{*}{G}{}^s=G,\quad \overset{*}{G}{}^t=G^{st},\quad \overset{*}{G}{}^{st}=G^t, \tag{5.5}$$

$$\overset{*}{S}(G,G^s,G^t,G^{st},\mathbf{X}',\mathbf{X},\tau')=S(G^s,G,G^{st},G^t,\mathbf{X},\mathbf{X}',\tau').$$

Introducing (5.1) into (4.17) to (4.20), we obtain

$$\eta = -\frac{1}{\rho_0}\int_0^\infty d\tau' \int_{V-\Sigma} \frac{\partial S}{\partial \theta} dV', \tag{5.6}$$

$${}_E T_{KL} = 2\int_0^\infty d\tau' \int_{V-\Sigma} \frac{\partial S}{\partial C_{KL}} dV', \tag{5.7}$$

$$\Pi_K = -\int_0^\infty d\tau' \int_{V-\Sigma} \frac{\partial S}{\partial \mathscr{E}_K} dV', \tag{5.8}$$

$$M_K = -\int_0^\infty d\tau' \int_{V-\Sigma} \frac{\partial S}{\partial B_K} dV'. \tag{5.9}$$

We also write constitutive equations for Q_K and $\mathscr{J}_K$

$$Q_K/\theta^2 = \int_0^\infty d\tau' \int_{V-\Sigma} K_K(G', G, \mathbf{X}', \mathbf{X}, \tau') dV', \tag{5.10}$$

$$\mathscr{J}_K/\theta = \int_0^\infty d\tau' \int_{V-\Sigma} L_K(G', G, \mathbf{X}', \mathbf{X}, \tau') dV'. \tag{5.11}$$

Polynomial constitutive equations of various degree, may be derived from (5.6) to (5.11) by expressing S, K_K, and L_K as polynomials in the vector and tensor variables.

The set of constitutive equations (5.1), (5.6)–(5.11) represents an alternative set to those obtained in Sec. IV. This set is more limited in one sense, namely, they are expressed in terms of single space-time integral while those of Sec. IV contain multifold integrals in space-time. Here the kernel functions are nonlinear functions, whereas in Sec. IV they consist of polynomials. In most nonlinear cases, the present representation should be adequate for possible calculations.

VI. FINITE-LINEAR CONSTITUTIVE EQUATIONS

For a large class of materials, the memory dependence on the past history of fields may be taken linear, yet the effect of local fields at the present time may be large. In this case, constitutive equations are called finite linear and they are obtained by taking

$$S = S^0 + \Sigma^1_{KL} C'_{KL} + \Sigma^2_K \mathscr{E}'_K + \Sigma^3_K B'_K, \tag{6.1}$$

where $S^0, \Sigma^1_{KL}, \Sigma^2_K$, and Σ^3_K are functions of $\mathbf{C}, \mathscr{E}, \mathbf{B}, \theta, \mathbf{X}, \mathbf{X}'$, and τ'. Upon substituting (6.1) into (5.6) to (5.9), and dropping nonlinear terms in G' arising from $\overset{*}{S}$, we obtain

$$\eta = -\frac{1}{\rho_0}\int_0^\infty d\tau' \int_{V-\Sigma} \left(\frac{\partial S^0}{\partial \theta} + \frac{\partial \Sigma^1_{KL}}{\partial \theta} C'_{KL} + \frac{\partial \Sigma^2_{KL}}{\partial \theta} \mathscr{E}'_K + \frac{\partial \Sigma^3_K}{\partial \theta} B'_K \right) dV', \tag{6.2}$$

$${}_E T_{KL} = 2\int_0^\infty d\tau' \int_{V-\Sigma} \left(\frac{\partial S^0}{\partial C_{KL}} + \frac{\partial \Sigma^1_{MN}}{\partial C_{KL}} C'_{MN} + \frac{\partial \Sigma^2_M}{\partial C_{KL}} \mathscr{E}'_M + \frac{\partial \Sigma^3_M}{\partial C_{KL}} B'_M \right) dV', \tag{6.3}$$

$$\Pi_k = -\int_0^\infty d\tau' \int_{V-\Sigma} \left(\frac{\partial S^0}{\partial \mathscr{E}_K} + \frac{\partial \Sigma^1_{MN}}{\partial \mathscr{E}_K} C'_{MN} + \frac{\partial \Sigma^2_M}{\partial \mathscr{E}_K \mathscr{E}'_M} + \frac{\partial \Sigma^3_M}{\partial \mathscr{E}_K} B'_M \right) dV', \tag{6.4}$$

$$M_K = -\int_0^\infty d\tau' \int_{V-\Sigma} \left(\frac{\partial S^0}{\partial B_K} + \frac{\partial \Sigma^1_{MN}}{\partial B_K} C'_{MN} + \frac{\partial \Sigma^2_M}{\partial B_K} \mathscr{E}'_M + \frac{\partial \Sigma^3_M}{\partial B_K} B'_M \right) dV'. \tag{6.5}$$

For the heat and electric conduction in Eqs. (5.10) and (5.11), we put

$$K_K = K^0_K + K^1_{KLM} C'_{LM} + K^2_{KL} \mathscr{E}'_L + K^3_{KL} B'_L + K^4_{KL} \theta'_{,L}, \tag{6.6}$$

$$L_K = L^0_K + L^1_{KLM} C'_{LM} + L^2_{KL} \mathscr{E}'_L + L^3_{KL} B'_L + L^4_{KL} \theta'_{,L}, \tag{6.7}$$

where $\mathbf{K}^\alpha$ and $\mathbf{L}^\alpha$ are functions of $\mathbf{C}, \mathscr{E}, \mathbf{B}, \theta, \theta_{,K}, \mathbf{X}, \mathbf{X}'$, and τ'. As usual, Eqs. (5.10) and (5.11) are subject to the entropy inequality (4.22).

We observe that first terms under the integral sign in these equations can be taken out of the integral by integrating S^0, K^0_K, and L^0_K over the volume and time since the argument fields $\mathbf{C}, \mathscr{E}, \mathbf{B}$, and θ are independent of $\mathbf{X}'$ and τ'.

The finite-linear theory should be useful in discussing problems related to nonlinear optics, magnetism, phase transition, and nonlinear piezoelectricity of materials with weak absorptions.

VII. LINEAR CONSTITUTIVE EQUATIONS

Linear constitutive equations are obtained by writing a second-degree polynomial for the free energy. To avoid lengthy expressions, we introduce abbreviation G_p for some members of G and T_p for some members of Z:

$$\begin{aligned} &G_1 = -T, \quad G_2 = E_{KL}, \quad G_3 = -\mathscr{E}_K, \quad G_4 = -B_K, \\ &T_1 = \rho_0 \eta, \quad T_2 = {}_E T_{KL}, \quad T_3 = \Pi_K, \quad T_4 = M_K. \end{aligned} \tag{7.1}$$

In the spirit of the linear theory, we replace the finite strain measure C_{KL} by the linear strain measure E_{KL} and consider small temperature changes T from an ambient temperature T_0, i.e.,

$$E_{KL} = \tfrac{1}{2}(C_{KL} - \delta_{KL}) \simeq \tfrac{1}{2}(U_{K,L} + U_{L,K}), \tag{7.2}$$

$$\theta = T_0 + T, \quad |T| \ll T_0, \quad T_0 > 0, \tag{7.3}$$

where U_K is the displacement vector.

Employing (4.29) and (4.31), a second-degree functional for the free energy may be expressed as

$$\rho_0\Psi = F = f^0_p G_p + \frac{1}{2} f^0_{pq} G_p G_q + \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} f_p(\mathbf{X}', t-\tau') \times \frac{\partial G_p(\mathbf{X}', \tau')}{\partial \tau'} dV' + G_p \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} f^1_{pq}(\mathbf{X}', t-\tau') \times \frac{\partial G_q}{\partial \tau'}(\mathbf{X}', \tau') dV' + \frac{1}{2} \int_{-\infty}^{t} d\tau' \int_{-\infty}^{t} d\tau'_1 \int_{V-\Sigma} \int_{V-\Sigma} \times f_{pq}(\mathbf{X}', t-\tau'; \mathbf{X}'_1, t-\tau'_1) \times \frac{\partial G_p(\mathbf{X}', \tau')}{\partial \tau'} \frac{\partial G_q(\mathbf{X}'_1, \tau'_1)}{\partial \tau'_1} \times dV'; dV'_1, \tag{7.4}$$

where $G_p \equiv G_p(\mathbf{X}, t)$ constitutive moduli $f^0_p, f^0_{pq}, f_p, f^1_{pq}$, and f_{pq} are also functions of $\mathbf{X}$. As discussed in Sec. IV, $G_p(\mathbf{X}', \tau')$ is assumed to belong to a Hilbert space with an influence function. Consequently,

$$\lim_{\tau' \to \infty} f_p = 0, \quad \lim_{\tau' \to \infty} f^1_{pq} = 0, \quad \lim_{\tau' \to \infty} f_{pq} = \lim_{\tau'_1 \to \infty} f_{pq} = 0. \tag{7.5}$$

Moreover, it is clear from (7.4) that f^0_{pq} and f_{pq} may be considered to possess the symmetry regulations

$$f^0_{pq} = f^0_{pq}, \quad f_{pq}(\mathbf{X}', \tau'; \mathbf{X}'_1, \tau'_1) = f_{qp}(\mathbf{X}'_1, \tau'_1; \mathbf{X}', \tau'). \tag{7.6}$$

If we now calculate the time rate of (7.4) and use (4.17) to (4.20), we obtain

$$T_p = \frac{\partial \dot{F}}{\partial \dot{G}_p} = f^0_p + f^0_{pq} G_q + \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} f^1_{pq} \frac{\partial G_q}{\partial \tau'} dV, \tag{7.7}$$

$$\delta F = \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} f_p \frac{\partial^2 G_p}{\partial \tau'^2} dV' + G_p \times \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} f^1_{pq} \frac{\partial^2 G_p}{\partial \tau'^2} dV' + \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} \int_{V-\Sigma} f_{pq}(\mathbf{X}', t-\tau'; \mathbf{X}'_1, 0) \times \frac{\partial G_p(\mathbf{x}', \tau')}{\partial \tau'} \frac{\partial G_q(\mathbf{X}'_1, t)}{\partial t} dV' dV'_1 + \frac{1}{2} \int_{-\infty}^{t} d\tau' \int_{-\infty}^{t} d\tau'_1 \times \int_{V-\Sigma} \int_{V-\Sigma} f_{pq} \frac{\partial G_p}{\partial \tau'} \frac{\partial G_q}{\partial \tau'_1} dV' dV'_1, \tag{7.8}$$

where, by means of by part integrations, we combined two integrals involving f_p and f^1_{pq} and used (7.6).

An alternative form for (7.7) would result by writing[29]

$$f'_{pq}(\mathbf{X}', t-\tau') = f^1_{pq}(\mathbf{X}', t-\tau') + f^0_{pq}(\mathbf{X}')\delta(\mathbf{X}' - \mathbf{X}), \tag{7.9}$$

where $\delta(\mathbf{X}' - \mathbf{X})$ is the Dirac delta measure

$$T_p = f^0_p + \int_{-\infty}^{t} f'_{pq}(\mathbf{X}', t-\tau') \frac{\partial G_q(\mathbf{X}', \tau')}{\partial \tau'} dV'. \tag{7.10}$$

The quantity f^0_p represents the fields at the natural state of the body. If the body is free of fields and underformed at the natural state, then $f^0_p = 0$.

The explicit forms of (7.10) may now be written as

$$\eta = \eta_0 + \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} \frac{1}{\rho_0} \left(A' \frac{\partial T}{\partial \tau'} + B'_{KL} \frac{\partial E_{KL}}{\partial \tau'} + \bar{\omega}'_K \frac{\partial \mathscr{E}_K}{\partial \tau'} + \Gamma'_K \frac{\partial B_K}{\partial \tau'} \right) dV', \tag{7.11}$$

$${}_E T_{KL} = T_{0KL} + \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} \left(-\bar{B}'_{KL} \frac{\partial T}{\partial \tau'} + \Sigma'_{KLMN} \frac{\partial E_{MN}}{\partial \tau'} - E'_{MKL} \frac{\partial \mathscr{E}_M}{\partial \tau'} - H'_{MKL} \frac{\partial B_M}{\partial \tau'} \right) dV', \tag{7.12}$$

$$\Pi_K = \Pi_{0K} + \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} \left(\bar{\bar{\omega}}'_K \frac{\partial T}{\partial \tau'} + \bar{E}'_{KLM} \frac{\partial E_{LM}}{\partial \tau'} + \chi'^E_{KL} \frac{\partial \mathscr{E}_L}{\partial \tau'} + \Lambda'_{KL} \frac{\partial B_L}{\partial \tau'} \right) dV', \tag{7.13}$$

$$M_K = M_{0K} + \int_{-\infty}^{t} d\tau' \int_{V-\Sigma} \left(\bar{\Gamma}'_K \frac{\partial T}{\partial \tau'} + \bar{H}'_{KLM} \frac{\partial E_{LM}}{\partial \tau'} + \bar{\Lambda}'_{LK} \frac{\partial \mathscr{E}_L}{\partial \tau'} + \chi'^B_{KL} \frac{\partial B_L}{\partial \tau'} \right) dV'. \tag{7.14}$$

The total free energy of the body is obtained by integrating (7.4) over the volume $V - \Sigma$. In this expression, the double volume integral containing f^1_{pq} can be integrated by part with respect to τ', resulting in two separate integrals. From one of these integrals containing $f^1_{pq}(\mathbf{X}', 0, \mathbf{X})$, with the use of (7.9) and an interchange of $\mathbf{X}'$ and $\mathbf{X}$, we deduce that

$$f'_{pq}(\mathbf{X}', 0, \mathbf{X}) = f'_{qp}(\mathbf{X}, 0, \mathbf{X}') \equiv f'^*_{pq}. \tag{7.15}$$

No such general expression can be obtained for $f'_{pq}(\mathbf{X}', t-\tau'; \mathbf{X})$, however, unless we invoke Onsager relations. This assumption is often used in classical (local) theory of viscoelastic solids. If we assume that it is also valid for the nonlocal theory, we will have

$$A' = A'^*, \quad \bar{B}'_{KL} = B'^*_{LK}, \quad \bar{\omega}'_K = \bar{\bar{\omega}}'_K, \quad \bar{\Gamma}'_K = \Gamma'^*_K,$$
$$\Sigma'_{KLMN} = \Sigma'^*_{MNKL}, \quad \bar{E}'_{KLM} = E'^*_{MKL},$$
$$\bar{H}'_{KLM} = H'^*_{MKL}, \tag{7.16}$$
$$\chi'^E_{KL} = \chi'^{*E}_{LK}, \quad \chi'^B_{KL} = \chi'^{*B}_{LK}, \quad \bar{\Lambda}'_{KL} = \Lambda'^*_{LK}.$$

Because of the symmetry of T_{KL} and E_{KL}, we also note the symmetry regulations

$$T_{0KL}=T_{0LK},\quad B'_{KL}=B'_{LK},\quad \overline{B}'_{KL}=\overline{B}'_{LK},$$
$$\Sigma'_{KLMN}=\Sigma'_{LKMN}=\Sigma'_{KLNM},\quad H'_{MKL}=H'_{MLK},\qquad (7.17)$$
$$\overline{E}'_{MKL}=\overline{E}'_{MLK},\quad \overline{H}'_{MKL}=\overline{H}'_{MLK}.$$

For homogeneous materials, constitutive equations must be invariant under translations of the material frame of reference. This implies that the constitutive moduli shall depend on $\mathbf{X}'$ and $\mathbf{X}$ only through $\mathbf{X}'-\mathbf{X}$, i.e.,

$$f^0_p=\text{const},\quad f'_p=f'_p(\mathbf{X}'-\mathbf{X},t-\tau') \qquad (7.18)$$
$$f'_{pq}=f'_{pq}(\mathbf{X}'-\mathbf{X},t-\tau'),$$
$$f_{pq}=f_{pq}(\mathbf{X}'-\mathbf{X},t-\tau';\mathbf{X}'_1-\mathbf{X},t-\tau'_1)$$

In terms of the specific moduli appearing in (7.11) to (7.14), we have for example,

$$T_{0KL}=\text{const},\quad B'_{KL}=B'_{KL}(\mathbf{X}'-\mathbf{X},t-\tau'), \qquad (7.19)$$
$$\Sigma'_{KLMN}=\Sigma'_{KLMN}(\mathbf{X}'-\mathbf{X},t-\tau'),$$
$$\Lambda'_{KL}=\Lambda'_{KL}(\mathbf{X}'-\mathbf{X},t-\tau'),\ldots.$$

Constitutive equations for heat and electric conduction may be expressed as

$$\frac{Q_k}{\theta^2}=\int_{-\infty}^{t}d\tau'\int_{V-\Sigma}\left(\kappa'_{KL}\frac{\partial T_{,L}}{\partial\tau'}+\kappa'^{E}_{KL}\frac{\partial\mathscr{E}_L}{\partial\tau'}\right.$$
$$\left.+\kappa'^{B}_{K}\frac{\partial B_L}{\partial\tau'}+G'_{KLM}\frac{\partial E_{LM}}{\partial\tau'}\right)dV', \qquad (7.20)$$
$$\frac{\mathscr{J}_K}{\theta}=\int_{-\infty}^{t}d\tau'\int_{V-\Sigma}\left(\Sigma'^{T}_{KL}\frac{\partial T_{,L}}{\partial\tau'}+\Sigma'_{KL}\frac{\partial\mathscr{E}_L}{\partial\tau'}\right.$$
$$\left.+\Sigma'^{B}_{KL}\frac{\partial B_L}{\partial\tau'}+\Gamma'_{KLM}\frac{\partial E_{LM}}{\partial\tau'}\right)dV', \qquad (7.21)$$

where the conduction moduli $\kappa'_{KL},\ldots,\Gamma'_{KLM}$ are functions of $\mathbf{X}'$, $\mathbf{X}$, and $t-\tau'$. For homogeneous materials, they depend on $\mathbf{X}'-\mathbf{X}$ and $t-\tau'$. Various moduli appearing in (7.20) represent the heat conduction and those in (7.21) represent the electric conduction due to various fields. Some of these moduli may vanish or be severely restricted due to the second law of thermodynamics.

The second law of thermodynamics (4.22) places restrictions on the constitutive moduli

$$\int_{V-\Sigma}\frac{1}{\theta}\left(\frac{1}{\theta}Q_K\theta_{,K}+\mathscr{J}_K\mathscr{E}_K-\delta F\right)dV\geqslant 0. \qquad (7.22)$$

In the special case of no conduction, we must have

$$-\int_{V-\Sigma}\frac{1}{\theta}\delta F\,dV\geqslant 0. \qquad (7.23)$$

This means that the total dissipation in the body must be non-negative. Employing (7.8), we see that the first two integrals involving f_p and f^1_{pq} are linear in $\partial^2 G_p/\partial\tau'^2$ so that they must vanish, i.e.,

$$\int_{V-\Sigma}(f_p,f^1_{pq})\,dV=0. \qquad (7.24)$$

The third integral in (7.23) involves $f_{pq}(\mathbf{X}',t-\tau';\mathbf{X}'_1,0)$. This integral is also linear in $\partial G_p(\mathbf{X}',\tau')/\partial\tau'$ except for $\tau'=t$. Hence it must vanish. When f_{pq} does not depend on τ', we would have

$$\int_{V-\Sigma}\int_{V-\Sigma}\int_{V-\Sigma}f_{pq}(\mathbf{X}',t;\mathbf{X}'_1,0;\mathbf{X})$$
$$\times G_p(\mathbf{X}',t)G_q(\mathbf{X}'_1,t)\,dV'\,dV'_1\,dV\geqslant 0. \qquad (7.25)$$

For the general case, however, we have

$$\frac{1}{2}\int_{-\infty}^{t}d\tau'\int_{-\infty}^{t}d\tau'_1\int_{V-\Sigma}\int_{V-\Sigma}\int_{V-\Sigma}\dot{f}_{pq}$$
$$\times\frac{\partial G_p}{\partial\tau'}\frac{\partial G_q}{\partial\tau'_1}\,dV'\,dV'_1\,dV\geqslant 0. \qquad (7.26)$$

Both of these inequalities can be made non-negative under the symmetry conditions (7.6) for f_{pq}. For example, a non-negative definite form for f_{pq}, will provide a sufficient condition for (7.25) and (7.26). Thus, if (7.23) is made non-negative, then all we need is to see that

$$\int_{V-\Sigma}\left(\frac{Q_k}{T_0^2}T_{,K}+\frac{1}{T_0}\mathscr{J}_K\mathscr{E}_K\right)dV\geqslant 0. \qquad (7.27)$$

From this and (7.20) and (7.21), it is clear that for fixed $T_{,K}(\mathbf{X},t)$ and $\mathscr{E}_K(\mathbf{X},t)$, the functionals Q_K and $\mathscr{J}_K$ can be varied arbitrarily over the past histories, consequently the conduction moduli must not depend on time. In this case, the integration on τ' can be carried out leading to

$$\frac{Q_K}{T_0^2}=\int_{V-\Sigma}(\kappa'_{KL}T_{,L}+\kappa'^{E}_{KL}\mathscr{E}_L$$
$$+\kappa'^{B}_{KL}B_L+G'_{KLM}E_{LM})\,dV', \qquad (7.28)$$
$$\frac{\mathscr{J}_K}{T_0}=\int_{V-\Sigma}(\Sigma'^{T}_{KL}T_{,L}+\Sigma'_{KL}\mathscr{E}_L+\Sigma'^{B}_{KL}B_L$$
$$+\Gamma'_{KLM}E_{LM})\,dV'. \qquad (7.29)$$

If these are substituted into (7.27), we see that the resulting expression contains terms which are linear in B'_L and E'_{LM}. Integrals containing these terms must vanish for $T_{,K}=0$ or $\mathscr{E}_K=0$. Hence,

$$\int_{V-\Sigma}\int_{V-\Sigma}\kappa'^{B}_{KL}(\mathbf{X}',\mathbf{X})T_{,K}(\mathbf{X})B_L(\mathbf{X}')dV'\,dV=0, \qquad (7.30)$$
$$\int_{V-\Sigma}\int_{V-\Sigma}\Sigma'^{B}_{KL}(\mathbf{X}',\mathbf{X})\mathscr{E}_K(\mathbf{X})B_L(\mathbf{X}')dV'\,dV=0, \qquad (7.31)$$
$$\int_{V-\Sigma}\int_{V-\Sigma}G'_{KLM}(\mathbf{X}',\mathbf{X})T_{,K}(\mathbf{X})E_{LM}(\mathbf{X}')dV'\,dV=0, \qquad (7.32)$$
$$\int_{V-\Sigma}\int_{V-\Sigma}\Gamma'_{KLM}(\mathbf{X}',\mathbf{X})\mathscr{E}_K(\mathbf{X})E_{LM}(\mathbf{X}')dV'\,dV=0, \qquad (7.33)$$
$$\int_{V-\Sigma}\int_{V-\Sigma}[\kappa'_{KL}(\mathbf{X}',\mathbf{X})T_{,K}(\mathbf{X})T_{,L}(\mathbf{X}')$$
$$+\Sigma'_{KL}(\mathbf{X}',\mathbf{X})\mathscr{E}_K(\mathbf{X})\mathscr{E}_L(\mathbf{X}')$$
$$+\kappa'^{E}_{KL}(\mathbf{X}',\mathbf{X})T_{,K}(\mathbf{X})\mathscr{E}_L(\mathbf{X}')$$
$$+\Sigma'^{T}_{KL}(\mathbf{X}',\mathbf{X})\mathscr{E}_K(\mathbf{X})T_{,L}(\mathbf{X}')]dV'\,dV\geqslant 0. \qquad (7.34)$$

It may happen that (7.30)–(7.33) can be satisfied identically, because of the symmetry regulations valid for $\kappa'^{B}_{KL},\ldots,\Gamma'_{KLM}$ (and/or because of some of the fields being absent), without these moduli vanishing point-wise, unlike in the case of local theory.[20]

Furthermore, (7.34) will not be violated when functions κ'_{KL}, Σ'_{KL}, κ'^{E}_{KL}, and Σ'^{T}_{KL} constitute a non-negative, definite

set. In local theory certainly, this is the case.[20] For $\mathscr{E}_K = 0$ if κ'_{KL} is a non-negative form throughout $V - \Sigma$, (7.34) will not be violated.

A similar condition is valid for Σ'_{KL} in the case $T_{,K} = 0$. These conditions are, however, sufficient but not necessary. Much less restricted conditions can be found for the conductivity moduli appearing in (7.34), irrespective of values of $\mathscr{E}_K$ and $T_{,K}$ through $V - \Sigma$. We also note that only the symmetric parts of these moduli contribute to the double integral, so that we take

$$\begin{aligned}
&\kappa'_{KL}(\mathbf{X}', \mathbf{X}) = \kappa'_{LK}(\mathbf{X}, \mathbf{X}'),\\
&\Sigma'_{KL}(\mathbf{X}', \mathbf{X}) = \Sigma'_{LK}(\mathbf{X}, \mathbf{X}'),\\
&\kappa'^{E}_{KL}(\mathbf{X}', \mathbf{X}) + \Sigma'^{T}_{LK}(\mathbf{X}, \mathbf{X}')\\
&= \kappa'^{E}_{LK}(\mathbf{X}, \mathbf{X}') + \Sigma'^{T}_{KL}(\mathbf{X}', \mathbf{X}).
\end{aligned} \tag{7.35}$$

Of course, the material symmetry imposes further restrictions on the conduction moduli.

The spatial forms of constitutive equations are obtained by using (3.5) and

$$\begin{aligned}
&E_{KL} = e_{Kl}x_{k,K}x_{l,L}, \quad R_{KL} = r_{kl}x_{k,K}x_{l,L},\\
&\rho_0/\rho \simeq 1 - e_{rr}, \quad x_{k,K} = (\delta_{MK} + E_{MK} + R_{MK})\delta_{Mk},
\end{aligned} \tag{7.36}$$

where δ_{Mk} is the Kronecker delta when the spatial and material frames are coincident and e_{kl} and r_{kl} are, respectively, the linear strain and rotation measures which are defined in terms of the spatial components u_k of the displacement vector by

$$e_{kl} = \tfrac{1}{2}(u_{k,l} + u_{l,k}), \quad r_{kl} = \tfrac{1}{2}(u_{k,l} - u_{l,k}). \tag{7.37}$$

We also introduce spatial material moduli by

$$\sigma_{kl} = \Sigma_{KL}\delta_{Kk}\delta_{Ll}, \ldots, \lambda_{kl} = \Lambda_{KL}\delta_{Kk}\delta_{Ll} \tag{7.38}$$

in (7.11) to (7.14), and drop nonlinear terms leading to

$$\begin{aligned}
\eta = \eta_0 + \int_{-\infty}^{t} d\tau' \int_{V-\sigma} \frac{1}{\rho_0}\Big(&A' \frac{\partial T}{\partial \tau'} + \beta'_{kL}\frac{\partial e_{kl}}{\partial \tau'}\\
&+ \overline{\omega}'_k \frac{\partial \mathscr{E}_k}{\partial \tau'} + \gamma'_k \frac{\partial B_k}{\partial \tau'}\Big) dv',
\end{aligned} \tag{7.39}$$

$$\begin{aligned}
{}_E T_{kl} = (1 - e_{rr})t_{0kl} &+ t_{0ml}(e_{km} + r_{km}) + t_{0km}(e_{lm} + r_{lm})\\
&+ \int_{-\infty}^{t} d\tau' \int_{V-\sigma} \Big(-\beta'_{kl}\frac{\partial T}{\partial \tau'} + \sigma'_{klmn}\frac{\partial e_{mn}}{\partial \tau'}\\
&- e'_{mkl}\frac{\partial \mathscr{E}_m}{\partial \tau'} - h'_{mkl}\frac{\partial B_m}{\partial \tau'}\Big) dv',
\end{aligned} \tag{7.40}$$

$$\begin{aligned}
P_k = (1 - e_{rr})\pi_{0k} + \pi_{0l}(e_{kl} + r_{kl}) + &\int_{-\infty}^{t} d\tau'\\
\times \int_{V-\sigma} \Big(&\overline{\omega}'_k \frac{\partial T}{\partial \tau'} + \chi'^{E}_{kl}\frac{\partial \mathscr{E}_l}{\partial \tau'} + e'_{klm}\frac{\partial e_{lm}}{\partial \tau'} + \lambda'_{kl}\frac{\partial \mathscr{E}_l}{\partial \tau'}\Big) dv',
\end{aligned} \tag{7.41}$$

$$\begin{aligned}
\mathscr{M}_k = (1 - e_{rr})m_{0k} + m_{0l}(e_{kl} + r_{kl}) + &\int_{-\infty}^{t} d\tau'\\
\times \int_{V-\sigma} \Big(&\gamma'_k \frac{\partial T}{\partial \tau'} + \chi'^{B}_{kl}\frac{\partial B_l}{\partial \tau'} + h'_{klm}\frac{\partial e_{lm}}{\partial \tau'}\\
&+ \lambda'_{lk}\frac{\partial \mathscr{E}_l}{\partial \tau'}\Big) dv'.
\end{aligned} \tag{7.42}$$

Equations of heat and electric conduction may be expressed as

$$q_k = \int_{V-\sigma} (\kappa'_{kl}T_{,l} + \kappa'^{E}_{kl}\mathscr{E}_l + \kappa'^{B}_{kl}B_l + g'_{klm}e_{lm}) dv', \tag{7.43}$$

$$\mathscr{J}_k = \int_{V-\sigma} (\sigma'^{T}_{kl}T_{,l} + \sigma'_{kl}\mathscr{E}_l + \sigma'^{B}_{kl}B_l + \gamma'_{klm}e_{lm}) dv', \tag{7.44}$$

where the conduction moduli $\kappa'_{kl}, \ldots, \gamma'_{klm}$ are functions of $\mathbf{x}'$ and $\mathbf{x}$ and they are subject to the spatial forms of (7.30) to (7.34), i.e.,

$$\int_{V-\sigma}\int_{V-\sigma} \kappa'^{B}_{kl}T_{,k}(\mathbf{x})B_l(\mathbf{x}') dv'\, dv = 0, \tag{7.45}$$

$$\int_{V-\sigma}\int_{V-\sigma} \sigma'^{B}_{kl}\mathscr{E}_k(\mathbf{x})B_l(\mathbf{x}') dv'\, dv = 0, \tag{7.46}$$

$$\int_{V-\sigma}\int_{V-\sigma} g'_{klm}T_{,k}(\mathbf{x})e_{lm}(\mathbf{x}') dv'\, dv = 0, \tag{7.47}$$

$$\int_{V-\sigma}\int_{V-\sigma} \gamma'_{klm}\mathscr{E}_k(\mathbf{x})e_{lm}(\mathbf{x}') dv'\, dv = 0, \tag{7.48}$$

$$\begin{aligned}
\int_{V-\sigma}\int_{V-\sigma} [&\kappa'_{kl}T_{,k}(\mathbf{x})T_{,l}(\mathbf{x}') + \sigma'_{kl}\mathscr{E}_k(\mathbf{x})\mathscr{E}_l(\mathbf{x}')\\
&+ \kappa^{E}_{kl}T_{,k}(\mathbf{x})\mathscr{E}_l(\mathbf{x}') + \sigma'^{T}_{kl}\mathscr{E}_k(\mathbf{x})T_{,l}(\mathbf{x}')] dv'\, dv \geqslant 0
\end{aligned} \tag{7.49}$$

with the symmetry regulations

$$\begin{aligned}
&\kappa'_{kl}(\mathbf{x}', \mathbf{x}) = \kappa'_{lk}(\mathbf{x}, \mathbf{x}'), \quad \sigma'_{kl}(\mathbf{x}', \mathbf{x}) = \sigma'_{lk}(\mathbf{x}, \mathbf{x}'),\\
&\kappa'^{E}_{kl}(\mathbf{x}', \mathbf{x}) + \sigma'^{T}_{lk}(\mathbf{x}, \mathbf{x}') = \kappa'_{lk}(\mathbf{x}, \mathbf{x}') + \sigma'^{T}_{kl}(\mathbf{x}', \mathbf{x}).
\end{aligned} \tag{7.50}$$

For homogeneous materials, conduction moduli are functions of $\mathbf{x}' - \mathbf{x}$.

In the special case of memory-independent materials, the material moduli appearing in (7.39) to (7.42) do not depend on τ'. In this case, the integrals on τ' can be carried out leading to constitutive equations of nonlocal piezoelectricity obtained before.[27]

The physical meanings of various material moduli are the same as in local theory, except that here they are volume densities:

η_0	entropy at the natural state
t_{0kl}	the stress at the natural state
A'	heat capacity
π^0_k	polarization at the natural state
m^0_k	magnetization at the natural state
β'_{kl}	thermal stress moduli
σ'_{klmn}	viscoelastic moduli
e'_{mkl}	piezoelectric moduli
h'_{mkl}	piezomagnetic moduli

$\overline{\omega}'_k$	pyroelectric polarizability
χ'^{E}_{kl}	dielectric susceptibility
λ'_{kl}	magnetic polarizability
γ'_k	pyromagnetic moduli
χ'^{B}_{kl}	magnetic susceptibility
κ'_{kl}	heat conduction moduli
σ'_{kl}	electric conduction moduli
κ'^{E}_{kl}	Peltier moduli
σ'^{T}_{kl}	Seebeck moduli.

In the absence of a better terminology, we christen the following four moduli as:

κ'^{B}_{kl}	thermomagnetic conductivity
g'_{klm}	thermoelastic conductivity
σ'^{B}_{kl}	magnetoelectric conductivity
γ'_{klm}	electroelastic conductivity

assuming that they exist.

The material moduli are Dirac-delta function sequence, so that in the limit when nonlocality in space-time vanishes, these equations revert to classical (local) forms. Note that the material moduli are densities in space-time. Consequently, they depend on a length scale a and a time scale τ so that when $a\to 0$ constitutive equations must revert to classical equations for memory-dependent materials and when $\tau\to 0$, they revert to equations of nonlocal elasticity. When both $a\to 0$ and $\tau\to 0$, we obtain classical theory EM elastic solids.

The internal characteristic length can be taken as the lattice parameter, granular distance, pore size, etc. The characteristic time could be considered as the relaxation time.

If the material possess certain symmetry represented by a group of orthogonal transformations $\{\mathbf{S}\}$, then the material moduli must obey the following types of functional relations:

$$\begin{aligned}
&S_{kp}S_{lq}\chi^{E}_{pq}(\boldsymbol{\kappa},\tau') = \chi^{E}_{kl}(\mathbf{S}\boldsymbol{\kappa},\tau'),\\
&S_{kp}S_{lq}S_{mr}e_{pqr}(\boldsymbol{\kappa},\tau') = e_{klm}(\mathbf{S}\boldsymbol{\kappa},\tau'),\\
&S_{kp}S_{lq}S_{mr}S_{ns}\sigma_{pqrs}(\boldsymbol{\kappa},\theta) = \sigma_{klmn}(\mathbf{S}\boldsymbol{\kappa},\tau'),\\
&S_{kp}S_{lq}\chi^{B}_{pq}(\boldsymbol{\kappa},\tau') = \chi^{B}_{kl}(\mathbf{S}\boldsymbol{\kappa},\tau')\det\mathbf{S}
\end{aligned} \tag{7.51}$$

for all members of the group $\{\mathbf{S}\}$, where $\boldsymbol{\kappa} = \mathbf{x}' - \mathbf{x}$. As a consequence of these, the material moduli will be restricted in their dependence on $\mathbf{x}' - \mathbf{x}$. For example, for the isotropic dielectrics, these imply that

$$\begin{aligned}
\chi'^{E}_{kl} &= \chi^{E}_1\delta_{kl} + \chi^{E}_2\kappa_k\kappa_l,\\
e'_{klm} &= e_1\kappa_k\delta_{lm} + e_2(\kappa_l\delta_{km} + \kappa_m\delta_{kl}) + e_3\kappa_k\kappa_l\kappa_m,\\
\sigma'_{klmn} &= \lambda_1\delta_{kl}\delta_{mn} + \mu_1(\delta_{km}\delta_{ln} + \delta_{kn}\delta_{lm})\\
&\quad + \sigma_1(\kappa_m\kappa_n\delta_{kl} + \kappa_k\kappa_l\delta_{mn})\\
&\quad + \sigma_2(\kappa_k\kappa_m\delta_{ln} + \kappa_k\kappa_n\delta_{ml}\\
&\quad + \kappa_l\kappa_m\delta_{kn} + \kappa_l\kappa_n\delta_{km}) + \sigma_3\kappa_k\kappa_l\kappa_m\kappa_n.
\end{aligned} \tag{7.52}$$

Similar expressions are valid for other moduli. Coefficients $\chi^{E}_1, \chi^{E}_2, e_1,\ldots,\sigma_3$ are functions of $|\mathbf{x}' - \mathbf{x}|$ and τ', e.g.,

$$\chi^{E}_1 = \chi^{E}_1(|\mathbf{x}' - \mathbf{x}|,\tau'),\quad \sigma_\alpha = (|\mathbf{x}' - \mathbf{x}|,\tau'). \tag{7.53}$$

The appearance of the material moduli χ^{1}_{2}, e_α, and σ_α indicates that even for isotropic solids, interatomic orientations cause piezoelectric effects which are missing in classical (local) field theories.

Finally, as noted before, the material moduli depend on characteristic length and time scales, e.g.,

$$\lambda = \lambda\,[(\mathbf{x}' - \mathbf{x})/a, \tau'/\tau], \tag{7.54}$$

so that

$$\begin{aligned}
&\lim_{a\to 0}\lambda = \lambda_m(\tau'/\tau)\delta(\mathbf{x}' - \mathbf{x}),\\
&\lim_{\tau\to 0}\lambda = \lambda_s[(\mathbf{x}' - \mathbf{x})/a]\delta(t - \tau'),\\
&\lim_{\tau\to 0}\lambda_m = \lim_{a\to 0}\lambda_s = \lambda_0,
\end{aligned} \tag{7.55}$$

where λ_0 is a local material constant.

If we also recall the attenuating neighborhood hypothesis as formalized by an influence function, we may employ such forms as

$$\lambda = A\exp[-(k^2/a^2)(\mathbf{x}' - \mathbf{x})\cdot(\mathbf{x}' - \mathbf{x}) - l^2(\tau'^2/\tau^2)], \tag{7.56}$$

where k, l, and A are constants, subject to the normalization

$$\int_0^\infty d\tau'\int_{V-\sigma}\lambda\,dv' = \lambda_0. \tag{7.57}$$

When the body extends to infinity in all directions in N-space dimensions, this gives

$$A = \pi^{-(N+1)/2}(k/a)^N(2l/\tau). \tag{7.58}$$

Of course, other possibilities exist. We may, for example, determine λ' by comparing the dispersion and absorption curves obtained in lattice dynamics with those calculated by means of nonlocal theory.[23–25,30]

VIII. DIELECTRICS

Most dielectrics are nonmagnetizable and all are nonconductors. Thus, the effect of B-field and conductions are ignored. Equations (7.40) and (7.41) are then the only relevant equations to consider. If the natural state is field-free, then we have

$${}_E t_{kl} = \int_{-\infty}^{t} d\tau'\int_{V-\sigma}\left(\sigma'_{klmn}\frac{\partial e_{mn}}{\partial\tau'} - e'_{mkl}\frac{\partial\mathscr{E}_m}{\partial\tau'}\right)dv', \tag{8.1}$$

$$P_k = \int_{-\infty}^{t} d\tau'\int_{V-\sigma}\left(e'_{klm}\frac{\partial e_{lm}}{\partial\tau'} + \chi'^{E}_{kl}\frac{\partial\mathscr{E}_l}{\partial\tau'}\right)dv'. \tag{8.2}$$

For rigid dielectric, the dependence on the strain tensor is ignored and we have the only constitutive equation

$$P_k = \int_{-\infty}^{t} d\tau'\int_{V-\sigma}\chi'^{E}_{kl}(\mathbf{x}' - \mathbf{x}, t - \tau')\frac{\partial\mathscr{E}_l(\mathbf{x}',\tau')}{\partial\tau'}dv'. \tag{8.3}$$

For unbounded solids, the Fourier transform of (8.1) and (8.2) are useful.

$$\begin{aligned}
{}_E\bar{t}_{kl} &= -i\omega[\bar{\sigma}_{klmn}(\boldsymbol{\xi},\omega)\bar{e}_{mn}(\boldsymbol{\xi},\omega) - \bar{e}_{mkl}(\boldsymbol{\xi},\omega)\overline{\mathscr{E}}_m(\boldsymbol{\xi},\omega)],\\
\overline{P}_k &= -i\omega[\bar{e}_{klm}(\boldsymbol{\xi},\omega)\bar{e}_{lm}(\boldsymbol{\xi},\omega) + \overline{\chi}^{E}_{kl}(\boldsymbol{\xi},\omega)\overline{\mathscr{E}}_l(\boldsymbol{\xi},\omega)],
\end{aligned} \tag{8.4}$$

where a superposed bar indicates the Fourier transform, e.g.,

$$\overline{F}(\xi,\omega)=(2\pi)^{-2}\int_0^\infty dt\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\times F(\mathbf{x},t)e^{i(\xi\cdot\mathbf{x}+\omega t)}\,d\mathbf{x}^3. \tag{8.5}$$

Note that the Fourier transform with respect to time is one sided.

Sometimes it is useful to replace χ^E_{kl} by the dielectric moduli ϵ_{kl} defined by

$$\epsilon_{kl}=\delta_{kl}+\chi^E_{kl}. \tag{8.6}$$

From $\overline{D}_k=\bar{\epsilon}_{kl}(\xi,\omega)\overline{E}_l$ and (8.5) it follows that the real **E** will cause real **D** if

$$\bar{\epsilon}_{kl}(\xi,\omega)=\bar{\epsilon}^*_{kl}(-\xi^*,-\omega^*), \tag{8.7}$$

where an asterisk denotes the complex conjugate. If ξ and ω are real, writing $\bar{\epsilon}_{kl}=\bar{\epsilon}'_{kl}+i\bar{\epsilon}''_{kl}$ this leads to

$$\begin{aligned}&\bar{\epsilon}_{kl}(-\xi,-\omega)=\epsilon'_{kl}(\xi,\omega),\\&\epsilon''_{kl}(-\xi,-\omega)=-\epsilon''_{kl}(\xi,\omega),\\&\bar{\epsilon}_{kl}(-\xi,-\omega)=\bar{\epsilon}'_{kl}(\xi,\omega)-i\bar{\epsilon}''_{kl}(\xi,\omega).\end{aligned} \tag{8.7a}$$

Similar identities are valid for $\bar{\sigma}_{klmn}$ and $\bar{e}_{klm}$.

When the memory dependence is negligible then the material moduli do not depend on τ' and (8.1) and (8.2) become

$$E t_{kl}=\int_{V-\sigma}[\sigma_{klmn}(\mathbf{x}'-\mathbf{x})e_{mn}(\mathbf{x}',t)-e_{mkl}(\mathbf{x}'-\mathbf{x})\mathscr{E}_m(\mathbf{x}',t)]dv', \tag{8.8}$$

$$P_k=\int_{V-\sigma}[e_{klm}(\mathbf{x}'-\mathbf{x})e_{lm}(\mathbf{x}',t)+\chi^E_{kl}(\mathbf{x}'-\mathbf{x})\mathscr{E}_l(\mathbf{x}',t)]dv'. \tag{8.9}$$

These are identical to those obtained in our previous work.[27]

Isotropic dielectrics

Frequently, the analysis in the Fourier domain is used for analytical and experimental purposes. The material moduli in the (ξ,ω) domain can be expressed by merely replacing κ in (7.52) by ξ, specifically,

$$\bar{\epsilon}_{kl}=\left(\delta_{kl}-\frac{\xi_k\xi_l}{\xi^2}\right)\epsilon_T(\xi^2,\omega)+\frac{\xi_k\xi_l}{\xi^2}\epsilon_L(\xi^2,\omega), \tag{8.10}$$

$$\bar{e}_{klm}=\xi^{-1}(\gamma_1\xi_k\delta_{lm}+\gamma_2\xi_l\delta_{km}+\gamma_2\epsilon_m\delta_{kl})+\gamma_3\xi^{-3}\xi_k\xi_l\xi_m, \tag{8.11}$$

$$\begin{aligned}\bar{\sigma}_{klmn}=&\bar{\lambda}\delta_{kl}\delta_{mn}+\bar{\mu}(\delta_{km}\delta_{ln}+\delta_{kn}\delta_{lm})\\&+\lambda_1\xi^{-2}(\xi_m\xi_n\delta_{kl}+\xi_k\xi_l\delta_{mn})\\&+\lambda_2\xi^{-2}(\xi_k\xi_m\delta_{ln}+\xi_k\xi_n\delta_{lm}\\&+\xi_l\xi_m\delta_{kn}+\xi_l\xi_n\delta_{km})\\&+\lambda_3\xi^{-4}\xi_k\xi_l\xi_m\xi_n,\end{aligned} \tag{8.12}$$

where ϵ_T and ϵ_L are the transverse and longitudinal dielectric moduli. γ_α, $\bar{\lambda}$, $\bar{\mu}$, and λ_α are functions of $\xi^2\equiv\xi\cdot\xi$ and ω only. These expressions are identical to the forms given before[27] except that here the material moduli depend on ω as well.

The dependence of $\bar{\epsilon}_{kl}$ on the wave vector ξ indicates the space dispersion and ω the presence of the longitudinal optical modes. In fact, as conjectured in our previous work, the dispersion relation for the longitudinal optical modes are given by

$$\epsilon_L(\xi^2,\omega)=0 \tag{8.13}$$

so that the phase velocity $v=\omega/\xi$ and the index of refraction $n=c/v$ depends on the wave length $2\pi/\xi$ indicating dispersion, which is particularly strong for high frequencies or short wavelengths.

From (8.4), it is also clear that isotropic solids may exhibit stress-optic effects since $\bar{e}_{mkl}\neq 0$. This is particularly true, again in the short-wavelength region, i.e., near the boundaries of the Brillouin zone. Thus, isotropic solids at these boundaries can be spatially active and/or *gyrotropic*. Such effects are, of course, ruled out in classical theories. Equation (8.4) indicate Brillouin scattering from an exciton which have been observed.

For anisotropic dielectrics material functions $\bar{\epsilon}_{kl}$, $\bar{e}_{klm}$, and $\bar{\sigma}_{klmn}$ acquire other forms compatible with their group symmetry. For example, for an uniaxial crystal with axis $\mathbf{i}_3$, $\bar{\epsilon}_{kl}(\xi,\omega)$ is of the form[27]

$$\bar{\epsilon}_{kl}=\left(\delta_{kl}-\frac{\xi_k\xi_l}{\xi^2}\right)\epsilon_T+\frac{\xi_k\xi_l}{\xi^2}\epsilon_L+\epsilon_0\delta_{3k}\delta_{3l}+\epsilon_R\xi^{-1}(\xi_k\delta_{3l}+\xi_l\delta_{3k}), \tag{8.14}$$

where ϵ_T, ϵ_L, ϵ_0, and ϵ_R are functions of ξ^2, ω, and T.

When the spatial dispersion is weak $\xi a\ll 1$, $\bar{\epsilon}_{kl}$ may be approximated by a power series expansion of the form

$$\bar{\epsilon}_{kl}(\xi,\omega)=\bar{\epsilon}^0_{kl}(\omega)+i\bar{\gamma}_{klm}(\omega)\xi_m+\bar{\alpha}_{klmn}(\omega)\xi_m\xi_n, \tag{8.15}$$

similar expansions being valid for $\bar{e}_{klm}$ and $\bar{\sigma}_{klmn}$. In a space-time domain this is equivalent to the operator

$$\epsilon_{kl}(\mathbf{x},t)=\epsilon^0_{kl}(t)-\gamma_{klm}(t)\frac{\partial}{\partial x^m}-\alpha_{klmn}(t)\frac{\partial}{\partial x^m}\frac{\partial}{\partial x^n}. \tag{8.16}$$

Other approximate forms have been suggested,[31] e.g.,

$$\bar{\epsilon}_{kl}(\xi,\omega)=\epsilon^0_{kl}+g_{kl}/[\omega-\omega_L-F(\xi)], \tag{8.17}$$

where ϵ^0_{kl}, g_{kl}, ω_L are constant and $F(\xi)$ is a suitable function of ξ, which may be approximated by a polynomial in ξ.

Rigid gyrotropic crystals possess no center of symmetry so that for these crystals one can neglect quadratic terms in powers of ξ also. The presence of linear terms in ξ, near the absorption lines, give rise to a new type of wave which is absent in classical theories.[31]

The complex frequency dependence of the material moduli can be used to study absorption of waves. In the case of electric conduction, the nonlocality and memory effects are very important and they lead to highly damped EM waves and the anomalous skin effect. Thus, we can employ the nonlocal theory to predict various results associated with electronic conduction in a lattice.

Eventually, by means of statistical mechanics, it should be possible to calculate the nonlocal moduli in terms of atomic parameters and force laws. Thus far, this has been possible only for very simple cases. Of course, experimental results can be used to find approximate expressions.

IX. INFRA-RED DISPERSION AND LATTICE VIBRATIONS

Here, we consider the interaction of electromagnetic waves with lattice vibrations. It is well known that acoustic branches of lattice vibrations are identical to those of elastic waves in the long-wavelength limit. However, the optical branches of lattice vibrations are not included in classical elasticity. It will be shown that both acoustic and optical branches are included in the nonlocal theory. Therefore, it should be possible to study the interactions of EM waves with acoustic and optical branches. Consequently, infrared dispersion and the retardation effect on lattice vibrations can be accounted for in the range of infrared frequencies. Moreover, when the memory dependence is considered, the damping of waves and the energy dissipation can be calculated. The dissipative effects are important, especially in the neighborhood of resonance frequencies. Initially, we assume that the memory dependence is absent. The magnetization, conductions, and body forces will be assumed to be negligible.

Fourier transforms of Maxwell's equations and equations of motion are given by

$$\boldsymbol{\xi}\cdot\overline{\mathbf{D}}=0, \tag{9.1}$$

$$\boldsymbol{\xi}\times\overline{\mathbf{E}}+(\omega/c)\,\overline{\mathbf{H}}=\mathbf{0}, \tag{9.2}$$

$$\boldsymbol{\xi}\cdot\overline{\mathbf{H}}=0, \tag{9.3}$$

$$\boldsymbol{\xi}\times\overline{\mathbf{H}}-(\omega/c)\,\overline{\mathbf{D}}=\mathbf{0}, \tag{9.4}$$

$$i\xi_k\,{}_E\bar{t}_{kl}-\rho\omega^2\bar{u}_l=0. \tag{9.5}$$

Constitutive equations for ${}_E\bar{t}_{kl}$ and $\overline{D}_k$ follow from (8.8) to (8.12)

$$\begin{aligned}{}_E\bar{t}_{kl}=&-i\{[(\bar{\lambda}+\lambda_1)\delta_{kl}\\&+\xi^{-2}(\lambda_1+2\lambda_2+\lambda_3)\xi_k\xi_l]\bar{\mathbf{u}}\cdot\boldsymbol{\xi}\\&+(\bar{\mu}+\lambda_2)(\xi_l\bar{u}_k+\xi_k\bar{u}_l)\}+\xi^{-1}[\gamma_2(E_l\xi_k+E_k\xi_l)\\&+\boldsymbol{\xi}\cdot\mathbf{E}(\gamma_1\delta_{kl}+\gamma_3\xi^{-2}\xi_k\xi_l)],\end{aligned} \tag{9.6}$$

$$\begin{aligned}\overline{D}_k=&\epsilon_T\overline{E}_k+(\epsilon_L-\epsilon_T)\xi^{-2}\xi_k\overline{\mathbf{E}}\cdot\boldsymbol{\xi}\\&-i\xi^{-1}[(\gamma_1+\gamma_2+\gamma_3)\\&\times\xi_k\bar{\mathbf{u}}\cdot\boldsymbol{\xi}+\gamma_2\xi^2\bar{u}_k],\end{aligned} \tag{9.7}$$

where a superposed bar represents the Fourier transform. Substituting these into (9.1) and (9.5), we have

$$\bar{\epsilon}_L\mathbf{E}\cdot\boldsymbol{\xi}-i\xi(\gamma_1+\gamma_2+\gamma_3)\bar{\mathbf{u}}\cdot\boldsymbol{\xi}=0, \tag{9.8}$$

$$\begin{aligned}&[\rho\omega^2-(\bar{\mu}+\lambda_2)\xi^2]\bar{u}_l\\&-(\bar{\lambda}+\bar{\mu}+2\lambda_1+3\lambda_2+\lambda_3)\xi_l\bar{\mathbf{u}}\cdot\boldsymbol{\xi}\\&-i\xi^{-1}[\gamma_2\xi^2\overline{E}_l+(\gamma_1+\gamma_2+\gamma_3)\xi_l\overline{\mathbf{E}}\cdot\boldsymbol{\xi}]=0.\end{aligned} \tag{9.9}$$

Scalar product of (9.9) with $\boldsymbol{\xi}$ gives

$$\begin{aligned}&-i\xi(\gamma_1+2\gamma_2+\gamma_3)\overline{\mathbf{E}}\cdot\boldsymbol{\xi}\\&+[\rho\omega^2-(\bar{\lambda}+2\bar{\mu}+2\lambda_1+4\lambda_2+\lambda_3)\xi^2]\bar{\mathbf{u}}\cdot\boldsymbol{\xi}=0.\end{aligned} \tag{9.10}$$

Equations (9.8) and (9.10) may have nonvanishing solutions for $\overline{\mathbf{E}}\cdot\boldsymbol{\xi}$ and $\bar{\mathbf{u}}\cdot\boldsymbol{\xi}$ if the determinant of their coefficients vanishes, i.e.,

$$\rho\bar{\epsilon}_L=\gamma^2[c_1^2-(\omega/\xi)^2]^{-1}, \tag{9.11}$$

where

$$c_1^2=(\bar{\lambda}+2\bar{\mu}+2\lambda_1+4\lambda_2+\lambda_3)/\rho, \tag{9.12}$$

$$\gamma=\gamma_1+2\gamma_2+\gamma_3.$$

Equation (9.11) is the dispersion relations for waves which possess longitudinal components $\overline{\mathbf{E}}\cdot\boldsymbol{\xi}$ and $\bar{\mathbf{u}}\cdot\boldsymbol{\xi}$. In the classical limit, $\bar{\epsilon}_L=0$ and $\gamma\neq0$, we have the dispersion relations of the irrotational modes of elastic waves.

If on the other hand, $\overline{\mathbf{E}}\cdot\boldsymbol{\xi}=0$, then it follows from (9.8) that, for $\gamma_1+\gamma_2+\gamma_3\neq0$, $\bar{\mathbf{u}}\cdot\boldsymbol{\xi}=0$. In this case, we have transverse waves only, and (9.6), (9.7), and (9.9) reduce to

$${}_E\bar{t}_{kl}=-i(\bar{\mu}+\lambda_2)(\xi_l\bar{u}_k+\xi_k\bar{u}_l)+\xi^{-1}\gamma_2(\overline{E}_l\xi_k+\overline{E}_k\xi_l), \tag{9.13}$$

$$\overline{D}_k=\epsilon_T\overline{E}_k-i\gamma_2\xi\bar{u}_k, \tag{9.14}$$

$$[\rho\omega^2-(\bar{\mu}+\lambda_2)\xi^2]\bar{u}_l-i\xi\gamma_2\overline{E}_l=\mathbf{0}. \tag{9.15}$$

Eliminating $\bar{\mathbf{u}}$ between (9.14) and (9.15), we have

$$\overline{D}=\epsilon\overline{E}_k, \tag{9.16}$$

where

$$\epsilon=\epsilon_T-\rho^{-1}\gamma^2\xi^2\{[(\bar{\mu}+\lambda_2)/\rho]\xi^2-\omega^2\}^{-1}. \tag{9.17}$$

This relation indicates clearly the frequency dependence of the dielectric constant, consequently, the dependence of the refraction of monochromatic wave on its frequency (the dispersion).

Comparison of this result with the classical treatment of photon–phonon interaction[32] shows that

$$\begin{aligned}&\epsilon_\infty=\epsilon_T,\quad \epsilon_0-\epsilon_\infty=-\gamma^2/(\bar{\mu}+\lambda_2),\\&\omega_0^2=[(\bar{\mu}+\lambda_2)/\rho]\,\xi^2.\end{aligned} \tag{9.18}$$

In the classical treatment ϵ_∞, ϵ_0, and ω_0^2 are considered constants. Here they are functions of ξ^2 so that the space dispersion is included. Of course, in some region of infrared frequencies, ξ-dependence is negligible and we may consider them contants in that region. With the identification (9.18), the nonlocal material moduli are determined since classical values of ϵ_∞, ϵ_0, and ω_0^2 are measured for various materials (e.g., alkali-halides).

Cross product of (9.4) with $\boldsymbol{\xi}$, using (9.2) and (9.7), leads to

$$[\epsilon_T(\omega^2/c^2)-\xi^2]\overline{\mathbf{H}}+i\gamma_2(\omega/c)\xi(\boldsymbol{\xi}\times\bar{\mathbf{u}})=\mathbf{0}. \tag{9.19}$$

Cross product of (9.9) with $\boldsymbol{\xi}$, with the use (9.2), gives

$$i\gamma_2(\omega/c)\xi\overline{\mathbf{H}}+[\rho\omega^2-(\bar{\mu}+\lambda_2)\xi^2]\boldsymbol{\xi}\times\bar{\mathbf{u}}=\mathbf{0}. \tag{9.20}$$

If (9.19) and (9.20) are to possess nonzero solutions for $\overline{\mathbf{H}}$ and $\boldsymbol{\xi}\times\bar{\mathbf{u}}$, we must have

$$\omega^4-\left(\frac{c^2}{\epsilon_T}+\frac{\bar{\mu}+\lambda_2}{\rho}-\frac{\gamma_2^2}{\rho\epsilon_T}\right)\xi^2\omega^2+\frac{\bar{\mu}+\lambda_2}{\rho}\frac{c^2}{\epsilon_T}\xi^4=0. \tag{9.21}$$

This is the dispersion relations for optical modes which possess transverse components $\boldsymbol{\xi}\times\bar{\mathbf{u}}$, $\boldsymbol{\xi}\times\overline{\mathbf{E}}=-\omega\overline{\mathbf{H}}/c$. They are also valid when $\boldsymbol{\xi}\cdot\overline{\mathbf{E}}=\boldsymbol{\xi}\cdot\bar{\mathbf{u}}=0$.

Note that the optical modes are brought into play through the presence of the space gradients of the displace-

ment vector, indicated by the presence of the operators ξ_k in the constitutive equation (9.6) and (9.7). In this way, terms involving λ_k and γ_k give rise to higher-order space gradients of the displacement and electric vectors since in the physical space, ξ_k corresponds to $-i\partial/\partial x_k$. Thus, for example, if we set $\gamma_2 = \lambda_2 = 0$ in (9.21), we can factor this equation leading to $(\omega^2 - c^2/\epsilon_T)(\omega^2 - \bar{\mu}/\rho) = 0$, which give classical dispersion relations of EM and irrotational elastic waves in the long wave limit (when c^2/ϵ_T and $\bar{\mu}/\rho$ are independent of $\boldsymbol{\xi}$ and ω).

If we use (9.18), Eq. (9.21) can be expressed as

$$\frac{\xi^2 c^2}{\omega^2} = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\infty}{\omega_0^2 - \omega^2}\omega_0^2, \tag{9.22}$$

which is identical in form to the classical result.[32]

If $\overline{\mathbf{H}} = \mathbf{0}$, then from (9.19) it follows that $\boldsymbol{\xi} \times \bar{\mathbf{u}} = \mathbf{0}$ and the waves will be longitudinal only. In this case, we have the dispersion relations (9.11).

In dispersion relations (9.11), (9.17) and (9.21) and (9.22), constitutive moduli are functions of ξ^2, when the memory dependence is neglected. When the memory dependence is included, the foregoing equations are modified by multiplying all material moduli by $-i\omega$. This factor arises from the time rates of e_{kl} and $\mathscr{E}_k$ in (8.1) and (8.2). In this case, the material moduli are functions of ξ^2 and ω so that we will have complex roots for $\omega = \omega(\xi)$, indicating dispersion with absorption. Thus, the so-called polariton dispersion with absorption is fully accounted for.

X. NATURAL OPTICAL ACTIVITY

The dependence of the dielectric moduli $\bar{\epsilon}_{kl}$ on the wave vector can give rise to optical activity. For this to occur, the crystal must not possess a center of symmetry.

The nonlocal constitutive equation for rigid anisotropic dielectrics is given by

$$P_k(\mathbf{x}, t) = \int_{-\infty}^{t} d\tau' \int_{\mathscr{V}} \chi^E_{kl}(\mathbf{x}' - \mathbf{x}, t - \tau') \frac{\partial E_l(\mathbf{x}', \tau')}{\partial \tau'} dv(\mathbf{x}'). \tag{10.1}$$

The total free energy, in the absence of other fields, is of the form

$$\int_{-\infty}^{t} d\tau \int_{\mathscr{V}} P_k(\mathbf{x}, \tau) \frac{\partial E_k(\mathbf{x}, \tau)}{\partial \tau} dv(\mathbf{x}). \tag{10.2}$$

If we carry (10.1) into (10.2), interchange k and l and the order of the integrals on $(\mathbf{x}, \tau)$ and $(\mathbf{x}', \tau')$, we see that we must have [see also (7.16)]

$$\chi^E_{kl}(\mathbf{x}' - \mathbf{x}, \tau - \tau') = \chi^E_{lk}(\mathbf{x} - \mathbf{x}', \tau' - \tau). \tag{10.3}$$

In terms of dielectric moduli, this is equivalent to

$$\epsilon_{kl}(\mathbf{x}' - \mathbf{x}, \tau - \tau') = \epsilon_{lk}(\mathbf{x} - \mathbf{x}', \tau' - \tau). \tag{10.4}$$

The Fourier transform of ϵ_{kl} is of the form

$$\bar{\epsilon}_{kl}(\boldsymbol{\xi}, \omega) = \bar{\epsilon}_{lk}(-\boldsymbol{\xi}, -\omega). \tag{10.5}$$

But from (8.7) we also have

$$\bar{\epsilon}_{kl}(\boldsymbol{\xi}, \omega) = \overset{*}{\bar{\epsilon}}_{kl}(-\overset{*}{\boldsymbol{\xi}}, -\overset{*}{\omega}). \tag{10.6}$$

If we expand (10.5) and (10.6) into power series about $\boldsymbol{\xi} = \mathbf{0}$ and $\overset{*}{\boldsymbol{\xi}} = \mathbf{0}$ and retain only the first two terms, we will have

$$\begin{aligned}\bar{\epsilon}_{kl}(\xi, \omega) &= \bar{\epsilon}_{kl}(0, \omega) + \bar{\epsilon}_{kl,m}(0, \omega)\xi_m \\ &= \bar{\epsilon}_{lk}(0, -\omega) - \bar{\epsilon}_{lk,m}(0, -\omega)\xi_m \\ &= \overset{*}{\bar{\epsilon}}_{kl}(0, -\overset{*}{\omega}) - \overset{*}{\bar{\epsilon}}_{kl,m}(0, -\overset{*}{\omega})\overset{*}{\xi}_m.\end{aligned} \tag{10.7}$$

Consequently,

$$\begin{aligned}&\bar{\epsilon}_{kl}(0, \omega) = \bar{\epsilon}_{lk}(0, -\omega) = \overset{*}{\bar{\epsilon}}_{kl}(0, -\overset{*}{\omega}), \\ &\bar{\epsilon}_{kl,m}(0, \omega) = -\bar{\epsilon}_{lk,m}(0, -\omega).\end{aligned} \tag{10.8}$$

This approximation is equivalent to

$$D_k = \epsilon^0_{kl}(\omega)E_l + \gamma_{klm}(\omega)\frac{\partial E_l}{\partial x_m}, \tag{10.9}$$

where

$$\epsilon^0_{kl}(\omega) = \bar{\epsilon}_{kl}(0, \omega), \quad \gamma_{klm}(\omega) = i\bar{\epsilon}_{kl,m}(0, \omega) \tag{10.10}$$

with the symmetry relations

$$\begin{aligned}&\epsilon^0_{kl}(\omega) = \epsilon^0_{lk}(-\omega) = \overset{*}{\epsilon}{}^0_{kl}(-\overset{*}{\omega}), \\ &\gamma_{klm}(\omega) = -\gamma_{lkm}(-\omega) = \overset{*}{\gamma}_{klm}(-\overset{*}{\omega}).\end{aligned} \tag{10.11}$$

Equation (10.9) is identical to the classical expression of the constitutive equation leading to natural optical activity if the memory dependence is ignored, i.e., $\omega = 0$ (cf. Landau and Lifschitz, Ref. 2, p. 338).

For memory-dependent materials, using (10.11), the inverse Fourier transform of (10.9) with respect to ω will show that the dielectric displacement $\mathbf{D}$ will depend on the memory of past electric fields. In this case, the crystal will exhibit absorption as well.

XI. ANOMALOUS SKIN EFFECT

Anomalous skin effect arises when a highly nonuniform field is established in a conductor. In this case the electric field varies rapidly within the skin depth and the nonlocal effects become important.

Consider a semi-infinite isotropic rigid conductor occupying the space $|x_1| \leqslant \infty\ |x_2| \leqslant \infty$, $0 \leqslant x_3$. We assume that the displacement current, magnetization and, temperature gradient are negligible. In this case the Fourier transforms of Maxwell's equation reduces to

$$\boldsymbol{\xi} \times \overline{\mathbf{E}} + (\omega/c)\mathbf{H} = \mathbf{0}, \quad \boldsymbol{\xi} \times \overline{\mathbf{H}} = (i/c)\bar{\mathbf{J}}, \tag{11.1}$$

where $\bar{\mathbf{J}}$ is given by the constitutive equation (7.44), i.e.,

$$\bar{J}_k = \bar{\sigma}_{kl}\bar{E}_l. \tag{11.2}$$

The Fourier transform of the conduction moduli σ_{kl}, for isotropic solids, are of the form

$$\bar{\sigma}_{kl} = \sigma_0\delta_{kl} + \sigma_1\xi^{-2}\xi_k\xi_l, \quad \xi^3 \equiv \boldsymbol{\xi} \cdot \boldsymbol{\xi}, \tag{11.3}$$

where σ_0 and σ_1 are functions of ξ^2 and ω. Eliminating $\overline{\mathbf{H}}$ and $\bar{\mathbf{J}}$ among (11.1) and (11.2) we obtain

$$(\xi_k\xi_l - \xi^2\delta_{kl} + i(\omega/c^2)\bar{\sigma}_{kl})\bar{E}_l = 0. \tag{11.4}$$

For the semi-infinite conductor under consideration, we take the applied E-field in x_1-direction and assume that EM fields are independent of x_2-coordinate. Consequently

$$\xi_k = \xi_3\delta_{k3}, \quad E_k = E_1\delta_{k1}, \quad H_k = H_2\delta_{k2}, \quad J_k = J_1\delta_{k1}, \tag{11.5}$$

with this (11.4) gives

$$i(\omega/c^2)\bar{\sigma}_{11}(\xi_3^2, \omega) - \xi_3^2 = 0. \tag{11.6}$$

When the memory effect is absent then $\bar{\sigma}_{11} = \bar{\sigma}_{11}(\xi_3^2)$ is a function of ξ_3^2 only. From (11.6) it is clear that the penetration depth depends on the frequency. This result can be used to determine σ_{11} by comparing (11.6) with atomic calculations, or surface impedance measurements.

The integro-differential equation for E_1 corresponding to the case (11.5) is given by

$$\frac{d^2E_1}{dx_3^2} = -\frac{i\omega}{c^2}\int_{-\infty}^{\infty} \sigma_{11}(x_3 - x_3', \omega)E_1(x_3')dx_3'. \tag{11.7}$$

Comparing this with the Reuter and Sondheimer's result,[33] for the case of specular surface reflexion, we see that

$$\sigma_{11}(x, \omega) = \sigma_0\{Ei_1[(1 - i\omega\tau)|x|] - Ei_3[(1 - i\omega\tau)|x|]\}, \tag{11.8}$$

where σ_0 and τ are constants and

$$Ei_n(x) = \int_1^{\infty} s^{-n}e^{-sx}\,ds. \tag{11.9}$$

In this way we establish a definite connection with the electron theory.

Finally note that when $\bar{\sigma}_{11}$ is independent of ξ_3 and ω, (11.6) gives the classical skin depth.

XII. SUPERCONDUCTIVITY

Below a critical temperature ranging from less than 1 °K to 18 °K large numbers of metals and alloys are superconductors. At this stage there is no resistance to the electric field inside of the metal. Here we show that the superconductivity is included in the nonlocal theory.

Two surviving Maxwell's equations have the form

$$\nabla\times\mathbf{H} = \mathbf{J}/c, \quad \nabla\cdot\mathbf{H} = 0. \tag{12.1}$$

Of course $\mathbf{H}$ and $\mathbf{J}$ are interpreted as the microscopic fields. For simplicity, we consider a rigid superconductor at a constant temperature and ignore the memory effects. From (7.44) it then follows that

$$J_k = \int_{\mathscr{V}-\sigma} \sigma_{kl}(\mathbf{x}' - \mathbf{x}, T)H_l(\mathbf{x}')dv(\mathbf{x}'), \tag{12.2}$$

where we wrote $\sigma_{kl}^B \equiv \sigma_{kl}$. Since $E_k = 0$, it is clear that the entropy inequality will not be violated. However, $\mathbf{J}$ is a polar vector, while $\mathbf{H}$ is an axial vector. Consequently σ_{kl} must be an axial tensor. This is possible since σ_{kl} depends on the vector $\mathbf{x}' - \mathbf{x}$. For an isotropic solid σ_{kl} has the form (Ref. 20, Appendix B)

$$\sigma_{kl} = \sigma_0\epsilon_{kml}(x_m - x_m'), \tag{12.3}$$

where σ_0 is a function of $|\mathbf{x} - \mathbf{x}'|$ and T. Consequently for isotropic materials (12.2) reads

$$\mathbf{J} = \int_{\mathscr{V}-\sigma} \sigma_0(|\mathbf{x}' - \mathbf{x}|, T)(\mathbf{x}' - \mathbf{x})\times\mathbf{H}(\mathbf{x}')dv(x'). \tag{12.4}$$

Carrying (12.4) into (12.1) we have

$$c\nabla\times\mathbf{H} = \int_{\mathscr{V}-\sigma} \sigma_0(R, T)\mathbf{R}\times\mathbf{H}'\,dv', \tag{12.5}$$

where we wrote $\mathbf{R} = \mathbf{x}' - \mathbf{x}$. This equation together with the second equation in (12.1) and the boundary conditions

$$\mathbf{H}\cdot\mathbf{n} = 0, \quad [\mathbf{H}\times\mathbf{n}] = \mathbf{0}, \quad \partial\mathscr{V} \tag{12.6}$$

are adequate to determine $\mathbf{H}$ when the kernel $\sigma_0(R, T)$ is known.

We now show that (12.5), in a special case, gives London's equation of superconductivity. In fact, we calculate curl of $\mathbf{J}$:

$$(\nabla\times\mathbf{J})_i = \int_{\mathscr{V}} \epsilon_{ijk}\epsilon_{klm}(\sigma_0R_l)_{,j}H_m'\,dv' = \int_{\mathscr{V}} [(\sigma_0R_i)_{,j}H_j' - (\sigma_0R_j)_{,j}H_i']dv'.$$

In the first term of the integrand we wrote $\partial/\partial x_j = -\partial x_j'$ so that

$$(\sigma_0R_i)_{,j} = -(\sigma_0R_iH_j')_{,j} + \sigma_0R_iH_{j,j}'.$$

Here the second term vanishes on account of (12.1) and the first term can be converted to a surface integral by means of the Green–Gauss theorem. Hence,

$$\nabla\times\mathbf{J} = -\int_{\partial\mathscr{V}} \sigma_0\mathbf{R}\mathbf{H}'\cdot d\mathbf{a}' - \int_{\mathscr{V}} \mathbf{H}'\nabla\cdot(\sigma_0\mathbf{R})dv'.$$

The integrand in the surface integral on $\partial\mathscr{V}$ vanishes because of (12.6) and we obtain

$$\nabla\times\mathbf{J} = -\int_{\mathscr{V}} \gamma(R, T)\mathbf{H}'\,dv', \tag{12.7}$$

where

$$\gamma(R, T) = \nabla\cdot(\sigma_0\mathbf{R}). \tag{12.8}$$

Combining (12.7) and (12.1), we obtain

$$\nabla\times\nabla\times\mathbf{H} + \frac{1}{c}\int_{\mathscr{V}} \gamma(R, T)\mathbf{H}(\mathbf{x}')dv' = \mathbf{0}. \tag{12.9}$$

This integro-differential equation reverts to London's equation in the classical limit when the $\gamma(R, T)$ becomes a Dirac delta measure. This is in perfect accord with the conditions (7.55) set on nonlocal kernels namely that they must be a Dirac delta sequence. Consequently, $\gamma\to\gamma_0(T)\delta(|\mathbf{x}' - \mathbf{x}|)$ in the limit when the internal characteristic length approaches zero, and (12.9) converts to

$$\nabla\times\nabla\times\mathbf{H} + (\gamma_0/c)\mathbf{H} = 0, \tag{12.10}$$

which is the London's equation with γ_0 appropriately identified.[5]

We now introduce vector potential $\mathbf{A}$ by

$$\mathbf{H} = \nabla\times\mathbf{A}, \quad \nabla\cdot\mathbf{A} = 0. \tag{12.11}$$

With this the second equation in (12.1) is satisfied and (12.4) may be written as

$$J_k = \int_{\mathscr{V}} \sigma_0\epsilon_{klm}\epsilon_{mrs}R_lA_{s,r}'\,dv' = \int_{\mathscr{V}} [(\sigma_0\epsilon_{klm}\epsilon_{mrs}R_lA_s)_{,r} - (\epsilon_{klm}\epsilon_{mrs}\sigma_0R_l)_{,r}A_s']dv'.$$

By means of the Green–Gauss theorem the first part of the volume integral is converted to a surface integral so that

$$\mathbf{J}=\int_{\partial\mathscr{V}}\sigma_0(\mathbf{R}\cdot\mathbf{A}'\,da'-\mathbf{R}\cdot d\mathbf{a}'\,\mathbf{A}')$$
$$-\int_{\mathscr{V}}\left[\frac{\partial\sigma_0}{\partial R}\mathbf{R}\cdot\mathbf{A}'\frac{\mathbf{R}}{R}-\left(R\frac{\partial\sigma_0}{\partial R}+2\sigma_0\right)\mathbf{A}'\right]dv'. \tag{12.12}$$

Clearly the functions σ_0 must be such that at large distance (large $R=|\mathbf{x}'-\mathbf{x}|$) from $\mathbf{x}$ the surface integral vanishes or is negligibly small. To assure this we take

$$\sigma_0=(C/2R^2)e^{-R/\xi_0},\quad R<\xi_0, \tag{12.13}$$

where C and ξ_0 are constants. With this (12.12) reduces to

$$\mathbf{J}=C\int_{\mathscr{V}}\left\{\frac{(\mathbf{A}'\cdot\mathbf{R})\mathbf{R}}{R^4}\right.$$
$$\left.+\frac{1}{2\xi_0 R}\left[\mathbf{A}'\cdot\mathbf{R}\frac{\mathbf{R}}{R^2}-\left(1+\frac{4\xi_0}{R}\right)\mathbf{A}'\right]\right\}e^{-R/\xi_0}\,dv'. \tag{12.14}$$

For large ξ_0 the second term may be neglected, resulting in the expression first given by Pippard[3]

$$\mathbf{J}=C\int_{\mathscr{V}}\frac{(\mathbf{A}'\cdot\mathbf{R})\mathbf{R}}{R^4}e^{-R/\xi_0}\,dv'. \tag{12.15}$$

Coefficient C is determined by considering that $\mathbf{A}$ varies slowly over a distance ξ_0 so that it can be taken out of the integral. If we employ (12.13) in (12.8) we obtain

$$\gamma(R,T)=(C/2R^2)(1-R/\xi_0)e^{-R/\xi_0}, \tag{12.16}$$

which is valid for $R\lesssim\xi_0$ and $\gamma=0$ for $R>\xi_0$.

It is clear that the origin of the superconductivity can be traced to the nonlocal theory. In fact, for an isotropic solid we can simply carry (12.4) into (12.1), resulting in an integro-partial differential equation for $\mathbf{H}$. For anisotropic solid it is necessary to determine the properly invariant form of σ_{kl}.

Uniaxial crystal

For a uniaxial crystal with axis in x_3 direction we take

$$\sigma_k=\sigma_{kl}(\mathbf{R},\mathbf{i}_3)H_l,\quad \mathbf{R}\equiv\mathbf{x}'-\mathbf{x}. \tag{12.17}$$

σ_k is an isotropic function of two vectors $\mathbf{R}$, $\mathbf{i}_3$ and a skew-symmetric tensor $H_{kl}=\epsilon_{klm}H_m$ (equivalent to $\mathbf{H}$); linear in $\mathbf{H}$. Hence, it has the general form (Ref. 20, Appendix B)

$$\sigma_k=\alpha_0R_k+\alpha_1H_{km}R_m, \tag{12.18}$$

where α_0 and α_1 are functions of the invariants

$$R^2,\ R_3,\ H^2,\ (\mathbf{R}\cdot\mathbf{H})^2,\ (\mathbf{R}\times\mathbf{H})_3,\ H^2R_3-(\mathbf{H}\cdot\mathbf{R})H_3. \tag{12.19}$$

From (12.17) it follows that

$$\sigma_{kl}=\left.\frac{\partial\sigma_k}{\partial H_l}\right|_{\mathbf{H}=\mathbf{0}}.$$

Hence,

$$\sigma_{kl}=\sigma_0\epsilon_{kl}R_m+\epsilon_{ml3}(\sigma_1\delta_{km}+\sigma_2R_m\delta_{k3}+\sigma R_kR_m), \tag{12.20}$$

where σ_α, $\alpha=0,1,2,3$ depend on R^2 and R_3. Consequently (12.2) for uniaxial solids read

$$\mathbf{J}=\int_{\mathscr{V}-\sigma}[\sigma_0\mathbf{R}\times\mathbf{H}'+\sigma_1\mathbf{H}'\times\mathbf{i}_3$$
$$+(\mathbf{R}\times\mathbf{H}')_3(\sigma_2\mathbf{i}_3+\sigma_3\mathbf{R})]dv'. \tag{12.21}$$

It is possible to construct appropriate kernels σ_α similar to that of Pippard, but the scope of the present work does not include such an investigation. It is rather to indicate how the superconductivity can be brought into the domain of field theories once the nonlocality is incorporated properly.

Thus far, we have ignored the possibility of heat conduction under constant temperature. This question may be raised since the constitutive equation (7.43) for the heat acquires the form

$$q_k=\int_{\mathscr{V}-\sigma}\kappa_{kl}(\mathbf{x}'-\mathbf{x},T)H_l'\,dv'. \tag{12.22}$$

Similar to (12.4) for isotropic solids we have

$$\mathbf{q}=\int_{\mathscr{V}-\sigma}\sigma_0(|\mathbf{x}'-\mathbf{x}|,T)(\mathbf{x}'-\mathbf{x})\times\mathbf{H}(\mathbf{x}')dv'. \tag{12.23}$$

From this it follows that

$$\nabla\cdot\mathbf{q}=0. \tag{12.24}$$

Hence, no super heat conduction is possible.

Finally, we remark that the superconductivity of the second kind can be treated by means of the present theory. In this case we need the nonlinear theory with no memory effect. Such an investigation is left to a future study.

XIII. PASSAGE TO LATTICE STRUCTURE

At the atomic scale, materials are inhomogeneous since the lattice nodes are not equivalent to other points in the body. This situation is also true near the surface of a body. Nonlocal continuum theory can still provide the appropriate tool for the discussion of the atomic scale phenomena by taking nonlocal kernels as functions of two points $\mathbf{x}$ and $\mathbf{x}'$, e.g.,

$$\epsilon_{kl}=\epsilon_{kl}(\mathbf{x},\mathbf{x}',t-\tau') \tag{13.1}$$

rather than a function of $\mathbf{x}-\mathbf{x}'$. Of course, similar functional relations are assumed to be valid for the other moduli. The Fourier transforms are then taken with respect to $\mathbf{x}$, $\mathbf{x}'$, and τ' leading to

$$\bar{\epsilon}_{kl}=\bar{\epsilon}_{kl}(\boldsymbol{\xi},\boldsymbol{\xi}',\omega). \tag{13.2}$$

For homogeneous medium $\epsilon(\mathbf{x},\mathbf{x}',t-\tau')=\epsilon(\mathbf{x}-\mathbf{x}',t-\tau')$ and the transform with respect to $\mathbf{x}$ and $\mathbf{x}'$ give $\bar{\epsilon}(\boldsymbol{\xi},\omega)\delta(\boldsymbol{\xi}-\boldsymbol{\xi}')$ and we get back $\bar{\epsilon}_{kl}(\boldsymbol{\xi},\omega)$.

In an infinite crystal ϵ_{kl} remains unchanged when $\mathbf{x}$ and $\mathbf{x}'$ are given a translation by a lattice vector $\mathbf{a}$, i.e.,

$$\epsilon_{kl}(\mathbf{x}',\mathbf{x},t-\tau')=\epsilon_{kl}(\mathbf{x}+\mathbf{a},\mathbf{x}'+\mathbf{a},t-\tau'). \tag{13.3}$$

Such a periodic function can be expressed as

$$\epsilon_{kl}(\mathbf{x},\mathbf{x}',t-\tau')=\sum_{\mathbf{b}}g_{\mathbf{b}}(\mathbf{x}-\mathbf{x}',t-\tau')\exp(-2\pi i\mathbf{b}\cdot\mathbf{x}'), \tag{13.4}$$

where $\mathbf{b}=n_i\mathbf{b}_i$. Here n_i are integers and $\mathbf{b}_i$ are three base vectors of the reciprocal lattice, $\exp(2\pi i\mathbf{a}\cdot\mathbf{b})=1$. Consequently,[31]

$$\bar{\epsilon}_{kl}(\boldsymbol{\xi},\boldsymbol{\xi}',\omega)=\sum_{\mathbf{b}}\epsilon_{kl}^{\mathbf{b}}(\boldsymbol{\xi},\omega)\delta(\boldsymbol{\xi}'-\boldsymbol{\xi}-2\pi\mathbf{b}), \tag{13.5}$$

where $\epsilon_{kl}^{\mathbf{b}}$ is given by

$$\epsilon_{kl}^{\mathbf{b}}(\xi, \omega) = \int g_{\mathbf{b}}(\mathbf{R}, \omega) e^{-i\mathbf{b}\cdot\mathbf{x}}\, dv(\mathbf{R}) \tag{13.6}$$

and $\mathbf{R} = \mathbf{x} - \mathbf{x}'$.

From (13.5), it is clear that $\epsilon_{kl}(\xi, \omega)$ can be used in the electrodynamics of crystals if only the first term in (13.5) is adequate. This situation prevails if

$$|\xi| \ll 2\pi|\mathbf{b}| \gtrsim 2\pi/a. \tag{13.7}$$

This problem is discussed further in Ref. 31.

This result is valid for other piezoelectric and piezomagnetic moduli.

ACKNOWLEDGMENTS

I am indebted to Professor Mikael Ciftan of Duke University for drawing my attention to Ref. 31 at the proof stage. My thanks to Betty Kaminski for the skillful typing of this paper.

[1]A. Cemal Eringen, J. Math. Phys. **25**, 717 (1984).
[2]L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960).
[3]A. B. Pippard, *The Dynamics of Conduction Electrons* (Gordon and Breach, New York, 1965).
[4]F. Seitz and D. Turnbull, *Solid State Physics* (Academic, New York, 1963), Vol. 15.
[5]P. G. DeGennes, *Superconductivity of Metals and Alloys* (Benjamin, New York, 1966).
[6]*The Physics of Metals 1. Electrons*, edited by J. M. Ziman (Cambridge U.P., Cambridge, 1969).
[7]D. Pines, *Elementary Excitations in Solids* (Benjamin, New York, 1964).
[8]D. R. Penn, Phys. Rev. **128**, 2098 (1962).
[9]A. A. Rukhadze and V. P. Silin, Usp. Fiz. Nauk **74**, 223 (1961) [Sov. Phys. Usp. **4**, 459 (1961)].
[10]A. Cemal Eringen and D. G. B. Edelen, Int. J. Eng. Sci. **10**, 233 (1972).
[11]A. Cemal Eringen, *Continuum Physics IV* (Academic, New York, 1976). For a variational approach see D. G. B. Edelen, Part II.
[12]A. Cemal Eringen, Int. J. Eng. Sci. **10**, 561 (1972).
[13]C. Speziale and A. Cemal Eringen, Comput. Math. Appl. **7**, 27 (1981).
[14]A. Cemal Eringen, Mol. Cryst. Liq. Cryst. **75**, 321 (1981).
[15]A. Cemal Eringen, J. Math. Phys. **14**, 733 (1973).
[16]L. E. Hajdo and A. Cemal Eringen, Lett. Appl. Eng. Sci. **17**, 785 (1979).
[17]B. M. McCay and M. N. L. Narisimhan, Arch. Mech. **33**, 3, 365 (1981).
[18]A. Cemal Eringen, C. Speziale, and B. S. Kim, J. Mech. Phys. Solids **25**, 339 (1977).
[19]A. Cemal Eringen, J. Phys. D. **10**, 671 (1977).
[20]A. Cemal Eringen, *Mechanics of Continua* (Krieger, New York, 1980), 2nd ed.
[21]A. Cemal Eringen, Int. J. Eng. Sci. **4**, 179 (1966).
[22]Functional dependence on rotations is excluded. For optical branches of lattice vibrations, it may be necessary to include the rotational dependence as well.
[23]A. Cemal Eringen and B. S. Kim, Cryst. Lattice Defects **7**, 51 (1977).
[24]A. Cemal Eringen, J. Appl. Phys. **54**, 4703 (1983).
[25]N. Ari and A. Cemal Eringen, Cryst. Lattice Defects Amorph. Mat. **10**, 33 (1983).
[26]I. N. Sneddon, *Continuum Physics I*, edited by A. C. Eringen (Academic, New York, 1971).
[27]In Ref. 1, constitutive equations [such as (4.18)] also contain additional volume integrals. This is due to accounting independent variables at $(\mathbf{X}, t)$ [such as $\mathbf{C}(\mathbf{X}, t)$] separately from those at $(\mathbf{X}', t)$ [such as $\mathbf{C}(\mathbf{X}', t)$].
[28]N. Friedman and M. Katz, Arch. Rational Mech. Anal. **21**, 49 (1960).
[29]In this case, the space of functions G' is the Hilbert space so that in (4.7), (4.9), (4.11), and (4.12) G', defined by the third equation in (4.6), replaces G^r.
[30]A. Cemal Eringen, *Nonlinear Equations in Physics and Mathematics*, edited by A. O. Barut (Reidel, Dordrecht, 1978), p. 271.
[31]V. M. Agranovich and V. L. Ginsburg, *Spatial Dispersion in Crystal Optics and The Theory of Excitons* (Wiley, New York, 1966), Chap. 3.
[32]M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford, New York, 1954), Sec. 8.
[33]G. E. H. Reuter and E. H. Sondheimer, Proc. R. Soc. London Ser. A **195**, 336 (1948).

Proof of Regge analyticity for power law potentials

D. P. Datta and S. Mukherjee
Physics Department, North Bengal University, Dist. Darjeeling, India 734430

(Received 2 February 1984; accepted for publication 30 March 1984)

The l-plane analyticity of Schrödinger energy levels $E_n(l)$ for power law potentials $V(r) = r^\alpha$, for $\alpha > 2$ has been proved by using the Kato–Rellich perturbation theory for linear operators.

PACS numbers: 03.65 − w, 02.30. + g

I. INTRODUCTION

Grosse and Martin (GM)[1] recently initiated the study of the l-plane analyticity of Schrödinger energy levels E_n for pure-power potentials

$$V(r) = r^\alpha. \tag{1.1}$$

They proved the analyticity of E_n in the physical half-plane $\mathrm{Re}\, l > -\frac{1}{2}$ only for potentials with $\alpha \leqslant 2$. The proof for $\alpha > 2$ does not, however, follow from the argument. The reason may be summarized as follows. The GM technique rests on the characterization of the energy level E_n by the number of zeros of the corresponding square-integrable solution $u_n(r)$ and to show that this representation remains valid even for complex l with $\mathrm{Re}\, l > -\frac{1}{2}$ within a sector [e.g., $|\phi| < \pi/(2+\alpha)$] in the complex $z(\equiv r)$ plane, where ϕ is the argument of z. One can then easily eliminate the existence of a branch point of $E_n(l)$ in the physical half-plane by showing that the same square-integrable solution u_n may be retrieved by traveling along any closed contour in the half-plane $\mathrm{Re}\, l > -\frac{1}{2}$, starting from and returning back to the same point on the l-plane. It should be remarked at this point that the other singularities of $E_n(l)$ may be discarded by the use of general theorems such as the Herglotz theorem. Now, the integral which plays an important role in discarding branch points is

$$\mathrm{Im}\, u_n' u_n^* = \int_0^t dt'' |u_n|^2 \left\{ \frac{\mathrm{Im}\,\lambda}{t''^2} + t''^\alpha \sin(2+\alpha)\phi - |E_n| \sin(2\phi + \arg E_n) \right\},$$

$$z = te^{i\phi},\ \lambda = l(l+1), \tag{1.2}$$

where a prime denotes the derivative with respect to z and u_n^*, the complex conjugate wave function. This integral is used to show the constancy of the number of zeros of u_n in the sector $|\phi| < \pi/(2+\alpha)$ (for $\alpha > 2$). The square-integrability of $u_n(z)$ in the sector $|\phi| < \pi/(2+\alpha)$ follows from the asymptotic behavior

$$u_n \sim z^{-\alpha/4} \exp\left\{ -\frac{z^{\alpha/2+1}}{\alpha/2+1} \right\}, \tag{1.3}$$

which guarantees as well the exponential vanishing of $\mathrm{Im}(u_n' u_n^*)$ in $|\phi| < \pi/(2+\alpha)$ as $|z| \to \infty$. For $\alpha > 2$, it follows from Eq. (1.3) that u_n is still square-integrable on the rays $|\phi| = \pi/(2+\alpha)$. However, the square-integrability, in this case, is very delicate ($|u_n|^2 \sim z^{-\alpha/2}$) and this does not ensure the vanishing of $\mathrm{Im}(u_n' u_n^*)$ for $|z| \to \infty$. In fact, one can verify that $\mathrm{Im}(u_n' u_n^*) \sim O(1)$. Now, in the GM technique one vital step is to show that u_n has no zero on $\phi = \pi/(2+\alpha)$ for $\alpha > 2$, where $\mathrm{Im}\,\lambda > 0$, which follows only if the left-hand side of Eq. (1.2) vanishes as $t \to \infty$. It transpires that this is not the case.

In Refs. 2 and 3, we have developed a method based on Kato–Rellich perturbation theory on a Hilbert space by which one can prove the analyticity of $E_n(l)$ for the superposition of potentials of the form

$$V(r) = r^\alpha + kr^\beta,\quad \alpha > 0,\ 0 < \beta < \alpha,\ k \text{ real}. \tag{1.4}$$

We have stated this result in the earlier paper.[3] Here we shall supply a proof for the same. It will be shown that a proof for the required analyticity [also for pure power potential (1.1) with $\alpha > 2$] will follow naturally from the argument.

II. THE PROOF

First we note that by following the GM technique which has been outlined in the introduction one can easily prove the analyticity of $E_n(l)$ in $\mathrm{Re}\, l > -\frac{1}{2}$ for the potential (1.4) with $k > 0$. For clarity, we further remark that the method makes use of the asymptotic behavior of u_n which, in the present case, assumes the form

$$u_n \sim z^{-\alpha/4} \exp\left\{ -\frac{z^{\alpha/2+1}}{\alpha/2+1} - \frac{k}{2}\frac{z^{\beta-\alpha/2+1}}{\beta-\alpha/2+1} + \text{lower-order terms} \right\}. \tag{2.1}$$

This shows that u_n is square-integrable in the sector $|\phi| \leqslant \pi/(2+\alpha)$, when $k > 0$. One can then use the integral (1.2) to show that u_n has n zeros with $|\phi| \leqslant \pi/(2+\alpha)$ for any l with $\mathrm{Re}\, l > -\frac{1}{2}$. For details we refer to Ref. 2.

Let us now consider the Hilbert space $L^2(0,\infty)$ of square-integrable functions. We choose the potential

$$V(r) = r^\alpha + \kappa\beta r^\beta,\quad k \text{ real},\ \alpha > 0,\ 0 < \beta < \alpha \tag{2.2}$$

for convenience. The corresponding well-defined Schrödinger operator is given by

$$H(\beta,\lambda) = -\frac{d^2}{dr^2} + \frac{\lambda}{r^2} + r^\alpha + k\beta r^\beta,\quad \lambda = l(l+1), \tag{2.3}$$

$$D(H) = \{u \,|\, u,u' \text{ absolutely continuous } |u,u' \in L^2 | u(0) = 0 | Hu \in L^2\}. \tag{2.4}$$

For l in the half-plane $\mathrm{Re}\, l > -\frac{1}{2}$, λ belongs to the complex plane cut along the real axis form $-\frac{1}{4}$ to $-\infty$. We shall now state several lemmas.

Lemma 2.1: For each fixed λ in the cut plane, the minimal operator $\dot{H}(0,\lambda)$, defined on the space of infinitely differ-

 0022-2488/84/113250-02$02.50

entiable functions with compact support C_0^∞ is a densely defined sectorial operator.

Remark: $H(0,\lambda)$ defined on $D(H)$ is then a closed (*m*-sectorial) extension of $\dot{H}(0,\lambda)$ for compacts of the λ cut plane.

Lemma 2.2: Let $u \in D(H)$. Then for any compact set in the λ cut plane we must have

$$\|x^\alpha u\| \leqslant \|H(0,\lambda)u\|. \tag{2.5}$$

Lemma 2.3: The maximal multiplication operator $r^\beta (0 \leqslant \beta < \alpha)$ in L^2 is $H(0,\lambda)$-bounded with relative bound zero.

The proof of these lemmas can be obtained by exactly following the proofs of the corresponding lemmas of Ref. 3. Collecting the results stated in the lemmas we can now write down the following theorem.

Theorem 2.1: For each fixed λ in the cut plane, $H(\beta,\lambda)$ is a holomorphic family of type A in β $(0 < \beta < \alpha)$ for each fixed real k and in k for each fixed β in the given interval, with compact resolvents.

We shall now consider the convergence of $H(\beta,\lambda)$ to $H(0,\lambda)$, when $k\beta \to 0^+$. We denote the resolvent $[H(\beta,\lambda) - z]^{-1}$ of $H(\beta,\lambda)$ by $R(\beta,z)$ and the corresponding resolvent set by $\rho(H(\beta,\lambda))$.

Theorem 2.2: Let $z \in \rho(H(0,\lambda))$. Then $z \in \rho(H(\beta,\lambda))$ for $k\beta$ sufficiently small and $R(\beta,z)$ converges to $R(0,z)$ in norm as $k\beta \to 0^+$:

$$\lim_{k\beta \to 0^+} \|R(\beta,z) - R(0,z)\| = 0. \tag{2.6}$$

The convergence is uniform on compacts in the λ cut plane.

Proof: Since

$$R(\beta,z) = R(\beta,z_0) \times [1 + (z_0 - z)R(\beta,z_0)]^{-1},$$

it is sufficient to prove the result for one $z_0 \in \rho(H(0,\lambda))$. For a compact subset Γ in the λ cut plane, the union $\cup$ of the numerical ranges [i.e., the set of values of $\langle u, H(\beta,\lambda)u\rangle$, $\|u\| = 1$] for $H(\beta,\lambda)$ over all $\lambda \in \Gamma$ is not the whole complex plane. Hence, we can choose z_0 so that

$$\operatorname{dist}(z_0, \cup) = d > 0.$$

Then we have[4]

$$\|R(\beta,z_0)\| \leqslant d^{-1}.$$

Next it follows that

$$R(\beta,z_0) - R(0,z_0) = -k\beta\,[r^{\beta/2}R(\beta,z_0)] \times [r^{\beta/2}R(0,z_0)]. \tag{2.7}$$

By Lemma 2.3 we can find positive constants a, b such that

$$\|r^{\beta/2}u\| \leqslant a\|[H(\beta,\lambda) - z_0]u\| + (b + a|z_0|)\|u\| \tag{2.8}$$

for all $\lambda \in \Gamma$ and $0 \leqslant \beta < \alpha$. Thus we have

$$\|r^{\beta/2}R(\beta,z_0)\| \leqslant a + d^{-1}(b + a|z_0|) = m \tag{2.9}$$

and also

$$\|r^{\beta/2}\|R(0,z_0) \leqslant a + d^{-1}(b + a|z_0|) = m. \tag{2.10}$$

Hence, from Eq. (2.7) it follows that

$$\|R(\beta,z_0) - R(0,z_0)\| \leqslant |k\beta|\, m^2 \to 0 \text{ as } k\beta \to 0^+.$$

According to Theorem 2.2, it follows that an eigenvalue $E_n(\beta,\lambda)$ of $H(\beta,\lambda)$ for each fixed λ in the cut plane, tends to the eigenvalue $E_n(0,\lambda)$ of $H(0,\lambda)$ as $k\beta \to 0^+$, at least asymptotically.

We shall now prove the analyticity of $E_n(l)$ in $\operatorname{Re} l > -\frac{1}{2}$ for the potential (1.1) when $\alpha > 2$. Since $E_n(\beta,\lambda)$ is asymptotic to $E_n(0,\lambda)$ on compacts of λ, we must have

$$E_n(\beta,\lambda) = E_n(0,\lambda) + O(k\beta). \tag{2.11}$$

Now for $k > 0$, the potential (2.2) is exactly identical to the case discussed in the first paragraph of this section and hence $E_n(\beta,\lambda)$ is analytic in the λ cut plane. Since by Eq. (2.11), any singularity of $E_n(0,\lambda)$ must be a singularity of $E_n(\beta,\lambda)$ the analyticity of $E_n(0,\lambda)$ in the λ cut plane is established.

The analyticity of $E_n(\beta,\lambda)$ for real k now follows directly from the general arguments of Ref. 2.

ACKNOWLEDGMENT

This work was supported by the Indian Space Research Organization.

[1] H. Grosse and A. Martin, Phys. Rep. **60**, 343 (1980).
[2] D. P. Datta and S. Mukherjee, J. Math. Phys. **23**, 665 (1982); **24**, 1013E (1983).
[3] D. P. Datta and S. Mukherjee, J. Math. Phys. **24**, 573 (1983).
[4] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976).

Functions analytic on the half-plane as quantum mechanical states

Thierry Paul
Centre de Physique Théorique Section 2, Centre National de la Recherche Scientifique, Luminy Case 907, F-13288 Marseille Cedex 9, France

(Received 29 December 1983; accepted for publication 6 April 1984)

A transform between the state space of one-dimensional quantum mechanical systems and a direct sum of two spaces of square integrable functions analytic on the open upper half-plane is constructed. It gives a representation of usual quantum mechanics on which the free evolution is trivial and the representation of canonical transformation very simple. Generalizations to higher dimensions are also discussed.

PACS numbers: 03.65. – w, 02.20 + b

I. INTRODUCTION

Though all irreducible representations of the canonical commutation relations

$$[Q,P] = i \tag{1.1}$$

are equivalent, other representations than the usual Schrödinger one have been helpful in quantum mechanics: Bargmann[1,2] used a space of entire functions, Itzykson used functions analytic on the disk,[3] and van Winter used functions analytic on sectors.[4]

In this paper we formulate nonrelativistic quantum mechanics of one degree of freedom (generalizations to higher dimension are also described) in the direct sum $\mathscr{H} = \mathscr{H}_{1/2} \oplus \mathscr{H}_{-1/2}$ of two spaces of analytic functions on the upper half-plane. The $\mathscr{H}_{\pm 1/2}$ are the spaces of functions $f(b + ia)$ analytic on the upper half-plane $a > 0$ and square integrable with respect to the measure $a^{\pm 1/2}\, da\, db$.

Some operators used in quantum mechanics are very transparent in this representation.

(i) The free Hamiltonian is just the operator of differentiation (times $-i$): $-i(\partial/\partial z)$; so the free evolution consists of a translation of the variable along the real axis.

(ii) The usual dilation operator $D(\sqrt{a})$, acting on $L^2(R,dp)$ by

$$(D(\sqrt{a})\psi)(p) = a^{1/4}\psi(\sqrt{a}p),$$

acts on $\mathscr{H}_{\pm 1/2}$ by

$$(D(\sqrt{a})f)(z) = a^{\lambda} f(a^{-1}z)$$

(with $\lambda = -\frac{5}{4}$ for $\mathscr{H}_{+1/2}$ and $\lambda = -\frac{3}{4}$ for $\mathscr{H}_{-1/2}$).

(iii) The "time operator" $(P^{-1}Q + QP^{-1})/2$ has also a simple expression.

Furthermore, as in the Bargmann space, the variable has a "classical" meaning which is here

$$z \sim \frac{q}{p} + \frac{i}{2p^2}, \quad \text{for } \mathscr{H}_{-1/2}$$

and

$$z \sim \frac{q}{p} + \frac{3i}{2p^2}, \quad \text{for } \mathscr{H}_{+1/2}.$$

So this representation consists of writing ordinary quantum mechanics in terms of the variable "$q/p + i\lambda/p^2$" ($\lambda = \frac{1}{2}$ or $\frac{3}{2}$).

These spaces allow us also to bring the representation of linear canonical transformations into a simple form. It should be noticed that the analytic functions used here do not reduce to the analytic functions in other formulations: they are analytic only on the half-plane (as contrasted to the entire functions in the Bargmann space) and they are square integrable with respect to a two-dimensional measure (in contrast to the Hardy functions studied by van Winter).

The transform from $L^2(R,dp)$ on the two spaces $\mathscr{H}_{+1/2}$ and $\mathscr{H}_{-1/2}$ is given by

$$f_e(z) = \frac{1}{\sqrt{2}\pi^{3/4}} \int_{-\infty}^{+\infty} p^2\, e^{iz(p^2/2)}\psi(p)dp,$$

$$f_0(z) = \frac{1}{2\pi^{3/4}} \int_{-\infty}^{+\infty} p e^{iz(p^2/2)}\psi(p)dp.$$

Let us now quickly indicate how we are led to this transform: from (1.1) it is easy to deduce that

$$\left[\frac{PQ+QP}{4}, \frac{P^2}{2}\right] = i\frac{P^2}{2}. \tag{1.2}$$

We recognize the Lie algebra of the affine group,[5] which can be identified with the upper half-plane. This group is nonunimodular but we know from the theory of square integrable representation of nonunimodular groups (see Refs. 6–9) the following facts.

(i) Let G be a locally compact nonunimodular group with left Haar measure $d\mu(g)$ and $U(g)$ be a representation of G in a Hilbert space $\mathscr{H}$ with scalar product $(\cdot,\cdot)$

(ii) Let φ be an element of $\mathscr{H}$; if $\int |(U(g)\varphi,\varphi)|^2\, d\mu(g) < \infty$ (admissibility condition) we write

$$C_\varphi = \frac{\int |(U(g)\varphi,\varphi)|^2\, d\mu(g)}{(\varphi,\varphi)}.$$

Then, for each $\psi \in \mathscr{H}$, the function $f(g)$ is defined by

$$f(g) = C_\varphi^{-1/2}(U(g)\varphi,\psi),$$

and is square integrable with respect to $d\mu(g)$ and

$$\int \overline{f(g)} f(g) d\mu(g) = (\psi,\psi).$$

We will see that in the case of the affine group such a transform, with a suitable choice of φ, gives us square integrable analytic functions on the half-plane.

In Sec. II we given elementary properties of the affine group; in Sec. III we construct the transform and prove its unitarity. Section IV is devoted to the study of a generaliza-

 0022-2488/84/113252-12$02.50

tion of the spaces $\mathscr{H}_{\pm 1/2}$. We give some application of this transform in quantum mechanics in Sec. IV. Different generalizations are described in Sec. VI.

II. THE "$ax+b$" (AFFINE) GROUP[7,10]

The affine group G is the set of couples

$$\{a,b\}, \quad a>0, \quad b\in R,$$

with the law

$$\{a',b'\}\{a,b\}=\{a'a,a'b+b'\}.$$

It is a locally compact nonunimodular group with right Haar measure $d\mu_R(a,b)=da\,db/a$ and left Haar measure $d\mu_L(a,b)=da\,db/a^2$.

The Lie algebra is given by

$$[A,B]=iB,$$

and it has been proved[11,12] that G admits only two irreducible inequivalent representations, one with $B>0$ and the other with $B<0$.

Here we consider the (reducible) representation on $L^2(R,dp)$ with generators given by $A=(PQ+QP/4)$ and $B=P^2/2$, where P and Q are the usual Heisenberg operators satisfying $[Q,P]=i$; the representation is given by

$$U(a,b)=\exp\left(-i\frac{bP^2}{2}\right)\exp\left(i\frac{LNa}{2}\cdot\frac{PQ+QP}{2}\right),$$

which acts on $L^2(R,dp)$ by

$$[U(a,b)\psi](p)=a^{1/4}e^{-i(bp^2/2)}\psi(\sqrt{a}p).$$

This representation permits to identify G as the group of upper triangular real matrices of determinant 1 [quotient of SL(2,R) by SO(2,R)] with the restriction of the metaplectic representation of SL(2,R) to G.

The identification is made by

$$\{a,b\}\Leftrightarrow\begin{pmatrix}\sqrt{a} & b/\sqrt{a}\\ 0 & 1/\sqrt{a}\end{pmatrix}.$$

The representation U is a continuous, unitary representation of G, irreducible on even ($L^2_e(R,dp)$) and odd ($L^2_o(R,dp)$) parts of $L^2(R,dp)$:

$$L^2_e(R,dp)=\left\{\psi,\ \text{functions on } R \text{ with } \psi(-p)=\psi(p) \text{ and } \int_{-\infty}^{+\infty}|\psi(p)|^2\,dp=2\int_0^{+\infty}|\psi(p)|^2\,dp<+\infty\right\},$$

$$L^2_o(R,dp)=\left\{\psi,\ \text{functions on } R \text{ with } \psi(-p)=-\psi(p) \text{ and } \int_{-\infty}^{+\infty}|\psi(p)|^2\,dp=2\int_0^{+\infty}|\psi(p)|^2\,dp<+\infty\right\}.$$

Now we compute the admissibility condition introduced in the Introduction. That is,

$$\int|(U(a,b)\varphi,\varphi)|^2\frac{da\,db}{a^2}<+\infty,$$

and

$$c_\varphi=\frac{\int|(U(a,b)\varphi,\varphi)|^2(da/db/a^2)}{\|\varphi\|^2},$$

for $\varphi\in L^2(R,dp)$, φ even or odd.

An easy computation gives

$$\int_0^{+\infty}\frac{|\varphi(p)|^2}{p^2}dp<+\infty$$

and

$$c_\varphi=8\pi\int_0^{+\infty}\frac{|\varphi(p)|^2}{p^2}dp.$$

III. THE TRANSFORM

A. Definition

According to the general theory described in the Introduction, we take two elements of $L^2(R,dp)$, one φ^e in $L^2_e(R,dp)$ and one φ^o in $L^2_o(R,dp)$, which satisfy the admissibility condition. Two candidates which will give us analytic functions are

$$\varphi^o(p)=pe^{-p^2/2},\qquad \varphi^e(p)=p^2e^{-p^2/2}. \tag{3.1}$$

Let us write

$$c_o=8\pi\int_0^{+\infty}\frac{|\varphi^o(p)|^2}{p^2}=4\pi^{3/2},\qquad c_e=8\pi\int_0^{\infty}\frac{|\varphi^e(p)|^2}{p^2}dp=2\pi^{3/2}. \tag{3.2}$$

We will associate to every $\psi\in L^2_o(R,dp)$ [resp. $L^2_e(R,dp)$] an analytic function in the following way: we know that the transform

$$\psi\in L^2_o(R,dp)\ [\text{resp. } L^2_e(R,dp)]\to\frac{1}{\sqrt{c_o}}(U(a,b)\varphi^o,\psi)$$

[resp. $(1/\sqrt{c_e})$ $(U(a,b)\varphi^e,\psi)$] is an isometry between $L^2_o(R,dp)$ and $L^2(G,d\mu_L(g))$, which is

$$L^2(\Pi,da\,dp/a^2),$$

where

$$\Pi=\{z\in\mathbb{C}/\mathrm{Im}(z)>0\}.$$

Here and in what follows we shall write

$$z=b+ia.$$

But

$$\frac{1}{\sqrt{c_o}}(U(a,b)\varphi^o,\psi)=\frac{a^{3/4}}{\sqrt{C_o}}\int_{-\infty}^{+\infty}pe^{iz(p^2/2)}\psi(p)dp,$$

$$\left(\text{resp. } \frac{1}{\sqrt{C_e}}(U(a,b)\varphi^e,\psi)=\frac{a^{5/4}}{\sqrt{C_e}}\int_{-\infty}^{+\infty}p^2e^{iz(p^2/2)}\psi(p)dp\right).$$

This suggests we consider the two following transforms:

$$(A_0\psi)(z)=\frac{1}{2\pi^{3/4}}\int_{-\infty}^{+\infty}pe^{iz(p^2/2)}\psi(p)dp, \quad \text{for } \psi\in L^2_o(R,dp), \tag{3.3}$$

$$(A_e\psi)(z)=\frac{1}{\sqrt{2\pi}^{3/4}}\int_{-\infty}^{+\infty}p^2e^{iz(p^2/2)}\psi(p)dp, \quad \text{for } \psi\in L^2_e(R,dp). \tag{3.4}$$

B. Properties

Before we give the main result of this section we introduce the following spaces.

Definition: $\mathscr{H}_{\pm 1/2}$ is the space of analytic functions defined on Π (Poincaré half-plane), square integrable with respect to the measure $[\mathrm{Im}(z)]^{\pm 1/2}\, d\,\mathrm{Re}\, z\, d\,\mathrm{Im}\, z$.

A deeper study of such spaces will be made in the following section; let us just mention that they are Hilbert spaces (see the Appendix), equipped with the scalar product

$$(f,g)_{\pm 1/2} = \int \overline{f(z)} g(z) d\mu_{\pm 1/2}(z)$$

where

$$d\mu_{\pm 1/2}(z) = a^{\pm 1/2}\, da\, db. \tag{3.5}$$

Theorem: A_o (resp. A_e) is a unitary transform between $L^2_o(R,dp)$ [resp. $L^2_e(R,dp)$] and $\mathscr{H}_{-1/2}$ (resp. $\mathscr{H}_{1/2}$) and the inverse transform is given by

$$(A_o^{-1} f)(p) = \frac{1}{2\pi^{3/4}} \lim_{\substack{\sigma\to\infty \\ \gamma\to\infty}} \int_{\substack{|\mathrm{Re}(z)|<\sigma \\ 1/\gamma<\mathrm{Im}(z)<\gamma}} p e^{-i(p^2/2)\bar{z}} \times f(z) a^{-1/2}\, da\, db, \quad \text{for } f \in \mathscr{H}_{-1/2} \tag{3.6}$$

and

$$(A_e^{-1} f)(p) = \frac{1}{\sqrt{2}\pi^{3/4}} \lim_{\substack{\sigma\to\infty \\ \gamma\to\infty}} \int_{\substack{|\mathrm{Re}(z)<\sigma \\ 1/\gamma<\mathrm{Im}(z)<\gamma}} p^2 e^{-i(p^2/2)\bar{z}} \times f(z) a^{1/2}\, da\, db. \tag{3.7}$$

Proof of the Theorem: The functions $(A_e\psi)$ [resp. $(A_0\psi)(z)$] are clearly analytic in z. We just need to prove the isometry and surjectivity.

Proof of isometry: We have

$$(A_o\psi)(z) = \frac{1}{2\pi^{3/4}} \int_{-\infty}^{+\infty} p e^{iz(p^2/2)} \psi(p) dp,$$

so, with $\psi,\varphi \in L^2_o(R,dp)$,

$$\begin{aligned}
&(A_o\psi, A_o\varphi)_{-1/2} \\
&= \frac{1}{4\pi^{3/4}} \int_{-\infty}^{+\infty} dp \int_{-\infty}^{+\infty} dp' \int_0^{+\infty} da \int_{-\infty}^{+\infty} db\, pp' \\
&\quad \times \exp\left[ib\left(\frac{p'^2-p^2}{2}\right)\right] \exp\left[-a\left(\frac{p^2+p'^2}{2}\right)\right] \\
&\quad \times \varphi(p')\bar{\psi}(p) \\
&= \frac{1}{2\sqrt{\pi}} \int pp' \left[\frac{\delta(p-p')}{|p|} + \frac{\delta(p+p')}{|p|}\right] \\
&\quad \times \exp\left[-a\left(\frac{p^2+p'^2}{2}\right)\right] \varphi(p')\,\overline{\psi(p)}\quad da\, dp\, dp'.
\end{aligned}$$

Now, since ψ and φ are odd we obtain

$$\begin{aligned}
&(A_o\psi, A_o\varphi)_{-1/2} \\
&= \frac{1}{2\sqrt{\pi}} \int \frac{p^2}{|p|} e^{-ap^2} \varphi(p)\,\overline{\psi(p)} a^{-1/2}\, da\, dp \\
&= \int \overline{\psi(p)}\varphi(p) dp = (\psi,\varphi)_{L^2(R,dp)}.
\end{aligned}$$

The calculation for A_e is exactly the same.

Proof of surjectivity: First of all we prove the inversion formula.

Consider the map $W_o^{\sigma,\gamma}$, given by $f \in \mathscr{H}_{-1/2}$,

$$(W_o^{\sigma,\gamma} f)(p) = \frac{1}{2\pi^{3/4}} \int_{\substack{|\mathrm{Re}(z)|<\sigma \\ \gamma>\mathrm{Im}(z)>1/\gamma}} p e^{-i(p^2/2)\bar{z}} f(z) \times a^{-1/2}\, da\, db.$$

Now $(W_o^{\sigma,\gamma} f)(p)$ is clearly the scalar product in $L^2(\Pi, a^{-1/2}\, da\, db)$ of f and the function h^p_σ defined by

$$h^p_{\sigma,\gamma}(z) = \begin{cases} \dfrac{1}{2\pi^{3/4}} p e^{-i(p^2/2)\bar{z}}, \\ \quad \text{if } |\mathrm{Re}(z)|<\sigma, 1/\gamma>\mathrm{Im}(z)>\gamma, \\ 0, \quad \text{if } |\mathrm{Re}(z)|>\sigma \text{ or } 1/\gamma<\mathrm{Im}(z) \\ \quad\quad \text{or } \mathrm{Im}(z)<\gamma, \end{cases}$$

$h^p_{\sigma,\gamma} \in L^2(\Pi, a^{-1/2}\, da\, db)$ so $(W_o^{\sigma,\gamma} f)(p)$ is defined for every $p \in R$ and $\sigma > 0$.

One can see that $W_o^{\sigma,\gamma} f \in L^2_o(R,dp)$ and that, if $f = A_0\psi$, $\lim_{\sigma\to\infty} \| W_0 f - \psi \| = 0$, where $\|\ \|$ is the norm in $L^2(R,dp)$, so we can write

$$(A_o^{-1} f)(p) = \lim_{\substack{\sigma\to\infty \\ \gamma\to\infty}} p \int_{\substack{|\mathrm{Re}(z)|<\sigma \\ 1/\gamma<\mathrm{Im}(z)<\gamma}} e^{-i\bar{z}(p^2/2)} f(z) a^{-1/2}\, da\, db,$$

where the limit is understood in the L^2 sense.

We prove now that, for every $f \in \mathscr{H}_{-1/2}$,

$$\|A_o^{-1} f\|^2 = (f,f)_{-1/2}.$$

From (3.5) we have

$$\begin{aligned}
&\int_{-\infty}^{+\infty} |(A_o^{-1} f)(p)|^2 dp \\
&= \frac{1}{4\pi^{3/2}} \iiint_{-\infty}^{+\infty} dp\, p^2 e^{i(p^2/2)[z-\bar{z}']} \\
&\quad \times d\mu_{-1/2}(z) d\mu_{-1/2}(z') \\
&= \int \frac{1}{8\pi} \left(\frac{z-\bar{z}'}{2i}\right)^{-3/2} f(z')\,\overline{f(z)} \\
&\quad \times d\mu_{-1/2}(z') d\mu_{-1/2}(z) \\
&= \int \rho_{-1/2}(z-\bar{z}') f(z')\,\overline{f(z)} \\
&\quad \times d\mu_{-1/2}(z') d\mu_{-1/2}(z),
\end{aligned}$$

where $\rho_{-1/2}$ is the reproducing kernel of $\mathscr{H}_{-1/2}$ (see the Appendix) so

$$\int_{-\infty}^{+\infty} |(A_o^{-1} f)(p)|^2 dp = \int |f(z)|^2 d\mu_{-1/2}(z) < +\infty$$

and the surjectivity is proved. The same calculation is possible for A_e^{-1} and the result is the same. The theorem is proved.

Remarks: (1) Every function in $L^2(R,dp)$ can be decomposed in even and odd parts $\psi = \psi_e + \psi_o$ with $\psi_e \in L^2_e(R,dp)$, $\psi_o \in L^2_o(R,dp)$. We have

$$\int_{-\infty}^{+\infty} p e^{iz(p^2/2)} \psi_e(p)\, dp = \int_{-\infty}^{+\infty} p^2 e^{iz(p^2/2)} \psi_o(p) dp = 0$$

so

$$(A_o\,\psi_o)(z) = \frac{1}{2\pi^{3/4}} \int_{-\infty}^{+\infty} p e^{izp^2/2} \psi(p) dp$$

and

$$(A_e\psi_e)(z)=\frac{1}{\sqrt{2}\pi^{3/4}}\int_{-\infty}^{+\infty}p^2e^{iz(p^2/2)}\psi(p)dp.$$

So we can consider the map A:

$$L^2(R,dp)\to\mathscr{H}_{+1/2}\oplus\mathscr{H}_{-1/2}$$

$$\psi\to\begin{pmatrix}f_e(z)\\f_0(z)\end{pmatrix}\equiv\begin{pmatrix}\dfrac{1}{\sqrt{2}\pi^{3/4}}\displaystyle\int_{-\infty}^{+\infty}p^2e^{iz(p^2/2)}\psi(p)dp\\ \dfrac{1}{2\pi^{3/4}}\displaystyle\int_{-\infty}^{\infty}pe^{iz(p^2/2)}\psi(p)dP\end{pmatrix}. \tag{3.8}$$

The inverse transform is given by

$$\psi(p)=\frac{1}{\sqrt{2}\pi^{3/4}}\lim_{\substack{\sigma\to\infty\\ \gamma\to\infty}}\int_{\substack{|\mathrm{Re}(z)|<\sigma\\ 1/\gamma<\mathrm{Im}(z)<\gamma}}p^2e^{-i\bar{z}p^2/2}f_e(z)d\mu_{1/2}(z)+\frac{1}{2\pi^{3/4}}\lim_{\substack{\sigma\to\infty\\ \gamma\to\infty}}\int_{\substack{|\mathrm{Re}(z)|<\sigma\\ 1/\gamma<\mathrm{Im}(z)<\gamma}}pe^{-i\bar{z}p^2/2}f_o(z)d\mu_{-1/2}(z). \tag{3.9}$$

(2) Because of the fast decrease of the kernels $pe^{iz(p^2/2)}$ and $p^2e^{iz(p^2/2)}$, the transform A can be extended to the space of tempered distributions on R, as is the case for the Bargmann transform. The transform of a tempered distribution gives also an analytic function on Π, but of course, in general, it is non-square-integrable. This study will be made in a forthcoming paper.

IV. THE FAMILY OF SPACES $\mathscr{H}_\alpha$

With a view towards studying the spaces $\mathscr{H}_{\pm1/2}$ we define the more general family of spaces $\mathscr{H}_\alpha$ and present some of their properties. Technical points are given in the Appendix.

A. Definition

For $\alpha>-1$ we define $\mathscr{H}_\alpha$ as the space of analytic functions defined on the open upper half-plane and square integrable with respect to the measure

$$\left(\frac{z-\bar{z}}{2i}\right)^\alpha\frac{dz\wedge d\bar{z}}{2i}\equiv d\mu_\alpha(z).$$

It is a Hilbert space with scalar product

$$\begin{aligned}(f,g)&=\int\overline{f(z)}g(z)\left(\frac{z-\bar{z}}{2i}\right)^\alpha\frac{dz\wedge d\bar{z}}{2i}\\ &\equiv\int\overline{f(z)}g(z)d\mu_\alpha(z)\\ &=\int_{b=-\infty}^{+\infty}\int_{a=0}^{+\infty}\overline{f(z)}g(z)a^\alpha\,da\,db,\end{aligned} \tag{4.1}$$

with $z=b+ia$ (see the Appendix). As other spaces of analytic functions, it possesses a reproducing kernel, i.e., an element $\rho_\alpha(z-\overline{Z})\mathscr{H}_\alpha$ such that

$$\forall f\in\mathscr{H}_\alpha,\ \int\rho_\alpha(z-\overline{Z})f(Z)d\mu_\alpha(Z)=f(z);$$

ρ_α is given by the formula (see the Appendix)

$$\rho_\alpha(z-\overline{Z})=c_\alpha[(z-\overline{Z})/2i]^{-(\alpha+2)}, \tag{4.2}$$

where $c_\alpha\in\mathbf{R}$, $c_\alpha=(\alpha+1)/4\pi$, and

$$\left(\frac{a-\overline{Z}}{2i}\right)^{-(\alpha+2)}=\left|\frac{z-\overline{Z}}{2}\right|^{-(\alpha+2)}\times\exp\left[-i(\alpha+2)\arg\left(\frac{z-\overline{Z}}{2i}\right)\right]$$

and

$$-\frac{\pi}{2}<\arg\left[\frac{z-\overline{Z}}{2i}\right]<\frac{\pi}{2}.$$

Directly from the definition, we see that

$$\overline{\rho_\alpha(z-\overline{Z})}=\rho_\alpha(Z-\bar{z}).$$

From now we will note $|Z,\alpha>\in\mathscr{H}_\alpha$, given by

$$|Z,\alpha\rangle=\rho_\alpha(\cdot-\overline{Z}),$$

and in general $|f\rangle$ each element of $\mathscr{H}_\alpha$.

With this notation, the reproducing property is written

$$\langle Z,\alpha|f\rangle=f(Z).$$

If there is no danger of confusion, we will abbreviate $|Z,\alpha\rangle$ to $|Z\rangle$.

For each f in $\mathscr{H}_\alpha$ the following estimate holds:

$$|f(z)|\leqslant\|f\|(\sqrt{\alpha+1}/4\sqrt{\pi})\{\mathrm{Im}(z)\}^{-\alpha/2-1}, \tag{4.3}$$

where $\mathrm{Im}(z)$ is the imaginary part of z and

$$\|f\|^2=\int\overline{f(z)}f(z)d\mu_\alpha(z).$$

B. Representation of SL(2,R); $|Z\rangle$ as coherent states

In this section we will see that the vectors $|Z\rangle$ are just, up to a normalization constant, coherent states of SL(2,R) (in the Perelomov sense[13]).

In $\mathscr{H}_\alpha$ we have the following irreducible unitary representation U of SL(2,R):

$$\text{if } g\in\mathrm{SL}(2,R)\text{ with }g^{-1}=\begin{pmatrix}a&b\\c&d\end{pmatrix},$$

$$(U(g)f)(z)=(cz+d)^{-\alpha-2}f[(az+b)/(cz+d)]. \tag{4.4}$$

The unitarity is easily shown by direct computation and the irreducibility has been proved by Berezin[14] [for the representation of SU(1,1) on analytic functions on the disk which is equivalent to this one (see Appendix)].

For α integer (>-1) we have a true representation and for α half-integer we have a projective representation with multiplier ±1.

It is well known that each element

$$S=\begin{pmatrix}a&b\\c&d\end{pmatrix}$$

of SL(2,R) can be decomposed in the form

$$\begin{pmatrix}a&b\\c&d\end{pmatrix}=\begin{pmatrix}\sqrt{y}&x/\sqrt{y}\\0&1/\sqrt{y}\end{pmatrix}\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix} \tag{4.5}$$

with

$$x+iy=(ai+b)/(ci+d)\equiv S\cdot i$$

and

$$\theta=\arg(ci+d)(-\pi<\theta\leqslant+\pi).$$

Now, denoting

$$R(\theta)=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}$$

an element of SO(2,R) we have

$$\begin{aligned}\langle Z|U(R(\theta))|i\rangle &= c_\alpha(\sin\theta Z+\cos\theta)^{-\alpha-2}\\ &\times\left[\left(\frac{\cos\theta Z-\sin\theta}{2i(\sin\theta Z+\cos\theta)}\right)+\frac{i}{2i}\right]^{-\alpha-2}\\ &= c_\alpha\left(\frac{\cos\theta Z-\sin\theta+i\sin\theta Z+i\cos\theta}{2i}\right)^{-\alpha-2}\\ &= (\cos\theta+i\sin\theta)^{-\alpha-2}c_\alpha[(Z+i)/2i]^{-\alpha-2}\\ &= e^{-i(\alpha+2)\theta}\langle Z|i\rangle,\end{aligned}$$

so

$$U(R(\theta))|i\rangle = e^{-i(\alpha+2)\theta}|i\rangle, \tag{4.6}$$

so the action of the subgroup SO(2,R) on the state $|i\rangle$ is just a multiplication by a phase factor. Then following Perelomov[13] we call "coherent states" of SL(2,R) associated with $|i\rangle$ the system of states $|xy\rangle$ defined by

$$|\tilde{xy}\rangle = U\left(\begin{pmatrix}\sqrt{y} & x/\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}\right)|i\rangle$$

which, since

$$\begin{pmatrix}y & x\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}^{-1}=\begin{pmatrix}1/\sqrt{y} & -x/\sqrt{y}\\ 0 & \sqrt{y}\end{pmatrix},$$

is given by

$$\begin{aligned}\langle Z|\tilde{xy}\rangle &= \langle Z|U\left(\begin{pmatrix}y & x/\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}\right)|i\rangle\\ &= c_\alpha(\sqrt{y})^{-\alpha-2}\left(\frac{1}{2i}\left[\left(\frac{Z}{\sqrt{y}}-\frac{x}{\sqrt{y}}\right)\frac{1}{\sqrt{y}}+i\right]\right)^{-\alpha-2}\\ &= (\sqrt{y})^{\alpha+2}\langle Z|x+iy\rangle.\end{aligned}$$

So we have

$$|\tilde{xy}\rangle = \left(\frac{\langle i|i\rangle}{\langle Z|Z\rangle}\right)^{1/2}|Z\rangle, \tag{4.7}$$

with $Z=x+iy$, so

$$|Z\rangle = \left(\frac{\langle Z|Z\rangle}{\langle i|i\rangle}\right)^{1/2}|\tilde{xy}\rangle. \tag{4.8}$$

C. Principal operators on $\mathscr{H}_\alpha$

A large class of operators H on $\mathscr{H}_\alpha$ can be expressed by kernels, i.e., functions of two complex variables, analytic in the first and antianalytic in the second: $h(z,\bar{z}')$ so that

$$(Hf)(z) = \int h(z,\bar{z})f(z')d\mu_\alpha(z'), \tag{4.9}$$

where $h(z,\bar{z}')=\langle z|Hz'\rangle$,

$|Hz'\rangle \equiv H|z'\rangle$.

The kernel of the adjoint operator H^* of H is given by the following formula, denoting $G=H^*$:

$$g(z,\bar{z}') = \overline{h(z',\bar{z})}. \tag{4.10}$$

We just will now briefly mention some elementary operators in $\mathscr{H}_\alpha$, see Ref. 15 for details of proof.

1. The operator η: $-i(\partial/\partial z)$

The operator η defined by

$$(\eta f)(z) = -i\frac{\partial}{\partial z}f(z) \tag{4.11}$$

is self-adjoint because it is the strong derivative (multiplied by $-i$) of the strongly continuous group $T(t)$ of unitary transformations given by

$$[T(t)f](z)=f(z-t). \tag{4.12}$$

Its domain is the set of functions of $\mathscr{H}_\alpha$ such that

$$\int\left|\frac{\partial f}{\partial z}(z)\right|^2 d\mu_\alpha(z)<+\infty.$$

2. The operator ξ multiplication by z

Here ξ is defined by

$$(\xi f)(z) = zf(z). \tag{4.13}$$

The resolvent $R(E)$ of ξ at point E is given by

$$(R(E)f)(z) = ((\xi-E)^{-1}f)(z) = (z-E)^{-1}f(z), \tag{4.14}$$

so we see that the spectrum of ξ is the upper half-plane. The adjoint ξ^* of ξ is given by

$$(\xi f)(z) = \int \rho_\alpha(z-\bar{z}')\bar{z}'f(z')d\mu_\alpha(z') \tag{4.15}$$

and its spectrum is the lower half-plane.

Note: We have

$$\left[\frac{\xi+\xi^+}{2},\eta\right]=i. \tag{4.16}$$

3. Dilations

If we restrict the representation (4.4) to the subgroup of SL(2,R) of diagonal matrix (dilations)

$$\begin{pmatrix}\sqrt{a} & 0\\ 0 & 1/\sqrt{a}\end{pmatrix},$$

we obtain the following unitary dilation operator $D(a)$:

$$(D(a)f)(z) = a^{\alpha/2+1}f(az). \tag{4.17}$$

So we obtain the generator δ of the dilations

$$(\delta f)(z) = -iz\frac{\partial}{\partial z}-i\left(\frac{\alpha}{2}+1\right). \tag{4.18}$$

V. APPLICATION TO QUANTUM MECHANICS

In this section we come back to the transform defined in Sec. III. We have constructed in Sec. III a unitary map between $L^2(R,dp)$ and $\mathscr{H}_{1/2}\oplus\mathscr{H}_{-1/2}\equiv\mathscr{H}$,

$$\psi\overset{A}{\longrightarrow}\begin{pmatrix}f_e\\ f_0\end{pmatrix}=\begin{pmatrix}A_e\psi\\ A_0\psi\end{pmatrix} \tag{5.1}$$

(here $\psi=\psi_e+\psi_o$ and $A_e\psi_o=0$ and $A_o\psi_e=0$ so that $A_e\psi_e=A_e\psi$ and $A_o\psi_o=A_o\psi$).

Each operator in $L^2(R\,dp)$ will be represented in $\mathscr{H}_{1/2}\oplus\mathscr{H}_{-1/2}$ by a "matrix-valued kernel," that is, four kernels

$$\begin{pmatrix}h_{ee} & h_{eo}\\ h_{oe} & h_{oo}\end{pmatrix}$$

such that

$$\underline{H}\begin{pmatrix} f_e \\ f_o \end{pmatrix}(z) = \begin{pmatrix} \int h_{ee}(z,\bar{z}')f_e(z')d\mu_{1/2}(z') + \int h_{eo}(z,\bar{z}')f_o(z')d\mu_{-1/2}(z') \\ \int h_{oe}(z,\bar{z}')f_e(z')d\mu_{1/2}(z') + \int h_{oo}(z,\bar{z}')f_o(z')d\mu_{-1/2}(z') \end{pmatrix}. \tag{5.2}$$

Here, if H is the operator in $L^2(R,dp)$,

$$\underline{H} = AHA^{-1} \tag{5.3}$$

and

$$h_{ij}(z,\bar{z}') = \langle \varphi^i_z | H | \varphi^z_j \rangle, \quad \text{with } i,j \in \{e,o\}, \tag{5.4}$$

and

$$|\varphi^z_j\rangle \frac{U(a',b')|\varphi_j\rangle}{\sqrt{C_j}a'^{c(j)}},$$

$$\text{with } c_j = \begin{cases} \frac{5}{4}, & \text{for } j = e, \\ \frac{3}{4}, & \text{for } j = o. \end{cases}$$

We will give the expression in $\mathscr{H}$ of simple operators in $L^2(R,dp)$; one of them (multiplication by the variable z) suggests a classical meaning of z, the free evolution and Schrödinger equation in the interaction picture are very simple in this representation.

A. Some operators

(i) $-i(\partial/\partial z)$ since

$$-i\frac{\partial}{\partial z}\int p^\beta e^{iz(p^2/2)}\psi(p)dp = \int p^\beta e^{iz(p^2/2)}\frac{p^2}{2}\psi(p)dp.$$

The operator $-i(\partial/\partial z)$ is equal to $A(P^2/2)A^{-1}$.

(ii) The dilation generator $(QP+PQ)/2$ [where P and Q are the Heisenberg operators on $L^2(R,dp)$] is expressed in

$$\mathscr{H}_{-1/2} \text{ by } -2iz\frac{\partial}{\partial z} - \frac{3}{2}i$$

and in

$$\mathscr{H}_{+1/2} \text{ by } -2iz\frac{\partial}{\partial z} - \frac{5}{2}i$$

so we get the matrix

$$\begin{pmatrix} -2iz(\partial/\partial z) - \frac{5}{2}i & 0 \\ 0 & -2iz(\partial/\partial z) - \frac{3}{2}i \end{pmatrix}.$$

The operators P and Q change the parity so their expressions are not very simple. We get for P, the matrix

$$\sqrt{2}\begin{pmatrix} 0 & -i(\partial/\partial z) \\ 1 & 0 \end{pmatrix},$$

and for Q, the matrix

$$\begin{pmatrix} 0 & i2\sqrt{2}[1+z(\partial/\partial z)] \\ -\sqrt{2}z + \dfrac{i[(z-\bar{z}')/2i]^{-3/2}}{4\sqrt{2}\pi} & 0 \end{pmatrix}.$$

B. The operator ξ and the classical meaning of z

From (4.13) we see that, for $f \in \mathscr{H}_{-1/2}$,

$$(\xi f)(z) = zf(z)$$

$$= \frac{1}{2\pi^{3/4}}\int zpe^{iz(p^2/2)}\psi(p)dp$$

$$= \int \overline{\left[\left(\frac{P^{-1}Q+QP^{-1}}{2} - \frac{i}{2}P^{-2}\right)\varphi^z_o\right](p)}\psi(p)dp$$

$$= \left(\left[\frac{P^{-1}Q+QP^{-1}}{2} - \frac{i}{2}P^{-2}\right]\varphi^z_o, \psi\right)_{L_o(R,dp)}$$

with

$$\varphi^z_o(P) = (1/2\pi^{3/4})pe^{-\bar{z}(p^2/2)}.$$

So we see that if ψ is in the natural domain of

$$\frac{P^{-1}Q+QP^{-1}}{2} - i\frac{P^{-2}}{2},$$

we have

$$(\xi f)(z) = A_o\left[\left(\frac{P^{-1}Q+QP^{-1}}{2} + \frac{i}{2}P^{-2}\right)\psi\right]. \tag{5.5}$$

For f in $\mathscr{H}_{1/2}$ we have the same computation

$$(\xi f)(z) = A_e\left[\left(\frac{P^{-1}Q+QP^{-1}}{2} + \frac{3}{2}iP^{-2}\right)\psi\right]. \tag{5.6}$$

This suggests, in analogy with the Schrödinger and Bargmann representations, to identify the variable z with the classical quantity $q/p + i/2p^2$ in the odd case and $q/p + \frac{3}{2}(i/2p^2)$ in the even case. Two other arguments suggest the same interpretation: it is easy to see that, if we call

$$\varphi^z_o = A_0^{-1}|z\rangle \quad (\text{resp. } \varphi^z_e = A_e^{-1}|z\rangle),$$

that is

$$\varphi^z_o(p) = \frac{1}{2\pi^{3/4}}pe^{-\bar{z}p^2} \quad \left(\text{resp. } \varphi^z_e = \frac{1}{\sqrt{2}\pi^{3/4}}p^2e^{-\bar{z}p^2}\right), \tag{5.7}$$

then

$$\frac{(\varphi^z_o, [(P^{-1}Q+QP^{-1})/2]\varphi^z_o)_{L^2_o}}{(\varphi^z_o, \varphi^z_o)_{L^2_o}} = \operatorname{Re} z \tag{5.8}$$

(and the same for φ^z_e), and

$$\frac{(\varphi^z_o, (P^{-2}/2)\varphi^z_o)_{L^2_o}}{(\varphi^z_o, \varphi^z_o)_{L^2_o}} = \operatorname{Im} z \tag{5.9}$$

$$\left(\text{resp. } \frac{(\varphi^z_e, \frac{3}{2}P^{-2}\varphi^z_e)_{L^2_e}}{(\varphi^z_e, \varphi^z_e)_{L^2_e}} = \operatorname{Im} z\right). \tag{5.10}$$

So by analogy with the usual (Weyl) coherent states $|p,q\rangle$, for which we have

$$\langle p,q|(Q+iP|p,q\rangle) = q + ip, \tag{5.11}$$

where Q and P are the Heisenberg operators, we can identify z with $q/p + i/2p^2$ (resp. $q/p + 3i/2p^2$).

Another argument for this "classical meaning" is the following: in the Weyl case we construct coherent states by

acting on an element $|\psi_o\rangle$ of $L^2(R)$ with a (projective) representation of the group of translations on phase space R^2, namely a Weyl coherent state is $|p,q\rangle = W(p,q)|\psi_o\rangle$ with $\psi_o(p) = e^{-p^2/2}$, and $W(p,q)$ is the usual Weyl operator, in this representation we have identified a translation in phase space with the point translated from this origin. The origin can be defined as the point (q_o, p_o) such that

$$p_o = \langle\psi_o|P|\psi_o\rangle = 0,$$

and

$$q_o = \langle\psi_0|Q|\psi_o\rangle = 0.$$

We want to show that we have here the same construction: this time the group acting is the subgroup of $Sp(2,R)$ consisting of the triangular matrices

$$\{a,b\} = \begin{pmatrix} \sqrt{a} & b/\sqrt{a} \\ 0 & 1/\sqrt{a} \end{pmatrix}.$$

This group acts on phase space as a subgroup of the symplectic group

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} \sqrt{a} & b/\sqrt{a} \\ 0 & 1/\sqrt{a} \end{pmatrix} \times \begin{pmatrix} q \\ p \end{pmatrix}.$$

The representation of this group in $L^2(R,dp)$ is of the form

$$U(a,b) = \exp\left(-ib\frac{p^2}{2}\right)\exp\left(+i\frac{\log a}{2}\frac{PQ+QP}{2}\right),$$

so b and $-\log a/2$ must be "identified" as conjugate variables of $p^2/2$ and qp. We obtain

$$b \simeq q/p$$

which gives by symmetrization the operator

$$\tfrac{1}{2}(P^{-1}Q + QP^{-1})$$

and

$$(-\log a)/2 \simeq \log p \Rightarrow a \simeq P^{-2}$$

which gives the operator P^{-2}.

Consider now the even case.

As "origin" of our phase space we take the point

$$\begin{pmatrix} q_o \\ p_o \end{pmatrix}$$

defined by

$$\frac{1}{p_o^2} = \frac{(\varphi_e, P^{-2}\varphi_e)}{(\varphi_e,\varphi_e)} = 2 \Rightarrow p_o = \frac{1}{\sqrt{2}},$$

and

$$\frac{q_o}{p_o} = \frac{(\varphi_e, [(P^{-1}Q + QP^{-1})/2]\varphi_e)}{(\varphi_e,\varphi_e)} = 0,$$

so

$$\begin{pmatrix} q_o \\ p_o \end{pmatrix} = \begin{pmatrix} 0 \\ 1/\sqrt{2} \end{pmatrix}.$$

Now, identifying $\{a,b\}$ with the point image of the origin by $\{a,b\}$ we have

$$\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} \sqrt{a} & b/\sqrt{a} \\ 0 & 1/\sqrt{a} \end{pmatrix} \times \begin{pmatrix} 0 \\ 1/\sqrt{2} \end{pmatrix},$$

which gives

$$a = 1/2p^2 \quad \text{and} \quad b = q/p.$$

The same computation is valid also for the odd case and gives

$$a = 3/2p^2 \quad \text{and} \quad b = q/p.$$

Remarks: (1) Arguments given here are essentially heuristic,

(2) The operator $(P^{-1}Q + QP^{-1})/2$ is not self-adjoint.

C. Free evolution

If we consider the free movement of a particle with mass 1 its Hamiltonian is

$$H_o = \frac{P^2}{2} \text{ and in } \mathscr{H} \text{ it is } \underline{H}_o = -i\frac{\partial}{\partial z}.$$

Then the free evolution $e^{-i\underline{H}_o t}$ is just a translation by $-t$ in the argument of the function

$$e^{-i\underline{H}_o t}\begin{pmatrix} f_e \\ f_o \end{pmatrix}(z) = \begin{pmatrix} f_e(z-t) \\ f_o(z-t) \end{pmatrix}. \tag{5.12}$$

Remark: Since the variable z has the classical meaning of $q/p + \lambda i/p^2$ with $\lambda = \frac{1}{2}$ or $\frac{3}{2}$, one could expect the free evolution to be a classical change corresponding to

$$p(t) = p \quad \text{and} \quad q(t) = tp$$

so

$$\frac{q(t)}{p(t)} + \frac{\lambda i}{p(t)^2} = \left(\frac{q}{p} + t\right) + \frac{\lambda i}{p^2}.$$

The reason for the presence of the minus sign will be explained in the next section (VI).

D. Schrödinger equation

Consider a particle of mass $m = 1$ with Hamiltonian

$$H = P^2/2 + V,$$

where V is an operator in $L^2(R,dp)$ (potential part). Suppose for simplicity that the operator V conserves parity [the general case can be treated with a matrix valued kernel as in (5.2)]. Then the Schrödinger equation is written on each $\mathscr{H}_{1/2}$ or $\mathscr{H}_{-1/2}$:

$$\begin{aligned} i\frac{\partial}{\partial t}f_e(z,t) &= -i\frac{\partial}{\partial z}f_e(z,t) + \int V_e(z,\bar{z}')f_e(z',t)d\mu_{1/2}(z'), \\ i\frac{\partial}{\partial t}f_o(z,t) &= -i\frac{\partial}{\partial z}f_o(z,t) + \int V_o(z,\bar{z}')f_o(z',t)d\mu_{-1/2}(z'), \end{aligned} \tag{5.13}$$

where

$$\begin{aligned} V_e(z,\bar{z}') &= (\varphi_e^z, V\varphi_e^{\bar{z}'}), \\ V_o(z,\bar{z}') &= (\varphi_o^z, V\varphi_o^{\bar{z}'}), \end{aligned} \tag{5.14}$$

with φ_e^z and φ_o^z given by (5.7).

If now we define $|g(t)\rangle = e^{i\underline{H}_o t}|f(t)\rangle$ that is explicitly here

$$g(z,t) = f(z+t,t), \tag{5.15}$$

we obtain easily the Schrödinger equation in interaction picture

$$i\frac{\partial}{\partial t}g_e(z,t)=\int V_e(z+t,\bar{z}'+t)g_e(z',t)d\mu_{1/2}(z'),$$

$$i\frac{\partial}{\partial t}g_0(z,t)=\int V_0(z+t,\bar{z}'+t)g_0(z',t)d\mu_{-1/2}(z'). \tag{5.16}$$

Application: Quantum mechanics of a system of two fermions interacting with a local potential: a system of two particles in one dimension can be described, after removing the center of mass, by a wave function in $L^2(R)$. If these particles are fermions, the wave function is odd (because of the Fermi statistics). The kernel of the potential is then

$$V_o(z,\bar{z}')=(\varphi_o^z,V\varphi_o^{z'})=(\tilde{\varphi}_o^z,\tilde{V}\tilde{\varphi}_o^{z'}),$$

which, if the potential is local, is equal to

$$V_0(z,\bar{z}')=\int_{-\infty}^{+\infty}\overline{\tilde{\varphi}_o^z(q)}\mathscr{V}(q)\tilde{\varphi}_o^{z'}(q)dq, \tag{5.17}$$

where

$$\tilde{\varphi}_o^z(q)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{iqp}\frac{1}{2\pi^{3/4}}pe^{-\bar{z}(p^2/2)}dp$$
$$=\frac{i^{-1/2}}{2\pi^{3/4}}(\bar{z})^{-3/2}qe^{(i/\bar{z})(q^2/2)}$$

so we obtain

$$V_o(z,\bar{z}')=\frac{1}{4\pi^{3/2}}(z\bar{z}')^{-3/2}$$
$$\times\int_{-\infty}^{+\infty}q^2\mathscr{V}(q)\exp\left[i\left(\frac{z-\bar{z}'}{z\bar{z}'}\right)\frac{q^2}{2}\right]dq. \tag{5.18}$$

Also, $V_o(z,\bar{z}')$ can be written

$$V_o(z,\bar{z}')=\frac{1}{4\pi^{3/2}}(z\bar{z}')^{-3/2}$$
$$\times\int_0^{+\infty}\mathscr{W}(\lambda)\exp\left[i\left(\frac{z-\bar{z}'}{2z\bar{z}'}\right)\lambda\right]d\lambda, \tag{5.19}$$

with $\mathscr{W}(\lambda)=\lambda^{1/2}\mathscr{V}(\lambda^{1/2})$. So $V_o(z,\bar{z}')$ is closely related to the Laplace transform of $\mathscr{W}(\lambda)$.

E. Orthogonal basis for $\mathscr{H}$

In this section we study in greater detail the complete orthogonal systems V_n defined in the Appendix. We shall see that the $V_n^{-1/2}$ in $\mathscr{H}_{-1/2}$ (odd case) are eigenvectors of the harmonic oscillator (odd Hermite functions) and we shall compute the elements of $\mathscr{H}_{1/2}$ corresponding to the even Hermite functions. We shall also see that $V_n^{+1/2}$ in $\mathscr{H}_{1/2}$ (even case) are eigenvectors of the harmonic oscillator with a centrifugal force.

1. The V_n in $\mathscr{H}_{-1/2}$

Let us compute first of all the operator in $\mathscr{H}_{-1/2}$ corresponding to the harmonic oscillator Hamiltonian H:

$$H=\frac{Q^2+P^2}{2}=\frac{1}{2}\left[p^2-\frac{\partial^2}{\partial p^2}\right].$$

Then we have

$$H\varphi_o^{-\bar{z}}=\left[-i(z^2+1)\frac{\partial}{\partial z}-\frac{3}{2}iz\right]\varphi_o^{-\bar{z}}, \tag{5.20}$$

so that the operator H in $\mathscr{H}_{-1/2}$ is

$$\underline{H}\colon -i(z^2+1)\frac{\partial}{\partial z}-\frac{3}{2}iz.$$

The eigenvalue equation for $\underline{H}$ is then $\underline{H}f=(2n+\frac{3}{2})f$ which is exactly Eq. (A14) for $\alpha=-\frac{1}{2}$. Then the f_n are the images of the odd Hermite functions:

$$V_n^{-1/2}=A^o(h_{2n+1}),$$
$$V_n^{-1/2}(z)=\sqrt{\frac{2\Gamma(n+3/2)}{\pi^{3/2}(n)!}}\frac{(z-i)^n}{(z+i)^{n+3/2}}.$$

Now, to compute the images of the even Hermite functions h we make the following remark: It is easy to verify that

$$ph_{h-1}=\sqrt{\frac{n}{2}}nh+\sqrt{\frac{n-1}{2}}h_{h-2}.$$

Then, for $n-1$ even, we have

$$\int p^2e^{i(zp^2/2)}h_{h-1}dp$$
$$=\int pe^{i(zp^2/2)}\left\{\sqrt{\frac{n}{2}}h_h+\sqrt{\frac{n-1}{2}}h_{h-2}\right\}(p)dp,$$

so that we have

$$\omega_n\equiv A^e(h_{2n})$$
$$=\sqrt{\frac{c_o}{c_e}}\left\{\sqrt{\frac{2n+1}{2}}A^o(h_{2n+1})+\sqrt{\frac{2n}{2}}A^o(h_{2n-1})\right\}$$
$$=\sqrt{2n+1}V_n^{-1/2}+\sqrt{2n}V_{n-1}^{-1/2}.$$

Explicitly we have

$$\omega_{n-1}=2\sqrt{\frac{2\Gamma(n+\frac{1}{2})}{\pi^{3/2}(n-1)!}}\left\{\frac{n+1}{\sqrt{n}}\left(\frac{z-i}{z+1}\right)\right.$$
$$\left.+\sqrt{n}\right\}\frac{(z-i)^{n-1}}{(z+i)^{n+1/2}}.$$

2. The W_n in $\mathscr{H}_{1/2}$

The $V_n^{+1/2}$ in $\mathscr{H}_{1/2}$ verify the following equation:

$$\left[-i(z^2+1)\frac{\partial}{\partial z}-i\frac{5}{2}z\right]V_n^{+1/2}(z)$$
$$=\left(2n+\frac{5}{2}\right)V_n^{+1/2}(z).$$

An easy computation shows that

$$\left(-i(z^2+1)\frac{\partial}{\partial z}-i\frac{5}{2}z\right)p^2e^{iz(p^2/2)}$$
$$=\left(-\frac{1}{2}\frac{\partial^2}{\partial p^2}+\frac{1}{p^2}+\frac{p^2}{2}\right)p^2e^{iz(p^2/2)}.$$

So the harmonic oscillator with a centrifugal force is in $\mathscr{H}_{1/2}$ the following:

$$-i(z^2+1)\frac{\partial}{\partial z}-i\frac{5}{2}z.$$

The

$$V_n^{+1/2}(z)=4\sqrt{\frac{\Gamma(n+\frac{5}{2})}{\pi^{3/2}n!}}\left(\frac{z-i}{z+i}\right)^n(z+i)^{-5/2}$$

are eigenvectors, with eigenvalues $(2n+\frac{5}{2})$ of this operator, which is a special case of operators studied by Calogero.[16]

VI. REPRESENTATION OF THE CANONICAL TRANSFORMATIONS

A. Odd case

Remember that every element C of SL(2,R) can be written as

$$C=\begin{pmatrix}\sqrt{y} & x/\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix},$$

with (6.1)

$$x+iy=\frac{ai+b}{ci+d}\equiv S\cdot i,\quad \theta=\arg[ci+d\,].$$

if $\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in$ SL(2,R) we have also

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}C=\begin{pmatrix}\sqrt{y'} & x'/\sqrt{y'}\\ 0 & 1/\sqrt{y'}\end{pmatrix}\begin{pmatrix}\cos\theta' & -\sin\theta'\\ \sin\theta' & \cos\theta'\end{pmatrix},$$

with (6.2)

$$x'+iy'\equiv C\cdot z\equiv\frac{az+b}{cz+d},\quad \theta'=\theta+\arg[cz+d\,].$$

Now, consider the metaplectic representation of SL(2,R),[17] $U(S)$, and consider the transform of $U(S)\psi$, with $\psi\in L^2(R,dp)$, ψ odd. We have

$$\sqrt{C_o}(A_oU(S)\psi)(x+iy)=y^{-3/4}(U(\{x,y\})\varphi_o,\,U(S)\psi),$$

where

$$\{x,y\}=\begin{pmatrix}\sqrt{y} & x/\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}$$
$$=y^{-3/4}(U(S^{-1})U(\{x,y\})\varphi_o,\psi)$$
$$=y^{-3/4}\left(U(S^{-1})\begin{pmatrix}\sqrt{y} & x/\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}\varphi_o,\psi\right).$$

Now we have, if $S^{-1}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$,

$$S^{-1}\begin{pmatrix}\sqrt{y} & x/\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}$$
$$=\begin{pmatrix}\sqrt{y'} & x'/\sqrt{y'}\\ 0 & 1/\sqrt{y'}\end{pmatrix}\times\begin{pmatrix}\cos\theta' & -\sin\theta'\\ \sin\theta' & \cos\theta'\end{pmatrix},$$

with $x'+iy'=(az+b)/(cz+d)$, where $z=x+iz$ and $\theta'=\arg[cz+d\,]$ so

$$\sqrt{C_o}(A_oU(S)\psi)(z)$$
$$=y^{-3/4}(U(\{x'+y'\})\,U\begin{pmatrix}\cos\theta' & -\sin\theta'\\ \sin\theta' & \cos\theta'\end{pmatrix}\varphi_o,\psi), \tag{6.3}$$

but, because φ^o is the first excited state of the harmonic oscillator and the generator of the subgroup SO(2,R) of SL(2,R) is the harmonic oscillator Hamiltonian, we have

$$U\begin{pmatrix}\cos\theta' & -\sin\theta'\\ \sin\theta' & \cos\theta'\end{pmatrix}\varphi_o=e^{-i(3/2)\theta'}\varphi_o,$$
$$y^{-3/4}e^{-i(3/2)\theta'}=y^{-3/4}(e^{i\theta'})^{-3/2}$$
$$=(cz+d)^{-3/2}\left[\frac{y}{(cz+d)(c\bar{z}+d)}\right]^{-3/4}$$
$$=(cz+d)^{-3/2}\left[\frac{(z-\bar{z})/2i}{(cz+d)(c\bar{z}+d)}\right]^{-3/4}$$
$$=(cz+d)^{-3/2}\left[\mathrm{Im}\left(\frac{az+b}{cz+d}\right)\right]^{-3/4}. \tag{6.4}$$

So we have

$$\sqrt{C_o}(A_oU(S)\psi)(z)$$
$$=\mathrm{Im}\left(\frac{az+b}{cz+d}\right)(cz+d)^{-3/2}(U(\{x',y'\})\varphi_o,\psi)$$
$$=(cz+d)^{-3/2}\sqrt{C_o}(A_o\psi)\left(\frac{az+b}{cz+d}\right).$$

Then we have the following result:

$$(A_o(U(S)\psi)(z)=(cz+d)^{-3/2}(A_o\psi)\left(\frac{az+b}{cz+d}\right).$$

We have then the following representation of SL(2,R):

$$(V(S)f)(z)=(cz+d)^{-3/2}f\left(\frac{az+b}{cz+d}\right), \tag{6.5}$$

with $S^{-1}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$.

Remarks: (1) The space $\mathscr{H}_{-1/2}$ has been used by Itzykson, but considered as the space of analytic functions on the disk, see Ref. 3.

(2) This representation is a projective representation of SL(2,R) with multiplier ± 1 given by the factor $(cz+d)^{-3/2}$.

(3) The representation of the subgroup of free evolution on phase space, that is $\begin{pmatrix}1 & t\\ 0 & 1\end{pmatrix}$, is then

$$\left[V\begin{pmatrix}1 & t\\ 0 & 1\end{pmatrix}f\right](z)=f(z-t).$$

This explains the minus sign in the free evolution (see 5.12) because SL(2,R) acts on the argument by inverse transform.

B. Even case

The calculation for the even case begins in the same manner and we obtain a formula identical to (6.3):

$$\sqrt{C_e}(A_eU(S)\psi)(z)$$
$$=y^{-5/4}(U(\{x'+y'\})U\left(\begin{pmatrix}\cos\theta' & -\sin\theta'\\ \sin\theta' & \cos\theta'\end{pmatrix}\right)\varphi_e,\psi). \tag{6.6}$$

Now of course φ_e is not an eigenvector of the harmonic oscillator but we remark that

$$\varphi^e=(\pi^{1/4}/2)[h_0+\sqrt{2}h_2], \tag{6.7}$$

where h_0 and h_2 are the normalized Hermite functions of order 0 and 2. So we obtain

$$U\begin{pmatrix}\cos\theta' & -\sin\theta'\\ \sin\theta' & \cos\theta'\end{pmatrix}\varphi_e$$
$$=e^{-i(5/2)\theta'}\varphi^e+\frac{\pi^{1/4}}{2}[e^{-i\theta'/2}-e^{-i(5/2)\theta'}]h_0.$$

The first term $(e^{-i(5/2)\theta}\varphi^e)$ gives in (5.6) the contribution

$$\sqrt{C_e}(cz+d)^{-5/2}f((az+b)/cz+d)).$$

For the second term we remark that

$$(U(\{x'+y'\})h_0)(p)=y'^{1/4}e^{-i\bar{z}'(p^2/2)}\pi^{-1/4}, \tag{6.8a}$$

with $z'=x'+iy'$ and [recall that $z'=(az+b/(cz+d)$]

$$(\pi^{1/4}/2)y^{-5/4}[e^{-i\theta'/2}-e^{-i(5/2)\theta'}]y'^{1/4}$$
$$=(2i)(\pi^{1/4}/2)(cz+d)^{-3/2}c, \tag{6.8b}$$

so the second term in the left-hand side of (6.6) is

$$(i)(cz+d)^{-3/2}c\int_{-\infty}^{+\infty} e^{iz'p2/2}\psi(p)dp$$
$$= -\sqrt{C_e}\,\frac{i}{2}(cz+d)^{-3/2}c\int_0^{+\infty} f(z'+i\lambda)d\lambda \qquad (6.9)$$

(because of estimates on f this integral always exists); so we obtain the final expression

$$(V(S)f)(z) = (cz+d)^{-5/2} f\left(\frac{az+d}{cz+d}\right)$$
$$+\frac{i}{2}c(cz+d)^{-3/2}\int_0^{+\infty} f\left(\frac{az+b}{cz+d}+i\lambda\right)d\lambda. \qquad (6.10)$$

Remarks: (1) For the even case, Itzykson used another space of analytic functions constructed on the same manner but with $\varphi_e(p) = e^{-p^2/2}$ which does not satisfy the admissibility condition (see Ref. 3).

(2) If we restrict the representation to the "$ax+b$" group then $c=0$ and we have the following representation:

$$(V\begin{pmatrix}\sqrt{a} & b/\sqrt{a}\\ 0 & 1/\sqrt{a}\end{pmatrix}f)(z) = = a^{5/4}f(az+b).$$

(3) The same argument for the free evolution as in the odd case holds.

VII. GENERALIZATION TO HIGHER DIMENSIONS

In this section we describe briefly the possible generalizations to a higher dimension.

The generalization of the Lobatschevski space in the Siegel half-plane, i.e., the set of symmetric complex matrices Z with an imaginary part positive definite.

A good candidate for the transform is in the even case,

$$f_e(Z): \int e^{i\bar{p}Zp}\psi(p)d^Np$$

(up to a constant), and in the odd case the N transforms

$$f_{o_i}(Z) = \int p_i e^{i\bar{p}Zp}\psi(p)d^Np$$

(up to a constant). We can transport the even and odd Hermite functions and define on this space of analytic functions in Z a Hilbert space structure by defining the system of the image of H_n as orthonormal (this space is the space used by Itzykson, but on the disk).[3]

In this space the metaplectic representation is the following: let S in SL(2N,R) with

$$S^{-1} = \begin{pmatrix}A & B\\ C & D\end{pmatrix},$$

we have for the even case

$$(U(S)f_e)(z) = \{\det[CZ+D]\}^{-1/2}$$
$$\times f_e((AZ+D)(CZ+D)^{-1}),$$

and for the odd case,

$$(U(S)\mathbf{f}_o)_i(z) = \det[CZ+D]^{-1/2}\sum_{j=1}^{N}(CZ+D)_{ij}^{-1}$$
$$\times f_{oj}((AZ+B)(CZ+D)^{-1}).$$

VIII. OTHER TRANSFORMS

(i) Instead of using the generators $P^2/2$ and $(PQ+QP)/4$ of the "$ax+b$" group, we can use the generator Q and $(PQ+QP)/2$ on $L^2(R^+,dq)$. This gives the family of unitary transforms between $L^2(R^+,dq)$ and $\mathscr{H}_{2\alpha-1}$:

$$f_\alpha(z) = \frac{2^\lambda}{\sqrt{2\pi\Gamma(2\lambda)}}\int_0^{+\infty} q^\alpha e^{izq}\psi(q)dq,$$

with $\alpha>0$.

This map is a unitary map between $L^2(R^+,dq)$ and the space $\mathscr{H}_{2\alpha-1}$ of analytic functions used by Berezin to quantize the Lobatchevski space.[14]

In this space, the variable z can be considered as the classical quantity $p+i\alpha/q$, $q>0$.

(ii) We can also consider the map A_α on the radial part of the wave function in $L^2(R^N)$, that is

$$L^2(R^N) = L^2(R^+,p^{N-1}dp)\otimes L^2(S^{N-1},d\Omega)$$
$$A_\alpha \downarrow$$
$$\mathscr{H}_{\alpha+N/2-2}\oplus L^2(S^{N-1},d\Omega),$$

by

$$\psi \xrightarrow{A_\alpha} \int p^\alpha e^{iz(p^2/2)}\psi(p,\Omega)p^{N-1}\,dp\,d\Omega$$

(up to a constant). This map is unitary and in this space the operator

$$\frac{1}{4H_0}\mathbf{P}\cdot\mathbf{X}+\mathbf{X}\cdot\mathbf{P}\frac{1}{4H_0}$$

(time operator), where $H_0 = p^2/2$ is expressed by the kernel $\rho_{\alpha+N/2-2}(z-\bar{z}')[(z+\bar{z}')/2]$.

APPENDIX: PROOFS OF STATEMENTS OF SEC. IV

In this Appendix we compute the reproducing kernel and an orthogonal basis of $\mathscr{H}_\alpha$, $\alpha > -1$ (see Refs. 13–15).

First of all we remark that $\mathscr{H}_\alpha = \{f$, analytic function on Π such that $\int|f(z)|^2\,d\mu_\alpha(z) < +\infty\}$, where $(z=b+ia)$ $d\mu_\alpha(z) = a^\alpha\,da\,db$ is unitary equivalent to the space $\mathscr{D}_\alpha$ of analytic functions on the disk $D = \{\beta, |\beta|<1\}$ square integrable with respect to the measure $(\beta = x+iy)d\nu_\alpha(\beta)$ $= ((1-\beta\bar{\beta})/2)^\alpha\,dx\,dy$.

The unitary transform between $\mathscr{H}_\alpha$ and $\mathscr{D}_\alpha$ is B:

$$\mathscr{H}_\alpha \xrightarrow{B} \mathscr{D}_\alpha,$$

$$f \to g = Bf,$$

with

$$g(\beta) = 2^{\alpha/2+1}\left(\frac{1-\beta}{i}\right)^{-\alpha-2} f\left(i\frac{1+\beta}{1-\beta}\right), \qquad (A1)$$

$$f(z) = 2^{\alpha/2+1}(z+i)^{-\alpha-2}g\left(\frac{z-i}{z+i}\right). \qquad (A2)$$

It is easy to show that in $\mathscr{D}_\alpha$ the system of functions u_n, defined by

$$u_n(\beta) = \frac{2^{\alpha/2}}{\sqrt{\pi}}\sqrt{\frac{\Gamma(n+\alpha+2)}{\Gamma(n+1)\Gamma(\alpha+1)}}\,\beta^n, \qquad (A3)$$

forms an orthogonal system.

Furthermore, this system is complete[18]: since g is analytic in D we can write

$$g(\beta)=\sum_{n=0}^{\infty} g_n \beta^n=\sum_{n=0}^{\infty} c_n u_n(\beta).$$

So we just need to show that the series converges in L^2 norm. Let $0<r<1$, then we have

$$\int_{|\beta|<r}|g(\beta)|^2 d\nu_\alpha(\beta)$$
$$=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} C_n\overline{C}_m \int_{|\beta|<r} u_n(\beta)\,\overline{u_m(\beta)}d\nu_\alpha(\beta).$$

The integration over angles gives a δ_{nm} and taking the limit $r\to 1$ we obtain

$$\int |g(\beta)|^2 d\nu_\alpha(\beta)=\sum_{n=0}^{\infty}|C_n|^2$$

so the series is L^2-convergent.

According to Bergmann,[19] the function defined by

$$r_\alpha(\beta,\bar{\beta}')=\sum_{n=0}^{\infty} u_n(\beta)\,\overline{u_n(\beta')} \tag{A4}$$

is a reproducing kernel for the space, i.e., we have for $g\in\mathscr{D}_\alpha$

$$\int r_\alpha(\beta,\bar{\beta}')g(\beta')d\nu_\alpha(\beta')=g(\beta). \tag{A5}$$

Since

$$\sum_{n=0}^{\infty}\frac{2^\alpha}{\pi}\frac{\Gamma(n+\alpha+2)}{\Gamma(n+1)\Gamma(\alpha+1)}(\beta\bar{\beta}')^n$$
$$=\left(\frac{\alpha+1}{4\pi}\right)\left(\frac{1-\beta\bar{\beta}'}{2}\right)^{-(\alpha+2)},$$

we have

$$r_\alpha(\beta,\bar{\beta}')=\left(\frac{\alpha+1}{4\pi}\right)\left(\frac{1-\beta\bar{\beta}'}{2}\right)^{-(\alpha+2)}. \tag{A6}$$

By the unitary transform B we obtain from (A2) the kernel for $\mathscr{H}_\alpha$:

$$\rho_\alpha(z-\bar{z}')=\left(\frac{\alpha+1}{4\pi}\right)\left(\frac{z-\bar{z}'}{2i}\right)^{-(\alpha+2)}, \tag{A7}$$

$$\forall f\in\mathscr{H}_\alpha,\quad \int f(z')\rho_\alpha(z-\bar{z}')d\mu_\alpha(z')=f(z). \tag{A8}$$

In (A6) and (A7) the power is taken as analytic continuation from the real axis for ρ_α and from the circle for r_α, i.e.,

$$\left(\frac{z-\bar{z}'}{2i}\right)^{-\alpha-2}=\left|\frac{z-\bar{z}'}{2i}\right|^{-(\alpha+2)}$$
$$\times\exp\left[-i(\alpha+2)\arg\left(\frac{z-\bar{z}'}{2i}\right)\right],$$

with

$$-\frac{\pi}{2}<\arg\left(\frac{z-\bar{z}'}{2i}\right)\leqslant+\frac{\pi}{2},$$

and

$$\left(\frac{1-\beta\bar{\beta}'}{2}\right)^{-(\alpha+2)}=\left|\frac{1-\beta\bar{\beta}'}{2}\right|^{-(\alpha+2)}$$
$$\times\exp\left[-i(\alpha+2)\arg\left(\frac{1-\beta\bar{\beta}'}{2}\right)\right].$$

From (A5) and (A8) we deduce, by the Cauchy–Schwartz inequality, that

$$|g(\beta)|\leqslant\{\|g\|\}_\alpha\sqrt{\frac{\alpha+1}{4\pi}}\left(\frac{1-|\beta|^2}{2}\right)^{-\alpha/2-1}, \tag{A9}$$

$$|f(z)|\leqslant\|f\|_\alpha\sqrt{\frac{\alpha+1}{4\pi}}(\mathrm{Im}(z))^{-\alpha-1/2}, \tag{A10}$$

where

$$\{\|g\|\}_\alpha^2=\int g(\theta)\,\overline{g(\theta)}d\nu_\alpha(\theta),$$

and

$$\|\rho\|_\alpha^2=\int f(z)\overline{f(z)}d\mu_\alpha(z).$$

From (A9) it is easy to see that $\mathscr{D}_\alpha$ is a Hilbert space: $\mathscr{D}_\alpha\subset L^2(D,d\nu_\alpha(\theta))$ which is a Hilbert space; so every Cauchy sequence in $\mathscr{D}_\alpha$ converges in $L^2(D,d\nu_\alpha(\theta))$. Now because of (A9) the restriction of a Cauchy sequence $\{g_n\}$ to a compact subset E of D verifies $\forall\theta\in E$,

$$|g_n(\theta)-g_m(\theta)|$$
$$\leqslant\{\|g_n-g_m\|\}_\alpha\sqrt{\frac{\alpha+1}{4\pi}}\left(\frac{1-|\theta|^2}{2}\right)^{-\alpha/2-1}$$
$$\leqslant\{\|g_n-g_m\|\}_\alpha\sqrt{\frac{\alpha+1}{4\pi}}\,C,$$

for some C so the restriction of $\{g_n\}$ to E is uniformly Cauchy convergent and $\{g_n\}$ is uniformly convergent on each compact. It is well known that every sequence of analytic functions uniformly convergent on each compact converges to an analytic function so $\mathscr{D}_\alpha$ is complete. The same argument holds for $\mathscr{H}_\alpha$.

With B we can also obtain an orthogonal complete basis for $\mathscr{H}_\alpha$; the result is

$$V_n(z)=\frac{2^{\alpha+1}}{\sqrt{\pi}}\sqrt{\frac{\Gamma(\alpha+2+n)}{\Gamma(n+1)\Gamma(\alpha+1)}}\frac{(z-i)^n}{(z+i)^{n+\alpha+2}}. \tag{A11}$$

Remark: It is easy to verify that the functions V_n verify the equation

$$\left[-i(z^2+1)\frac{\partial}{\partial z}-i(\alpha+2)z\right]V_n(z)=(2n+\alpha+2)V_n(z). \tag{A12}$$

Special cases: For $\mathscr{H}_{-1/2}$ we obtain

$$V_n^{-1/2}(z)=\sqrt{\frac{2\Gamma(n+3/2)}{(\pi)^{3/2}n!}}\left(\frac{z-i}{z+i}\right)^n(z+i)^{-3/2}, \tag{A13}$$

which verify

$$\left(-i(z^2+1)\frac{\partial}{\partial z}-\frac{3}{2}iz\right)V_n^{-1/2}=\left(2n+\frac{3}{2}\right)V_n^{-1/2}. \tag{A14}$$

For $\mathscr{H}_{1/2}$ we have

$$V_n^{+1/2}(z)=4\sqrt{\frac{\Gamma(n+5/2)}{\Pi^{3/2}n!}}\left(\frac{z-i}{z+i}\right)^n(z+i)^{-5/2}, \tag{A15}$$

which verify

$$\left[-i(z^2+1)\frac{\partial}{\partial z} - \frac{5}{2}iz \right] V_n^{+1/2} = \left(2n + \frac{5}{2}\right) V_n^{+1/2}. \tag{A16}$$

ACKNOWLEDGMENTS

It is a great pleasure to thank Alex Grossmann for motivation on this work, many helpful discussions during its elaboration, and reading the manuscript.

The final draft of this paper was written in Bielefeld where the author enjoyed the stimulating atmosphere of the Zentrum für interdisziplinäre Forschung der Universität Bielefeld.

The author received financial support from Direction Generale de la Recherche Scientifique et Technique (DGRST).

[1]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[2]V. Bargmann, Commun. Pure Appl. Math. **20**, 1 (1967).
[3]C. Itzykson, Commun. Math. Phys. **4**, 92 (1967).
[4]C. van Winter, Trans. Math. Soc. **162**, 103 (1971).
[5]E. W. Aslasken and J. R. Klauder, J.Math. Phys. **10**, 2267 (1969).
[6]M. Duflo and C. C. Moore, J. Funct. Anal. **21**, 209 (1976).
[7]A. Grossman and J. Morlet (to appear in SIAM Math. Anal.).
[8]A. Grossman, J.Morlet, and T. Paul (in preparation).
[9]A. L. Carey, Bull. Austral. Math. Soc. **15**, 1 (1976).
[10]N. Y. Vilenkin, *Special functions and the theory of group representations* (A.M.S., Providence, RI, 1968).
[11]E. W. Aslasken and J. R. Klauder, J.Math. Phys. **9**, 206 (1968).
[12]I. M. Gelfand and M. A. Naimark, Dokl. Akad. Nauk SSSR **55**, 570 (1947).
[13]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[14]F. A. Berezin, Commun. Math. Phys. **40**, 153 (1975).
[15]T. Paul, Thèse de troisième cycle, Université Paris VI, 1984.
[16]F. Calogero, J. Math. Phys. **10** (1969).
[17]P. Kramer, M. Moshinsky, and T. H. Seligman, "Complex extensions of canonical transformations," in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. III.
[18]S. Lang, *SL (2,R)* (Addision–Wesley, Reading, MA, 1975).
[19]S. Bergman, *The Kernel Function and Conformal Mapping, Mathematical Surveys No. 5* (A.M.S., New York, 1950).

Quantum systems with time-dependent harmonic part and the Morse index

Jorge Rezende
Fakultät für Physik, Universität Bielefeld, D-4800 Bielefeld 1, Federal Republic of Germany

(Received 9 August 1983; accepted for publication 6 April 1984)

A simple relation between two quantum systems with a time-dependent, respectively, time-independent, harmonic part is established. Using this we give a computation, valid for all times, of the Green's functions of the time-dependent harmonic oscillator with and without a perturbation of the type g/x^2. The asymptotic expansion of the wave function in powers of Planck's constant is discussed using a new representation of the Morse index.

PACS numbers: 03.65.Ge, 03.65.Sq

I. INTRODUCTION

A canonical change of variables has been used by several authors, see, e.g., Refs. 1–10, both in classical and quantum mechanics, in order to reduce a dynamical system containing a time-dependent harmonic potential to a simpler one.

To set up the framework developed in the present paper we give first a simple example of such a tranformation. Consider the Schrödinger equation

$$\left[ih\frac{\partial}{\partial t}+\frac{h^2}{2}\frac{d^2}{dx^2}-\frac{\lambda^2}{2}x^2\right]\psi_\lambda(x,t)=0, \tag{1.1}$$

where h is Planck's constant divided by 2π and $\lambda\geqslant 0$. We distinguish two cases: the free case ($\lambda=0$) and the time-independent harmonic oscillator case (TIHO: $\lambda>0$). Let $\psi_0(x,t)$ be a solution of (1.1) for $\lambda=0$. Then it is easy to see that $\psi_\lambda(x,t)$ is given by

$$\psi_\lambda(x,t)=(\cos\lambda t)^{-1/2}\exp[-i\lambda\tan\lambda t x^2/2h]$$
$$\times\psi_0(x/\cos\lambda t,\tan\lambda t/\lambda). \tag{1.2}$$

But it must be emphasized that the solution (1.2) of (1.1) is valid only for $|t|<\pi/2\lambda$. In other words, the change of variables under discussion does not give the time evolution of the wave function ψ_λ for all times if $\lambda>0$. A similar difficulty arises in the corresponding classical situation.

A general expression, valid for each t such that $\cos\lambda t\neq 0$, is obtained by taking into account the Maslov correction[11–23] and is given by

$$\psi_\lambda(x,t)=|\cos\lambda t|^{-1/2}\exp[-i\pi m/2-i\lambda\tan\lambda t x^2/2h]$$
$$\times\psi_0(x/\cos\lambda t,\tan\lambda t/\lambda), \tag{1.3}$$

where m is the Morse (Maslov) index, $m=\mathrm{int}(1/2+\lambda t/\pi)$; int ϵ is the integer number such that $\epsilon-1<\mathrm{int}\,\epsilon\leqslant\epsilon$.

We learn from this example that the TIHO system can be essentially reduced to a countable number of free systems using a countable number of canonical transformations chosen in order to have continuity for the time evolution of the wave function. This rule introduces the Maslov correction consisting in a jump of the phase at every half-period.

This phenomenon has been observed long ago as pointed out in Ref. 14 (see also the references therein and Ref. 17): "In 1890 Gouy observed and explained the phase gained by a wave as it goes through a focus. Similar phase shifts occur in the wave function of quantum systems; they have been derived by Keller from the single valuedness of the wave function and by Gutzwiller who established their relationship with the Morse index of the corresponding classical trajectory."

If we replace λ^2 in Eq. (1.1) by a real continuous function $p(t)$ we obtain the Schrödinger equation corresponding to the quantum mechanical system describing the time-dependent harmonic oscillator (TDHO). We shall prove in this paper that one can reduce every TDHO system to every TIHO one by a simple canonical transformation without worrying anymore with the Maslov correction. We shall also apply these ideas to more general systems containing a time-dependent harmonic potential.

Let us now describe shortly the structure of the paper.

In Sec. II we discuss the TDHO differential equation

$$\ddot{u}(t)+p(t)u(t)=0, \tag{1.4}$$

which has been already studied by several authors in relation with the theory of exact invariants for the TDHO.[2–8,24–58] Here we use the results stated in Refs. 59 and 60. For a class of solutions of (1.4) we introduce the notion of an index function.

Section III is devoted to a detailed analysis of the canonical transformation which reduces a quantum system containing a time-dependent harmonic potential to a simpler one. We obtain in this way the quantum analog of the transformation established by Perelomov[9] in the classical case.

In Sec. III we apply the results of the previous sections to calculate some Green's functions with Maslov corrections. Let us now mention some previous related work and the connections with ours. The Green's function of the TDHO, for small times, was first calculated in Ref. 61 using the Magnus formula.[62] The Maslov correction was given in Ref. 63. (See also Refs. 21 and 29.) Our global (i.e., valid for all times) and canonical calculation is based upon well-established mathematical results.[13,22] The Green's function of the TDHO with a perturbation of the type g/x^2 for small times, was given in Ref. 29 using the Feynman path integral formulation.[64] Our result includes the Maslov correction.

In Sec. V we wish to comment upon the connection between Maslov's results and the TDHO. We point out that the problem of defining, in the semiclassical approximation, the Morse index, arises in all its complexity already in the TDHO case. Finally, we investigate the asymptotic behavior, at time t and as $h\to 0$, of the wave function using the fact that, in this case, the Schrödinger equation is explicitly soluble. Let us add the following remark to emphasize the advantage of studying the TDHO; in fact, this approach allows us to use the method of stationary phase in finite dimension,[65,66] avoiding, therefore, the complexity of this method in Hilbert space.[67–69]

 0022-2488/84/113264-06$02.50

II. THE DIFFERENTIAL EQUATION OF THE TIME-DEPENDENT HARMONIC OSCILLATOR

We consider the differential equation in $\mathbb{R}$

$$\ddot{u}(t) + p(t)u(t) = 0, \tag{2.1}$$

where $p(t)$ is a real-valued continuous function. The solutions of this equation have beautiful properties. For example, the zeros of any nontrivial solution are isolated and simple. For more details see, e.g., Refs. 59 and 60.

In what follows u_1 and u_2 denote two independent solutions of (2.1) with Wronskian $w(u_1,u_2) = w = u_1\dot{u}_2 - \dot{u}_1u_2$. As $w \neq 0$, we have that if one of these two solutions vanishes for some $t \in \mathbb{R}$ the other is automatically different from zero in this point.

Let s and c be the two independent solutions of (2.1) such that $s(0) = \dot{c}(0) = 0$ *and* $c(0) = \dot{s}(0) = 1$, which implies that $w(c,s) = 1$. We define now

$$\xi = Pc^2 + 2Qcs + Rs^2 = \bar{P}u_1^2 + 2\bar{Q}u_1u_2 + \bar{R}u_2^2, \tag{2.2}$$

for $P,Q,R,\bar{P},\bar{Q},\bar{R} \in \mathbb{R}$ where (P,Q,R) and $(\bar{P},\bar{Q},\bar{R})$ are related by some one-to-one correspondence.

Lemma 1: (i) $w^3(\bar{P}\bar{R} - \bar{Q}^2) = PR - Q^2$, where $w = w(u_1,u_2)$. (ii) We assume that $P,R \geqslant 0$ and $P + R > 0$. Then $PR - Q^2 \geqslant 0$ implies $\xi(t) \geqslant 0$; the same result holds if $\geqslant$ is replaced by $>$.

Proof: Lemma 1 follows immediately from the fact that c and s do not vanish simultaneously.

From now on we assume that $PR - Q^2 = \lambda^2$ with $\lambda \geqslant 0$ and that $P, R \geqslant 0$, $P + R > 0$.

For every such λ we have the following important relation involving ξ (see Ref. 59):

$$2\xi\dddot{\xi} - \dot{\xi}^2 + 4p\xi^2 - 4\lambda^2 = 0. \tag{2.3}$$

We denote by $]m_2, M_2[$ a maximal interval where ξ does not vanish. For $\theta \in]m_2, M_2[$ we define the function η: $]m_2, M_2[\to]m_1, M_1[$ by

$$\eta(t) = \int_\theta^t \xi(\tau)^{-1}\, d\tau, \tag{2.4}$$

with $M_1 = \lim_{t \to M_2} \eta(t)$, $m_1 = \lim_{t \to m_2} \eta(t)$, where $M_1 \in]0, +\infty]$ and $m_1 \in [-\infty, 0[$. We emphasize that η is well defined because $\xi(t) > 0$ for every $t \in]m_2, M_2[$ and that η is strictly increasing in $]m_2 M_2[$. Finally η is a smooth bijection between $]m_2, M_2[$ and $]m_1, M_1[$.

We distinguish now two cases: $\lambda > 0$ and $\lambda = 0$.

(a) $\lambda > 0$: From Lemma 1 we get that $m_2 = -\infty$ and $M_2 = +\infty$. The general theory discussed in Ref. 59 establishes that

$$u(t) = \xi(t)^{1/2}(A\cos(\lambda\eta(t)) + B\sin(\lambda\eta(t))) \tag{2.5}$$

is the general solution of (2.1), where A and B are arbitrary constants. We remark that (see, e.g., Refs. 8 and 35) $v = \pm\xi^{1/2}$ is the general solution of the equation

$$\ddot{v} + pv - \lambda^2 v^{-3} = 0. \tag{2.6}$$

Equations (2.1) and (2.6) appear in the theory of exact invariants for the time-dependent harmonic oscillator[2-8,24-58] but we do not go into this subject here.

(b) $\lambda = 0$: In this case we have that

$$\xi = (P^{1/2}c \pm R^{1/2}s)^2 = u_1^2, \tag{2.7}$$

where u_1 is some solution of (2.1). It is now easy to see that there exists a unique u_2, with $w(u_1, u_2) = 1$, such that

$$\eta = u_2/u_1. \tag{2.8}$$

Now we recall that for $z \in \mathbb{C}$, with $-\pi < \arg z < \pi$, $\log z$ is defined such that $\operatorname{Im} \log z = \arg z$. For $\epsilon \in \mathbb{R}$, we define as usual $z^\epsilon = \exp[\epsilon \log z]$. Let u be a solution of the equation (2.1). We denote by $m(u,t)$: $\mathbb{R} \to \mathbb{Z}$ the *index function* of u; $m(u,t)$ is such that for $t_2 > t_1$, $m(u,t_2) - m(u,t_1)$ is the number of zeros of u in the interval $]t_1, t_2]$ and $m(u,0) = 0$.

If there exists some $\bar{t} > 0$ such that $\tau \in]0,\bar{t}[$ implies $u(\tau) > 0$, we define for ϵ, $t \in \mathbb{R}$

$$u^\epsilon(t) = |u(t)|^\epsilon \exp[i\epsilon\pi m(u,t)]. \tag{2.9}$$

Example: If we put in (2.2) $P = 1$, $Q = 0$, $R = \lambda^2$ and in (2.4) $\theta = 0$ we have $\xi = c^2 + \lambda^2 s^2$ and

$$s(t) = \lambda^{-1}\xi(t)^{1/2}\sin(\lambda\eta(t)), \tag{2.10}$$

$$c(t) = \xi(t)^{1/2}\cos(\lambda\eta(t)). \tag{2.11}$$

Let first $p(t) = \mu^2$ with $\mu > 0$. Then we have $s(t) = \mu^{-1}\sin\mu t$ and $c(t) = \cos\mu t$. In this situation $m(s,t) = \operatorname{int}(\mu t/\pi)$ [and $m(c,t) = \operatorname{int}(\mu t/\pi + 1/2)$.] Here one has $\xi(t) = \cos^2\mu t + \lambda^2\mu^{-2}\sin^2\mu t$ and $\xi(t) \geqslant \min\{1, \lambda^2/\mu^2\}$. For $\mu = 1$ one can easily check that (cf. Ref. 20)

$$\sin^{1/2} t = \exp[i(t/2 - \pi/4)](1 - i\cot t)^{-1/2}, \tag{2.12}$$

$$\cos^{1/2} t = \exp[i(t/2)](1 + i\tan t)^{-1/2}. \tag{2.13}$$

Now in the general case, where p is not constant, using formulas (2.10) and (2.11) we have $c^{1/2}(t) = \xi(t)^{1/4}\cos^{1/2}(\lambda\eta(t))$, $m(c,t) = \operatorname{int}(\lambda\eta(t)/\pi + 1/2)$, and so on.

III. A SIMPLE RELATION BETWEEN TWO QUANTUM SYSTEMS

In this section the notations are the same as in Sec. II. Let D_2 be a domain (open and connected set) such that $D_2 \subset \mathbb{R} \times]m_2, M_2[$. For $(x,t) \in D_2$ we introduce the following change of variables:

$$(x,t) \to (\sigma\xi(t)^{-1/2}x, \eta(t)), \tag{3.1}$$

where $\sigma = \pm 1$. It is obvious that this transformation defines a smooth bijection between D_2 and its image $D_1 \subset \mathbb{R} \times]m_1, M_1[$.

Consider now the two quantum systems associated with the following Schrödinger equations defined, respectively, in D_1 and D_2:

$$\left[ih\frac{\partial}{\partial t} + \frac{h^2}{2}\Delta - \frac{\lambda^2}{2}x^2 - V(x,t)\right]\phi(x,t) = 0, \tag{3.2}$$

$$\left[ih\frac{\partial}{\partial t} + \frac{h^2}{2}\Delta - \frac{p(t)}{2}x^2 - \xi(t)^{-1}V(\sigma\xi(t)^{-1/2}x, \eta(t))\right]\psi(x,t) = 0, \tag{3.3}$$

where $V(x,t)$ is a real-valued continuous function in D_1 and $\phi(x,t)$ and $\psi(x,t)$ are continuously differentiable with respect to t and twice continuously differentiable with respect to x.

Using (2.3) one can now prove the following theorem.

Theorem 1: Let $\phi(x,t)$ be a solution in D_1 of Eq. (3.2). Then

$$\psi(x,t) = \xi(t)^{-1/4}\exp[i\dot{\xi}(t)x^2/4h\xi(t)] \times \phi(\sigma\xi(t)^{-1/2}x, \eta(t)) \tag{3.4}$$

is a solution in D_2 of Eq. (3.3).

For $\lambda = 0$, using (2.7) and (2.8), we have that $\xi(t) = u_1(t)^2, \xi(t)^{-1/2} = |u_1(t)|^{-1}, \eta(t) = u_2(t)/u_1(t)$, in (3.1) and (3.3), and formula (3.4) becomes

$$\psi(x,t) = |u_1(t)|^{-1/2} \exp[iu_1(t)x^2/2hu_1(t)] \times \phi(\sigma|u_1(t)|^{-1}x, u_2(t)/u_1(t)). \quad (3.5)$$

This is the quantum change of variables corresponding to the one used in Ref. 9 in the classical case. See also formulas (5) in Ref. 10 and (3.7) in Ref. 7.

Remark: In Ref. 70 Husimi describes a canonical transformation which reduces a quantum system with a time-dependent linear part to a simpler one (Taniuti's transformation). This transformation together with the one just stated can be used to generalize some results of Sec. IV. For example, formula (3.66) in Ref. 64 and partial results of Refs. 30 and 71 can be obtained in this way.

IV. APPLICATIONS: THE CALCULATION OF GREEN'S FUNCTIONS AND THE MORSE INDEX

A. Harmonic oscillators

Let us begin by introducing the following notations:

$$H_0 = -(h^2/2)\Delta, \quad H_1 = (-h^2/2)\Delta + (\lambda^2/2)x^2,$$
$$H_2 = -(h^2/2)\Delta + (p(t)/2)x^2$$

are the Hamiltonians on $L^2(\mathbb{R})$ in the free, TIHO, and TDHO cases, respectively (here we let $\lambda \neq 0$); and $U_0(t)$, $U_1(t)$, and $U_2(t)$ are the corresponding time evolution operators on $L^2(\mathbb{R})$. This means that one has $\psi(x,t) = U_j(t)\psi(x,0)$, i.e., $ih(dU_j(t)/dt) = H_j U_j(t)$ and $U_j(0) = \mathbb{1}$ for $j = 0,1,2$.

Moreover for $j = 0$, 1, $U_j(t) = \exp[-(i/h)tH_j]$ and the Green's function in the free case is given by

$$K_0(x,t;y,0) = (2\pi iht)^{-1/2} \exp[(i/2ht)(x-y)^2], \quad \text{for } t \neq 0 \quad (4.1)$$

$$K_0(x,0;y,0) = \delta(y-x), \quad (4.2)$$

where δ is the Dirac measure. In order to calculate the Green's function for the TIHO K_1 from K_0 we put $\lambda = 0$, $V = 0$ in (3.2), $p = \lambda^2$ in (3.3), and $u_1 = \cos\lambda t$, $u_2 = \lambda^{-1}\sin\lambda t$ in (3.5). Let $A_m =]-\pi/2 + m\pi, \pi/2 + m\pi[$, $m \in \mathbb{Z}$; here every A_m plays the role of $]m_2, M_2[$ of Sec. II and III. For such an A_m we choose $\sigma = (-1)^m$. By Theorem 1 if $\phi(x,t)$, $(x,t) \in \mathbb{R}^2$, is a solution of the Schrödinger equation in the free case we have that

$$\psi(x,t) = \epsilon_m |\cos\lambda t|^{-1/2} \exp[-i\lambda \tan\lambda t x^2/2h] \times \phi(x/\cos\lambda t, \tan\lambda t/\lambda) \quad (4.3)$$

is a solution of the Schrödinger equation in the TIHO case for $(x,t) \in \mathbb{R} \times A_m$, where $\epsilon_m \in \mathbb{C}$. We let $\epsilon_0 = 1$ which implies $\psi(x,0) = \phi(x,0)$. From (4.3) we have

$$\psi(x, m\pi/\lambda) = \epsilon_m \phi((-1)^m x, 0). \quad (4.4)$$

Now we can choose the ϵ_m such that $\psi(x,t)$ is continuous at the points $\pi/2 + m\pi$, $m \in \mathbb{Z}$; (4.1) and (4.3) imply that

$$\psi(x,t) = \epsilon_m |\cos\lambda t|^{-1/2} (\lambda/2\pi ih \tan\lambda t)^{1/2} \times \int_{\mathbb{R}} \exp[\cdots]\psi(y,0)dy, \quad (4.5)$$

where

$$[\cdots] = (i\lambda/2h \sin\lambda t)[(x^2+y^2)\cos\lambda t - 2xy]. \quad (4.6)$$

A simple calculation shows that $\epsilon_m = \exp[-i\pi m/2]$ and finally we have

$$K_1(x,t;y,0) = (\lambda/2\pi ih)^{1/2} \sin^{-1/2}\lambda t \exp[\cdots], \quad t \neq m\pi/\lambda, \quad (4.7)$$

$$K_1(x, m\pi/\lambda; y, 0) = \exp[-im\pi/2]\delta(y - (-1)^m x), \quad (4.8)$$

where $\sin^{-1/2}$ is defined as in Sec. II, $[\cdots]$ is defined by (4.6), and δ is the Dirac measure; (4.7) is the well-known Feynman's formula with the so-called Maslov correction.[13,17,20,22,64] We have deduced once more this known formula because this new method can be used in different situations as we shall see in Sec. IV B.

We can obtain now the Green's function for the TDHO K_2 from K_1. As in the example of the Sec. II we put in (2.2) $P = 1, Q = 0, R = \lambda^2$ and in (2.4) $\theta = 0$. From Theorem 1 we get that

$$U_2(t) = U(t)U_1(\eta(t)), \quad (4.9)$$

where $U(t)$ is the unitary operator on $L^2(\mathbb{R})$ defined by

$$U(t): \psi(x) \to \xi(t)^{-1/4} \exp[i\dot{\xi}(t)x^2/4h\xi(t)] \times \psi(\xi(t)^{-1/2}x). \quad (4.10)$$

Using (2.10), (2.11), (4.7), (4.8), and (4.10) we can now prove the following theorem.

Theorem 2: The Green's function $K_2(x,t;y,0)$ for the Schrödinger equation of the TDHO is given by

$$K_2(x,t;y,0) = (2\pi ih)^{-1/2} s^{-1/2}(t) \times \exp[(i/2hs(t))(\dot{s}(t)x^2 + c(t)y^2 - 2xy)], \quad t \neq t_m, \quad (4.11)$$

$$K_2(x,t_m;y,0) = c^{-1/2}(t_m)\exp[i\dot{c}(t_m)x^2/2hc(t_m)], \times \delta(y - x/c(t_m)) \quad (4.12)$$

where $s^{-1/2}$, $c^{-1/2}$ are defined as in Sec. II, δ is the Dirac measure, and t_m is the zero of s such that $m(s, t_m) = m$.

Remark: For related results see Refs. 10, 13, 15, 17, 20, 29, 30, 61, 63, 64, and 70–78.

B. Harmonic oscillators with an inverse quadratic potential

Putting $V(x,t) = g/x^2$, $g \in \mathbb{R}$, in (3.2) we have $\xi(t)^{-1}V(\sigma\xi(t)^{-1/2}x, \eta(t)) = g/x^2$ in (3.3). So, starting from g/x^2 and doing the change of variables of Sec. III, we arrive at the same time-independent singular perturbation. As we can see the results of Sec. III are well adapted to the study of the quantum systems associated with the Hamiltonians $H_j^p = H_j + g/x^2$, $x > 0$, $j = 0,1,2$. For more details about such quantum systems see Refs. 29 and 79–84.

In the following we denote by K_j^p, $j = 0,1,2$, the corresponding Green's functions and we are going to deduce K_1^p and K_2^p from K_0^p.

According to Ref. 29 we have

$$K_0^p(x,t;y,0) = ((xy)^{1/2}/iht)I_a(xy/iht) \times \exp[(i/2ht)(x^2+y^2)], \quad (4.13)$$

where $x,y>0$, $t\neq 0$, $g>-\hbar^2/8$, and $a=(1+8g/\hbar^2)^{1/2}/2$. Here I_a is the modified Bessel function (see Ref. 85),

$$I_a(z)=\sum_{m=0}^{\infty}\frac{(z/2)^{2m+a}}{m!(m+a)!},\quad -\pi<\operatorname{Arg} z<\pi. \tag{4.14}$$

In order to calculate K_1^p and K_2^p we use the method of Sec. IV A with the exception that we let $\sigma=1$ in every A_m. This prescription is motivated by the fact that $x>0$, the mathematical structure of (4.13), and is in agreement with Ref. 79. For example, the formula corresponding to (4.5) is, in this case,

$$\psi(x,t)=(-1)^m\epsilon_m\frac{\lambda}{i\hbar\sin\lambda t}\int_0^{+\infty}(xy)^{1/2}\times I_a\left(\frac{xy\lambda\cot\lambda t}{i\hbar|\cos\lambda t|}\right)\exp[\cdots]\psi(y,0)dy, \tag{4.15}$$

where

$$[\cdots]=(i\lambda/2\hbar)\cot\lambda t(x^2+y^2). \tag{4.16}$$

Using the fact that $I_a(-i\zeta)=\exp[-i\pi a]\,I_a(i\zeta)$, for $\zeta>0$, and the same argument of continuity at the points $\pi/2+m\pi$ as in Sec. IV A, we conclude that $\epsilon_m=\exp[-i\pi m(a+1)]$. Moreover a careful calculation gives the following theorem (cf. Ref. 29).

Theorem 3: The Green's function $K_2^p(x,t;y,0)$ for the Schrödinger equation of the TDHO with perturbation of the type g/x^2 is given by

$$K_2^p(x,t;y,0)=\frac{(xy)^{1/2}}{i\hbar s(t)}I_a\left(\frac{xy}{i\hbar s(t)}\right)\exp\left[-2i\pi a\,\mathrm{int}\left(\frac{n}{2}+\frac{1}{2}\right)+\frac{i(\dot{s}(t)x^2+c(t)y^2)}{2\hbar s(t)}\right],\quad t\neq t_m, \tag{4.17}$$

$$K_2^p(x,t_m;y,0)=|c(t_m)|^{-1/2}\exp\left[-i\pi m(a+1)+\frac{i\dot{c}(t_m)x^2}{2\hbar c(t_m)}\right]\times\delta\left(y-\frac{x}{|c(t_m)|}\right), \tag{4.18}$$

where δ is the Dirac measure, $n=m(s,t)$, and t_m is the zero of s such that $m(s,t_m)=m$.

Remark: Formula (4.17) differs from the one of Ref. 29 in as much as it takes into account the Maslov correction.

V. THE SEMICLASSICAL APPROXIMATION

We consider the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi(x,t)=-\frac{\hbar^2}{2}\Delta\psi(x,t)+V(x)\psi(x,t), \tag{5.1}$$

where Δ is the Laplacian on $\mathbb{R}^n$ and $\hbar$ is Planck's constant divided by 2π; the wave function $\psi(x,t)$ describes the state at time t of a quantum mechanical particle with mass one, in $\mathbb{R}^n$, under the influence of a potential $V(x)$; for each fixed t, $\psi(x,t)$ belongs to $L^2(\mathbb{R}^n)$.

Following Ref. 19, we set up the Cauchy data at time $t=0$,

$$\psi(x,0)=\psi_0(x)\exp[(i/\hbar)S_0(x)]. \tag{5.2}$$

Functions $\psi_0(x)$, $S_0(x)$, and $V(x)$ are real valued and

$$\psi_0(x)\in C_0^\infty(\mathbb{R}^n),\quad S_0(x)\in C^\infty(\mathbb{R}^n), \tag{5.3}$$

$$V(x)\in\mathscr{P}(\mathbb{R}^n),\quad \inf\{V(x):x\in\mathbb{R}^n\}>-\infty. \tag{5.4}$$

For more details see Ref. 19, Sec. 12. For $y\in\mathbb{R}^n$, consider the path $\gamma(y,t)$ which satisfies the differential equation

$$\ddot{\gamma}(y,t)+V'(\gamma(y,t))=0, \tag{5.5}$$

and the Cauchy initial conditons

$$\gamma(y,0)=y\quad\text{and}\quad\dot{\gamma}(y,0)=S_0'(y). \tag{5.6}$$

In this section the prime denotes the derivative in $\mathbb{R}^n$, e.g., $V'=DV$, and, as before, the dot denotes differentiation with respect to t.

We define now $u(y,t)=\gamma'(y,t)$ and $J(y,t)=\det u(y,t)$. For $y\in\mathbb{R}^n$, the matrix valued function $u(y,t)$ is the solution of the Jacobi equation

$$\ddot{u}(y,t)+V''(\gamma(y,t))u(y,t)=0, \tag{5.7}$$

such that $u(y,0)=\mathbb{1}$ and $\dot{u}(y,0)=S_0''(y)$.

In Ref. 19 asymptotic formulas are given for the solutions of the Schrödinger equation at nonfocal points. For $y\in\mathbb{R}^n$, a point t is called a *focus* of the trajectory $\gamma(y,\tau)$ if $J(y,t)=0$. The *multiplicity* of the focal point t is the co-rank of the matrix $u(y,t)$.

For $y\in\mathbb{R}^n$ and $t>0$, the *Morse index* m of the trajectory $\gamma(y,\tau)$, with $\tau\in[0,t]$, is the number of focal points on $[0,t]$ counted with their multiplicity.

We fix $x\in\mathbb{R}^n$ and $t>0$, and we suppose that $y_j\in\mathbb{R}^n$ are all points for which $\gamma(y_j,t)=x$; (x,t) is called focal if t is focal for at least one of the trajectories $\gamma(y_j,\tau)$. As we assume conditions (5.3) and (5.4) one can prove (see Ref. 19) that the number of points y_j is finite, provided that (x,t) is not a focal point.

We have now the following theorem (for the proof see Ref. 19, Sec. 12).

Theorem 4: We assume that (x,t) is not a focal point. Then

$$\psi(x,t)=\sum_j\left(\psi_0(y_j)|J(y_j,t)|^{-1/2}\exp\left[-i\frac{\pi}{2}m_j\right]+O(\hbar)\right)\exp\left[\frac{i}{\hbar}S_j(x,t)\right], \tag{5.8}$$

as $\hbar\to 0$, where $S_j(x,t)$ is the action along the classical trajectory joining the points y_j and x, i.e.,

$$S_j(x,t)=S_0(y_j)+\int_0^t\left[\frac{1}{2}\dot{\gamma}(y_j,\tau)^2-V(\gamma(y_j,\tau))\right]d\tau, \tag{5.9}$$

and m_j is the Morse index of the trajectory.

Remark: Expression (5.8) is called the semiclassical asymptotic of the wave function.[14,18,19,68,86] The same kind of asymptotic expansions arise in the WKB approximation.[14,21,64]

In order to simplify notations we let $n=1$ from now on. If we replace $V(x)$ in Eq. (5.1) by the time-dependent harmonic potential $\frac{1}{2}p(t)x^2$, where $p(t)$ is a real continuous function, we obtain the Schrödinger equation corresponding to the quantum mechanical system describing the time-dependent harmonic oscillator.

In this case Eq. (5.5) becomes

$$\ddot{\gamma}(y,t)+p(t)\gamma(y,t)=0, \tag{5.10}$$

with $\gamma(y,t)$ satisfying conditions (5.6).

If $p(t)=V''(\gamma(y_j,t))$, where $\gamma(y_j,t)$ is one of the above classical paths related with the potential $V(x)$, then we have that $\gamma(y_j,t)$ satisfies also (5.10). We shall see that the contribution of this trajectory to the asymptotic expansion of the wave function of the TDHO is given by

$$(\psi_0(y_j)|J(y_j,t)|^{-1/2} \times\exp[-i(\pi/2)m_j]+O(h))\exp[i/h)\bar{S}_j(x,t)], \tag{5.11}$$

where

$$\bar{S}_j(x,t)=S_0(y_j)+\int_0^t\frac{1}{2}[\dot{\gamma}(y_j,\tau)^2-p(\tau)\gamma(y_j,\tau)^2]d\tau. \tag{5.12}$$

As we see, the difference between (5.11) and each summand in (5.8) consists only of the part that contains the classical action of the two different potentials. So, the problem of defining the Morse index arises in all its complexity already in the case of the TDHO. This is the reason why we are now going to investigate the asymptotic behavior, at time t and as $h\to 0$, of the wave function ψ for the Cauchy problem (5.2) in the TDHO case. We begin by recalling the following lemma.[65,66]

Lemma 2: Let $f\in C^\infty(\mathbb{R})$ and $g\in C_0^\infty(\mathbb{R})$ be real-valued functions such that $f'(x_j)=0$ for a finite number of x_j in the support of g. Suppose that $f''(x_j)\neq 0$ for every x_j. We have then

$$h^{-1/2}\int_{\mathbb{R}}\exp\left[\frac{i}{h}f(x)\right]g(x)dx = \sum_j\exp\left[\frac{i}{h}f(x_j)\right]\left\{\left(\frac{2\pi i}{f''(x_j)}\right)^{1/2}g(x_j)+O(h)\right\},$$

as $h\to 0$.

We use now this lemma to prove the following and final theorem.

Theorem 4': We assume that conditions (5.3) hold for $\psi_0(x)$ and $S_0(x)$. Let the point (x,t) be such that (i) the equation $x=c(t)y+S_0'(y)s(t)$ has only a finite number of solutions y_j in the support of ψ_0; and (ii) for every y_j let $c(t)+S_0''(y_j)s(t)\neq 0$. Then

$$\psi(x,t)=\sum_j(\psi_0(y_j)J^{-1/2}(y_j,t)+O(h)) \times\exp\left[\frac{i}{h}\bar{S}_j(x,t)\right],\quad \text{as } h\to 0,$$

where $J(y,\tau)$ is the solution of the Jacobi equation such that $J(y,0)=1$ and $\dot{J}(y,0)=S_0''(y)$; $J^{-1/2}$ is defined by (2.9) and $\bar{S}_j(x,t)$ is as in (5.12).

Proof: In order to obtain the asymptotic expansion of the wave function we use Theorem 2. The case $t=t_m$ is obvious. For $t\neq t_m$ we use Lemma 2. Then we have

$$g(y)=(2\pi i)^{-1/2}s^{-1/2}(t)\psi_0(y),$$
$$f(y)=(2s(t))^{-1}(\dot{s}(t)x^2+c(t)y^2-2xy)+S_0(y),$$
$$\gamma(y,\tau)=c(\tau)y+S_0'(y)s(\tau),$$
$$u(y,\tau)=J(y,\tau)=c(\tau)+S_0''(y)s(\tau),$$
$$f'(y)=s(t)^{-1}(\gamma(y,t)-x),$$
$$f''(y)=s(t)^{-1}u(y,t).$$

From assumptions (i) and (ii) we get that in support of ψ_0 there exist a finite number of stationary points y_j [i.e., $f'(y_j)=0$] which are nondegenerate [i.e., $f''(y_j)\neq 0$]. From the above formulas we conclude that $\gamma(y_j,\tau)$ is the classical path joining y_j and x and with initial momentum $S_0'(y_j)$. We can easily see that $f(y_j)=\bar{S}_j(x,t)$, i.e., $f(y_j)$ is the action along the calssical path $\gamma(y_j,\tau)$. Finally, a trivial but cautious computation using (2.10)–(2.13) shows that $(f''(y_j)/2\pi i)^{1/2}(2\pi i)^{1/2}s^{1/2}(t)=J^{1/2}(y_j,t)$, where $J^{1/2}(y_j,\tau)$ is defined by (2.9).

Remark: This result can be generalized to the case of the n-dimensional TDHO in agreement with Maslov's results.[19]

ACKNOWLEDGMENTS

I thank the referee for valid suggestions he made in order to improve the presentation of this paper.

I am indebted to Professor Ph. Blanchard for his constant encouragement and scientific support as well as to Professor S. Albeverio for many suggestions during the preparation of this paper. I have benefitted from conversations with Professor R. Lima to whom I am sincerely grateful.

Financial support by the DAAD (Deutscher Akademischer Austauschdienst) is gratefully acknowledged.

[1] H. U. Kranold, J. Math. Phys. **24**, 2345 (1983).
[2] P. G. L. Leach, SIAM J. Appl. Math. **34**, 496 (1978).
[3] P. G. L. Leach, J. Math. Phys. **18**, 1902 (1977).
[4] P. G. L. Leach, J. Math. Phys. **19**, 446 (1978).
[5] P. G. L. Leach, J. Math. Phys. **20**, 96 (1979).
[6] P. G. L. Leach, J. Phys. A **13**, 1991 (1980).
[7] P. G. L. Leach, J. Phys. A **16**, 3261 (1983).
[8] H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165 (1982).
[9] A. M. Perelomov, Commun. Math. Phys. **63**, 9 (1978).
[10] V. S. Popov and A. M. Perelomov, Sov. Phys. JETP **30**, 910 (1970).
[11] V. I. Arnold, Funct. Anal. Appl. **1**, 1 (1967).
[12] V. S. Buslaev, Funct. Anal. Appl. **3**, 181 (1969).
[13] Ph. Combe, G. Rideau, R. Rodriguez, and M. Sirugue-Collin, Rep. Math. Phys. **13**, 279 (1978).
[14] C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rev. **50**, 255 (1979).
[15] J. P. Eckmann and R. Sénéor, Arch. Rat. Mech. Anal. **61**, 153 (1976).
[16] P. Exner, J. Math. Phys. **24**, 1129 (1983).
[17] P. A. Horváthy, Int. J. Theor. Phys. **18**, 245 (1979).
[18] V. P. Maslov, *Théorie des Perturbations et Méthodes Asymptotiques* (Dunod, Paris, 1972).
[19] V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Dordrecht, 1981).
[20] J. Rezende, Lett. Math. Phys. **7**, 75 (1983).
[21] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley–Interscience, New York, 1981).
[22] J. M. Souriau, *Lecture Notes in Physics* (Springer, Berlin, 1976), Vol. 50, pp. 117–148.
[23] A. Voros, Ann. Inst. H. Poincaré A **24**, 31 (1976).
[24] N. J. Günther and P. G. L. Leach, J. Math. Phys. **18**, 527 (1977).
[25] J. G. Hartley and J. R. Ray, Phys. Rev. A **25**, 2388 (1982).
[26] J. G. Hartley and J. R. Ray, Phys. Rev. A **24**, 2873 (1981).
[27] J. G. Hartley and J. R. Ray, Phys. Rev. D **25**, 382 (1982).
[28] G. H. Katzin and J. Levine, J. Math. Phys. **18**, 1267 (1977).
[29] D. C. Khandekar and S. V. Lawande, J. Math. Phys. **16**, 384 (1975).
[30] D. C. Khandekar and S. V. Lawande, J. Math. Phys. **20**, 1870 (1979).
[31] P. G. L. Leach, J. Math. Phys. **18**, 1608 (1977).
[32] P. G. L. Leach, J. Math. Phys. **21**, 300 (1980).

[33]P. G. L. Leach, J. Math. Phys. **22**, 456 (1981).
[34]H. R. Lewis, Jr., Phys. Rev. Lett. **18**, 510 636 (1967).
[35]H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968).
[36]H. R. Lewis, Jr. and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[37]H. R. Lewis, Jr., Phys. Rev. **172**, 1313 (1968).
[38]M. Lutzky, J. Phys. A **11**, 249 (1978).
[39]M. Lutzky, Phys. Lett. A **68**, (1978).
[40]M. Lutzky, Phys. Lett. A **78**, 301 (1980).
[41]M. Lutzky, J. Math. Phys. **21**, 1370 (1980).
[42]G. E. Prince and C. J. Eliezer, J. Phys. A **13**, 815 (1980).
[43]J. R. Ray, Phys. Lett. A **78**, 4 (1980).
[44]J. R. Ray, J. Phys. A **13**, 1969 (1980).
[45]J. R. Ray, Lett. Nuovo Cimento **27**, 424 (1980).
[46]J. R. Ray, Phys. Rev. A **26**, 729 (1982).
[47]J. R. Ray and J. G. Hartley, Phys. Lett. A **88**, 125 (1982).
[48]J. R. Ray and J. L. Reid, Phys. Lett. A **71**, 317 (1979).
[49]J. R. Ray and J. L. Reid, Phys. Lett. A **74**, 23 (1979).
[50]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979).
[51]J. R. Ray and J. L. Reid, J. Math. Phys. **21**, 1583 (1980).
[52]J. R. Ray and J. L. Reid, J. Math. Phys. **22**, 91 (1981).
[53]J. R. Ray and J. L. Reid, Lett. Math. Phys. **4**, 235 (1980).
[54]J. R. Ray and J. L. Reid, Nuovo Cimento A **59**, 134 (1980).
[55]J. R. Ray, J. L. Reid, and M. Lutzky, Phys. Lett. A **84**, 42 (1981).
[56]W. Sarlet, Phys. Lett. A **82**, 161 (1981).
[57]W. Sarlet and J. R. Ray, J. Math. Phys. **22**, 2504 (1981).
[58]L. S. Wollenberg, Phys. Lett. A **79**, 269 (1980).
[59]J. M. Hill, *Solution of Differential Equations by Means of One-Parameter Groups. Research Notes in Mathematics* (Pitman, Boston, 1982), Vol. 63.
[60]E. Hille, *Lecture on Ordinary Differential Equations* (Addison–Wesley, Reading, MA, 1969).
[61]P. Pechukas and J. Light, J. Chem. Phys. **44**, 3897 (1966).
[62]W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[63]P. Pechukas, Phys. Rev. **181**, 166 (1969).
[64]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965).
[65]L. Hörmander, Acta Math. **127**, 79 (1971).
[66]L. Hörmander, Math. Z. **146**, 69 (1976).
[67]S. Albeverio, Ph. Blanchard, and R. Høegh-Krohn, Commun. Math. Phys. **83**, 49 (1982).
[68]S. Albeverio and R. Høegh-Krohn, Invent. Math. **40**, 59 (1977).
[69]J. Rezende (to appear).
[70]K. Husimi, Progr. Theor. Phys. **9**, 381 (1953).
[71]M. J. Goovaerts, Physica **77**, 379 (1974).
[72]R. H. Cameron and W. T. Martin, Bull. Am. Math. Soc. **51**, 73 (1945).
[73]X. Chapuisat and J. Manz, Mol. Phys. **26**, 1577 (1973).
[74]J. E. Littlewood, Ann. Phys. **21**, 233 (1963).
[75]J. E. Littlewood, Ann. Phys. **29**, 13 (1964).
[76]K. Nishiwada, Proc. Jpn. Acad. A **56**, 362 (1980).
[77]V. S. Popov and A. M. Perelomov, Sov. Phys. JETP **29**, 738 (1969).
[78]S. Solimeno, P. Di Porto and B. Crosignani, J. Math. Phys. **10**, 1922 (1969).
[79]F. Calogero, J. Math. Phys. **10**, 2191 (1969).
[80]F .Calogero, J. Math. Phys. **12**, 419 (1971).
[81]P. Camiz, A. Gerardi, C. Marchioro, E. Presutti, and E. Scacciattelli, J. Math. Phys. **12**, 2040 (1971).
[82]M. J. Goovaerts, J. Math. Phys. **16**, 720 (1975).
[83]D. C. Khandekar and S. V. Lawande, J. Phys. A **5**, 812 (1972).
[84]D. C. Khandekar and S. V. Lawande, J. Phys. A **5**, L57 (1972).
[85]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1966).
[86]A. Truman, *Lecture Notes in Physics* (Springer, Berlin, 1979), Vol. 106, pp. 73–102.

Nonexistence of asymptotically free solutions for a nonlinear Schrödinger equation

Jacqueline E. Barab[a)]
Department of Mathematics, Kansas State University, Manhattan, Kansas 66506

(Received 21 November 1983; accepted for publication 13 April 1984)

Let u be a nontrivial, smooth solution to $iu_t = \Delta u - |u|^{p-1}u$. If $n = 1$ and $2 < p \leqslant 3$, then there does not exist any finite energy free solution v such that $\|u(t) - v(t)\|_2 \to 0$ as $t \to +\infty$. This extends a theorem of Strauss in which the same result was proved for $1 < p \leqslant 2$.

PACS numbers: 03.65.Nk, 42.65. – k

I. INTRODUCTION

Consider the perturbed (nonlinear) Schrödinger equation

$$iu_t = \Delta u - g|u|^{p-1}u \tag{1}$$

and the corresponding free (linear) equation

$$iv_t = \Delta v. \tag{2}$$

Here $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, $p > 1$, and the constant g is positive unless stated otherwise. For each $q \in [1, \infty)$, the norm $\|\cdot\|_q$ is the usual spatial $L^q(\mathbb{R}^n)$-norm and the dual variable is denoted by q' (so $1/q + 1/q' = 1$). The suppressed notation

$$L^q = L^q(\mathbb{R}^n), \quad \int = \int_{\mathbb{R}^n} d^n x, \quad w = w(t) = w(x,t)$$

will be used.

Definition: A solution u to (1) is *asymptotically free* if there exist L^2-solutions $v_\pm$ to (2), decaying sufficiently rapidly, such that $\|u(t) - v_\pm(t)\|_2 \to 0$ as $t \to \pm\infty$. (The necessary decay is made precise in Remark 2 following Lemma 2.)

It has been shown (Strauss[1]) that if p is large enough, then a substantial class of solutions to (1) are asymptotically free. Strauss then proved the following partial converse.

Theorem 1 (Strauss[1]): If either

(i) $n \geqslant 2$ and $1 < p \leqslant 1 + 2/n$, or

(ii) $n = 1$ and $1 < p \leqslant 2$,

then the only asymptotically free solution to (1) is identically zero.

The main result in this paper, Theorem 2, is that the condition for nonexistence of asymptotically free solutions to (1) includes the case $n \geqslant 1$ and $1 < p \leqslant 1 + 2/n$; i.e., the one-dimensional case is not really exceptional. The general idea in the proofs of both this result and Theorem 1 was originally used by Glassey[2] to prove the analogous theorem for the nonlinear Klein–Gordon equation. The proof is by contradiction. A bilinear form H is defined so that $t \mapsto |H[u(t), v(t)]|$ is uniformly bounded for t sufficiently large. All the lemmas are preliminaries to the establishment of the key estimate which is essentially $dH/dt \geqslant c/t > 0$; integration leads immediately to the contradiction.

The new ingredient which makes this extension to Strauss' theorem possible is an estimate (Lemma 3) derived from the pseudoconformal identity of Ginibre and Velo.[3] Kadekawa[4] first used this estimate to obtain positive scattering results in higher dimensions.

[a)] This work constitutes part of the author's Ph.D. thesis in the Department of Mathematics at Indiana University, August 1982.

II. PRELIMINARIES

The lemmas consist of identities and estimates satisfied by solutions to (1) and/or (2). Since it is expected that such statements will continue to be basic tools in future work, they are proved for $n \geqslant 1$ even though Theorem 2 is a one-dimensional result. For simplicity, solutions to (1) and (2) are assumed to be smooth; however, much less stringent assumptions would suffice (see, e.g., Ginibre and Velo[3] for details). Therefore, the hypotheses stated here are intentionally redundant to emphasize those features of the smoothness assumptions that are crucial.

The following well-known conservation laws obtain for both free and perturbed solutions.

Lemma 1 (Conservation Laws): If w is a smooth solution to (1) with $g \in \mathbb{R}$ and $w(0) \in H^1 \cap L^{p+1}$, then for all t

$$\frac{d}{dt}(\|w(t)\|_2) = 0, \tag{3}$$

$$\frac{d}{dt}\left(\|\nabla w(t)\|_2^2 + \frac{2g}{p+1}\|w(t)\|_{p+1}^{p+1}\right) = 0. \tag{4}$$

Proof: Replace u by w in (1), multiply by $2\bar{w}$, and take the imaginary part to obtain

$$\frac{d}{dt}(|w(t)|^2) = \nabla \cdot \mathrm{Im}(2\bar{w}\nabla w). \tag{5}$$

Integration over all space implies (3). To establish (4), multiply (1) by $2\bar{w}_t$, integrate over $\mathbb{R}^n$, and take the real part of the result. ∎

Part (i) of Lemma 2 is well known. The statement and idea for the proof of part (ii) appeared in Strauss.[1]

Lemma 2 (Estimates for Free Solutions): If v is a smooth solution to (2) with $0 \neq \phi = v(0) \in L^1 \cap L^2$ and $2 \leqslant q \leqslant \infty$, then (i) there exists a constant $c = c(\|\phi\|_{q'})$ such that

$$\|v(t)\|_q \leqslant ct^{-n(q-2)/2q}, \quad \forall t > 0, \tag{6}$$

and (ii) there exist positive constants $B = B(n,q,\phi)$ and $T_0 = T_0(\phi)$ such that

$$\|v(t)\|_q \geqslant Bt^{-n(q-2)/2q}, \quad \forall t \geqslant T_0. \tag{7}$$

When $q = \infty$, the power of t is $-n/2$.

Proof: The classical representation of v,

$$v(x,t) = (4\pi it)^{-n/2}\int_{\mathbb{R}^n} e^{-i|x-y|^2/4t} v(y,0)\, d^n y \tag{8}$$

(obtained via Fourier transform), implies

$$\|v(t)\|_\infty \leqslant t^{-n/2}\|\phi\|_1.$$

This estimate and an application of (3) to v together satisfy

 0022-2488/84/113270-04$02.50

the hypothesis of the Riesz–Thorin theorem (Reed and Simon,[5] p. 27) whose conclusion is (6) with $c = \|\phi\|_{q'}$.

If $q = 2$, then (3) implies (7) with $B = \|\phi\|_2$. If $q > 2$, then for any $k > 0$ (yet to be chosen), Hölder's inequality implies

$$\left(\int_{|x|<kt} |v(x,t)|^2\,dx\right)^{q/2}$$
$$\leqslant \left(\int_{|x|<kt} dx\right)^{(q-2)/2} \int_{|x|<kt} |v(x,t)|^q\,dx$$
$$= C(k,n,q)t^{n(q-2)/2}\|v(t)\|_q^q.$$

Therefore, to prove (7) it suffices to show that there exist $k > 0$, $T_0 < \infty$, and $C_1 > 0$ such that

$$C_1 \leqslant \int_{|x|<kt} |v(x,t)|^2\,dx \equiv I(t), \quad \forall t \geqslant T_0. \tag{9}$$

Replace v by representation (8), change variables ($\xi = x/2t$), and use $|e^{i\alpha}| = 1$ with $\alpha = -|x|^2/4t$ to see that

$$I(t) = \int_{|\xi|<k/2} \left|(2\pi)^{-n/2} \int_{\mathbf{R}^n} e^{i\xi\cdot y}[e^{-i|y|^2/4t}\phi(y)]\,dy\right|^2 d\xi.$$

Then

$$I(t) = \int_{|\xi|<k/2} |\mathscr{F}^{-1}f(\xi,t)|^2\,d\xi,$$

where $f(y,t) \equiv e^{-i|y|^2/4t}\phi(y)$ and $\mathscr{F}^{-1}$ denotes the inverse Fourier transform in the space variable. It follows from the dominated convergence theorem that $f(\cdot,t) \to \phi(\cdot)$ in $L^2(\mathbb{R}^n)$ as $t \to \infty$. Since $\mathscr{F}$ is an isomorphism on L^2, $\mathscr{F}^{-1}f \to \mathscr{F}^{-1}\phi$ in $L^2(\mathbb{R}^n)$ and hence also in $L^2(\{|\xi| \leqslant k/2\})$. Therefore, for each $k > 0$,

$$I(t) \to \int_{|\xi|<k/2} |\mathscr{F}^{-1}\phi|^2\,d\xi \equiv I(\infty) \quad \text{as } t \to \infty.$$

Now $0 \neq \|\phi\|_2^2 = \|\mathscr{F}^{-1}\phi\|_2^2$, so there exists k large enough for $I(\infty) \geqslant \frac{1}{2}\|\phi\|_2^2 > 0$. Fix such a k and define $C_1 \equiv \frac{1}{2}I(\infty)$. Since $I(t) \to 2C_1 > 0$ as $t \to \infty$, there exists $T_0 < \infty$ such that $I(t) \geqslant C_1$ for all $t \geqslant T_0$. This establishes (9) which implies (7) with $B = \sqrt{C_1}/[C(k,n,q)]^{1/q}$. ■

Remark 1: Note that k and hence also B and T_0 depend on the function ϕ (the support of $\mathscr{F}^{-1}\phi$) not only on $\|\phi\|_2$.

Remark 2: The decay in the definition of asymptotically free is that stated in (6). If $v \in L^2$, then the additional weak hypothesis that $v(x,0) \in L^1$ is sufficient to guarantee this decay.

If perturbed solutions enjoyed the same decay (6) as free solutions, in particular in the L^∞-norm, the proof of the main theorem would be straightforward. Though such a strong estimate is not expected, the following key lemma establishes *some* decay which will suffice.

Lemma 3 (Decay of Perturbed Solutions): If u is a smooth solution to (1) with $1 < p < 1 + 4/n$, $\phi(x) = u(x,0) \in H^1 \cap L^{p+1}$, and $\|x\phi(x)\|_2 < \infty$, then there exists $c > 0$ such that

$$\|u(t)\|_{p+1} \leqslant ct^{-n(p-1)/2(p+1)}, \quad \forall t > 0.$$

Proof: The first main step is to derive the pseudoconformal identity

$$\frac{d}{dt}\int \left(|xu - 2it\nabla u|^2 + \frac{8g}{p+1}t^2|u|^{p+1}\right)dx$$
$$= \frac{4g[4 - n(p-1)]}{p+1}\,t\int |u|^{p+1}\,dx. \tag{10}$$

Here, $x = (x_1,\ldots,x_n)$, $r = |x|$, and $u_k = \partial_k u = \partial u/\partial x_k$. Multiply (1) by $2r\bar{u}_r$ and integrate the real part over $\mathbf{R}^n$. The result can be expressed as

$$\mathrm{I} = \mathrm{II} + \mathrm{III},$$

where

$$\mathrm{I} = 2\,\mathrm{Re}\,i\int \left(\sum_k x_k \bar{u}_k u_t\right)dx,$$

$$\mathrm{II} = 2\,\mathrm{Re}\int r\bar{u}_r \Delta u\,dx,$$

$$\mathrm{III} = -\frac{2g}{p+1}\int r\partial_r(|u|^{p+1})dx.$$

Integrate by parts to find that

$$\mathrm{II} = (n-2)\int |\nabla u|^2\,dx,$$

$$\mathrm{III} = \frac{2gn}{p+1}\int |u|^{p+1}\,dx.$$

Term I can be rewritten as

$$\mathrm{I} = \mathrm{Re}\left[i\int \sum_k x_k(\bar{u}_k u_t - u_k \bar{u}_t)dx\right]$$
$$= \frac{d}{dt}\mathrm{Re}\left[i\int \sum_k x_k(\partial_t(\bar{u}_k u) - \partial_k(u\bar{u}_t))dx\right]$$
$$= \frac{d}{dt}\mathrm{Re}\left[i\int r\bar{u}_r u\,dx\right] + \mathrm{Re}\left[in\int u\bar{u}_t\,dx\right].$$

Substitute for $i\bar{u}_t$ from (1) to get

$$\mathrm{I} = \frac{d}{dt}\mathrm{Im}\left[\int ru_r\bar{u}\,dx\right] + n\int (|\nabla u|^2 + g|u|^{p+1})dx.$$

Hence, the equation I = II + III becomes

$$\frac{d}{dt}\left(\mathrm{Im}\int ru_r\bar{u}\right) = -2\int |\nabla u|^2 - \frac{ng(p-1)}{p+1}\int |u|^{p+1}. \tag{11}$$

In light of the lhs, multiply (5) by $|x|^2$ and integrate over $\mathbf{R}^n$ to obtain

$$\frac{d}{dt}\int |xu|^2 = -4\,\mathrm{Im}\int ru_r\bar{u}. \tag{12}$$

Now multiply (11) by $4t$ and rewrite as

$$\frac{d}{dt}\left(4t\,\mathrm{Im}\int ru_r\bar{u}\right) - 4\,\mathrm{Im}\int ru_r\bar{u}$$
$$= \frac{d}{dt}\left(-4t^2\int |\nabla u|^2\right) + 4t^2\frac{d}{dt}\int |\nabla u|^2$$
$$- \frac{4ng(p-1)}{p+1}\,t\int |u|^{p+1}.$$

Use (12) in the second term and (4) in the fourth term; then rewrite to obtain

$$\frac{d}{dt}\int (|x|^2|u|^2 + 4t^2|\nabla u|^2 - \mathrm{Re}\, 4tiru_r\bar{u})dx$$
$$= \frac{d}{dt}\left(\frac{-8g}{p+1}t^2\int |u|^{p+1}\right)$$
$$+ \frac{16g}{p+1}t\int |u|^{p+1} - \frac{4ng(p-1)}{p+1}t\int |u|^{p+1}.$$

Simplification yields (10).

The remainder of the proof is the following Gronwall argument. Integrate (10) over $[0,t]$ and use the positivity of the first term:

$$\frac{8gt^2}{p+1}\|u(t)\|_{p+1}^{p+1}$$
$$\leqslant \|xu(0)\|_2^2 + \frac{4g[4-n(p-1)]}{p+1}\int_0^t \tau\|u(\tau)\|_{p+1}^{p+1}d\tau.$$

Thus,

$$t^2\|u(t)\|_{p+1}^{p+1} \leqslant \alpha' + \frac{4-n(p-1)}{2}\int_1^t \tau\|u(\tau)\|_{p+1}^{p+1}d\tau, \qquad (13)$$

where

$$\alpha' \equiv \frac{p+1}{8g}\left(\|xu(0)\|_2^2 + \frac{4g[4-n(p-1)]}{p+1}\int_0^1 \tau\|u(\tau)\|_{p+1}^{p+1}d\tau\right).$$

Since integration of (4) over $[0,t]$ implies

$$\|\nabla u(t)\|_2^2 + \frac{2g}{p+1}\|u(t)\|_{p+1}^{p+1} = \|\nabla\phi\|_2^2 + \frac{2g}{p+1}\|\phi\|_{p+1}^{p+1},$$

it is easy to see that

$$\alpha' \leqslant \alpha \equiv c(g,p)(\|x\phi\|_2^2 + \|\nabla\phi\|_2^2 + \|\phi\|_{p+1}^{p+1}),$$

which is finite by hypothesis. Then (13) can be written as

$$F(t) \leqslant \alpha + \int_1^t \beta(\tau)F(\tau)d\tau,$$

where

$$F(t) \equiv t^2\|u(t)\|_{p+1}^{p+1}$$

and

$$\beta(t) \equiv [4-n(p-1)]/2t, \quad \text{for } t \geqslant 1.$$

Here, F, $\alpha \geqslant 0$ by definition and $\beta \geqslant 0$ by hypothesis. Since F and β are continuous on $[1,\infty)$, Gronwall's lemma implies that

$$F(t) \leqslant \alpha \exp\left(\int_1^t \beta(\tau)d\tau\right), \quad \forall t > 1;$$

i.e.,

$$t^2\|u(t)\|_{p+1}^{p+1} \leqslant \alpha \exp\left(\int_1^t \frac{[4-n(p-1)]}{2\tau}d\tau\right).$$

Simplify to get

$$\|u(t)\|_{p+1}^{p+1} \leqslant \alpha t^{-n(p-1)/2}, \quad \forall t > 1.$$

The hypothesis on $u(0)$ and integration of (4) together imply that $\|u(t)\|_{p+1}^{p+1}$ is bounded uniformly for all t, in particular for $0 \leqslant t \leqslant 1$. Hence, there exists a constant $c = c(g,p,\|x\phi\|_2,\|\nabla\phi\|_2,\|\phi\|_{p+1})$ such that

$$\|u(t)\|_{p+1} \leqslant ct^{-n(p-1)/2(p+1)}, \quad \forall t > 0. \qquad \blacksquare$$

III. THE MAIN RESULT

Theorem 2 (Nonexistence of Asymptotically Free Solutions): If $n \geqslant 1$ and $1 < p \leqslant 1 + 2/n$, then the only smooth, asymptotically free solution to (1) is identically zero.

Proof: In light of Strauss' result (Theorem 1), it suffices to consider the case $n = 1$ and $2 < p \leqslant 3$. Assume u is a smooth, asymptotically free solution to (1). Then there exists a smooth L^2-solution v of (2) such that

$$\left.\begin{aligned}\|u(t) - v(t)\|_2 &\to 0 \\ \|v(t)\|_\infty &= O(t^{-1/2})\end{aligned}\right\} \text{ as } t \to +\infty. \qquad (14),(15)$$

Since the conservation of the L^2-norm (3) and statement (14) together imply

$$\|u(t)\|_2 = \|v(t)\|_2 \equiv A, \quad \forall t, \qquad (16)$$

it suffices to show that $v(0) = 0$. The proof is by contradiction, so suppose $v(0) \neq 0$. Let B and T_0 be as in Lemma 2 (ii), which will be applied to v. Note that (15) implies

$$\|v(t)\|_\infty \leqslant ct^{-1/2}, \quad \forall t > T_1 \geqslant T_0. \qquad (17)$$

Now for $t > T_1$, define

$$H(t) \equiv \int_{-\infty}^{\infty} u(x,t)\bar{v}(x,t)dx.$$

Differentiate H with respect to t, substitute from the differential equations (1) and (2) for u_t and v_t, respectively, and integrate by parts to get

$$\dot{H}(t) \equiv \frac{dH}{dt} = ig\int |u|^{p-1}u\bar{v}\,dx.$$

Add and subtract $ig\int |v|^{p+1}dx$; then take the imaginary part

$$\mathrm{Im}\,\dot{H}(t) = g\int |v|^{p+1}dx + \mathrm{Re}\, g\int (|u|^{p-1}u\bar{v} - |v|^{p+1})dx.$$

Lemma 2 (ii), applied to v, implies

$$\mathrm{Im}\,\dot{H}(t) \geqslant gBt^{-(p-1)/2} - gI, \quad \forall t > T_1, \qquad (18)$$

where

$$I \equiv \left|\mathrm{Re}\int (|u|^{p-1}u\bar{v} - |v|^{p+1})dx\right|.$$

The following estimate shows that $I = o(t^{-(p-1)/2})$ as $t \to \infty$, so that $\mathrm{Im}\,\dot{H}(t) \geqslant ct^{-(p-1)/2} > 0$ for all large t. The contradiction arises by showing, after integration, that the lhs of (18) is bounded above while the rhs has an arbitrarily large lower bound.

Use the Minkowski inequality and the mean value theorem (since $0 < p - 2$) to estimate

$$I \leqslant \left|\int (|u|^{p-1} - |v|^{p-1})u\bar{v}\right| + \left|\int |v|^{p-1}(u-v)\bar{v}\right|$$
$$\leqslant c\int (|u|^{p-2} + |v|^{p-2})|u-v||u||v| + \int |v|^{p-1}|u-v||v|.$$

So

$$I \leqslant J_1 + J_2 + J_3,$$

where

$$J_1 \equiv c \int |u - v||u|^{p-1}|v|dx,$$

$$J_2 \equiv c \int |v|^{p-1}|u - v||u|dx,$$

$$J_3 \equiv \int |v|^{p-1}|u - v||v|dx.$$

First consider J_k for $k = 2,3$ each of which has the form $c\int |v|^{p-1}|u - v||w|$, where $w = u,v$ for $k = 2,3$, respectively. From Schwarz's inequality, the decay of the free solution (17), and the L^2-bound (16) it follows that

$$J_k \leqslant c\|v(t)\|_\infty^{p-1}\|u(t) - v(t)\|_2 w(t)\|_2$$
$$\leqslant c(t^{-1/2})^{p-1}\|u(t) - v(t)\|_2.$$

The asymptotic assumption (14) implies

$$J_k = o(t^{-(p-1)/2}) \quad \text{as } t \to \infty, \quad \text{for } k = 2,3. \tag{19}$$

To estimate J_1 use the generalized Hölder inequality to get

$$J_1 \leqslant c\|u(t) - v(t)\|_2 \|u(t)\|_{p+1}^{p-1}\|v(t)\|_{2(p+1)/(3-p)}.$$

(The positive numbers $\frac{1}{2}$, $(p-1)/(p+1)$, and $(3-p)/2(p+1)$ sum to 1.) Apply the decay estimates Lemmas 3 and 2 (ii) to u and v, respectively, to see that

$$J_1 \leqslant c\|u(t) - v(t)\|_2 t^{-(p-1)^2/2(p+1)} t^{-(q-2)/2q},$$

with $q = 2(p+1)/(3-p)$. (The assumption $2 < p \leqslant 3$ implies that $2 \leqslant q \leqslant \infty$.) Simplification yields

$$J_1 \leqslant c\|u(t) - v(t)\|_2 t^{-(p-1)/2},$$

so

$$J_1 = o(t^{-(p-1)/2}).$$

Recall (19) to conclude that

$$I \leqslant J_1 + J_2 + J_3 = o(t^{-(p-1)/2}) \quad \text{as } t \to \infty.$$

This estimate together with (18) implies that there exists $T > \max\{1, T_1\}$ and a positive constant C ($C < gB$) such that

$$\operatorname{Im} \dot{H}(t) \geqslant Ct^{-(p-1)/2}, \quad \forall t \geqslant T.$$

Keeping C and T fixed, let K be a positive integer (yet to be chosen) and integrate this inequality over $T \leqslant t \leqslant KT$ to get

$$\int_T^{KT} \frac{d}{dt}[\operatorname{Im} H(t)]dt \geqslant \int_T^{KT} Ct^{-(p-1)/2}\,dt \geqslant C\int_T^{KT} t^{-1}\,dt.$$

Note the use of the hypothesis $p \leqslant 3$, crucial for the validity of the second inequality. Therefore,

$$\operatorname{Im} H(KT) - \operatorname{Im} H(T) \geqslant C \ln K.$$

The bound (*uniform* for $K > 1$) for the lhs follows from the definition of H, Schwarz's inequality, and (16):

$$|\operatorname{Im} H(t)| \leqslant |H(t)| \equiv \left| \int_{-\infty}^{\infty} u(x,t)\bar{v}(x,t)dx \right|$$
$$\leqslant \|u(t)\|_2\|v(t)\|_2 \equiv A^2, \quad \forall t > T.$$

So

$$C \ln K \leqslant |\operatorname{Im} H(KT)| + |\operatorname{Im} H(T)| \leqslant 2A^2.$$

Choose $K > \exp(2A^2/C)$ to reach the desired contradiction. Hence, $v(0) = 0$ which implies via (16) that $u(t) = 0$ in L^2 for all t. The smoothness of u implies $u(x,t) \equiv 0$, and the theorem is proved. ■

Though Lemma 3 is sufficient for the proof of Theorem 2, the following corollary says that perturbed solutions decay in the L^q-norm at the same rate as free solutions provided q is not too large.

Corollary: Under the hypothesis of Lemma 3,

$$\|u(t)\|_q \leqslant ct^{-n(q-2)/2q}, \quad \text{for } 2 \leqslant q \leqslant p+1,$$

where c depends on the same parameters as in Lemma 3.

Proof: Interpolate to get

$$\|u(t)\|_q \leqslant \|u(t)\|_2^{1-\theta}\|u(t)\|_{p+1}^{\theta}, \quad \text{for } \theta \in (0,1),$$

satisfying $1/q = (1-\theta)/2 + \theta/(p+1)$. An elementary computation shows $\theta = (q-2)(p+1)/q(p-1)$. Use of the decay in the L^{p+1}-norm and the conservation of the L^2-norm [Lemmas 3 and 2 (i)] establishes the corollary. ■

IV. CONCLUSION

The nonexistence theorem precludes the development of a scattering theory in the case $1 < p \leqslant 1 + 2/n$ and $n \geqslant 1$. Current scattering theories demand that p be sufficiently large. It has just been proved (Tsutsumi and Yajima[6]) that for a large class of data and $1 + 2/n < p$ there exist asymptotically free solutions. (If $n \geqslant 3$, there is an additional upper bound restriction on p.) Hence, the nonexistence theorem is sharp.

[1] W. A. Strauss "Nonlinear Scattering Theory", in *Scattering Theory in Mathematical Physics*, edited by J. A. Lavita and J-P. Marchand (Reidel, Dordrecht, Holland, 1974), pp. 53–78.

[2] R. T. Glassey, Trans. AMS **182**, 187 (1973).

[3] J. Ginibre and G. Velo, "Existence of Solutions and Scattering for the Nonlinear Schrödinger Equation," in Proceedings of the *International Conference on Operator Algebras, Ideals, and Their Applications in Theoretical Physics*, edited by J. Baumgärtel *et al.* (Teubner, Leipzig, 1978), pp. 320–334.

[4] S. Kadekawa, Ph.D. Thesis, Indiana University, 1980.

[5] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, II: Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).

[6] Y. Tsutsumi and K. Yajima, "The Asymptotic Behavior of Nonlinear Schrödinger Equations," preprint, 1983.

Anisotropic fluids and conformal motions in general relativity

L. Herrera,[a] J. Jiménez, L. Leal, and J. Ponce de León[b]
Departmento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela

M. Esculpi
Departamento de Física Aplicada, Facultad de Ingeniería, Universidad Central de Venezuela, Caracas 1051, Venezuela

V. Galina
Departamento de Física, Universidad Simón Bolívar, Sartenejas, Venezuela

(Received 25 January 1984; accepted for publication 30 March 1984)

We study the consequences of the existence of a one-parameter group of conformal motions for anisotropic matter, in the context of general relativity. It is shown that for a class of conformal motions (special conformal motions), the equation of state is uniquely determined by the Einstein equations. For spherically symmetric and static distributions of matter we found two analytical solutions of the Einstein equations which correspond to isotropic and anisotropic matter, respectively. Both solutions can be matched to the Schwarzschild exterior metric and possesses positive energy density larger than the stresses, everywhere within the sphere.

PACS numbers: 04.20. – q

I. INTRODUCTION

In this paper we attempt to study space-times which admit a one-parameter group of conformal motions generated by a vector field ξ^α, such that

$$\underset{\xi}{L} g_{\alpha\beta} = \psi g_{\alpha\beta}, \tag{1}$$

where the left-hand side is the Lie derivative of the metric tensor with respect to the vector field ξ, and ψ is an arbitrary function of the coordinates. For $\psi = 2$ and perfect fluids we recover the self-similar solutions which have been extensively studied in the past.[1–5]

In the present work we generalize the discussion on self-similar space-times; specifically: (1) we consider conformal motions with an arbitrary choice of the function ψ in Eq. (1); (2) instead of perfect fluids we shall consider anisotropic matter (principal stresses unequal).

The use of general conformal motions, instead of homothetic motions (ψ = constant), allow us to find static and spherically symmetric distributions of matter which may be fitted to the exterior Schwarzschild metric.[6] We shall see that for the case of homothetic motions this fitting cannot be accomplished.

The introduction of anisotropic matter is suggested by recent theoretical works on more realistic equations of state and stellar models,[7,8] which indicate that some of these objects could have anisotropic pressures. Beside, it has been shown that some properties of anisotropic spheres may differ drastically from the properties of isotropic ones.[9–14] Anisotropy could be introduced by the existence of a solid core, by the presence of type P superfluid, or by the existence of an external field. Also, if the fluid is composed of two perfect fluids with different four-velocities, then the energy-momentum tensor can be cast into the standard form for anisotropic fluids.[15]

In this paper we do not discuss the mechanisms for inducing anisotropy. Rather we concentrate on the following two questions: (a) what kind of constraints on the hydrodynamical variables does the existence of the conformal motion impose? (b) How to generate exact solutions for anisotropic matter?

Discussion on the first question, as well as the conventions used, are given in Sec. II. In Sec. III we exhibit two solutions for both perfect fluid and anisotropic matter. Finally the results are discussed in the last section.

II. CONFORMAL MOTIONS AND THE HYDRODYNAMICAL VARIABLES

Let us consider a space-time whose metric tensor $g_{\mu\nu}$ is a solution of the Einstein equations for a distribution of matter represented by an anisotropic fluid. Further we shall assume that the space-time under consideration admits a one-parameter group of conformal motions [i.e., the metric satisfies Eq. (1)].

To find out the constraints that our assumptions impose on the hydrodynamical variables, let us start by taking the Lie derivative of Einstein equations

$$\underset{\xi}{L}(R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R) = -8\pi \underset{\xi}{L} T_{\mu\nu}, \tag{2}$$

where the energy-momentum tensor can be given as

$$T_{\mu\nu} = (\rho + P_\perp)U_\mu U_\nu - P_\perp g_{\mu\nu} + (P - P_\perp)\chi_\mu\chi_\nu,$$

where U^μ is the four-velocity, χ^μ is a unit spacelike vector orthogonal to U^μ, ρ is the energy density, P is the pressure in the direction of χ_μ, and $P_\perp$ is the pressure on the two-space orthogonal to χ_μ.

The Lie derivative of the energy momentum tensor $T_{\mu\nu}$ can be written as

$$\begin{aligned}\underset{\xi}{L} T_{\mu\nu} = &\left[\underset{\xi}{L}\rho + \underset{\xi}{L} P_\perp + \psi(\rho + P_\perp)\right] \\ &\times U_\mu U_\nu - g_{\mu\nu}(\underset{\xi}{L} P_\perp + \psi P_\perp) \\ &+ \left[\underset{\xi}{L} P - \underset{\xi}{L} P_\perp + \psi(P - P_\perp)\right]\chi_\mu\chi_\nu,\end{aligned}$$

where we have used the expressions

$$\begin{aligned}\underset{\xi}{L} U_\mu &= (\psi/2)U_\mu, \\ \underset{\xi}{L}\chi_\mu &= (\psi/2)\chi_\mu,\end{aligned} \tag{3}$$

[a] Postal address: Apartado 80793, Caracas 1080A, Venezuela.
[b] Centro de Física, I. V. I. C., Apdo. 1827, Caracas 1030-A, Venezuela.

 0022-2488/84/113274-05$02.50

which follow from the symmetry of the problem. In fact,

$$\underset{\xi}{L} U^{\mu} = \underset{\xi}{L} \frac{dx^{\mu}}{ds} = -U^{\mu} \frac{\underset{\xi}{L}\, ds}{ds}, \tag{4}$$

where we have used

$$\underset{\xi}{L}\, dx^{\mu} = 0 \tag{5}$$

(see Ref. 16, p. 89). On the other hand,

$$\underset{\xi}{L}\, ds^2 = 2\, ds \underset{\xi}{L}\, ds = \underset{\xi}{L} g_{\mu\nu}\, dx^{\mu}\, dx^{\nu} = \psi\, ds^2. \tag{6}$$

Feeding (6) back into (4) we get

$$\underset{\xi}{L} U^{\mu} = -(\psi/2)U^{\mu} \tag{7}$$

or, for the covariant components,

$$\underset{\xi}{L} U_{\mu} = (\psi/2)U_{\mu}. \tag{8}$$

In a similar way, the expression for $\underset{\xi}{L} \chi^{\mu}$ can be obtained.

Next, let us calculate the left-hand side of Eq. (2). For the Lie derivative of the Ricci tensor we have

$$\underset{\xi}{L} R_{\mu\nu} = \tfrac{1}{2} g^{\sigma\rho}(\nabla_{\nu}\nabla_{\mu} \underset{\xi}{L} g_{\sigma\rho} - \nabla_{\sigma}\nabla_{\mu} \underset{\xi}{L} g_{\nu\rho} - \nabla_{\sigma}\nabla_{\nu} \underset{\xi}{L} g_{\mu\rho} + \nabla_{\sigma}\nabla_{\rho} \underset{\xi}{L} g_{\mu\nu}) \tag{9}$$

(see Ref. 16, p. 52), where ∇ denotes covariant differentiation. Using (1) in (9) we get

$$\underset{\xi}{L} R_{\mu\nu} = \nabla_{\nu}\nabla_{\mu}\psi + \tfrac{1}{2} g_{\mu\nu} \Box\psi, \tag{10}$$

where

$$\Box\psi \equiv g^{\sigma\rho}\nabla_{\sigma}\nabla_{\rho}\psi.$$

Next, for the Lie derivative of the Ricci scalar R we obtain, using the expression

$$\underset{\xi}{L} g^{\alpha\beta} = -g^{\alpha\rho} g^{\beta\sigma} \underset{\xi}{L} g_{\rho\sigma}$$

and (10)

$$\underset{\xi}{L} R = \underset{\xi}{L} (g^{\mu\nu} R_{\mu\nu}) = 3\Box\psi - \psi R. \tag{11}$$

Thus Eq. (2) reads

$$\begin{aligned} &\nabla_{\mu}\nabla_{\nu}\psi - \tfrac{1}{2} g_{\mu\nu}\Box\psi \\ &= -8\pi\Big\{\Big[\underset{\xi}{L}\rho + \underset{\xi}{L} P_{\perp} + \psi(\rho + P_{\perp})\Big] \\ &\quad \times U^{\mu}U^{\nu} - g_{\mu\nu}(\underset{\xi}{L} P_{\perp} + \psi P_{\perp}) \\ &\quad + \Big[\underset{\xi}{L} P - \underset{\xi}{L} P_{\perp} + \psi(P - P_{\perp})\Big]\chi_{\mu}\chi_{\nu}\Big\} \end{aligned} \tag{12}$$

or, taking projections,

$$U^{\mu} U^{\nu} \nabla_{\mu}\nabla_{\nu}\psi - \tfrac{1}{2}\Box\psi = -8\pi(\underset{\xi}{L}\rho + \psi\rho), \tag{13}$$

$$\chi^{\mu}\chi^{\nu}\nabla_{\mu}\nabla_{\nu}\psi + \tfrac{1}{2}\Box\psi = -8\pi(\underset{\xi}{L} P + \psi P), \tag{14}$$

$$S^{\mu} S^{\nu} \nabla_{\mu}\nabla_{\nu}\psi + \tfrac{1}{2}\Box\psi = -8\pi(\underset{\xi}{L} P_{\perp} + \psi P_{\perp}), \tag{15}$$

$$-\Box\psi = -8\pi(\underset{\xi}{L} T + \psi T), \tag{16}$$

where $T \equiv T^{\mu}_{\mu} = \rho - 2P_{\perp} - P$, and S^{μ} is a unit spacelike vector orthogonal to χ^{μ} and U^{μ}.

Thus the transformation properties (the Lie derivatives) of the hydrodynamical variables are determined by the expressions (13)–(16). Moreover, there is one case, for which a specific equation of state for the stresses and the energy density appears as a consequence of the conformal motion. We have in mind the so-called special conformal motions.[17] For this subgroup of the conformal motions, which includes the homothetic motions as a special subcase, the function ψ satifies the condition

$$\nabla_{\nu}\nabla_{\mu}\psi = 0. \tag{17}$$

In this case, it follows at once from (13)–(16), that

$$\underset{\xi}{L}\rho + \psi\rho = 0, \tag{18}$$

$$\underset{\xi}{L} P + \psi P = 0, \tag{19}$$

$$\underset{\xi}{L} T + \psi T = 0, \tag{20}$$

$$\underset{\xi}{L} P_{\perp} + \psi P_{\perp} = 0. \tag{21}$$

Also it is easy to prove that the following law of conservation holds

$$\nabla_{\mu}(R^{\mu}_{\ \nu}\xi^{\nu}) = 0. \tag{22}$$

In fact, developing the left-hand side of (22) and using the Bianchi identities, we get

$$\nabla_{\mu}(R^{\mu}_{\ \nu}\xi^{\nu}) = \tfrac{1}{2}\frac{\partial R}{\partial x^{\nu}}\xi^{\nu} + R^{\mu\nu}\nabla_{\mu}\xi_{\nu}. \tag{23}$$

Then using (20), the Einstein equations, and

$$\nabla_{\nu}\xi_{\mu} + \nabla_{\mu}\xi_{\nu} = \psi g_{\mu\nu}$$

we obtain (22).

We shall further restrict our special conformal motions to two specific subcases, namely (a) the vector field ξ^{μ} is collinear with χ^{μ} (i.e., $\xi^{\mu} = \lambda\chi^{\mu}$, λ being an arbitrary function of the coordinates); (b) the vector field ξ^{μ} is spacelike and orthogonal to U^{μ} and χ^{μ}.

Let us consider the first case. Using the Einstein equations and the fact that $\xi^{\mu} = \lambda\chi^{\mu}$, we get

$$R^{\mu}_{\ \nu}\xi^{\nu} = 8\pi\xi^{\mu}(-P_{\perp} + \rho/2 + P/2). \tag{24}$$

Taking divergence of (24), using the expressions (18)–(21), the definition of the Lie derivative of a scalar, and

$$\nabla_{\mu}\xi^{\mu} = 2\psi,$$

we finally obtain

$$\nabla_{\mu}(R^{\mu}_{\ \nu}\xi^{\nu}) = 8\pi\psi(\rho/2 + P/2 - P_{\perp}). \tag{25}$$

Now, since the left-hand side of (25) vanishes, according to (22), then

$$\rho = 2P_{\perp} - P. \tag{26}$$

Thus the relationship between the stresses and the density is

given in a unique way by Eq. (26), provided the space-time admits a one-parameter group of special conformal motions, with the vector field ξ^μ parallel to χ^μ. Observe that for perfect fluids ($P_\perp = P$), the equation of state becomes

$$P = \rho. \tag{27}$$

This equation of state has been widely used in general relativity to obtain stellar and cosmological models for ultradense matter.[3,18–21]

Let us now consider the case when

$$\xi^\alpha U_\alpha = \xi^\alpha \chi_\alpha = 0. \tag{28}$$

Then from the Einstein equations we get

$$R^\mu_\nu \xi^\nu = (8\pi/2)\xi^\mu(\rho - P). \tag{29}$$

Taking divergence of (29), and using the conservation law (22) and the expressions (18), (19), we obtain

$$P = \rho. \tag{30}$$

No constraint involving the tangential pressure $P_\perp$ was found in this case.

Thus for perfect fluids we get again the stiff equation of state, pressure equal to the energy density.

Let us now see what kind of constraints are derived from the transformation law of the four-velocity [Eq.(8)].

Using the definition of the Lie derivative of a covariant vector, we have

$$\underset{\xi}{L} U_\nu = \xi^\alpha \nabla_\alpha U_\nu + U_\alpha \nabla_\nu \xi^\alpha, \tag{31}$$

or, from the fact that ξ^α and U^α are orthogonal to each other,

$$\underset{\xi}{L} U_\nu = \xi^\alpha(\nabla_\alpha U_\nu - \nabla_\nu U_\alpha). \tag{32}$$

Let us now introduce the tensor $\gamma_{\mu\nu}$, as

$$\gamma_{\mu\nu} = g_{\mu\nu} - U_\mu U_\nu, \tag{33}$$

which defines the projection operator onto the three-space quotient to the streamlines.

We can now define the kinematic quantities which characterize the streamlines, they are the acceleration

$$a^\mu = U^\alpha \nabla_\alpha U^\mu, \tag{34}$$

the expansion

$$\Theta = \nabla_\mu U^\mu, \tag{35}$$

the shear

$$\sigma_{\mu\nu} = \gamma^\alpha_\mu \gamma^\beta_\nu \{\nabla_{(\alpha} U_{\beta)} - \tfrac{1}{3}\Theta\gamma_{\alpha\beta}\}, \tag{36}$$

and the vorticity bivector

$$\omega_{\mu\nu} = \gamma^\alpha_\mu \gamma^\beta_\nu \nabla_{[\alpha} U_{\beta]}. \tag{37}$$

(Round brackets and square brackets between the indices denote symmetrization and antisymmetrization, respectively.) It follows from these definitions that

$$\nabla_\nu U_\mu = a_\mu U_\nu + \omega_{\mu\nu} + \sigma_{\mu\nu} + \tfrac{1}{3}\Theta\gamma_{\mu\nu}. \tag{38}$$

Feeding (38) back into (32) and comparing the result with (8), we get

$$\underset{\xi}{L} U_\nu = -\xi^\alpha a_\alpha U_\nu + 2\xi^\alpha \omega_{\nu\alpha} = (\psi/2)U_\nu. \tag{39}$$

Contracting this last equation with U^ν, it follows that

$$\psi = -2a^\nu \xi_\nu$$

which implies, using (39) again, that

$$\xi^\alpha \omega_{\alpha\nu} = 0. \tag{40}$$

This relation is valid for all kind of conformal motions provided ξ^α is orthogonal to U^α.

III. SOME EXACT SOLUTIONS FOR STATIC AND SPHERICALLY SYMMETRIC DISTRIBUTIONS OF MATTER

In this section we shall assume that the metric tensor not only admits the one-parameter group of conformal motions but also is static and spherically symmetric. In the usual Schwarzschild coordinates the line element may be written as

$$ds^2 = A^2(r)dt^2 - B^2(r)dr^2 - r^2(d\Theta^2 + \sin^2\Theta\, d\phi^2) \tag{41}$$

and in comoving coordinates we may choose

$$U^\mu(U^0,0,0,0), \quad \chi^\mu(0,\chi^1,0,0).$$

Since $U^\mu U_\mu = -\chi^\mu \chi_\mu = 1$, then

$$U^0 = 1/A(r), \quad \chi^1 = 1/B(r),$$

and the components of the energy-momentum tensor are

$$T^0_0 = \rho, \quad T^1_1 = -P, \quad T^2_2 = T^3_3 = -P_\perp.$$

It is easy to prove that by virtue of the spherical symmetry and the independence of the metric tensor on the timelike coordinate the most general form of ξ^α is

$$\xi^\alpha = \lambda\chi^\alpha, \tag{42}$$

where λ is an arbitrary function of r.

Now, the corresponding field equations are given as

$$8\pi P = \frac{1}{B^2}\left(\frac{2A'}{Ar} + \frac{1}{r^2}\right) - \frac{1}{r^2}, \tag{43}$$

$$8\pi P_\perp = \frac{1}{B^2}\left\{\frac{A''}{A} - \frac{A'B'}{AB} + \frac{1}{r}\left(\frac{A'}{A} - \frac{B'}{B}\right)\right\}, \tag{44}$$

$$8\pi\rho = \frac{1}{B^2}\left(\frac{2B'}{Br} - \frac{1}{r^2}\right) + \frac{1}{r^2} \tag{45}$$

(primes denote differentiation with respect to r).

The functions A and B are further restricted by the condition (1), which implies, in our case

$$A(r) = C_1 r, \tag{46}$$

$$B(r) = 2C_2/\psi, \tag{47}$$

$$\lambda = C_2 r, \tag{48}$$

where C_1 and C_2 are two constants, and we have used (41) and (42). Feeding (46)–(48) back into the field equations (43)–(45) we get

$$8\pi P = \frac{3\psi^2}{4C_2^2}\frac{1}{r^2} - \frac{1}{r^2}, \tag{49}$$

$$8\pi P_\perp = \frac{\psi^2}{4C_2^2}\left\{\frac{2\psi'}{r\psi} + \frac{1}{r^2}\right\}, \tag{50}$$

$$8\pi\rho = \frac{1}{r^2} - \frac{\psi^2}{4C_2^2}\left\{\frac{2\psi'}{r\psi} + \frac{1}{r^2}\right\}. \tag{51}$$

Thus, different choices of the function ψ will lead to different classes of solutions with the symmetry properties

specified above. If one desires to match any of those solutions to the exterior Schwarzschild metric on the boundary of the source, then the radial pressure should vanish for some finite value of the radial coordinate (say $r = r_0$), and the functions $A(r)$ and $B(r)$ should be continuously joined to the corresponding values of the vacuum Schwarzschild metric.[6] Thus,

$$A^2(r_0) = C_1^2 r_0^2 = 1 - 2M/r_0, \tag{52}$$

$$B^2(r_0) = \frac{4C_2^2}{\psi^2(r_0)} = \frac{1}{1 - 2M/r_0} \tag{53}$$

(where M is the total mass), whereas from the vanishing pressure condition we get

$$\psi^2(r_0) = 4C_2^2/3. \tag{54}$$

Combining (53) and (54) we get

$$M/r_0 = \tfrac{1}{3}. \tag{55}$$

Furthermore, this value for the ratio M/r_0 may be obtained just by integration of the energy density, given by (51), over the sphere of radius r_0, in fact

$$M = \int_0^{r_0} 4\pi r^2 \rho \, dr = \int_0^{r_0} r^2 \left\{ \frac{1}{2r^2} - \frac{1}{8C_2^2 r^2}(2\psi\psi' r + \psi^2) \right\} dr \tag{56}$$

or

$$M = \frac{r_0}{2} - \frac{1}{8C_2^2} \int_0^{r_0} (\psi^2 r)' \, dr. \tag{57}$$

Assuming $\psi(0) < \infty$, we get

$$M = r_0/2 - (1/8C_2^2)\psi^2(r_0) r_0, \tag{58}$$

and using (54) in (58) we finally obtain

$$M/r_0 = \tfrac{1}{3}.$$

Two main conclusions follow from the results above.

(a) All solutions, for any choice of the function ψ (bounded in the interval $0 \le r \le r_0$) will have the same gravitational potential on the boundary, provided the boundary is a vanishing pressure (radial) surface.

(b) The existence of the vanishing pressure (radial) surface ensures the fulfillment of the condition (53). Furthermore, since the condition (52) may always be satisfied by an appropriated choice of the constant C_1, we can conclude that the existence of the vanishing pressure surface ensures the matching of any of the solutions to the vacuum Schwarzschild solution on the boundary of the sphere.

Let us now specialize the choice of ψ.

A. $\psi = 2$

In this case we get

$$P = \frac{1}{8\pi}\left\{ \frac{3}{C_2^2} - 1 \right\} \frac{1}{r^2}, \tag{59}$$

$$P_\perp = 1/8\pi C_2^2 r^2, \tag{60}$$

$$\rho = \frac{1}{8\pi}\left\{ 1 - \frac{1}{C_2^2} \right\} \frac{1}{r^2}, \tag{61}$$

and for the metric functions

$$A = C_1 r, \quad B = C_2. \tag{62}$$

Next, from (59)–(61) it follows that

$$\rho = 2P_\perp - P, \tag{63}$$

which was to be expected according to the precedent section ($\psi = 2$, defines a special conformal motion and $\xi^\mu = \lambda\chi^\mu$). The solution given by (59)–(62) is the generalization to anisotropic matter of previously known solutions,[1,2] and is a special case of one of the Bayin solutions.[13] When the constant C_2 is chosen to be $C_2 = \sqrt{2}$, matter becomes a perfect fluid and

$$P = P_\perp = \rho.$$

Finally, observe that the solution above cannot be matched to the exterior Schwarzschild metric, as it can be seen from the fact, that the pressure (radial) does not vanish for any finite value of the radial coordinate.

B. The perfect fluid solution

One important point emerging from Eqs. (49)–(51) is that there exists only one choice of the function ψ (which includes $\psi = 2$, as a special subcase) for which there exists a perfect fluid solution. In fact, from the condition

$$P = P_\perp$$

and using (49) and (50) we get the equation

$$r\psi\psi' + 2C_2^2 - \psi^2 = 0, \tag{64}$$

whose general solution is

$$\psi^2 = C_2^2(Cr^2 + 2), \tag{65}$$

where C is a constant. For $C = 0$; $C_2^2 = 2$, we recover the $\psi = 2$ subcase mentioned above.

If we now demand the radial pressure to vanish for some finite value of the radial coordinate (say $r = r_0$), then we get from (49) and (65)

$$r_0^2 = -2/3C, \quad C < 0. \tag{66}$$

We can now write the expressions for the pressure, the density, and the line element

$$P = P_\perp = \frac{1}{8\pi}\left\{ \frac{1}{2r^2} - \frac{1}{2r_0^2} \right\}, \tag{67}$$

$$\rho = \frac{1}{8\pi}\left\{ \frac{1}{2r^2} + \frac{1}{2r_0^2} \right\}, \tag{68}$$

$$ds^2 = C_1^2 r^2 dt^2 - \frac{4\,dr^2}{(Cr^2 + 2)} - r^2(d\Theta^2 + \sin^2\Theta\, d\phi^2) \quad (0 \le r \le r_0) \tag{69}$$

from which it can be seen at once that

$$\rho \ge P \ge 0. \tag{70}$$

As it was shown above the total mass of the sphere will be

$$M = r_0/3.$$

Thus it will be less compact than the interior Schwarzschild sphere for which

$$M/r_0 = 0.44. \tag{71}$$

Finally observe that the singular surface $r^2 = -2/C$ in the line element (69) is outside the sphere, whose radius is given by Eq. (66).

C. An anisotropic solution

As we have just seen the perfect fluid case occurs whenever the function ψ is chosen according to the expression (65). Thus, in principle, there exists an anisotropic solution for any other choice of the function ψ.

Here, we shall give one example, whose physical properties are reasonable and which possesses vanishing pressure surface.

Let us take

$$\psi^2 = C_2^2(Cr^2 + 2) + C_2^2 H, \tag{72}$$

where H is a constant which measures the anisotropy. Feeding (72) back into (49)–(51) we get

$$8\pi P = \frac{3}{4}C + \left(\frac{3H}{4} + \frac{1}{2}\right)\frac{1}{r^2}, \tag{73}$$

$$8\pi P_\perp = \frac{3}{4}C + \left(\frac{H}{4} + \frac{1}{2}\right)\frac{1}{r^2}, \tag{74}$$

$$8\pi\rho = -\frac{3}{4}C + \left(\frac{1}{2} - \frac{H}{4}\right)\frac{1}{r^2}, \tag{75}$$

and for the radius of the sphere

$$r_0^2 = -\frac{2}{3C}\left(1 + \frac{3H}{2}\right), \quad C<0 \tag{76}$$

In order to ensure the positiveness of the energy density and the stresses, we have to restrict H to the interval

$$-\tfrac{3}{2} \le H \le 0. \tag{77}$$

Also, in that interval the energy density will be larger than the stresses. As in the perfect fluid case the function ψ vanishes for a value of the radial coordinate which is bigger than r_0.

IV. CONCLUSIONS

We have seen so far that, under the assumptions of Sec. II, the existence of a one-parameter group of conformal motions introduces specific restrictions on the hydrodynamical variables. Furthermore, for the case of special conformal motions, the stiff equation of state (pressure equal to the energy density) is singled out in a unique way, provided the vector field ξ^α is orthogonal to the four-velocity. So far we do not know whether this link between the stiff equation of state and the group of special conformal motions is "casual" or if there is any "deeper" physical meaning behind it.

Concerning the solutions presented in Sec. III, they could serve as initial (or final) configurations in a self-similar evolution scenario. They share reasonable physical properties (positiveness of the energy density and stresses, energy density larger than stresses) and possess vanishing pressure surfaces. We would like to stress the fact that the gravitational potential at the surface will be the same for all solutions, provided the boundary is a vanishing pressure surface. We recall that for others' previously known anisotropic solutions,[11,13,14] the ratio M/r_0 depends on the degree of anisotropy, and it can approach the limit value $1/2$ as close as one desires, provided large amounts of anisotropy are allowed.

Finally we would like to stress the point, that for any choice of ψ different from (65) and (72), different anisotropic solutions could be obtained.

[1] M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).
[2] R. N. Henriksen and P. S. Wesson, Astrophys. Space Sci. **53**, 429 (1978).
[3] G. V. Bicnell and R. N. Henriksen, Astrophys. J. **219**, 1043 (1978).
[4] G. V. Bicnell and R. N. Henriksen, Astrophys. J. **225**, 237 (1978).
[5] Douglas M. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[6] L. Herrera and J. Jiménez, Phys. Rev. D **28**, 2987 (1983).
[7] V. Canuto, "Neutron Stars: General Review," presented at the Solvay Conference on Astrophysics and Gravitation, Brussels, 1973.
[8] M. Ruderman, Ann. Rev. Astron. Astrophys. **10**, 427 (1972).
[9] R. Bowers and E. Liang, Astrophys. J. **188**, 657 (1974).
[10] L. Herrera, G. J. Ruggeri, and L. Witten, Astrophys. J. **234**, 1094 (1979).
[11] M. Consenza, L. Herrera, M. Esculpi, and L. Witten, J. Math. Phys. **22**, 118 (1981).
[12] M. Consenza, L. Herrera, M. Esculpi, and L. Witten Phys. Rev. D **25**, 2527 (1982).
[13] S. Bayin, Phys. Rev. D **26**, 1262 (1982).
[14] B. W. Stewart, J. Phys. A **15**, 2419 (1982).
[15] S. J. Putterman, *Superfluid Hydrodynamics* (North-Holland, Amsterdam, 1974).
[16] K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1957).
[17] G. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).
[18] R. Tabensky, and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).
[19] Ya. B. Zeldovich, Zh. Eksp. Teor. Fiz. **41**, 1609 (1961) [Sov. Phys. JETP **14**, 1143 (1962)].
[20] P. S. Wesson, J. Math. Phys. **19**, 2283 (1978).
[21] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton University, Princeton, NJ, 1975).

Kantowski–Sachs cosmological models approaching isotropy

Edmond Weber
Institut de Physique Théorique, Université Catholique de Louvain, B-1348-Louvain-la-Neuve, Belgium and Université du Burundi, Bujumbura, Burundi[a)]

(Received 25 August 1983; accepted for publication 23 March 1984)

In this article we analyze a particular class of anisotropic cosmological models, the Kantowski–Sachs models, in the presence of a nonzero cosmological constant Λ. We study them qualitatively by means of autonomous systems with two and three dimensions. The plane autonomous system gives a new class of empty Kantowski–Sachs cosmologies, with $\Lambda > 0$ and $\Lambda < 0$. We find two new types of singularity points. The autonomous system with three dimensions yields a set of solutions of nonzero measure becoming isotropic in an infinite cosmological time.

PACS numbers: 04.20. – q, 98.80.Dr

I. INTRODUCTION

During the last fifteen years, spatially homogeneous cosmological models belonging to the Bianchi class have been studied in the framework of general relativity.[1–3] An exceptional case to these cosmologies was discovered by Kantowski and Sachs.[4] The isometry group for these models contains a three-parameter Lie group G_3 whose orbits are two-dimensional. The curvature of such an orbit is constant and positive. An improved version of the proofs of these statements has been given by Collins.[5] He analyzes the Kantowski–Sachs cosmologies containing a perfect fluid with a zero cosmological constant. The field equations can be transformed in this case into a plane autonomous system. This permits an easy qualitative study of the evolution of the Kantowski–Sachs cosmological models. This method has been applied to more types of Bianchi models.[6–8]

In this article we introduce a further parameter, the cosmological constant Λ, which gives rise to a three-dimensional autonomous system when we consider nonempty Kantowski–Sachs models containing a perfect fluid with an equation of state of the form $p = (\gamma - 1)\mu$, where μ is the density of matter, p the pressure, and γ a constant whose values lie in the range $1 \leqslant \gamma \leqslant 2$. We will derive this three-dimensional system in Sec. II below, and discuss the singular points at finite distance that it exhibits. A more general result of the geodesic incompleteness of the Kantowski–Sachs models than that presented by Collins[5] can be given. In fact, for a null or negative cosmological constant, these models have a past or future singularity of the cigar, pancake, barrel, or point-type.

When we consider only empty Kantowski–Sachs models in the presence of a cosmological constant, our three-dimensional autonomous system reduces to a plane one. We will study extensively this case in Sec. III.

It is the simultaneous presence of matter and of a cosmological constant which leads to a three-dimensional autonomous system. The latter will be studied in Sec. IV. An interesting result emerges (Sec. V), namely the existence of a set of nonzero measure of Kantowski–Sachs solutions to the field equations which approach isotropy in an infinite cosmological time, contrary to a claim made by Collins and Hawking,[9] who did not consider the influence of a cosmological constant.

Finally, the theorems necessary for the study of our three-dimensional autonomous system are mentioned in the Appendix.[10–12]

II. QUALITATIVE ANALYSIS

The Kantowski–Sachs metric[4] takes the form

$$ds^2 = dt^2 - X^2(t)\,dr^2 - Y^2(t)(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{2.1}$$

in the coordinates (t, r, θ, ϕ), where t is the cosmic time coordinate, r a radial coordinate, and θ, ϕ the usual spherical coordinates. Here $X(t)$ and $Y(t)$ are two unknown functions of t. Einstein's field equations with a cosmological constant Λ can be written as follows (μ and p being the fluid parameters described above):

$$2\frac{\dot{X}\dot{Y}}{XY} + \left(\frac{1+\dot{Y}^2}{Y^2}\right) - \Lambda = \mu, \tag{2.2}$$

$$2\frac{\ddot{Y}}{Y} + \left(\frac{1+\dot{Y}^2}{Y^2}\right) - \Lambda = -p, \tag{2.3}$$

$$\frac{\ddot{Y}}{Y} + \frac{\ddot{X}}{X} + \frac{\dot{X}\dot{Y}}{XY} - \Lambda = -p. \tag{2.4}$$

A dot in these equations denotes differentiation with respect to t.

By using the volume expansion $\theta = \dot{X}X^{-1} + 2\dot{Y}Y^{-1}$ and the shear $\sigma = 3^{-1/2}(\dot{X}X^{-1} - \dot{Y}Y^{-1})$, we can express the field equations as follows:

$$\dot{\theta} + 3^{-1}\theta^2 + 2\sigma^2 + 2^{-1}(\mu + 3p) - \Lambda = 0, \tag{2.5}$$

$$\dot{\sigma} + \sigma\theta - 3^{-1/2}Y^{-2} = 0, \tag{2.6}$$

$$3^{-1}\theta^2 - \sigma^2 + Y^{-2} - \Lambda = \mu. \tag{2.7}$$

Equation (2.5) is the well-known Raychaudhuri's equation in the case of a perfect fluid and a nonzero cosmological constant. We shall analyze Kantowski–Sachs models for which the pressure p and the matter density μ of the perfect fluid are related by the barotropic equation of state $p = (\gamma - 1)\mu$. The values of the constant γ lie in the range $1 \leqslant \gamma \leqslant 2$. We shall use the variables $\Omega(t)$ and $\beta(t)$ instead of $X(t)$ and $Y(t)$, defined by $X \equiv \exp(-\Omega + \beta)$ and $Y \equiv \exp(-\Omega - \beta/2)$. The time variable will be Ω and a

[a)] Present address.

 0022-2488/84/113279-07$02.50

prime will denote differentiation with respect to Ω. We then introduce quantities measuring respectively the dynamical importance of the shear, i.e., $\beta' = -(2\sqrt{3}\,\sigma)/\theta$, and the dynamical importance of the fluid, i.e., $x = 3\mu/\theta^2 = \mu/3\dot{\Omega}^2$, as well as a third quantity $z = 3x/\mu$.

These definitions allow us to reexpress the field equations (2.2)–(2.4) together with the conservation equation

$$\dot{\mu} + (\mu + p)\,\theta = 0 \tag{2.8}$$

in the form of a three-dimensional autonomous system and a constraint equation, with three dependent variables β', x, z, and with Ω as the independent variable.

Equation (2.2) becomes

$$\beta'^2 + 4x - 4 + \tfrac{4}{3}\Lambda z = (\tfrac{4}{3}\dot{\Omega}^2)\exp(2\Omega + \beta). \tag{2.9}$$

By eliminating $\ddot{\Omega}$ from (2.3) and (2.4) and substituting the expression of the $\exp(2\Omega + \beta)$ term obtained from (2.9) we have

$$\beta'' = \tfrac{1}{2}\beta'[4 - \beta'^2 - (3\gamma - 2)x + \tfrac{2}{3}\Lambda z] - \tfrac{1}{2}[4 - 4x - \beta'^2 - \tfrac{4}{3}\Lambda z]. \tag{2.10}$$

The elimination of $\ddot{\beta}$ in (2.3) and (2.4) yields

$$-6\ddot{\Omega} + 9\dot{\Omega}^2 + \tfrac{9}{4}\dot{\beta}^2 - 3\Lambda + \exp(2\Omega + \beta) = -3p. \tag{2.11}$$

By using (2.8) and (2.9) we get

$$x' = x[(3\gamma - 2)(1 - x) - \beta'^2 + 2\Lambda z/3]. \tag{2.12}$$

Differentiation of z with respect to Ω gives

$$z' = -2z[1 + \tfrac{1}{2}(3\gamma - 2)x + \tfrac{1}{2}\beta'^2 - \Lambda z/3]. \tag{2.13}$$

Equations (2.10), (2.12), and (2.13) form a three-dimensional autonomous system of ordinary differential equations. The qualitative behavior of the solutions will be drawn in the (x, β', z) phase space. The region of interest is given by ($x > 0$, $z > 0$, and $\beta'^2 + 4x - 4 + \tfrac{4}{3}\Lambda z > 0$).

The singular points at finite distance, also called critical points, i.e., the points (x, β', z) where the right-hand sides of the system vanish simultaneously, are different according to the constant γ, or do not exist (as real numbers) according to the value of the cosmological parameter Λ. By taking account of the physical region we have the situation represented in Table I. The study of the three-dimensional autono-

TABLE I. This table summarizes all critical points in the two cases: $1 \leqslant \gamma < 2$ and $\gamma = 2$. In the first case we have divided the points according to their appurtenance to the x- or β'-axis, or (β', z) plane when $\Lambda > 0$. In the second case there is a continuous line of critical points in the (x, β') plane. There are no (real) critical points in the (β', z) plane when $\Lambda \leqslant 0$.

	$\Lambda > 0$ (x, β', z)	$\Lambda \leqslant 0$
$1 \leqslant \gamma < 2$	(1,0,0)	(1,0,0)
	(0,2,0)	(0,2,0)
	(0, − 2,0)	(0, − 2,0)
	(0,0,3/Λ)	...
	(0, − 2,9/Λ)	...
$\gamma = 2$	$(z = 0,\ 4 - 4x - \beta'^2 = 0)$	$(z = 0,\ 4 - 4x - \beta'^2 = 0)$
	(0,0,3/Λ)	...
	(0, − 2,9/Λ)	...

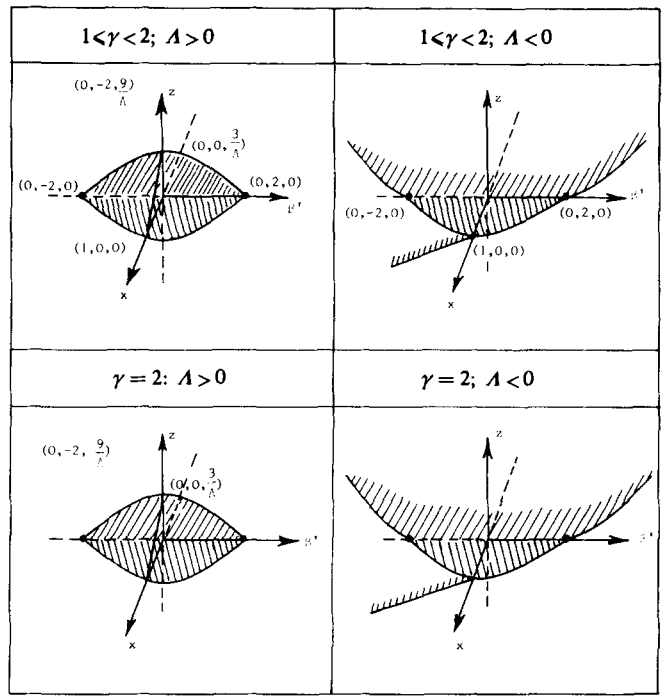


FIG. 1. We have indicated the region of physical interest for the three-dimensional autonomous system, as well as the singular points at finite distance according to the value of γ and of the cosmological constant Λ. Striped parts are outside the region of physical interest.

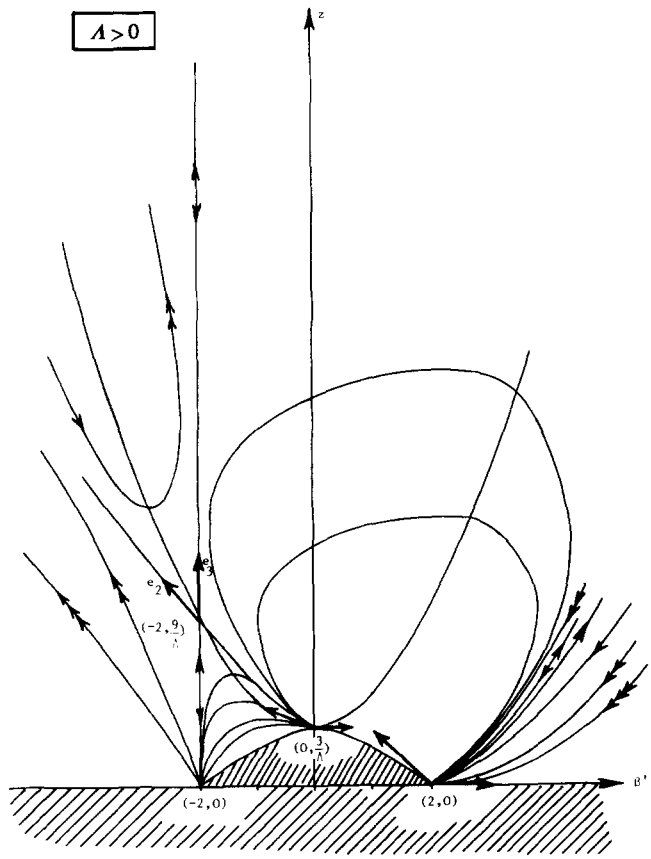


FIG. 2. Qualitative description of the evolution of the empty Kantowski–Sachs models when $\Lambda > 0$. Each curve represents the evolution of a model for a fixed set of initial conditions. The variable $\beta' = -2\sqrt{3}(\sigma/\theta)$ measures the relative dynamical importance of the fluid shear. The associate types of arrows indicate the entire course of evolution. The time reverse is also possible.

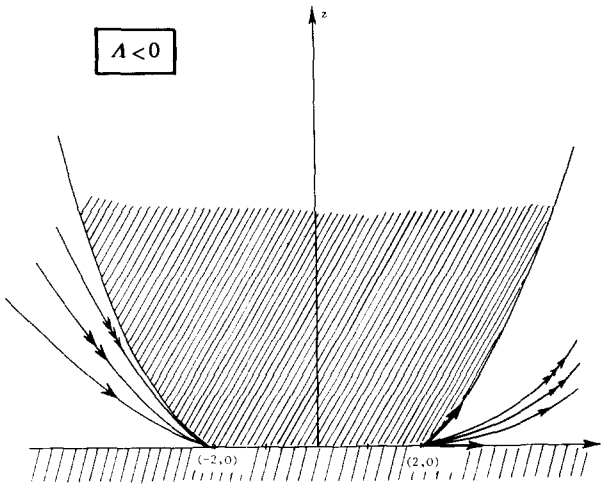


FIG. 3. Qualitative description of the evolution of the empty Kantowski–Sachs models when $\Lambda < 0$.

mous system can then be subdivided into four different cases: $(1 \leqslant \gamma < 2, \Lambda > 0)$; $(1 \leqslant \gamma < 2, \Lambda < 0)$; $(\gamma = 2, \Lambda > 0)$; and $(\gamma = 2, \Lambda < 0)$. We have represented the singular points in Fig. 1.

When the cosmological constant vanishes, we have a plane autonomous system in the variables x, β'. It has been studied in detail by Collins.[4] By setting $x = 0$ in (2.10), (2.12), and (2.13) we obtain a new autonomous system

$$\beta'' = \tfrac{1}{2}\beta'[4 - \beta'^2 + \tfrac{2}{3}\Lambda z] - \tfrac{1}{2}[4 - \beta'^2 - \tfrac{4}{3}\Lambda z], \quad (2.14)$$

and

$$z' = -2z[1 + \tfrac{1}{2}\beta'^2 - \Lambda z/3]. \quad (2.15)$$

It describes empty Kantowski–Sachs models in the presence of a cosmological constant. The region of physical interest is given by $(z > 0, \beta'^2 + \frac{4}{3}\Lambda z - 4 > 0)$ and the singular points at finite distance according to the value of Λ are $(\beta' = \pm 2, z = 0)$, $(\beta' = 0,\ z = 3/\Lambda)$, and $(\beta' = -2,\ z = 9/\Lambda)$ when $\Lambda > 0$, and $(\beta' = \pm 2,\ z = 0)$ when $\Lambda < 0$.

In his article,[5] Collins has shown that all perfect-fluid-filled Kantowski–Sachs models are geodesically incomplete both to the future and the past. This theorem can be generalized to negative values of Λ.

III. PLANE AUTONOMOUS SYSTEM FOR EMPTY KANTOWSKI–SACHS MODELS WITH A COSMOLOGICAL CONSTANT

In order to study qualitatively the plane autonomous system (2.14) and (2.15) we examine the behavior of integral curves in the neighborhood of the critical points at finite distance as well as at infinity and we join the two regions.

The critical points at finite distance are simple.[13,14] Two of them, (2,0) and $(0,3/\Lambda)$, are improper nodes; the point $(-2,0)$ is a proper node, and the point $(-2,9/\Lambda)$ is a saddle point.

TABLE II. Empty Kantowski–Sachs models with $\Lambda > 0$ and $\Lambda < 0$. We indicate for these models the singularity type as well as the asymptotic behavior of physically relevant variables: the average length scale l, the length scales X and Y, the fluid expansion θ, the fluid shear σ, the Ricci scalar ${}^3R$, and the integrated shear β. Here cte denotes a nonzero finite limit.

Λ	(β', z)	Singularity type	$l = e^{-\alpha}$	Cosmological singularity	X	Y	θ	σ	${}^3R = 2/Y^2$	β	Dominant terms in Raychaudhuri equation
$\Lambda > 0$	(2,0)	cigar $X \to \infty$, $Y \to 0$	$(\pm t)^{1/3}$	$\Omega \to \infty : t \to 0_{\pm}$	$(\pm t)^{-1/3}$	$(\pm t)^{+2/3}$	t^{-1}	$-t^{-1}$	$t^{-4/3}$	$-\ln\lvert t\rvert$	$\sigma^2, \theta^2 \sim t^{-2}$
$\Lambda < 0$	$(-2,0)$	pancake $X \to 0$, $Y \to$ cte	$(\pm t)^{1/3}$	$\Omega \to \infty : t \to 0_{\pm}$	$\pm t$	cte	$\pm t^{-1}$	t^{-1}	cte	$\ln\lvert t\rvert$	$\sigma^2, \theta^2 \sim t^{-2}$
	$(0,3/\Lambda)$	infinity $X \to \infty$, $Y \to \infty$, e_2 direction	$e^{\mp\sqrt{\Lambda/3}t}$	$\Omega \to -\infty : t \to \mp\infty$	$e^{\mp\sqrt{\Lambda/3}t}$	$e^{\mp\sqrt{\Lambda/3}t}$	$\mp\sqrt{3\Lambda}$	$e^{\pm 2\sqrt{\Lambda/3}t}$	$e^{\pm 2\sqrt{\Lambda/3}t}$	$e^{\pm 2\sqrt{\Lambda/3}t}$	Λ and $\theta^2 = 3\Lambda$
$\Lambda > 0$	$(-2,9/\Lambda)$	pancake $X \to 0$, $Y \to$ cte, e_3 direction	$e^{\pm\sqrt{\Lambda/3}t}$	$\Omega \to \infty : t \to \mp\infty$	$e^{\pm\sqrt{\Lambda}t}$	cte	$\pm\sqrt{\Lambda}$	$\pm\sqrt{\Lambda}$	cte	$\pm 2\sqrt{\Lambda/3}t$	$\sigma^2, \theta^2, \Lambda$
		infinite barrel $X \to \infty$, $Y \to$ cte	$e^{\mp\sqrt{\Lambda/3}t}$	$\Omega \to \infty : t \to \mp\infty$	$e^{\pm\sqrt{\Lambda}t}$	cte	$\mp\sqrt{\Lambda}$	$\mp\sqrt{\Lambda}$	cte	$\mp 2\sqrt{\Lambda/3}t$	$\sigma^2, \theta^2, \Lambda$

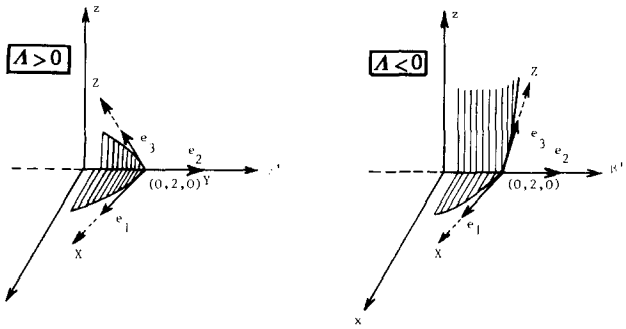


FIG. 4. We indicate the region of physical interest which is outside the striped part in the neighborhood of the critical point (0,2,0) for $\Lambda > 0$ and $\Lambda < 0$. Here e_1, e_2, e_3 are the three characteristic vectors.

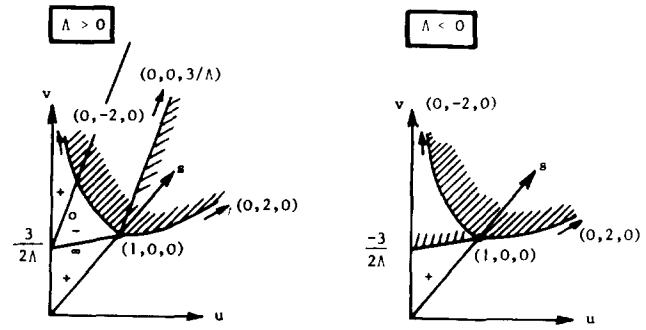


FIG. 5. We indicate the singular points at infinity: $(s = 0, u = 0, v > 0)$ when $\Lambda > 0$, and $(s = 0, u = 0, 0 \leqslant v \leqslant -3/\Lambda)$ when $\Lambda < 0$. Striped parts are outside the region of physical interest.

The two Poincaré transformations $(\beta' = s^{-1}, z = us^{-1})$ and $(\beta' = vs^{-1}, z = s^{-1})$ enable us to study the critical points at infinity. In the variables (u, s) we have the simple singular point (0,0) whose topological structure is a saddle point. On the z-axis at infinity, we find $(v = 0, s = 0)$, which is a double singular point. The theory is well established for such multiple equilibrium states.[13] We find three directions of approach (in polar coordinates) $\psi = 0$, arctan $(-\frac{1}{2})$, and π when $\Lambda > 0$, and two directions $\psi = 0$ and π when $\Lambda < 0$.

The global picture is given by the integral curves of the system (2.14) and (2.15) drawn in Figs. 2 and 3. In each diagram there are integral curves starting at a finitely distant singular point (in the variables β' and z), extending to infinitely large values of z and coming back to another singular point at finite distance. When $\Lambda > 0$ we have such behaviors as well as other ones like curves which are time symmetric and curves for which we do not have $z \to \infty$. The arrows depict the entire course of evolution, but the time reverses are also possible.

The critical point $(\beta' = 2, z = 0)$ is a "cigar" singularity type, whereas $(\beta' = -2, z = 0)$ is a "pancake" singularity. The Raychaudhuri equation (2.5) tells us that at these points the shear σ and the expansion θ are dominant and Λ is negligible. The point $(0, 3\Lambda^{-1})$ is an "infinite" singularity, so called because $X \to \infty$ and $Y \to \infty$ from the metric point of view, and as autonomous system it is a singular point. The quantities Λ and θ are dominant and σ is negligible. The saddle point $(-2, 9\Lambda^{-1})$ is a "pancake" singularity in one direction, and an "infinite barrel" $(X \to \infty, Y \to \text{const})$ in the other direction, which is a new type of singularity; σ, θ, and Λ are equally important. The asymptotic behavior of all these quantities as a function of t is given in Table II.

IV. NONEMPTY KANTOWSKI–SACHS MODELS WITH A COSMOLOGICAL CONSTANT

We will follow for the study of the three-dimensional system the same pattern as in two dimensions. All the critical points at finite distance are simple. The point $(x = 0, \beta' = 2, z = 0)$ is a node; the characteristic roots $\lambda = 3\gamma - 6$, $\mu = -2$, $\nu = -6$ are negative when $1 \leqslant \gamma < 2$. The orbits starting at a sphere (see the Appendix) centered at (0,2,0) tend to this point for $\Omega \to \infty$. We distinguish three cases:

$$1 \leqslant \gamma < \tfrac{4}{3}: \quad |\mu| < |\lambda| < |\nu|, \tag{4.1}$$

$$\gamma = \tfrac{4}{3}: \quad |\mu| = |\lambda| < |\nu|, \tag{4.2}$$

$$\tfrac{4}{3} < \gamma < 2: \quad |\lambda| < |\mu| < |\nu|. \tag{4.3}$$

When $\gamma \neq \frac{4}{3}$ we have the three characteristic vectors $e_1 = (1, -1, 0)$ $e_2 = (0,1,0)$, and $e_3 = (0, -1, 3/\Lambda)$ corresponding to λ, μ, and ν, respectively. When $\gamma = \frac{4}{3}$ we have an infinity of characteristic vectors in the plane (x, β'). The coordinate system (X, Y, Z) with origin at the point (0,2,0) is associated with these vectors (see Fig. 4).

In the case (4.1) there is a double infinity of orbits starting at the sphere and tending to (0,2,0) alongside the vectors e_2; in the case (4.2) we find a double infinity of orbits tending to (0,2,0) along the plane (X, Y); and in the case (4.3) we have a double infinity of orbits tending to (0,2,0) along the vector e_1. These are all the orbits within the physical region tending to the critical point (0,2,0).

The point $(0, -2, 0)$ is also a node with characteristic roots $\lambda = 3\gamma - 6$ and $\mu = \nu = -6$, all negative for $1 \leqslant \gamma < 2$. The vector $e_1 = (1,1,0)$ corresponds to λ, and there is an infinity of vectors for $\mu = \nu$ in the plane (β', z). A double infinity of orbits tends to $(0, -2, 0)$ alongside the vector e_1.

The singular point (1,0,0) is a saddle point with $\lambda = -3\gamma + 2, \mu = 3 - \frac{3}{2}\gamma, \nu = -3\gamma$ and with corresponding vectors $e_1 = (1, -2/(1 + \frac{3}{2}\gamma), 0)$, $e_2 = (0,1,0)$, $e_3 = (1,0, -3/\Lambda)$. We find a simple infinity of orbits tending to (1,0,0) along the vector e_1. All these results are valid for $\Lambda > 0$ and $\Lambda < 0$. When $\gamma = 2$ there is a continuous line of singular points in the plane (x, β'): $(4 - 4x - \beta'^2 = 0, z = 0)$. In this case we have a simple infinity of orbits tending to each singular point along a characteristic vector in this plane.

In the plane (β', z) we have the singular point $(0,0,3/\Lambda)$ which is a node. The characteristic roots $\lambda = 3\gamma, \mu = 3$, and $\nu = 2$ have the corresponding vectors $e_1 = (1,0, -3/\Lambda)$, $e_2 = (0,1,0)$, and $e_3 = (0, -1, 3/2\Lambda)$ with $\Lambda > 0$ and $1 \leqslant \gamma \leqslant 2$. There is a double infinity of orbits tending to $(0,0,3/\Lambda)$ along the vector e_3.

Finally we have the singular point $(0, -2, 9/\Lambda)$ which is a saddle point with $\lambda = 3\gamma, \mu = -3$, and $\nu = 6$. The corresponding vectors are

TABLE III. Nonempty Kantowski–Sachs models with $\Lambda > 0$ and $\Lambda < 0$. We indicate the different singularity types as well as the asymptotic behavior of physically relevant variables. Comparing with Table II, we add the fluid density μ.

$p=(\gamma-1)\mu$ / Λ	(x,β',z)	Singularity type	$l=e^{-\Omega}$	Cosmological singularity	X	Y	μ	θ	σ	${}^3R=2/Y^2$	β	Dominant terms in Raychaudhuri equation
$1\leqslant\gamma<2$ $\Lambda>0$ $\Lambda<0$	(0,2,0)	cigar $X\to\infty$ $Y\to 0$	$(\pm t)^{+1/3}$	$\Omega\to\infty: t\to 0_{\pm}$	$(\pm t)^{-1/3}$	$(\pm t)^{2/3}$	$(\pm t)^{-\gamma}$	t^{-1}	$-t^{-1}$	$t^{-4/3}$	$-\ln\lvert t\rvert$	$\sigma^2,\theta^2\sim t^{-2}$
	$(0,-2,0)$	pancake $X\to 0$ $Y\to$ cte	$(\pm t)^{1/3}$	$\Omega\to\infty: t\to 0_{\pm}$	$\pm t$	cte	$(\pm t)^{-\gamma}$	t^{-1}	t^{-1}	cte	$\ln\lvert t\rvert$	$\sigma^2,\theta^2\sim t^{-2}$
	(1,0,0)	point $X\to 0$ $Y\to 0$	$(\pm t)^{2/3\gamma}$	$\Omega\to\infty: t\to 0_{\pm}$	$(\pm t)^{2/3\gamma}$	$(\pm t)^{2/3\gamma}$	$(\pm t)^{-2}$	t^{-1}	$t^{(3\gamma-4)/3\gamma}$	$t^{-4/3\gamma}$	$t^{(2/3\gamma)(3\gamma-2)}$	$\mu,\theta^2\sim t^{-2}$
$\gamma=2$ $\Lambda>0$ $\Lambda<0$	(x,β',z) such as $-2<\beta'<2$ $x=1-\frac{\beta'^2}{4}$ $z=0$	$-2<\beta<1$: point $X\to 0$ $Y\to 0$ $\beta'=1$: barrel $X\to$ cte $Y\to 0$ $1<\beta'<2$: cigar $X\to\infty$ $Y\to 0$	$(\pm t)^{1/3}$	$\Omega\to\infty: t\to 0_{\pm}$	$(\pm t)^{(1-\beta_0')/3}$	$(\pm t)^{(2+\beta_0')/6}$	$(\pm t)^{-2}$	t^{-1}	t^{-1} $(\beta_0'\neq 0)$ $t^{1/3}$ $(\beta_0'=0)$	$t^{-(2+\beta_0')/3}$	$\ln\lvert t\rvert$ $t^{4/3}(\beta_0'\neq 0))$ $t^{4/3}(\beta_0'=0)$	$\mu,\theta^2,\sigma^2\sim t^{-2}$
$1\leqslant\gamma<2$ $\Lambda>0$	$(0,0,\frac{3}{\Lambda})$	infinity $X\to\infty$ $Y\to\infty$	$e^{\mp\sqrt{\Lambda/3}t}$	$\Omega\to-\infty: t\to{}^{-}_{+}\infty$	$e^{\mp\sqrt{\Lambda/3}t}$	$e^{\mp\sqrt{\Lambda/3}t}$	$e^{\mp\sqrt{3\Lambda}\gamma t}$	$\mp\sqrt{3\Lambda}$	$e^{\pm 2\sqrt{\Lambda/3}t}$	$e^{\pm 2\sqrt{\Lambda/3}t}$	$e^{\pm 2\sqrt{\Lambda/3}t}$	Λ and $\theta^2=3\Lambda$
	$(0,-2,\frac{9}{\Lambda})$	infinite barrel $X\to\infty$ $Y\to$ cte	$e^{\mp\sqrt{\Lambda/3}t}$	$\Omega\to-\infty: t\to{}^{-}_{+}\infty$	$e^{\mp\sqrt{\Lambda}t}$	cte	$e^{\pm\sqrt{\Lambda}\gamma t}$	$\mp\sqrt{\Lambda}$	$\mp\sqrt{\Lambda}$	cte	$\pm 2\sqrt{\Lambda/3}t$	$\sigma^2,\theta^2,\Lambda$

$$e_1 = \left(\frac{(1+\gamma)(2-\gamma)}{\gamma[12/\Lambda - (3/\Lambda)(3\gamma-2)(1+\gamma)/\gamma]}, \frac{2-\gamma}{12/\Lambda - (3/\Lambda)(3\gamma-2)(1+\gamma)/\gamma}, -1 \right),$$

$e_2 = (0, -\Lambda/4, 1)$ and $e_3 = (0,0,1)$ with $\Lambda > 0$ and $1 \leqslant \gamma \leqslant 2$. There is a simple infinity of orbits tending to $(0, -2,9/\Lambda)$ along the vector e_1 for $1 \leqslant \gamma < 2$. When $\gamma = 2$, there is only one orbit tending to $(0, -2,9/\Lambda)$ in every direction of the plane (e_1, e_3).

We generalize to three dimensions the Poincaré transformations used in Sec. III for the study of the critical points at infinity. Setting $x = s^{-1}$, $\beta' = us^{-1}$, $z = vs^{-1}$ we find the double singular points at infinity ($s = 0, u = 0, v \geqslant 0$) which are not in the plane (β', z) (see Fig. 5). The general expression of the directions of approach (see the Appendix), not in the plane ($s = 0$) of the singular points ($s = 0, u = 0, v = v_0$) is given by

$$\{\omega_i\} = \left\{ \frac{1}{3\gamma - 2 - (2\Lambda/3)v_0}, \frac{2(2 + (2\Lambda/3)v_0)}{(3\gamma - 2 - (2\Lambda/3)v_0)^2}, \frac{-3\gamma v_0}{(3\gamma - 2 - (2\Lambda/3)v_0)^2} \right\}.$$

Notice that $\tan\phi = \omega_1/\omega_2$ is $(3\gamma - 2)/4$ when $v_0 = 0$ and $-\frac{1}{2}$ when $v_0 \to \infty$, which corresponds to the two plane autonomous systems. We have $\tan\phi = 0$, when $v_0 = \frac{3}{2}\Lambda$ with $\Lambda > 0$ and $\tan\phi \to \infty$ when $v_0 = -3/\Lambda$ with $\Lambda < 0$. The regions of physical interest for positive and negative values of the cosmological constant are indicated in Fig. 5.

By analyzing the three surfaces $\{dx/d\Omega = 0\}$, $\{d\beta'/d\Omega = 0\}$, $\{dz/d\Omega = 0\}$, we obtain a global picture of the orbits. We distinguish four different cases, as we have seen in Sec. II. The first case divides into three subcases according to the value of γ.

When $\Lambda > 0$ and $1 \leqslant \gamma < 2$ we have a double infinity of orbits starting at (0,2,0) and tending to $(0, -2,0)$, becoming tangent to the singular line at infinity with $\tan\phi = 0$. There is a simple infinity which tends to the saddle point (1,0,0). Likewise we had a time symmetric orbit in the plane (x, β) we have now a time symmetric surface of orbits, approaching the singular points at infinity with $\phi = \arctan \omega_1/\omega_2$, as long as $v_0 \leqslant \frac{3}{2}\Lambda$. There is further a double infinity of orbits starting at (0,2,0) extending to infinity with $\phi = \pi$ and coming back to the same critical point. From the points $(0, \pm 2, 0)$ and (1,0,0) orbits start and tend to $(0,0,3/\Lambda)$ as a double infinity with x, β', z being finite for the whole evolution. There is finally a simple infinity tending to $(0, -2,9/\Lambda)$. When $\gamma = 2$ we have a simple infinity of orbits coming from each singular point in the plane (x, β') and tending either to the singular line at infinity or to $(0,0,3/\Lambda)$. From $(0, -2,9/\Lambda)$ comes a simple infinity for every characteristic direction which tends to the singular line at infinity. For $\Lambda < 0$ and $1 \leqslant \gamma \leqslant 2$, we have a global picture similar to the plane case when $\Lambda = 0$. The asymptotic behavior of the models around the singular points at finite distance is indicated in Table III.

V. CONCLUSION

We have carried out a detailed analysis of Kantowski–Sachs models in the presence of a cosmological constant. Two new types of singularity points have been found, namely, an infinite singularity (from the metric point of view and as an autonomous system) and an infinite barrel.

One should notice, in particular, the models tending to $(0,0,3/\Lambda)$ when $\Lambda > 0$. Their asymptotic behavior (see Table III) gives the same length scales $X = Y = \exp(\mp\sqrt{\Lambda/3}\,t)$ tending to infinity when $t \to \mp\infty$, whence these models become isotropic in an infinite cosmological time. There is a set of nonzero measure of such models when they come from (0,2,0) or $(0, -2,0)$, and a set of zero measure for those models coming from (1,0,0). The point $(0,0,3/\Lambda)$ is an infinite singularity (as defined above) and the Raychaudhuri equation (2.5) tells us that at this point the cosmological constant Λ and the expansion θ are dominant whereas the shear σ and the density of matter μ are negligible. The time variable Ω goes to $-\infty$ (see Appendix) and Eq. (2.9) can be transformed in order to give the cosmic time dependence of Ω, i.e., $\dot{\Omega} = \pm\sqrt{\Lambda/3}$. The average length scale $l = (XY^2)^{1/3} = \exp(-\Omega) = \exp(\mp\sqrt{\Lambda/3}\,t)$ shows then that for $\Omega \to -\infty$ we have $l \to \infty$ and $t \to \mp\infty$: the point $(0,0,3/\Lambda)$ is not a cosmological singularity, because it takes an infinite cosmic time to get there.

This class of Kantowski–Sachs models becoming isotropic in the presence of a cosmological constant motivates a further study of three-dimensional systems in order to obtain more general results, in particular for the Bianchi cosmological models.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor D. Speiser for stimulating this work,[15] as well as J. Demaret and M. Goethals for helpful discussions.

APPENDIX: THEOREMS FOR THREE-DIMENSIONAL AUTONOMOUS SYSTEMS[10–12]

Consider a three-dimensional autonomous system. The definitions for the simple and multiple singular points are the same as for a plane one. When two of the three real characteristic roots λ, μ, ν in the case of a simple singular point have the same sign and the other one the opposite sign, we have a saddle point. If the three real characteristic roots have the same sign, we have a node.

We analyze first a saddle point with the autonomous system in the form

$$X' = \lambda X + f, \quad Y' = \mu Y + g, \quad Z' = \nu Z + h, \qquad \text{(A1)}$$

where f, g, h are the nonlinear terms of the system and where the singular point is at the origin of the coordinates (X,Y,Z). The sign of ν is denoted by $\sigma(\nu)$. We indicate only those theorems which are needed in this article. θ, ϕ intervening in these theorems corresponds to the angles in spherical coordinates ($X = r\sin\theta\cos\phi$, $Y = r\sin\theta\sin\phi$, $Z = r\cos\theta$).

Theorem 1: The autonomous system (A1) with $\lambda > 0$, $\mu > 0$, $\nu < 0$ has (a) only one orbit tending to the origin with $\sigma(\nu)\,\Omega \to -\infty$ and $\theta_\infty \equiv \lim_{\sigma\Omega \to -\infty} \theta = 0$ and only one orbit with $\theta_\infty = \pi$; and (b) an orbit γ satisfying the following properties: γ and its projection on the plane (X,Y) are homeomor-

phic to a circle, and every orbit starting at one point of γ is tending to the origin with $\sigma(\nu)\,\Omega \to +\infty$ and $\lim_{\sigma\Omega \to +\infty} \theta = \pi/2$.

Theorem 2: The autonomous system (A1) with $\lambda \neq \mu$ and $|\mu| < |\lambda|$ has an infinity of orbits starting from γ with $\phi_\infty = -\pi/2$. There is only one with $\phi_\infty = 0$ and only one with $\phi_\infty = \pi$.

Theorem 3: The autonomous system (A1) with $\lambda = \mu$ and with $f,g = O(r^{1+\epsilon})$ has for every $\phi_0 \in [0,2\pi]$ only one orbit with $\phi_\infty = \phi_0$.

We consider now a node for the same autonomous system.

Theorem 4: If the characteristic roots are all of the same sign, there is a sphere centered at the origin such that every orbit starting at its surface tends to the singular point with $\sigma\Omega \to -\infty$.

Theorem 5: If the characteristic roots are such that $|\nu| < |\mu| \leqslant |\lambda|$, then every orbit which tends to the origin does go alongside the positive Z-axis or the negative one, except those orbits which start at an orbit γ which is homeomorphic to a circle, as well as its projection on the plane (X,Y); in this case $\theta_\infty = \pi/2$.

Theorem 6: If the characteristic roots are such that $|\nu| < |\mu| < |\lambda|$, then every orbit starting at γ tends to the origin with $\phi_\infty = \pm\pi/2$, except only one with $\phi_\infty = 0$, and only one with $\phi_\infty = \pi$.

Theorem 7: If the characteristic roots are such that $|\nu| < |\lambda| = |\mu|$, then for every $\phi_0 \in]0,2\pi]$, there exists only one orbit starting at γ, tending to the origin with $\phi_\infty = \phi_0$, when $f,g = O(r^{1+\epsilon})$ when $r \to 0$.

When we have a multiple singular point for the autonomous system

$$\frac{dx_i}{d\Omega} = f_i(x_1, x_2, x_3), \quad i = 1,2,3, \tag{A2}$$

we define the direction of approach to the singular point (at the origin) as follows.

Definition: Let $\{x_i(\Omega);\ 1 \leqslant i \leqslant 3\}$ be an orbit of (A2) and l defined by $l^2 = \Sigma_i (dx_i/d\Omega)^2$. When $\{x_i(\Omega)\}$ tends to the origin for $\Omega \to \infty$ and when $\lim_{\Omega \to \infty} x_i(\Omega)/l = \omega_i$ with $\{\omega_i\}$ a nonzero vector, then we say that $\{\omega_i\}$ is a direction of approach of the orbit to the origin.

If $l = O(r^n)$ when $r \to 0$, we say that the multiple singular point of order n has a regular principal part, and in this case, the ω_i are the roots of the equation

$$\frac{1}{n-1}\,\omega_i = \frac{\partial^n f_i}{\partial x_i^p \cdots \partial x_j^q}(0)\,\omega_i^p \cdots \omega_j^q.$$

[1] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[2] L. Bianchi, *Lezioni sulla teoria dei gruppi continui finiti transformazioni* (Spoerri, Pisa, 1918).
[3] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, NJ, 1975).
[4] R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).
[5] C. B. Collins, J. Math. Phys. **18**, 2116 (1977).
[6] C. B. Collins, Commun. Math. Phys. **23**, 137 (1971).
[7] C. B. Collins, Commun. Math. Phys. **27**, 37 (1972).
[8] C. B. Collins, Commun. Math. Phys. **39**, 131 (1974).
[9] C. B. Collins and S. W. Hawking, Astrophys. J. **180**, 317 (1973).
[10] M. Goethals, private communication.
[11] I. Bihari, Stud. Sci. Math. Hung. **1**, 261 (1966).
[12] G. Couper, J. Diff. Eq. **10**, 27 (1971).
[13] A. A. Andronov, E. A. Leontovich, I. I. Gordon, and A. G. Maier, *Qualitative Theory of Second-order Dynamic Systems* (Wiley, New York, 1973).
[14] V. V. Nemytskii and V. V. Stepanov, *Qualitative Theory of Differential Equations* (Princeton U.P., Princeton, NJ, 1960).
[15] E. Weber, Ph.D. thesis, Université Catholique de Louvain, Louvain-la-neuve, 1981.

Statistical mechanics of the gravitational field in a conformally static setting

Henry E. Kandrup
Center for Studies in Statistical Mechanics and Center for Relativity, University of Texas, Austin, Texas 78712

(Received 2 February 1984; accepted for publication 20 April 1984)

This paper formulates a statistical description of a collection of N identical classical particles that interact relativistically via linear fields in a fixed background space-time that admits a conformal timelike Killing field. Attention focuses upon the special cases of a simple scalar interaction and a linearized gravitation interaction which should suffice to model many systems of astrophysical interest. The fundamental object of the theory is a complicated distribution function that depends upon appropriate variables for both the particles and the fields. By assuming that, in a first approximation, this distribution factorizes into an infinite product of reduced distribution functions, one recovers the type of mean-field theory developed by such authors as Ipser and Thorne. Alternatively, one may derive various exact and approximate relations which contain information about the interparticle correlations.

PACS numbers: 04.20.Cv, 05.20.Gg, 98.80.Dr

I. INTRODUCTION

Recently, Israel and Kandrup have developed a new, manifestly covariant approach to nonequilibrium statistical mechanics in classical general relativity which could be applied to the study of such problems as galaxy clustering or the dynamics of a collection of stars.[1-4] Although the formalism which they constructed is rather complicated in its details, the basic physical ingredients may be stated very simply.

(i) Following the viewpoint developed by Hakim,[5,6] it is argued that a collection of N gravitationally interacting particles may be characterized by an N-particle distribution function, defined in an $8N$-dimensional phase space, which satisfies a collection of N conservation equations.

(ii) By mapping the "true" physics—particles following geodesics in the "true" space-time manifold—onto a fictitious "background" space-time, which may be chosen to satisfy some "average" field equations, one then obtains a useful covariant notion of "evolution" in response to a "fluctuating gravitational force."

(iii) It is assumed that the deviations between the true and background space-times are in some sense small, so that they may be described by linear field equations.

(iv) It is assumed further that these gravitational forces, which derive from linear field equations, may be modeled by a direct interaction that involves only the coordinates and momenta of the various particles (one assumes, therefore, that incoherent radiative effects may be neglected).

Given these four premises, it is straightforward to define a statistical mechanics for the system, and, in particular, to formulate various exact equations for appropriately defined reduced distribution functions. Thus, for example, by introducing a preferred time coordinate, one can derive an exact closed equation for the evolution of the one-particle distribution function very much analogous to the sort of relation that arises in a Newtonian theory.[7] By implementing, in a suitable fashion, a relativistic analog of an "impulse" or "dilute gas" approximation, one is then led to a covariant analog of the standard Landau (or Fokker–Planck) equation.[1,4]

This approach, albeit a legitimate one, has the obvious disadvantage of failing to include from the outset explicit reference to the degrees of freedom of the gravitational field. The objective of ongoing research is, therefore, the development of a more general, "field theoretic" approach which will embrace, in a natural way, these gravitational degrees of freedom. This can, for example, be done by introducing generalized coordinates and momenta for the fields, and by defining as the basic object of the theory a more complicated distribution function involving both particle and field variables.

Given the obvious desire for the formulation of a covariant statistical description, it is of course natural to approach this general problem in terms of the theory of constrained Hamiltonian dynamics,[8] as applied to general relativity by such authors as Kuchař[9] or Arnowitt, Deser, and Misner.[10] This can (at least in principle) and should be done in a completely general context. It should, however, be clear that the problem simplifies enormously if the underlying space-time admits a preferred time coordinate, e.g., if one is concerned with astrophysical processes in the context of a Friedmann cosmology or if one is concerned with a cluster of stars that is nearly static. In this case, it would not seem all that unreasonable to break manifest covariance by implementing the obvious $3+1$ decomposition, and, indeed, one obtains thereby, equations of motion that are more easily tractable.

For this reason, this paper is devoted to the general problem of formulating a statistical description of a many-particle system interacting via linear fields in a conformally static background space-time, i.e., for a space-time that admits a conformal timelike Killing field. Attention will focus upon (i) the linearized gravitational theory introduced by Israel and Kandrup[1,4] and (ii) the simpler example of a scalar field which has been studied in a special relativistic context by such authors as Hakim[5,6] and Kandrup.[2] The case of the scalar field can, and will, be described without great difficulties for an arbitrary conformally static space-time. The gravitational interaction is made more complicated by the tensorial character of the field and, therefore, for simplicity, attention will be restricted to the special case of a $k=0$ spatially flat Friedmann cosmology. A number of useful results

 0022-2488/84/113286-11$02.50

appropriate for gravitational interactions in a static, spherically symmetric space-time are implicit in the work of Ipser and Thorne[11] and Kandrup.[3]

At this stage, it is convenient to record the appropriate equations of motion for the gravitationally interacting system in a form that is manifestly covariant. Let $g'_{\alpha\beta}$ denote the metric appropriate for the "true" space-time, in which the particles follow geodesics, and let $g_{\alpha\beta}$ denote the metric for the "background" space-time to which these trajectories are to be referred. The particle equations then take the forms[1]

$$\frac{dx^\alpha}{d\tau}=\frac{P^\alpha}{m},$$

and

$$\frac{dP_\alpha}{d\tau}=\frac{1}{m}\Gamma^\lambda_{\alpha\mu}P_\lambda P^\mu-\frac{1}{m}\Delta_{\alpha\lambda}\delta\Gamma^\lambda_{\mu\nu}P^\mu P^\nu$$
$$(\alpha,\beta,\ldots=0,1,2,3), \tag{1.1}$$

where τ denotes the proper time of the particle (for the metric $g_{\alpha\beta}$), $\Gamma^\lambda_{\alpha\mu}$ and $\Gamma'^\lambda_{\alpha\mu}$ denote, respectively, the Christoffel symbols for the background and true space-times,

$$\delta\Gamma^\lambda_{\mu\nu}(x^\alpha)\equiv\Gamma'^\lambda_{\mu\nu}(x^\alpha)-\Gamma^\lambda_{\mu\nu}(x^\alpha), \tag{1.2}$$

and the quantities P^α, m, and $\Delta_{\alpha\beta}$ are to be viewed as functions of the covariant momentum P_α and the inverse metric $g^{\alpha\beta}$:

$$P^\alpha\equiv g^{\alpha\beta}P_\beta,\quad m\equiv(-g^{\mu\nu}P_\mu P_\nu)^{1/2},$$
$$\text{and}\ \Delta_{\alpha\beta}=g_{\alpha\beta}+m^{-2}P_\alpha P_\beta. \tag{1.3}$$

These equations are derived from the equations of motion for a free particle in the true space-time by identifying space-time points but rescaling momenta so as to preserve the mass shell constraints and cross sections of the tangent bundle. The quantity $\delta\Gamma^\lambda_{\mu\nu}$ will of course transform as a tensor and, therefore, it is clear that the "gravitational force" possesses an invariant geometric meaning.

Granted that the difference between $g'_{\alpha\beta}$ and $g_{\alpha\beta}$ may be treated as small, all the fields of interest will depend linearly upon the quantity

$$h_{\alpha\beta}(x^\mu)\equiv g'_{\alpha\beta}(x^\mu)-g_{\alpha\beta}(x^\mu). \tag{1.4}$$

Thus, in the context of the linearized theory,

$$\delta\Gamma^\lambda_{\mu\nu}=\nabla_{(\mu}h_{\nu)}{}^\lambda-\tfrac{1}{2}\nabla^\lambda h_{\mu\nu}, \tag{1.5}$$

$h_{\mu\nu}$ being derived as a solution to the linearized field equation

$$\delta G_\alpha{}^\beta[h]=8\pi\delta T_\alpha{}^\beta. \tag{1.6}$$

Here $\delta G_\alpha{}^\beta[h]$, the perturbed Einstein tensor constructed from $h_{\mu\nu}$, takes the form[12]

$$\delta G_\alpha{}^\beta=\tfrac{1}{2}\nabla_\mu\nabla^\beta h_\alpha{}^\mu+\tfrac{1}{2}\nabla_\mu\nabla_\alpha h^{\beta\mu}-\tfrac{1}{2}\nabla^\mu\nabla_\mu h_\alpha{}^\beta$$
$$-\tfrac{1}{2}\nabla^\beta\nabla_\alpha h_\mu{}^\mu-h^{\mu\beta}R_{\mu\alpha}-\tfrac{1}{2}\delta_\alpha{}^\beta(\nabla^\mu\nabla_\nu h_\mu{}^\nu$$
$$-\nabla^\mu\nabla_\mu h_\nu{}^\nu-h^{\mu\nu}R_{\mu\nu}), \tag{1.7}$$

where $R_{\mu\nu}$ and ∇_μ denote, respectively, the Ricci curvature and the covariant derivative operator associated with $g_{\alpha\beta}$, and the quantity

$$\delta T_\alpha{}^\beta\equiv T_\alpha{}^\beta-[T_\alpha{}^\beta]_B \tag{1.8}$$

denotes the difference between the stress-energy sources for the true and background space-times.

The equations of motion for the scalar interaction to be considered are very similar. In this case, the particle equations take the form[2]

$$\frac{dx^\alpha}{d\tau}=\frac{P^\alpha}{m}$$

and

$$\frac{dP_\alpha}{d\tau}=\frac{1}{m}\Gamma^\lambda_{\alpha\mu}P_\lambda P^\mu-\lambda\Delta_{\alpha\lambda}F^\lambda, \tag{1.9}$$

where λ is a coupling constant, and the "force" F^α is constructed as the gradient of the scalar field $\Phi(x^\mu)$:

$$F^\alpha(x^\mu)=\nabla^\alpha\Phi(x^\mu). \tag{1.10}$$

This Φ is in turn assumed to satisfy an inhomogeneous wave equation of the form

$$\nabla_\alpha\nabla^\alpha\Phi+nR\Phi=4\pi\lambda\mu=4\pi(\rho/m), \tag{1.11}$$

where $\rho=\mu m$ is the mass density, R is the scalar curvature, and n is an arbitrary numerical constant. The case $n=0$ corresponds to the minimally coupled field. The case $n=-\frac{1}{6}$ is conformally invariant.

The program of this paper is as follows. Section II formulates the basic equations for the scalar interaction in an arbitrary conformally static background space-time in a fashion that exploits the natural 3 + 1 decomposition. Section III then constructs a statistical mechanics appropriate for the interacting system. A particle-field distribution function is defined as a probability density in a suitably constructed infinite-dimensional phase space, an infinite-dimensional Liouville equation is formulated, and it is seen that, in the most naive possible approximation, one recovers a relativistic analog of the ordinary self-consistent field approximation. Section IV examines various properties of interest for the one-particle mean-field description. Section V then considers an analogous description for a gravitationally interacting system in a $k=0$ Friedmann cosmology.

A final summary of notation is in order. Unless noted otherwise, all conventions follow Misner *et al.*[12] This implies a metric with signature $(-,+,+,+)$. Greek letters $\alpha,\beta,\ldots$ label space-time indices 0,1,2,3, whereas lower case Latin letters $a, b,\ldots$ label spatial indices 1, 2,3. Capital letters $A,B,\ldots$ label the field oscillators introduced in Sec. II. Lower case letters $i,j,\ldots$ label individual particles. The quantities t and η denote, respectively, the "true" and "conformal" times, $\Omega(\eta)$ is the conformal factor, so that $dt=\Omega\,d\eta$, and $\gamma_{\mu\nu}(x^a)=\Omega^{-2}(\eta)g_{\mu\nu}(x^\alpha)$ denotes the conformally static metric.

II. A SIMPLE SCALAR FIELD

Granted that the space-time is conformally static, the line element can of course be taken to be of the form

$$ds^2=g_{\mu\nu}\,dx^\mu\,dx^\nu=\Omega^2(\eta)\gamma_{\mu\nu}(x^c)dx^\mu\,dx^\nu$$
$$=\Omega^2(\eta)(\gamma_{\eta\eta}d\eta^2+\gamma_{ab}\,dx^a\,dx^b), \tag{2.1}$$

where $\gamma_{ab}(x^c)$ denotes the time-independent spatial metric.

The starting point for the analysis is the total action S, for the system of particles and fields, which is assumed to be of the form

$$\begin{aligned} S = & -\sum_{i=1}^{N}\int d\tau_i\,[m+\lambda\Phi(x)] \\ & -\int d^4x(-g)^{1/2}\frac{1}{8\pi}(\nabla_\mu\Phi\nabla^\mu\Phi+nR\Phi^2) \\ = & -\sum_i m\int d\tau_i-\int d^4x(-g)^{1/2} \\ & \times\left[\frac{\lambda}{m}\rho\Phi+\frac{1}{8\pi}(\nabla_\mu\Phi\nabla^\mu\Phi+nR\Phi^2)\right]. \end{aligned} \tag{2.2}$$

Here m denotes the mass of the N identical particles, λ is a coupling constant associated with the scalar field Φ, τ_i denotes the proper time of the ith particle, and, in analogy with the electromagnetic interaction, the scalar density ρ may be written as[13]

$$\rho(x^a,\eta)=m\sum_i\frac{d\tau_i}{d\eta}(-g)^{-1/2}\delta^{(3)}[x^a-x_i{}^a(\eta)], \tag{2.3}$$

where $\delta^{(3)}$ denotes a three-dimensional Dirac delta distribution.

Given this action and the space-time metric $g_{\mu\nu}$ of Eq. (2.1), the Lagrangian for any given particle is of course

$$\begin{aligned} L_p &= -(m+\lambda\Phi)(-g_{\eta\eta}-g_{ab}v^av^b)^{1/2} \\ &= -(m+\lambda\Phi)\Omega(-\gamma_{\eta\eta}-\gamma_{ab}v^av^b)^{1/2}, \end{aligned} \tag{2.4}$$

where $v^a\equiv dx^a/d\eta$ denotes the ordinary three-velocity. Associated with this Lagrangian is the canonical three-momentum

$$\pi_a\equiv\gamma_{ab}\pi^b=\frac{(m+\lambda\Phi)\Omega\gamma_{ab}v^b}{(-\gamma_{\eta\eta}-\gamma_{cd}v^cv^d)^{1/2}}. \tag{2.5}$$

By inverting Eq. (2.5), one finds that

$$v^a\equiv\gamma^{ab}v_b=\frac{(-\gamma_{\eta\eta})^{1/2}\gamma^{ab}\pi_b}{[(m+\lambda\Phi)^2\Omega^2+\gamma^{cd}\pi_c\pi_d]^{1/2}}, \tag{2.6}$$

where γ^{ab} is defined so that $\gamma^{ab}\gamma_{bc}=\delta^a{}_c$. It will be observed that, in Eqs. (2.5) and (2.6), spatial indices are raised and lowered with the conformal metric γ_{ab}.

Given these relations, one can immediately identify the particle Hamiltonian

$$\begin{aligned} H &= (-\gamma_{\eta\eta})^{1/2}[(m+\lambda\Phi)^2\Omega^2+\gamma^{ab}\pi_a\pi_b]^{1/2} \\ &= \frac{-(m+\lambda\Phi)\Omega\gamma_{\eta\eta}}{(-\gamma_{\eta\eta}-\gamma_{ab}v^av^b)^{1/2}}\equiv-\pi_\eta. \end{aligned} \tag{2.7}$$

Here the introduction of the notation $H=-\pi_\eta$ is to be viewed simply as a convenient definition, to assist with the summation convention. As, e.g., in the paper of Balescu and Kotera,[14] it is not to be assumed that π_η transforms as the component of a four-vector. One may also observe for future reference that

$$\frac{d\tau}{d\eta}=\frac{-\Omega^2\gamma_{\eta\eta}(m+\lambda\Phi)}{\pi_\eta}=\Omega(-\gamma_{\eta\eta}-\gamma_{ab}v^av^b)^{1/2}. \tag{2.8}$$

It follows at once from the Hamiltonian (2.7) that the particle equations of motion take the forms

$$\frac{dx^a}{d\eta}=\frac{d\tau}{d\eta}\frac{1}{\Omega^2(m+\lambda\Phi)}\gamma^{ab}\pi_b$$

and

$$\frac{d\pi_a}{d\eta}=\frac{d\tau}{d\eta}\left[\frac{1}{2\Omega^2(m+\lambda\Phi)}\pi_\mu\pi_\nu\,\partial_a\gamma^{\mu\nu}-\lambda\,\partial_a\Phi\right], \tag{2.9}$$

where $\partial_a\equiv\partial/\partial x^a$ denotes an ordinary partial derivative.

Related to π_μ is the "physical" momentum p_μ defined by the relation

$$p_\mu\equiv(1+\lambda\Phi/m)^{-1}\pi_\mu. \tag{2.10}$$

This p_μ coincides with the P_μ entering into the covariant equation (1.9). If, however, one adheres to the convention of raising and lowering indices with the conformal metric, it follows that $p^\mu\equiv\gamma^{\mu\nu}p_\nu=\Omega^2P^\mu$. In any case, by re-expressing the particle equations of motion in terms of p_μ, one finds that

$$\frac{dp_a}{d\eta}=\frac{d\tau}{d\eta}\left[\frac{-1}{2m\Omega^2}p_\mu p_\nu\,\partial_a\gamma^{\mu\nu}-\lambda\left(1+\frac{\lambda\Phi}{m}\right)^{-1}\Delta_a{}^\mu\,\partial_\mu\Phi\right], \tag{2.11}$$

where, now,

$$\Delta_a{}^\mu\equiv\delta_a{}^\mu+(1/\Omega^2m^2)p_ap^\mu \tag{2.12}$$

is the spatial projection tensor viewed as a function of p_μ. Because of the factor $(1+\lambda\Phi/m)^{-1}$, Eq. (2.11) is nonlinear and, therefore, will be difficult to analyze. However, in the limit that $|\lambda\Phi/m|$ is small, that factor may of course be neglected, and, in that case, the equations of motion for the ith particle assume the form

$$\frac{dx_i{}^a}{d\eta}=\frac{d\tau_i}{d\eta}\frac{1}{\Omega^2m}\gamma^{ab}(i)p^i{}_b$$

and

$$\frac{dp^i{}_a}{d\eta}=\frac{d\tau_i}{d\eta}\left[\frac{-1}{2m\Omega^2}p^i{}_\mu p^i{}_\nu\,\partial^i{}_a\gamma^{\mu\nu}(i)-\lambda\Delta_a{}^\mu(i)\partial^i{}_\mu\Phi(i)\right]. \tag{2.13}$$

Equations (2.13) will be taken as the starting point for all subsequent discussion.

By varying the action (2.2) with respect to the field configuration, one obtains the field equation

$$\nabla_\mu\nabla^\mu\Phi+nR[g]\Phi=4\pi(\lambda/m)\rho(x^a,\eta), \tag{2.14}$$

where ∇_μ and $R[g]$ denote, respectively, the covariant derivative operator and scalar curvature associated with $g_{\mu\nu}$. For the $g_{\mu\nu}$ of Eq. (2.1), one finds explicitly that

$$\begin{aligned} &\gamma^{\eta\eta}\partial_\eta{}^2\Phi+\gamma^{\eta\eta}(2\Omega'/\Omega)\partial_\eta\Phi+(-\gamma)^{-1/2}\partial_a(-\gamma)^{1/2} \\ &\quad\times\gamma^{ab}\partial_b\Phi+nR[g]\Omega^2\Phi=4\pi(\lambda/m)\rho\Omega^2, \end{aligned} \tag{2.15}$$

where now a prime $'$ denotes differentiation with respect to η. The scalar curvature $R[g]$ is related to the $R[\gamma]$ appropriate for $\gamma_{\mu\nu}$ by the simple formula[15]

$$\begin{aligned} R[g] &= \Omega^{-2}R[\gamma]-6\Omega^{-3}D_\mu D^\mu\Omega \\ &= \Omega^{-2}R[\gamma]-6\Omega^{-2}\gamma^{\eta\eta}(\Omega''/\Omega), \end{aligned} \tag{2.16}$$

where D_μ is the covariant derivative operator associated with $\gamma_{\mu\nu}$, and thus, one sees immediately that

$$\gamma^{\eta\eta}\left[\partial_\eta{}^2\Phi + \frac{2\Omega'}{\Omega}\partial_\eta\Phi - 6n\frac{\Omega''}{\Omega}\Phi - \Delta\Phi\right] = 4\pi\frac{\lambda}{m}\rho\Omega^2, \tag{2.17}$$

where

$$\Delta \equiv (-\gamma^{\eta\eta})^{-1}\{(-\gamma)^{-1/2}\partial_a(-\gamma)^{1/2}\gamma^{ab}\partial_b + nR[\gamma]\} \tag{2.18}$$

denotes a convenient generalization of the flat space Laplacian. In terms of the rescaled field

$$\chi = \Omega\Phi \tag{2.19}$$

Eq. (2.17) takes the form

$$\partial_\eta{}^2\chi - (1+6n)(\Omega''/\Omega)\chi - \Delta\chi = -4\pi(\lambda/m)\rho\Omega^3(-\gamma^{\eta\eta})^{-1}. \tag{2.20}$$

For the special case when $n = -\frac{1}{6}$, one finds that

$$D_\mu D^\mu\chi + nR[\gamma]\chi = 4\pi(\lambda/m)\rho\Omega^3. \tag{2.21}$$

At this stage, it is convenient to re-express the rescaled χ as a superposition of appropriately defined field oscillators.[16] The first step in the procedure is to view Eq. (2.20) as an operator equation on a Hilbert space with an inner product

$$(\xi,\zeta) \equiv \int d^3x(-\gamma)^{1/2}(-\gamma^{\eta\eta})\xi(x^a)\zeta(x^a) \tag{2.22}$$

(the form of the inner product is all that is relevant here; domain issues, completeness, etc., will be neglected in what follows). The important point, then, is that if one can integrate by parts and neglect all surface terms, the operator Δ will be symmetric:

$$(\xi,\Delta\zeta) = (\Delta\xi,\zeta). \tag{2.23}$$

If, for example, $\gamma_{\mu\nu}$ is flat, it is customary to ensure this by imposing periodic boundary conditions.

Granted that Δ is symmetric, it follows at once that its eigenvectors are orthogonal, and that its eigenvalues are real. And thus, if one assumes, in the usual way,[16] that the eigenvectors $\psi_A(x^a)$ form a complete set, $\chi(x^a,\eta)$ may be expressed as a linear combination of the ψ_A's:

$$\chi(x^a,\eta) = \sum_{A=1}^{\infty} q_A(\eta)\psi_A(x^a), \tag{2.24}$$

where

$$\Delta\psi_A + \omega_A{}^2\psi_A = 0. \tag{2.25}$$

Here the notation is predicated upon the assumption that the modes are in fact discrete. If this is not true, the summation in Eq. (2.24) must of course be replaced by a Stieltjes integral. Without loss of generality, one is free to impose the normalization

$$(\psi_A,\psi_B) = 4\pi\delta_{AB}, \tag{2.26}$$

and, given this normalization, the assumption of completeness implies that

$$\sum_A \psi_A(x^a)\psi_B(y^a) = 4\pi(-\gamma^{\eta\eta})^{-1}(-\gamma)^{-1/2}\delta^{(3)}(x^a - y^a). \tag{2.27}$$

It is now trivial to obtain equations of motion for the individual oscillators. All that one need do is substitute the expansion (2.24) into Eq. (2.20), take the inner product with ψ_B, and divide by 4π. In this fashion, one finds that

$$\begin{aligned} q_A'' - (1+6n)\frac{\Omega''}{\Omega}q_A + \omega_A{}^2 q_A &= -\frac{\lambda}{m}\int d^3x(-\gamma)^{1/2}(-\gamma^{\eta\eta})\rho\Omega^3\psi_A(x^a)/(-\gamma^{\eta\eta}) \\ &= -\frac{\lambda}{\Omega}\sum_i\int d^3x\frac{d\tau_i}{d\eta}\delta^{(3)}[x^a - x_i{}^a(\eta)]\psi_A(x^a) \\ &= -\frac{\lambda}{\Omega}\sum_i\frac{d\tau_i}{d\eta}\psi_A(x_i{}^a). \end{aligned} \tag{2.28}$$

Alternatively, the second-order Eq. (2.28) may be viewed as a coupled first-order system:

$$\frac{dq_A}{d\eta} = -p_A$$

and

$$\frac{dp_A}{d\eta} = \left[\omega_A{}^2 - (1+6n)\frac{\Omega''}{\Omega}\right]q_A + \frac{\lambda}{\Omega}\sum_i\frac{d\tau_i}{d\eta}\psi_A(i). \tag{2.29}$$

In the limit that Ω is a constant, Eqs. (2.29) are completely analogous to the equations for an electromagnetic field in flat space.

III. A STATISTICAL DESCRIPTION

The first step in the formulation of a statistical description is of course the construction of an appropriate phase space.

If one starts from the particle equations of motion in a form that is manifestly covariant, the "natural" configuration space for any given particle is the four-dimensional space-time manifold itself: this comes equipped with a volume element $(-g)^{1/2}d^4x$. Similarly, the momenta live naturally in the cotangent space, which comes equipped with the volume element $(-g)^{-1/2}d^4p$. The natural eight-dimensional one-particle solution space, the cotangent bundle associated with the space-time manifold, thus comes equipped with the volume element[1]

$$\begin{aligned} d^8V &= (-g)^{1/2}d^4x(-g)^{-1/2}d^4p = d^4x\,d^4p \\ &= d^3x\,d^3p\,d\eta\,dp_\eta. \end{aligned} \tag{3.1}$$

The mass shell constraint will of course restrict the physics to a particular seven-dimensional hypersurface in the cotangent bundle (so that p_η may be thought of as a function of the spatial p_a's) and, by focusing upon a given instant of time, i.e., by setting η = constant, one selects out a six-dimensional analog of the ordinary Newtonian phase space. This space comes equipped with the natural volume element

$$\begin{aligned} d^6V &= d^3x\,d^3p = (-g)^{1/2}d^3x(-g)^{-1/2}d^3p \\ &= (-\gamma)^{1/2}d^3x(-\gamma)^{-1/2}d^3p, \end{aligned} \tag{3.2}$$

which can of course be written in a form that is manifestly covariant[17]:

$$d^8V = d^6V\,d\eta\,dp_\eta = d^6V\,d\tau\,dm. \tag{3.3}$$

It should be clear that this d^6V is the natural element associated with the noncovariant equations (2.13).

The full $6N$-dimensional, N-particle phase space is constructed as the direct product of N identical copies of the

one-particle phase space, and comes equipped with the volume element

$$d\mathscr{V}_p = \prod_{i=1}^{N} d^6V_i = \prod_{i=1}^{N} d^3x_i\, d^3p^i. \tag{3.4}$$

The embedding of this space in the "natural" $8N$-dimensional covariant solution space has been considered by such authors as Israel and Kandrup[1] or Hakim.[5]

It is perhaps most natural to think of the oscillator phase space as being the cotangent bundle associated with an infinite-dimensional flat manifold for which the q_A's are the generalized coordinates (the structure of curved, infinite-dimensional manifolds is of course a very subtle affair). The volume element then takes the form

$$d\mathscr{V}_f = \prod_{A=1}^{\infty} dq_A\, dp_A. \tag{3.5}$$

The total particle-field phase space is constructed as a direct product of particle and field phase spaces, equipped with the volume element

$$d\Theta \equiv d\mathscr{V}_p\, d\mathscr{V}_f = \left(\prod_{i=1}^{N} d^3x_i\, d^3p^i\right)\left(\prod_{A=1}^{\infty} dq_A\, dp_A\right). \tag{3.6}$$

Given the introduction of the element $d\Theta$, one may define the distribution function μ, the basic object of the theory, by the statement that

$$d\mathscr{P} = \mu(x_1{}^a, p^1{}_a, \ldots, q_1, p_1, \ldots; \eta)\, d\Theta \tag{3.7}$$

represents the probability that, at some time η, each given particle i has coordinates and momenta centered about the values $x_i{}^a$ and $p^i{}_a$, and that each oscillator A has coordinates and momenta centered about q_A and p_A. Conservation of probability in the infinite-dimensional phase space, i.e., the invariance of $d\mathscr{P}$ under Lie transport along the trajectories defined by the equations of motion, then implies that

$$\frac{\partial \mu}{\partial \eta} + \sum_i \frac{\partial}{\partial x_i{}^a}\left(\frac{dx_i{}^a}{d\eta}\mu\right) + \sum_i \frac{\partial}{\partial p^i{}_a}\left(\frac{dp^i{}_a}{d\eta}\mu\right) + \sum_A \frac{\partial}{\partial q_A}\left(\frac{dq_A}{d\eta}\mu\right) + \sum_A \frac{\partial}{\partial p_A}\left(\frac{dp_A}{d\eta}\mu\right) = 0. \tag{3.8}$$

It follows at once from Eq. (3.8) that one can impose the normalization

$$\int \mu\, d\Theta = 1 \tag{3.9}$$

for all values of η.

Given the fundamental distribution function μ, one can of course define appropriate reduced distributions. Thus, for example, one may define the N-particle distribution

$$F(x_1{}^a, p^1{}_a, \ldots; \eta) \equiv \int \mu\, d\mathscr{V}_f \tag{3.10}$$

and the field distribution

$$G(q_1, p_1, \ldots; \eta) \equiv \int \mu\, d\mathscr{V}_p. \tag{3.11}$$

Of particular interest are the irreducible one-particle and field distributions, $f(i)$ and $g(A)$, obtained by integrating out all the degrees of freedom except for one of the particles or oscillators:

$$f(x_i{}^a, p^i{}_a; \eta) \equiv f(i) = \int F \prod_{j \neq i} d^6V_j \tag{3.12}$$

and

$$g(q_A, p_A; \eta) \equiv g(A) = \int G \prod_{B \neq A} dq_B\, dp_B. \tag{3.13}$$

These will of course satisfy the normalization

$$\int f(i) d^6V_i = 1 = \int g(A)\, dq_A\, dp_A. \tag{3.14}$$

At this stage, it should be straightforward to proceed in analogy with the standard Newtonian analyses to obtain various exact equations for the evolution of the reduced distribution functions. Thus, for example, by integrating over the degrees of freedom of some subset of the particles and the oscillators, one can obtain various relations among the reduced distributions, and, therefore, one might hope to obtain some useful analog of the standard BBGKY hierarchy of coupled equations.[18] Similarly, one might try to construct some analog of the projection operator techniques developed by Balescu, Prigogine, and their co-workers[18,19] (this will, however, be complicated by the facts that the "forces" will in general be explicitly time dependent, and that even a "free" particle will experience time-dependent effects). These sorts of issues will be considered in a later paper.

The object here is to demonstrate simply that, in the most naive possible limit, one can in fact recover a relativistic analog of the ordinary self-consistent field (or Vlasov) approximation.[7,18] This approximation amounts simply to the assumption that, to lowest order, the full distribution μ may be taken as an infinite product of irreducible f's and g's:

$$\mu \simeq \prod_{i=1}^{N} f(i) \prod_{A=1}^{\infty} g(A). \tag{3.15}$$

If one inserts this Ansatz into Eq. (3.8) and integrates over the degrees of freedom for all of the oscillators and for $N-1$ of the particles, one obtains an equation of the form

$$\frac{\partial f(i)}{\partial \eta} + \frac{\partial}{\partial x_i{}^a}\left(\frac{d\tau_i}{d\eta}\frac{1}{\Omega^2 m}\gamma^{ab}(i) p^i{}_b f\right) - \frac{\partial}{\partial p^i{}_a}\left(\frac{d\tau_i}{d\eta}\frac{1}{2\Omega^2 m} p^i{}_\mu p^i{}_\nu\, \partial_a \gamma^{\mu\nu}(i) f\right) - \frac{\partial}{\partial p^i{}_a}\left(\frac{d\tau_i}{d\eta}\lambda \Delta_a{}^\mu(i)\langle F_\mu(i)\rangle f\right) = 0, \tag{3.16}$$

where, now,

$$\langle F_\mu(i)\rangle \equiv \partial^i{}_\mu \langle \Phi(i)\rangle = \partial^i{}_\mu [\Omega^{-1}\langle \chi(i)\rangle], \tag{3.17}$$

and

$$\langle \Phi(i)\rangle = \Omega^{-1}\sum_A \int dq_A\, dp_A\, g(A)\psi_A(i) q_A \tag{3.18}$$

denotes an "average" value for the scalar field Φ, defined with respect to the one-particle $g(A)$'s. In a similar fashion, one may obtain an equation for the evolution of each $g(A)$. Thus, if one supposes that μ is symmetric under particle interchange, one finds that

$$\frac{\partial g(A)}{\partial \eta} - p_A \frac{\partial g}{\partial q_A} + \left[\omega_A{}^2 - (1+6n)\frac{\Omega''}{\Omega}\right] q_A \frac{\partial g}{\partial p_A} + \frac{N\lambda}{\Omega}\int d^3x_1\, d^3p^1 f(1)\frac{d\tau_1}{d\eta}\psi_A(1)\frac{\partial g}{\partial p_A} = 0. \tag{3.19}$$

By constructing an appropriate moment of Eq. (3.19), it will become apparent that the coupled system of Eqs. (3.16) and (3.19) really do imply a simple mean-field description. The proof is straightforward if one observes that

$$\frac{\partial}{\partial \eta}\int dq_A\, dp_A\, g(A) q_A \psi_A(k) = -\int dq_A\, dp_A\, g(A) p_A \psi_A(k). \tag{3.20}$$

Equation (3.20) is easy enough to verify. Thus, if one views Eq. (3.19) as an operator equation

$$\frac{\partial g}{\partial \eta} = \Lambda g, \tag{3.21}$$

it follows immediately by an integration by parts that

$$\begin{aligned}\int dq_A\, dp_A\, q_A \psi_A(k)\frac{\partial g(A)}{\partial \eta} &= \int dq_A\, dp_A\, q_A \psi_A(k)\Lambda g(A) \\ &= \int dq_A\, dp_A\, q_A \psi_A(k) p_A \frac{\partial g}{\partial q_A} \\ &= -\int dq_A\, dp_A\, \psi_A(k) p_A g(A).\end{aligned} \tag{3.22}$$

What one must do now is multiply Eq. (3.19) by the quantity $\psi_A(i)p_A$, integrate over the variables q_A and p_A, and then sum over all the A's. By virtue of Eq. (3.20), it follows that

$$\begin{aligned}&\sum_A \int dq_A\, dp_A\, \psi_A(i) p_A \left[\frac{\partial g(A)}{\partial \eta} - (1+6n)\frac{\Omega''}{\Omega} q_A \frac{\partial g(A)}{\partial p_A}\right] \\ &= \left[-\partial_\eta{}^2 + (1+6n)\frac{\Omega''}{\Omega}\right]\langle \chi(i)\rangle \\ &= \left[-\partial_\eta{}^2 + (1+6n)\frac{\Omega''}{\Omega}\right]\Omega\langle \Phi(i)\rangle.\end{aligned} \tag{3.23}$$

Similarly, it is easy to see that

$$\begin{aligned}&\sum_A \int dq_A\, dp_A\, \psi_A(i) p_A \omega_A{}^2 q_A \frac{\partial g(A)}{\partial p_A} \\ &= -\sum_A \int dq_A\, dp_A\, g(A) q_A \omega_A{}^2 \psi_A(i) \\ &= \sum_A \int dq_A\, dp_A\, g(A) q_A \Delta \psi_A(i) \\ &= \Delta\langle \chi(i)\rangle = \Delta\Omega\langle \Phi(i)\rangle.\end{aligned} \tag{3.24}$$

And finally, a somewhat more complicated calculation shows that

$$\begin{aligned}&\frac{N\lambda}{\Omega}\sum_A \int dq_A\, dp_A\, \psi_A(i) p_A \int d^3x_1\, d^3p^1 f(1)\frac{d\tau_1}{d\eta}\psi_A(1)\frac{\partial g(A)}{\partial p_A} \\ &= -\frac{N\lambda}{\Omega}\sum_A\left[\int dq_A\, dp_A\, g(A)\right] \\ &\quad\times \int d^3x_1\, d^3p^1 f(1)\frac{d\tau_1}{d\eta}\psi_A(1)\psi_A(i) \\ &= -\frac{4\pi N\lambda}{\Omega}\int d^3x_1\, d^3p^1 f(1)\frac{d\tau_1}{d\eta} \\ &\quad\times(-\gamma^{\eta\eta})^{-1}(-\gamma)^{-1/2}\delta^{(3)}(x_1{}^a - x_i{}^a) \\ &= -\frac{4\pi\lambda\Omega^3}{m}\frac{\langle \rho(i)\rangle}{-\gamma^{\eta\eta}},\end{aligned} \tag{3.25}$$

where

$$\langle \rho(i)\rangle \equiv N\int d^3p \frac{d\tau_i}{d\eta}(-g)^{-1/2} m f(i) = Nm\int d\omega^i f(i), \tag{3.26}$$

and $d\omega^i$ denotes the invariant three-dimensional momentum space volume element associated with particle i.[17] The quantity $\langle \rho(i)\rangle$ may of course be interpreted as an "average" density defined with respect to the one-particle distribution.

By combining Eqs. (3.22)–(3.26), one concludes that

$$\begin{aligned}&\partial_\eta{}^2\langle \chi\rangle - (1+6n)(\Omega''/\Omega)\langle \chi\rangle - \Delta\langle \chi\rangle \\ &\quad = 4\pi(\lambda/m)\Omega^3\langle \rho\rangle/(-\gamma^{\eta\eta})\end{aligned} \tag{3.27}$$

or, equivalently, that

$$\nabla_\mu \nabla^\mu \langle \Phi\rangle + nR[g]\langle \Phi\rangle = 4\pi(\lambda/m)\langle \rho\rangle. \tag{3.28}$$

The coupled system of Eqs. (3.16) and (3.28) constitute a simple relativistic analog of the ordinary self-consistent field approximation. A gravitational analog of these equations is in fact the starting point for the stability theory of "collisionless stellar dynamics" developed by such authors as Ipser and Thorne,[11] in which one is concerned with the behavior of linearized perturbations of a static background space-time.[3] There is also an obvious connection with the sort of formalism developed by such authors as Ehlers, Ellis, Israel, and Stewart.[17,20–22] It should, however, be observed that these authors were concerned with the full nonlinear Einstein equations and, as such, they could not have expressed their fields as a linear superposition of oscillators.

At this stage, it is also straightforward to derive hydrodynamic moment equations involving the particle distributions. Thus, for example, a trivial integration of Eq. (3.16) shows that the "average" current

$$\begin{aligned}\langle J^\mu\rangle &\equiv N\int d^3p(-g)^{-1/2} m\frac{dx^\mu}{d\eta} f \\ &= N\int d^3p(-g)^{-1/2}\frac{d\tau}{d\eta}\frac{1}{\Omega^2}\gamma^{\mu\nu}p_\nu f \\ &= N\int d\omega\, P^\mu f\end{aligned} \tag{3.29}$$

will satisfy the relation

$$\nabla_\mu \langle J^\mu\rangle = 0. \tag{3.30}$$

This is nothing more than an expression of probability conservation. Similarly, if one multiplies Eq. (3.16) by $dx^\nu/d\tau$ before integrating, one finds that

$$\nabla_\mu \langle T^{\mu\nu}\rangle = -\frac{\lambda}{m}\nabla_\mu\langle\Phi\rangle(g^{\mu\nu}\langle\rho\rangle + \langle T^{\mu\nu}\rangle), \tag{3.31}$$

where

$$\langle T^{\mu\nu}\rangle = \frac{N}{m}\int d\omega\, P^\mu P^\nu f. \tag{3.32}$$

Equation (3.31) is the energy-momentum balance equation.

IV. TWO LIMITING CASES

A. A static background space-time

Equation (3.16) assumes a particularly simple form for the special case of a space-time that is actually static. In this case, one may of course arrange that $\Omega = 1$, so that $\eta = t$, and, therefore, the mean-field or Vlasov equation takes the form

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x^a}\left(\frac{d\tau}{dt}\frac{\gamma^{ab}p_b}{m}f\right) - \frac{\partial}{\partial p_a}\left(\frac{d\tau}{dt}\frac{1}{2m}p_\mu p_\nu\,\partial_a\gamma^{\mu\nu}f\right) - \frac{\partial}{\partial p_a}\left(\frac{d\tau}{dt}\lambda\Delta_a{}^\mu\nabla_\mu\langle\Phi\rangle f\right) = 0. \tag{4.1}$$

It is useful to rewrite this expression in a form that is manifestly covariant. In the limit that $\Omega = 1$, one knows that

$$-p_t = (-\gamma_{tt})^{1/2}(m^2 + \gamma^{ab}p_a p_b)^{1/2} \tag{4.2}$$

and

$$\frac{dt}{d\tau} = \frac{\gamma^{tt}p_t}{m}, \tag{4.3}$$

and, therefore, one finds that

$$\frac{\partial}{\partial p_a}\left(\frac{dt}{d\tau}\right) = -\frac{1}{mp_t}\gamma^{ab}p_b$$

and

$$\frac{\partial}{\partial x^a}\left(\frac{dt}{d\tau}\right) = -\frac{1}{2mp_t}p_\mu p_\nu\,\partial_a\gamma^{\mu\nu}. \tag{4.4}$$

Equation (4.1) thus takes the form

$$\frac{\partial}{\partial t}\left(\frac{\gamma^{tt}}{m}p_t f\right) + \frac{\partial}{\partial x^a}\left(\frac{\gamma^{ab}}{m}p_b f\right) - \frac{\partial}{\partial p_a}\left(\frac{1}{2m}p_\mu p_\nu\,\partial_a\gamma^{\mu\nu}f\right) - \lambda\Delta_a{}^\mu\nabla_\mu\langle\Phi\rangle\frac{\partial f}{\partial p_a} - \frac{3\lambda}{m}p^\mu\nabla_\mu\langle\Phi\rangle f = 0. \tag{4.5}$$

It is, however, easy to see that

$$\frac{\partial}{\partial p_\alpha}\Delta_{\alpha\mu} = \frac{3p_\mu}{m} \tag{4.6}$$

and

$$-\tfrac{1}{2}\,p_\mu p_\nu\,\partial_\alpha g^{\mu\nu} = \Gamma^\lambda_{\alpha\mu}p_\lambda p^\mu, \tag{4.7}$$

where $\Gamma^\lambda_{\alpha\mu}$ is the Christoffel symbol associated with $g_{\mu\nu}$, and one therefore concludes that

$$\frac{\partial}{\partial x^\alpha}\left(\frac{p^\alpha}{m}f\right) + \frac{\partial}{\partial p_\alpha}\left(\frac{1}{m}\Gamma^\lambda_{\alpha\mu}p_\lambda p^\mu f\right) - \frac{\partial}{\partial p_\alpha}(\lambda\Delta_\alpha{}^\mu\nabla_\mu\langle\Phi\rangle f) = 0, \tag{4.8}$$

or, equivalently, that

$$\frac{p^\alpha}{m}\frac{\partial f}{\partial x^\alpha} + \frac{1}{m}\Gamma^\lambda_{\alpha\mu}p_\lambda p^\mu\frac{\partial f}{\partial p_\alpha} - \frac{\partial}{\partial p_\alpha}(\lambda\Delta_\alpha{}^\mu\nabla_\mu\langle\Phi\rangle f) = 0. \tag{4.9}$$

In obtaining these relations, one has of course exploited the fact that f depends upon only three independent components of momentum. Equations (3.28) and (4.9) define a mean-field theory that is manifestly covariant. The relativistic theory of "collisionless stellar dynamics" admits to an analogous formulation.

One of the crucial bits of folklore underlying one's intuition in kinetic theory is the idea that the mean-field equations should admit the isothermal distribution as an exact, stationary solution. In other words, one anticipates that Eq. (4.9) should admit a solution of the form

$$\log f = a(\mathbf{x}) + B^\mu(\mathbf{x})\langle\pi_\mu\rangle = a(\mathbf{x}) + (1 + \lambda\langle\Phi\rangle/m)B_\mu(\mathbf{x})p^\mu, \tag{4.10}$$

where $\langle\pi_\mu\rangle$ denotes the canonical momentum associated with the "average" $\langle\Phi\rangle$.

That this is in fact the case is not difficult to see. If one substitutes Eq. (4.10) into Eq. (4.9) and, consistent with the linearized equations (2.13), neglects contributions quadratic in λ, one finds that

$$\frac{p^\mu}{m}\nabla_\mu\left(a - \frac{3\lambda}{m}\langle\Phi\rangle\right) + \left(1 + \frac{\lambda\langle\Phi\rangle}{m}\right)\frac{p^\mu p^\nu}{m}\nabla_{(\mu}B_{\nu)} - B^\mu\nabla_\mu\langle\Phi\rangle = 0. \tag{4.11}$$

This relation will be satisfied identically if the coefficients of p^μ and $p^\mu p^\nu$ and the terms independent of $\mathbf{p}$ each vanish separately. In other words, Eq. (4.10) will in fact provide a solution if (i) $\nabla_{(\mu}B_{\nu)} = 0$, (ii) $B^\mu\nabla_\mu\langle\Phi\rangle = 0$, and (iii) $a(\mathbf{x}) - 3\lambda\langle\Phi\rangle/m = \text{const.}$[23]

Condition (i) requires simply that B_μ be a Killing field. If, moreover, the hydrodynamic moments of f are to exist, one must demand that B_μ be timelike.[17] That such a Killing field exists is of course guaranteed since, by assumption, the space-time is static. Condition (ii) requires that B_μ Lie derive the average $\langle\Phi\rangle$, i.e., that $\langle\Phi\rangle$ be "time independent." If one supposes that $\mathbf{B} = \beta\,\partial_t$, Eq. (4.10) takes the form

$$\log f = \text{const} + \frac{3\lambda\langle\Phi\rangle}{m} - \beta(-\gamma_{tt})^{1/2}\left(1 + \frac{\lambda\langle\Phi\rangle}{m}\right)(m^2 + \gamma^{ab}p_a p_b)^{1/2}$$
$$= \text{const} + \frac{3\lambda\langle\Phi\rangle}{m} - \frac{1}{kT(\mathbf{x})}\left(1 + \frac{\lambda\langle\Phi\rangle}{m}\right)(m^2 + \gamma^{ab}p_a p_b)^{1/2}, \tag{4.12}$$

where, now,

$$kT(\mathbf{x})(-\gamma_{tt})^{1/2} \equiv \beta^{-1} = \text{const}. \tag{4.13}$$

It should be observed that the coefficient of β in Eq. (4.12) is nothing other than the Hamiltonian (2.7), viewed as a function of the average $\langle\Phi\rangle$. The $T(\mathbf{x})$ defined by Eq. (4.13) is of course the "physical," red-shifted temperature.[17,24] The presence of the factor $3\lambda\langle\Phi\rangle/m$, required by condition (iii), is a manifestation of the noninvariance of the six-dimensional volume element: $\partial F_\alpha/\partial p_\alpha \neq 0$.

B. Radiative modes in a spatially flat Friedmann cosmology

The mean field description developed in Sec. III may also be used to describe such phenomena as radiative modes in a $k = 0$ Friedmann cosmology. Here, of course, the metric may be written in the form

$$ds^2 = \Omega^2(\eta)(-d\eta^2 + \delta_{ab}\, dx^a\, dx^b). \tag{4.14}$$

And thus, if one supposes (i) that the conformal factor $\Omega \propto t^p \propto \eta^{p/(1-p)}$, and (ii) that $\langle \rho \rangle = 0$ (i.e., that only fields need be considered), one is led to an homogeneous wave equation of the form

$$\partial_\eta{}^2\langle\chi\rangle - [\,p(2p-1)/(1-p)^2](1+6n)\langle\chi\rangle - \Delta\,\langle\chi\rangle = 0. \tag{4.15}$$

If one assumes that $p = \frac{2}{3}$, i.e., that the cosmology is dominated by a zero-pressure fluid, and looks for solutions with spatial dependence $\exp(ik_a x^a)$, one obtains the amplitude equation

$$\tilde{\chi}'' - [2(1+6n)/\eta^2]\tilde{\chi} + k^2\tilde{\chi} = 0. \tag{4.16}$$

In terms of the rescaled

$$\zeta(\eta) = \eta^{-1/2}\tilde{\chi}(\eta), \tag{4.17}$$

this relation takes the form

$$\zeta'' + (1/\eta)\zeta' + (k^2 - \nu^2/\eta^2)\zeta = 0, \tag{4.18}$$

where

$$\nu = \tfrac{1}{2}(1 + 16n/3)^{1/2}. \tag{4.19}$$

Equation (4.18) is of course Bessel's equation and, therefore, the general solution takes the form

$$\langle\Phi(x^a,\eta)\rangle = \exp(ik_a x^a)\eta^{-3/2}[AJ_\nu(k\eta) + BN_\nu(k\eta)]. \tag{4.20}$$

In the limit of short-wavelength disturbances, Eq. (4.20) implies an oscillatory time dependence. Alternatively, in the limit of long wavelengths, one obtains a simple power law behavior. For the conformally invariant wave equation, with $n = -\frac{1}{6}$ and $\nu = \frac{1}{2}$, the long-wavelength limit implies that $\langle\Phi\rangle \propto \eta^{-1}$ or η^{-2}. In other words, one is led to damped solutions

$$\langle\Phi\rangle \propto t^{-1/3} \text{ or } t^{-2/3}. \tag{4.21}$$

For the minimally coupled equation, with $n = 0$ and $\nu = \frac{3}{2}$, the long-wavelength modes exhibit a time-dependence $\langle\Phi\rangle \propto \eta^0$ or η^{-3}, i.e.,

$$\langle\Phi\rangle \propto t^0 \text{ or } t^{-1}. \tag{4.22}$$

Again, in this case, there are no growing modes. If, however, one allows for any $n > 0$, one will in fact obtain solutions that grow with time. Thus, in particular, the case when $n = +\frac{1}{6}$ implies that $\langle\Phi\rangle \propto t^{1/3}$ or $t^{-4/3}$.

V. LINEARIZED GRAVITATIONAL INTERACTIONS IN A SPATIALLY FLAT FRIEDMANN COSMOLOGY

Turn now to the linearized gravitational interaction described in Sec. I. For the special case of a spatially flat Friedmann cosmology, with a metric given by Eq. (4.14), the particle equations of motion take the forms

$$\frac{dx_i{}^a}{d\eta} = \frac{d\tau_i}{d\eta}\frac{1}{\Omega^2 m}\delta^{ab}p^i{}_b$$

and

$$\frac{dp^i{}_a}{d\eta} = -\frac{d\tau_i}{d\eta}\frac{1}{\Omega^4 m}\Delta_{a\lambda}(i)\delta\Gamma^\lambda_{\mu\nu}(i)p_i{}^\mu p_i{}^\nu, \tag{5.1}$$

where, recall, $\delta\Gamma^\lambda_{\mu\nu}$ is defined by Eq. (1.2), $p_i{}^\mu = \eta^{\mu\nu}p^i{}_\nu$, and

$$\Delta_{\mu\nu} = \Omega^2\eta_{\mu\nu} + m^{-2}p_\mu p_\nu, \tag{5.2}$$

where $\eta_{\mu\nu}$ is the Minkowski metric.

When considering the field equation for $h_\mu{}^\nu$, it is very convenient to impose the so-called synchronous gauge condition[25,26]

$$h_\eta{}^\eta = h_a{}^\eta = 0. \tag{5.3}$$

This condition implies that one need only consider the spatial components of the field equation, and, moreover, that they may be viewed as a relation involving the 3×3 spatial tensor $h_a{}^b$. The perturbed Einstein tensor $\delta G_\alpha{}^\beta$ may of course be generated from the perturbed Ricci tensor by the relation

$$\delta G_\alpha{}^\beta = \delta R_\alpha{}^\beta - \tfrac{1}{2}\delta_\alpha{}^\beta\delta R_\mu{}^\mu, \tag{5.4}$$

and it is straightforward to calculate[13,26] that, for the metric (4.14),

$$\delta R_b{}^a = \frac{1}{2\Omega^2}(\partial_c\,\partial^a h_b{}^c + \partial^c\,\partial_b h_c{}^a - \partial^c\,\partial_c h_b{}^a - \partial^a\,\partial_b h_c{}^c) + \frac{1}{2\Omega^2}\partial_\eta{}^2 h_b{}^a + \frac{\Omega'}{\Omega^3}\partial_\eta h_b{}^a + \frac{\Omega'}{2\Omega^3}\delta_b{}^a\,\partial_\eta h_c{}^c \tag{5.5}$$

and

$$\delta R_\eta{}^\eta = (1/2\Omega^2)\partial_\eta{}^2 h_c{}^c + (\Omega'/2\Omega^3)\partial_\eta h_c{}^c. \tag{5.6}$$

It is useful to re-express Eqs. (5.5) and (5.6) in terms of the new function

$$\xi_a{}^b = h_a{}^b - \delta_a{}^b h_c{}^c, \tag{5.7}$$

for which

$$h_a{}^b = \xi_a{}^b - \tfrac{1}{2}\delta_a{}^b\xi_c{}^c. \tag{5.8}$$

Thus, with this substitution, one finds that

$$\delta G_a{}^b = (1/2\Omega^2)[\partial_\eta{}^2\xi_a{}^b + (2\Omega'/\Omega)\partial_\eta\xi_a{}^b - (\Delta\xi)_a{}^b], \tag{5.9}$$

where[27]

$$(\Delta\xi)_a{}^b = \partial^c\,\partial_c\xi_a{}^b - \partial^b\,\partial_c\xi_a{}^c - \partial^c\,\partial_a\xi_c{}^b + \delta_a{}^b\partial^d\,\partial_c\xi_d{}^c + \partial_a\,\partial^b\xi_c{}^c - \tfrac{1}{2}\delta_a{}^b\partial^c\,\partial_c\xi_d{}^d. \tag{5.10}$$

In Eq. (5.10), one is again instructed to raise and lower indices with the conformal metric $\gamma_{\mu\nu}$, so that $\partial^c = \gamma^{c\mu}\partial_\mu = \eta^{c\mu}\partial_\mu = \partial_c$. These relations lead to an inhomogeneous wave equation of the form

$$\partial_\eta{}^2\xi_a{}^b + (2\Omega'/\Omega)\partial_\eta\xi_a{}^b - (\Delta\xi)_a{}^b = 16\pi\Omega^2(T_a{}^b - [T_a{}^b]_B), \tag{5.11}$$

where, recall, $T_\mu{}^\nu$ and $[T_\mu{}^\nu]_B$ denote, respectively, the

stress energies appropriate for the "true" and "background" space-times.

It is convenient at this stage to eliminate explicit reference to $[T_a{}^b]_B$ by rewriting $\xi_a{}^b$ in the form

$$\xi_a{}^b = H_a{}^b - [H_a{}^b]_B, \tag{5.12}$$

where

$$\partial_\eta{}^2[H_a{}^b]_B + (2\Omega'/\Omega)\partial_\eta[H_a{}^b]_B - (\Delta[H]_B)_a{}^b = 16\pi\Omega^2[T_a{}^b]_B. \tag{5.13}$$

This decomposition of $\xi_a{}^b$ is of course not unique. In any case, Eqs. (5.12) and (5.13) imply that

$$\partial_\eta{}^2 H_a{}^b + (2\Omega'/\Omega)\partial_\eta H_a{}^b - (\Delta H)_a{}^b = \frac{16\pi}{m\Omega^4}\sum_i \frac{d\tau_i}{d\eta} p^i{}_a p_i{}^b \delta^{(3)}[x^c - x_i{}^c(\eta)], \tag{5.14}$$

where one has inserted the explicit form of $T_a{}^b$.

One is now in a position to expand the linear field $H_a{}^b$ in terms of appropriately defined field oscillators. In analogy with the discussion in Sec. III, the idea is to view Eq. (5.14) as an operator equation on a Hilbert space with inner product

$$(\xi^a{}_b, \zeta_a{}^b) = \int d^3x\, \xi^a{}_b(x^c)\zeta_a{}^b(x^c). \tag{5.15}$$

The important point again is that if one can justify neglecting all surface terms when integrating by parts, one may show that Δ is in fact symmetric[28]:

$$(\xi, \Delta\zeta) = (\Delta\xi, \zeta). \tag{5.16}$$

This implies that the eigenvalues of Δ will be real and that the eigenvectors will be orthogonal. Had one worked instead with $h_a{}^b$, rather than with $\xi_a{}^b$, the analysis would have been more complicated.

If, as in Sec. III, one assumes that the modes are both discrete and complete, $H_a{}^b$ may be expressed as a linear combination of the form

$$H_a{}^b(x^c, \eta) = \sum_{A=1}^{\infty} q_A(\eta)\psi_{Aa}{}^b(x^c), \tag{5.17}$$

where

$$(\Delta\psi_A)_a{}^b + \omega_A{}^2\psi_{Aa}{}^b = 0. \tag{5.18}$$

If one then imposes the normalization

$$(\psi_A, \psi_B) = 4\pi\delta_{AB}, \tag{5.19}$$

the assumption of completeness requires that

$$\sum_A \psi_A{}^a{}_b(x^e)\psi_A{}^c{}_d(x^e) = 4\pi\delta^{(a}{}_b\delta^{c)}{}_d\delta^{(3)}(x^e - y^e). \tag{5.20}$$

At this stage, it is easy to obtain the equations of motion for the individual oscillators. All that one need do is substitute the expansion (5.17) into Eq. (5.14), take the inner product with $\psi_B(i)$, and divide through by 4π:

$$q_B'' + \frac{2\Omega'}{\Omega} q_B' + \omega_B{}^2 q_B = \frac{4}{m\Omega^4}\sum_i \frac{d\tau_i}{d\eta} p^i{}_a p_i{}^b \psi_{Ab}{}^a(i) \equiv \sum_i \sigma_a{}^b(i)\psi_{Ab}{}^a(i). \tag{5.21}$$

Alternatively, Eq. (5.21) may be viewed as a coupled first-order system of the form

$$\frac{dq_A}{d\eta} = -p_A$$

and

$$\frac{dp_A}{d\eta} = -\frac{2\Omega'}{\Omega}p_A + \omega_A{}^2 q_A - \sum_i \sigma_a{}^b(i)\psi_{Ab}{}^a(i). \tag{5.22}$$

Given the equations of motion (5.1) and (5.22), it is straightforward to define a distribution function $\mu(x_1{}^a, p^1{}_a, \ldots, q_A, p_A, \ldots; \eta)$ and to formulate an N-particle conservation equation. In analogy with Eq. (3.8), this relation takes the form

$$\frac{\partial\mu}{\partial\eta} + \sum_i \frac{\partial}{\partial x_i{}^a}\left(\frac{d\tau_i}{d\eta}\frac{1}{\Omega^2 m}p_i{}^a\mu\right) - \sum_i \frac{\partial}{\partial p^i{}_a}\left(\frac{d\tau_i}{d\eta}\frac{1}{\Omega^4 m}\Delta_{a\lambda}(i)\delta\Gamma^\lambda_{\mu\nu}(i)p_i{}^\mu p_i{}^\nu\right) - \sum_A \frac{\partial}{\partial q_A}(p_A\mu) + \sum_A \frac{\partial}{\partial p_A}\left\{\left[-\frac{2\Omega'}{\Omega}p_A + \omega_A{}^2 q_A - \sum_i \sigma_a{}^b(i)\psi_{Ab}{}^a(i)\right]\mu\right\} = 0. \tag{5.23}$$

Such matters as the definitions of reduced distributions will of course be trivial.

It is also simple enough to explore the consequences of a self-consistent field approximation of the form generated by Eq. (3.15). Thus, in this approximation, it is easy to see that

$$\frac{\partial f(i)}{\partial\eta} + \frac{d\tau_i}{d\eta}\frac{p_i{}^a}{m\Omega^2}\frac{\partial f}{\partial x_i{}^a} - \frac{\partial}{\partial p^i{}_a}\left(\frac{d\tau_i}{d\eta}\frac{1}{\Omega^4 m}\Delta_{a\lambda}(i)\langle\delta\Gamma^\lambda_{\mu\nu}(i)\rangle p_i{}^\mu p_i{}^\nu f\right) = 0, \tag{5.24}$$

where, now, $\langle\delta\Gamma^\lambda_{\mu\nu}\rangle$ denotes the "perturbed Christoffel symbol" associated with the average $\langle h_{\mu\nu}\rangle$ [c.f. Eq. (1.5)]:

$$\langle\delta\Gamma^\lambda_{\mu\nu}\rangle \equiv \tfrac{1}{2}(\nabla_\mu\langle h_\nu{}^\lambda\rangle + \nabla_\nu\langle h_\mu{}^\lambda\rangle - \nabla^\lambda\langle h_{\mu\nu}\rangle). \tag{5.25}$$

Consistent with the gauge conditions (5.3), $\langle h_{\mu\nu}\rangle$ is constrained so that

$$\langle h_\eta{}^\eta\rangle = \langle h_\eta{}^a\rangle = \langle h_a{}^\eta\rangle = 0, \tag{5.26}$$

whereas the spatial components satisfy the relation

$$\langle h_a{}^b\rangle = \langle\xi_a{}^b\rangle - \delta_a{}^b\langle\xi_c{}^c\rangle, \tag{5.27}$$

where

$$\langle\xi_a{}^b\rangle = \langle H_a{}^b\rangle - [H_a{}^b]_B = \sum_A \int dq_A\, dp_A\, g(A) q_A \psi_{Aa}{}^b - [H_a{}^b]_B. \tag{5.28}$$

The quantity $\langle H_a{}^b\rangle$ may be viewed as an average value for $H_a{}^b$ defined with respect to the one-particle $g(A)$'s. In a similar fashion, one finds that, in this approximation,

$$\frac{\partial g(A)}{\partial\eta} - p_A\frac{\partial g}{\partial q_A} - \frac{2\Omega'}{\Omega}\frac{\partial}{\partial p_A}(p_A g) + \omega_A{}^2 q_A\frac{\partial g}{\partial p_A} - N\int d^3x_1\, d^3p^1 f(1)\sigma_a{}^b(1)\psi_{Ab}{}^a(1)\frac{\partial g}{\partial p_A} = 0. \tag{5.29}$$

It is again straightforward to verify that Eq. (5.29) implies a well-defined "average" field equation. Thus, in analogy with Eq. (3.20), it is easy to see that

$$\frac{\partial}{\partial\eta}\int dq_A\, dp_A\, g(A)q_A\psi_{Aa}{}^b = -\int dq_A\, dp_A\, g(A)p_A\psi_{Aa}{}^b, \tag{5.30}$$

and, therefore, one may verify that

$$\partial_\eta{}^2\langle H_a{}^b\rangle + (2\Omega'/\Omega)\partial_\eta\langle H_a{}^b\rangle - (\Delta\langle H\rangle)_a{}^b = 16\pi\Omega^2\langle T_a{}^b\rangle, \tag{5.31}$$

or, equivalently, that

$$\delta G_a{}^b[\langle H\rangle] = 8\pi\langle T_a{}^b\rangle, \tag{5.32}$$

where $\delta G_a{}^b[\langle H\rangle]$ denotes the linearized Einstein tensor associated with $\langle H_a{}^b\rangle$, and $\langle T_a{}^b\rangle$ is defined by Eq. (3.32).

As a simple example, one may suppose that the "average" $\langle T_\mu{}^\nu\rangle$ is characterized completely by an average four-velocity $\langle U^\mu\rangle$, an isotropic pressure $\langle P\rangle$, and an average energy density $\langle\epsilon\rangle = \langle\rho\rangle + 3\langle P\rangle$ (this will, for example, be the case if the one-particle distribution is a local Maxwellian[29]). In this event, $\langle T_\mu{}^\nu\rangle$ takes the form appropriate for a perfect fluid:

$$\langle T_\mu{}^\nu\rangle = (\langle\epsilon\rangle + \langle P\rangle)\langle U_\mu\rangle\langle U^\nu\rangle + \langle P\rangle\delta_\mu{}^\nu. \tag{5.33}$$

If one then assumes further that, in the usual way, $[T_\mu{}^\nu]_B$ is itself given as a perfect fluid, and recalls that the quantity $\langle T_a{}^b\rangle - [T_a{}^b]_B$ is to be treated as a linearized perturbation, one concludes that

$$\langle\delta T_a{}^b\rangle \equiv \langle T_a{}^b\rangle - [T_a{}^b]_B = \delta_a{}^b\langle\delta P\rangle, \tag{5.34}$$

so that the "average" field equation takes the form

$$\delta G_a{}^b[\langle\xi\rangle] = 8\pi\langle\delta T_a{}^b\rangle = 8\pi\delta_a{}^b\langle\delta P\rangle. \tag{5.35}$$

In a similar fashion, one may record hydrodynamic moment equations involving the particle distribution. Thus, for example, it is easy to see that the average current $\langle J^\mu\rangle$ satisfies the relation

$$\nabla_\mu\langle J^\mu\rangle = 0. \tag{5.36}$$

And, similarly, one finds that the energy-momentum balance equation takes the form

$$\nabla_\alpha\langle T^{\alpha\beta}\rangle = -\langle\delta\Gamma^\lambda_{\mu\nu}\rangle(\delta_\lambda{}^\beta\langle T^{\mu\nu}\rangle + \langle T_\lambda{}^{\beta\mu\nu}\rangle), \tag{5.37}$$

where, now,

$$\langle T_\lambda{}^{\beta\mu\nu}\rangle = \frac{N}{m^3}\int d\omega\, P_\lambda P^\beta P^\mu P^\nu f. \tag{5.38}$$

It is of some interest to consider solutions to Eq. (5.35) for the special case when $\langle\delta T_a{}^b\rangle = 0$. If one supposes that $\Omega\propto\eta^2\propto t^{2/3}$, and looks for plane-wave solutions, one obtains an amplitude equation of the form

$$\tilde\xi_a{}^{b\,\prime\prime} + (4/\eta)\tilde\xi_a{}^{b\,\prime} + k^2\tilde\xi_a{}^b - k^b k_c\tilde\xi_a{}^c - k^c k_a\tilde\xi_c{}^b + \delta_a{}^b k^d k_c\tilde\xi_d{}^c + k^b k_a\tilde\xi_c{}^c - \tfrac{1}{2}\delta_a{}^b k^2\tilde\xi_d{}^d = 0, \tag{5.39}$$

where, explicitly, $k^2 = k_c k^c = k_c k_c$. Equation (5.39) is still rather complicated, since it couples together the various components $\tilde\xi_a{}^b$. If, however, one restricts attention to traceless modes with $h_c{}^c = 0$, so that

$$\langle\tilde h_a{}^b\rangle = \langle\tilde\xi_a{}^b\rangle, \tag{5.40}$$

and supposes further that $\nabla_a\langle h^a{}_b\rangle = 0$, so that

$$k_a\tilde\xi_b{}^a = 0, \tag{5.41}$$

one obtains the comparatively simple relation

$$\tilde h_a{}^{b\,\prime\prime} + \frac{4}{\eta}\tilde h_a{}^{b\,\prime} + k^2\tilde h_a{}^b = 0. \tag{5.42}$$

It follows immediately from Eq. (5.42) that

$$\langle h_a{}^b(x^c,\eta)\rangle = \eta^{-3/2}\exp(ik_c x^c)[A_a{}^b J_{3/2}(k\eta) + B_a{}^b N_{3/2}(k\eta)], \tag{5.43}$$

where $A_a{}^b$ and $B_a{}^b$ are constants so chosen that $k_a A_b{}^a = k_a B_b{}^a = 0$ and $A_a{}^a = B_a{}^a = 0$. In the limit of short-wavelength disturbances, those solutions will exhibit an oscillatory behavior. Alternatively, in the limit of long wavelengths, they will exhibit a time dependence of the form

$$\langle h_a{}^b\rangle \propto t^0 \text{ or } t^{-1}, \tag{5.44}$$

so that $\langle h_{\mu\nu}\rangle \equiv g_{\mu\alpha}\langle h_\nu{}^\alpha\rangle \propto t^{4/3}$ or $t^{1/3}$. These are of course well-known results.[25]

ACKNOWLEDGMENTS

Limited financial support was provided by the National Science Foundation and the Robert A. Welch Foundation.

[1]W. Israel and H. E. Kandup, Ann. Phys. (NY) **152**, 30 (1984).
[2]H. E. Kandrup, Ann. Phys. (NY) **153**, 44 (1984).
[3]H. E. Kandrup, Astrophys. J. **282** (in press).
[4]H. E. Kandrup, Physica A (in press).
[5]R. Hakim, J. Math. Phys. **8**, 1315 (1967).
[6]R. Hakim, J. Math. Phys. **8**, 1379 (1967).
[7]H. E. Kandrup, Astrophys. J. **244**, 316 (1981).
[8]See, e.g., A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rome, 1976).
[9]K. Kuchař, in *Relativity, Astrophysics, and Cosmology*, edited by W. Israel (Reidel, Dordrecht, 1973).
[10]R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), and references contained therein.
[11]See, e.g., J. R. Ipser and K. S. Thorne, Astrophys. J. **154**, 251 (1968), and references contained therein.
[12]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[13]L. D. Landau and E. M. Lifshitz, *Classical Theory of Fields* (Pergamon, Oxford, 1975).
[14]R. Balescu and T. Kotera, Physica **33**, 581 (1967).
[15]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge University, Cambridge, 1973).
[16]See, e.g., W. Heitler, *The Quantum Theory of Radiation* (Clarendon, Oxford, 1954) or Ref. 14 for an analogous treatment of the electromagnetic field in Minkowski space.
[17]W. Israel, in *General Relativity: Papers in Honour of J. L. Synge*, edited by L. O'Raifertaigh (Clarendon, Oxford, 1972).
[18]See, e.g., R. Balescu, *Equilibrium and Non-Equilibrium Statistical Mechanics* (Wiley, New York, 1975).
[19]See, e.g., R. Balescu and J. Wallenborn, Physica **54**, 477 (1971); and the references to the earlier work of Balescu and Prigogine cited therein.
[20]J. Ehlers, in *General Relativity and Gravitation: Enrico Fermi Course XLVII* (Academic, New York, 1971).
[21]G. F. R. Ellis, D. R. Matravers, and R. Treciokas, Ann. Phys. (NY) **150**, 455 (1983).
[22]J. M. Stewart, *Non-Equilibrium Relativistic Kinetic Theory* (Springer, Berlin, 1971); and references contained therein.
[23]Had one considered a nonlinear equation of motion, analogous to Eq. (2.11), but with Φ replaced by $\langle\Phi\rangle$, conditions (i) and (ii) would remain unchanged, but (iii) would take the form $a(\mathbf{x}) - 3\log(1+\lambda\langle\Phi\rangle/m) = \text{const}$. See Ref. 2.
[24]H. E. Kandrup, Astrophys. Space Sci. (in press).

[25]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[26]E. M. Lifshitz and I. M. Khalatnikov, Adv. Phys. **12**, 208 (1963).
[27]Given Eq. (6.18) in Ref. 26, it is easy to see that if one were to consider a more general Friedman cosmology, $(\Delta\xi)_a{}^b$ would acquire an additional contribution of the form $4k(\xi_a{}^b + \frac{1}{2}\delta_a{}^b\xi_c{}^c)$, where $k = +1, -1$, respectively, for the spatially closed and open models. This additional contribution will not alter such properties of Δ as the symmetry expressed by Eq. (5.16).
[28]Because the conformal metric is flat, the formal analysis of Δ should proceed in complete analogy with the standard treatment of the electromagnetic interaction in Minkowski space. One therefore anticipates that rigorous proofs of symmetry, completeness, etc., should all go through without significant complications.
[29]See, e.g., J. M. Stewart, Proc. R. Soc. London, Ser. A **357**, 59 (1977).

A cosmological solution of Einstein's equations

Paul S. Wesson
Department of Physics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

(Received 22 February 1984; accepted for publication 8 June 1984)

An exact solution of the equations of general relativity is given which tends to Einstein–de Sitter at late times. The energy density is inhomogeneous at early times, and is singular on two spherical surfaces which may be related to the "bubbles" typical of inflationary universe models.

PACS numbers: 04.20.Jb, 98.80.Dr, 95.30.Sf

I. INTRODUCTION

The universe at the present time can be reasonably well described by a solution of Einstein's equations in which the energy density of matter is homogeneous and isotropic and the pressure of matter is zero. The simplest such solution is the Einstein–de Sitter one. However, it is now widely acknowledged that the universe at early times cannot be well described by solutions of this type: they lead to serious astrophysical problems, of which the best known are those to do with flatness, the horizon, and galaxy formation. These problems have been known for a while, and can in principle be resolved by a model which is Einstein–de Sitter at late times but differs from this at early times. This has been one of the motivations for the study of inhomogeneous and anisotropic solutions of Einstein's equations (see Refs. 1–4 for reviews). Attempts to resolve the classical astrophysical problems have been recently renewed, and have led to the proposal of the "inflationary" universe models.[5,6] Many solutions of Einstein's equations which are inflationary in nature are also inhomogeneous and anistropic,[7] so the motivation for studying solutions of the latter type has also been recently renewed.

Although there is strong motivation for studying solutions of Einstein's equations which are Einstein–de Sitter at late times but differ from this at early times, most solutions of the desired type are too complicated to work with easily in astrophysical contexts. Therefore, an attempt has been made to find solutions of the desired type which are relatively simple. The procedure which has been adopted is naive but effective. In the Einstein–de Sitter solution, distances vary as $f^{2/3}$, where $f=t$ and t is the time. An attempt has been made to find "generalized" Einstein–de Sitter solutions which are spherically symmetric and have $f=k_1t+k_2g(R)+k_3$, where the k's are constants and $g(R)$ is a function of a radial space coordinate. In what follows, a particularly simple solution of this type will be presented.

II. A COSMOLOGICAL SOLUTION

A spherically symmetric metric is taken in the form

$$ds^2=e^{\sigma}dt^2-e^{\omega}dR^2-r^2(d\theta^2+\sin^2\theta\,d\phi^2). \tag{1}$$

Here, the metric coefficients σ, ω, and r are all functions of the time t and a comoving radial space coordinate R. Derivatives with respect to t and R will be denoted by $(\cdot)$ and $(')$, respectively. With (1) and a perfect-fluid energy-momentum tensor, Einstein's equations can be expressed[8–10] as five relations (of which the first is really a definition). These are

$$2m/r=1+e^{-\sigma}\dot{r}^2-e^{-\omega}r'^2, \tag{2a}$$

$$\dot{m}=-4\pi r^2\dot{r}p, \tag{2b}$$

$$m'=4\pi r^2r'\epsilon, \tag{2c}$$

$$\sigma'=-2p'/(p+\epsilon), \tag{2d}$$

$$\dot{\omega}=-2\dot{\epsilon}/(p+\epsilon)-4\dot{r}/r. \tag{2e}$$

Here, m is the mass, p is the pressure, and ϵ is the energy density. Units have been chosen in which the magnitudes of the Newtonian gravitational constant and the velocity of light are unity.

It may be verified by direct substitution that a solution of (2) is given by

$$f\equiv k_1t+k_2\log_e R+k_3, \tag{3a}$$

$$e^{\sigma/2}=1,\quad e^{\omega/2}=f^{2/3}(1+2k_2/3f),\quad r=f^{2/3}R, \tag{3b}$$

$$m=2k_1^2R^3/9,\quad p=0,\quad \epsilon=k_1^2/2\pi f(2k_2+3f). \tag{3c}$$

Here, as mentioned above, the k's are constants.

III. DISCUSSION

The solution (3) has some interesting properties, of which the main ones may be mentioned here. For $t\to\infty$, $f\sim k_1t$, $e^{\omega/2}\sim f^{2/3}$, $\epsilon\sim(6\pi t^2)^{-1}$, and is homogeneous, and the solution tends to Einstein–de Sitter. For $t\to 0$, ϵ is inhomogeneous. The energy density is singlular ($\epsilon\to\infty$) on two hypersurfaces, namely $f=0$ and $(2k_2+3f)=0$. On these hypersurfaces, the metric is also singular, since $r\to 0$ and $e^{\omega/2}\to 0$, respectively. This behavior appears to be typical of solutions like this, since in another solution of this type[9] a similar behavior is found. It is not necessarily unphysical, since by a suitable choice of parameters such solutions can be interpreted as cosmological models.[11] Indeed, the existence of singular hypersurfaces in (3) may be regarded as an asset. These surfaces are spherical in ordinary three-dimensional space, and may therefore be related to the "bubbles" typical of inflationary universe models.[5–7] These latter models are characterized by two or more regions separated by singular surfaces, where at least one of the regions has a finite vacuum energy density. The latter can be incorporated into the solution (3) by the usual device[7] of assuming that the total energy density (ϵ) and total pressure (p) can be expressed as sums of a matter part and a vacuum part: $\epsilon\equiv\epsilon_m+\epsilon_v$, $p\equiv p_m+p_v$. For $\epsilon_v=-p_v=\Lambda/8\pi$, this is equivalent to introducing the cosmological constant Λ. It is hoped to report in greater detail on the astrophysical applications of (3) and its relation to inflationary universe models in future work.

ACKNOWLEDGMENTS

This work was suggested by an earlier collaboration with R. N. Henriksen.

 0022-2488/84/113297-02$02.50

It was supported by the Natural Sciences and Engineering Research Council of Canada.

[1]M. MacCallum, in *Cargèse Lectures in Physics 6*, edited by E. Schatzman (Gordon and Breach, London, 1973), pp. 61–174.
[2]L. C. Shepley, in Ref. 1, pp. 227–260.
[3]D. Kramer, H. Stephani, E. Herlt, M. MacCallum, and E. Schmutzer, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980).
[4]P. S. Wesson, *Gravity, Particles, and Astrophysics* (Reidel, Dordrecht, 1980), p. 147.
[5]A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[6]A. H. Guth and E. J. Weinberg, Phys. Rev. D **23**, 876 (1981).
[7]R. N. Henriksen, A. G. Emslie, and P. S. Wesson, Phys. Rev. D **27**, 1219 (1983).
[8]M. A. Podurets, Sov. Astron. **8**, 19 (1964).
[9]R. N. Henriksen and P. S. Wesson, Astrophys. Space Sci. **53**, 429 (1978).
[10]P. S. Wesson, J. Math. Phys. **19**, 2283 (1978).
[11]P. S. Wesson, Astrophys. J. **228**, 647 (1979).

Static space-time in general scalar tensor theory

N. Banerjee, S. B. Duttachoudhury, and A. Banerjee
Department of Physics, Jadavpur University, Calcutta 700032, India

(Received 9 January 1984; accepted for publication 23 March 1984)

The problem of the coupled Einstein–Maxwell scalar field in the framework of general scalar tensor theory of Nordtvedt is completely solved for plane symmetric static matter-free space-time. Special cases are considered for some special choices of ω as functions of the scalar field. Solutions are given also in the absence of the electromagnetic field and these are shown to generate a few special cases of Bianchi I cosmological models when subject to a set of complex transformations of coordinates.

PACS numbers: 04.50. + h, 04.20.Jb

I. INTRODUCTION

Amongst different scalar tensor theories of gravitation Brans–Dicke[1] theory attracted attention at a certain stage since the theory could incorporate Mach's principle in its framework and also because there is no *a priori* reason to exclude the introduction of a scalar field in the universe. However, the results of recent experiments pointed towards very large values of ω at the present stage of the universe making only a very little deviation from Einstein's theory. In this context the general scalar tensor theory in which ω is a function of the scalar field and is a variable (Nordtvedt[2]) may be worth investigating. The implication of the said theory in cosmological models appear to be particularly appealing in view of the fact that the parameter ω, which is apparently quite large at the present stage of evolution, might be small at other epochs and small ω will introduce a distinct difference in the dynamics of a model from that of the corresponding model in the absence of the scalar field. Because of such a possible role of the scalar field in the generalized scalar tensor theory exact static and nonstatic solutions should be studied. In fact there are already some in the literature in Nordtvedt's theory (Banerjee and Duttachoudhury,[3] Barkar,[4] Banerjee and Santos,[5,6] Rao and Reddy,[7] and Van den Bergh[8]).

In the present paper we propose to find exact solutions of the gravitational field equations for the electrovac in Nordtvedt's scalar-tensor theory for a plane symmetric static space-time, which by definition admits three parameter groups with minimum varieties as zero-curvature two-dimensional surfaces. The work of Amundsen and Grøn[9] contains exhaustive references of existing solutions for plane symmetric static and nonstatic solutions, none of which, however, refers to Nordtvedt's scalar tensor theory.

In Sec. II we consider the plane symmetric metric in the form of Taub and completely solve the field equations in presence of electromagnetic field and scalar field. Solutions are given for different choices of ω as functions of ψ. The functional forms are chosen as examples from the different theories as given clearly in Van den Bergh's paper previously mentioned and the corresponding solutions are given. They include the Brans–Dicke solution as a special case for $\omega = \text{const}$.

In Sec. III exact solutions in the absence of electromagnetic field are obtained.

Lastly, in Sec. IV a few special cases of Bianchi I cosmological solutions are constructed from the static plane symmetric solutions by complex transformations of the coordinates and some of their properties are indicated.

II. SOLUTIONS OF THE FIELD EQUATIONS

We consider the line element in the form (Taub)

$$ds^2 = e^{2\alpha}(dt^2 - dx^2) - e^{2\beta}(dy^2 + dz^2), \tag{2.1}$$

where α and β are functions of x alone. The coupled Nordtvedt–Maxwell field equations in matter-free space may be written as

$$2\beta'' + 3\beta'^2 - 2\alpha'\beta' = -\frac{e^{-2\alpha}\phi'^2}{\psi} - \frac{\omega\psi'^2}{2\psi^2} + (\alpha' - 2\beta')\frac{\psi'}{\psi} - \frac{\psi''}{\psi}, \tag{2.2}$$

$$\alpha'' + \beta'' + \beta'^2 = \frac{e^{-2\alpha}\phi'^2}{\psi} - \frac{\omega\psi'^2}{2\psi^2} - \beta'\frac{\psi'}{\psi} - \frac{\psi''}{\psi}, \tag{2.3}$$

$$\beta'^2 + 2\alpha'\beta' = -\frac{e^{-2\alpha}\phi'^2}{\psi} + \frac{\omega\psi'^2}{2\psi^2} - (\alpha' + 2\beta')\frac{\psi'}{\psi}, \tag{2.4}$$

where a prime indicates differentiation with respect to x. Further the wave equation for the scalar field ψ and the Maxwell equation for the electric potential ϕ are given, respectively, by

$$\psi'' + 2\beta'\psi' = -\omega'\psi'/(2\omega + 3), \tag{2.5}$$

and

$$[e^{-2(\alpha-\beta)}\phi']' = 0. \tag{2.6}$$

The Maxwell equation (2.6) can be readily integrated to yield

$$e^{2(\beta-\alpha)}\phi' = q, \tag{2.7}$$

where q is an arbitrary constant which is seen to be related to the charge contained by the source. Adding (2.2) and (2.4) together we obtain

$$2\beta''\psi + 4\beta'^2\psi + \psi'' + 4\beta'\psi' = -2e^{-2\alpha}\phi'^2. \tag{2.8}$$

Multiplying Eq. (2.8) throughout by $e^{2\beta}$ and using Eq. (2.7) we arrive at the equation

$$[e^{2\beta}\psi]'' + 2q\phi' = 0, \tag{2.9}$$

which in turn yields on integration

$$[e^{2\beta}\psi]' + 2q\phi = p, \tag{2.10}$$

where p is another arbitrary constant. Next we multiply Eq.

 0022-2488/84/113299-04$02.50

(2.3) by 2 and add the result with the equation obtained by subtracting Eq. (2.4) from Eq. (2.2). This gives

$$(2\alpha'' + 4\alpha'\beta')\psi + (2\alpha' + 2\beta')\psi' + \psi'' - 2e^{-2\alpha}\phi'^2 = 0. \tag{2.11}$$

By means of Eq. (2.6) we now eliminate β' from Eq. (2.11) and obtain

$$(e^{2\alpha}\psi)''\phi' - (e^{2\alpha}\psi)'\phi'' = 2\phi'^3, \tag{2.12}$$

which on integration gives

$$e^{2\alpha}\psi = (\phi^2 + a\phi + b), \tag{2.13}$$

where a and b are arbitrary constants. Again Eq. (2.10) may also be written as

$$e^{2\beta}\psi(2\beta + \psi'/\psi) = p - 2q\phi. \tag{2.14}$$

Dividing Eq. (2.14) by (2.13) and using Eq. (2.7) we get

$$2\beta' + \frac{\psi'}{\psi} = \frac{(p/q - 2\phi)\phi'}{\phi^2 + a\phi + b}, \tag{2.15}$$

which can be integrated in three different cases, viz. $a^2 > 4b$, $a^2 = 4b$, and $a^2 < 4b$, to yield the following relations:

(a) $a^2 > 4b$,

$$e^{2\beta}\psi = (\phi^2 + a\phi + b)^{-1} \times \left[\frac{2\phi + a - \sqrt{a^2 - 4b}}{2\phi + a + \sqrt{a^2 - 4b}}\right]^{(p/q + a)/\sqrt{a^2 - 4b}}; \tag{2.16a}$$

(b) $a^2 = 4b$,

$$e^{2\beta}\psi = (2\phi + a)^{-2}\exp\left[-\frac{2(p/q + a)}{(2\phi + a)}\right]; \tag{2.16b}$$

(c) $a^2 < 4b$,

$$e^{2\beta}\psi = (\phi^2 + a\phi + b)^{-1} \times \exp\left[\frac{2(p/q + a)}{\sqrt{4b - a^2}}\tan^{-1}\frac{(2\phi + a)}{\sqrt{4b - a^2}}\right]. \tag{2.16c}$$

Equations (2.13) and (2.16) relate the metric with the electric potential and the scalar field. In view of (2.7), (2.13), and (2.16) one gets equations for the electric potential ϕ, which on integration yield the following different solutions in different cases:

(i) $a^2 > 4b$, $p/q \neq -a$, $p/q \neq -a \pm \sqrt{a^2 - 4b}$,

$$\frac{(2\phi + p/q + 2a)^2 + (p^2/q^2 + 2ap/q + 4b)}{4(p/q + a)(p^2/q^2 + 2ap/q + 4b)} \times (2\phi + a - \sqrt{a^2 - 4b})^{[(p/q + a)/\sqrt{a^2 - 4b}] - 1} \times (2\phi + a + \sqrt{a^2 - 4b})^{-[(p/q + a)/\sqrt{a^2 - 4b}] - 1} = (q/8)x + n_1; \tag{2.17a}$$

(ii) $a^2 > 4b$ $p/q = -a$,

$$-\frac{1}{2(a^2 - 4b)}\left[\frac{(2\phi + a)}{4(\phi^2 + a\phi + b)} + \frac{1}{2\sqrt{a^2 - 4b}}\ln\left(\frac{2\phi + a - \sqrt{a^2 - 4b}}{2\phi + a + \sqrt{a^2 - 4b}}\right)\right] = (q/8)x + n_2; \tag{2.17b}$$

(iii) $a^2 > 4b$, $p/q = -a \pm \sqrt{a^2 - 4b}$,

$$\frac{1}{\pm 4\sqrt{a^2 - 4b}}\left[\frac{1}{\{(2\phi + a) \pm \sqrt{a^2 - 4b}\}^2} + \frac{1}{(\pm\sqrt{a^2 - 4b})}\left\{\frac{1}{2\phi + a \pm \sqrt{a^2 - 4b}} + \frac{1}{2\sqrt{a^2 - 4b}}\ln\left(\frac{2\phi + a - \sqrt{a^2 - 4b}}{2\phi + a + \sqrt{a^2 - 4b}}\right)\right\}\right] = (q/8)x + n_3; \tag{2.17c}$$

(iv) $a^2 = 4b$, $p/q \neq -a$,

$$\frac{1}{4(p/q + a)^3}[(2\phi + a)(2\phi + 3a + 2p/q) + 2(p/q + a)^2]e^{-2(p/q + a)/2(\phi + a)} = (q/2)x + n_4; \tag{2.17d}$$

(v) $a^2 = 4b$, $p/q = -a$,

$$\phi = \tfrac{1}{2}[(-3q/2)x + n_5]^{-1/3} - a/2; \tag{2.17e}$$

(vi) $a^2 < 4b$, $p/q \neq -a$,

$$\exp\left[\frac{2(2\phi + a)}{\sqrt{4b - a^2}}\tan^{-1}\left(\frac{2\phi + a}{\sqrt{4b - a^2}}\right)\right] \times [4(p/q + a)(p^2/q^2 + 2ap/q + 4b)]^{-1} \times \left[\frac{2(p/q + a)^2}{(4b - a^2)}\cos^2\left\{\tan^{-1}\left(\frac{2\phi + a}{\sqrt{4b - a^2}}\right)\right\} + \frac{(p/q + a)}{\sqrt{4b - a^2}}\sin 2\left\{\tan^{-1}\left(\frac{2\phi + a}{\sqrt{4b - a^2}}\right)\right\} + 1\right] = (q/8)x + n_6; \tag{2.17f}$$

(vii) $a^2 < 4b$, $p/q = -a$,

$$\frac{1}{2(4b - a^2)}\left[\frac{2\phi + a}{4(\phi^2 + a\phi + b)} + \frac{1}{\sqrt{4b - a^2}}\tan^{-1}\left(\frac{2\phi + a}{\sqrt{4b - a^2}}\right)\right] = (q/8)x + n_7. \tag{2.17g}$$

In the above $n_1, n_2, n_3, \ldots, n_7$ are arbitrary constants. It may be noted that except for the case (v) the solutions for ϕ given by (2.17e) are in transcendental forms preventing us from obtaining algebraically ϕ as a function of x.

We next use relations (2.13) and (2.16) to eliminate α and the derivatives of α and β from the field equations and arrive at the equation

$$(2\omega + 3)\frac{\psi'^2}{\psi^2} = (p^2/q^2 + 2ap/q + 4b)\frac{\phi'^2}{(\phi^2 + a\phi + b)^2}. \tag{2.18}$$

Now with ϕ given by (2.17) we can solve Eq. (2.18) for ψ, at least in principle, provided an exact functional form of $\omega(\psi)$ is known. It then follows from relations (2.13) and (2.16) that it is formally possible to obtain the explicit forms for the metric.

Again with Eqs. (2.13) and (2.16)–(2.18) at hand it is not difficult to discuss the corresponding cases in the absence of the scalar field. Now in view of (2.18), ψ' vanishes when $(p^2/q^2 + 2ap/q + 4b) = 0$. Bearing this in mind while eliminating ϕ between (2.16a) and (2.17c) one arrives at the expression for $a^2 > 4b$, $p/q = -a \pm \sqrt{a^2 - 4b}$,

$$e^{2\beta}\psi + [2/(p/q+a)]e^{\beta}\psi^{1/2} + [2/(p/q+a)^2]\ln[1-(p/q+a)e^{\beta}\psi^{1/2}] = 2(p/q+a)(qx+8n_3). \quad (2.19)$$

On the other hand using (2.17e) in (2.13) and (2.16b), respectively, one obtains for $a^2 = 4b$ and $p/q = -a$

$$e^{2\alpha} = (1/4\psi)(-(3q/2)x + n_5)^{-2/3}, \quad (2.20)$$

$$e^{2\beta} = (1/\psi)(-(3q/2)x + n_5)^{2/3}. \quad (2.21)$$

Equations (2.19)–(2.21) with $\psi = \text{const}$ can be easily recognized as Patnaik's results for the electrovac in general relativity.[10] It is not difficult to see that here $\beta = \alpha$ leads us to $\phi' = 0$, which means that field equations in this case do not admit an electrovac solution.

III. GRAVITATIONAL FIELDS DUE TO AN UNCHARGED SOURCE

In the absence of the electric field, $\phi' = 0$ and Eqs. (2.8) and (2.11) yield, respectively, on integration

$$e^{2\beta}\psi = cx + d, \quad (3.1)$$

and

$$e^{2\beta}\psi(2\alpha' + \psi'/\psi) = f, \quad (3.2)$$

c, d, and f being arbitrary constants. Now combining (3.1) and (3.2) one has

$$2\alpha' + \psi'/\psi = f/(cx+d), \quad (3.3)$$

which in turn integrates to give, after suitable coordinate transformations absoring the integration constant,

$$e^{2\alpha}\psi = (cx+d)^{f/c}. \quad (3.4)$$

Now eliminating the derivatives of α and β by means of (3.1) and (3.4) one obtains from the field equations

$$(2\omega+3)\frac{\psi'^2}{\psi^2} = \left(1+\frac{2f}{c}\right)\frac{c^2}{(cx+d)^2}. \quad (3.5)$$

It follows from Eq. (3.5) that the constant $(1+2f/c)$ and $(2\omega+3)$ must have the same sign. In what follows we shall, however, choose $(2\omega+3)>0$. This choice corresponds to the scalar field with positive energy density of the contribution from the scalar field.

By virtue of the above analysis integration of the field equations is essentially reduced to the task of solving Eq. (3.5) with a suitable functional form of $\omega(\psi)$. Equations (3.1) and (3.4) then give the explicit forms for the metric provided the integration of (3.5) yields ψ as an exact function of x.

Different theories suggest different forms of ω as functions of the scalar field (see Ref. 8 and references therein). We now integrate Eq.(3.5) and give explicit forms for the metric by means of (3.1) and (3.4) in some of these theories.

A. Brans–Dicke theory: $\omega = \text{const}$

Here,

$$\psi = \psi_0(cx+d)^k, \quad (3.6a)$$

$$e^{2\beta} = \psi_0^{-1}(cx+d)^{1-k}, \quad (3.6b)$$

$$e^{2\alpha} = \psi_0^{-1}(cx+d)^{f/c-k}, \quad (3.6c)$$

with $k^2 = [(1+2f/c)/(2\omega+3)]$ and ψ_0 an arbitrary constant. In this context it may be mentioned that the solutions given by Reddy[11] in this case are erroneous owing to use of an incorrect set of field equation in the work.

B. Barkar theory: $\omega = (4-3\psi)/2(\psi-1)$

Here,

$$\psi = \sec^2[\ln\{A(cx+d)^l\}], \quad (3.7a)$$

$$e^{2\beta} = (cx+d)\cos^2[\ln\{A(cx+d)^l\}], \quad (3.7b)$$

$$e^{2\alpha} = (cx+d)^{f/c}\cos^2[\ln\{A(cx+d)^l\}], \quad (3.7c)$$

with $l^2 = \frac{1}{4}(1+2f/c)$ and A an arbitrary constant. Solutions (3.7) were previously given by Banerjee and Duttachoudhury.[3]

C. Schwinger theory: $\omega = (1-3n\psi)/2n\psi$, n being a constant $[n>0]$

Here,

$$\psi = [\ln\{B(cx+d)^m\}]^{-2}, \quad (3.8a)$$

$$e^{2\beta} = (cx+d)[\ln\{B(cx+d)^m\}]^2, \quad (3.8b)$$

$$e^{2\alpha} = (cx+d)^{f/c}[\ln\{B(cx+d)^m\}]^2, \quad (3.8c)$$

with $m^2 = \frac{1}{4}[n(1+2f/c)]$ and B an arbitrary constant.

D. Models with curvature coupling: $\omega = 3\psi/2(1-\psi)$

Here,

$$\psi = 4D(cx+d)^n[1+D(cx+d)^n]^{-2}, \quad (3.9a)$$

$$e^{2\beta} = (1/4D)(cx+d)^{1-n}[1+D(cx+d)^n]^2, \quad (3.9b)$$

$$e^{2\alpha} = (1/4D)(cx+d)^{f/c-n}[1+D(cx+d)^n]^2. \quad (3.9c)$$

In the above $n^2 = 1/3(1+2f/c)$ and D is an arbitrary constant. It may be pointed out here that any other scalar defined to be proportional to $(1-\psi)^{1/2}$ will satisfy a conformally invariant equation $(\Box + \frac{1}{6}R)\gamma = 0$ (see Penrose[12]).

IV. COSMOLOGICAL SOLUTION

The static solutions (3.6)–(3.9) can yield special cases of Bianchi I homogeneous solutions when one performs a set of complex coordinate transformations

$$x \to i(t - d/c), \quad t \to ix, \quad c \to -ic, \quad f \to -if.$$

In the following we give the final forms of a few cosmological solutions obtained in the above manner.

A. Brans–Dicke theory ($\omega = \text{const}$)

Here,

$$e^{2\beta} = \psi_0^{-1}c^{1-k}t^{1-k}, \quad (4.1a)$$

$$e^{2\alpha} = \psi_0^{-1}c^{(f/c-k)}t^{(f/c-k)}, \quad (4.1b)$$

$$\psi = \psi_0 c^k t^k, \quad (4.1c)$$

$$R^3 = e^{(\alpha+2\beta)} = \psi_0^{-3/2}c^{k_1}t^{k_1}, \quad (4.1d)$$

and the Ricci scalar $g^{\mu\nu}R_{\mu\nu}$ is given by

$$g^{\mu\nu}R_{\mu\nu} = -\psi_0 c^{2+k_2}k^2\omega t^{k_2}, \quad (4.1e)$$

where

$$k^2 = (1+2f/c)(2\omega+3)^{-1},$$

$$k_1 = \tfrac{1}{4}[2\omega k^2 + 3(k-1)^2],$$

$$k_2 = -[(\omega+1)k^2 + \tfrac{1}{2}(k-1)^2 + 1].$$

In the above situation it is very clear for $\omega > 0$. We note that $R \to 0$ and $g^{\mu\nu}R_{\mu\nu} \to \infty$ as $t \to 0$. The model is monotonically expanding or contracting without having a turning point. At the initial epoch ψ is either vanishingly small or infinitely large depending on the sign of the constant K. The same conclusion was also previously arrived at by Matzner and Ryan.[13]

B. Barkar theory

The cosmological solutions corresponding to static solutions (3.7) are

$$e^{2\beta} = ct\cos^2(l\ln t + \lambda), \tag{4.2a}$$

$$e^{2\alpha} = c^{f/c}t^{f/c}\cos^2(l\ln t + \lambda), \tag{4.2b}$$

$$\psi = \sec^2(l\ln t + \lambda), \tag{4.2c}$$

$$R^3 = e^{(\alpha+2\beta)} = c^{l_1}t^{l_1}\cos^3(l\ln t + \lambda), \tag{4.2d}$$

and

$$g^{\mu\nu}R_{\mu\nu} = -8c^{2(1-l_1)}l^2t^{-2l_1}\sec^2(l\ln t + \lambda) \tag{4.2e}$$

with

$$l^2 = \tfrac{1}{4}(1 + 2f/c), \quad \lambda = \ln(Ac^l), \quad l_1 = (l^2 + 3/4).$$

Here we see that as $t \to 0$, we have $R^3 \to 0$, i.e., the spatial volume of the model is zero and at a subsequent time when $\sec(l\ln t + \lambda) \to 0$, we have again $R^3 \to 0$. There is a turning point inbetween at $t = \exp[(1/l)\{\tan^{-1}(l_1/3l) - \lambda\}]$, where the spatial volume attains a maximum value. It is also seen that ψ has a minimum value that is $\psi = 1$ at $t = \exp(-\lambda/l)$. At this epoch, since the parameter ω becomes infinitely large, there is little difference from the corresponding situation in general relativity. One may also readily verify that $\psi > 0$ when $R = 0$. In other words the expansion halts after the scalar field crosses its minimum.

C. Schwinger theory

In view of the static solutions (3.8) the corresponding cosmological solutions are

$$e^{2\beta} = T[\ln(BT^m)]^2, \tag{4.3a}$$

$$e^{2\alpha} = T^{f/c}[\ln(BT^m)]^2, \tag{4.3b}$$

$$\psi = [\ln(BT^m)]^{-2}, \tag{4.3c}$$

$$R^3 = e^{(\alpha+2\beta)} = T^{(1+f/2c)}[\ln(BT^m)]^3, \tag{4.3d}$$

and the Ricci scalar

$$g_{\mu\nu}R_{\mu\nu} = (-2/n)m^2c^2T^{-2(1+f/2c)}[\ln(BT^m)]^{-2}, \tag{4.3e}$$

where the symbol T stands for ct.

Here we observe that $R^3 \to 0$ and $g^{\mu\nu}R_{\mu\nu} \to \infty$ as t approaches 0 and also $1/(CB^{1/m})$ and $\dot{R} = 0$ at an intermediate stage. It can be easily shown that for $m < 0$, $\ddot{R} < 0$ at the turning point $\dot{R} = 0$ and thus we have a maximum of the spatial volume at this epoch. Further the scalar field ψ approaches zero or infinity as the spatial volume vanishes, i.e., $R^3 \to 0$.

D. Curvature coupling

Bianchi I type cosmological solutions are obtained from (3.9) in the form

$$e^{2\beta} = \tfrac{1}{4}D^{-1}T^{(1-n)}[1 + DT^n]^2, \tag{4.4a}$$

$$e^{2\alpha} = \tfrac{1}{4}D^{-1}T^{(f/c-n)}[1 + DT^n]^2, \tag{4.4b}$$

$$\psi = 4DT^n[1 + DT^n]^{-2}, \tag{4.4c}$$

$$R^3 = e^{(\alpha+2\beta)} = \frac{1}{8}D^{3/2}T^{(3/4)(n+1)^2}\left[\frac{1 + DT^n}{DT^n}\right]^3, \tag{4.4d}$$

and $g^{\mu\nu}R_{\mu\nu} = 0$ with T written for ct.

The above model has a monotonic time behavior indicating that there is no turning point anywhere in the evolution from $t = 0$ to $t = \infty$. One should note that the vanishing of ψ at any instant indicates the infinitely large value of the gravitational constant, which, however, ensures the existence of singularity.

We have seen that in none of the cosmological models cited above the existence of singularity can be avoided. For all the above models expressed in the metrics given in (4.5)–(4.8) if one puts $\dot{\psi} = 0$ one finds immediately that the line element in each case reduces to the form

$$ds^2 = (ct)^{-1/2}(dt^2 - dx^2) - ct(dy^2 + dz^2). \tag{4.5}$$

Now introducing new time and space coordinates by

$$\bar{t} = \tfrac{4}{3}c^{-1/4}t^{3/4}, \quad \bar{x} = (3c/4)^{-1/3}x,$$

$$\bar{y} = (3c/4)^{1/3}y, \quad \bar{z} = (3c/4)^{1/3}z,$$

and then dropping bars over coordinates it is not difficult to show that (4.5) reduces finally to the form

$$ds^2 = dt^2 - t^{-2/3}dx^2 - t^{4/3}(dy^2 + dz^2). \tag{4.6}$$

The above metric can easily be recognized as a special case of the Kasner universe.[14]

ACKNOWLEDGMENT

One of the authors (N.B.) is grateful to the Council of Scientific and Industrial Research for the award of a Research Fellowship.

[1] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[2] K. Nordtvedt, Jr., Astrophys. J. **161**, 1059 (1970).
[3] A. Banerjee and S. B. Duttachoudhury, J. Math. Phys. **21**, 2799 (1980).
[4] B. M. Barkar, Astrophys. J. **219**, 5 (1978).
[5] A. Banerjee and N. O. Santos, Phys. Rev. D. **23**, 2111 (1981).
[6] A. Banerjee and N. O. Santos, Nuovo Cimento B **71**, 55 (1982).
[7] V. U. M. Rao and D. R. K. Reddy, Gen. Rel. Grav. **14**, 1017 (1982).
[8] N. Van den Bergh, Gen. Rel. Grav. **14**, 17 (1982).
[9] P. A. Amundsen and Ø. Grøn, Phys. Rev. D **27**, 1731 (1983).
[10] S. N. Patnaik, Proc. Cambridge Philos. Soc. **67**, 127 (1970).
[11] D. R. K. Reddy, J. Phys. A; Math. Gen. **10**, 55 (1977).
[12] R. Penrose, Proc. R. Soc. London Ser. A **284**, 159 (1965).
[13] R. A. Matzner and M. P. Ryan, Nuovo Cimento B **14**, 161 (1973).
[14] A. K. Raychaudhuri, *Theoretical Cosmology* (Clarendon, Oxford, 1979).

Probability measures and Hamiltonian models on Bethe lattices. I. Properties and construction of MRT probability measures

Fulvio Peruggi
Gruppo Nazionale di Struttura della Materia and Istituto di Fisica Teorica, Mostra d'Oltremare, pad. 19, 80125 Napoli, Italy

(Received 26 July 1983; accepted for publication 27 January 1984)

The properties of one-step Markov, rotationally and m-step ($m = 1$ or 2) translationally invariant (MRT) probability measures on q-state-site (qSS) Bethe lattices are studied. A theorem is proven, which completely defines such measures in terms of $m(q^2 + q)$ fundamental probabilities. These are explicitly calculated for any MRT–qSS Hamiltonian model. As a consequence of our approach, the dychotomy between alternative solutions of Hamiltonian models on Bethe lattices is solved.

PACS numbers: 05.50 + q, 02.50. + s, 64.60. − i

I. INTRODUCTION

Hamiltonian models on regular lattices are the subjects of great attention in statistical mechanics, since they are schematizations which retain the most relevant physical properties of many real systems. Despite their apparent simplicity, such models are not exactly solved, except for few cases. This is due to the problem of taking into account correlation effects, which is an enormous task for systems with only nearest-neighbor interactions, too. These difficulties justify the fast development of new approximation methods and the continuous refinements of old ones in order to extract useful information about the models under investigation. An alternative strategy is the modification of the topological structure of lattices, provided this preserves the main physical features of the original systems and (possibly) gives exact solutions. This further schematization can be done by studying Hamiltonian models on hierarchical lattices and Bethe lattices. As regards to the former ones, which can be obtained through iterated decoration and miniaturization of an initial structure, refer to the recent papers by Griffiths and Kaufman,[1] and references therein. Although finite portions of a Bethe lattice could be seen as hierarchical lattices[1] we shall follow a distinct approach here.

The main feature of Bethe lattices is their thin structure. Only one path joins every pair of sites, so that correlations can be taken into account exactly for any m-step Markov system.[2] Due to this property, Hamiltonian models on this kind of lattices were studied in detail by many authors.[3–21] It is known (see, e.g., Ref. 20) that distinct results can be obtained for the same system in the thermodynamic limit: they are usually referred to as (i) Cayley tree solutions and (ii) Bethe lattice solutions. These labels are somewhat misleading, and it is important to realize that they do not refer to distinct topological objects (see the Appendix A), but to distinct ways of calculating the per site values of extensive functions in the thermodynamic limit. In fact, the huge number of surface sites in any finite tree gives rise to non-negligible surface effects which are retained by type (i) approaches and neglected by type (ii) approaches. From the mathematical point of view type (i) methods seem to be correct, while the others are criticizable since they do not give formal justifications for the rejection of surface effects. From the physical point of view the situation is reversed, since residual surface effects in the thermodynamic limit can be seen as undesired contributions from a "ghost surface" which actually does not exist in the infinite system.

The previously described dichotomy between alternative solutions is related to the concept of convergence in the sense of Van Hove.[22] Roughly speaking, the per site free energy of an infinite system is not defined univocally, but depends on the sequence of finite subsystems used to take the thermodynamic limit: Van Hove's convergence condition is a criterion to select those sequences which give the "physically good" result. In this connection, Bethe lattices are very singular objects, since the use of this criterion in its standard form (see Sec. VII) implies that no sequence of finite trees converges to an infinite tree in the sense of Van Hove. As a consequence two interpretations arise: (I) Van Hove's convergence condition is considered meaningless on Bethe lattices, and therefore it may be disregarded [as it is done by the authors who follow type (i) approaches]; (II) this criterion is considered meaningful on Bethe lattices, which implies that "good" solutions perhaps can be obtained by "artificial" deletion of surface effects [this way is followed by all type (ii) approaches]. From the formal point of view this double interpretation is unsatisfactory, since it seems to imply that statistical mechanics cannot be extended to Hamiltonian models on Bethe lattices without uncertainty.

It is our aim to remark that this uncertainty may be removed, and that a new definition and a univocal interpretation of Van Hove's convergence condition may be found, by means of some results of rigorous statistical mechanics, which have been recently collected, generalized, and unified by Ruelle.[23] The main point of interest of Ruelle's approach, in the present case, is the fact that the thermodynamic limit of probability measures always exists on any discrete state countable lattice, and does not depend on the choice of the sequence of finite subsystems used to take the limit. In view of this property, the probability measure approach can be seen as the canonical method for the solution of Hamiltonian models on lattices, and Van Hove's convergence condition can be interpreted as a test of compatibility of the thermodynamic limits of the free energy with that of Gibbs probability measures. This subject is considered explicitly by Ruelle for Hamiltonian models on hypercubic lattices, and will be studied here on Bethe lattices by means of formal applications of

 0022-2488/84/113303-13$02.50

the probability measure approach. In this paper we will also prove that, owing to the simple structure of Bethe lattices, the probability measure approach can be applied explicitly, and gives the complete analytical solution of a large class of Hamiltonian models [including as very special cases the Ising model, the Potts model, the vector (or planar) Potts model, the Ashkin–Teller model, the $Z(q)$ (or clock) model, and all the previous models with annealed site dilution]. To do this, some tools of measure theory, probability theory, and graph theory will be applied.

We first study the properties of a one-step Markov (M) probability measure μ defined on a q-state-site (qSS) Bethe lattice: $q = 2,3,\ldots$. We prove a decomposition rule which completely defines μ in terms of two classes of elementary objects ("site" probabilities and "bond" probabilities). If μ is also rotationally and one- or two-step translationally invariant (RT) on the Bethe lattice, the distinct elementary probabilities reduce to $m(q^2+q)$ fundamental probabilities, where $m = 1$ ($m = 2$) if μ is one-step (two-step) translationally invariant. Then we introduce the most general Hamiltonian model characterized by MRT interactions on a qSS Bethe lattice, and show that the physical expectation for RT Gibbs probability measure(s) describing pure phase(s) of the system in the thermodynamic limit, corresponds to the mathematical requirement that a certain piecewise contracting property holds. When this takes place (which, e.g., is the case for the ferromagnetic or antiferromagnetic Ising and Potts models) the fundamental probabilities are easily calculated, and allow us to construct the measure μ relative to each phase that appears in the system. Furthermore, any thermodynamic limit characterized by RT breakdown is shown to be a "mixture," i.e., a probability measure which describes phase mixing. One can prove that the results obtained are the same (but much more detailed) as those given by the Bethe–Peierls cluster approximation[24] on standard lattices, while general agreement with those of the previously described type (ii) methods can be checked easily. These results and the proof that the thermodynamic limit of the free energy [as it is done in type (i) approaches] is affected by a topological discrepancy, show that Van Hove's convergence condition is extendible in its standard form to Bethe lattices, and is meaningful on these graphs. All these subjects are contained in the present paper.

Although the knowledge of μ formally solves every configurational problem on the qSS Bethe lattice, the thermal properties of the model under investigation could have no simple relations with Gibbs probability measures. This problem is entirely solved by finding the "correct" free energy of the infinite system in terms of the fundamental probabilities. One can do that, and actually one can find the correct thermodynamic limit of any extensive function on Bethe lattices, through a localization procedure which takes advantage of the above-mentioned decomposition property of μ. In fact, as regards to the thermodynamic limit of local quantities, it is irrelevant whether Van Hove's convergence condition holds or does not. The subjects above and other topics (e.g., the solution of all polychromatic MRT-correlated-site/random-bond percolation models) are collected in a following paper[25] (which we call paper II from now on).

The outline of this paper is as follows. In Sec. II the general terminology is established. A theorem concerning the definition of a MRT probability measure μ on the qSS Bethe lattice in terms of the above-mentioned fundamental probabilities is proven in Sec. III. The M probability measures on finite trees are studied in Sec. IV, while the problem of the construction of M Gibbs probability measures on finite trees is considered in Sec. V. The thermodynamic limit is treated in Sec. VI. Conclusions and comparisons with other methods of solution are given in Sec. VII. Useful definitions concerning trees, and their relations with other terminologies, are reported in Appendix A. The infinite-dimensional character of Bethe lattices, and the "classical" character of related solutions are briefly considered in Appendix B. Specialization of our general formalism in order to obtain some relevant Hamiltonian models is described in Appendix C. A proof concerning arguments in Sec. VI is given in Appendix D.

Finally, we remark that preliminary reading of Secs. I and VIII of paper II could be very useful for physicists who are more interested in the applications of the present procedure, rather than its mathematical justification.

II. NOTATION AND GENERAL REMARKS

Let us consider an infinite, connected, and locally finite graph[26] $G = (V,E)$ with a countable set of sites. We suppose that every site $i \in V$ assumes q states which will be labeled by the variable $\nu_i = 1,2,\ldots,q$. Each configuration of V can be represented by the collection $\{V^{(1)}, V^{(2)},\ldots,V^{(q)}\}$, where each set $V^{(r)}$ contains all the sites $i \in V$ such that $\nu_i = r$; $r = 1,\ldots,q$. Therefore, the set of all the configurations of V is in one–one correspondence with the set $\|V\|$ of all ordered partitions of V in q sets. Let $A_1, A_2, \ldots, A_q$ be finite sets of sites of G. The local event $\{A_1, A_2, \ldots, A_q\}$ on V will be defined as the set of all the configurations in $\|V\|$ which attribute the state $\nu_i = r$ to every site $i \in A_r$; $r = 1, \ldots, q$. When there is no confusion we use the notation $\mathscr{E}_A$ to denote any local event $\{A_1, \ldots, A_q\}$ with basis $A \equiv \cup_{r=1}^{q} A_r$. Notice that, if $A_s \cap A_t \neq \varnothing$ at least for one pair of indices $s \neq t$, $\{A_1, \ldots, A_q\}$ is incompatible with all the configurations of V and is the null event $\varnothing$. The global event containing all the configurations of V is $\{\varnothing,\varnothing,\ldots,\varnothing\} = \|V\|$. The following inclusion rule and composition rule follow by definition:

$$\{A_1,\ldots,A_q\} \supset \{A'_1,\ldots,A'_q\} \Leftrightarrow A_r \subset A'_r,\ r = 1,\ldots,q,$$

$$\{A_1,\ldots,A_q\} \cap \{A''_1,\ldots,A''_q\} = \{A_1 \cup A''_1,\ldots,A_q \cup A''_q\}. \quad (1)$$

Let V_0 be a finite nonvoid subset of V. Using relations (1) it is easy to see that the set of all local events $\mathscr{E}_A$ such that $A \subset V$ ($A \subseteq V_0$) is a semiring[27] R (R_0), i.e., the simplest collection of sets where (probability) measures[27,28] can be defined. We may also consider generalized local and nonlocal events on V (generalized local events on V_0), i.e., the elements of the smallest σ-field[27] $F \supset R$ [$F_0 \supset R_0$ (actually R_0 contains a finite number of elements, and F_0 is a ring)[27]]. This follows by the fact that any measure μ^* defined on a semiring R^* can be extended to the σ-field $F^* \supset R^*$ of all μ^*-measurable events.[27] Notice that the collection $\|V_0\|$ of all local events $\mathscr{E}_{V_0} \in R_0$ such that $\mathscr{E}_{V_0} \neq \varnothing$ may be seen as the set of all the configurations of V_0. Obviously, the knowledge of $\mu_0(\mathscr{E}_{V_0})$

for every $\mathscr{E}_{V_0} \in \|V_0\|$ completely defines the probability measure μ_0 on R_0.

Up to now the bonds in G were not used: we need them to introduce a metric in V. We define the length of a walk[26] in G as its number of bonds, and the distance between any pair i,j of sites in V as the length of the shortest walk(s) connecting i to j in G. Given a finite nonvoid $X \subset V$ and a fixed number $m > 0$, we define the internal boundary ΔX (the external boundary ∂X) of X as the set of all the sites in X ($V - X$) such that their distance from at least one point in $V - X(X)$ is less than or equal to m. Let Y be another finite subset of V such that $X \cap Y = \emptyset$ and $\partial X \subseteq Y$; and let $\mathscr{E}_X$, $\mathscr{E}_{\partial X}$, $\mathscr{E}_Y$ be local events such that $\mathscr{E}_Y \subseteq \mathscr{E}_{\partial X}$. We say that the probability measure μ on R is m-step Markov[2] if and only if the equality between conditional probabilities[28]

$$\mu(\mathscr{E}_X | \mathscr{E}_Y) = \mu(\mathscr{E}_X | \mathscr{E}_{\partial X}) \tag{2}$$

is verified for every choice of the sets of sites and the associated events which satisfies the above-mentioned conditions, and does not hold if m is decreased by 1. The same definition can be done for μ_0 on R_0 by substituting V_0 to V.

We are interested in the construction of Gibbs probability measures on R associated to Hamiltonian models on G. To this end, let us introduce an interaction I in the system, which associates a certain energy to every local configuration of V. In other terms I may be considered as a real function defined on the elements of R, which is zero on the global event. The general theory requires that I is properly normed.[23] This will be satisfied here by assuming that $I(\mathscr{E}) = 0$ for every $\mathscr{E} \in R$ whose basis contains at least two sites at a distance larger than a fixed number $m > 0$. Notice that the hypothesis above also implies that any Gibbs probability measure on $R_0(R)$ generated by I is m-step Markov. Given two local events $\mathscr{E}_{V_0}$, $\mathscr{E}_{\partial V_0}$, the Hamiltonian on V_0 relative to $\mathscr{E}_{V_0}$ is defined as[29]

$$-\beta \mathscr{H}_0(\mathscr{E}_{V_0}) \equiv \sum_{\mathscr{E} \in R: \mathscr{E} \supseteq \mathscr{E}_{V_0}} I(\mathscr{E}), \tag{3}$$

while the interface Hamiltonian on ΔV_0, relative to the "boundary condition" $\mathscr{E}_{\partial V_0}$, is defined as[29]

$$-\beta \Delta \mathscr{H}_0(\mathscr{E}_{V_0}; \mathscr{E}_{\partial V_0}) \equiv \sum_{\mathscr{E} \in R: \mathscr{E} \supseteq \mathscr{E}_{V_0} \cap \mathscr{E}_{\partial V_0}; \mathscr{E} \not\supseteq \mathscr{E}_{V_0}; \mathscr{E} \not\supseteq \mathscr{E}_{\partial V_0}} I(\mathscr{E}). \tag{4}$$

The Gibbs probability measure μ_0 on R_0 is obtained in the conventional way by assuming the following probability for every $\mathscr{E}_{V_0} \in R_0$:

$$\mu_0(\mathscr{E}_{V_0}) \equiv (1/Z_0) \exp[-\beta \mathscr{H}_0(\mathscr{E}_{V_0})], \tag{5}$$

$$Z_0 \equiv \sum_{\mathscr{E}_{V_0} \in R_0} \exp[-\beta \mathscr{H}_0(\mathscr{E}_{V_0})].$$

In a similar way we define the (conditional) Gibbs probability measure μ_{0c} on R_0 (given that $\mathscr{E}_{\partial V_0}$ is an event with probability 1):

$$\mu_{0c}(\mathscr{E}_{V_0} | \mathscr{E}_{\partial V_0}) \equiv (1/Z_{0c}) \exp[-\beta \mathscr{H}_0(\mathscr{E}_{V_0}) - \beta \Delta \mathscr{H}_0(\mathscr{E}_{V_0}; \mathscr{E}_{\partial V_0})], \tag{6}$$

$$Z_{0c} \equiv \sum_{\mathscr{E}_{V_0} \in R_0} \exp[-\beta \mathscr{H}_0(\mathscr{E}_{V_0}) - \beta \Delta \mathscr{H}_0(\mathscr{E}_{V_0}; \mathscr{E}_{\partial V_0})].$$

Let us consider a sequence $\{V_n\}_{n=1}^{\infty}$ of finite subsets of V (ordered by inclusion) such that $\cup_{n=1}^{\infty} V_n = V$. Furthermore, let μ_n be a probability measure on R_n, and let n_A be an index such that $A \subset V_n$ for every $n > n_A$. The following can be shown.[23]

Proposition 1: One can choose a subsequence $\{V_{n'}\}_{n'=1}^{\infty}$ of $\{V_n\}_{n=1}^{\infty}$ such that the limit

$$\lim_{n'_A < n' \to \infty} \mu_{n'}(\mathscr{E}_A) \equiv \mu(\mathscr{E}_A)$$

exists for every $\mathscr{E}_A \in R$, and defines the probability measure μ on R.

We say that μ is a thermodynamic limit of the μ_n's, and use the conventional notation

$$\lim_{n \to \infty} \mu_n = \mu. \tag{7}$$

When every μ_n is a probability measure given by (6), we obtain Gibbs probability measures (≡Gibbs states) describing the infinite system governed by the interaction I. In this case the need of subsequences in Proposition 1 is better understood by noting that distinct limits may be obtained according to the selected (subsequence $\{\mu_{n'}\}_{n'=1}^{\infty}$ and the corresponding) boundary conditions. The closed hull of the thermodynamic limits found in such a way is the set $\mathscr{K}_I$ of all Gibbs states[23] [in fact it contains, as well, the limits obtained when every μ_n is given by (5)]. The set $\mathscr{K}_I$ is convex and compact, and is a simplex.[23] In other terms, any nonextremal element of $\mathscr{K}_I$ has a unique decomposition in terms of extremal Gibbs states, i.e., it is a convex or integral combination of these probability measures.

III. PROPERTIES OF MRT PROBABILITY MEASURES ON *q*SS BETHE LATTICES

A Bethe lattice $L = (V,E)$ is an infinite connected tree[26] whose sites have the same coordination number $\sigma + 1$. We shall identify L with the graph G of the preceding section, in order to study the special properties that any one-step Markov probability measure (not necessarily a Gibbs state) μ on R exhibits is such a case. In fact, the following theorem holds.

Theorem 1: If μ is MRT on L, i.e., if it is one-step Markov (M), and invariant under any elementary rotation and m-step translation ($m = 1$ or 2) that carry L onto itself (RT), then μ is completely defined by a set of $m(q^2 + q)$ fundamental probabilities.

To begin, we give a formal definition of m-step translations and elementary rotations of L. The first step in our procedure is the construction of a proper labeling for every site of L. To do this we choose a reference site $i \in V$ and consider a partition of V in shells surrounding i. The l th shell contains all the sites $h \in V$ at distance l [$\equiv l(i,h) \equiv$number of bonds of the unique walk connecting i to h] from i; $l = 1,2,\ldots$. Let us consider a numerical representation using $\sigma + 2$ digits, and assign the digit 0 to i, and the digits $1,2,\ldots,[\sigma]$, $[\sigma + 1]$ to the sites of the first shell. Then proceed in the following way: (1) execute steps (2)–(4) for $l = 1,2,\ldots$; (2)

choose a site k belonging to the l th shell surrounding i; (3) assign the digits 1,..., $[\sigma]$ to the sites of the $(l+1)$-th shell which are adjacent to k; and (4) repeat steps (2)–(3) for every k belonging to the l th shell. Let $i \equiv v_0, v_1, \ldots, v_l \equiv h$ be the sites of the walk $W_l \subseteq L$ connecting i to h, and let $[a_t]$ be the digit assigned to $v_t; t = 1,\ldots,l$. The site h will be labeled with the number $[a_1][a_2]\cdots[a_l]$ which we denote $a_i(h)$ to stress its dependence on the reference site i. Let $i \in V$ and $j \in V$ be two sites of L at distance n. The n-step translation $\mathcal{T}_{ij}$ of L onto itself that carries i into j is defined as the isomorphism that carries every $h \in V$ onto the site $k \in V$ such that $a_i(h) = a_j(k)$. Let $\mathcal{P}_{[s][t]}$ be the permutation of the digits 1, ..., $[\sigma+1]$ which interchanges the digits $[s]$ and $[t]$ and leaves unchanged the others. The elementary rotation $\mathcal{R}_{i[s][t]}$ of L around the site i is defined as the isomorphism that carries each $h \in V - \{i\}$ labeled by $a_i(h) \equiv [a_1][a_2]\ldots[a_l]$ onto the site $k \in V - \{i\}$ labeled by $a_i(k) \equiv \mathcal{P}_{[s][t]}([a_1])[a_2]\ldots[a_l]$. Note that, since every bond of L is identified univocally by its terminal sites, bond transformation of L under $\mathcal{T}_{ij}$ and $\mathcal{R}_{i[s][t]}$ is trivially implied by site transformation.

Proof of Theorem 1: For every $h \in V$ let us consider the local events $\{\{h\},\varnothing,\ldots,\varnothing\}, \{\varnothing,\{h\},\varnothing,\ldots,\varnothing\},\ldots, \{\varnothing,\ldots,\varnothing,\{h\}\}$, which will be denoted, respectively, by $\mathscr{E}_{hv_h}; v_h = 1,2,\ldots,q$. For a given local event $\mathscr{E}_A = \{A_1,\ldots,A_q\}$, let $T_A = (V_A, E_A)$ be the smallest connected subgraph of L such that the basis A of $\mathscr{E}_A$ is contained in V_A, let $\sigma_j + 1$ be the coordination number of $j \in V_A$ in T_A, and let $\bar{A} \equiv V_A - A$. It is clear that the local events in $\|\bar{A}\|$ are mutually disjoint, and that their union is the global event $\mathscr{E}_\varnothing = \{\varnothing,\ldots,\varnothing\}$. Then we have

$$\mu(\mathscr{E}_A) = \mu(\mathscr{E}_A \cap \mathscr{E}_\varnothing) = \sum_{\mathscr{E}_{\bar{A}} \in \|\bar{A}\|} \mu(\mathscr{E}_A \cap \mathscr{E}_{\bar{A}}). \tag{8}$$

Since T_A is a tree, and owing to the M property, we may interpret $\mathscr{E}_A \cap \mathscr{E}_{\bar{A}}$ as a branching process.[28] It starts from the source site $i \in V_A$, propagates along the bonds $|hk\rangle \in E_A$ (oriented along the running direction), and stops into the surface sites of T_A (see Appendix A). Then we have

$$\mu(\mathscr{E}_A) = \sum_{\|\bar{A}\|} \mu(\mathscr{E}_{iv_i}) \prod_{|hk\rangle \in E_A} \mu(\mathscr{E}_{kv_k} | \mathscr{E}_{hv_h}) \tag{9a}$$

$$= \sum_{\|\bar{A}\|} \frac{\Pi_{\langle hk\rangle \in E_A} \mu(\mathscr{E}_{hv_h} \cap \mathscr{E}_{kv_k})}{\Pi_{j \in V_A} [\mu(\mathscr{E}_{jv_j})]^{\sigma_j}}, \tag{9b}$$

where, by definition, $z \in \bar{A}_r \cup A_r$ implies $v_z = r (r = 1,\ldots,q)$; while the second equality shows that $\mu(\mathscr{E}_A)$ does not depend on the choice of the source. Therefore, we have proven that the probability measure of any local event in R (therefore of any event in F) can be determined by means of the site probabilities,

$$\mu(\mathscr{E}_{iv_i}), \quad i \in V, \quad v_i = 1,\ldots,q; \tag{10}$$

and the bond probabilities,

$$\left.\begin{matrix} \mu(\mathscr{E}_{kv_k} | \mathscr{E}_{hv_h}) \\ \mu(\mathscr{E}_{hv_h} | \mathscr{E}_{kv_k}) \end{matrix}\right\} \langle hk\rangle \in E, \ v_h, v_k = 1,\ldots,q. \tag{11}$$

Now, let us choose two adjacent reference sites u, v of L, and let us apply the RT property for $m = 1$. By repeated application of one-step translations we have

$$\mu(\mathscr{E}_{ir}) = \mu(\mathscr{E}_{ur}) \equiv p_r, \quad i \in V, \quad r = 1,\ldots,q. \tag{12}$$

By application of the translation that carries k in h, and of an elementary rotation, we prove that $\mu(\mathscr{E}_{ks} | \mathscr{E}_{hr}) = \mu(\mathscr{E}_{hs} | \mathscr{E}_{kr})$, for every $\langle hk\rangle \in E$, and $r,s = 1,\ldots,q$. Then, applying other isomorphisms of L onto itself, we find

$$\mu(\mathscr{E}_{ks} | \mathscr{E}_{hr}) = \mu(\mathscr{E}_{us} | \mathscr{E}_{ur}) \equiv p_{rs}, \quad \langle hk\rangle \in E, \quad r,s = 1,\ldots,q, \tag{13}$$

proving the theorem for $m = 1$. Now, let us consider the case $m = 2$. It is easy to see that two-step translations induce a partition of V in two subsets V^e and V^o such that all the sites of L which are adjacent to $j \in V^e (j \in V^o)$ belong to $V^o (V^e)$. Moreover, V^e and V^o do not mix under rotations of L. We use again two adjacent reference sites u, v of L, such that $u \in V^e$. As in the preceding case, repeated applications of elementary rotations and two-step translations give

$$\left.\begin{matrix} \mu(\mathscr{E}_{ir}) = \mu(\mathscr{E}_{ur}) \equiv p^e_r, & i \in V^e \\ \mu(\mathscr{E}_{ir}) = \mu(\mathscr{E}_{vr}) \equiv p^o_r, & i \in V^o \end{matrix}\right\} r = 1,\ldots,q; \tag{14a}$$

$$\left.\begin{matrix} \mu(\mathscr{E}_{ks} | \mathscr{E}_{hr}) = \mu(\mathscr{E}_{us} | \mathscr{E}_{ur}) \equiv p^e_{rs}, \\ \mu(\mathscr{E}_{hs} | \mathscr{E}_{kr}) = \mu(\mathscr{E}_{us} | \mathscr{E}_{vr}) \equiv p^o_{rs}, \end{matrix}\right\} \begin{matrix} h \in V^e, \langle hk\rangle \in E, \\ r,s = 1,\ldots,q. \end{matrix} \tag{14b}$$

Thus, in the present case, we have $2q + 2q^2$ fundamental probabilities.

Q.E.D.

We also state the following.

Corollary 1: There are at most $q^2 - 1$ independent fundamental probabilities for $m = 2$; they are at most $(q+2)(q-1)/2$ for $m = 1$.

Proof: For $m = 2$ the following relations hold:

$$\sum_{r=1}^{q} p^x_r = 1, \quad x = e,o; \tag{15a}$$

$$\sum_{s=1}^{q} p^x_{rs} = 1, \quad x = e,o; \ r = 1,\ldots q; \tag{15b}$$

$$p^e_r p^e_{rs} = \mu(\mathscr{E}_{ur} \cap \mathscr{E}_{vs}) = p^o_s p^o_{sr}, \quad r,s = 1,\ldots,q. \tag{15c}$$

Using the normalization condition (15b) for $x = o$, and the symmetry condition (15c), we see that

$$p^o_s = \sum_{t=1}^{q} p^e_t p^e_{ts}, \quad r,s = 1,\ldots,q; \tag{16}$$

$$p^o_{sr} = \frac{p^e_r p^e_{rs}}{\Sigma^q_{t=1} p^e_t p^e_{ts}},$$

i.e., o-probabilities depend on e-probabilities. These must satisfy relations (15a)–(15b) for $x = e$, which imply that only $q^2 - 1$ of them are independent. For $m = 1$, relations (15) still hold, provided we delete all the upper indices. In such a case, the symmetry condition can be used only to prove that $p_{sr} = p_r p_{rs}/p_s; r \neq s$. Then we can consider every p_{sr} with $s > r$ as a function of the other $(q+1)q/2 + q$ fundamental probabilities. The normalization conditions allow us to eliminate other $q+1$ terms, thus proving the statement. ■

IV. PROPERTIES OF M PROBABILITY MEASURES ON *q*SS FINITE TREES

Let μ_T be a one-step Markov probability measure on the semiring R_T of all qSS local events defined on a finite tree

$T=(V_T,E_T)$. We choose a reference site $u\in V_T$ and orientate every bond $\langle ij\rangle\in E_T$ in such a way that $l(u,i)<l(u,j)$.

Lemma 1: μ_T is completely defined by the probabilities

$$\mu_T(\mathscr{E}_{ur}), \qquad r=1,...,q-1,$$
$$\mu_T(\mathscr{E}_{js}|\mathscr{E}_{ir}), \quad r=1,...,q,\ s=1,...,q-1, \quad |ij\rangle\in E_T.$$

Proof: It is enough to use relation (9a), provided we require that T_A must contain the site u, and that u is chosen always as source site. Some probabilities are not needed since they are given by obvious normalization conditions. ■

We call $V_T^e(V_T^o)$ the set of all the sites in V_T such that their distance from u is an even (odd) number. Let $v\in V_T^o$ be a site adjacent to u, and $w\in V_T^e$ be a site adjacent to v and distinct from u. We say that μ_T satisfies two-step partial invariance (PI) on T if, for every $|ij\rangle\in E_T$, we have

$$\mu_T(\mathscr{E}_{js}|\mathscr{E}_{ir})=\mu_T(\mathscr{E}_{vs}|\mathscr{E}_{ur}),\ i\in V_T^e,$$
$$\mu_T(\mathscr{E}_{js}|\mathscr{E}_{ir})=\mu_T(\mathscr{E}_{ws}|\mathscr{E}_{vr}),\ i\in V_T^o,$$
$$r,s=1,...,q. \tag{17}$$

One-step PI is recovered if $\mu_T(\mathscr{E}_{vs}|\mathscr{E}_{ur})=\mu_T(\mathscr{E}_{ws}|\mathscr{E}_{vr})$ for every pair of indices. Remark that, as a trivial consequence of Lemma 1 and relation (9a), PI implies that $\mu_T(\mathscr{E}_{hr})=\mu_T(\mathscr{E}_{kr})$ $(r=1,...,q)$ for every pair h,k of sites in T such that $l(u,h)=l(u,k)$. We say that μ_T satisfies two-step (one-step) global invariance (GI) if relations as (14) [(12)–(13)] are verified on T.

Lemma 2: μ_T is m-step GI on T if it is m-step PI and

$$\mu_T(\mathscr{E}_{ur}\cap\mathscr{E}_{vs})=\mu_T(\mathscr{E}_{vs}\cap\mathscr{E}_{wr}),\ m=2,$$
$$\mu_T(\mathscr{E}_{ur}\cap\mathscr{E}_{vs})=\mu_T(\mathscr{E}_{us}\cap\mathscr{E}_{vr}),\ m=1,$$
$$r,s=1,...,q.$$

Proof: The case $m=2$. Note that

$$\begin{aligned}\mu_T(\mathscr{E}_{ur})&=\sum_{s=1}^{q}\mu_T(\mathscr{E}_{ur}\cap\mathscr{E}_{vs})\\&=\sum_{s=1}^{q}\mu_T(\mathscr{E}_{vs}\cap\mathscr{E}_{wr})\\&=\mu_T(\mathscr{E}_{wr}),\ r=1,...,q.\end{aligned} \tag{18}$$

Let $u\equiv z_0, z_1,...,z_d$ be the sites of a walk connecting u to the surface site z_d of T. Suppose the following equality holds:

$$\mu_T(\mathscr{E}_{z_ls})=\mu_T(\mathscr{E}_{z_{l+2}s}),\ s=1,...,q. \tag{19}$$

Then PI and (19) imply

$$\begin{aligned}\mu_T(\mathscr{E}_{z_{l+1}r})&=\sum_{s=1}^{q}\mu_T(\mathscr{E}_{z_ls})\mu_T(\mathscr{E}_{z_{l+1}r}|\mathscr{E}_{z_ls})\\&=\sum_{s=1}^{q}\mu_T(\mathscr{E}_{z_{l+2}s})\mu_T(\mathscr{E}_{z_{l+3}r}|\mathscr{E}_{z_{l+2}s})\\&=\mu_T(\mathscr{E}_{z_{l+3}r}),\ r=1,...,q;\\ \mu_T(\mathscr{E}_{z_2s})&=\mu_T(\mathscr{E}_{ws}),\ s=1,...,q.\end{aligned} \tag{20}$$

From (18) and (20) it follows by induction that (19) is true for $l=0,1,...,d-2$; i.e., relations as (14a) hold. Then GI for μ_T will be achieved if we show that

$$\mu_T(\mathscr{E}_{z_lr}|\mathscr{E}_{z_{l+1}s})=\mu_T(\mathscr{E}_{z_{l+2}r}|\mathscr{E}_{z_{l+1}s}),$$
$$\mu_T(\mathscr{E}_{z_{l+1}s}|\mathscr{E}_{z_lr})=\mu_T(\mathscr{E}_{z_{l+1}s}|\mathscr{E}_{z_{l+2}r}),$$
$$0\leqslant l\leqslant d-2,\ r,s=1,...,q. \tag{21}$$

Remark that $\mu_T(\mathscr{E}|\mathscr{E}')=\mu_T(\mathscr{E}'|\mathscr{E})\mu_T(\mathscr{E})/\mu_T(\mathscr{E}')$. Therefore PI and (19) imply that it is enough to prove the second equality (21) for $l=0$. Without loss of generality we can choose $z_1\equiv v$ and $z_2\equiv w$. In this case, our starting hypothesis and (18) imply the desired result. The case $m=1$ can be proved in the same way. ■

We are interested in the determination of the probability measure μ of the preceding section through a thermodynamic limit. Therefore, we introduce a sequence $\{T_n\}_{n=1}^{\infty}$ of connected section graphs of the Bethe lattice $L=(V,E)$, which are constructed as follows. We choose a reference site $u\in V$ and define T_1 as the tree formed by u, its adjacent sites in L, and the bonds connecting them. The tree T_{n+1} is obtained by adding to $T_n=(V_n,E_n)$ the sites in V belonging to the shell at distance $n+1$ from u, and the bonds in E connecting them to sites in V_n. The whole sequence follows by induction. Note that the sets V_n^e,V_n^o defined as above, automatically induce the partition of the sites of L into subsets V^e,V^o.

Let $\tilde{\mu}_T(\tilde{\mu}_n)$ be the restriction of the probability measure μ on R to the semiring $R_T(R_n)$ of all local events defined on the finite tree $T\subset L$ $(T_n\subset L,\ n=1,2,...)$. Furthermore, let $\{T_{n'}\}_{n'=1}^{\infty}$ be a subsequence of $\{T_n\}_{n=1}^{\infty}$.

Lemma 3: μ is MRT on the Bethe lattice L if $\tilde{\mu}_{n'}$ is MGI on $T_{n'}$ for every index n'.

Proof: Remark that, as a consequence of Theorem 1, any MGI probability measure μ_T on R_T spans a MRT probability measure on R. Therefore, μ is MRT on L if, for every T, $\tilde{\mu}_T$ is characterized by the same set $\{p\}$ of fundamental probabilities. We see that all the $\tilde{\mu}_{n'}$'s correspond to the same $\{p\}$, since, for every index n_0', $\tilde{\mu}_{n_0'}$ can be considered as the restriction to $R_{n_0'}$ of any $\tilde{\mu}_{n'}$ with $n'>n_0'$. Moreover, any $\tilde{\mu}_T$ corresponds to the same $\{p\}$, since, for every T, there exists large enough n_1' such that $T\subset T_{n_1'}$ and $\tilde{\mu}_T$ can be seen as the restriction of $\tilde{\mu}_{n_1'}$. ■

Finally, we look for the structure of the most general interaction I on R which is compatible with a MRT Gibbs probability measure μ. According to the M property, the maximum range of I must be $m=1$, i.e., it is nonzero only on maximal local events $\mathscr{E}_A\in R$ such that A contains only one site (≡external fields), or a pair of adjacent sites (≡nearest neighbor coupling terms). Adding to this the request for rotational and two-step translational invariance, we deduce that the Hamiltonian (3) relative to each tree T_n is given by

$$\begin{aligned}-\beta\mathscr{H}_n(\mathscr{E}_{V_n})&=\sum_{\mathscr{E}_{iv_i}\cap\mathscr{E}_{jv_j}\supset\mathscr{E}_{V_n};\langle ij\rangle\in E_n}I(\mathscr{E}_{iv_i}\cap\mathscr{E}_{jv_j})+\sum_{\mathscr{E}_{iv_i}\supset\mathscr{E}_{V_n}}I(\mathscr{E}_{iv_i})\\&\equiv\sum_{\langle ij\rangle\in E_n}K_{v_iv_j}^{x}+\sum_{i\in V_n}H_{v_i}^{x}.\end{aligned} \tag{22}$$

Here x is e (o) if $i\in V^e$ $(i\in V^o)$;

$$K_{rs}^{e}\equiv K_{sr}^{o},\ r,s=1,...,q; \tag{23}$$

and one-step translational invariance can be recovered by deleting sublattice dependence, i.e., by dropping everywhere the upper indices. As regards to the interface Hamiltonian (4), the M property implies that it contains only coupling

terms associated to pairs of adjacent sites lying, respectively, in ΔV_n and ∂V_n, i.e., to bonds connecting surface sites of T_n to their adjacent sites in $V - V_n$ (≡the perimeter[26] of T_n). Without loss of generality, we will represent them by means of effective fields acting on each site of the surface of T_n:

$$-\beta\Delta\mathscr{H}_n(\mathscr{C}_{V_n};\mathscr{C}_{\partial V_n}) = \sum_{\mathscr{C}_{iv_i}\supset\mathscr{C}_{v_n};\mathscr{C}_{jv_j}\supset\mathscr{C}_{\partial v_n}:\langle ij\rangle\in E} I(\mathscr{C}_{iv_i}\cap\mathscr{C}_{jv_j}) \equiv \sum_{i\in\Delta V_n} B^{(n)}_{iv_i}. \tag{24}$$

For convenience we set $H^{(n)}_{iv_i}\equiv H^x_{v_i} + B^{(n)}_{iv_i}$ for every $i\in\Delta V_n$, and introduce the shortened notation

$$-\beta(\mathscr{H}_n + \Delta\mathscr{H}_n) = \sum_{\langle ij\rangle\in E_n} K^x_{v_iv_j} + \sum_{i\in V_n - \Delta V_n} H^x_{v_i} + \sum_{i\in\Delta V_n} H^{(n)}_{iv_i}. \tag{25}$$

V. CONSTRUCTION OF M GIBBS PROBABILITY MEASURES ON qSS FINITE TREES

Let u be the common central site of the trees of the sequence $\{T_n\}_{n=1}^{\infty}$. Given an index n, we consider a site h of T_n which is distinct from u and does not belong to the surface of T_n. Deletion[26] of h disconnects the tree T_n. Let $T_n(h) = (V_n(h), E_n(h))$ be the tree formed by h, the disconnected parts not containing u, and the bonds of T_n joining them to h. For $h\in\Delta V_n$ we set $T_n(h)\equiv(\{h\},\varnothing)$. Finally, for $h\equiv u$, and for every k such that $l(u,k) = 1$, we introduce the tree $T_n(h,k)$ which we obtain by T_n through deletion of all the sites in $T_n(k)$. We write again $T_n(h)$, instead of $T_n(h,k)$, when there is no confusion. For every $T_n(h)$ we also define

$$\Delta V_n(h)\equiv V_n(h)\cap\Delta V_n,$$

$$\Lambda_n(h,v_h)\equiv \sum_{\|V_n(h)-\{h\}\|} \exp\Bigg(\sum_{\langle ij\rangle\in E_n(h)} K^x_{v_iv_j} + \sum_{i\in V_n(h)-\Delta V_n(h)} H^x_{v_i} + \sum_{i\in\Delta V_n(h)} H^{(n)}_{iv_i}\Bigg). \tag{26}$$

Let $V(h)$ be the set of sites in $T_n(h)$ which are adjacent to h. The hierarchic relations

$$\Lambda_n(h,v_h) = \exp(H^x_{v_h}) \prod_{k\in V(h)} \Bigg(\sum_{v_k=1}^{q} \exp(K^x_{v_hv_k})\Lambda_n(k,v_k)\Bigg) \tag{27}$$

give the values of all the Λ's in terms of

$$\Lambda_n(i,v_i) = \exp(H^{(n)}_{iv_i}), \quad i\in\Delta V_n. \tag{28}$$

For every $X\subseteq V(h), X\neq\varnothing$, we also introduce

$$\Gamma_n(h,v_h;X)\equiv\prod_{k\in X}\Bigg(\sum_{v_k=1}^{q} \exp(K^x_{v_hv_k})\Lambda_n(k,v_k)\Bigg). \tag{29}$$

Let j be a site of T_n, and let $u\equiv z_0, z_1,\ldots,z_d,i,j$ be the sites of the walk connecting u to j. We find

$$\begin{aligned}\mu_n(\mathscr{C}_{ir}) &= \frac{1}{Z_n}\sum_{\zeta_0=1}^{q}\Lambda_n(u,\zeta_0)\\ &\quad\times\sum_{\zeta_1=1}^{q}\exp(K^e_{\zeta_0\zeta_1} + H^o_{\zeta_1})\\ &\quad\times\Gamma_n(z_1,\zeta_1;V(z_1)-\{z_2\})\\ &\quad\times\cdots\\ &\quad\times\sum_{\zeta_d=1}^{q}\exp(K^x_{\zeta_{d-1}\zeta_d} + H^y_{\zeta_d})\\ &\quad\times\Gamma_n(z_d,\zeta_d;V(z_d)-\{i\})\\ &\quad\times\exp(K^y_{\zeta_dr} + H^x_r)\Gamma_n(i,r;V(i)-\{j\})\\ &\quad\times\sum_{t=1}^{q}\exp(K^x_{rt})\Lambda_n(j,t)\\ &\equiv\mathscr{Z}_n(i,r)\sum_{t=1}^{q}\exp(K^x_{rt})\Lambda_n(j,t)\end{aligned} \tag{30}$$

$$\mu_n(\mathscr{C}_{ir}\cap\mathscr{C}_{js}) = \mathscr{Z}_n(i,r)\exp(K^x_{rs})\Lambda_n(j,s).$$

Therefore Lemma 1 implies that μ_n is completely defined, since we have

$$\mu_n(\mathscr{C}_{ur}) = \frac{\Lambda_n(u,r)\sum_{t=1}^{q}\exp(K^e_{rt})\Lambda_n(v,t)}{\sum_{s=1}^{q}\Lambda_n(u,s)\sum_{t=1}^{q}\exp(K^e_{st})\Lambda_n(v,t)}, \tag{31}$$

$$\mu_n(\mathscr{C}_{js}|\mathscr{C}_{ir}) = \frac{\exp(K^x_{rs})\Lambda_n(j,s)}{\sum_{t=1}^{q}\exp(K^x_{rt})\Lambda_n(j,t)}.$$

Notice that these probabilities are homogeneous functions (of degree zero) of the Λ's, and can be expressed in terms of

$$\lambda_n(h,r)\equiv\frac{\Lambda_n(h,r)}{\Lambda_n(h,1)}, \quad h\in V_n, \quad r = 1,\ldots,q. \tag{32}$$

These new parameters satisfy the hierarchic relations

$$\begin{aligned}\lambda_n(h,r) &= \exp(H^x_r - H^x_1)\prod_{k\in V(h)}\Bigg(\frac{\sum_{s=1}^{q}\exp(K^x_{rs})\lambda_n(k,s)}{\sum_{s=1}^{q}\exp(K^x_{1s})\lambda_n(k,s)}\Bigg)\\ &\equiv\Psi^x_r(\lambda_n(k,s)), \quad r = 2,\ldots,q,\end{aligned} \tag{33}$$

that give all of them in terms of

$$\lambda_n(i,t) = \exp(H^{(n)}_{it} - H^{(n)}_{i1}), \quad i\in\Delta V_n, \quad t = 2,\ldots,q. \tag{34}$$

Let us consider now the special case of uniform boundary conditions, i.e., $H^{(n)}_{it} = H^{(n)}_t$ for every $i\in\Delta V_n$, $t = 1,\ldots,q$. These give shell symmetry on T_n, and imply that site dependence is substituted by shell index (≡distance from u) dependence in every Λ, Γ, and λ. Moreover, the hierarchic relations (33) are substituted by the recursive relations

$$\lambda_n(l-1,r) = \Phi^x_r(\lambda_n(l,s)), \quad r = 2,\ldots,q, \tag{35}$$

where x is $e(o)$ if $l-1$ is even (odd), and

$$\Phi^x_r(\lambda_n(l,s))\equiv\exp(H^x_r - H^x_1)\Bigg(\frac{\sum_{s=1}^{q}\exp(K^x_{rs})\lambda_n(l,s)}{\sum_{s=1}^{q}\exp(K^x_{1s})\lambda_n(l,s)}\Bigg)^{\sigma}, \quad r = 2,\ldots,q, \quad 1\leqslant l\leqslant n. \tag{36}$$

We will say that $\{\lambda^e(r)\}_{r=2}^{q}$ is a fixed point of the system $(T_n;\mathscr{H}_n + \Delta\mathscr{H}_n)$ if and only if

$$\lambda^e(r) = \Phi^e_r(\Phi^o_s(\lambda^e(t))), \quad r = 2,\ldots,q; \tag{37}$$

and define the auxiliary fixed point parameters $\{\lambda^o(s)\}_{s=2}^{q}$ as

$$\lambda^o(s)\equiv\Phi^o_s(\lambda^e(t)), \quad s = 2,\ldots,q. \tag{38}$$

Note that every fixed point corresponds to a set of uniform boundary conditions such that the probability measure μ_n is (two-step or, in case, one-step) PI on T_n. In the particular case of an interaction such that

$$K^e_{rs} = K^o_{rs}\equiv K_{rs}, \quad H^e_r = H^o_r\equiv H_r, \quad r,s = 1,\ldots,q, \tag{39}$$

μ_n will be one-step PI on T_n if and only if

$$\lambda^{e}(r) = \lambda^{o}(r) \equiv \lambda(r), \quad r = 2,...,q. \tag{40}$$

We state the following.

Lemma 4: The probability measure μ_n is GI on T_n at every fixed point of the system.

Proof: We will show that μ_n is at least two-step GI on T_n. Because of Lemma 2 it is enough to prove that

$$\mu_n(\mathscr{E}_{ur} \cap \mathscr{E}_{vs}) = \mu_n(\mathscr{E}_{vs} \cap \mathscr{E}_{wr}), \quad r,s = 1,...,q, \tag{41}$$

where $v \in V_1 - \{u\}$ and $w \in V_2 - V_1$ are two adjacent sites of T_n; $n = 2, 3,...$. This relation is equivalent to

$$\begin{aligned} &\Lambda_n(u,r)\exp(K^{e}_{rs} + H^{o}_{s})\Gamma_n(v,s;V(v) - \{w\}) \\ &\quad \times \sum_{t=1}^{q} \exp(K^{o}_{st})\Lambda_n(w,t) \\ &= \sum_{t=1}^{q} \Lambda_n(u,t)\exp(K^{e}_{ts} + H^{o}_{s}) \\ &\quad \times \Gamma_n(v,s;V(v) - \{w\})\exp(K^{o}_{sr})\Lambda_n(w,r). \end{aligned} \tag{42}$$

Using (23) it becomes

$$\sum_{t=1}^{q} \exp(K^{o}_{st})[\Lambda_n(u,r)\Lambda_n(w,t) - \Lambda_n(u,t)\Lambda_n(w,r)] = 0. \tag{43}$$

By means of (32) we see that the term in the brackets is zero if

$$\frac{\lambda_n(u,r)}{\lambda_n(u,t)} = \frac{\lambda_n(w,r)}{\lambda_n(w,t)}, \tag{44}$$

which is certainly satisfied at every fixed point since $l(u,w) = 2$. ■

When (39) holds, applying again Lemma 2 for $m = 1$ and other simple procedures, one can also prove that condition (40) is necessary and sufficient to assure that μ_n is one-step GI on T_n.

Finally, we remark that, if μ_n is GI on T_n, it is completely defined by the following set of fundamental probabilities:

$$\begin{aligned} p^{x}_{r} &= \frac{\lambda^{x}(r)\sum_{t=1}^{q}\exp(K^{x}_{rt})\lambda^{y}(t)}{\sum_{s=1}^{q}\lambda^{x}(s)\sum_{t=1}^{q}\exp(K^{x}_{st})\lambda^{y}(t)}, \\ p^{x}_{rs} &= \frac{\exp(K^{x}_{rs})\lambda^{y}(s)}{\sum_{t=1}^{q}\exp(K^{x}_{rt})\lambda^{y}(t)}, \\ &\quad x,y = e,o \text{ or } o,e, \quad r,s = 1,...,q, \end{aligned} \tag{45}$$

where $\{\lambda^{x}(r)\}_{r=2}^{q}$ are fixed points for $x = e$, and the associated auxiliary parameters for $x = o$.

VI. THE THERMODYNAMIC LIMIT OF M SYSTEMS ON *q*SS BETHE LATTICES

In the first part of this section we limit our considerations to thermodynamic limits (see Proposition 1) obtained with sequences $\{\{H^{(n)}_{t}\}_{t=1}^{q}\}_{n=1}^{\infty}$ of uniform boundary conditions on the trees T_n.

We denote $\mathscr{R}^{q-1}_{>}$ ($\mathscr{R}^{q-1}_{\geq}$) the sector of the $(q-1)$-dimensional real space which contains all the points with nonnegative (positive) coordinates. Let $\Phi^{x}: \mathscr{R}^{q-1}_{\geq} \to \mathscr{R}^{q-1}_{\geq}$ be the functions with the components $\{\Phi^{x}_{r}\}_{r=2}^{q}$ defined by (36). We construct a sequence $\{g_m\}_{m=1}^{\infty}$ of functions by setting $g_1 \equiv \Phi^{e} \circ \Phi^{o}; g_{m+1} \equiv g_1 \circ g_m$; $m = 1,2,...$. The properties of thermodynamic limits on Bethe lattices are related to the asymptotic behavior of $\{g_m\}_{m=1}^{\infty}$. On principle, for every $\{\lambda(r)\}_{r=2}^{q} \equiv \lambda \in \mathscr{R}^{q-1}_{>}$, the sequence $\{g_m(\lambda)\}_{m=1}^{\infty}$ of points in $\mathscr{R}^{q-1}_{>}$ converges, or exhibits ergodic, turbulent, or cyclic behavior. We say that the iterative equations (35) are piecewise contracting (PC) if $\{g_m(\lambda)\}_{m=1}^{\infty}$ has a limit point $\lambda^{*} \equiv g(\lambda)$ for every $\lambda \in \mathscr{R}^{q-1}_{>}$. In other terms, the PC property holds if we have the pointwise convergence $g_m \to g$ on $\mathscr{R}^{q-1}_{>}$.

Theorem 2: The probability measure $\mu = \lim_{n\to\infty} \mu_n$ is RT on the Bethe lattice if the recursive relations (35) are PC.

Proof: Let $\{T_{n'}\}_{n'=1}^{\infty}$ be the subsequence of $\{T_n\}_{n=1}^{\infty}$ which defines μ (see Proposition 1). We denote $\lambda_{n'}(h) \equiv \{\lambda_{n'}(h,r)\}_{r=2}^{q}$ $[\lambda^{*}(h) \equiv \{\lambda^{*}(h,r)\}_{r=2}^{q}]$ the set of parameters associated by $\mu_{n'}(\mu)$ to the site $h \in V_{n'}$. We assume that $\Delta V_{n'} \subset V^{e}$ for every tree $T_{n'}$. This can be done without loss of generality, because any $T_{n'}$ with odd n' and uniform boundary conditions $\lambda_{n'}(n')$ on $\Delta V_{n'}$ can be substituted by $T_{n'-1}$ with the boundary conditions $\lambda_{n'-1}(n'-1)$ $\equiv \lambda_{n'}(n'-1) = \Phi^{e}(\lambda_{n'}(n'))$ on $\Delta V_{n'-1}$ [obtained by means of (35)]. Finally, let us consider a tree $T = (V_T, E_T)$ belonging to the sequence $\{T_{n'}\}_{n'=1}^{\infty}$.

The convergence $\mu_{n'} \to \mu$ implies that, for every $\epsilon > 0$, there exists large enough n'_0 such that $T \subset T_{n'}$ and

$$[\lambda_{n'}(i);\lambda^{*}(i)] < \epsilon/3, \tag{46}$$

for every $n' > n'_0$ and $i \in \Delta V_T$ (here the brackets represent the Euclidean distance in $\mathscr{R}^{q-1}$). Due to the continuity of Φ^{e} and Φ^{o}, the relation above can be extended to every $i \in V_T$ for n' greater than a certain $n'_1 \geq n'_0$. On the other hand, the PC property implies that, given $\lambda \in \mathscr{R}^{q-1}_{>}$, for every $\eta > 0$ there exists large enough $m(\lambda)$ such that

$$[g_{m'}(\lambda);g_{m''}(\lambda)] < \eta, \tag{47}$$

for every $m' > m(\lambda)$ and $m'' > m(\lambda)$. This means that, given T, for every $\epsilon > 0$ there exists large enough $n'_2(\lambda)$ such that $T \subset T_{n'}$ and

$$[\lambda_{n'}(i);\lambda_{n'}(j)] < \epsilon/3, \tag{48}$$

for every $n' > n'_2(\lambda)$, $i \in V^{e}_{T}$, and $j \in V^{e}_{T}$ (now λ is the boundary condition on $\Delta V_{n'}$). Relations (46) and (48) imply

$$\begin{aligned} [\lambda^{*}(i);\lambda^{*}(j)] &\leq [\lambda^{*}(i);\lambda_{n'}(i)] \\ &\quad + [\lambda_{n'}(i);\lambda_{n'}(j)] \\ &\quad + [\lambda_{n'}(j);\lambda^{*}(j)] < \epsilon, \end{aligned} \tag{49}$$

for every $\epsilon > 0$, $i \in V^{e}_{T}$, and $j \in V^{e}_{T}$. By Lemma 4 it follows that the restriction $\tilde{\mu}$ on R_T of the probability measure μ on R is MGI on T for every T belonging to $\{T_{n'}\}_{n'=1}^{\infty}$. Therefore Lemma 3 applies, thus completing our proof. ■

Remark that every point of the graph of g is a fixed point of the system, since $\lambda^{*} = g(\lambda)$ satisfies

$$\lambda^{*} = g_1(\lambda^{*}), \tag{50}$$

for every $\lambda \in \mathscr{R}^{q-1}_{>}$. In general we expect a finite set $\{\lambda^{*}_{\alpha}\}_{\alpha=1}^{a}$ of distinct solutions for (50), i.e., g is a step function on $\mathscr{R}^{q-1}_{>}$. This means that there exists a partition $\{D_\alpha\}_{\alpha=1}^{a}$ of $\mathscr{R}^{q-1}_{>}$ (in domains of attraction) such that every seed λ goes into the same fixed point $\lambda^{*}_{\alpha} = g(\lambda)$, for every $\lambda \in D_\alpha$; $\alpha = 1,...,a$. As a consequence of Theorem 1, Theorem 2, and Lemma 4, we state the following.

Corollary 2: If the recursive relations (35) are PC, the fundamental probabilities which define μ are given by (45) at each fixed point of the system.

The proof that the PC property holds in the special case of the ferromagnetic or antiferromagnetic Ising and Potts models is reported in Appendix D. We do not give here a general proof that the iterative equations (35) are PC.[30] However, it is easy to see that the pointwise convergence $g_m \to g$ holds at least in a domain $D \subseteq \mathscr{R}_{>}^{q-1}$. Notice that the functions Φ_r^x defined by (36) have finite upper bounds for all real K_{rs}^x and H_r^x. This implies that there exists a hypercube $U_b \equiv \{\lambda \in \mathscr{R}_{>}^{q-1} : \lambda(r) \leqslant b;\ r = 2,...,q\}$, and large enough $b > 0$, such that $g_1(\lambda) \in U_b$ for every $\lambda \in U_b$. Therefore, Brouwer's theorem[31] applies, and there is at least one fixed point satisfying (50) in U_b, whose domain of attraction is the above-mentioned set D. This fixed point is certainly unique (and $D = \mathscr{R}_{>}^{q-1}$) in the infinite temperature limit, since our solution must recover RT Bernoulli measures associated to random distributions of states. As a matter of fact, when $K_{rs}^x \to 0$ for every x,r,s, relations (36)–(38) give[32]

$$\lambda^x(r) = \Phi_r^x(\lambda^y(s)) = \exp(\overline{H}_r^x - \overline{H}_1^x),$$
$$x,y = e,o \text{ or } o,e,\ r,s = 1,...,q \tag{51}$$

(where $\overline{H}_r^x$ is the infinite temperature limit of H_r^x), and relations (45) become

$$p_r^x = p_{sr}^y = \frac{\lambda^x(r)}{\sum_{t=1}^{q} \lambda^x(t)},$$
$$x,y = e,o \text{ or } o,e,\ r,s = 1,...,q. \tag{52}$$

The existence of several fixed points (corresponding to the same set of coupling terms and external fields, and distinct sets of boundary conditions) is related to distinct orderings of the system. Each of them corresponds to a pure thermodynamic phase described by a certain MRT probability measure. This would be the typical case at low temperatures.

Now we briefly consider the case of thermodynamic limits with nonuniform boundary conditions. It is easy to convince oneself that, when (35) are PC, this procedure does not give rise to new pure phases. This is mainly due to the mixing/damping properties of the functions Ψ's defined by (33). When (50) admits only one fixed point, it is clear that any choice of nonuniform boundary conditions will generate in the thermodynamic limit the same state for the system through the Ψ's. When (50) admits several fixed points, two cases arise. The mixing/damping properties may prevail over the nonuniformity of the boundary conditions, and then bring the system into one of the above-mentioned fixed points. Due to the branching structure of the trees T_n, we can also find certain boundary conditions whose nonuniformities evolve towards distinct fixed points and do not mix until the center (or any finite central zone) of the tree is reached. In the thermodynamic limit this corresponds to the breakdown of rotational and translational invariance associated to phase mixing. The appearance of these mixtures has the maximum physical relevance when they describe actual first-order transitions between pure phases, i.e., when each component pure phase is characterized by the same free energy.[25,33]

Using the terminology introduced in Sec. II, we say that every probability measure corresponding to a fixed point of the system is an extremal point of $\mathscr{K}_I$. Other extremal points describing non-RT Gibbs states will be found in the set of all the mixtures.

Finally, we remark that (a preliminary version of) the present procedure has been applied by Peruggi, di Liberto, and Monroy[34] to solve the Potts model. The reader is referred to that reference for a simple realization of our general results. Connections with the solution of the Potts model are given in Appendix D, while addenda to, and some remarks on, the results of Ref. 34 are given, respectively, in paper II and in the following section.

VII. FINAL REMARKS

From the physical point of view it is interesting to compare the results obtained by means of the present approach to the solution of Hamiltonian models on Bethe lattices, with the results given by the methods described in the introduction.

For what concerns type (ii) approaches, simple checks show that there is general agreement with our procedure. Here we focus mainly on those type (ii) methods which reduce to adaptations of the Bethe–Peierls cluster approximation[24] (BPCA) on Bethe lattices, and on the BPCA itself on regular infinite lattice graphs[26] of coordination number $\sigma + 1$. Notice that the above-mentioned "cluster" and the relative Hamiltonian are, in the present terminology, the systems $(T_1, \mathscr{H}_1 + \Delta\mathscr{H}_1)$ or $(T_2, \mathscr{H}_2 + \Delta\mathscr{H}_2)$ with uniform boundary conditions. The state of these systems is characterized by imposing m-step consistency conditions on the respective "magnetization(s)," i.e.,

$$\left.\begin{aligned} &\mu_1(\mathscr{E}_{ur}) = \mu_1(\mathscr{E}_{vr}), \quad m = 1 \\ &\mu_2(\mathscr{E}_{ur}) = \mu_2(\mathscr{E}_{wr}), \quad m = 2 \end{aligned}\right\} r = 1,...,q. \tag{53a, 53b}$$

The physical idea under (53) is that they would select GI probability measure(s) which describe approximately the translationally invariant equilibrium state(s) of the regular lattice. In particular, (53a) is used when (39) holds and the system $(T_1, \mathscr{H}_1 + \Delta\mathscr{H}_1)$ is expected to be described by a one-step GI probability measure. By means of Lemma 4 we deduce that every MRT solution on the Bethe lattice is a solution for the BPCA. Conversely, it is easy to see that (53a) implies $\lambda_1(u,r) = \lambda_1(v,r)$ for $r = 2,...,q$; i.e.,

$$\lambda^* = \Phi(\lambda^*). \tag{54}$$

In regard to (53b), it could be satisfied at nonfixed points, too. One expects that these solutions (if any) do not correspond to physical states. The uncertainty can be removed by substituting condition (53b) with (41) for $n = 2$.

In conclusion, these results imply that the probability measure μ (i.e., the exact solution obtained for a MRT–qSS Hamiltonian model on Bethe lattices with the present method) is the same as the MRT extension of the probability measure μ_1 (μ_2) (i.e., the approximate solution obtained for that model on regular lattices with the BPCA).[35] Let us emphasize the fact that we have not introduced just a formally more satisfying version than BPCA and type (ii) methods: our procedure is a substantial improvement. In fact it gives in a very simple way the free energy[25,33] of the system under investiga-

tion in terms of the fundamental probabilities, which in turn provides complete information about the physical properties of the system. Furthermore, the knowledge of the probability measure describing a system allows us to solve almost all problems concerning that system, not only thermal ones (see, e.g., the topics studied in paper II).

The above proofs also suggest that our results and those of type (i) approaches do not agree. Notwithstanding this, in the following we will study in detail the problems concerning the thermodynamic limits of extensive functions on Bethe lattices. In fact, the decomposition rule (9) makes more transparent the underlying reasons for the above-mentioned disagreement, and also suggests the way to follow in order to obtain limits consistent with the probability measure approach. First, we add some definitions and remarks to the general formalism introduced in Sec. II. Let an m-step Markov qSS Hamiltonian model be defined on an infinite, connected, and locally finite graph[26] $G=(V,E)$ with a countable set of sites. We say that the sequence $\{V_n\}_{n=1}^{\infty}$ of finite subsets of V, tends to V in the sense of Van Hove if the boundary-to-bulk ratio $\rho_n \equiv |\Delta V_n|/|V_n|$ of its components satisfies the condition

$$\lim_{n\to\infty} \rho_n = 0. \tag{55}$$

The thermodynamic formalism for translationally invariant systems (typically d-dimensional hypercubic lattices; $d=1,2,\ldots$) requires the use of sequences which satisfy Van Hove's convergence condition when the thermodynamic limits of extensive functions, as the entropy or the free energy, are taken.[23] This assures their existence, their uniqueness (in the sense that the limits obtained do not depend on the choice of the sequence), and their consistence with the limit of Gibbs probability measures. [We remark again that no requirement as (55) is needed when the $n\to\infty$ limit of probability measures is taken; see Secs. I–II and Ref. 23.] Roughly speaking, the convergence in the sense of Van Hove assures that we do not retain systematical errors when the thermodynamic limit is used to find the per site expectation values of extensive functions. When condition (55) does not hold for the sequence $\{V_n\}_{n=1}^{\infty}$, we expect that the $n\to\infty$ limit of

$$\langle f\rangle_n \equiv \sum_{\mathscr{C}_{V_n}\in\|V_n\|} f(\mathscr{C}_{V_n})\mu_n(\mathscr{C}_{V_n}), \tag{56}$$

in general is not equal to

$$\langle f\rangle \equiv \int_{\|V\|} f\,d\mu, \tag{57}$$

where f is a per site extensive physical observable defined on the configurations of $V_n(V)$, and $\mu=\lim_{n\to\infty}\mu_n$. Simple checks may be done by looking for the internal energy, whose expression (57) for (one-step) translationally invariant infinite systems reduces, for every $i\in V$, to

$$\beta\mathscr{U} = \sum_{\mathscr{C}_A\in R:\mathscr{C}_A\neq\emptyset;A\ni i} \frac{1}{|A|}\langle I(\mathscr{C}_A)\rangle. \tag{58}$$

Let us apply the arguments above to the Bethe lattice L. We see that no sequence $\{T_n\}_{n=1}^{\infty}$ of finite trees tends to L in the sense of Van Hove, since, for every m-step Markov system, one finds

$$\lim_{n\to\infty} \rho_n \geqslant (\sigma-1)/\sigma, \tag{59}$$

where the equality holds for $m=1$. For major clarity let us assume that both the interaction I on R (which characterizes the Hamiltonian model on L) and the probability measure μ on R (which describes that model) are MRT, one-step translationally invariant, on L. Then (58) and (22) or (25) give

$$\beta\mathscr{U} = -\sum_{r=1}^{q} H_r p_r - \frac{\sigma+1}{2}\sum_{r,s=1}^{q} K_{rs}p_r p_{rs}, \tag{60}$$

where we have used the fact that $\delta_{\nu_i r}\delta_{\nu_j s}(\delta_{\nu_i r})$ is the indicator of the event $\mathscr{C}_{ir}\cap\mathscr{C}_{js}(\mathscr{C}_{ir})$, i.e., the function defined on $\|V\|$ which is 1 on the configurations belonging to $\mathscr{C}_{ir}\cap\mathscr{C}_{js}(\mathscr{C}_{ir})$, and 0 elsewhere. On the other hand, on T_n we have

$$\begin{aligned}\frac{1}{|V_n|}\beta\mathscr{U}_n &\equiv \frac{1}{|V_n|}\langle\beta\mathscr{H}_n\rangle_n \\ &= -\frac{1}{|V_n|}\sum_{i\in V_n}\sum_{r=1}^{q} H_r\mu_n(\mathscr{C}_{ir}) \\ &\quad -\frac{|E_n|}{|V_n|}\frac{1}{|E_n|}\sum_{\langle ij\rangle\in E_n}\sum_{r,s=1}^{q} K_{rs}\mu_n(\mathscr{C}_{ir}\cap\mathscr{C}_{js}).\end{aligned} \tag{61}$$

Note that this relation holds in both cases of free boundary conditions [Hamiltonian (22)] or fixed boundary conditions [Hamiltonian (25)]. In fact, the interface Hamiltonian (24) does not "belong" to the system [definition (25) was introduced for notation convenience only], and has influence on its state only through the values of site and bond probabilities. Relation (61) implies

$$\lim_{n\to\infty}\frac{1}{|V_n|}\beta\mathscr{U}_n = -\sum_{r=1}^{q} H_r p_r - \sum_{r,s=1}^{q} K_{rs}p_r p_{rs}. \tag{62}$$

As was expected in view of (59), we see that $\beta\mathscr{U}$ and the thermodynamic limit of $\beta\mathscr{U}_n$ are not equal. This result can be extended to the case of non-RT μ too,[36,37] since the terms p_r and $p_r p_{rs}$ in (60) and (62) may be considered, in such a case, as the averages of $\mu(\mathscr{C}_{ir})$ and $\mu(\mathscr{C}_{ir}\cap\mathscr{C}_{js})$, respectively, over all $i\in V$ and $\langle ij\rangle\in E$; $r,s=1,\ldots,q$. It is easy to see that the above disagreement is due to a discrepancy [related to (59)] between the asymptotic topological properties of $\{T_n\}_{n=1}^{\infty}$ and those of L. Let us define the local number of bonds per site $\psi_i(G^*)$, relative to the site i of a locally finite graph G^*, as half the coordination number of i in G^*, and let the mean number of bonds per site $\psi(G^*)$ be the average of $\psi_i(G^*)$ on G^*. We have

$$\psi_i(T_n) = \begin{cases}(\sigma+1)/2, & i\in V_n - \Delta V_n,\\ \frac{1}{2}, & i\in\Delta V_n,\end{cases} \tag{63a}$$

$$\psi(T_n) = \frac{1}{|V_n|}\sum_{i\in V_n}\psi_i(T_n) = 1+\frac{1}{|V_n|}, \tag{63b}$$

$$\psi_i(L) = (\sigma+1)/2 = \lim_{n\to\infty}\psi_i(T_n),\quad i\in V \tag{63c}$$

$$\psi(L) = \psi_i(L) \neq \lim_{n\to\infty}\psi(T_n). \tag{63d}$$

Relation (61) depends on $\psi(T_n)$, while (60) depends on $\psi(L)$: (60) and (62) do not agree because of the "surface effect" (63d). The entropy on T_n (which is explicitly calculated in paper II) depends on $\psi(T_n)$: this implies that (63d) affects its

$n \to \infty$ limit, too, and therefore the limit of the free energy.

In conclusion, in this section we have found the following results: (a) the BPCA and type (ii) methods agree with the probability measure approach, while type (i) methods do not; (b) Van Hove's convergence condition, in its standard form, is not satisfied by Hamiltonian models on Bethe lattices; and (c) the thermodynamic limit of the free energy, as is calculated by type (i) approaches, is affected by a topological discrepancy. In agreement with the point of view which we explained in the Introduction, our interpretation of (a)–(c) is the following: (1) Van Hove's convergence condition can be extended to, and is meaningful on, Bethe lattices; and (2) the probability measure approach should be seen as the canonical method for the solution of models on Bethe lattices, since it is the unique method which is, at the same time, rigorous and unaffected by surface effects.

The above picture seems to imply that we are unable to find the correct free energy of systems defined on Bethe lattices, which, in fact, is the common lack of all type (ii) methods except for one special case,[20] where an integration procedure, starting from the equation of state, is used. However, in the context of a type (i) method, Baumgärtel and Müller–Hartmann[19] found the generating function of the random cluster model and its corresponding form in a BPCA context. Their assumptions, in the present terminology, can be expressed as the heuristic rule

$$\lim_{n \to \infty} \psi(T_n) \equiv \psi(L). \tag{64}$$

Furthermore, for the solution of the Potts model with (a preliminary version of) the present method, Peruggi, di Liberto, and Monroy[34] also introduced (64) to find the free energy, and verified that it was consistent with the probability measure approach. Although the present results and (64) can be used to find the free energy of any MRT Hamiltonian model on qSS Bethe lattices, the procedure is formally unsatisfactory. In paper II we will take full advantage of the decomposition rule (9), which allows us to proceed to the localization of extensive functions (i.e., the association of bond terms to the sites). The main point of interest of the localization procedure is that the thermodynamic limit of local quantities is not affected by topological discrepancies, as can be seen by (63c), thus providing a rigorous limit procedure which has no need of heuristic rules as (64).

ACKNOWLEDGMENTS

The author is very indebted to Professor A. Coniglio and Professor L. Russo for enlightening suggestions, Professor B. Preziosi for constant encouragement, F. di Liberto and G. Monroy for helpful discussions. G. Monroy is also acknowledged for critical reading of the manuscript.

This work was supported by Ministero della Pubblica Istruzione (MPI) and the Consiglio Nazionale delle Ricerche (CNR).

APPENDIX A: CAYLEY TREES, BETHE LATTICES, AND RELATED TOPICS

In the physical literature concerning trees, i.e., the graphs of main interest in this paper, many distinct, and somewhat confusing, terminologies are used to denote the same objects. In order to make easier the connections and comparisons of our results with those of other authors, we give here our definitions[26] of such objects, together with those usually found in other works.

A tree is a connected graph without polygons. In the mathematical literature a tree is a Cayley tree, and vice versa. We follow this convention, and always drop the name "Cayley." A tree $T = (V_T, E_T)$ is finite if the set $V_T(E_T)$ of its sites (bonds) contains a finite number of elements. The surface of a tree is the set of all the sites whose coordination number is equal to 1. The interior of a tree is the set of all the sites which do not belong to its surface. Given a sequence $\{T_m\}_{m=1}^{\infty}$ of finite trees, we say that it is strictly increasing if the interior of T_{m+1} contains V_m for every index m. Any strictly increasing sequence of finite trees defines an infinite tree, namely,

$$L \equiv (V,E) \equiv \left(\bigcup_{m=1}^{\infty} V_m, \bigcup_{m=1}^{\infty} E_m \right) \equiv \bigcup_{m=1}^{\infty} T_m. \tag{A1}$$

To be concise we say that $\{T_m\}_{m=1}^{\infty}$ tends to L. This type of convergence may be used in thermodynamic limit procedures (see Sec. II), where the knowledge of the behavior of a model on growing finite systems (e.g., T_m; $m = 1,2,...$) allows one to deduce its properties on an infinite system (e.g., L). Actually this is done in all the papers concerning the present subject, where strictly increasing sequences of trees are always used. Remark that relation (A1) implies that every $i \in V$ is an interior site of L: in fact, by definition, it belongs to infinitely many trees in the sequence $\{T_m\}_{m=1}^{\infty}$. Although obvious, this fact clearly shows that the surface or the "boundary" of the infinite (Cayley) tree, which are often mentioned in the physical literature, are not real topological objects: their authors refer only to the retaining of surface effects when the thermodynamic limit of extensive functions, defined on the elements of sequences as $\{T_m\}_{m=1}^{\infty}$, is taken (see Secs. I and VII).

Let us consider, now, the connection between other terminologies and ours. Some authors reserve the label "(finite) Cayley trees" to denote those finite trees whose interior sites have the same coordination number $\sigma + 1$. As a trivial consequence of the remark above, we see that all the sites of the tree, generated by a strictly increasing sequence of such graphs, have the same coordination number $\sigma + 1$, i.e., in the present case, the "(infinite) Cayley tree" is exactly the same as the Bethe lattice (see the definition at the beginning of Sec. III). Other authors reserve the label "(regular) Cayley tree" to denote those finite trees whose interior sites have the same coordination number $\sigma + 1$, except the "central" site, whose coordination number is σ. The "(infinite) Cayley tree" or "regular Bethe lattice" generated by a strictly increasing sequence of such graphs may be called, in our terminology, single-defect Bethe lattice. In fact, it is a Bethe lattice of coordination number $\sigma + 1$ which has one site (the defect) with coordination number σ. It is physically intuitive that the defect has no relevance with respect to the properties of the model under investigation: actually, the authors who introduce these graphs use them only in order to make easier calculations.

APPENDIX B: THE DIMENSIONALITY OF BETHE LATTICES

Throughout this paper, graphs and trees are only used as topological supports for abstract Hamiltonian models. However it is customary for physicists to think of these objects in more concrete terms, i.e., to consider realizations of the graphs (≡drawings of sites and bonds as points and segments) in a Euclidean space where the interactions defining the models have proper spatial distributions. Since the set of sites of a Bethe lattice L is countable (see Sec. III), and any finite tree is planar,[26] it follows that one can consider realizations of L in the plane (see Fig. 1), and actually in every d-dimensional real space $\mathscr{R}^d$. In this connection, the question arises concerning the effective dimensionality of Bethe lattices.

One can provide answers from the topological point of view (notice that topology and interactions are strictly related here). The fact that no polygons are present in the Bethe lattice seems to suggest a one-dimensional character. On the other hand, the dimensionality can be defined as the smallest positive integer d such that a regular representation of L can be drawn in $\mathscr{R}^d$. Since it is known that (for every $\sigma > 1$) the angles between bonds, and the lengths of bonds, in a representation of L in $\mathscr{R}^d$ cannot be bounded below by any positive constant for every d,[38] one may deduce that L has an infinite-dimensional character. Also the method described by Baxter,[20] which gives correct results for the dimensionality of all the regular two- and three-dimensional lattices, gives $d = \infty$. In both cases the result obtained is due to the fact that the number of sites of L at distance l from a given site (see Sec. III) grows exponentially when l increases.

Other suggestions about the dimensionality of Bethe lattices may be argued by analysis of solutions. For convenience we refer to the class of models which satisfy the symmetry condition:

$$\sum_{s=1}^{q} \exp(K^{x}_{rs}) = \sum_{s=1}^{q} \exp(K^{x}_{1s}), \quad x = e,o;\ r = 2,\ldots,q. \tag{B1}$$

In such a case, at zero external fields and for any temperature, Eqs. (37) admit the fixed point $\lambda^{e}(2) = \lambda^{e}(3) = \cdots = \lambda^{e}(q) = 1$. By means of (38) and (45) we see that it describes the disordered state of the system, and by means of (34) we see that it is always obtained with free boundary conditions ($H^{(n)}_{t} = 0$; $t = 1,\ldots,q$; $n = 1,2,\ldots$). In a sense we can say that it is the "natural" phase of the system, thus deducing a one-dimensional character of Bethe lattices. (Let us remark that, for $\sigma = 1$, our formalism recovers the one-dimensional chain, which is always found in such a phase, in agreement with Ruelle's proof[23] of absence of phase transitions in a class of models which includes those studied here.) However, for $\sigma > 1$ and sufficiently low temperatures, the "natural" phase may correspond to a repulsive fixed point, i.e., (arbitrarily small) fixed boundary conditions bring the system into attractive fixed points that describe ordered phases which actually minimize the free energy (see Sec. III of paper II). Therefore, we are led again to an infinite-dimensional character of Bethe lattices, associated to solutions described by "classical" exponents at the critical point(s). This is supported by the equivalence of our results and the Bethe–Peierls cluster approximation on standard lattices (see Sec. VII), since it is known that the latter belongs to the same class of approximations, and actually is an improvement, of the mean field theory (further details about these topics will be found in the works by Domb[4] and Baxter[20]). Another connection between the infinite-dimensional character of Bethe lattices and the classical values of the critical exponents γ and ν is discussed by Peruggi, di Liberto, and Monroy.[39]

In conclusion, we think that the effective dimensionality of Bethe lattices, at least for the MRT-qSS Hamiltonian models studied here, is neither "subjective," as stated by Hughes and Sahimi,[38] nor "quasi-one-dimensional," as stated by Moraal,[17] but actually is $d = \infty$. The doubts about this fact may be related, in technical terms, to the unusual presence of the "natural" phase at low temperatures, too. In fact, such an extremal point of $\mathscr{K}_I$ is not expected on hypercubic lattices,[23] and certainly does not exist for the Ising model on the square lattice.[40] This subject is discussed extensively, in connection with the free energy properties, in Sec. III of paper II.

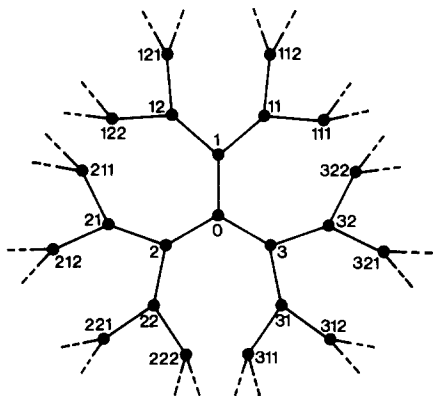


FIG. 1. Finite portion of a realization on the plane of the Bethe lattice with coordination number $\sigma + 1 = 3$. Lexicographic ordering of the sites is shown.

APPENDIX C: SOME RELEVANT MODELS RECOVERED BY THE PRESENT FORMALISM

A large class of Hamiltonian models can be expressed in the form (22). Complete characterization of a certain system is obtained by means of proper assignments of symmetry conditions and/or relative ratios to the coupling terms K^{x}_{rs}. As regards to the external fields H^{x}_{r}, no special prescriptions are needed since they were introduced in order to make accessible the largest set of values for the fundamental probabilities (from a physical point of view they can be seen as chemical potentials which govern the relative densities of states).

The Potts model is recovered by the following specialization of the general formalism:

$$K^{x}_{rs} = K\delta_{rs}, \quad x = e,o, \quad r,s = 1,\ldots,q. \tag{C1}$$

The ferromagnetic (antiferromagnetic) model is characterized by $K > 0$ ($K < 0$); (a model isomorph to) the Ising model

is found for $q = 2$.

The vector (or planar) Potts model is recovered through the choice

$$K^x_{rs} = K\cos(2\pi|r-s|/q), \quad x = e,o, \quad r,s = 1,...,q \tag{C2}$$

[remark that for $q = 2$ ($q \to \infty$) one obtains the Ising model (the classical Heisenberg model)]. The solution of the models (C1) and (C2) can be found in the series of papers on the Potts model by Peruggi, di Liberto, and Monroy.[34,41,42]

Another interesting model contained as a special case in the present formalism is the Ashkin–Teller model. This is defined for $q = 4$ and is characterized by the following interaction matrix:

$$K^x_{rs} = \begin{pmatrix} K_0 & K_1 & K_2 & K_3 \\ K_1 & K_0 & K_3 & K_2 \\ K_2 & K_3 & K_0 & K_1 \\ K_3 & K_2 & K_1 & K_0 \end{pmatrix}, \quad x = e,o, \quad r,s = 1,2,3,4. \tag{C3}$$

The $Z(q)$ (or clock) model, which has recently attracted great attention in field theory, is also recovered. Its interaction matrix is cyclic:

$$K^x_{rs} = K_{|r-s|}, \quad x = e,o, \quad r,s = 1,...,q \tag{C4}$$

[note that (C2) is a special case of (C4)].

APPENDIX D: PROOF OF THE PC PROPERTY FOR THE POTTS MODEL

We limit our considerations to the Potts model with only one external field, which is recovered by adding to (C1) the following relation:

$$H^x_r = H\delta_{r1}, \quad x = e,o, \quad r = 1,...,q. \tag{D1}$$

In view of (C1) and (D1), the iterative functions (36) have no sublattice dependence, and reduce to

$$\Phi_r(\lambda^x(s)) = e^{-H}\left(\frac{1 + e^K\lambda^x(r) + \Sigma^q_{t=2(t\neq r)}\lambda^x(t)}{e^K + \Sigma^q_{t=2}\lambda^x(t)}\right)^\sigma, \quad r = 2,...,q. \tag{D2}$$

Remark that the model of Ref. 34 was constructed with the total equivalence between the states $r = 2,...,q$. Although its Hamiltonian is exactly the same, the present model is the Potts limit of a general model and retains (possible) distinctions. These will be disregarded, and complete identification will be achieved, if we also set

$$\lambda^x(2) = \lambda^x(3) = \cdots = \lambda^x(q) \equiv \lambda^x, \quad x = e,o. \tag{D3}$$

This implies that all the relations (D2) become

$$\Phi(\lambda^x) = e^{-H}\left(\frac{1 + (e^K + q - 2)\lambda^x}{e^K + (q-1)\lambda^x}\right)^\sigma, \tag{D4}$$

and that the definitions of Sec. VI are relative, now, to the positive (nonnegative) real line $\mathscr{R}_>$ ($\mathscr{R}_\geq$). A little algebra shows that the function $g_1: \mathscr{R}_> \to \mathscr{R}_>$ is monotone increasing and has finite positive lower and upper bounds for every real K and H. As a consequence, Eq. (50) admits at least one solution, i.e., there is at least one fixed point for the system. Let $\{\lambda^*_\alpha\}^a_{\alpha=1}$ be the solutions of Eq. (50) in increasing order. We consider the open intervals $(0,\lambda^*_1)$; $(\lambda^*_1,\lambda^*_2)$; ...; (λ^*_a,∞). Let (b,c) be one of these intervals. Observe that we have $g_1(\lambda) > \lambda$ for every $\lambda \in (b,c)$, or $g_1(\lambda) < \lambda$ for every $\lambda \in (b,c)$. Suppose the first (second) case is true. The definition of the functions g_m, and the monotoneness of g_1 imply that, for every $\lambda \in (b,c)$, $\{g_m(\lambda)\}^\infty_{m=1}$ is an increasing (decreasing) sequence of points in (b,c) whose least upper bound is c (greatest lower bound is b). It follows that $g_m(\lambda) \to g(\lambda) = c$ [$g_m(\lambda) \to g(\lambda) = b$] for every $\lambda \in (b,c)$, i.e., the PC property holds.

The classification of every λ^*_α as an attractive, repulsive, or mixed fixed point, and the partition $\{D_\alpha\}^a_{\alpha=1}$ of $\mathscr{R}_>$ follow trivially from the preceding results. Furthermore, although our proof does not depend on the number of fixed points, we can also see that in the present case $a \leqslant 3$.

A numerical study of the general relations (D2), without the one-parameter condition (D3), can be found in Ref. 41.

[1] M. Kaufman and R. B. Griffiths, Phys. Rev. B **24**, 496 (1981); R. B. Griffiths and M. Kaufman, Phys. Rev. B **26**, 5022 (1982); also see M. Kaufman and R. B. Griffiths, Phys. Rev. B **26**, 5282 (1982).

[2] See, e.g., Y. G. Sinai, *Theory of Phase Transitions: Rigorous Results* (Pergamon, Oxford, 1982). Also see the definition of Markov processes in Ref. 28.

[3] M. Kurata, R. Kikuchi, and T. Watari, J. Chem. Phys. **21**, 434 (1953).

[4] C. Domb, Adv. Phys. **9**, 149 (1960).

[5] G. W. Woodbury, Jr., J. Chem. Phys. **47**, 270 (1967); L. K. Runnels, J. Math. Phys. **8**, 2081 (1967).

[6] T. Obokata and T. Oguchi, J. Phys. Soc. Jpn. **25**, 322 (1968).

[7] J. C. Wheeler and B. Widom, J. Chem. Phys. **52**, 5334 (1970).

[8] T. P. Eggarter, Phys. Rev. B **9**, 2989 (1974).

[9] J. von Heimburg and H. Thomas, J. Phys. C **7**, 3433 (1974); M. Matsuda, Prog. Theor. Phys. **51**, 1053 (1974).

[10] S. Katsura and M. Takizawa, Prog. Theor. Phys. **51**, 82 (1974); S. Katsura, J. Phys. A **8**, 252 (1975).

[11] E. Müller-Hartmann and J. Zittartz, Phys. Rev. Lett. **33**, 893 (1974); E. Müller-Hartmann and J. Zittartz, Z. Phys. B **22**, 59 (1975); E. Müller-Hartmann, Z. Phys. B **27**, 61 (1977).

[12] H. Falk, Phys. Rev. B **12**, 5184 (1975).

[13] A. Coniglio, J. Phys. A **8**, 1773 (1975); A. Coniglio, Phys. Rev. B **13**, 2194 (1976).

[14] T. Morita and T. Horiguchi, Prog. Theor. Phys. **54**, 982 (1975); T. Morita and T. Horiguchi, J. Stat. Phys. **26**, 665 (1981).

[15] S. Muto and T. Oguchi, Prog. Theor. Phys. **55**, 81 (1976).

[16] Y. K. Wang and F. Y. Wu, J. Phys. A **9**, 593 (1976).

[17] H. Moraal, Physica A **85**, 457 (1976); H. Moraal, Physica A **92**, 305 (1978); H. Moraal, Physica A **105**, 472 (1981); H. Moraal, Physica A **113**, 44 (1982); H. Moraal, Physica A **113**, 67 (1982); H. Moraal, Z. Phys. B **45**, 237 (1982). Also see H. Moraal, J. Phys. C **15**, L55 (1982); H. Moraal, Phys. Lett. A **89**, 310 (1982); H. Moraal, Physica A **122**, 313 (1983).

[18] C. Thompson, J. Stat. Phys. **27**, 441 (1982).

[19] H. G. Baumgärtel and E. Müller-Hartmann, Z. Phys. B **46**, 227 (1982).

[20] R. J. Baxter, *Exactly solved models in Statistical Mechanics* (Academic, New York, 1982).

[21] The set of Refs. 3–20 is not exhaustive.

[22] This concept was first introduced for classical systems [see, e.g., the proofs of van Hove's theorem and the theorems of Yang and Lee given by K. Huang, *Statistical Mechanics* (Wiley, New York, 1963)]; its generalization to lattice systems can be found in D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), also see Ref. 23. The historical development of thermodynamic limit procedures, van Hove's convergence condition, and many related topics were discussed in detail by R. B. Griffiths, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 1.

[23] D. Ruelle, *Thermodynamic Formalism, Encyclopedia of Mathematics and its Applications* (Addison–Wesley, Reading, MA, 1978), Vol. 5.

[24]H. A. Bethe, Proc. R. Soc. London, Ser. A **150**, 122 (1935); R. Peierls, Proc. Camb. Phyl. Soc. **32**, 471 (1936). Also see the reviews by Domb (Ref. 4) and by D. M. Burley, in *Phase transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 2.
[25]F. Peruggi, J. Math. Phys. **25**, 3316 (1984).
[26]We refer to the terminology of graph theory established by J. W. Essam and M. E. Fisher, Rev. Mod. Phys. **42**, 272 (1970).
[27]For definitions and results of abstract measure theory we refer to the book by A. C. Zaanen, *Integration* (North–Holland, Amsterdam, 1967).
[28]For definitions and results of probability theory we refer to the books by W. Feller, *An Introduction to Probability Theory and its Applications* (New York, 1968), Vols. 1 and 2; and M. Loève, *Probability Theory* (Springer, New York, 1977), Vols. 1 and 2.
[29]Note that the basis of the null event is not defined univocally, but the value of $I(\emptyset)$ is irrelevant since it is not included in (3) and (4). Also note that the energy associated to the global event, which appears only in (3), is a shift of the zero-energy level. Therefore the assumption $I(\mathscr{E}_\emptyset)\equiv 0$ was done without loss of generality.
[30]This and other subjects will be treated in a following paper III.
[31]See, e.g., L. Collatz, *Functional Analysis and Numerical Mathematics* (Academic, New York, 1966).
[32]We remark that coupling terms and external fields depend on the temperature through the included Boltzmann factor β.
[33]Extensive functions as the free energy will be always considered per site of the lattice.
[34]F. Peruggi, F. di Liberto, and G. Monroy, J. Phys. A **16**, 811 (1983).
[35]In other words the BPCA is exact on Bethe lattices and corresponds to the direct search for the fixed points of the system. This extends to every MRT–qSS Hamiltonian model (satisfying the PC property) the consistency checks given in Refs. 3, 4, 10, and 16 for the Ising and Potts models, and justifies the name of "Bethe lattice."
[36]Therefore the suggestion in Ref. 34, that the results of type (i) methods could correspond to non-RT probability measures, is erroneous.
[37]Also the condition that the interaction I is RT may be relaxed. See F. Peruggi, J. Phys. A **16**, L713 (1983).
[38]B. D. Hughes and M. Sahimi, J. Stat. Phys. **29**, 781 (1982).
[39]F. Peruggi, F. di Liberto, and G. Monroy, Physica A **123**, 175 (1984).
[40]J. Lebowitz and A. Martin-Löf, Commun. Math. Phys. **25**, 276 (1972); A. Messager and S. Miracle-Sole, Commun. Math. Phys. **40**, 187 (1975). Also see A. Coniglio, C. R. Nappi, F. Peruggi, and L. Russo, Commun. Math. Phys. **51**, 315 (1976).
[41]F. Peruggi, F. di Liberto, and G. Monroy, "Critical behavior in three-state Potts antiferromagnets" (to be published); F. Peruggi, F. di Liberto, and G. Monroy, "The Potts model on Bethe lattices. II: New results" (to be published).
[42]F. Peruggi, F. di Liberto, and G. Monroy, "The Potts model on Bethe lattices. III: The vector model" (in preparation).

Probability measures and Hamiltonian models on Bethe lattices. II. The solution of thermal and configurational problems

Fulvio Peruggi
Gruppo Nazionale di Struttura della Materia and Istituto di Fisica Teorica, Mostra d'Oltremare, pad. 19, 80125 Napoli, Italy

(Received 26 July 1983; accepted for publication 27 January 1984)

In a previous paper we introduced a method for the construction of rotationally and translationally invariant probability measures generated by one-step Markov Hamiltonian models on q-state-site Bethe lattices. Here, the corresponding thermal problems are solved by finding the relative free energy, which gives complete information on the properties of the models under study. Configurational problems also can be solved with the present tools. As an example, the solution of polychromatic correlated-site/random-bond percolation models is found.

PACS numbers: 05.50. + q, 02.50. + s, 64.60. − i, 05.70.Jk

I. INTRODUCTION

The interest of Bethe lattices and hierarchical lattices in statistical mechanics is related to their iterative topological properties.[1] In particular, the very simple structure of Bethe lattices suggests that Hamiltonian models on these kind of lattices can be solved with a reduced number of technical problems.

In a previous paper[2] (which will be called paper I, from now on) we studied the properties of general Hamiltonian models defined on q-state-site (qSS) Bethe lattices, and characterized by one-step Markov (M), rotationally and one- or two-step translationally invariant (RT) interactions. We showed that these models generate at least one RT Gibbs probability measure, and that only RT Gibbs probability measures, or their mixtures, are obtained if a certain piecewise contracting (PC) property holds. Furthermore, any RT probability measure was explicitly calculated in terms of a small number of known fundamental probabilities.

All previous methods of solution of Hamiltonian models on Bethe lattices[3] do not use explicitly probability measures. They must tackle the common technical problem that no sequence of finite trees tends to an infinite tree in the sense of Van Hove[2,4,5] (roughly speaking, this means that the bulk properties of Hamiltonian models on a finite tree are dominated by its surface). As a consequence, each of these methods can be classified according to whether, in the calculation of thermal extensive functions, (i) it takes into account the non-negligible surface effects, or (ii) it deletes them by means of proper artifices. The distinguishing feature of our approach is the exclusive use of Gibbs probability measures. This is very important when the limit towards an infinite system is taken. In fact the thermodynamic limit of probability measures does not depend on the topological properties of the sequences of finite subsystems chosen to take the limit.[4] On Bethe lattices this implies that our approach is (A) a rigorous limit procedure, which is (B) unaffected by surface effects. On the other hand, type (i) approaches satisfy only property (A), while type (ii) approaches satisfy only property (B). Our interpretation of these facts was extensively discussed in Secs. I and VII of paper I.

In the present article the results of paper I will be used for the solution of thermal and configurational problems (concerning models which satisfy the PC property). The localization procedure of extensive functions, which was suggested in paper I in order to obtain thermodynamic limits not affected by topological discrepancies, is introduced and applied. This provides a rigorous limit procedure which gives the internal energy, the entropy, and the free energy[6] in terms of the fundamental probabilities. The analytic expression of the free energy provides, in turn, other thermal functions of interest. Furthermore, a configurational approach is used to reduce the evaluation of the pair correlation function to the diagonalization of a known matrix.

As compared with other approaches, our procedure has two striking advantages: the properties of the fundamental probabilities are physically intuitive and their use is very simple; the free energy is easily obtained and gives complete information about the behavior of the models under investigation. As a matter of fact, the solution of the Potts model obtained by Peruggi, di Liberto, and Monroy[7] with (a preliminary version of) the present method can be compared with other papers[8] where both type (i) and (ii) approaches are followed.

A generalization of our method is considered in the final part of the present paper. It is used to solve a special configurational problem, i.e., polychromatic correlated-site/random-bond (CS/RB) percolation models, whose physical relevance in several theoretical and experimental contexts has been recently investigated.[9]

The outline of this paper is as follows. The entropy of M systems on qSS finite trees is found in Sec. II. The free energy of every MRT–qSS Hamiltonian model on Bethe lattices is calculated in Sec. III. Any thermal function of interest can be found by means of the free energy, except the pair correlation function (because of the RT property). However, it is found directly in Sec. IV by means of an alternative approach. Some applications of the preceding results to the Potts model, which also show another useful property of the general procedure, are collected in Sec. V. In Sec. VI we consider general infinite connected graphs, introduce q-state-site/two-state-bond events and the relevant probability measures, define formally the characteristic functions of percolation models, and prove a useful sum rule relating them. These subjects are used in Sec. VII, where MRT polychro-

 0022-2488/84/113316-08$02.50

matic CS/RB percolation models on Bethe lattices are solved. A sort of user's guide is given in Sec. VIII.

Finally, we remark that the present work is the natural continuation of paper I, whose reading must be considered as a prerequisite. The same notation will be used here, but, to avoid unnecessary repetitions, we shall not give again the definitions. However, reference to the relative sections (and to the formulas) of paper I will be made (the symbol I is added before their identification numbers).

II. THE ENTROPY OF M SYSTEMS ON qSS FINITE TREES

Let $T=(V_T, E_T)$ be a finite tree, R_T be the semiring of all qSS local events on V_T, and μ_T be a M probability measure defined on R_T (I/Sec. II). We introduce an ensemble Ω_T formed by ω copies of T. Let us consider an assignment of one configuration to each element of Ω_T. The most probable distribution of configurations (MPDC) in Ω_T is realized by all the assignments such that every configuration $\mathscr{C}_{V_T} \in \|V_T\| \subset R_T$ appears $\omega\mu_T(\mathscr{C}_{V_T})$ times in the ensemble. Starting from one of these assignments we can generate all the others using the $\omega!$ permutations of its configurations over the elements of Ω_T. Since the permutations which interchange identical configurations do not generate new assignments, the total number of distinct assignments satisfying the MPDC on Ω_T is given by

$$\Xi = \frac{\omega!}{\prod_{\mathscr{C}_{V_T}\in\|V_T\|}\{[\omega\mu_T(\mathscr{C}_{V_T})]!\}}. \tag{1}$$

Therefore the expected number of ways to obtain the MPDC for one copy of T is $\Xi^{1/\omega}$, and the entropy of the system (T, μ_T) is

$$\mathscr{S}_T = \kappa \ln \Xi^{1/\omega}$$

$$= \frac{\kappa}{\omega}\left(\ln(\omega!) - \sum_{\mathscr{C}_{V_T}\in\|V_T\|}\ln\{[\omega\mu_T(\mathscr{C}_{V_T})]!\}\right), \tag{2}$$

where κ is the Boltzmann constant. Letting $\omega\to\infty$, and applying Stirling's formula and the normalization condition

$$\sum_{\mathscr{C}_{V_T}\in\|V_T\|}\mu_T(\mathscr{C}_{V_T}) = 1, \tag{3}$$

relation (2) gives

$$\mathscr{S}_T = -\kappa \sum_{\mathscr{C}_{V_T}\in\|V_T\|}\mu_T(\mathscr{C}_{V_T})\ln\mu_T(\mathscr{C}_{V_T}). \tag{4}$$

Obviously, this expression is the same as the formal definition of the entropy.[4] Using [I/(9b)] we have

$$\ln\mu_T(\mathscr{C}_{V_T}) = \sum_{\langle hk\rangle\in E_T}\ln\mu_T(\mathscr{C}_{h\nu_h}\cap\mathscr{C}_{k\nu_k}) - \sum_{j\in V_T}\sigma_j \ln\mu_T(\mathscr{C}_{j\nu_j}). \tag{5}$$

On the other hand, the definition of conditional probabilities implies

$$\mu_T(\mathscr{C}_{V_T}) = \mu_T(\mathscr{C}_{j\nu_j})\mu_T(\mathscr{C}_{V'_T}|\mathscr{C}_{j\nu_j}) = \mu_T(\mathscr{C}_{h\nu_h}\cap\mathscr{C}_{k\nu_k})\mu_T(\mathscr{C}_{V''_T}|\mathscr{C}_{h\nu_h}\cap\mathscr{C}_{k\nu_k}), \tag{6}$$

where $V'_T \equiv V_T - \{j\}$; $V''_T \equiv V_T - \{h,k\}$; $\mathscr{C}_{V'_T}\cap\mathscr{C}_{j\nu_j} \equiv \mathscr{C}_{V''_T}\cap\mathscr{C}_{h\nu_h}\cap\mathscr{C}_{k\nu_k} \equiv \mathscr{C}_{V_T}$. Inserting (5) and (6) into relation (4), and introducing the function $\mathscr{L}(a)\equiv a\ln a$, the entropy of (T, μ_T) becomes

$$\mathscr{S}_T = \kappa\Bigg(-\sum_{\langle hk\rangle\in E_T}\sum_{\nu_h,\nu_k=1}^{q}\mathscr{L}(\mu_T(\mathscr{C}_{h\nu_h}\cap\mathscr{C}_{k\nu_k})) \times \sum_{\mathscr{C}_{V''_T}\in\|V''_T\|}\mu_T(\mathscr{C}_{V''_T}|\mathscr{C}_{h\nu_h}\cap\mathscr{C}_{k\nu_k}) + \sum_{j\in V_T}\sigma_j\sum_{\nu_j=1}^{q}\mathscr{L}(\mu_T(\mathscr{C}_{j\nu_j})) \times \sum_{\mathscr{C}_{V'_T}\in\|V'_T\|}\mu_T(\mathscr{C}_{V'_T}|\mathscr{C}_{j\nu_j})\Bigg). \tag{7}$$

The conditional probability that everything happens on V'_T (V''_T) provided that j is in the state ν_j (h is in the state ν_h and k in the state ν_k) is equal to 1. Therefore (7) reduces to

$$\mathscr{S}_T = \kappa\Bigg(-\sum_{\langle hk\rangle\in E_T}\sum_{r,s=1}^{q}\mathscr{L}(\mu_T(\mathscr{C}_{hr}\cap\mathscr{C}_{ks})) + \sum_{j\in V_T}\sigma_j\sum_{r=1}^{q}\mathscr{L}(\mu_T(\mathscr{C}_{jr}))\Bigg). \tag{8}$$

III. THE FREE ENERGY OF MRT HAMILTONIAN MODELS ON qSS BETHE LATTICES

Let us consider the sequence $\{T_n\}_{n=1}^{\infty}$ (I/Sec. IV) of finite trees tending to the Bethe lattice $L=(V,E)$. Suppose that the interaction I (I/Sec. II) defined on R generates the Hamiltonian with boundary terms [I/(25)], and let μ_n on R_n be the Gibbs probability measure (I/Sec. II) associated to the system $(T_n, \mathscr{H}_n + \Delta\mathscr{H}_n)$; $n=1,2,\ldots$. According to [I/(56)], and proceeding as for the calculation of [I/(61)], the internal energy on T_n is given by

$$\beta\mathscr{U}_n = -\sum_{\langle ij\rangle\in E_n}\sum_{r,s=1}^{q}K^x_{rs}\mu_n(\mathscr{C}_{ir}\cap\mathscr{C}_{js}) - \sum_{i\in V_n}\sum_{r=1}^{q}H^x_r\mu_n(\mathscr{C}_{ir}). \tag{9}$$

We see that the entropy (8) and the internal energy (9) relative to M Gibbs probability measures on finite trees are formed by bond and site terms. As suggested in (I/Sec. VII), we proceed to their localization associating bond contributions to the sites. As a matter of fact, we can write the internal energy relative to the site $i\in V^x_n$ ($x=e,o$) as

$$\beta\mathscr{U}_{in} = \begin{cases} -\sum_{j:\langle ij\rangle\in E_n}b_n(i,j)\sum_{r,s=1}^{q}K^x_{rs}\mu_n(\mathscr{C}_{ir}\cap\mathscr{C}_{js}) - \sum_{r=1}^{q}H^x_r\mu_n(\mathscr{C}_{ir}), & i\in V_n - \Delta V_n, \\ -b_n(i,j)\sum_{r,s=1}^{q}K^x_{rs}\mu_n(\mathscr{C}_{ir}\cap\mathscr{C}_{js}) - \sum_{r=1}^{q}H^x_r\mu_n(\mathscr{C}_{ir}), & i\in\Delta V_n, \end{cases} \tag{10}$$

where j in the third row is the unique site adjacent to i in T_n, and the topological weight factors $b_n(i,j)$ were introduced for

maximum generality in the assignment of bond terms. Three obvious physical requirements are imposed to the weights. They are non-negative. They must be normalized, i.e., $b_n(i,j) + b_n(j,i) \equiv 1$ for every $n \geqslant 1$ and $\langle ij \rangle \in E_n$. Furthermore, for every $\langle ij \rangle \in E$, we want that the $n \to \infty$ limit of $b_n(i,j)$ exists. Notice that these hypotheses and the rotational and one-step translational invariance of the Bethe lattice itself (I/Sec. III) imply

$$\lim_{n\to\infty} b_n(i,j) = \tfrac{1}{2}. \tag{11}$$

Similar weight factors (possibly the same) can be used for the entropy associated to the site i, which is

$$\frac{1}{k}\mathcal{S}_{in} = \begin{cases} -\sum_{j:\langle ij\rangle\in E_n} b'_n(i,j) \sum_{r,s=1}^{q} \mathcal{L}(\mu_n(\mathcal{C}_{ir}\cap\mathcal{C}_{js})) \\ \qquad + \sigma \sum_{r=1}^{q} \mathcal{L}(\mu_n(\mathcal{C}_{ir})), \quad i \in V_n - \Delta V_n \\ -b'_n(i,j) \sum_{r,s=1}^{q} \mathcal{L}(\mu_n(\mathcal{C}_{ir}\cap\mathcal{C}_{js})), \quad i \in \Delta V_n. \end{cases} \tag{12}$$

Now, suppose that the recursive relations [I/(35)] are PC (I/Sec. VI), and that the probability measure μ, obtained in the thermodynamic limit, describes a pure phase, i.e., it is MRT on L. In the case of two-step translational invariance, for every $i \in V^x$, we have

$$\beta\mathcal{U}^x = -\frac{\sigma+1}{2}\sum_{r,s=1}^{q} K^x_{rs} p^x_r p^x_{rs} - \sum_{r=1}^{q} H^x_r p^x_r, \tag{13a}$$

$$\frac{1}{k}\mathcal{S}^x = -\frac{\sigma+1}{2}\sum_{r,s=1}^{q} \mathcal{L}(p^x_r p^x_{rs}) + \sigma \sum_{r=1}^{q} \mathcal{L}(p^x_r), \quad x = e,o. \tag{13b}$$

It follows that the mean internal energy, the mean entropy, and the mean free energy on L are given by

$$\begin{aligned} \beta\mathcal{U} &= \tfrac{1}{2}(\beta\mathcal{U}^e + \beta\mathcal{U}^o), \\ k^{-1}\mathcal{S} &= \tfrac{1}{2}(k^{-1}\mathcal{S}^e + k^{-1}\mathcal{S}^o), \\ \beta\mathcal{F} &= \tfrac{1}{2}(\beta\mathcal{F}^e + \beta\mathcal{F}^o), \end{aligned} \tag{14}$$

where the sublattice free energies are defined as

$$\beta\mathcal{F}^x \equiv \beta\mathcal{U}^x - k^{-1}\mathcal{S}^x, \quad x = e,o. \tag{15}$$

If I and μ are one-step translationally invariant on L, there is no sublattice dependence, and all the upper indices disappear.[10,11]

The main features of the models under investigation will be obtained by studying the properties of the free energy. It is useful to consider also (a) its first derivatives, which (because of formal deductions) must satisfy the following relations:

$$-\frac{\partial\beta\mathcal{F}}{\partial H^x_r} = p^x_r, \quad \begin{cases} \text{constant temperature,} \\ x = e, o, \ r = 1,\ldots,q; \end{cases}$$

$$\beta\frac{\partial\beta\mathcal{F}}{\partial\beta} = \beta\mathcal{U}, \quad \text{constant fields;} \tag{16}$$

and (b) the second derivatives with respect to the fields or the temperature, i.e., the generalized isothermal susceptibilities and the specific heat at constant fields. Since the "magnetizations" and the internal energy are known functions, we see that an analytical or numerical check of relations (16) provides a test of internal consistence for our procedure. Furthermore, the second derivatives may be used to verify whether the usual convexity properties are satisfied by $\beta\mathcal{F}$. These controls where done in Ref. 7, where (a preliminary version of) the present method was applied to solve the Ising and Potts models: (16) were verified analytically; and evidence of nonconvexity of the free energy was found, because the "natural" phase (I/Appendix B) below the critical temperature is characterized by negative susceptibility. Actually we expect that, given a certain set of external fields and coupling terms, several phases characterized by distinct values of $\beta\mathcal{F}$ may be found in the system, for more general models too. We classify these RT Gibbs states (I/Sec. II) as unstable states (negative specific heat and/or susceptibilities, high free energy), metastable states (positive specific heat and susceptibilities, intermediate free energy), and stable states (positive specific heat and susceptibilities, the lowest free energy). We will give analytic expressions of the second derivatives of the free energy, and will study their properties, together with those of the above-mentioned phases, in a following paper III. Here we focus on the nonconvexity of the free energy, which would be irrelevant in the context of an approximation method, but needs special consideration in the present case, where we are dealing with a rigorous approach. From a mathematical point of view, this unusual property is easily understood. Remark that standard proofs of existence and convexity of the free energy on hypercubic lattices[4,5] make use of the thermodynamic limit on sequences of nested boxes, and subsequently are extended to any sequence which converges to the infinite lattice in the sense of Van Hove. On the other hand, we know that such sequences do not exist on Bethe lattices (I/Sec. VII), so that the previously described localization procedure must be introduced to avoid the topological discrepancy [I/(63d)]. As a consequence, generalization of the usual convexity proofs to the free energy on Bethe lattices is prevented. To understand this fact from a physical point of view, we remind the reader of the following exclusive properties of any Bethe lattice L.

(A) for every finite tree $T = (V_T, E_T) \subset L$ we have $|\partial V_T| > |\Delta V_T| > |V_T - \Delta V_T|$, i.e., any fixed or free boundary condition on T (I/Sec. II and IV) actually is a "bulk" condition.

(B) No closed walk exists on L, i.e., no correlations propagate from a site to itself along external paths.

(A) and (B) imply that on a finite tree one has the maximum sensitivity to the boundary conditions, which means that, in the thermodynamic limit, the Bethe lattice may be forced into Gibbs states which would be "unphysical" on standard lattices.

In conclusion, we deduce that the properties of Bethe lattices, which allowed us to solve MRT-qSS Hamiltonian models, also imply some peculiar features. Their study will be useful for intrinsic interest and in connection with nonequilibrium statistical mechanics. Furthermore, we emphasize that use of Bethe lattices as suggested in the Introduction of paper I, i.e., as a tool to extract information concerning models of interest on standard lattices, will also be fruitful. In such a case it is sufficient to concentrate one's

attention on the stable state(s) of the system, in order to find the nature of phase transitions, their location, and the complete phase diagram of the model under study. The equivalence with the Bethe–Peierls cluster approximation (I/Sec. VII) assures that our exact results, although described by classical critical exponents (I/Appendix B), will be good approximations on (bipartite) d-dimensional lattices for every $d \geqslant 3$, with increasing accuracy for high d (as a matter of fact see Ref. 7).

IV. THE PAIR CORRELATION FUNCTION OF MRT SYSTEMS ON qSS BETHE LATTICES

Let us consider the Bethe lattice $L = (V, E)$, and a MRT probability measure μ on the semiring R of all qSS local events on L. The pair correlation function relative to the sites i and j of L, in the states $\nu_i = r$ and $\nu_j = s$, is defined as

$$\mathscr{G}_{ij}(r,s) \equiv \mu(\mathscr{E}_{ir} \cap \mathscr{E}_{js}) - \mu(\mathscr{E}_{ir})\mu(\mathscr{E}_{js}). \tag{17}$$

Let $W = (\{i,h,k,\ldots z,j\}, \{\langle ih\rangle, \langle hk\rangle, \ldots, \langle zj\rangle\}) = (V_W, E_W) \subset L$ be the walk connecting i to j. Then, using [I/(9a)], we find

$$\mathscr{G}_{ij}(r,s) = \mu(\mathscr{E}_{ir})\Bigg(\sum_{\| V_W - \{i,j\} \|} \mu(\mathscr{E}_{h\nu_h} | \mathscr{E}_{ir}) \mu(\mathscr{E}_{k\nu_k} | \mathscr{E}_{h\nu_h}) \times \cdots \mu(\mathscr{E}_{js} | \mathscr{E}_{z\nu_z}) - \mu(\mathscr{E}_{js}) \Bigg). \tag{18}$$

If μ is one-step translationally invariant on L, we see that the function (18) does not depend on i and j, but on their distance $l \equiv l(i,j)$ (I/Sec. III). Moreover, defining the matrix

$$\langle r|\mathbf{T}|s\rangle \equiv p_{rs}, \quad r,s = 1,\ldots,q, \tag{19}$$

we find

$$\mathscr{G}_l(r,s) = p_r(\langle r|(\mathbf{T})^l|s\rangle - p_s). \tag{20}$$

Thus the evaluation of the pair correlation function reduces to the determination of the l th power of $\mathbf{T}$, i.e., to its diagonalization. This can be done numerically with standard procedures, after calculation of the fundamental probabilities [I/(45)], or analytically in some simple cases (an example is given in the next section).

If μ is two-step translationally invariant on L, it is useful to preserve the dependence of the pair correlation function by the distance l, only, defining

$$\mathscr{G}_l(r,s) \equiv \tfrac{1}{2}[\mathscr{G}^e_l(r,s) + \mathscr{G}^o_l(r,s)]. \tag{21}$$

Here the upper indices mean that the relative (partial) correlation function must be calculated by choosing the site i on the corresponding sublattice. This time we introduce

$$\langle r|\mathbf{T}^x|s\rangle \equiv p^x_{rs}, \quad x = e,o, \tag{22a}$$

$$\langle r|\mathbf{T}^{eo}|s\rangle \equiv \bar{p}_{rs} \equiv \sum_{t=1}^{q} p^e_{rt} p^o_{ts}, \quad r,s = 1,\ldots,q, \tag{22b}$$

so we find

$$\mathscr{G}^e_l(r,s) = \begin{cases} p^e_r[\langle r|(\mathbf{T}^{eo})^{l/2}|s\rangle - p^e_s], \\ \text{even } l, \\ p^e_r[\langle r|(\mathbf{T}^{eo})^{(l-1)/2}\mathbf{T}^e|s\rangle - p^o_s], \\ \text{odd } l, \end{cases} \tag{23}$$

and similar relations for $\mathscr{G}^o_l(r,s)$. Therefore the calculation of the pair correlation function reduces again to the diagonalization of a matrix.

V. SOME APPLICATIONS TO THE POTTS MODEL

The Potts model is recovered by our general formalism if we set

$$K^x_{rs} = K\delta_{rs}, \quad H^x_r = H\delta_{r1}, \quad x = e,o, \quad r,s = 1,\ldots,q. \tag{24}$$

We limit our considerations to the one-parameter model (I/Appendix D and Ref. 7). Its properties and the normalization conditions between the fundamental probabilities imply

$$\begin{aligned} &p^x_r = (1 - p^x_1)/(q-1), \quad p^x_{1r} = (1 - p^x_{11})/(q-1), \\ &\quad x = e,o, \quad r = 2,\ldots,q; \\ &p^x_{r1} = p^x_{21}, \quad p^x_{rr} = p^x_{22}, \\ &\quad x = e,o, \quad r = 3,\ldots,q; \\ &p^x_{rs} = (1 - p^x_{21} - p^x_{22})/(q-2), \\ &\quad x = e,o, \quad r \neq s, \quad r,s = 2,\ldots,q. \end{aligned} \tag{25}$$

In other terms only eight fundamental probabilities may be used to define each probability measure μ associated to the model, namely: $\{p^x_1, p^x_{11}, p^x_{21}, p^x_{22}\}_{x=e,o}$. These often reduce to four (when sublattice dependence disappears): in fact, a procedure similar to that in Appendix D of paper I can be used to show that two-step translationally invariant probability measures arise only in the antiferromagnetic ($K < 0$) model at sufficiently low temperatures. [Actually, (I/Corollary 1) also shows that the above-mentioned fundamental probabilities are not independent, but this fact will not be used here, since it is more convenient to follow the same notation as in Ref. 7.]

Relations (25) allow us to find explicitly the pair correlation function for the Potts model. It is easy to see that the matrices (22), in the present case, can be partitioned as

$$\begin{aligned} &\mathbf{T}^x = \begin{pmatrix} \alpha^x & \beta^x \mathbf{R} \\ \gamma^x \mathbf{C} & \zeta^x \mathbf{S} + \xi^x \mathbf{I} \end{pmatrix}, \quad x = e,o; \\ &\mathbf{T}^{eo} = \begin{pmatrix} \alpha_1 & \beta \mathbf{R} \\ \gamma \mathbf{C} & \zeta_1 \mathbf{S} + \xi \mathbf{I} \end{pmatrix}; \\ &(\mathbf{T}^{eo})^m = \begin{pmatrix} \alpha_m & \eta_m \beta \mathbf{R} \\ \eta_m \gamma \mathbf{C} & \zeta_m \mathbf{S} + \xi^m \mathbf{I} \end{pmatrix}, \quad m = 1,2,\ldots; \end{aligned} \tag{26}$$

where $\mathbf{I}$ is the $(q-1)\times(q-1)$ identity matrix, $\mathbf{R}$ is a row vector whose $q-1$ elements are equal to 1, $\mathbf{C} \equiv \tilde{\mathbf{R}}$, and $\mathbf{S} \equiv \mathbf{CR}$. The equalities

$$(\mathbf{T}^{eo})^m = (\mathbf{T}^{eo})^{m-1}\mathbf{T}^{eo} = \mathbf{T}^{eo}(\mathbf{T}^{eo})^{m-1} \tag{27}$$

imply

$$\alpha_m = \alpha_1 \alpha_{m-1} + (q-1)\beta\gamma\eta_{m-1}, \tag{28a}$$

$$\eta_m = \alpha_{m-1} + [(q-1)\zeta_1 + \xi]\eta_{m-1} = \alpha_1 \eta_{m-1} + (q-1)\zeta_{m-1} + \xi^{m-1}, \tag{28b}$$

$$\zeta_m = \beta\gamma\eta_{m-1} + [(q-1)\zeta_1 + \xi]\zeta_{m-1} + \zeta_1 \xi^{m-1}. \tag{28c}$$

Since (28b) gives

$$\zeta_{m-1} = [1/(q-1)]\{\alpha_{m-1} - [\alpha_1 - (q-1)\zeta_1 - \xi]\eta_{m-1} - \xi^{m-1}\}, \tag{29}$$

(28c) can be neglected.[12] Therefore our problem reduces to the diagonalization of a 2×2 matrix $\mathbf{X}$, because (28a) and (28b) imply

$$\begin{pmatrix}\alpha_m \\ \eta_m\end{pmatrix} = \mathbf{X}\begin{pmatrix}\alpha_{m-1} \\ \eta_{m-1}\end{pmatrix} = \mathbf{X}^{m-1}\begin{pmatrix}\alpha_1 \\ 1\end{pmatrix}, \tag{30}$$

where

$$\mathbf{X} \equiv \begin{pmatrix}\alpha_1 & (q-1)\beta\gamma \\ 1 & (q-1)\zeta_1 + \xi\end{pmatrix}. \tag{31}$$

Expressing all the terms as functions of the jump probabilities $\bar{p}_{rs}$ defined by (22b), and applying elementary algebra, one finds

$$\begin{aligned}\langle r|(\mathbf{T}^{eo})^m|s\rangle &= p_s^e + [1/(q-1)][\pi(r,s)\Delta^m \\ &\quad + \bar{\pi}(r,s)(q-2)\xi^m], \\ \langle r|(\mathbf{T}^{eo})^m\mathbf{T}^e|s\rangle &= p_s^o + [1/(q-1)][\pi(r,s)\Delta^m\Delta^e \\ &\quad + \bar{\pi}(r,s)(q-2)\xi^m\xi^e],\end{aligned} \tag{32}$$

where

$$\begin{aligned}&\Delta \equiv \bar{p}_{11} - \bar{p}_{21}, \\ &\xi \equiv [(q-1)\bar{p}_{22} + \bar{p}_{21} - 1]/(q-2) \\ &\Delta^x \equiv p_{11}^x - p_{21}^x, \\ &\xi^x \equiv [(q-1)p_{22}^x + p_{21}^x - 1]/(q-2) \\ &\pi(r,s) \equiv (q\delta_{1s} - 1)(\delta_{r1} - p_1^e), \\ &\bar{\pi}(r,s) \equiv (1-\delta_{r1})(1-\delta_{1s})/[(q-1)\delta_{rs} - (q-2)], \\ &\qquad x = e,o.\end{aligned} \tag{33}$$

Finally, noticing that $\Delta = \Delta^e\Delta^o$, $\xi = \xi^e\xi^o$, and introducing (32) in (23) we obtain

$$\mathscr{G}_l^e(r,s) = \begin{cases} \dfrac{1-p_1^e}{(q-1)^2}[\epsilon(r,s)p_1^e(\Delta^e\Delta^o)^{l/2} \\ \quad + \bar{\epsilon}(r,s)(\xi^e\xi^o)^{l/2}], \quad \text{even } l, \\ \dfrac{1-p_1^e}{(q-1)^2}[\epsilon(r,s)p_1^e(\Delta^e)^{(l+1)/2}(\Delta^o)^{(l-1)/2} \\ \quad + \bar{\epsilon}(r,s)(\xi^e)^{(l+1)/2}(\xi^o)^{(l-1)/2}], \quad \text{odd } l, \end{cases} \tag{34}$$

where

$$\begin{aligned}&\epsilon(r,s) \equiv q^2\delta_{r1}\delta_{1s} - q(\delta_{r1} + \delta_{1s}) + 1, \\ &\bar{\epsilon}(r,s) \equiv (q-1)\delta_{rs} - q\delta_{r1} + (\delta_{r1} + \delta_{1s}) - 1.\end{aligned} \tag{35}$$

Besides its physical interest,[13] the pair correlation function is relevant since it gives another check on the internal consistency of our method. As a matter of fact, using the standard and staggered fluctuation relations

$$\begin{aligned}&\chi = \sum_{j\in V}\mathscr{G}_l(1,1), \\ &\chi_S = \sum_{j\in V^e}\mathscr{G}_l(1,1) + \sum_{s=2}^{q}\sum_{j\in V^o}\mathscr{G}_l(1,s),\end{aligned} \tag{36}$$

it is easy to verify that one recovers the standard and staggered susceptibilities obtained in Ref. 7 through differentiation of the free energy.

VI. SITE/BOND EVENTS AND CHARACTERISTIC FUNCTIONS OF PERCOLATION MODELS

In this section we consider the infinite, connected, and locally finite graph $G = (V,E)$ of (I/Sec. II), and make the auxiliary hypothesis that it has no multiedges.[14] We introduce more general events on G than those defined in (I/Sec. II) and studied up to now. Besides the q states assumed by each site in V, also bonds in E are allowed to take up two states, which we call "active" and "nonactive," and label with the indices 1,0, respectively. Bond configurations are represented by the ordered partitions $\{E^{(0)}, E^{(1)}\}$ of E in two subsets (of nonactive and active bonds), whose collection will be denoted by $\|E\|$. Given two finite subsets B_0, B_1 of E, the local bond event $\{B_0, B_1\}$ is formed by all the configurations in $\|E\|$ such that every $\langle ij\rangle \in B_0$ is nonactive, and every $\langle ij\rangle \in B_1$ is active. The same procedure as in (I/Sec. II) can be used to define the null and global events, and the semiring $\overline{R}$ of all local bond events. Site/bond configurations (site/bond local events) on G are the elements of the set $\|G\| \equiv \|V\| \otimes \|E\|$ (of the semiring $\hat{R} \equiv R \otimes \overline{R}$).

Let us introduce, now, a percolation model[15] on the graph G. We consider a (self-avoiding) walk of length $l \geqslant 0$ in G, i.e., a subgraph $W_l = (\{v_0,v_1,...,v_l\}, \{\langle v_0v_1\rangle, \langle v_1v_2\rangle, ..., \langle v_{l-1}v_l\rangle\}) = (V_W, E_W)$. We say that W_l is an r-walk in each configuration of G such that all the sites in V_W are in the state r, and all the bonds in E_W are active. Given a configuration in $\|G\|$, we define an r-cluster as each maximal subgraph[14] of G such that all its sites are connected by r-walks. The size m of a cluster is, by definition, the number of its sites. The following events are needed to define percolative functions:

$$\begin{aligned}&\mathscr{C}_i^m \equiv \{c \in \|G\| : i \in V \text{ belong to a cluster of size } m\}, \\ &\mathscr{C}_i^F \equiv \bigcup_{m=1}^{\infty}\mathscr{C}_i^m, \\ &\mathscr{C}_i^\infty \equiv \|G\| - \mathscr{C}_i^F, \\ &\mathscr{C}_{ij}^F \equiv \{c \in \|G\| : i \in V \text{ and } j \in V \text{ belong to the same finite cluster}\}.\end{aligned} \tag{37}$$

Note that these are generalized local events or nonlocal events which belong to the smallest σ-field[16] containing $\hat{R}$. Therefore they are $\hat{\mu}$-measurable events[16] for every probability measure $\hat{\mu}$ defined on $\hat{R}$. The percolation probability, the pair connectedness, the mean size of finite clusters, and the mean number of finite clusters are, respectively, defined as

$$\begin{aligned}&P_{ir} \equiv \hat{\mu}(\mathscr{C}_i^\infty | \mathscr{C}_{ir}) = 1 - \hat{\mu}(\mathscr{C}_i^F | \mathscr{C}_{ir}), \\ &P_{ijr} \equiv \hat{\mu}(\mathscr{C}_{ij}^F \cap \mathscr{C}_{ir}), \\ &S_{ir} \equiv \frac{\sum_{m=1}^{\infty} m\hat{\mu}(\mathscr{C}_i^m \cap \mathscr{C}_{ir})}{\sum_{m=1}^{\infty}\hat{\mu}(\mathscr{C}_i^m \cap \mathscr{C}_{ir})}, \\ &N_{ir} \equiv \sum_{m=1}^{\infty}\frac{1}{m}\hat{\mu}(\mathscr{C}_i^m \cap \mathscr{C}_{ir}), \\ &\qquad i \in V, \quad j \in V, \quad r = 1,...,q.\end{aligned} \tag{38}$$

Clearly, each cluster species r can be regarded as a color, thus justifying the name of polychromatic percolation model.[17] Also note that the dependence of the functions (38) by the choice of the site $i \in V$ is due to the absence of hypotheses about the spatial symmetries of G and/or $\hat{\mu}$. Furthermore,

we remark that, here and in the following, no requirements are made about the properties of $\hat{\mu}$ (e.g., there is no need that $\hat{\mu}$ is m-step Markov on G).

Finally, we prove a useful sum rule which relates the mean size of finite clusters to the pair connectedness and the percolation probability. Let γ_{ir}^{m} and γ_{ijr}^{F} be, respectively, the indicators of the events $\mathscr{C}_i^m \cap \mathscr{C}_{ir}$ and $\mathscr{C}_{ij}^F \cap \mathscr{C}_{ir}$, i.e., the functions which assume the value 1 on the configurations contained in those events, and the value 0 on the others. The following relation holds for every site/bond configuration of G and for every $i \in V$:

$$\sum_{m=1}^{\infty} (m-1)\gamma_{ir}^{m} = \sum_{j \in V - \{i\}} \gamma_{ijr}^{F}. \tag{39}$$

Averaging over all the configurations with respect to the probability measure $\hat{\mu}$ (see the definition [I/(57)]), we have

$$\sum_{m=1}^{\infty} m \langle \gamma_{ir}^{m} \rangle \hat{} = \sum_{m=1}^{\infty} \langle \gamma_{ir}^{m} \rangle \hat{} + \sum_{j \in V - \{i\}} \langle \gamma_{ijr}^{F} \rangle \hat{}, \tag{40}$$

which is equivalent to

$$S_{ir} = 1 + \frac{\Sigma_{j \in V - \{i\}} P_{ijr}}{\Sigma_{m=1}^{\infty} \hat{\mu}(\mathscr{C}_i^m \cap \mathscr{C}_{ir})}. \tag{41}$$

The events $\{\{\mathscr{C}_i^m \cap \mathscr{C}_{ir}\}_{r=1}^{q}\}_{m=1}^{\infty}$ are mutually disjoint. Therefore, using the countable additivity[16] of $\hat{\mu}$, we obtain the normalization condition

$$\sum_{m=1}^{\infty} \hat{\mu}(\mathscr{C}_i^m \cap \mathscr{C}_{ir}) = \hat{\mu}(\mathscr{C}_i^F \cap \mathscr{C}_{ir}) = \hat{\mu}(\mathscr{C}_{ir}) \hat{\mu}(\mathscr{C}_i^F | \mathscr{C}_{ir}), \tag{42}$$

which implies

$$S_{ir} = 1 + \frac{\Sigma_{j \in V - \{i\}} P_{ijr}}{\hat{\mu}(\mathscr{C}_{ir})(1 - P_{ir})}. \tag{43}$$

This relation was already obtained by Essam[18] for random systems.

Up to now we have considered standard site/bond percolation. However the definitions and results above may be easily generalized to $\mathscr{A}\mathscr{B}$ site/bond percolation models.[19,20] Let us subdivide the set $\{1,2,...,q\}$ of site states in two disjoint classes $\mathscr{A}$ and $\mathscr{B}$. For every walk $W_l = (V_W, E_W) \subset G$, of length $l \geqslant 0$, we will say that it is an $\mathscr{A}\mathscr{B}$-walk in each configuration of G such that the sites in (the ordered set) V_W are alternately in states belonging to $\mathscr{A}$ and $\mathscr{B}$, while all the bonds in E_W are active. Given a configuration in $\|G\|$, we define an $\mathscr{A}\mathscr{B}$-cluster as each maximal subgraph of G such that all its sites are connected by $\mathscr{A}\mathscr{B}$-walks. Relations (38) can be used to define $\mathscr{A}\mathscr{B}$-percolation functions, too, when the following changes are done in their interpretation. The index r now stands for the classes $\mathscr{A}$ or $\mathscr{B}$; in the definitions of $\mathscr{C}_i^m$ and $\mathscr{C}_{ij}^F$ the word cluster must be replaced by $\mathscr{A}\mathscr{B}$-cluster; the event $\mathscr{C}_{ir}$ is defined as

$$\mathscr{C}_{ir} \equiv \{c \in \|G\| : v_i \in r\}, \quad r = \mathscr{A}, \mathscr{B}. \tag{44}$$

It is clear that the sum rule (43) remains true for $\mathscr{A}\mathscr{B}$-percolative functions. We remark again that in standard polychromatic percolation models we can find simultaneously in G q distinct colors of clusters labeled by the index r. On the contrary, in any $\mathscr{A}\mathscr{B}$-percolation model the graph G is completely filled by the unique species of $\mathscr{A}\mathscr{B}$-clusters defined above.

VII. THE SOLUTION OF MRT POLYCHROMATIC CS/RB PERCOLATION MODELS ON BETHE LATTICES

Suppose that the site/bond events, just introduced in the preceding section, are defined on a Bethe lattice $L = (V,E)$ of coordination number $\sigma + 1$. We make three hypotheses about the structure of the probability measure $\hat{\mu}$ on $\hat{R}$. The first one is that site states do not depend on bond states, and vice versa. Therefore we have

$$\hat{\mu} = \mu \otimes \bar{\mu}, \tag{45}$$

i.e., $\hat{\mu}$ is a product measure whose components are two independent probability measures μ on R and $\bar{\mu}$ on $\bar{R}$. The second hypothesis is that μ is MRT. The third one is that bond states are randomly distributed, and are rotationally and one-step translationally invariant on L. This gives explicitly $\bar{\mu}$:

$$\bar{\mu}(\{B_0,B_1\}) = p_b^{|B_1|}(1-p_b)^{|B_0|}, \quad \varnothing \neq \{B_0,B_1\} \in \bar{R}, \tag{46}$$

where the external parameter p_b is the probability that a bond is active. The physical interest of the associated CS/RB percolation models in several problems is discussed elsewhere.[9] Here we solve only standard (polychromatic) percolation models, assuming that μ is one-step translationally invariant. The solution of $\mathscr{A}\mathscr{B}$-percolation models, in the case of a two-step translationally invariant probability measure μ, is given by Peruggi, di Liberto, and Monroy.[20] Mixed cases follow easily.

From a "percolative" point of view we are interested only in connectivity properties, which will be completely described by means of the site probabilities $\{p_r\}_{r=1}^{q}$, and the transition probabilities

$$t_{rr} \equiv p_b p_{rr}, \quad r = 1,...,q. \tag{47}$$

As a matter of fact, let us evaluate the probability of a finite r-cluster. In formal terms: given the finite connected subgraph $T = (V_T, E_T)$ of L, we want to calculate the measure of the generalized local site/bond event $\mathscr{C}_{Tr} \equiv \{c \in \|L\| : T \text{ is a } r\text{-cluster}\}$. The perimeter of T is by definition the pair $(\partial V_T, \partial E_T)$, where ∂V_T is the perimeter of V_T as defined in (I/Sec. IV), while $\partial E_T \subset E$ is the set of all the bonds which connect sites in V_T to sites in ∂V_T. For fixed size $m = |V_T| = |E_T| + 1$, we have $|\partial V_T| = |\partial E_T| = m(\sigma+1) - 2(m-1)$ for every shape of T. Applying a simple generalization of [I/(9a)] we have

$$\hat{\mu}(\mathscr{C}_{Tr}) = p_r t_{rr}^{m-1}(1-t_{rr})^{m(\sigma-1)+2}. \tag{48}$$

The percolation threshold of our model can be found by considering the realization of an infinite r-cluster as a branching process[21] which gives rise to at least one r-walk of infinite length. Suppose this r-walk has reached the site $i \in V$: the process does not extinguish if at least one of the following σ steps reaches a site in the state r passing through an active bond. Therefore infinite r-clusters exist if

$$\sigma t_{rr} \geqslant 1. \tag{49}$$

The equality gives the condition which characterizes the r-threshold.

Now, we calculate the percolative functions defined in the preceding section: to do this we use the generalization of a method previously introduced for random systems.[18] Let u, v be two adjacent sites of L, and consider the nonlocal event $\mathscr{C}_{uv}^{W} \equiv \{c \in \|L\| :$ the r-walks ($r = 1,...,q$) starting from u

towards v extinguish in a finite number of steps}. The knowledge of the probabilities

$$Q_r \equiv \hat{\mu}(\mathscr{E}^W_{uv} | \mathscr{E}_{ur}), \quad r = 1,...,q \tag{50}$$

is essential, because they allow us to distinguish the contributions of finite r-clusters to the percolative functions. Each Q_r can be obtained by the corresponding recursion relation

$$Q_r = (1 - t_{rr}) + t_{rr} Q_r^{\sigma}, \quad r = 1,...,q, \tag{51}$$

which means that a finite r-walk starting from the site u (provided $\nu_u = r$) towards v extinguishes at the first step, or after a finite number of steps in the σ branches departing from v. Equation (51) always admits the solution $Q_r = 1$. However, when condition (49) is satisfied, another solution $Q_r \leqslant 1$ appears, which is the physical one in that range since it goes to 1 for $t_{rr} \to 1/\sigma$, and goes to 0 for $t_{rr} \to 1$.

In terms of Q_r the percolation probability is given by

$$P_r = 1 - Q_r^{\sigma+1}, \quad r = 1,...,q \tag{52}$$

(we omit the index representing the reference site because of the one-step RT property of $\hat{\mu}$ and L itself).

Let us consider two sites i, j of L at distance l. The perimeter of the walk connecting i to j is formed by $(l-1)(\sigma-1) + 2\sigma$ sites (bonds). Then the pair connectedness is

$$P_{ijr} = p_r t_{rr}^{l} Q_r^{(l-1)(\sigma-1)+2\sigma}, \quad r = 1,...,q. \tag{53}$$

The mean size of finite r-clusters can be obtained by using the preceding results and the sum rule (43):

$$S_r = 1 + \frac{\sum_{l=1}^{\infty} (\sigma+1)\sigma^{l-1} p_r t_{rr}^{l} Q_r^{(l-1)(\sigma-1)+2\sigma}}{p_r Q_r^{\sigma+1}}$$

$$= \frac{1 + t_{rr} Q_r^{\sigma-1}}{1 - \sigma t_{rr} Q_r^{\sigma-1}}, \quad r = 1,...,q. \tag{54}$$

The mean number of finite r-clusters needs a special procedure. We will use the relation which gives the cyclomatic number[14] C_F of any finite graph $G_F = (V_F, E_F)$ in terms of $|V_F|$, $|E_F|$, and the number M_F of its components[14]:

$$C_F = |E_F| - |V_F| + M_F. \tag{55}$$

Let us consider the tree $T_n = (V_n, E_n)$ of the sequence $\{T_n\}_{n=1}^{\infty}$ (I/Sec. IV), and the configuration $c_0 \in \| L \|$ of the Bethe lattice. We denote V_{nr} (E_{nr}) the set of all the sites (bonds) in V_n (E_n) which belong to finite r-clusters in c_0. Applying (55) to the graph $G_{nr} = (V_{nr}, E_{nr})$ we have

$$M_{nr} = |V_{nr}| - |E_{nr}| \tag{56}$$

($C_{nr} = 0$ because $G_{nr} \subset L$ has no cycles[14]). Let γ^F_{ir} and $\gamma^F_{\langle ij \rangle r}$ be the indicators of the events $\mathscr{E}^F_i \cap \mathscr{E}_{ir}$ and $\mathscr{E}^F_{\langle ij \rangle} \cap \mathscr{E}_{ir}$ $\equiv \{c \in \| L \| : \langle ij \rangle$ belongs to a finite r-cluster$\}$, respectively. Then, introducing topological weights as in Sec. III, (56) becomes

$$M_{nr} = \sum_{i \in V_n} \left(\gamma^F_{ir} - \sum_{j:\langle ij \rangle \in E_n} b''_n(i,j) \gamma^F_{\langle ij \rangle r} \right). \tag{57}$$

Averaging over all the configurations of L, we obtain the expected number N_{nr} of finite r-clusters on L having at least one site in T_n. At local level this is written as

$$N_{nir} = \hat{\mu}(\mathscr{E}^F_i \cap \mathscr{E}_{ir}) - \sum_{j:\langle ij \rangle \in E_n} b''_n(i,j) \hat{\mu}(\mathscr{E}^F_{\langle ij \rangle} \cap \mathscr{E}_{ir}), \tag{58}$$

which, in the $n \to \infty$ limit, gives

$$N_r = p_r Q_r^{\sigma+1} - [(\sigma+1)/2] p_r t_{rr} Q_r^{2\sigma}, \quad r = 1,...,q, \tag{59}$$

for every $i \in V$.

It is easy to check that all previous solutions[22] of percolation models on Bethe lattices are recovered.

VIII. FINAL REMARKS

This section is devoted to those physicists whose main interest is related to the use of our procedure in practical cases, rather than its mathematical details. For maximum clarity and concision we give the following menu concerning the utilization of the general results contained in paper I and the present paper II.

(1) Write the one-step Markov, rotationally and one- or two-step translationally invariant (MRT) Hamiltonian of interest in the form [I/(22)]. (Boundary terms as in [I/(25)] are not needed. Note that site-diluted annealed Hamiltonians may be considered, too.)

(2) Verify the piecewise contracting (PC) property defined in (I/Sec. VI). (When this control is done numerically, check that any seed introduced into the recursive relations [I/(35)] tends to a fixed point as defined by [I/(37)] and [I/(38)].)

(3) Find the fixed point(s) corresponding to the given set of coupling term(s) and external field(s). (Note that both of them depend on the temperature via the included Boltzmann factor.)

(4) Evaluate the fundamental probabilities [I/(45)]. [This gives the MRT probability measure(s) describing the system in the prescribed conditions.]

(5) Evaluate the internal energy, the entropy, and the free energy of the system, as given by relations (13)–(15).

(6) Evaluate other thermal functions of interest. (If this is done analytically, use Eqs. [I/(37)] and [I/(38)] to find the derivatives of the fixed point parameters with respect to the external fields and the temperature.)

(7) Diagonalize the matrix (22b) or (19) and find the pair correlation function (20) or (21)–(23).

(8) Define the active bond probability p_b, evaluate the transition probabilities (47), and solve equations (51). The percolative functions are given by (52)–(54) and (59).

Finally, let us emphasize again that the knowledge of the probability measure(s) describing a certain system is sufficient, on principle, for the solution of every problem concerning that system. The procedure needed in such a case is the same we applied for the solution of thermal and percolative problems. One has to write all the "questions" in terms of local and/or generalized local events, and to find their probability measures by means of relations [I/(9)] (the measures of nonlocal events must be expressed in terms of the above probabilities, directly or through iterative equations).

ACKNOWLEDGMENTS

The author would like to thank Professor A. Coniglio, Professor B. Preziosi, Professor L. Russo, G. Monroy, and F. di Liberto for stimulating discussions. The author is very indebted and very grateful to Pina for her patience.

This work was supported by Ministero della Pubblica Istruzione (MPI) and the Consiglio Nazionale delle Ricerche (CNR).

[1]As regards to Bethe lattices see Ref. 14; Ref. 2; and references therein. As regards to hierarchical lattices see M. Kaufman and R. B. Griffiths, Phys. Rev. B **24**, 496 (1981); R. B. Griffiths and M. Kaufman, Phys. Rev. B **26**, 5022 (1982); M. Kaufman and R. B. Griffiths, Phys. Rev. B **26**, 5282 (1982); and references therein.

[2]F. Peruggi, J. Math. Phys. **25**, 3303 (1984).

[3]An extended set of references will be found in Ref. 2.

[4]D. Ruelle, "Thermodynamic formalism" *Encyclopedia of Mathematics and its Applications* (Addison–Wesley, Reading, MA, 1978), Vol. 5.

[5]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969); R. B. Griffiths, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 1; and references therein.

[6]Extensive functions on infinite lattices will be always considered per site.

[7]F. Peruggi, F. di Liberto, and G. Monroy, J. Phys. A **16**, 811 (1983).

[8]Y. K. Wang and F. Y. Wu, J. Phys. A **9**, 593 (1976); H. G. Baumgärtel and E. Müller-Hartmann, Z. Phys. B **46**, 227 (1982); also see H. Moraal, Z. Phys. B **45**, 237 (1982).

[9]A. Coniglio, H. E. Stanley, and W. Klein, Phys. Rev. Lett. **42**, 518 (1979); A. Coniglio and W. Klein, J. Phys. A **13**, 2775 (1980); A. Coniglio, F. di Liberto, G. Monroy, and F. Peruggi, Phys. Lett. A **87**, 189 (1982); A. Coniglio and F. Peruggi, J. Phys. A **15**, 1873 (1982); A. Coniglio, H. E. Stanley, and W. Klein, Phys. Rev. B **25**, 6805 (1982). Also see H. E. Stanley, "Disordered systems and Localization," in *Proceedings of the International Conference*, Rome 1981, edited by C. Castellani, C. di Castro, and L. Peliti (Springer-Verlag, Heidelberg, 1981); D. Stauffer, A. Coniglio, and M. Adam, Adv. Polymer Sci. **44**, 103 (1982); and references therein.

[10]In such a case there is no distinction between sublattice functions and mean functions. Also remark that (13a) reduces to relation [I/(60)], while (13b) is the same as the entropy given by the first-order approximation of the cluster variation method (see Ref. 11), which is known to be equivalent to the Bethe–Peierls cluster approximation, whose equivalence with our probability measure approach was shown in (I/Sec. VII).

[11]R. Kikuchi, Phys. Rev. **81**, 988 (1951); R. Kikuchi, J. Chem. Phys. **53**, 2713 (1970); and references therein.

[12]It can be shown that (28c) and (29) are equivalent.

[13]We remark that this expression of the pair correlation function is meaningful for every real K and H, in the ordered phases too.

[14]We refer to the terminology of graph theory established by J. W. Essam and M. E. Fisher, Rev. Mod. Phys. **42**, 272 (1970).

[15]About this topic, see the following reviews: D. Stauffer, Phys. Rep. **54**, 1 (1979); J. W. Essam, Rep. Progr. Phys. **43**, 833 (1980); "Percolation structures and processes," in *Annals of the Israel Physical Society*, edited by G. Deutscher, R. Zallen, and J. Adler (Adam Hilger, Bristol, 1983), Vol. 5; also see Ref. 18.

[16]For definitions and results of abstract measure theory we refer to the book by A. C. Zaanen, *Integration* (North-Holland, Amsterdam, 1967).

[17]See, e.g., R. Zallen, Phys. Rev. B **16**, 1426 (1977); J. W. Halley and W. K. Holcomb, Phys. Rev. Lett. **40**, 1670 (1978); H. E. Stanley, J. Phys. A **12**, L329 (1979); H. E. Stanley and J. Texeira, J. Chem. Phys. **73**, 3404 (1980); A.Coniglio and F. Peruggi, J. Phys. A **15**, 1873 (1982).

[18]J. W. Essam, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 2.

[19]M. Barma, J. W. Halley, and M. J. Stephen, "AB-percolation," 1981 (unpublished); A. Coniglio, F. di Liberto, G. Monroy, and F. Peruggi, Phys. Lett. A **87**, 189 (1982); G. Monroy, F. di Liberto, and F. Peruggi, Z. Phys. B **49**, 239 (1982); L. Turban, J. Phys. A **16**, L 15 (1983); F. Sevšek, J. M. Debierre, and L. Turban, J. Phys. A **16**, 801 (1983); C. Amitrano, F. di Liberto, R. Figari, and F. Peruggi, J. Phys. A **16**, 3925 (1983).

[20]F. Peruggi, F. di Liberto, and G. Monroy, Physica A **123**, 175 (1984).

[21]See, e.g., W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1968), Vols. 1 and 2; or M. Loeve, *Probability Theory* (Springer, New York, 1977), Vols. 1 and 2.

[22]M. E. Fisher and J. W. Essam, J. Math. Phys. **2**, 609 (1961) for random systems (also see Refs. 18 and 15); A. Coniglio, J. Phys. A **8**, 1773 (1975), and A. Coniglio, Phys. Rev. B **13**, 2194 (1976), for Ising-correlated-site systems; A. Coniglio (private communication, 1981) for Potts-correlated-site systems.

Gribov copies and the Faddeev–Popov formula in lattice gauge theories

Bruce Sharpe
Department of Mathematics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Y4

(Received 3 November 1983; accepted for publication 15 March 1984)

We present a rigorous, nonperturbative derivation of a lattice version of the Faddeev–Popov integral. This derivation shows that Gribov copies can occur in the lattice theory for certain gauges, but these copies do not affect *normalized* functional integrals in the lattice theory. Furthermore, taking the formal limit as the lattice spacing tends to zero leads to the usual continuum Faddeev–Popov integral.

PACS numbers: 11.15.Ha

I. INTRODUCTION

The gauge-fixed functional integral for nonabelian gauge field theories as formulated by Faddeev and Popov[1] and generalized by 't Hooft[2] appears to provide a suitable starting point for the rigorous construction of such a theory. However, the discovery of Gribov copies[3] has shown that there are difficulties with the Faddeev–Popov (FP) method even at a formal level.[4]

The simplest reworking of the FP argument which takes into account Gribov copies unfortunately leads to a replacement of the elegant FP functional integral with an unwieldy expression involving the inverse of a sum of inverse determinants.[5] Several other modifications of and alternatives to the FP technique have also been put forward.[6]

The references just cited amply demonstrate that Gribov copies have an effect on the functional integrals of the theory, but so far no one has been able to say in any generality what that effect is. Part of the difficulty arises from trying to work in the continuum theory where many of the fundamental quantities are not well defined.

In this paper, we formulate the FP technique for lattice gauge theories and give a rigorous proof of a FP formula in that setting. We show that Gribov copies can also occur in the lattice theory and determine what their influence on the relevant functional integrals is. Our work is based on some observations of Hirschfeld,[7] who argued that the FP formula is correct in spite of the existence of Gribov copies.

Our analysis supports this conclusion. Specifically, we show that in lattice gauge theories, the functional integral of a gauge invariant function differs from the value it would have if there were no Gribov copies only by a multiplicative constant. This constant, which we call the *gauge degree*, depends only on the choice of the gauge-fixing condition. As a consequence, the FP formula, because it involves a *ratio* of such integrals, is not affected by Gribov copies.

This conclusion, that Gribov copies do not invalidate the FP formula, is of course a very desirable one. However, it should be noted that it depends in an essential way on cancellations between positive and negative contributions to the FP integral and the lack of a positive density would be a hindrance in numerical studies of these integrals. The approaches taken in the articles of Ref. 6 do not share this difficulty. In particular, in the last article of Ref. 6, the FP integral is truncated in such a way that the density that remains is manifestly positive.

In Sec. II we give a more detailed explanation of the claim that the unmodified FP formula is correct and present an intuitive argument for this conclusion. A rigorous derivation of the lattice FP formula is given in Sec. III. In Sec. IV we discuss various examples and applications. As a partial justification of our choice of definitions in the lattice theory, we show in Sec. V that the formal continuum limit of the lattice expressions leads to the usual continuum FP formula. We derive an expression for the gauge degree in the Appendix.

II. FORMAL DESCRIPTION

In this section, we give the idea of our lattice argument using the more familiar language of the continuum theory.

The quantity of central interest is the formal functional integral for the expectation of a gauge invariant function. In the Euclidean theory this is

$$\langle f\rangle = \frac{\int f(A)\det M(A)E\circ F(A)e^{-S(A)}\mathscr{D}A}{\int \det M(A)E\circ F(A)e^{-S(A)}\mathscr{D}A}. \tag{1}$$

Here, f and the action S are invariant under the gauge transformation $A\to {}^{g}A$. The measure $\mathscr{D}A$ is the (nonexistent) infinite product of Lebesgue measures $\Pi_{x,\mu,j}dA_\mu^j(x)$. The gauge-fixing function F might be for example $F(A)=\partial_\mu A_\mu$ (Landau gauge) or $F(A)=A_0$ (axial gauge); the factor $E\circ F(A)$ is typically a gauge-fixing term $\delta(F(A))$ or a damping factor $\exp(a\int \mathrm{Tr}\, F(A)^2\,dx)$. We are denoting by $M(A)$ the usual FP operator, which arises as the Jacobian for a change of variables determined by F. We shall refer to Eq. (1) as the "FP formula."

It is the FP formula which has been cast into doubt by Gribov's discovery[3] that when the gauge group G is nonabelian, there can be distinct gauge-related solutions in A ("Gribov copies") of a gauge-fixing condition such as $F(A)=0$. More pictorially, this is described by saying a gauge orbit $\{{}^{g}A\}$ can intersect a gauge-fixing surface $\{A:F(A)=0\}$ more than once. The conventional derivations of the FP formula[1,8] assume that this is not the case.

Our aim is to show that the FP formula is nevertheless "correct" in the sense that it is equivalent to the naive functional integral expression for $\langle f\rangle$. That is, we show

 0022-2488/84/113324-07$02.50

$$\frac{\int f(A)e^{-S(A)}\mathscr{D}A}{\int e^{-S(A)}\mathscr{D}A} = \frac{\int f(A)\det M(A)E\circ F(A)e^{-S(A)}\mathscr{D}A}{\int \det M(A)E\circ F(A)e^{-S(A)}\mathscr{D}A}. \tag{2}$$

Let G denote the global gauge group. Equation (2) is a consequence of the following.

Theorem 1: There exists a constant η such that for any gauge invariant function f,

$$\eta\left(\int E(C)\mathscr{D}C\right)\left(\int f(A)e^{-S(A)}\mathscr{D}A\right) = \left(\int \mathscr{D}g\right)\left(\int f(A)\det M(A)E\circ F(A)e^{-S(A)}\mathscr{D}A\right), \tag{3}$$

where $\mathscr{D}g$ denotes the infinite product of Haar measures on G. □

We obtain Eq. (2) by normalizing Eq. (3),

$$\frac{\eta(\int E(C)\mathscr{D}C)(\int f(A)e^{-S(A)}\mathscr{D}A)}{\eta(\int E(C)\mathscr{D}C)(\int e^{-S(A)}\mathscr{D}A)} = \frac{(\int \mathscr{D}g)(\int f(A)\det M(A)E\circ F(A)e^{-S(A)}\mathscr{D}A)}{(\int \mathscr{D}g)(\int \det M(A)E\circ F(A)e^{-S(A)}\mathscr{D}A)},$$

and cancelling identical factors. As in the original FP argument, the volume of the gauge group cancels in the normalization. What is different in our derivation is the factor η which appears in some of the intermediate steps.

The constant η is given by

$$\eta = \int \det M({}^gA)\delta(F({}^gA) - C)\mathscr{D}g, \tag{4}$$

where the integral extends over all gauge transformations g. It would appear from Eq. (4) that η depends on A and C, but one of the main results of this paper is to show that this is not the case. Assuming for the moment that η is indeed a constant, we can obtain Eq. (3) by integrating both sides of Eq. (4) against $E(C)f(A)\exp(-S(A))$, which gives

$$\eta\left(\int E(C)\mathscr{D}C\right)\left(\int f(A)e^{-S(A)}\mathscr{D}A\right) = \int\int f(A)\det M({}^gA)E\circ F({}^gA)e^{-S(A)}\mathscr{D}A\,\mathscr{D}g. \tag{5}$$

Now change the integration variable on the right-hand side of Eq. (5) form A to ${}^gA$. Because f, S, and the measure $\mathscr{D}A$ are invariant under this transformation, the result is Eq. (3).

As we have just seen, the crux of our argument is the assertion that η is independent of A and C. We now discuss the reasons why this assertion is true. For the moment, we shall use the notation $\eta(A,C)$ instead of η. Suppose that for each copy ${}^{g_k}A$ for which $F({}^{g_k}A) = C$, we can find a small neighborhood U_k of g_k which contains no other g_j. Then by Eq. (4),

$$\eta(A,C) = \sum_k \int_{U_k} \det M({}^gA)\delta(F({}^gA) - C)\mathscr{D}g.$$

Within each neighborhood U_k we need only integrate over those g which are infinitesimally close to g_k. Such g can be written as $g = (\mathbf{1} + \gamma + O(\gamma^2))g_k$ and the measure $\mathscr{D}g$ as $\mathscr{D}\gamma = \Pi_{x,i}d\gamma^i(x)$, where $\gamma(x) = \Sigma\gamma^i(x)\,t_i$ is an element of the Lie algebra of the gauge group written in terms of the generators $\{t_i\}$. Also,

$$F({}^gA) - C = F({}^{(\mathbf{1}+\gamma+O(\gamma^2))g_k}A) - F({}^{g_k}A) = M({}^{g_k}A)\gamma + O(\gamma^2),$$

by the definition[1] of $M({}^{g_k}A)$. Consequently,

$$\int_{U_k} \det M({}^gA)\delta(F({}^gA) - C)\mathscr{D}g = \det M({}^{g_k}A)\int_{U_k'} \delta(M({}^{g_k}A)\gamma)\mathscr{D}\gamma = \frac{\det M({}^{g_k}A)}{|\det M({}^{g_k}A)|}.$$

Thus

$$\eta(A,C) = \sum_k \operatorname{sgn}\det M({}^{g_k}A). \tag{6}$$

As Hirschfeld[7] has pointed out, in discussions involving the FP formula, one cannot afford to be careless about absolute value signs. For example, in deriving Eq. (6) it is essential not to neglect the fact that in a change of variables the volume element changes by the *absolute value* of the Jacobian determinant [in this case, $|\det M({}^{g_k}A)|$].

Hirschfeld uses Eq. (6) to argue that $\eta(A,0)$ is independent of A by identifying it with the oriented intersection number of the gauge orbit $\{{}^gA\}$ with the gauge-fixing surface $Z = \{A:F(A) = 0\}$. To have this identification, however, one must make assumptions about the gauge-fixing surface [i.e., assume that $C = 0$ is a regular value of $F(A)$] and about the orbit–surface intersections (i.e., that they are transversal).

We offer a different interpretation of η based on Eq. (6), namely that it is the oriented degree[9] of the map $g\to F({}^gA)$. This requires neither of the assumptions mentioned above and allows us to show easily that $\eta(A,C)$ is independent of both A and C.

There is a simple geometric interpretation of Eq. (6) which shows intuitively why η has this property. Consider the simple case of a lattice model with only one point and gauge group $G = \mathrm{U}(1)$. The function $g\to F({}^gA)$ is then a map from U(1) to U(1) [see Eq. (9) below] and its graph can be drawn in the unit square with opposite edges identified (Fig. 1). In this case, $\operatorname{sgn}\det M({}^{g_k}A)$ is $+1$ (resp., -1) when the slope of the function $g\to F({}^gA)$ is positive (resp., negative) at g_k.

Figure 1 shows that $\eta(A,C)$ is independent of C because as the horizontal line determined by C moves up or down, the corresponding points of intersection are created and destroyed in positive–negative pairs. Thus the value of the sum in Eq. (6) does not change.

Moreover, given two gauge fields A and B, change A into B in a continuous way so that the graph of $F({}^gA)$ is continuously deformed into that of $F({}^gB)$. Using the same reasoning as above, we see that $\eta(A,C) = \eta(B,C)$, i.e., $\eta(A,C)$ is independent of A.

The notion of degree allows us to carry out the preceding argument in general. We do so in Sec. III where we give a

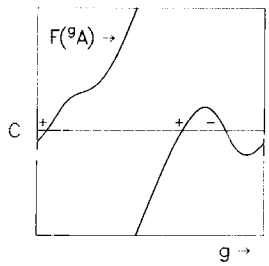


FIG. 1. Sign of $\det M({}^{g_k}A)$ when $F({}^{g_k}A) = C$.

rigorous derivation of a lattice version of the FP formula. Besides providing a regularization so that the relevant integrals are well defined, the lattice model fits the framework of degree theory more closely than does the continuum theory, the main difference being that the space of gauge fields is compact on the lattice.

III. THE FP FORMULA IN LATTICE GAUGE THEORIES

We now give the version of the argument presented in Sec. II for a lattice theory in $s+1$ space-time dimensions.

The gauge group G is taken to be a compact, connected Lie group which for simplicity we assume is a group of unitary matrices. Its Lie algebra is denoted by L. Let Λ be a finite lattice of points of the form $x = (n_0\epsilon,...,n_s\epsilon)$ with $n_\mu \in \{-N_\mu, -N_\mu+1,...,N_\mu\}$. The gauge fields are functions $a(x,y)$ from the bonds $\langle x,y\rangle$ in Λ into G. They satisfy

$$a(x,y) = a(y,x)^{-1}. \tag{7}$$

The gauge transformations are functions $g(x)$ from Λ to G, and they act by

$$({}^g a)(x,y) = g(x)a(x,y)g(y)^{-1}. \tag{8}$$

We usually think of a gauge transformation g as a point in the product group $\mathcal{G} \equiv \Pi_x G$. Similarly, a gauge field a is regarded as a point in $\mathcal{G}^* \equiv \Pi_{\langle x,y\rangle} G$. We denote by $\mathcal{L}$ and $\mathcal{L}^*$ the Lie algebras of $\mathcal{G}$ and $\mathcal{G}^*$, respectively.

The expectation of a function $f(a)$ of the gauge fields is given by

$$\langle f\rangle = \frac{\int f(a)e^{-S(a)}\,da}{\int e^{-S(a)}\,da},$$

where da denotes the product over all bonds of the normalized Haar measures $da(x,y)$; $S(a)$ is the lattice action. The action is a smooth, gauge-invariant function of a.

The lattice gauge-fixing function F will have the form

$$F(a)(x) = \prod_y a(x,y)^{m(x,y)}, \tag{9}$$

where $m(x,y)$ is an integer. [When the gauge group is nonabelian the order in which the product is taken in Eq. (9) must be specified.] For example, to obtain the lattice verison of the Landau gauge with

$$E\circ F(A) = \exp\left(\alpha \int \mathrm{Tr}(\partial_\mu A_\mu)^2\, d^4x\right),$$

one can choose

$$F(a)(x) = a(x,x+e_0)a(x,x-e_0)\cdots a(x,x+e_s)a(x,x-e_s), \tag{10a}$$

$$E(c) = \exp\left(2\alpha\lambda^{-2}\sum_x \mathrm{Tr}(c(x)-I)\right), \tag{10b}$$

For the axial gauge, take

$$F(a)(x) = a(x,x+e_0), \tag{11}$$

which is defined for those x for which $-\epsilon N_0 \leqslant x_0 < \epsilon N_0$.

Given a gauge-fixing function F and a gauge field a, define the map φ: $\mathcal{G}\to\mathcal{G}$ by

$$\varphi(g) = F({}^g a)F(a)^{-1}.$$

The *lattice Faddeev–Popov operator* is the map $M(a)$: $\mathcal{L}\to\mathcal{L}$ defined by $M(a) = d\varphi_1$. This operator plays the role of the Jacobian at $g = \mathbb{1}$ for the change of variables from g to $F({}^g a)$ in a Haar integral. The *lattice Faddeev–Popov determinant* is $\det M(a)$.

Examples of the FP determinant are discussed in Secs. IV and V.

Theorem 2: (Lattice FP formula) There is a constant η depending only on F such that for any smooth gauge-invariant function f on $\mathcal{G}^*$ and any smooth function E on $\mathcal{G}$,

$$\eta\left(\int E(c)dc\right)\left(\int f(a)e^{-S(a)}\,da\right) = \int \det M(a) f(a) E\circ F(a) e^{-S(a)}\,da. \tag{12}$$

Hence if $\eta(\int E(c)dc) \neq 0$,

$$\langle f\rangle = \frac{\int f(a)e^{-S(a)}\,da}{\int e^{-S(a)}\,da} = \frac{\int \det M(a) f(a) E\circ F(a) e^{-S(a)}\,da}{\int \det M(a) E\circ F(a) e^{-S(a)}\,da}. \tag{13}$$

The proof of this theorem is based on the notion of the degree of a map. Recall that if M and N are compact, connected and oriented manifolds of dimension n and $f: M\to N$ is a smooth map, then the *degree*[9] of f is the number for which the equation

$$\int_M f^*\omega = \deg(f)\int_N \omega \tag{14}$$

holds for any n-form on N.

Lemma 1: Let $\eta(a)$ be the degree of the map $f: \mathcal{G}\to\mathcal{G}$ defined by $f(g) = F({}^g a)$. Then for all smooth functions E on $\mathcal{G}$,

$$\int_{\mathcal{G}} \det M({}^g a) E\circ F({}^g a)\,dg = \eta(a)\int_{\mathcal{G}} E(h)\,dh. \tag{15}$$

Proof: This is essentially the same as Proposition XIV of Ref. 9. □

The connection of Eq. (4) with Eq. (15) can be made by taking $E(h) = \delta(hc^{-1})$ in the latter equation.

Hirschfeld[7] considered the oriented intersection number of the orbit and surface manifolds. This quantity is closely related to $\eta(a)$. As he pointed out, its utility depends on the following lemma.

Lemma 2: The quantity $\eta(a)$ defined in the preceding Lemma is independent of a.

Proof: This is a consequence of the fact that degree is a homotopy invariant. By assumption, G is connected and hence so is $\mathcal{G}^*$. Thus if a_i $(i = 0,1)$ are any two gauge fields

there is a path in $\mathcal{G}^*$ from a_0 to a_1. This path yields a homotopy of the two maps

$$f_i(g) = F({}^g a_i) \quad (i = 0,1) .$$

Hence $\deg(f_0) = \deg(f_1)$. □

The proof of the lattice FP formula is now simply an adaptation of the original continuum argument.[1,8]

Proof of Theorem 2: The upshot of the lemmas is that for some constant η depending only on F,

$$\eta \int_{\mathcal{G}} E(h)\,dh = \int_{\mathcal{G}} \det M({}^g a) E \circ F({}^g a)\,dg . \tag{16}$$

Multiply both sides of Eq. (16) by

$$\int_{\mathcal{G}^*} f(a) e^{-S(a)}\,da ,$$

and apply Fubini's theorem to obtain

$$\begin{aligned}
&\eta \left(\int_{\mathcal{G}} E(h)\,dh\right)\left(\int_{\mathcal{G}^*} f(a) e^{-S(a)}\,da\right) \\
&\quad = \left(\int_{\mathcal{G}} \det M({}^g a) E \circ F({}^g a)\,dg\right)\left(\int_{\mathcal{G}^*} f(a) e^{-S(a)} da\right) \\
&\quad = \int_{\mathcal{G}} \int_{\mathcal{G}^*} \det M({}^g a) f(a) E \circ F({}^g a) e^{-S(a)}\,da\,dg . \qquad (17)
\end{aligned}$$

Now make a change of variables on the right-hand side of Eq. (17) a to ${}^g a$. Because the Haar measure is translation invariant and f and S are gauge invariant, we obtain

$$\begin{aligned}
&\int_{\mathcal{G}^*} \det M({}^g a) f(a) E \circ F({}^g a) e^{-S(a)}\,da \\
&\quad = \int_{\mathcal{G}^*} \det M({}^g a) f({}^g a) E \circ F({}^g a) e^{-S({}^g a)}\,da \\
&\quad = \int_{\mathcal{G}^*} \det M(a) f(a) E \circ F(a) e^{-S(a)}\,da ,
\end{aligned}$$

and Eq. (17) becomes

$$\begin{aligned}
&\eta \left(\int_{\mathcal{G}} E(h)\,dh\right)\left(\int_{\mathcal{G}^*} f(a) e^{-S(a)}\,da\right) \\
&\quad = \int_{\mathcal{G}} \int_{\mathcal{G}^*} \det M(a) f(a) E \circ F(a) e^{-S(a)}\,da\,dg \\
&\quad = \int_{\mathcal{G}^*} \det M(a) f(a) E \circ F(a) e^{-S(a)}\,da ,
\end{aligned}$$

where we have used the fact that $\int_{\mathcal{G}} dg = 1$. □

Definition: The *gauge degree* η associated with the gauge fixing function F is the degree of the map $f: \mathcal{G} \to \mathcal{G}$, $f(g) = F({}^g a)$ for any $a \in \mathcal{G}^*$.

IV. EXAMPLES AND APPLICATIONS

(*a*) *Formula for the gauge degree*: Although η does not appear in the expression (13) for normalized lattice expectations, it is worth knowing because it contains information about the orbit–surface intersections. We shall use this information to analyze gauge-fixing on the lattice for the axial and Landau gauges. Also, we wish to know when $\eta = 0$ since in that case the gauge-fixing procedure described in Sec. III breaks down.

For gauge-fixing functions F of the form given in Eq. (9) we have

$$F({}^g \mathbf{1})(x) = \prod_y g(y)^{n(x,y)} , \tag{18}$$

for some integers $n(x,y)$. [We are ignoring the noncommutative nature of group multiplication in Eq. (16); the value of η is the same regardless of the order in which the product is taken.] We show in the Appendix that

$$\eta = (\det N)^r , \tag{19}$$

where $N_{x,y} = n(x,y)$ and r is the rank of G. [The rank of a Lie group is the dimension (as a manifold) of a maximal abelian subgroup.]

(*b*) *Axial gauge*: Axial gauge [Eq. (11)] is the easiest to handle with these methods. We show first that the FP determinant is identically 1.

One way to calculate $M(a)$ is to use the fact that if $g = \mathbf{1} + \gamma + O(\gamma^2)$,

$$F({}^g a)\, F(a)^{-1} = \mathbb{1} + M(a)\gamma + O(\gamma^2) . \tag{20}$$

[We are suppressing indices to simplify the formulas somewhat. For example, the matrix $M(a)$ acts on $\mathcal{L}$ and so is indexed both by the lattice points x and color indices i. Thus, $M(a)\gamma$ stands for $[M(a)\gamma]^i_x = \Sigma_{j,y} M(a)^{ij}_{xy} \gamma^j(y)$.] For axial gauge, we have

$$\begin{aligned}
&F({}^g a)(x)\, F(a)(x)^{-1} \\
&\quad = g(x) a(x,x+e_0) g(x+e_0)^{-1} a(x,x+e_0)^{-1} \\
&\quad = (I+\gamma)(x)) a(x,x+e_0)(I - \gamma(x+e_0)) \\
&\qquad \times a(x,x+e_0)^{-1} + O(\gamma^2) \\
&\quad = I + \gamma(x) - \mathrm{ad}(a(x,x+e_0))\gamma(x+e_0) + O(\gamma^2) ,
\end{aligned}$$

where $\mathrm{ad}(a)\gamma = a\gamma a^{-1}$. Thus

$$\begin{aligned}
&M(a)_{xx} = I , \\
&M(a)_{x,x+e_0} = -\mathrm{ad}(a(x,x+e_0)), \\
&M(a)_{x,y} = 0 , \quad \text{if } y \neq x_0 \quad \text{or} \quad x+e_0 .
\end{aligned}$$

(I is the identity operator on L.) Since M is an upper triangular matrix whose diagonal elements are equal to 1,

$$\det M(a) = 1 \quad \text{(axial gauge)} ,$$

for every gauge field.

Now we show that the gauge degree for axial gauge is also equal to 1. In this case,

$$F(a)(x) = a(x,x+e_0) ,$$

so that

$$F({}^g \mathbf{1})(x) = g(x) g(x+e_0)^{-1} .$$

The matrix elements of N are

$$\begin{aligned}
&N_{x,x} = 1 , \\
&N_{x,x+e_0} = -1 , \\
&N_{x,y} = 0 , \quad \text{if } y \neq x \quad \text{or} \quad x+e_0 .
\end{aligned}$$

By an argument similar to that used for the FP determinant, $\det N = 1$ and so $\eta = 1$. Note that if we take E to be a δ-function, we obtain the formula of Ref. 10

$$\int f(a) e^{-S(a)}\,da = \int \prod_x \delta(a(x,x+e_0)) f(a) e^{-S(a)}\,da .$$

(c) *Landau gauge*: The lattice FP determinant for the Landau gauge [Eq. (10)] is discussed in Sec. V. For the moment, we calculate just the gauge degree.

When F is given by Eq. (10a), the matrix N is given by

$$(Nf)(x) = \sum_{\mu=0}^{s} [\, -f(x+e_\mu) + 2f(x) - f(x-e_\mu)] \,. \tag{21}$$

We have not yet said how to define $F(a)(x)$ when x is on the boundary of Λ. For such, x, the definition (10a) does not make sense since some of the bonds referred to are not in Λ. It takes some care on this point to arrange that $\eta \neq 0$.

We shall proceed as follows. Define $F(a)(x)$ as in (10a) but only for x in the interior of Λ. In integrals such as the one defining η [Eq. (12)], integrate over $\Pi_x dg(x)$ only for x in the interior of Λ. This is equivalent to requiring that all gauge transformations be the identity on the boundary of Λ. The result is that N is given by Eq. (21) with the convention that $f(y) = 0$ when y is on the boundary of Λ. In other words

$$N = \epsilon^2 \Delta^{D}_{(\epsilon)} \,,$$

where $\Delta^{D}_{(\epsilon)}$ is the finite-difference Laplacian operator on Λ with Dirichlet boundary conditions. The eigenvalues of $\Delta^{D}_{(\epsilon)}$ are (e.g., see Sec. 9.5 of Ref. 11)

$$\lambda_k = \sum_{\mu=0}^{s} 4\epsilon^{-2} \sin^2\left(\frac{k_\mu}{4N_\mu}\right), \tag{22}$$

with $k_\mu = 1,2,\ldots,2N_\mu - 1$. In particular, no eigenvalue is 0 so $\det N \neq 0$ and $\eta \neq 0$.

We remark that if the boundary values of $F(a)$ had been defined by imposing periodic boundary conditions or by simply omitting any terms $a(x,y)$ for which the bond $\langle x,y\rangle$ is not in Λ, $\Delta^{D}_{(\epsilon)}$ would have been replaced by the Laplacian with periodic or Neumann boundary conditions respectively. In both cases, we would have $\eta = 0$.

(d) *Existence and uniqueness of orbit-surface intersections*: We now use our knowledge of η to answer for the lattice theory two questions raised by Gribov[3] for the continuum theory. This analysis is based on an alternative formula for the degree of a map[9]: if c is any regular value of f, then

$$\deg(f) = \sum_{g \in f^{-1}(c)} \epsilon(f,g) \,, \tag{23}$$

where $\epsilon(f,g)$ is $+1$ if f preserves orientation at g and -1 if f reverses orientation at g.

The first question is, given a gauge field a and an arbitrary function c, does the gauge orbit $\{{}^g a\}$ intersect the gauge-fixing surface $F = c$? In other words, does there exist a gauge transformation g so that

$$F({}^g a) = c?$$

It follows from the definition of η that the answer is yes if $\eta \neq 0$, for by Eq. (23) any map which is not surjective has degree 0.

The second question is, are there any Gribov copies? That is, if

$$F(a) = c \,,$$

are there any nontrivial gauge transformations g for which

$$F({}^g a) = c?$$

Again the answer is yes if $|\eta| \neq 1$ as can be seen from the definition of η and Eq. (23). If $|\eta| \neq 1$, there must be more than one term in the sum (23). Moreover, because η is the same for all a and c, if $|\eta| \neq 1$ or 0 then every orbit intersects every surface more than once. (If $\eta = 0$ there may be some orbit-surface combinations which do not intersect.)

For Landau gauge as discussed in part (c) we have from Eq. (22)

$$\det N = \prod_k \epsilon^2 \lambda_k \,,$$

from which it can be shown that $\det N \to \infty$ and hence $\eta \to \infty$ as $N_\mu \to \infty$. This shows that for the Landau gauge at least, Gribov copies occur in the lattice theory.

(e) *Gauge invariance of the FP determinant*: In the original formulations of the FP formula it appeared that the FP determinant was gauge invariant. The discovery of Gribov copies has made this doubtful. We now show that in the lattice theory, gauge invariance holds only in very special cases.

Of course, one such case is that where G is abelian, for then $M(a)$ is independent of a [see Eq. (20)] and the FP determinant is a constant. If G is nonabelian, take $E = 1$ in Eq. (15) to obtain

$$\eta = \int_{\mathcal{G}} \det M({}^g a)\, dg \,.$$

If $\det M(a)$ were gauge invariant, then we would have

$$\eta = \det M(a) \,,$$

for every field a, and the FP determinant would be a constant. Moreover, suppose k is the dimension of G. By substituting Eq. (18) into Eq. (20) we find that

$$M(\mathbb{1})_{xy} = N_{xy} I \,,$$

but $\det M(\mathbf{1}) = \eta = (\det N)^r$ and $r = k$ if and only if G is abelian. Hence it must be that $\det N = 0$ or $|\det N| = 1$. To summarize, if G is nonabelian, then the FP determinant is gauge invariant if and only if it is a constant and that constant is 0, $+1$, or -1. In particular, the Landau-gauge FP determinant is not gauge invariant.

V. CONTINUUM LIMIT

In this section, we illustrate how the continuum FP formula can be obtained formally by taking the limits $N_\mu \to \infty$ and $\epsilon \to 0$ in the finite-volume lattice expression (13). The limit is taken assuming that we obtain the lattice gauge field from a smooth continuum gauge field A_μ by the relation

$$a(x, x \pm e_\mu) = e^{\pm \lambda \epsilon A_\mu [x + (1/2)e_\mu]} \,, \tag{24}$$

where λ is the coupling constant.

For concreteness, we treat the case of Landau gauge [Eq. (10)]. The continuum expression in this case is

$$\langle f \rangle = \frac{\int f(A) \det M(A) e^{-S_\alpha(A)}\, \mathcal{D}A}{\int \det M(A)\, e^{-S_\alpha(A)} \mathcal{D}A} \,, \tag{25}$$

in which

$$S_\alpha(A) = -\frac{1}{2} \int \mathrm{tr}(F_{\mu\nu})^2\, d^4x + \alpha \int \mathrm{tr}(\partial_\mu A_\mu)^2\, d^4x \,, \tag{26}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + \lambda\,[A_\mu, A_\nu]$ and $M(A)$ acts on Lie algebra-valued functions B by

$$M(A)(B) = -\Delta B - \lambda\,\partial_\mu\,[A_\mu, B]\,. \tag{27}$$

To begin, take Eq. (13) for the finite lattice and let $N_\mu \to \infty$, so that all expressions now refer to an infinite lattice with spacing ϵ. We assume that the action $S(a)$ has been chosen so that, as $\epsilon \to 0$,

$$S(a) \to -\frac{1}{2}\int \mathrm{tr}\,(F_{\mu\nu})^2\, d^4x\,. \tag{28}$$

A simple argument based on Eq. (24) shows that

$$E \circ F(A) \to \exp\left(\alpha \int \mathrm{tr}\,(\partial_\mu A_\mu)^2\, d^4x\right) \tag{29}$$

as $\epsilon \to 0$ so that

$$E \circ F(A)\, e^{-S(A)} \to e^{-S_\alpha(A)}\,.$$

The remainder of this section is devoted to showing that the matrix $\epsilon^{-2}M(a)$ converges to the operator $M(A)$ defined by Eq. (27). In the following, the summation convention for repeated indices does not apply.

Given a point x in the lattice, define

$$\mathrm{ad}(\pm\mu) \equiv \mathrm{ad}(a(x,x+e_0)a(x,x-e_0) \times a(x,x+e_1)\cdots a(x,x\pm e_\mu))\,.$$

For F given by Eq. (10a) and $g = 1 + \gamma + O(\gamma^2)$ we have

$$\begin{aligned} F({}^g a)(x) F(a)(x)^{-1} &= (I+\gamma(x))\,a(x,x+e_0)(I-\gamma(x+e_0)) \\ &\quad \times (I+\gamma(x))a(x,x-e_0)(I-\gamma(x-e_0))\times\cdots \\ &\quad \times (I+\gamma(x))a(x,x-e_s)(I-\gamma(x-e_s)) \\ &\quad \times [a(x,x+e_0)\cdots a(x,x-e_s)]^{-1} + O(\gamma^2) \\ &= I + \{I - \mathrm{ad}(-s) + \sum_\mu [\mathrm{ad}(+\mu) + \mathrm{ad}(-\mu)]\}\gamma(x) \\ &\quad - \sum_\mu \mathrm{ad}(+\mu)\gamma(x+e_\mu) - \sum_\mu \mathrm{ad}(-\mu)\gamma(x-e_\mu) \\ &\quad + O(\gamma^2)\,. \end{aligned}$$

The matrix element $M(a)_{xy}$ is the coefficient of $\gamma(y)$ in the above expression. Thus if B is a Lie algebra-valued function,

$$\begin{aligned} \epsilon^{-2}(M(a)B)(x) &= \epsilon^{-2}\sum_y M(a)_{xy}\,B(y) \\ &= \epsilon^{-2}[I - \mathrm{ad}(-s)]\,B(x) \\ &\quad + \epsilon^{-2}\sum_\mu \mathrm{ad}(+\mu)[B(x) - B(x+e_\mu)] \\ &\quad + \epsilon^{-2}\sum_\mu \mathrm{ad}(-\mu)[B(x) - B(x-e_\mu)] \\ &= \epsilon^{-2}[I - \mathrm{ad}(-s)]\,B(x) \\ &\quad - \epsilon^{-1}\sum_\mu \mathrm{ad}(+\mu)\delta_\mu\, B\left(x+\frac{1}{2}e_\mu\right) \\ &\quad + \epsilon^{-1}\sum_\mu \mathrm{ad}(-\mu)\delta_\mu\, B\left(x-\frac{1}{2}e_\mu\right), \end{aligned} \tag{30}$$

where $\delta_\mu f(x) \equiv \epsilon^{-1}[f(x+\frac{1}{2}e_\mu) - f(x-\frac{1}{2}e_\mu)]$.

We now wish to expand the right-hand side of Eq. (30) to $O(\epsilon)$, using Eq. (24) to write

$$\begin{aligned} a(x,x\pm e_\mu) &= e^{\pm\lambda\epsilon A_\mu[x\pm(1/2)e_\mu]} \\ &= I \pm \lambda\epsilon A_\mu(x\pm\tfrac{1}{2}e_\mu) \\ &\quad + \tfrac{1}{2}[\lambda\epsilon A_\mu(x\pm\tfrac{1}{2}e_\mu)]^2 + O(\epsilon^3)\,, \end{aligned}$$

so that, for example,

$$a(x,x+e_\mu)a(x,x-e_\mu) = I + \lambda\epsilon^2\delta_\mu A_\mu(x) + O(\epsilon^3)\,.$$

The result for the terms appearing in Eq. (30) is

(i) $\epsilon^{-2}[I - \mathrm{ad}(-s)]\,B(x)$

$$= -\lambda\sum_\mu [\delta_\mu A_\mu(x), B(x)] + O(\epsilon)\,,$$

(ii) $-\epsilon^{-1}\,\mathrm{ad}(+\mu)\delta_\mu\, B(x+\tfrac{1}{2}e_\mu)$

$$\begin{aligned} &= -\epsilon^{-1}\delta_\mu\, B(x+\tfrac{1}{2}e_\mu) \\ &\quad - \lambda\,[A_\mu(x+\tfrac{1}{2}e_\mu), \delta_\mu B(x+\tfrac{1}{2})e_\mu)] \\ &\quad + O(\epsilon)\,, \end{aligned}$$

(iii) $\epsilon^{-1}\,\mathrm{ad}(-\mu)\delta_\mu\, B(x-\tfrac{1}{2}e_\mu)$

$$= \epsilon^{-1}\delta_\mu\, B(x-\tfrac{1}{2}e_\mu) + O(\epsilon)\,.$$

Putting these equations into Eq. (30), we get

$$\begin{aligned} \epsilon^{-2}(M(a))\,B(x) &= \sum_\mu \{-\lambda\,[\delta_\mu A_\mu(x), B(x)] - \delta_\mu\delta_\mu\, B(x) \\ &\quad - \lambda\,[A_\mu(x+\tfrac{1}{2}e_\mu), \delta_\mu B(x+\tfrac{1}{2}e_\mu)]\} + O(\epsilon) \\ &\to -\Delta B(x) - \lambda\sum_\mu \delta_\mu\,[A_\mu(x), B(x)]\,, \end{aligned}$$

as $\epsilon \to 0$.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Lon Rosen, for generously giving of his time on several occasions to discuss the material in this paper. I also wish to thank Lex Renner for making a suggestion which led to the results in the Appendix.

This work was supported in part by a Natural Sciences and Engineering Research Council of Canada 1967 Science Scholarship.

APPENDIX: FORMULA FOR THE GAUGE DEGREE

We indicate here how one obtains the expression (17) for the gauge degree for gauge-fixing functions F of the form given in Eq. (9). The gauge degree is the degree of the function f defined by

$$f(g) = F({}^g\mathbb{1})\,. \tag{A1}$$

Thus the problem to be solved is that of finding the degree of a function f on the product group $G^k = \Pi_1^k\, G$ of the form [cf. Eq. (16)]

$$f(g_1,\ldots,g_k) = (f_1(g_1,\ldots,g_k),\ldots,f_k(g_1,\ldots,g_k))\,, \tag{A2}$$

where

$$f_i(g_1,\ldots,g_k) = \prod_{j=1}^k g_j^{n_{ij}}\,. \tag{A3}$$

[The order in which the multiplications in Eq. (A3) occur does not affect deg f.]

Theorem 3: With f as described above,

$$\deg f = (\det N)^r ,$$

where r is the rank of G and $N_{ij} = n_{ij}$.

The proof of this theorem requires some knowledge of the cohomology of compact Lie groups. The relevant points are reviewed very briefly below following the notation of Ref. 9 to which the reader is referred for more information.

Let Π_j denote the projection map

$$\Pi_j(g_1,\ldots,g_k) = g_j ,$$

and M the multiplication map

$$M(g_1,\ldots,g_k) = g_1 g_2 \cdots g_k .$$

An element ω of the cohomology algebra $H(G)$ is *primitive* if

$$M^{\#}\omega = \sum_{j=1}^{k} \Pi_j^{\#}\omega . \tag{A4}$$

It follows from Eq. (A4) that if P_m is the m-power map $P_m(g) = g^m$ and ω is primitive, then

$$P_m^{\#} = m\omega .$$

Any primitive has odd degree [as an element of the graded algebra $H(G)$] so if ω_1 and ω_2 are primitive

$$\omega_1\omega_2 = -\omega_2\omega_1 . \tag{A5}$$

Moreover, there are primitives $\omega_1,\ldots,\omega_r$ so that

$$\omega_1\omega_2\cdots\omega_r \tag{A6}$$

is a nonzero element of the top cohomology group of G.

The degree of f is given by the equation

$$f^{\#}\omega = \deg(f)\cdot\omega ,$$

where ω is any member of the top cohomology group of G^k. To prove the theorem we shall construct a particular nonzero $\bar{\omega}$ and show that

$$f^{\#}\bar{\omega} = (\det N)^r\bar{\omega} .$$

Suppose ω is primitive and consider

$$(\Pi_i \circ f)^{\#}\omega .$$

Now

$$\Pi_i \circ f = M \circ Q_i ,$$

where

$$Q_i(g_1,\ldots,g_k) = (g_1^{n_{i1}},\ldots,g_k^{n_{ik}}) .$$

Consequently,

$$(\Pi_i \circ f)^{\#}\omega = Q_i^{\#} \circ M^{\#}\omega = \sum_{j=1}^{k} (\Pi_j Q_i)^{\#}\omega$$

$$= \sum_{j=1}^{k} (P_{n_{ij}} \circ \Pi_j)^{\#}\omega = \sum_{j=1}^{k} n_{ij}\,\Pi_j^{\#}\omega . \tag{A7}$$

Equations (A5) and (A7) imply that

$$f^{\#}\omega(k) = \det(N)\omega^{(k)} , \tag{A8}$$

where $\omega^{(k)} = (\Pi_1^{\#}\omega)(\Pi_2^{\#}\omega)\cdots(\Pi_k^{\#}\omega)$.

Define

$$\bar{\omega} = \omega_1^{(k)}\cdots\omega_r^{(k)} ,$$

where the ω_j are those of the expression (A6). Then by Eq. (A8),

$$f^{\#}\bar{\omega} = (f^{\#}\omega_1^{(k)})\cdots(f^{\#}\omega_r^{(k)}) = (\det N)^r\bar{\omega} ,$$

which proves the theorem.

[1] L. Faddeev and V. Popov, Phys. Lett. B **25**, 29 (1967).
[2] G. 't Hooft, Nucl. Phys. B **157**, 173 (1978).
[3] V. N. Gribov, Nucl. Phys. B **139**, 1 (1978).
[4] I. M. Singer, Commun. Math. Phys. **60**, 7 (1978); A. Chodos and V. Moncrief, J. Math. Phys. **21**, 364 (1980); O. Babelon and C. M. Viallet, Commun. Math. Phys. **81**, 515 (1981).
[5] C. M. Bender, T. Eguchi, and H. Pagels, Phys. Rev. D **17**, 1086 (1978); J. P. Greensite, Phys. Rev. D **18**, 3842 (1978).
[6] G. Parisi and Y. S. Wu, Sci. Sin. **24**, 483 (1981); I. O. Stamatescu, U. Wolff, and D. Swanziger, Nucl. Phys. B **225**, 377 (1983); D. Zwanziger, Nucl. Phys. B **192**, 259 (1981); D. Zwanziger, Nucl. Phys. B **209**, 336 (1982).
[7] P. Hirschfeld, Nucl. Phys. B **157**, 37 (1978).
[8] E. S. Abers and B. W. Lee, Phys. Rep. C **9**, 1 (1973); C. Itzykson and J-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980); P. Ramond, *Field Theory: A Modern Primer* (Benjamin, Reading, MA 1981).
[9] W. Greub, S. Halperin, and R. Vanstone, *Connections, Curvature and Cohomology* (Academic, New York, 1973), Vol. 2.
[10] K. Osterwalder and E. Seiler, Ann. Phys. **110**, 440 (1978).
[11] J. Glimm and A. Jaffe, *Quantum Physics* (Springer, New York, 1981).

Casimir invariants for the eight-dimensional subgroups of the Poincaré group $P(1,4)$

M. Leveille
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, C.P. 6128, Succ. A, Montréal, Québec, Canada H3C 3J7

(Received 27 July 1983; accepted for publication 6 April 1984)

The generalized Casimir operators of all eight-dimensional subgroups of the Poincaré group $P(1,4)$ are found. These operators include polynomials in the enveloping algebra of the considered eight-dimensional Lie algebra, quotients of such polynomials, and also more complicated functions of the infinitesimal operators.

PACS numbers: 11.30.Cp, 02.20 + b

The existence of invariants (Casimir) operators is one of the most important properties of Lie algebras mainly because of the role the operators play in representation theory. They are well known only for semisimple Lie algebras. Every invariant operator of a simple Lie algebra is a polynomial in certain basic ones. The basic invariant operators are homogenous of known degree[1] and their number is equal to the rank of the algebra. For nonsemisimple algebras the invariant operators have to be found by lengthy computations and not all of them can be written as polynomials in the elements of the Lie algebra. It is known that the number of the basic invariant operators is equal to the dimension of the algebra modulo 2.

An abundant source of physically interesting nonsemisimple Lie algebras is found in the Lie algebras of groups of inhomogeneous transformations, in particular the Lie algebra $P(3,1)$ of the Poincaré group which is a semidirect product of the Lorentz group with a four-parameter abelian group of space-time translations. A complete (infinite) list of conjugacy classes of subalgebras of $P(3,1)$ exists.[2] The corresponding invariant operators were found in Ref. 3 using the method of Ref. 4. In a number of applications it is useful to consider also the Lorentz and Poincaré groups in larger spaces. Thus Fedorchuk[5,6] found the subalgebras of the Poincaré Lie algebra $P(4,1)$ in (4,1)-dimensional space-time. In this article, besides a large number of entries found in $P(3,1)$ and those which are of the splitting kind, the most interesting ones are the nonsplitting subalgebras of larger dimensions. Reference 6 contains all eight-dimensional nonsplitting subalgebras of $P(4,1)$. The purpose of this paper is to find the invariant operators for the Lie algebra of Ref. 6 and those whom are eight-dimensional of Ref. 7. The algebras of Ref. 7 are of the splitting kind. Our method is esssentially that of Ref. 4 with some improvements from Ref. 7.

The eight-dimensional algebras of Ref. 6 and 7 are, respectively, those of Tables I and II. The generators G, L_i, P_i, C_i, X_ν $(i = 1, 2, 3;\ \nu = 1, 2, 3,4)$ which occur in these algebras are linear combinations of the 15 generators $M_{\mu\nu} = -M_{\nu\mu}$ and the P_μ $(\mu,\nu = 0,...4)$ satisfy the following commutation relations:

$$[P'_\mu,P'_\nu] = 0; \quad [M_{\nu\nu},P'_\sigma] = g_{\mu\sigma}P'_\nu - g_{\nu\mu}P'_\mu,$$
$$[M_{\mu\nu},M_{\rho\sigma}] = g_{\mu\rho}M_{\nu\sigma} + g_{\nu\sigma}M_{\mu\rho} - g_{\nu\rho}M_{\mu\sigma} - g_{\mu\sigma}M_{\nu\rho}, \tag{1}$$

where $g_{\mu\nu}$ is the metric tensor with components $g_{00} = -g_{\sigma\sigma} = 1$ and $g_{\mu\nu} = 0$ (if $\mu \neq \nu$) $(\sigma = 1,...4\,;\mu,\nu = 0,...4)$. The generators of the algebras are explicitly

$$G = M_{40}, \quad L_1 = M_{32}, \quad L_2 = -M_{31}, \quad L_3, = M_{21},$$
$$P_a = M_{4a} - M_{a0}, \quad L_a = M_{4a} + M_{a0},$$
$$X_0 = \tfrac{1}{2}(P'_0 - P'_4), \quad X_4 = \tfrac{1}{2}(P'_0 + P'_4), \quad X_k = X'_k,$$

where $a, k = 1,2,3$.

The commutation relations in terms of the generators of the algebras are

TABLE I. Subalgebras of the Poincaré algebra of dimension 8 of the nonsplitting kind.

Generators	Range of the parameters	Invariants
$G + aX_3, L_3 + bX_3; P_1, P_2, P_3, X_1, X_2, X_4$	$b \neq 0, \forall a$	none
$G + aX_0, L_3; P_1, P_2, P_3, X_1, X_2, X_4$	$a \neq 0$	$X_4 \exp\left(\frac{-P_3}{2aX_4}\right), \frac{P_1X_2 - P_2X_1 - 2X_4L_3}{X_4}$
$L_3 + X_0, P_3 + \tilde{h}_0X_0; P_1, P_2, X_1, X_2, X_3, X_4$	$\forall\tilde{h}_0$	$X_4, X_1(X_1 - 2P_2) + X_2(X_2 + 2P_1) - 4X_4(L_3 + X_0) + X_3^2$
$L_3 - X_0, P_3 + \tilde{h}_0X_0; P_1, P_2, X_1, X_2, X_3, X_4$	$\forall\tilde{h}_0$	$X_4, X_1(X_1 + 2P_2) + X_2(X_2 - 2P_1) + 4X_4(L_3 - X_0) + X_3^2$
$L_3, P_3 + \epsilon X_0, X_3; P_1, P_2, X_1, X_2, X_4$	$\epsilon = \pm 1$	$X_4, P_1X_2 - P_2X_1 - 2L_3X_4$
$G + aX_3, L_3 + dX_3; P_1, P_2, X_0, X_1, X_2, X_4$	$\forall a_3, d_3$	$X_1^2 + X_2^2 - 4X_4X_0, [P_1X_2 - P_2X_1 - 2X_4(L_3 + dX_3)]/X_4$
$\tilde{A} \equiv L_3 - (\epsilon/2)(P_3 + C_3) + \tilde{\mu}_0X_0;$	$\tilde{\mu}_0 \neq 0, \epsilon = \pm 1$	$X_1^2 + X_2^2 + X_3^2 + (X_4 - X_0)^2,$
$B_i \equiv L_i + (\epsilon/2)(P_i + C_i), X_1, X_2, X_3, X_4 - X_0$	$i = 1, 2, 3$	$(B_3 + \tilde{A} + (\tilde{\mu}_2/2)(X_4 - X_0))((X_4 - X_0)^2 + X_3^2 + X_2^2 + X_1^2)$ $+ 2\epsilon(X_4 - X_0)(X_1B_2 - X_2B_1) + 2X_2(X_3B_2 - X_2B_3)$ $+ 2X_1(X_3B_1 - X_1B_3)$

 0022-2488/84/113331-03$02.50

TABLE II. Subalgebras of the Poincaré algebra of dimension 8 of the splitting kind.

Generators	Range of the parameters	Invariants
$G; L_1, L_2, L_3, P_1, P_2, P_3, X_4$		$(\mathbf{P}\cdot\mathbf{P})/X_4$, $(\mathbf{P}\cdot\mathbf{L})/X_4$
$G, L_3; P_1, P_2, P_3, X_1, X_2, X_4$		P_3/X_4, $(X_1P_2 - X_2P_1 + 2X_4L_3)/X_4$
$L_3, P_3, X_3; X_1, X_2, X_4, P_1, P_2$		X_4, $X_1P_2 - X_2P_1 + 2X_4L_3$
$G; L_1, L_2, L_3, X_0, X_1, X_2, X_3$		$\mathbf{X}\cdot\mathbf{L}$, $\mathbf{X}\cdot\mathbf{X}$
$G; L_1, L_2, L_3, X_1, X_2, X_3, X_4$		$\mathbf{X}\cdot\mathbf{L}$, $\mathbf{X}\cdot\mathbf{X}$
$G, X_1, X_2, X_3; X_4, P_1, P_2, P_3$		none
$G, L_3 + bG; X_1, X_2, X_4, P_1, P_2, P_3$	$b>0$	none
$G, L_3; X_0, X_1, X_2, X_4, P_1, P_2$		$X_1^2 + X_2^2 - 4X_0X_4$, $(X_1P_2 - X_2P_1 + 2X_4L_3)/X_4$
$G, L_3, X_3; X_1, X_2, X_4, P_1, P_2$		X_3, $(X_1P_2 - X_2P_1 - 2X_4L_3)/X_4$
$P_1, P_2, P_3, X_0; X_1, X_2, X_3, X_4$		X_4, $X_1^2 + X_2^2 + X_3^2 - 4X_0X_4$
$G, L_3; X_0, X_1, X_2, X_3, X_4, P_3$		$X_1^2 + X_2^2$, $X_3^2 - 4X_0X_4$
$G, X_3; X_0, X_1, X_2, X_3, P_1, P_2$		X_3, $X_1^2 + X_2^2 - 4X_0X_4$
$X_0, X_3, L_3; X_1, X_2, X_4, P_1, P_2$		X_3, X_4, $X_1P_2 - X_2$, $P_1 + 2X_4L_3$, $X_1^2 + X_2^2 - 4X_0X_0$
$X_0, L_3 + \epsilon P_3; P_1, P_2, X_1, X_2, X_3, X_4$	$\epsilon = \pm 1$	X_4, $X_1^2 + X_2^2 + X_3^2 - 4X_0X_4$
$X_3, L_3 + cG; X_0, X_1, X_2, X_4, P_1, P_2$	$c>0$	X_3, $X_1^2 + X_2^2 - 4X_0X_4$
$X_0, X_4; L_1, L_2, L_3, X_1, X_2, X_3$		X_0, X_4, $\mathbf{X}\cdot\mathbf{L}$, $\mathbf{X}\cdot\mathbf{X}$

$$
\begin{aligned}
&[G,L_i] = 0, \quad [G,P_i] = -P_i, \quad [G,X_i] = 0 \quad (i = 1,2,3),\\
&[G,X_0] = X_0, \quad [G,X_4] = -X_4,\\
&[L_i,P_k] = \epsilon_{ikl}P_l \quad (i,k,l = 1,2,3),\\
&[L_3,X_1] = X_2, \quad [L_3,X_2] = -X_1,\\
&[L_3,X_3] = 0, \quad [L_3,X_4] = 0,\\
&[P_m,X_\nu] = 2X_4S_{m\nu} \quad (m = 1,2,3;\ \nu = 1,2,3,4),\\
&[X_\mu,X_\nu] = 0 \quad (\mu,\nu = 0,1,2,3,4),\\
&[B_i,B_j] = 2\epsilon_{ijk}B_k, \quad \text{where } B_i \equiv L_i + (\epsilon/2)(P_i + C_i),\\
&[B_i,\tilde{\mathrm{A}}] = (\epsilon/2)\tilde{\mu}_0X_i, \quad \text{where}\\
&\quad \tilde{\mathrm{A}} \equiv L_3 - (\epsilon/2)(P_3 + C_3) + \tilde{\mu}_0X_0. \qquad (2)
\end{aligned}
$$

These commutation relations allow us to determine the structure constants for each algebra. The structure constants c_{ij}^k are defined by

$$[X_i,X_j] = \sum_k c_{ij}^k X_k. \qquad (3)$$

Let us give a brief outline of the method used to calculate the Casimir invariants. For a more detailed description of the method we refer to Refs. 3 or 4. Let $\{X_i\}$ be a basis for the Lie algebra. We attempt to find a function $F(X_i)$ of the generators, such that

$$[X_i,F(X_k)] = 0, \qquad (4)$$

where the "i" subscript sums over all basis elements. The way to proceed is to replace the X_i by differential operators

$$\sum_{l,k=1}^{n} c_{ik}^l x_l \frac{\partial}{\partial x_k}. \qquad (5)$$

We now have a set of differential operators which act on real variables. Our problem is translated into one of finding a set of $F(x_k)$ satisfying the system of partial differential equations:

$$X_iF(x_k) = 0, \quad i,k = l,\ldots,n, \qquad (6)$$

where we are dealing now with real variables. In the solution of the system (6), we replace the real variables x_k, by the generators X_k. However, we must symmetrize the results to take into account the noncommutativity of the X_k's in the substitution. The resulting expressions are the Casimir invariants satisfying (4). Table I lists all the invariant operators for the subalgebra of Ref. 6. In the first column we find all eight-dimensional subalgebras classified in Ref. 6. In Table II we find the eight-dimensional subalgebras of Ref. 7. The generators which appear at the right of the semicolon belong to the derived algebras.

The second column is the range of the parameters of the algebras of the first column which in fact may represent several of these classified in Ref. 6. Our result, the full set of invariant operators for each algebra, is contained in the last column.

Remark: In all cases encountered when the invariants are not polynomials, the operator products occurring in the invariant expressions commute among themselves. In the last algebra of Table I, one of the polynomial Casimir operators involves noncommutative products of operators. It seems that this phenomenon arises when the algebra has a nonabelian subalgebra as is the case here for the generators B_i satisfying the commutation relation

$$[B_i,B_j] = 2\epsilon_{ijk}B_k. \qquad (7)$$

In Ref. 4 such a case arises under similar conditions: the algebra in question being denoted $A_{5,40}$.

ACKNOWLEDGMENTS

The author is grateful to M. Perroud for fruitful discussions and the help provided by his method of Ref. 7 and D. Rand for his computer program.

This research was supported by funding from the Fonds F.C.A.C. Pour l' aide et le Soutien a la Recherche, Province de Québec.

[1]D. B. Lichtenberg, *Unitary Symmetry and Elementary Particles* (Academic, New York, 1970).
[2]J. Patera and P. Winternitz, J. Math. Phys. **16**, 1597 (1975).
[3]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977 (1976).
[4]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[5]V. M. Fedorchuk, Ukrainian Math. J. **V31**, 554 (1979).
[6]V. M. Fedorchuk, Ukrainian Math. J. **V33**, 535 (1981).
[7]M. Perroud, "The fundamental invariants of inhomogeneous classical groups," Departement de mathematiques appliquees, Ecole Polytechnique de Montreal report, 1982.

Finite- and infinite-dimensional representations of the orthosymplectic superalgebra OSP(3,2)

Joris Van der Jeugt[a)]
Queen Mary College, Department of Applied Mathematics, Mile End Road, London E1 4NS, England

(Received 23 February 1984; accepted for publication 20 April 1984)

The shift operator technique is used to give a complete analysis of all finite- and infinite-dimensional irreducible representations of the orthosymplectic superalgebra osp(3,2). For all cases, the star or grade star conditions for the algebra are investigated. Only two finite-dimensional representations are grade star representations, if the representation space is required to be a graded Hilbert space. When the even part is $so(3) \oplus sp(2,\mathbb{R}) \approx su(2) \oplus su(1,1)$, an infinite class of infinite-dimensional star representations is found. One of them can be realized in terms of two-valued functions of a complex variable. This representation reduces to the sum of two metaplectic representations of sp(2). We show that it is precisely this "metaplectic representation for osp(3,2)" which gives the spin-energy eigenstates for the one-dimensional harmonic oscillator with spin $\frac{1}{2}$ states.

PACS numbers: 11.30.Pb, 02.20. + b, 02.90. + p

I. INTRODUCTION

Since the simple Lie superalgebras were classified by Kac[1] and other authors,[2,3] they have been the subject of several papers. Finite-dimensional representations of superalgebras have been studied in general.[4] There are also several works in which representations of specific Lie superalgebras were investigated (see Ref. 5 and references therein). Much less is known about infinite-dimensional representations of superalgebras.

In the present paper we analyze both finite- and infinite-dimensional irreducible representations (irreps) of the orthosymplectic superalgebra osp(3,2). We use the shift operator technique, developed by Hughes and Yadegar,[6] and used by Hughes to classify representations of the superalgebra osp(1,2).[7]

The even part of osp(3,2) is the semisimple Lie algebra $so(3) \oplus sp(2)$, isomorphic to $su(2) \oplus su(2)$. Hence, the shift operators are $su(2) \oplus su(2)$ shift operators, which change both su(2) labels by certain numbers when acting on a basis state of an $su(2) \oplus su(2)$ representation. The reduction $osp(3,2) \supset so(3) \oplus sp(2)$ is considered and we show that in general any osp(3,2) irrep decomposes into eight irreducible representations of the subalgebra, a result which was proven in a more general way by Kac[4] for the finite-dimensional representations. In the present paper, we show that this property is true also for the infinite-dimensional osp(3,2) irreps. The two independent invariants of osp(3,2), I_2 and I_4, are explicitly constructed and their eigenvalues for an osp(3,2) irrep $(p;q)$ are given in terms of p and q. We show that the I_2 and I_4 eigenvalues do not specify the superalgebra representations uniquely.

The generalization of a Hermitian operation for a Lie algebra is a star or grade star operation for a Lie superalgebra.[8] The study of star and grade star operations is in fact equivalent to the study of the real forms of the complex superalgebra osp(3,2).[3] We investigate all possible star and grade star operations of osp(3,2), and find that each of the four Hermitian operations on the even part $osp(3,2)_{\bar{0}}$ can be extended in two possible ways to a star or a grade star operation. Then we consider whether the osp(3,2) irreps are star (respectively, grade star) representations, which is the analog of anti-Hermitian representations for Lie algebras. It turns out that this problem is closely related to the choice of a nondegenerate Hermitian form $\langle | \rangle$ on the representation space V. Moreover we prove that if one requires the Hermitian form to be positive definite, then the finite-dimensional representations are not grade star (nor star) representations, except for the five-dimensional irrep $(0;\frac{1}{2})$ and the eight-dimensional irrep $(\frac{1}{2};\frac{1}{2})$ [the notation $(p;q)$ for an osp(3,2) irrep is explained in Sec. V]. It is possible, however, to choose a nondegenerate Hermitian form which is not positive definite, such that all the finite-dimensional osp(3,2) irreps are grade star representations.

We prove that a class of infinite-dimensional osp(3,2) irreps are star representations, which are consistent with a positive definite Hermitian form on the representation space. These representations are finite-dimensional with respect to the so(3) part, but infinite-dimensional with respect to the sp(2) part in the reduction $osp(3,2) \rightarrow so(3) \oplus sp(2)$. Elsewhere,[9] this algebra has been denoted by $Osp(3|2,\mathbb{R})$. It is the algebra of transformations in a five-dimensional space with one "bosonic" degree of freedom (the coordinate x and the momentum p) and three "fermionic" degrees of freedom (C_1,C_2,C_3). The transformations generated by $Osp(3|2,\mathbb{R})$ leave the (anti-)commutation relations

$$\begin{aligned} &\{C_j,C_k\} = 2\delta_{jk}, \\ &[C_j,x] = [C_j,p] = 0, \quad (j,k \in \{1,2,3\}) \\ &[x,p] = i, \end{aligned} \tag{1.1}$$

invariant. One of the representations of this algebra decomposes into only two subalgebra irreps, for which the so(3) label is $\frac{1}{2}$ and the $sp(2) \approx su(1,1)$ label is $-\frac{1}{4}$ or $-\frac{3}{4}$. We call this representation the "metaplectic representation" of osp(3,2), since on restriction to su(1,1) it yields the direct sum of two metaplectic representations of su(1,1). Metaplectic representations of $SU(k,l)$ have been studied in general by

[a)] Research assistant N.F.W.O. (Belgium). Permanent address: Seminarie voor Wiskundige Natuurkunde, R.U.G., Krijgslaan 281-S9, B-9000 Gent, Belgium.

 0022-2488/84/113334-16$02.50

Sternberg and Wolf.[10] For the metaplectic representation of osp(3,2), we give the explicit actions of the superalgebra generators on the basis states. We show how the basis states can be realized in terms of elements of $\mathscr{H}(\mathbb{C},\mathbb{C}^2)$. This is the space of holomorphic functions $f:\mathbb{C}\to\mathbb{C}^2$, with components f_1 and f_2, which satisfy

$$\int(|f_1(z)|^2+|f_2(z)|^2)\exp(-|z|^2)d\lambda(z)<\infty,$$

where λ is the Lebesgue measure on $\mathbb{C}$. The Lie superalgebra generators are then realized as operators acting on $\mathscr{H}(\mathbb{C},\mathbb{C}^2)$. The results are similar to the properties obtained for the metaplectic representation of osp(1,2).[7]

The metaplectic representation for osp(3,2) leads to a physical interpretation: we show that the energy eigenstates of the one-dimensional harmonic oscillator, which are simultaneously spin $\frac{1}{2}$ states, are precisely the basis vectors of the metaplectic representation. This implies that osp(3,2) is the spectrum generating algebra for the harmonic oscillator with spin $\frac{1}{2}$ states.

In a final section, we discuss the Jordan structure of osp(3,2), and give the explicit form of its underlying Jordan superalgebra.

II. THE LIE SUPERALGEBRA OSP(3,2) OR B(1,1)

The even part $\mathrm{osp}(3,2)_{\bar{0}}$ of the Lie superalgebra osp(3,2) consists of the direct sum Lie algebra $\mathrm{so}(3)\oplus\mathrm{sp}(2)$, which is isomorphic to $A_1\oplus A_1$. We denote the generator basis by $s_0,s_\pm$ and $t_0,t_\pm$, respectively, which have the following commutators:

$$\begin{aligned}&[s_0,s_\pm]=\pm s_\pm,\quad [s_+,s_-]=2s_0,\\ &[t_0,t_\pm]=\pm t_\pm,\quad [t_+,t_-]=2t_0,\\ &[s_\mu,t_\nu]=0\quad(\mu,\nu=0,\pm).\end{aligned}\tag{2.1}$$

The odd part $\mathrm{osp}(3,2)_{\bar{1}}$ consists of the tensor product of a three-dimensional tensor operator of so(3) and a two-dimensional sp(2) tensor. We denote its components by $R_{\alpha,\beta}(\alpha=-1,0,1;\beta=-\frac{1}{2},+\frac{1}{2})$; these satisfy the following relations:

$$\begin{aligned}&[s_0,R_{\alpha,\beta}]=\alpha R_{\alpha,\beta},\\ &[s_\pm,R_{\alpha,\beta}]=[(1\mp\alpha)(2\pm\alpha)]^{1/2}R_{\alpha\pm1,\beta},\\ &[t_0,R_{\alpha,\beta}]=\beta R_{\alpha,\beta},\\ &[t_\pm,R_{\alpha,\beta}]=[(\tfrac{1}{2}\mp\beta)(\tfrac{3}{2}\pm\beta)]^{1/2}R_{\alpha,\beta\pm1}.\end{aligned}\tag{2.2}$$

The multiplication in a Lie superalgebra $L=L_{\bar{0}}\oplus L_{\bar{1}}$ satisfies[1]

$$[A,B]=-(-1)^{\alpha\beta}[B,A],\tag{2.3a}$$

$$(-1)^{\gamma\alpha}[A,[B,C]]+(-1)^{\alpha\beta}[B,[C,A]]+(-1)^{\beta\gamma}[C,[A,B]]=0,\tag{2.3b}$$

and

$$[L_\alpha,L_\beta]\subset L_{\alpha+\beta},\tag{2.3c}$$

where $A\in L_\alpha$, $B\in L_\beta$, and $C\in L_\gamma(\alpha,\beta,\gamma\in\{\bar{0},\bar{1}\})$. The Cartan subalgebra H of $\mathrm{osp}(3,2)_{\bar{1}}$ is spanned by $\{s_0,t_0\}$. A form $\alpha\in H^*$ is a root of osp(3,2) if and only if

$$L_\alpha=\{X\in\mathrm{osp}(3,2)\,|\,[h,X]=\alpha(h)X,\ \forall h\in H\}\neq\{0\}.\tag{2.4}$$

Obviously, the roots of osp(3,2) are the roots of $\mathrm{so}(3)\oplus\mathrm{sp}(2)$ and the weights of the tensor representation $R^{[1,1/2]}$. If α and β are two roots of the Lie superalgebra, and if $X\in L_\alpha$ and $Y\in L_\beta$, then we have[1]

$$[X,Y]\in L_{\alpha+\beta}.\tag{2.5}$$

In order to obtain the multiplication table for $\mathrm{osp}(3,2)_{\bar{1}}$, we make use of (2.5) and the graded Jacobi identity (2.3b). This determines the product on $\mathrm{osp}(3,2)_{\bar{1}}$ unambiguously (up to an overall multiplicative constant for the tensor components). The nonvanishing products among the tensor components are

$$\begin{aligned}&[R_{1,1/2},R_{0,-1/2}]=(1/\sqrt{2})s_+,\\ &[R_{0,1/2},R_{0,1/2}]=2t_+,\\ &[R_{1,1/2},R_{-1,1/2}]=-2t_+,\\ &[R_{0,1/2},R_{0,-1/2}]=-2t_0,\\ &[R_{1,1/2},R_{-1,-1/2}]=-s_0+2t_0,\\ &[R_{0,1/2},R_{-1,-1/2}]=-(1/\sqrt{2})s_-,\\ &[R_{1,-1/2},R_{0,1/2}]=-(1/\sqrt{2})s_+,\\ &[R_{0,-1/2},R_{0,-1/2}]=-2t_-,\\ &[R_{1,-1/2},R_{-1,1/2}]=s_0+2t_0,\\ &[R_{0,-1/2},R_{-1,1/2}]=(1/\sqrt{2})s_-.\\ &[R_{1,-1/2},R_{-1,-1/2}]=2t_-.\end{aligned}\tag{2.6}$$

The Lie superalgebra osp(3,2) is then completely determined by (2.1), (2.2), (2.6), and (2.3a).

III. INVARIANTS AND SUBALGEBRA SCALARS

In this section we shall consider some special elements of the enveloping algebra of osp(3,2), namely the invariants and some scalar operators with respect to the $\mathrm{osp}(3,2)_{\bar{0}}$ subalgebra. The subalgebra $\mathrm{osp}(3,2)_{\bar{0}}$ is isomorphic to the Lie algebra $\mathrm{su}(2)\oplus\mathrm{su}(2)$, and its Casimir invariants are given by

$$\begin{aligned}&S^2=s_+s_-+s_0^2-s_0,\\ &T^2=t_+t_-+t_0^2-t_0.\end{aligned}\tag{3.1}$$

In order to determine the subalgebra scalar operators, we define

$$(R\times R)^{[k,l]}_{\alpha,\beta}=\sum\langle 1\alpha_1 1\alpha_2|k\alpha\rangle\langle\tfrac{1}{2}\beta_1\tfrac{1}{2}\beta_2|l\beta\rangle R_{\alpha_1\beta_1}R_{\alpha_2\beta_2},\tag{3.2}$$

where $\langle\cdots|\cdots\rangle$ is an su(2) Clebsch–Gordan coefficient. The second-order invariant I_2 of osp(3,2) must be a linear combination of S^2, T^2, and $(R\times R)^{[0,0]}_{0,0}$, and we find

$$I_2=\sqrt{6}(R\times R)^{[0,0]}_{0,0}+S^2-4T^2,$$

or, explicitly

$$\begin{aligned}I_2=2(&R_{1,1/2}R_{-1,-1/2}-R_{0,1/2}R_{0,-1/2}\\&-R_{1,-1/2}R_{-1,1/2}+s_0-t_0)+S^2-4T^2.\end{aligned}\tag{3.3}$$

The following subalgebra scalars of fourth degree in the generators are defined in terms of (3.2):

$$\begin{aligned}C^{(2,0;2)}=&-\frac{2}{\sqrt{6}}(R\times R)^{[2,0]}_{0,0}(s_+s_--2s_0^2-s_0)\\&+(R\times R)^{[2,0]}_{-1,0}s_+(2s_0+1)\\&-(R\times R)^{[2,0]}_{1,0}s_-(2s_0-1)\end{aligned}$$

$$+ (R \times R)^{[2,0]}_{-2,0} s_+^2 + (R \times R)^{[2,0]}_{2,0} s_-^2, \quad (3.4)$$

$$C^{(1,1;2)} = 2(R \times R)^{[1,1]}_{0,0} s_0 t_0 + \sqrt{2}[(R \times R)^{[1,1]}_{0,-1} t_+ - (R \times R)^{[1,1]}_{0,1} t_-] s_0 + \sqrt{2}[(R \times R)^{[1,1]}_{-1,0} s_+ - (R \times R)^{[1,1]}_{1,0} s_-] t_0 + [(R \times R)^{[1,1]}_{-1,-1} s_+ - (R \times R)^{[1,1]}_{1,-1} s_-] t_+ - [(R \times R)^{[1,1]}_{-1,1} s_+ - (R \times R)^{[1,1]}_{1,1} s_-] t_-, \quad (3.5)$$

$$C^{(0,0;4)} = \sum_\alpha \langle 2\alpha 2 - \alpha | 00 \rangle (R \times R)^{[2,0]}_{\alpha,0} (R \times R)^{[2,0]}_{-\alpha,0}. \quad (3.6)$$

The fourth-order invariant of osp(3,2) is a linear combination of fourth-order subalgebra scalars (which include products of I_2, S^2, and T^2), and we obtain a solution for the coefficients of the linear combination by requiring that the invariant must commute with all the tensor components $R_{\alpha,\beta}$. We find (up to an overall multiplicative constant) two independent solutions, one of which is precisely $(I_2)^2$. The second solution gives the fourth-order invariant I_4:

$$I_4 = \tfrac{1}{12}[6C^{(2,0;2)} + 24C^{(1,1;2)} + 14I_2S^2 + 24I_2T^2 - 5(S^2)^2 + 8S^2T^2 + 48(T^2)^2 + 30S^2 - 36T^2]. \quad (3.7)$$

For completeness, we also mention the first relation:

$$(I_2)^2 = -\frac{6}{\sqrt{5}} C^{(0,0;4)} + 2I_2(S^2 - 4T^2) - (S^2)^2 + 8S^2T^2 - 16(T^2)^2 + 6S^2 + 12T^2. \quad (3.8)$$

The operators I_2, I_4, S^2, and T^2 are four independent commuting operators. From (3.7), one would have the impression that one of the scalars $C^{(2,0;2)}$ or $C^{(1,1;2)}$ is still independent of the four operators mentioned. However, there is a relation, namely

$$9[C^{(2,0;2)}]^2 + 12[C^{(1,1;2)}]^2 + 3C^{(2,0;2)} \times \{I_2(-4S^2 + 4T^2 + 15) + 4(S^2)^2 + 16(T^2)^2 - 20S^2T^2 + 76T^2 - 15S^2 + 18\} + 6C^{(1,1;2)}\{2I_2 - 10S^2 + 4T^2 + 9\} + (I_2)^2S^2(4S^2 + 4T^2 - 3) + 2I_2S^2\{-4(S^2)^2 + 12S^2T^2 + 16(T^2)^2 + 3S^2 - 4T^2\} + S^2\{4(S^2)^3 - 28(S^2)^2T^2 + 32S^2(T^2)^2 + 64(T^2)^3 - 3(S^2)^2 - 184S^2T^2 - 128(T^2)^2 - 144S^2 + 204T^2 + 108\} = 0, \quad (3.9)$$

which shows, together with (3.7), that neither $C^{(2,0;2)}$ nor $C^{(1,1;2)}$ are functionally independent of I_2, I_4, S^2, and T^2. This is consistent with the fact that there is no missing label problem in the reduction osp(3,2)→su(2) ⊕ su(2), as we shall see in Sec. V. Actually, the method of Sec. V showing that there is no degeneracy for osp(3,2)→su(2) ⊕ su(2), provided the clue to the existence of a relation of type (3.9).

IV. SHIFT OPERATORS FOR osp(3,2) ⊃ su(2) ⊕ su(2)

The Lie algebra $\mathrm{osp}(3,2)_{\bar{0}}$ is a subalgebra of osp(3,2), hence every irreducible representation of osp(3,2) is also a representation of $\mathrm{osp}(3,2)_{\bar{0}} \approx \mathrm{su}(2) \oplus \mathrm{su}(2)$. Consequently, the su(2) ⊕ su(2) labels can be used to classify the basis states of an osp(3,2) irrep. In Sec. V we shall see that there is no degeneracy in the reduction osp(3,2)→su(2) ⊕ su(2). Hence the states of an osp(3,2) irrep are completely determined by

$$|s,m,t,n\rangle, \quad (4.1)$$

where $s(s+1)$, m, $t(t+1)$, and n are the eigenvalues of the operators S^2, s_0, T^2, and t_0, respectively. If m and n are irrelevant labels, which happens, for instance, in calculations which contain exclusively su(2) ⊕ su(2) scalars, whose eigenvalues are independent of m and n anyhow, we summarily denote the kets $|s,m,t,n\rangle$ as $|s,t\rangle$. Formulas containing this shorthand notation should be understood as being valid for all permissable m- and n-values.

The shift operators we shall need in order to analyze the irreps of osp(3,2) are su(2) ⊕ su(2) shift operators. Their explicit forms follow from the general analysis of su(2) shift operators by Hughes and Yadegar[6] and the method to obtain su(2) ⊕ su(2) shift operators from them.[11] The following expressions are obtained:

$$O^{-1,-1/2,-1/2}_{s,t,n} = -R_{-1,-1/2} s_+ (s+m)(t+n) + R_{-1,1/2} s_+ t_- (s+m) - R_{1,-1/2} s_- (s-m)(t+n) + R_{1,1/2} s_- t_- (s-m) + \sqrt{2}[R_{0,-1/2}(t+n) - R_{0,1/2} t_-](s+m)(s-m), \quad (4.2)$$

$$O^{-1,1/2,1/2}_{s,t,n} = R_{-1,1/2} s_+ (s+m)(t+n+1) + R_{-1,-1/2} s_+ t_+ (s+m) + R_{1,1/2} s_- (s-m)(t+n+1) + R_{1,-1/2} s_- t_+ (s-m) - \sqrt{2}[R_{0,1/2}(t+n+1) + R_{0,-1/2} t_+](s+m)(s-m), \quad (4.3)$$

$$O^{0,-1/2,-1/2}_{s,t,n} = R_{-1,-1/2} s_+ (t+n) - R_{-1,1/2} s_+ t_- - R_{1,-1/2} s_- (t+n) + R_{1,1/2} s_- t_-, \quad (4.4)$$

$$O^{0,1/2,1/2}_{s,t,n} = -R_{-1,1/2} s_+ (t+n+1) - R_{-1,-1/2} s_+ t_+ + R_{1,1/2} s_- (t+n+1) + R_{1,-1/2} s_- t_+, \quad (4.5)$$

$$O^{1,-1/2,-1/2}_{s,t,n} = R_{-1,-1/2} s_+ (s-m+1)(t+n) - R_{-1,1/2} s_+ t_- (s-m+1) + R_{1,-1/2} s_- (s+m+1)(t+n) - R_{1,1/2} s_- t_- (s+m+1) + \sqrt{2}[R_{0,-1/2}(t+n) - R_{0,1/2} t_-](s+m+1)(s-m+1), \quad (4.6)$$

$$O^{1,1/2,1/2}_{s,t,n} = -R_{-1,1/2}s_+(s-m+1)(t+n+1) - R_{-1,-1/2}s_+t_+(s-m+1)$$
$$-R_{1,1/2}s_-(s+m+1)(t+n+1) - R_{1,-1/2}s_-t_+(s+m+1)$$
$$-\sqrt{2}[R_{0,1/2}(t+n+1) + R_{0,-1/2}t_+](s+m+1)(s-m+1). \tag{4.7}$$

Note that in the right-hand sides of (4.2)–(4.7) the labels s and t could have been replaced by operators $\hat{s}$ and $\hat{t}$, respectively, whose actions on su(2) ⊕ su(2) states are given by

$$\hat{s}|s,m,t,n\rangle = s|s,m,t,n\rangle, \quad \hat{t}\,|s,m,t,n\rangle = t\,|s,m,t,n\rangle. \tag{4.8}$$

The operators (4.2)–(4.7) shift the eigenvalue s by $-1,0$ or $+1$, and shift t and n by $\pm\frac{1}{2}$:

$$O^{i,j,j}_{s,t,n}|s,m,t,n\rangle \sim |s+i,m,t+j,n+j\rangle \quad (i=0,\pm1, j=\pm\tfrac{1}{2}). \tag{4.9}$$

It is convenient to use the normalized operators whose actions on eigenstates $|s,m,t,n\rangle$ are related to those of the above operators by

$$A^{1,\pm1/2}_{s,t} = [(s+m+1)(s-m+1)(t+n+\tfrac{1}{2}\pm\tfrac{1}{2})]^{-1/2}O^{1,\pm1/2,\pm1/2}_{s,t,n},$$
$$A^{0,\pm1/2}_{s,t} = (t+n+\tfrac{1}{2}\pm\tfrac{1}{2})^{-1/2}O^{0,\pm1/2,\pm1/2}_{s,t,n}, \tag{4.10}$$
$$A^{-1,\pm1/2}_{s,t} = [(s+m)(s-m)(t+n+\tfrac{1}{2}\pm\tfrac{1}{2})]^{-1/2}O^{-1,\pm1/2,\pm1/2}_{s,t,n}.$$

If we consider the set of quadratic products of shift operators, it is easy to see that six scalar operators $A^{-i,-j}_{s+i,t+j}A^{i,j}_{s,t}$ belong to this set. Of course, these six product operators are not all independent. The relations which exist among those products are the basis of our analysis. It turns out that for every two scalar products there exists a combination which can be expressed in terms of the invariants I_2, I_4 and the subalgebra Casimirs S^2 and T^2. The relations are

$$(2s+4t+3)A^{1,1/2}_{s-1,t-1/2}A^{-1,-1/2}_{s,t} - (2s-4t-1)A^{1,-1/2}_{s-1,t+1/2}A^{-1,1/2}_{s,t}$$
$$-(2s-1)(2t+1)\{2I_4 - [s(s+3)+4t(t+1)]I_2 - \tfrac{1}{2}(s-2t)(s-2t+1)(s+2t+2)(s+2t+3)\} = 0, \tag{4.11}$$
$$(4t+3)A^{1,-1/2}_{s-1,t+1/2}A^{-1,1/2}_{s,t} + (2s+4t+3)A^{0,-1/2}_{s,t+1/2}A^{0,1/2}_{s,t}$$
$$-2s(t+1)\{2I_4 - [3s(s+2)+4t(t+1)+4st]I_2 + \tfrac{1}{2}(s+2t)(s+2t+2)(s-2t+1)(3s+2t+3)\} = 0, \tag{4.12}$$
$$-(2s-4t-1)A^{0,-1/2}_{s,t+1/2}A^{0,1/2}_{s,t} + (4t+3)A^{-1,-1/2}_{s+1,t+1/2}A^{1,1/2}_{s,t}$$
$$+2(s+1)(t+1)\{2I_4 - [3(s+1)(s-1)+4t^2-4st]I_2 + \tfrac{1}{2}(s-2t+1)(s-2t-1)(s+2t)(3s-2t)\} = 0, \tag{4.13}$$
$$(2s+4t+3)A^{-1,-1/2}_{s+1,t+1/2}A^{1,1/2}_{s,t} - (2s-4t-1)A^{-1,1/2}_{s+1,t-1/2}A^{1,-1/2}_{s,t}$$
$$+(2s+3)(2t+1)\{2I_4 - [(s+1)(s-2)+4t(t+1)]I_2 - \tfrac{1}{2}(s+2t+1)(s+2t)(s-2t-1)(s-2t-2)\} = 0, \tag{4.14}$$
$$-(4t+1)A^{-1,1/2}_{s+1,t-1/2}A^{1,-1/2}_{s,t} - (2s+4t+3)A^{0,1/2}_{s,t-1/2}A^{0,-1/2}_{s,t}$$
$$-2(s+1)t\{2I_4 - [(s+1)(3s+1)+4t(t+1)+4t(s+1)]I_2$$
$$+\tfrac{1}{2}(s+2t+3)(s+2t+1)(s-2t-2)(3s+2t+2)\} = 0, \tag{4.15}$$
$$(2s-4t-1)A^{0,1/2}_{s,t-1/2}A^{0,-1/2}_{s,t} - (4t+1)A^{1,1/2}_{s-1,t-1/2}A^{-1,-1/2}_{s,t}$$
$$+2st\{2I_4 - [s(3s+2)+4t(t+1)-4st]I_2 + \tfrac{1}{2}(s-2t-2)(s-2t)(s+2t+3)(3s-2t+1)\} = 0. \tag{4.16}$$

Besides the relations among scalar products, there are also a set of relations connecting the nonscalar products of shift operators. Their explicit forms will be extremely useful in the analysis of the osp(3,2) representations. We obtain

$$A^{\pm1,\pm1/2}_{s\pm1,t\pm1/2}A^{\pm1,\pm1/2}_{s,t} = A^{\pm1,\mp1/2}_{s\pm1,t\mp1/2}A^{\pm1,\mp1/2}_{s,t} = 0, \tag{4.17}$$
$$(t+1)A^{\pm1,1/2}_{s\pm1,t-1/2}A^{\pm1,-1/2}_{s,t}$$
$$+tA^{\pm1,-1/2}_{s\pm1,t+1/2}A^{\pm1,1/2}_{s,t} = 0, \tag{4.18}$$
$$(s-1)A^{-1,\pm1/2}_{s,t\pm1/2}A^{0,\pm1/2}_{s,t}$$
$$+(s+1)A^{0,\pm1/2}_{s-1,t\pm1/2}A^{-1,\pm1/2}_{s,t} = 0, \tag{4.19}$$
$$(s+2)A^{+1,\pm1/2}_{s,t\pm1/2}A^{0,\pm1/2}_{s,t} + sA^{0,\pm1/2}_{s+1,t\pm1/2}A^{1,\pm1/2}_{s,t} = 0, \tag{4.20}$$
$$A^{+1,\pm1/2}_{s-1,t\pm1/2}A^{-1,\pm1/2}_{s,t} - (2s-1)A^{0,\pm1/2}_{s,t\pm1/2}A^{0,\pm1/2}_{s,t} = 0, \tag{4.21}$$
$$A^{-1,\pm1/2}_{s+1,t\pm1/2}A^{1,\pm1/2}_{s,t} + (2s+3)A^{0,\pm1/2}_{s,t\pm1/2}A^{0,\pm1/2}_{s,t} = 0, \tag{4.22}$$
$$(s+2t+1)A^{0,1/2}_{s-1,t-1/2}A^{-1,-1/2}_{s,t} + 2stA^{0,-1/2}_{s-1,t+1/2}A^{-1,1/2}_{s,t}$$
$$+(s-1)(2t+1)A^{-1,1/2}_{s,t-1/2}A^{0,-1/2}_{s,t} = 0, \tag{4.23}$$
$$(s-2t-1)A^{0,-1/2}_{s-1,t+1/2}A^{-1,1/2}_{s,t} - 2s(t+1)A^{0,1/2}_{s-1,t-1/2}$$
$$\times A^{-1,-1/2}_{s,t} - (s-1)(2t+1)A^{-1,-1/2}_{s,t+1/2}A^{0,1/2}_{s,t} = 0, \tag{4.24}$$
$$(-s+2t)A^{0,1/2}_{s+1,t-1/2}A^{1,-1/2}_{s,t} - 2(s+1)tA^{0,-1.2}_{s+1,t+1/2}A^{1,1/2}_{s,t}$$
$$-(s+2)(2t+1)A^{1,1/2}_{s,t-1/2}A^{0,-1/2}_{s,t} = 0, \tag{4.25}$$
$$-(s+2t+2)A^{0,-1/2}_{s+1,t+1/2}A^{1,1/2}_{s,t} + 2(s+1)(t+1)A^{0,1/2}_{s+1,t-1/2}$$
$$+(s+2)(2t+1)A^{1,-1/2}_{s,t+1/2}A^{0,1/2}_{s,t} = 0. \tag{4.26}$$

V. ANALYSIS OF osp(3,2) IRREPS

We analyze the osp(3,2) irreps in the reduction osp(3,2)→su(2) ⊕ su(2). The su(2) ⊕ su(2) irreps which appear in the decomposition of an osp(3,2) irrep are denoted by (s,t). We shall also give a pair of numbers which label the osp(3,2) irreps uniquely.

We do not restrict ourselves to finite-dimensional representations of osp(3,2). Therefore, the su(2) ⊕ su(2) irreps in which such a representation decomposes may also be infinite-dimensional. Hence, we must bear in mind that the eigenvalues s and t can take on negative values, or even complex values.

We shall consider the representations of osp(3,2) for which the states can be connected by consecutive actions of the shift operators $A^{ij}_{s,t}$ (or $O^{i,j,j}_{s,t,n}$). Since the matrix elements of the shift operators are proportional to reduced matrix elements of the tensor $R^{[1,1/2]}$,[6] this method will give us all the irreducible representations of osp(3,2).

The nonscalar equations (4.17) show that if (s,t) and $(s-1,t-\frac{1}{2})$ are parts of an osp(3,2) irrep [always in the reduction to su(2) ⊕ su(2)], then neither $(s+1,t+\frac{1}{2})$ nor $(s-2,t-1)$ belong to that representation, because the square of $A^{\pm 1,\pm 1/2}$ is zero. Similarly, if (s,t) and $(s+1,t-\frac{1}{2})$ appear in the decomposition of an osp(3,2) irrep, then $(s-1,t+\frac{1}{2})$ and $(s+2,t-1)$ do not. Hence, we may assume that in general the irrep contains the representations (s',t'), $(s'-1,t'-\frac{1}{2})$, and $(s'+1,t'-\frac{1}{2})$, such that the actions of $A^{1,1/2}$ and $A^{-1,1/2}$ upon states $|s',t'\rangle$ vanish. Suppose that $A^{0,1/2}|s',t'\rangle$ would not be zero, which means that $(s',t'+\frac{1}{2})$ is a representation of the osp(3,2) irrep under consideration, then the application of (4.20) upon the states $|s'-1,t'-\frac{1}{2}\rangle$ and of (4.19) upon $|s'+1,t'-\frac{1}{2}\rangle$ show that $(s'-1,t')$ and $(s'+1,t')$ are also parts of the osp(3,2) irrep. Obviously, the actions of $A^{\pm 1,1/2}$ upon $|s',t'+\frac{1}{2}\rangle$ states are zero. But now also the action of $A^{0,1/2}$ upon $|s',t'+\frac{1}{2}\rangle$ must vanish, because otherwise the application of (4.20) upon $|s'-1,t'\rangle$ or of (4.19) upon $|s'+1,t'\rangle$ would imply that the representations $(s'-1,t'+\frac{1}{2})$ and $(s'+1,t'+\frac{1}{2})$ are parts of the osp(3,2) irrep, which contradicts our assumption. Hence, we have shown that an osp(3,2) irrep always contains an su(2) ⊕ su(2) irrep $(p,q)=(s',t'+\frac{1}{2})$ such that the actions of $A^{\pm 1,1/2}$ and $A^{0,1/2}$ upon $|p,q\rangle$ states vanish.

Let us first consider the cases where $|p|>\frac{3}{2}$ and $|q|>\frac{3}{2}$. Since the actions of $A^{\pm 1,1/2}$ and $A^{0,1/2}$ upon $|p,q\rangle$ are zero, the application of Eqs. (4.12) and (4.13) immediately gives the following solutions for the eigenvalues of the invariants:

$$I_2|p,q\rangle = (p+2q)(p-2q+1)|p,q\rangle,$$
$$I_4|p,q\rangle = \tfrac{1}{4}(p+2q)(p-2q+1)[3p(p+1) + 2(q+1)(2q-3)]|p,q\rangle. \tag{5.1}$$

The states $|p,q\rangle$ are connected to states $|p+1,q-\frac{1}{2}\rangle$ by means of the operator $A^{1,-1/2}_{p,q}$. The action of the scalar relation (4.14) then implies

$$A^{-1,1/2}_{p+1,q-1/2}A^{1,-1/2}_{p,q}|p,q\rangle = 2(p+1)(2p+3)(2q+1)(p+2q)|p,q\rangle, \tag{5.2}$$

which shows that (p,q) and $(p+1,q-\frac{1}{2})$ are parts of the same osp(3,2) irrep unless $p+2q=0$.

Now we proceed with the analysis starting from $(p+1,q-\frac{1}{2})$. The action of $A^{0,1/2}$ upon $|p+1,q-\frac{1}{2}\rangle$ states vanishes, because otherwise the application of (4.26) upon $|p,q\rangle$, namely

$$(p+2q+2)A^{0,-1/2}_{p+1,q+1/2}A^{1,1/2}_{p,q}|p,q\rangle = 2(p+1)(q+1)A^{0,1/2}_{p+1,q-1/2}A^{1,-1/2}_{p,q}|p,q\rangle, \tag{5.3}$$

would imply that $A^{1,1/2}|p,q\rangle\neq 0$. Similarly, the action of $A^{1,1/2}$ upon $|p+1,q-\frac{1}{2}\rangle$ vanishes since otherwise (4.18) would show that $A^{1,1/2}|p,q\rangle\neq 0$. Obviously, the action of $A^{1,-1/2}$ upon $|p+1,q-\frac{1}{2}\rangle$ vanishes because of (4.17): $A^{1,-1/2}_{p+1,q-1/2}A^{1,-1/2}_{p,q}|p,q\rangle = 0$. But $(p+1,q-\frac{1}{2})$ can be connected to $(p+1,q-1)$, and so we get from Eqs. (4.15) and (5.1) the relation

$$A^{0,1/2}_{p+1,q-1}A^{0,-1/2}_{p+1,q-1/2}|p+1,q-\tfrac{1}{2}\rangle = 4(p+2)^2q(2q-1)|p+1,q-\tfrac{1}{2}\rangle. \tag{5.4}$$

We continue our analysis by investigating which su(2) ⊕ su(2) representations can be connected to $(p+1,q-1)$. The action of Eq. (4.25) upon the states $|p+1,q-\frac{1}{2}\rangle$ shows that $A^{1,1/2}_{p+1,q-1}|p+1,q-1\rangle$ vanishes. Similarly, the action of Eq. (4.20) upon $|p+1,q-\frac{1}{2}\rangle$ implies that $A^{1,-1/2}_{p+1,q-1}|p+1,q-1\rangle = 0$, and the action of (4.22) that $A^{0,-1/2}_{p+1,q-1}|p+1,q-1\rangle = 0$. We then consider the action $A^{-1,-1/2}_{p+1,q-1}|p+1,q-1\rangle$. This is in general not zero, and the scalar relation (4.16) shows

$$A^{1,1/2}_{p,q-3/2}A^{-1,-1/2}_{p+1,q-1}|p+1,q-1\rangle = -4(p+1)(2p+3)(q-1)(p-2p+1)\times|p+1,q-1\rangle, \tag{5.5}$$

implying that $(p+1,q-1)$ and $(p,q-\frac{3}{2})$ are connected to each other unless $p-2q+1=0$. Using some of the nonscalar relations (4.17)–(4.26), it is then straightforward to show that the actions of $A^{1,-1/2}$, $A^{0,-1/2}$, and $A^{-1,-1/2}$ upon $|p,q-\frac{3}{2}\rangle$ all vanish. Hence, the analysis of the enveloping polygon (see Fig. 1) of the multiplicity diagram for the reduction osp(3,2)→su(2) ⊕ su(2) is completed for the side $s\geqslant p$, which consists of the points (p,q), $(p+1,q-\frac{1}{2})$, $(p+1,q-1)$, and $(p,q-\frac{3}{2})$.

The analysis of the reduction of the osp(3,2) irrep for $s\leqslant p$ is quite analogous, and therefore, we do not enter into the details of this investigation. The result is that on this side, the enveloping polygon consists of the points (p,q), $(p-1,q-\frac{1}{2})$, $(p-1,q-1)$, and $(p,q-\frac{3}{2})$. Hence, the enveloping polygon for the multiplicity diagram of the reduction of an osp(3,2) irrep into su(2) ⊕ su(2) irreps is completely determined. This analysis also shows that the osp(3,2) irrep itself is uniquely labeled by the pair $(p;q)$, which is the label (s,t) of the su(2) ⊕ su(2) irrep with highest t-value and corresponding highest s-value that appears in the reduction. The correspondence with the Kac–Dynkin labels[4] $(a_1, a_2; b=a_1-\frac{1}{2}a_2)$, for the finite-dimensional osp(3,2) irreps, is given by

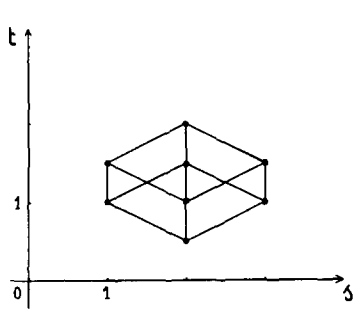


FIG. 1. Multiplicity diagram for a general osp(3,2) irreps in the reduction osp(3,2)→so(3) ⊕ sp(2), and the shift operators connecting the corresponding so(3) ⊕ sp(2) irreps.

$p = a_2/2,$

$q = b/2 = (a_1 - \frac{1}{2}a_2)/2.$

Inside the enveloping polygon, there are only two points which can correspond to an su(2) ⊕ su(2) irrep of the osp(3,2) representation, namely $(p,q-\frac{1}{2})$ and $(p,q-1)$. The states $|p,q-\frac{1}{2}\rangle$ always appear, since we find

$$A^{0,1/2}_{p,q-1/2}A^{0,-1/2}_{p,q}|p,q\rangle = 2p(p+1)(2q-1)(2q+1)|p,q\rangle. \tag{5.6}$$

Whether the representation $(p,q-1)$ appears in the reduction or not, depends on the vanishing of $p+2q$ and $p-2q+1$, since an appropriate use of the scalar relations (4.11)–(4.16) shows that

$$A^{0,1/2}_{p,q-1}A^{0,-1/2}_{p,q-1/2}|p,q-\tfrac{1}{2}\rangle = -2(p+2q)(p-2q+1)|p,q-\tfrac{1}{2}\rangle. \tag{5.7}$$

The multiplicity of the irreps (s,t) which correspond to points on the enveloping polygon is of course equal to one. But for the irreps $(p,q-\frac{1}{2})$ and $(p,q-1)$ there could appear a degeneracy. In order to investigate their multiplicity, we have to consider in how many independent ways a $|p,q-\frac{1}{2}\rangle$ state (resp. a $|p,q-1\rangle$ state) can be obtained from the "highest states" $|p,q\rangle$. Therefore, let us define

$$\begin{aligned} &A^{0,-1/2}_{p,q}|p,q\rangle = |(a);p,q-\tfrac{1}{2}\rangle, \\ &A^{-1,1/2}_{p+1,q-1}A^{0,-1/2}_{p+1,q-1/2}A^{1,-1/2}_{p,q}|p,q\rangle = |(b);p,q-\tfrac{1}{2}\rangle, \end{aligned} \tag{5.8}$$

where (a) and (b) are supplementary labels to distinguish between different $(p,q-\frac{1}{2})$ irreps if its multiplicity were greater than one. The previous analysis then shows

$$|(b);p,q-\tfrac{1}{2}\rangle \sim A^{-1,1/2}_{p+1,q-1}|p+1,q-1\rangle. \tag{5.9}$$

From (4.23) and (4.24) we obtain the relation

$$\begin{aligned} &(2t+1)A^{0,-1/2}_{s-1,t+1/2}A^{-1,1/2}_{s,t} + 2s(t+1)A^{-1,1/2}_{s,t-1/2}A^{0,-1/2}_{s,t} \\ &\quad -(s+2t+1)A^{-1,-1/2}_{s,t+1/2}A^{0,1/2}_{s,t} = 0, \end{aligned} \tag{5.10}$$

whose action upon the state $|p+1,q-\frac{1}{2}\rangle$ produces

$$\begin{aligned} &2qA^{0,-1/2}_{p,q}A^{-1,1/2}_{p+1,q-1/2}|p+1,q-\tfrac{1}{2}\rangle \\ &\quad = -2(p+1)(q+\tfrac{1}{2})A^{-1,1/2}_{p+1,q-1}A^{0,-1/2}_{p+1,q-1/2} \\ &\qquad \times|p+1,q-\tfrac{1}{2}\rangle. \end{aligned} \tag{5.11}$$

For the left-hand side of (5.11) we find

$$\begin{aligned} &A^{0,-1/2}_{p,q}A^{-1,1/2}_{p+1,q-1/2}|p+1,q-\tfrac{1}{2}\rangle \sim A^{0,-1/2}_{p,q}|p,q\rangle \\ &\quad \sim |(a);p,q-\tfrac{1}{2}\rangle, \end{aligned}$$

while the right-hand side gives

$$\begin{aligned} &A^{-1,1/2}_{p+1,q-1}A^{0,-1/2}_{p+1,q-1/2}|p+1,q-\tfrac{1}{2}\rangle \\ &\quad \sim A^{-1,1/2}_{p+1,q-1}|p+1,q-1\rangle \\ &\quad \sim |(b);p,q-\tfrac{1}{2}\rangle, \end{aligned}$$

from which we obtain that

$$|(a);p,q-\tfrac{1}{2}\rangle \sim |(b);p,q-\tfrac{1}{2}\rangle.$$

This shows that the two ways defined in (5.8) are not independent. It is easy to prove that all the other ways in which a $|p,q-\frac{1}{2}\rangle$ state might be defined are finally proportional to $A^{0,-1/2}_{p,q}|p,q\rangle$. Hence, the multiplicity of the $(p,q-\frac{1}{2})$ irrep is equal to one. The same reasoning is valid for the irrep $(p,q-1)$, so that also for this representation the multiplicity is one. As a consequence, there is no degeneracy for the reduction osp(3,2)→su(2) ⊕ su(2), and the states of an osp(3,2) irrep $(p;q)$ are completely labeled by the su(2) ⊕ su(2) labels $|s,m,t,n\rangle$, where $(s,t)\in\{(p,q),(p,q-\frac{1}{2}),\ (p\pm1,q-\frac{1}{2}),\ (p,q-1),\ (p\pm1,q-1),\ (p,q-\frac{3}{2})\}$. Figure 1 shows the reduction of a general osp(3,2) irrep, and the ways in which the corresponding states can be connected by means of the shift operators. Note that these eightfold patterns appear in the finite-dimensional case ($2p$ and $2q$ non-negative integers) as well as in the infinite-dimensional case (where p and/or q can be negative real numbers). There are, however, two main exceptions to this eightfold reduction pattern. Indeed, in the analysis of the osp(3,2) irrep we have mentioned that if $p-2q+1=0$ or $p+2q=0$, some states are not connected to each other. In order to obtain a better insight into these situations, we summarize the matrix elements of the scalar product operators which connect the states to each other. Their expressions follow from the scalar relations (4.11)–(4.16):

$$A^{-1,1/2}_{p+1,q-1/2}A^{1,-1/2}_{p,q}|p,q\rangle = 2(p+1)(2p+3)(2q+1)(p+2q)|p,q\rangle, \tag{5.12}$$

$$A^{0,1/2}_{p,q-1/2}A^{0,-1/2}_{p,q}|p,q\rangle = 2p(p+1)(2q-1)(2q+1)|p,q\rangle, \tag{5.13}$$

$$A^{1,1/2}_{p-1,q-1/2}A^{-1,-1/2}_{p,q}|p,q\rangle = -2p(2p-1)(2q+1)(p-2q+1)|p,q\rangle, \tag{5.14}$$

$$A^{0,1/2}_{p+1,q-1}A^{0,-1/2}_{p+1,q-1/2}|p+1,q-\tfrac{1}{2}\rangle = 4(p+2)^2q(2q-1)|p+1,q-\tfrac{1}{2}\rangle, \tag{5.15}$$

$$A^{1,1/2}_{p,q-1}A^{-1,-1/2}_{p+1,q-1/2}|p+1,q-\tfrac{1}{2}\rangle = -2p(2p+3)(2q-1)(p-2q+1)|p+1,q-\tfrac{1}{2}\rangle, \tag{5.16}$$

$$A^{-1,1/2}_{p+1,q-1}A^{1,-1/2}_{p,q-1/2}|p,q-\tfrac{1}{2}\rangle = 4pq(2p+3)(p+2q)|p,q-\tfrac{1}{2}\rangle, \tag{5.17}$$

$$A^{0,1/2}_{p,q-1}A^{0,-1/2}_{p,q-1/2}|p,q-\tfrac{1}{2}\rangle = -2(p+2q)(p-2q+1)|p,q-\tfrac{1}{2}\rangle, \tag{5.18}$$

$$A^{1,1/2}_{p-1,q-1}A^{-1,-1/2}_{p,q-1/2}|p,q-\tfrac{1}{2}\rangle = -4(2p-1)(p+1)q(p-2q+1)|p,q-\tfrac{1}{2}\rangle, \tag{5.19}$$

$$A^{-1,1/2}_{p,q-1}A^{1,-1/2}_{p-1,q-1/2}|p-1,q-\tfrac{1}{2}\rangle = 2(p+1)(2p-1)(2q-1)(p+2q)|p-1,q-\tfrac{1}{2}\rangle, \tag{5.20}$$

$$A^{0,1/2}_{p-1,q-1}A^{0,-1/2}_{p-1,q-1/2}|p-1,q-\tfrac{1}{2}\rangle = 4(p-1)^2q(2q-1)|p-1,q-\tfrac{1}{2}\rangle, \tag{5.21}$$

$$A^{1,1/2}_{p,q-3/2}A^{-1,-1/2}_{p+1,q-1}|p+1,q-1\rangle = -4(p+1)(2p+3)(q-1)(p-2q+1)|p+1,q-1\rangle, \tag{5.22}$$

$$A^{0,1/2}_{p,q-3/2}A^{0,-1/2}_{p,q-1}|p,q-1\rangle = 8p(p+1)q(q-1)|p,q-1\rangle, \tag{5.23}$$

$$A^{-1,1/2}_{p,q-3/2}A^{1,-1/2}_{p-1,q-1}|p-1,q-1\rangle = 4p(2p-1)(q-1)(p+2q)|p-1,q-1\rangle. \tag{5.24}$$

Now it is easy to see that if $p-2q+1=0$ (and $p+2q\neq0$), the eightfold pattern decomposes into two irreducible repre-

sentations of osp(3,2), since the actions (5.14), (5.16), (5.18), (5.19), and (5.22) are zero. They correspond to two parallelograms in Fig. 1. The first irrep, which we label by $(p;q)$, decomposes into the su(2) ⊕ su(2) irreps (p,q), $(p, q-\frac{1}{2})$, $(p+1,q-\frac{1}{2})$, and $(p+1,q-1)$, and the second osp(3,2) irrep, which we label by $(p-1;q-\frac{1}{2})$, decomposes into $(p-1,q-\frac{1}{2})$, $(p-1,q-1)$, $(p,q-1)$, and $(p,q-\frac{3}{2})$. These "fourfold" representations will turn out to be so-called atypical representations,[4] whereas the "eightfold" irreps are typical. In general, these two fourfold representations, which we labeled by (p,q) and $(p-1;q-\frac{1}{2})$, are parts of a reducible but indecomposable osp(3,2) representation, where the representation space of $(p;q)$ is the factor space, and that of $(p-1;q-\frac{1}{2})$ is the invariant space.[4,5] Let V be the indecomposable (finite-dimensional) representation space, and $V = V_1 \oplus V_2$, where V_1 is the factor space and V_2 is the invariant space. When a basis is chosen in V_1 and V_2, the indecomposable representation ρ is schematically described by

$$\rho: X \to \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} \quad [X \in \mathrm{osp}(3,2)].$$

What we find are the irreducible representations $\rho_1: X \to (A)$ and $\rho_2: X \to (D)$. Note that for all the fourfold representations the I_2 and I_4 eigenvalues are zero, since $p - 2q + 1 = 0$. This shows that the eigenvalues of the invariants do not specify the finite-dimensional irreps of the superalgebra uniquely, whereas for Lie algebras they always do. Analogously, if $p + 2q = 0$, the eightfold pattern decomposes into the two osp(3,2) irreps, the first one labeled by $(p;q)$ and containing the su(2) ⊕ su(2) irreps (p,q), $(p,q-\frac{1}{2})$, $(p-1,q-\frac{1}{2})$, and $(p-1,q-1)$, and the second labeled by $(p+1;q-\frac{1}{2})$ and containing the subalgebra representations $(p+1,q-\frac{1}{2})$, $(p+1,q-1)$, $(p,q-1)$, and $(p,q-\frac{3}{2})$.

Until now we have supposed that $|p| > \frac{3}{2}$ and $|q| > \frac{3}{2}$. The analysis of the remaining cases is similar to the previous one, and we shall not enter into the detailed calculations. An important difference, however, is that we shall have to distinguish between finite-dimensional and infinite-dimensional representations of su(2) ⊕ su(2). In order to understand this, let us consider the finite-dimensional su(2) ⊕ su(2) irrep $(s,0)$ $(2s \in \mathbf{N})$, as part of a certain osp(3,2) irrep. Because of the finite-dimensionality the representation $(s,0)$ contains the states $|s,m,0,0\rangle$, where $m = -s, -s+1, \ldots, s$. But then the expressions (4.2)–(4.7) show that the actions of the shift operators $O^{i,-1/2,-1/2}_{s,t,n}(i = -1,0,1)$ all vanish, since $t_\mu |s,m,0,0\rangle = 0$ for $\mu = 0, \pm$ and $(t+n) = 0$. Hence, the values $t = 0$, and similarly $s = 0$, are limits which cannot be exceeded by the shift operators in the finite-dimensional case.On the contrary, if $(s,0)$ is an infinite-dimensional su(2) ⊕ su(2) irrep, as part of an infinite-dimensional osp(3,2) irrep, the actions of the shift operators usually do not vanish. Suppose, for instance, that the su(2) irrep labeled by s $(s < -1)$ is a discrete positive series D^+ with minimum m-value $\underline{m} = -s$, and that the su(2) irrep labeled by 0 is a discrete positive series D^+ with minimum n-value $\underline{n} = 1$. Then the states of the irrep $(s,0)$ are labeled by $|s,m,0,n\rangle$, where m and n take on an infinity of values: $m = -s, -s+1, \ldots; n = 1,2,3, \ldots$. Obviously, the action of $t_\mu (\mu = 0, \pm)$ or $(t+n)$ on such states is in general different from zero, and hence the actions of the shift operators usually do not vanish. As a consequence, there are no limits for the shift operators if we consider infinite-dimensional representations.

In the case $|p| > \frac{3}{2}$ and $|q| > \frac{3}{2}$, we did not have to deal with the problem of finite- or infinite-dimensionality, since the eightfold patterns did not intersect the lines $s = 0$ or $t = 0$. Let us now consider the case $|p| \leqslant \frac{3}{2}$ or $|q| \leqslant \frac{3}{2}$, and we first investigate the finite-dimensional representations of osp(3,2), which, of course, decompose into finite-dimensional su(2) ⊕ su(2) irreps (s,t). Then s and t are non-negative integers or half-odd integers. A detailed analysis by means of the relations (4.11)–(4.26) then shows that for $p \in \{1, \frac{3}{2}\}$ and $q \geqslant \frac{3}{2}$, or for $q = \frac{3}{2}$ and $p \geqslant 1$, we still have the common eightfold patterns [or the fourfold patterns when $p - 2q + 1 = 0$, i.e., for $(p,q) = (2,\frac{3}{2})$], and the expressions (5.12)–(5.24) are still valid. Because the shift operators cannot exceed the $s = 0$ and $t = 0$ axes, the remaining cases will mainly consist of "truncated" eightfold patterns. We summarize the results.

A1. $q = 1, p > 1$. The osp(3,2) irrep $(p;1)$ reduces to the su(2) ⊕ su(2) irreps $(p,1)$, $(p,\frac{1}{2})$, $(p \pm 1,\frac{1}{2})$, $(p,0)$, and $(p \pm 1,0)$.

A2. $q = 1, p = 1$. The osp(3,2) irrep (1;1) is an atypical representation (a fourfold pattern) and decomposes into (1,1), $(1,\frac{1}{2})$, $(2,\frac{1}{2})$, and (2,0).

A3. $q = 1, p = \frac{1}{2}$. This representation reduces to the su(2) ⊕ su(2) irreps $(\frac{1}{2},1)$, $(\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0)$, $(\frac{3}{2},1)$, and $(\frac{3}{2},0)$.

A4. $q = 1, p = 0$. The irrep (0;1) decomposes into the subalgebra representations (0,1), $(1,\frac{1}{2})$, and (1,0). Hence, this representation corresponds to the 12-dimensional adjoint representation of osp(3,2).

B1. $q = \frac{1}{2}, p \geqslant 1$. The osp(3,2) irreps $(p;\frac{1}{2})$ decompose into the sum of only three su(2) ⊕ su(2) irreps, namely $(p,\frac{1}{2})$, $(p-1,0)$ and $(p+1,0)$.

B2. $q = \frac{1}{2}, p = \frac{1}{2}$. The irrep $(\frac{1}{2};\frac{1}{2})$ of osp(3,2) reduces to $(\frac{1}{2},\frac{1}{2})$ and $(\frac{3}{2},0)$.

B3. $q = \frac{1}{2}, p = 0$. This representation reduces to $(0,\frac{1}{2})$ and (1,0), and is the five-dimensional standard representation by which the osp(3,2) algebra is usually defined.

C1. $q = 0, p = 0$: the trivial representation of osp(3,2).

D1. $p = \frac{1}{2}, q > 1$. The $(\frac{1}{2};q)$ irreps of osp(3,2) decompose into the subalgebra representations $(\frac{1}{2},q)$ $(\frac{1}{2},q-\frac{1}{2})$, $(\frac{1}{2},q-1)$, $(\frac{1}{2},q-\frac{3}{2})$, $(\frac{3}{2},q-\frac{1}{2})$, and $(\frac{3}{2},q-1)$.

E1. $p = 0, q > 1$. These representations, finally, reduce to the su(2) ⊕ su(2) irreps $(0,q)$, $(1,q-\frac{1}{2})$, $(1,q-1)$, and $(0,q-\frac{3}{2})$.

These situations complete the classification of the finite-dimensional osp(3,2) irreps. It is obvious that all these cases are "parts" of the general eightfold pattern, and it is worthwhile remarking that for all these cases the corresponding parts of (5.12)–(5.24) are still valid. For instance, if we consider the representations B1, the expressions (5.2) and (5.14) which correspond to the only possible ways by which the su(2) ⊕ su(2) irreps may be connected, are still correct. Also note that (5.13), (5.15), (5.16), (5.20), and (5.21) would vanish in the case B1, which shows again that the B1 series reduce to only three su(2) ⊕ su(2) irreps.

We still have to investigate the infinite-dimensional osp(3,2) irreps in the case $|p| \leqslant \frac{3}{2}$ or $|q| \leqslant \frac{3}{2}$. These representa-

tions reduce to infinite-dimensional su(2) ⊕ su(2) irreps. Whether we have finite- or infinite-dimensional subalgebra representations depends in fact on the Hermiticity conditions we choose for the subalgebra generators. If we take the usual conditions

$$s_0^\dagger = s_0, \quad s_\pm^\dagger = s_\mp, \quad t_0^\dagger = t_0, \quad t_\pm^\dagger = t_\mp, \tag{5.25}$$

then this corresponds to the real compact forms of the two su(2) algebras, and we find only finite-dimensional representations for which $s = 0, \frac{1}{2}, 1, \ldots$ and $t = 0, \frac{1}{2}, 1, \ldots$. If we impose the Hermiticity conditions

$$s_0^\dagger = s_0, \quad s_\pm^\dagger = -s_\mp, \quad t_0^\dagger = t_0, \quad t_\pm^\dagger = t_\mp, \tag{5.26}$$

then the first su(2) generators are a basis for the noncompact form of su(2), while the second su(2) algebra is compact. The unitary representations of this algebra, which we denote by su(1,1) ⊕ su(2), are infinite-dimensional with respect to the first su(2)-subalgebra, and finite-dimensional with respect to the second. This means that the label s can take on negative real values (or, in the general case, also complex values $-\frac{1}{2} + i\rho$, where $\rho \in \mathbb{R}$), but for t we still have: $2t \in \mathbb{N}$. The two remaining possibilities are denoted by the obvious notations su(2) ⊕ su(1,1) and finally su(1,1) ⊕ su(1,1).

Before continuing the analysis of the infinite-dimensional osp(3,2) representations, we have to make an important remark. It is known that there exists a relation between the matrix elements of the shift operators and the reduced matrix elements of the tensor.[6] Such a relation reads

$$O^{i,j,j}_{s,t,n}|s,m,t,n\rangle \sim \langle s+i, t+j \| R^{[1,1/2]} \| s,t\rangle |s,m,t,n\rangle, \tag{5.27}$$

where the actual value of the coefficient is a well-known expression in s, m, t, n, i, and j. Once a Hermitian operation is chosen for the Lie superalgebra (which is a so-called star or grade star operation), the matrix elements of the shift operators can be determined from the expressions of the matrix elements of the product operators (5.12)–(5.24). We will do this explicitly in Sec. VII [see Eq. (7.12)]. Then the reduced matrix elements of $R^{[1,1/2]}$ are known, and from the Wigner–Eckart theorem[12] all the actions of its components are obtained:

$$\begin{aligned} R_{\alpha,\beta}|s,m,t,n\rangle = \sum_{s',m'} \sum_{t',n'} (-1)^{s'-m'} \begin{pmatrix} s' & 1 & s \\ -m' & \alpha & m \end{pmatrix} \\ \times (-1)^{t'-n'} \begin{pmatrix} t' & \frac{1}{2} & t \\ -n' & \beta & n \end{pmatrix} \\ \times \langle s',t' \| R^{[1,1/2]} \| s,t\rangle |s',m',t',n'\rangle, \end{aligned} \tag{5.28}$$

where the symbols on the right-hand side are Wigner $3j$-symbols with one j equal to 1 or $\frac{1}{2}$, for which explicit expressions are given in the literature.[12] In the finite-dimensional case this procedure does not give rise to any problems. In the infinite-dimensional case, the Wigner $3j$-symbols are "analytical continuations" of the expressions for integral or half-integral j's.[13] Then problems might occur because of the appearance of factors like $(2s+1)$ or $(2t+1)$ in the coefficient for (5.27) or in the denominators of the $3j$-symbols. Sometimes, these difficulties are solved because the same coefficients appear in the expression of the shift operator matrix element (which contributes finally to the numerator), and hence the singularities are dissolved by "taking the limit."

A detailed study of the coefficients in (5.27) and the $3j$-symbols in (5.28) finally showed that we have to exclude the infinite-dimensional osp(3,2) irreps $(p;q)$ for which the $\mathrm{su}(1,1)_s$ label p is $0, -\frac{1}{2}$ or -1, or for which the $\mathrm{su}(1,1)_t$ label q is 0 or $\frac{1}{2}$, because in these cases singularities appear which cannot be removed.

We can now give a summary of the remaining infinite-dimensional osp(3,2) representations, where $|p| \leqslant \frac{3}{2}$ or $|q| \leqslant \frac{3}{2}$.

(1) su(1,1) ⊕ su(2). It is sufficient to consider $p < 0$, since an su(1,1) representation labeled by s is equivalent to the representation labeled by $-s-1$.

For $p < -1$, $-1 < p < -\frac{1}{2}$ or $-\frac{1}{2} < p < 0$, and $q \geqslant \frac{3}{2} (q \in \frac{1}{2}\mathbb{N})$, we still have the usual eightfold decomposition diagram, except when $p + 2q = 0$, which gives the fourfold decomposition pattern existing of (p,q), $(p,q-\frac{1}{2})$, $(p-1,q-\frac{1}{2})$, and $(p-1,q-1)$.

For $p < 0$ (and $p \neq -1, -\frac{1}{2}$), and $q = 1$ we have truncated diagrams, decomposing into $(p,1)$, $(p,\frac{1}{2})$, $(p \pm 1,\frac{1}{2})$, $(p,0)$, and $(p \pm 1,0)$, except for $p = -2$, which gives again a fourfold diagram decomposing into $(-2,1)$, $(-2,\frac{1}{2})$, $(-3,\frac{1}{2})$, and $(-3,0)$.

For $p < 0$ $(\neq -1, -\frac{1}{2})$ and $q = \frac{1}{2}$, the representations reduce into the su(1,1) ⊕ su(2) irreps $(p,\frac{1}{2})$, $(p-1,0)$, and $(p+1,0)$. This completes the analysis for su(1,1) ⊕ su(2).

(2) su(2) ⊕ su(1,1). Now $2p \in \mathbb{N}$, and because of the equivalence for $t \to -t-1$ we have only to consider q-values which satisfy $q < \frac{1}{2}$.

For $p = 1$ or $\frac{3}{2}$, and $q < \frac{1}{2} (q \neq 0)$, the representation decomposes into the usual eightfold pattern, except when $p + 2q = 0$. Then $(p;q) = (\frac{3}{2}; -\frac{3}{4})$ decomposes into $(\frac{3}{2}, -\frac{3}{4})$, $(\frac{3}{2}, -\frac{5}{4})$, $(\frac{1}{2}, -\frac{5}{4})$, and $(\frac{1}{2}, -\frac{7}{4})$, and $(p;q) = (1; -\frac{1}{2})$ into $(1, -\frac{1}{2})$, $(1, -1)$, $(0, -1)$, and $(0, -\frac{3}{2})$.

For $p = \frac{1}{2}$ and $q < \frac{1}{2} (q \neq 0)$, the $(p;q)$ irrep reduces to the su(2) ⊕ su(1,1) representations $(\frac{1}{2},q)$, $(\frac{1}{2},q-\frac{1}{2})$, $(\frac{1}{2},q-1)$, $(\frac{1}{2},q-\frac{3}{2})$, $(\frac{3}{2},q-\frac{1}{2})$, and $(\frac{3}{2},q-1)$. There is only one exception to this, i.e., when $p + 2q = 0$. This is the representation $(\frac{1}{2}; -\frac{1}{4})$, and it decomposes into a doublet of subalgebra irreps, namely $(\frac{1}{2}, -\frac{1}{4})$ and $(\frac{1}{2}, -\frac{3}{4})$.

Finally, for $p = 0$ and $q < \frac{1}{2}$ $(q \neq 0)$, the $(p;q)$ irrep decomposes into the su(2) ⊕ su(1,1) irreps $(0,q)$, $(1,q-\frac{1}{2})$, $(1,q-1)$, and $(0,q-\frac{3}{2})$.

(3) su(1,1) ⊕ su(1,1). Because of the afore-mentioned symmetry, we can restrict ourselves to the values $p < 0$ and $q < \frac{1}{2}$. For all these values $(p \neq -1, -\frac{1}{2}, q \neq 0)$ the osp(3,2) irreps $(p;q)$ decompose into the usual set of eight su(1,1) ⊕ su(1,1) representations, except when $p - 2q + 1 = 0$ or $p + 2q = 0$. If $p - 2q + 1 = 0$, the decomposition contains only (p,q), $(p,q-\frac{1}{2})$, $(p+1,q-\frac{1}{2})$, and $(p+1,q-1)$, and if $p + 2q = 0$, it contains (p,q), $(p,q-\frac{1}{2})$, $(p-1,q-\frac{1}{2})$, and $(p-1,q-1)$. Note that $p + 2q = 0$ and $p - 2q + 1 = 0$ cannot occur simultaneously, because we had to exclude the possibility $p = -\frac{1}{2}$.

This completes the analysis of osp(3,2) irreducible representations. We would like to remark that the expressions (5.12)–(5.24) are still valid for the above-mentioned representations if we first use them to obtain expressions like (5.28), and then "take the limit." In Sec. IV we will give an example of this for the osp(3,2) irrep $(\frac{1}{2}; -\frac{1}{4})$.

VI. STAR AND GRADE STAR REPRESENTATIONS

Star and grade star operations for Lie superalgebras are the equivalents of Hermitian operations for Lie algebras, and

have been discussed in general by Scheunert *et al.*[8] If $L = L_{\bar{0}} \oplus L_{\bar{1}}$ is a Lie superalgebra, then the operation $^{\dagger}: L_\alpha \to L_\alpha$ ($\alpha = \bar{0}, \bar{1}$) is a star operation if

$$\begin{aligned} &(aA + bB)^{\dagger} = a^{*}A^{\dagger} + b^{*}B^{\dagger}, \\ &[A,B]^{\dagger} = [B^{\dagger},A^{\dagger}], \\ &(A^{\dagger})^{\dagger} = A, \end{aligned} \tag{6.1}$$

for all elements A, B of L and for all complex numbers a,b (the notation $*$ denotes the complex conjugate). The operation $^{\ddagger}: L_\alpha \to L_\alpha$ ($\alpha = \bar{0}, \bar{1}$) is a grade star operation if

$$\begin{aligned} &(aA + bB)^{\ddagger} = a^{*}A^{\ddagger} + b^{*}B^{\ddagger}, \\ &[A,B]^{\ddagger} = (-1)^{\alpha\beta}[B^{\ddagger},A^{\ddagger}], \\ &(A^{\ddagger})^{\ddagger} = (-1)^{\alpha}A, \end{aligned} \tag{6.2}$$

for all homogeneous elements A, B of L and for all complex numbers a,b. In (6.2) α (resp., β) is the degree of A (resp., B). The definitions (6.1) and (6.2) imply that the restriction of a star or grade star operation to the even part $L_{\bar{0}}$ is a Hermitian operation of the Lie algebra $L_{\bar{0}}$. Hence, in the case of osp(3,2) we have only to consider all possible Hermitian operations on the even part, and investigate whether it is possible to extend them to a star or grade star operation for the Lie superalgebra. In Sec. V we have already mentioned four independent Hermitian operations on the Lie algebra su(2) ⊕ su(2). For each of them we give the possible extensions:

(1) su(2) ⊗ su(2). No star operation can be defined for the Lie superalgebra, when the adjoint operation on the even part is as in (5.25). However, two grade star operations are consistent with (5.25), and we find

$$\begin{aligned} &s_0^{\ddagger} = s_0, \quad s_{\pm}^{\ddagger} = s_{\mp}, \quad t_0^{\ddagger} = t_0, \quad t_{\pm}^{\ddagger} = t_{\mp}, \\ &R_{\lambda,\mu}^{\ddagger} = (-1)^{\lambda+\mu+1/2}\epsilon R_{-\lambda,-\mu}, \end{aligned} \tag{6.3}$$

where ϵ is $+1$ or -1.

(2) su(1,1) ⊕ su(2). Again, no star operation can be given if the adjoint operation is as in (5.26), but the following two grade star operations satisfy the definition (6.2):

$$\begin{aligned} &s_0^{\ddagger} = s_0, \quad s_{\pm}^{\ddagger} = -s_{\mp}, \quad t_0^{\ddagger} = t_0, \quad t_{\pm}^{\ddagger} = t_{\mp}, \\ &R_{\lambda,\mu}^{\ddagger} = (-1)^{\mu+1/2}\epsilon R_{-\lambda,-\mu}, \end{aligned} \tag{6.4}$$

where $\epsilon \in \{+1, -1\}$.

(3) su(2) ⊕ su(1,1). If the adjoint operation on the even part is the one corresponding to su(2) ⊕ su(1,1), the situation is opposite to the previous ones, and only two star operations can be defined:

$$\begin{aligned} &s_0^{\dagger} = s_0, \quad s_{\pm}^{\dagger} = s_{\mp}, \quad t_0^{\dagger} = t_0, \quad t_{\pm}^{\dagger} = -t_{\mp}, \\ &R_{\lambda,\mu}^{\dagger} = (-1)^{\lambda+1}\epsilon R_{-\lambda,-\mu} \quad (\epsilon = +1 \text{ or } -1). \end{aligned} \tag{6.5}$$

(4) su(1,1) ⊕ su(1,1). The operations consistent with the adjoint operation on the even part are two star operations

$$\begin{aligned} &s_0^{\dagger} = s_0, \quad s_{\pm}^{\dagger} = -s_{\mp}, \quad t_0^{\dagger} = t_0, \quad t_{\pm}^{\dagger} = -t_{\mp}, \\ &R_{\lambda,\mu}^{\dagger} = -\epsilon R_{-\lambda,-\mu} \quad (\epsilon = +1 \text{ or } -1). \end{aligned} \tag{6.6}$$

Let ρ be a representation of the Lie superalgebra L into a graded representation space $V = V_{\bar{0}} \oplus V_{\bar{1}}$. Suppose that a nondegenerate Hermitian form $\langle | \rangle$ on V is given such that

$$\langle V_{\bar{0}} | V_{\bar{1}} \rangle = \{0\}. \tag{6.7}$$

If the nondegenerate Hermitian form $\langle | \rangle$ on V is positive definite, then V is called a graded Hilbert space.[8] Let $pl(V)$ be the space of linear mappings of V into itself. It is known that also $pl(V)$ is a graded vector space.[1] For every homogeneous element A of $pl(V)$, the adjoint operator $A^{\dagger}$, respectively, grade adjoint operator $A^{\ddagger}$, is defined by[8]

$$\langle A^{\dagger}x | y \rangle = \langle x | Ay \rangle, \quad \forall x,y \in V, \tag{6.8}$$

resp.,

$$\langle A^{\ddagger}x | y \rangle = (-1)^{\alpha\xi}\langle x | Ay \rangle, \quad \forall x \in V_{\xi}, \quad \forall y \in V, \tag{6.9}$$

where α = degree (A). Then, the representation $\rho: L \to V$ is a star, resp., a grade star, representation if

$$\rho(A^{\dagger}) = (\rho(A))^{\dagger}, \quad \forall A \in L_\alpha \quad (\alpha = \bar{0}, \bar{1}), \tag{6.10}$$

resp.,

$$\rho(A^{\ddagger}) = (\rho(A))^{\ddagger}, \quad \forall A \in L_\alpha \quad (\alpha = \bar{0}, \bar{1}). \tag{6.11}$$

In the following sections we shall investigate which of the representations, considered in Sec. V, are star or grade star representations. In particular we are interested for which of the irreps the representation space is a graded Hilbert space.

VII. FINITE-DIMENSIONAL GRADE STAR REPRESENTATIONS OF osp(3,2)

The only possible adjoint operations are given in (6.3). We first consider the general $(p;q)$ irreps with $p \geqslant 1$ and $q \geqslant \frac{3}{2}$, and $p - 2q + 1 \neq 0$, which reduce in su(2) ⊕ su(2) irreps (s,t), where

$$\begin{aligned} (s,t) \in I = \{&(p,q),(p,q-\tfrac{1}{2}),(p \pm 1,q-\tfrac{1}{2}), \\ &(p,q-1),(p \pm 1,q-1),(p,q-\tfrac{3}{2})\}. \end{aligned} \tag{7.1}$$

The representation space V is then spanned by

$$\begin{aligned} \{ |s,m,t,n\rangle \,|\, &(s,t) \in I; m = -s, -s+1, \ldots s; \\ &n = -t, -t+1, \ldots, +t \}. \end{aligned} \tag{7.2}$$

The adjoint operation on the even part of the superalgebra implies that the nondegenerate Hermitian form on V is of the form

$$\langle s',m',t',n' | s,m,t,n \rangle = g(s,t)\delta_{s',s}\delta_{t',t}\delta_{m',m}\delta_{n',n}, \tag{7.3}$$

where $g(s,t) \in \mathbb{R}$. By appropriate rescaling of the basis states, the Hermitian form can be chosen such that for all $(s,t) \in I$ the factor $g(s,t)$ satisfies $|g(s,t)| = 1$, hence $g(s,t) = \pm 1$. *If* $g(s,t) = 1$ for all (s,t), then V is a graded Hilbert space.

Using some general properties of the shift operators[6] and making use of (6.3), we can show that the operators (4.2)–(4.7) satisfy

$$\begin{aligned} &(O^{1,\pm 1/2,\pm 1/2})^{\ddagger}(2\hat{s}+1)(2\hat{t}+1) \\ &\quad = \pm\epsilon(O^{-1,\mp 1/2,\mp 1/2})(2\hat{s}-1)(2\hat{t}+1\mp 1), \\ &(O^{0,\pm 1/2,\pm 1/2})^{\ddagger}(2\hat{t}+1) \\ &\quad = \pm\epsilon(O^{0,\mp 1/2,\mp 1/2})(2\hat{t}+1\mp 1), \\ &(O^{-1,\pm 1/2,\pm 1/2})^{\ddagger}(2\hat{s}+1)(2\hat{t}+1) \\ &\quad = \pm\epsilon(O^{1,\mp 1/2,\mp 1/2})(2\hat{s}+3)(2\hat{t}+1\mp 1). \end{aligned} \tag{7.4}$$

We did not write the indices s,t,n for the shift operators, because they are supposed to be expressed in terms of the osp(3,2) generators $\hat{s}$ and $\hat{t}$ [i.e., in the expressions (4.2)–(4.7), s,m,t, and n are replaced by $\hat{s}$, s_0, $\hat{t}$, and t_0, respectively], and hence they can act on any state of V. The first relation in (7.4)

implies:

$$\langle s-1,t\mp\tfrac{1}{2}|A^{-1,\mp 1/2}_{s,t}|s,t\rangle(2s-1)(2t+1\mp 1)$$
$$=\pm\epsilon(2s+1)(2t+1)\langle s-1,t\mp\tfrac{1}{2}|(A^{1,\pm 1/2}_{s-1,t\mp 1/2})^{\dagger}|s,t\rangle, \tag{7.5}$$

and similar expressions are obtained for the rest of the relations in (7.4). Since the shift operators are represented by odd operators, we deduce by means of relations like (7.5)

$$\langle s,t|A^{\mp 1,\mp 1/2}_{s\pm 1,t\pm 1/2}A^{\pm 1,\pm 1/2}_{s,t}|s,t\rangle$$
$$=\mp\epsilon\frac{(2s+1\pm 2)(2t+1\pm 1)}{(2s+1)(2t+1)}(-1)^{\sigma}g(s\pm 1,t\pm\tfrac{1}{2})$$
$$\times|\langle s\pm 1,t\pm\tfrac{1}{2}|A^{\pm 1,\pm 1/2}_{s,t}|s,t\rangle|^2, \tag{7.6}$$

$$\langle s,t|A^{0,\mp 1/2}_{s,t\pm 1/2}A^{0,\pm 1/2}_{s,t}|s,t\rangle$$
$$=\mp\epsilon\frac{(2t+1\pm 1)}{(2t+1)}(-1)^{\sigma}g(s,t\pm\tfrac{1}{2})$$
$$\times|\langle s,t\pm\tfrac{1}{2}|A^{0,1/2}_{s,t}|s,t\rangle|^2, \tag{7.7}$$

$$\langle s,t|A^{\pm 1,\mp 1/2}_{s\mp 1,t\pm 1/2}A^{\mp 1,\pm 1/2}_{s,t}|s,t\rangle$$
$$=\mp\epsilon\frac{(2s+1\mp 2)(2t+1\pm 1)}{(2s+1)(2t+1)}(-1)^{\sigma}$$
$$\times g(s\mp 1,t\pm\tfrac{1}{2})|\langle s\mp 1,t\pm\tfrac{1}{2}|A^{\mp 1,\pm 1/2}_{s,t}|s,t\rangle|^2, \tag{7.8}$$

where σ is the degree of state $|s,t\rangle$ [sometimes denoted as $\sigma(s,t)$].

Equations (7.6)–(7.8) are applied on the irreps $(p;q)$ under consideration. We deduce from (5.12) and (7.8) that

$$2(p+1)(2q+1)(p+2q)g(p,q)$$
$$=(-1)^{\sigma(p,q)}\epsilon(p+1,q-\tfrac{1}{2})$$
$$\times(2q)|\langle p+1,q-\tfrac{1}{2}|A^{1,-1/2}_{p,q}|p,q\rangle|^2/(2p+1)(2q+1). \tag{7.9}$$

We first consider the operation with $\epsilon=+1$. For the degree of the states $|p,q\rangle$ there are two possibilities:

$$\text{(a)}\quad \sigma(p,q)=\bar{0},$$
$$\text{(b)}\quad \sigma(p,q)=\bar{1}, \tag{7.10}$$

If we are in case (a) [resp., (b)], Eq. (7.9) implies

$$\text{(a)}\quad g(p+1,q-\tfrac{1}{2})=g(p,q),$$

resp., (7.11)

$$\text{(b)}\quad g(p+1,q-\tfrac{1}{2})=-g(p,q),$$

and

$$|\langle p+1,q-\tfrac{1}{2}|A^{1,-1/2}_{p,q}|p,q\rangle|^2$$
$$=2(p+1)(2p+1)\frac{(2q+1)^2}{2q}(p+2q). \tag{7.12}$$

From (5.14) and (7.6), we obtain

$$-2p(2q+1)(p-2q+1)g(p,q)$$
$$=(-1)^{\sigma(p,q)}g(p-1,q-\tfrac{1}{2})(2q)$$
$$\times|\langle p-1,q-1|A^{-1,-1/2}_{p,q}|p,q\rangle|^2/(2p+1)(2q+1). \tag{7.13}$$

This shows that the sign of $p-2q+1$ plays a significant role. If $p-2q+1<0$, we have [(a) and (b) are referring to (7.10)]

$$\text{(a)}\quad g(p-1,q-\tfrac{1}{2})=g(p,q).$$
$$\text{(b)}\quad g(p-1,q-\tfrac{1}{2})=-g(p,q), \tag{7.14}$$

whereas for $p-2q+1>0$, we obtain

$$\text{(a)}\quad g(p-1,q-\tfrac{1}{2})=-g(p,q),$$
$$\text{(b)}\quad g(p-1,q-\tfrac{1}{2})=g(p,q). \tag{7.15}$$

In this way, the whole pattern is analyzed. We do not give the details of the calculations, but summarize the results for $\epsilon=+1$ in Table I. For the "highest states" $|p,q\rangle$ we have chosen, without loss of generality, $g(p,q)=+1$. For $\epsilon=-1$, we obtain the same table where all $\bar{0}$ and $\bar{1}$ are interchanged. The remaining finite-dimensional osp(3,2) irreps, i.e., the irreps $(p;q)$ for which $p-2q+1=0$, $p\leqslant\frac{1}{2}$ or $q\leqslant 1$, are known to be "parts" of the general eightfold pattern, and a detailed investigation shows that for such representations we can just copy the corresponding parts of Table I.

The main conclusion is the following: For none of the general osp(3,2) irreps $(p;q)$, which reduce in the usual eight su(2) ⊕ su(2) irreps, is the representation space a graded Hilbert space. There always exists, however, a nondegenerate Hermitian form on *V which is not positive definite*, such that all the irreps $(p;q)$ are grade star representations.

A closer look at the "truncated" representations A1–E1, considered in Sec. V, and their corresponding parts in Table I, shows that there are only two exceptions, namely the five-dimensional irrep $(0;\frac{1}{2})$ and the eight-dimensional irrep $(\frac{1}{2};\frac{1}{2})$ (the situations B3 and B2, respectively). For $\epsilon=\pm 1$, the first representation is spanned by the even states $|0,0,\frac{1}{2},\pm\frac{1}{2}\rangle$ and the odd states $|1,m,0,0\rangle$ $(m=0,\pm 1)$. The second representation is spanned by the even states

TABLE I. Degrees of the states $|s,t\rangle$ and of the factor $g(s,t)$ in (7.3).

irreps (s,t)	$p-2q+1<0$		$p-2q+1<0$		$p-2q+1>0$		$p-2q+1>0$	
of $(p;q)$	deg$\lvert s,t\rangle$	$g(s,t)$	deg$\lvert s,t\rangle$	$g(s,t)$	deg$\lvert s,t\rangle$	$g(s,t)$	deg$\lvert s,t\rangle$	$g(s,t)$
(p,q)	$\bar{0}$	$+1$	$\bar{1}$	$+1$	$\bar{0}$	$+1$	$\bar{1}$	$+1$
$(p-1,q-\frac{1}{2})$	$\bar{1}$	$+1$	$\bar{0}$	-1	$\bar{1}$	-1	$\bar{0}$	$+1$
$(p,q-\frac{1}{2})$	$\bar{1}$	$+1$	$\bar{0}$	-1	$\bar{1}$	$+1$	$\bar{0}$	-1
$(p+1,q-\frac{1}{2})$	$\bar{1}$	$+1$	$\bar{0}$	-1	$\bar{1}$	$+1$	$\bar{0}$	-1
$(p-1,q-1)$	$\bar{0}$	-1	$\bar{1}$	-1	$\bar{0}$	$+1$	$\bar{1}$	$+1$
$(p,q-1)$	$\bar{0}$	-1	$\bar{1}$	-1	$\bar{0}$	$+1$	$\bar{1}$	$+1$
$(p+1,q-1)$	$\bar{0}$	-1	$\bar{1}$	-1	$\bar{0}$	-1	$\bar{1}$	-1
$(p,q-\frac{3}{2})$	$\bar{1}$	-1	$\bar{0}$	$+1$	$\bar{1}$	$+1$	$\bar{0}$	-1

$|\frac{1}{2},m,\frac{1}{2},n\rangle$ $(m,n=\pm\frac{1}{2})$ and the odd states $|\frac{3}{2},m,0,0\rangle$ $(m=\pm\frac{3}{2},\pm\frac{1}{2})$. It is only for these two cases that the Hermitian form of the grade star representation is positive definite.[14]

It might seem rather strange that even the adjoint representation is not a grade star representation if the Hermitian form is required to be positive definite, although it is possible to define a grade star operation for the Lie superalgebra. Therefore, we consider this in a second way. From Sec. VI we deduce that a grade star representation ρ satisfies

$$\langle\rho(A^{\ddagger})x|y\rangle=(-1)^{\alpha\xi}\langle x|\rho(A)y\rangle, \tag{7.16}$$

for $A\in L_\alpha$, $x\in V_\xi$, $y\in V$. Let us apply this for $\rho=ad$. Then the space V is osp(3,2) itself, and we can choose the basis states of V proportional to the generators of Sec. II. For $A=R_{1,1/2}$, $x=R_{0,1/2}$, and $y=s_-$, (7.16) becomes

$$\epsilon/\sqrt{2}\langle s_-|s_-\rangle=\sqrt{2}\langle R_{0,1/2}|R_{0,1/2}\rangle; \tag{7.17}$$

for $A=R_{1,1/2}$, $x=t_+$ and $y=R_{-1,1/2}$, we obtain

$$\epsilon\langle R_{-1,1/2}|R_{-1,1/2}\rangle=-2\langle t_+|t_+\rangle. \tag{7.18}$$

These two relations show that, whatever choice of ϵ we make in (6.3), the grade star representation is not consistent with a positive definite Hermitian form.

The main reason why the general grade star representations are not consistent with a positive definite Hermitian form is the following. For grade star representations we have the rule that a minus sign is placed "whenever two odd objects are interchanged." But a Hermitian form on V satisfies

$$\langle y|x\rangle^*=\langle x|y\rangle,$$

for all x and y, and hence does not distinguish between "interchanging" even or odd states.

Finally, we give the relations between matrix elements of the shift operators and reduced matrix elements of the tensor $R^{[1,1/2]}$. They follow from the general analysis of shift operators[6]):

$$\begin{aligned}
&\langle s+1,t\pm\tfrac{1}{2}\|R^{[1,1/2]}\|s,t\rangle\\
&\quad=\left[\frac{(2s+3)}{(2s+2)(2s+1)}\frac{(2t+1\pm1)}{(2t+1)}\right]^{1/2}g(s+1,t\pm\tfrac{1}{2})\\
&\qquad\times\langle s+1,t\pm\tfrac{1}{2}|A^{1,\pm1/2}_{s,t}|s,t\rangle,\\
&\langle s,t\pm\tfrac{1}{2}\|R^{[1,1/2]}\|s,t\rangle\\
&\quad=\left[\frac{2(2s+1)}{(2s+2)(2s)}\frac{(2t+1\pm1)}{(2t+1)}\right]^{1/2}g(s,t\pm\tfrac{1}{2})\\
&\qquad\times\langle s,t\pm\tfrac{1}{2}|A^{0,\pm1/2}_{s,t}|s,t\rangle,\\
&\langle s-1,t\pm\tfrac{1}{2}\|R^{[1,1/2]}\|s,t\rangle\\
&\quad=-\left[\frac{(2s-1)}{(2s)(2s+1)}\frac{(2t+1\pm1)}{(2t+1)}\right]^{1/2}g(s-1,t\pm\tfrac{1}{2})\\
&\qquad\times\langle s-1,t\pm\tfrac{1}{2}|A^{-1,\pm1/2}_{s,t}|s,t\rangle.
\end{aligned} \tag{7.19}$$

Now all the necessary elements are given to deduce explicit expressions for the actions of all the generators of osp(3,2) upon the basis states of an irrep $(p;q)$. Indeed, (5.12)–(5.24) give the matrix elements of the shift operator products, from which the absolute values of the shift operator matrix elements are obtained by means of (7.6)–(7.8). The actual choice of the phase factor $(+1$ or $-1)$ for the shift operator matrix elements is determined by the nonscalar relations (4.17)–(4.26). Then (7.19) gives the expressions for all the reduced matrix elements of $R^{[1,1/2]}$, and finally the actions of the tensor components are obtained from (5.28). The actions of s_0, $s_\pm$, t_0, and $t_\pm$ are, of course, no problem. The only reason why we do not write down the explicit expressions for the actions $R_{\alpha,\beta}|s,m,t,n\rangle$, is that this would be a tedious summing up of 48 formulas, which the reader can easily obtain himself by following the above-mentioned procedure.

VIII. INFINITE-DIMENSIONAL STAR AND GRADE STAR REPRESENTATIONS OF osp(3,2)

Before analyzing the star and grade star conditions on the infinite-dimensional osp(3,2) irreps, we give a brief view of infinite-dimensional unitary representations of su(1,1), which is the Lie algebra generated by $j_0,j_\pm$ $([j_0,j_\pm]=\pm j_\pm,[j_+,j_-]=2j_0)$ whose basis elements satisfy the adjoint condition $j_0^\dagger=j_0$ and $j_\pm^\dagger=-j_\mp$. Those representations are grouped into four series (see Ref. 13, p. 182–188):

(a) The principal series D^P, where $j=-1/2+i\rho$ $(\rho\in\mathbb{R})$, and m takes on an infinity of values. Here, of course, $j(j+1)$ is the eigenvalue of the Casimir J^2, and m is the j_0-eigenvalue. No principal series will occur for osp(3,2), since if $j=s$ (resp., $j=t$) is an irrep, then also $s\pm1$ (resp., $t\pm\frac{1}{2}$) are irreps appearing in the osp(3,2) representation, which would give complex numbers that cannot characterize su(1,1) irreps.

(b) The continuous series D^S, where $-1<j<0$, and $m\in\{-\infty,\ldots,m_0-1,m_0,m_0+1,\ldots,+\infty\}$. Here, $m_0\in\mathbb{R}$ satisfies $-j-1<m_0<j+1$ if $-1<j\leqslant-\frac{1}{2}$, and $j<m_0<-j$ if $-\frac{1}{2}\leqslant j<0$.

(c) The discrete positive series D^+, where $j\in\mathbb{R}$ and m is bounded from below: $m\in\{\underline{m},\underline{m}+1,\underline{m}+2,\ldots,+\infty\}$. If $j\geqslant0$, then $\underline{m}=j+1$, if $j\leqslant-1$ then $\underline{m}=-j$, and if $-1<j<0$ we have two possibilities: $\underline{m}=j+1$ or $\underline{m}=-j$.

(d) The discrete negative series D^-, where $j\in\mathbb{R}$ and m is bounded from above: $m\in\{-\infty,\ldots,\overline{m}-2,\overline{m}-1,\overline{m}\}$, if $j>0$ then $\overline{m}=-j-1$, if $j\leqslant-1$ then $\overline{m}=j$, and if $-1<j<0$ we have again two possibilities $=\overline{m}=-j-1$ or $\overline{m}=+j$. The basis states of an su(1,1) irrep are the simultaneous eigenstates of J^2 and j_0, and therefore they are rather characterized by $j(j+1)$ and m instead of (j,m). This shows that we have the symmetry $j\to-j-1$, and consequently we can restrict ourselves to representations for which $j\leqslant-\frac{1}{2}$.

For the infinite-dimensional cases, we shall analyze only the representations for which the representation space is a graded Hilbert space. Hence, the Hermitian form on V is positive definite, and we may assume

$$\langle s,m,t,n|s',m',t',n'\rangle=\delta_{ss'}\delta_{mm'}\delta_{tt'}\delta_{nn'}. \tag{8.1}$$

We shall investigate whether the star (resp., grade star) operations are consistent with this Hermitian form.

We use the shift operators $O^{i,j,j}_{s,t,n}$ rather than the normalized operators $A^{i,j}_{s,t}$ because the internal structure of infinite-dimensional su(2) irreps is somewhat more complicated.

A. The case su(1,1) ⊗ su(2)

The grade star operation (6.4) implies that the Hermiticity properties for the shift operators are finally the same as in (7.4). From this we obtain

$$(2s+1)(2t+1)\langle s,m,t,n|O^{\mp 1,\mp 1/2,\mp 1/2}_{s\pm 1,t\pm 1/2,n\pm 1/2}O^{\pm 1,\pm 1/2,\pm 1/2}_{s,t,n}|s,m,t,n\rangle$$
$$= \mp\epsilon(-1)^{\sigma}(2s+1\pm 2)(2t+1\pm 1)|\langle s\pm 1,m,t\pm\tfrac{1}{2},n\pm\tfrac{1}{2}|O^{\pm 1,\pm 1/2,\pm 1/2}_{s,t,n}|s,m,t,n\rangle|^2, \tag{8.2}$$

$$(2t+1)\langle s,m,t,n|O^{0,\mp 1/2,\mp 1/2}_{s,t\pm 1/2,n\pm 1/2}O^{0,\pm 1/2,\pm 1/2}_{s,t,n}|s,m,t,n\rangle$$
$$= \mp\epsilon(-1)^{\sigma}(2t+1\pm 1)|\langle s,m,t\pm\tfrac{1}{2},n\pm\tfrac{1}{2}|O^{0,\pm 1/2,\pm 1/2}_{s,t,n}|s,m,t,n\rangle|^2, \tag{8.3}$$

$$(2s+1)(2t+1)\langle s,m,t,n|O^{\pm 1,\mp 1/2,\mp 1/2}_{s\mp 1,t\pm 1/2,n\pm 1/2}O^{\mp 1,\pm 1/2,\pm 1/2}_{s,t,n}|s,m,t,n\rangle$$
$$= \mp\epsilon(-1)^{\sigma}(2s+1\mp 2)(2t+1\pm 1)|\langle s\mp 1,m,t\pm\tfrac{1}{2},n\pm\tfrac{1}{2}|O^{\mp 1,\pm 1/2,\pm 1/2}_{s,t,n}|s,m,t,n\rangle|^2, \tag{8.4}$$

where $\sigma =$ degree $(|s,t\rangle)$. Let us consider the situation where $\epsilon = -1$. Since in this case $t\in\frac{1}{2}\mathbb{N}$, Eqs. (8.2)–(8.4) imply that

$O^{-1,\mp 1/2,\mp 1/2}_{s+1,t\pm 1/2,n\pm 1/2}O^{1,\pm 1/2,\pm 1/2}_{s,t,n}$ has the sign of $\pm(-1)^{\sigma}$ if $s\leqslant -\frac{3}{2}$ or $s> -\frac{1}{2}$, [resp., of $\mp(-1)^{\sigma}$ if $-\frac{3}{2}\leqslant s< -\frac{1}{2}$]; (8.5)

$O^{0,\mp 1/2,\mp 1/2}_{s,t\pm 1/2,n\pm 1/2}O^{0,\pm 1/2,\pm 1/2}_{s,t,n}$ has the sign of $\pm(-1)^{\sigma}$ for all s; (8.6)

$O^{+1,\mp 1/2,\mp 1/2}_{s-1,t\pm 1/2,n\pm 1/2}O^{-1,\pm 1/2,\pm 1/2}_{s,t,n}$ has the sign of $\pm(1)^{\sigma}$ if $s< -\frac{1}{2}$ or $s\geqslant\frac{1}{2}$, [resp., of $\mp(-1)^{\sigma}$ if $-\frac{1}{2}<s\leqslant\frac{1}{2}$]. (8.7)

We shall also make use of the general property

$$\langle s,m,t,n|O^{-i,-j,-j}_{s+i,t+j,n+j}O^{i,j,j}_{s,t,n}|s,m,t,n\rangle = \langle s+i,m,t+j,n+j|O^{i,j,j}_{s,t,n}O^{-i,-j,-j}_{s+i,t+j,n+j}|s+i,m,t+j,n+j\rangle, \tag{8.8}$$

where $i\in\{0,\pm 1\}$ and $j\in\{\pm\frac{1}{2}\}$.

Let us first consider the osp(3,2) irreps $(p;q)$, where $p< -1$ and $q\geqslant\frac{3}{2}$, which always reduce to eight subalgebra irreps unless $p+2q=0$, in which case they decompose in only four irreps. Equation (5.14) implies

$$\langle p,m,q,n|O^{1,1/2,1/2}_{p-1,q-1/2,n-1/2}O^{-1,-1/2,-1/2}_{p,q,n}|p,m,q,n\rangle = (p-m)(p+m)(q+n)(-2p)(2p-1)(2q+1)(2q+1)(p-2q+1). \tag{8.9}$$

Because of (8.7), this should have the sign of $-(-1)^{\sigma(p,q)}$. If $\sigma(p,q)=\bar{0}$ then (8.9) should be negative, hence $(p-m)(p+m)$ should be negative for all m, which gives two possibilities for $s=p$, namely a D^+ with $\underline{m}=-p$ or a D^- with $\overline{m}=p$. If $\sigma(p,q)=\bar{1}$, then (8.9), or $(p-m)(p+m)$, should be positive for all m, which cannot be satisfied because of the adjoint condition on $su(1,1)_s$. So we must have: $\sigma(p,q)=\bar{0}$. Now Eq. (5.13) gives

$$\langle p,m,q,n|O^{0,1/2,1/2}_{p,q-1/2,n-1/2}O^{0,-1/2,-1/2}_{p,q,n}|p,m,q,n\rangle$$
$$= (q+n)(2p)(p+1)(2q-1)(2q+1), \tag{8.10}$$

which is always positive. But according to (8.6), this should have the sign of $-(-1)^{\sigma(p,q)}$, and this leads to a contradiction. This shows that for the osp(3,2) irreps under consideration the grade star condition is not consistent with the positive definite Hermitian form. A detailed analysis of all the other cases ($-1<p<0$ or $q\leqslant 1$) showed that none of the representations $(p;q)$ are grade star representations for the choice (8.1) of the Hermitian form. The situation $\epsilon=+1$ leads to the same conclusion.

B. The case su(2) ⊗ su(1,1) or Osp(3|2,ℝ)

The star operation (6.5) implies the following Hermiticity properties of the shift operators:

$$(O^{i,j,j})^{\dagger}(2\hat{s}+1)(2\hat{t}+1)$$
$$= +\epsilon(O^{-i,-j,-j})(2\hat{s}+1-2i)(2\hat{t}+1-2j), \tag{8.11}$$

where $i\in\{-1,0,+1\}$, $j\in\{-\frac{1}{2},\frac{1}{2}\}$. Then we obtain

$$(2s+1)(2t+1)\langle s,m,t,n|O^{-i,-j,-j}_{s+i,t+j,n+j}O^{i,j,j}_{s,t,n}|s,m,t,n\rangle$$
$$= +\epsilon(2s+1+2i)(2t+1+2j)|\langle s+i,m,t+j,n+j|O^{i,j,j}_{s,t,n}|s,m,t,n\rangle|^2. \tag{8.12}$$

If we choose $\epsilon=+1$, then (8.12) requires that (if $s\geqslant\frac{1}{2}$):

$O^{-i,-1/2,-1/2}_{s+i,t+1/2,n+1/2}O^{i,1/2,1/2}_{s,t,n}$ is positive for $t< -1$ or $t> -\frac{1}{2}$, and negative for $-1<t< -\frac{1}{2}$, (8.13)

$O^{-i,1/2,1/2}_{s+i,t-1/2,n-1/2}O^{i,-1/2,-1/2}_{s,t,n}$ is positive for $t< -1/2$ or $t>0$, and negative for $-1/2<t<0$. (8.14)

Let us first consider the osp(3,2) irreps for which $p=1,\frac{3}{2},2,\ldots$, and $q< -\frac{1}{2}$. Eq. (5.14) shows

$$\langle p,m,q,n|O^{1,1/2,1/2}_{p-1,q-1/2,n-1/2}O^{-1,-1/2,-1/2}_{p,q,n}|p,q,n\rangle = (p-m)(p+m)(q+n)(-2p)(2p-1)(2q+1)(p-2q+1), \tag{8.15}$$

and according to (8.14) this should be positive. This implies $q+n\geqslant 0$ for all n, and hence we have a D^+ representation with $\underline{n}=-q$. If $-1<q<-\frac{1}{2}$ we have a second possibility, namely a D^+ with $\underline{n}=q+1$. Equations (8.8) and (8.15) imply that also for $(p-1,q-\frac{1}{2})$ and su(1,1) part is a D^+ with $\underline{n}=-q+\frac{1}{2}$. Since $q-\frac{1}{2}<-1$ we have only one possibility, and hence also for (p,q) the situation $\underline{n}=q+1$ must be excluded. Equation (5.13) gives

$$\langle p,m,q,n|O^{0,1/2,1/2}_{p,q-1/2,n-1/2}O^{0,-1/2,-1/2}_{p,q,n}|p,m,q,n\rangle$$
$$=(q+n)(2p)(p+1)(2q-1)(2q+1), \tag{8.16}$$

which shows again that $q+n\geqslant 0$. Because of (8.8) this also implies that for $(p,q-\frac{1}{2})$ the su(1,1) part is a D^+ representation with $\underline{n}=-q+\frac{1}{2}$. A similar investigation of (5.21) and (5.19) shows that also for $(p-1,q-1)$ we have a D^+ representation with $\underline{n}=-q+1$. Hence, if $p+2q=0$, this completes the analysis and we find that the four (s,t) irreps, in which such an osp(3,2) irrep $(p;q)$ decomposes, all consist of a finite-dimensional su(2) irrep s and a discrete positive su(1,1) irrep t for which $\underline{n}=-t$.

If $p+2q\neq 0$, the osp(3,2) irrep reduces to eight (s,t) representations, and we have to continue our analysis step by step. Application of (5.12) gives

$$\langle p,m,q,n|O^{-1,1/2,1/2}_{p+1,q-1/2,n-1/2}O^{1,-1/2,-1/2}_{p,q,n}|p,m,q,n\rangle$$
$$=(p-m+1)(p+m+1)(q+n)$$
$$\times 2(p+1)(2p+3)(2q+1)(p+2q). \tag{8.17}$$

Because of (8.13) this expression must be positive. Since we have a D^+ for $t=q$ in (p,q), $q+n$ is positive, and (8.17) implies that $p+2q$ must be negative. In this case the analysis can be continued, and we find that none of the Eqs. (5.12)–(5.24) leads to a contradiction. We conclude: If $p+2q<0$, the osp(3,2) $(p;q)$ is a star representation decomposing into eight su(2) $\oplus$ su(1,1) irreps (s,t) all consisting of a finite-dimensional su(2) irreps s and a positive discrete representation t with $\underline{n}=-t$. On the other hand, if $p+2q>0$, the star conditions are not consistent with the Hermitian form (8.1). The analysis of the remaining cases for su(2) $\oplus$ su(1,1) shows that the only other representations which are star, are the irreps $(p;q)$, where $p=\frac{1}{2}$ and $q\leqslant -\frac{1}{4}$. If $q<-\frac{1}{4}$ they decompose into the six (s,t) irreps, summarized in Sec. V, all consisting of a finite-dimensional su(2) irrep s and a positive discrete su(1,1) representation t with $\underline{n}=-t$. If $q=-\frac{1}{4}$, the representation reduces to only two subalgebra irreps, namely $(\frac{1}{2},-\frac{1}{4})$ and $(\frac{1}{2},-\frac{3}{4})$. Then we have two possible solutions; a D^+ for $t=-\frac{1}{4}$ with $\underline{n}=\frac{1}{4}$ connected to a D^+ for $t=-\frac{3}{4}$ with $\underline{n}=\frac{3}{4}$, or a D^+ for $t=-\frac{1}{4}$ with $\underline{n}=\frac{3}{4}$ connected to a D^+ for $t=-\frac{3}{4}$ with $\underline{n}=\frac{1}{4}$. Because of the symmetry $t\rightarrow -t-1$, these two solutions are equivalent. This is the case of the "metaplectic representation," which we shall consider in detail in Sec. IX.

For $\epsilon=-1$, we obtain the same selection of osp(3,2) irreps which are star representations. The only difference is that all D^+ representations t with $\underline{n}=-t$ are replaced by D^- representations t with $\bar{n}=t$.

In order to obtain general expressions for the actions of $R_{\alpha,\beta}$ upon the states $|s,m,t,n\rangle$, we can use the method described at the end of Sec. VII, except that we have to take more care of the signs of the factors appearing under the square root, as in (7.19).

C. The case su(1,1) $\oplus$ su(1,1)

The star conditions (6.6) on the Lie superalgebra imply that exactly the same relations (8.11) and (8.12) are valid in this case. The positivity and negativity conditions for the shift operator products follow from (8.12). For instance: $O^{1,1/2,1/2}_{s-1,t-1/2,n-1/2}O^{-1,-1/2,-1/2}_{s,t,n}$ is positive if $(s<-\frac{1}{2}$ or $s>\frac{1}{2})$ and $(t<-\frac{1}{2}$ or $t>0)$ or if $-\frac{1}{2}<s<\frac{1}{2}$ and $-\frac{1}{2}<t<0$ and negative in the other cases. Similar conclusions can be made for the remaining shift operator products. We consider the general osp(3,2) irreps $(p;q)$ with $p<-1$ and $q<-\frac{1}{2}$. The expressions (5.13) and (5.12) show that the following relations are valid:

$$\langle p,m,q,n|O^{0,1/2,1/2}_{p,q-1/2,n-1/2}O^{0,-1/2,-1/2}_{p,q,n}|p,m,q,n\rangle$$
$$=(q+n)(2p)(p+1)(2q-1)(2q+1), \tag{8.18}$$

$$\langle p,m,q,n|O^{-1,1/2,1/2}_{p+1,q-1/2,n-1/2}O^{1,-1/2,-1/2}_{p,q,n}|p,m,q,n\rangle$$
$$=(p+m+1)(p-m+1)(q+n)$$
$$\times 2(p+1)(2p+3)(2q+1)(p+2q). \tag{8.19}$$

From (8.18) it follows that $q+n\geqslant 0$, since $O^{0,1/2,1/2}_{s,t-1/2,n-1/2}O^{0,-1/2,-1/2}_{s,t,n}$ must be positive for $t<-\frac{1}{2}$. The positivity and negativity conditions on $O^{-1,1/2,1/2}_{s+1,t-1/2,n-1/2}O^{1,-1/2,-1/2}_{s,t,n}$ show that it should be positive for $s<-\frac{3}{2}$ (and $t<-\frac{1}{2}$) and negative for $-\frac{3}{2}<s<-1$ (and $t<-\frac{1}{2}$). In both cases, (8.19) implies

$$(p+m+1)(p-m+1)\geqslant 0 \tag{8.20}$$

for all m-values. Obviously, such a condition is only satisfied by a finite-dimensional su(2) representation, and never by an infinite-dimensional su(1,1) irrep. Consequently, the representations are not star representations.

The remaining representations are analyzed in a similar way. The conclusion is that for none of the osp(3,2) irreps $(p;q)$, where p and q are su(1,1) labels, are the star conditions on the Lie algebra consistent with the Hermitian form (8.1).

IX. THE METAPLECTIC REPRESENTATION

In this section we consider the osp(3,2) irrep$(\frac{1}{2};-\frac{1}{4})$. We refer to this irrep as the "metaplectic representation," since in the reduction osp(3,2) $\supset$ su(2) $\otimes$ su(1,1) $\supset$ su(1,1), it decomposes as the sum of two metaplectic representations of su(1,1).[10]

The states of this representation are $|\frac{1}{2},\pm\frac{1}{2},-\frac{1}{4},n\rangle$, where $n=\frac{1}{4},\frac{1}{4}+1,\frac{1}{4}+2,\ldots$, and $|\frac{1}{2},\pm\frac{1}{2},-\frac{3}{4},n\rangle$, where $n=\frac{3}{4},\frac{3}{4}+1,\frac{3}{4}+2,\ldots$. We used the method described at the end of Sec. VII in order to obtain the explicit expressions for the actions of $R_{\alpha\beta}$ upon the basis states. They are given by

$$R_{1,\pm 1/2}|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{1}{4},n\rangle=0, \tag{9.1}$$

$$R_{1,\pm 1/2}|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{4},n\rangle=\pm(1/\sqrt{2})(4n\pm 1)^{1/2}|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{3}{4},n\pm\tfrac{1}{2}\rangle, \tag{9.2}$$

$$R_{0,\pm 1/2}|\tfrac{1}{2},m,-\tfrac{1}{4},n\rangle$$
$$=\mp m(4n\pm 1)^{1/2}|\tfrac{1}{2},m,-\tfrac{3}{4},n\pm\tfrac{1}{2}\rangle \quad (m=\pm\tfrac{1}{2}), \tag{9.3}$$

$$R_{-1,\pm 1/2}|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{1}{4},n\rangle$$
$$=\mp(1/\sqrt{2})(4n\pm 1)^{1/2}|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{3}{4},n\pm\tfrac{1}{2}\rangle, \tag{9.4}$$

$$R_{-1,\pm 1/2}|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{4},n\rangle=0, \tag{9.5}$$

$$R_{1,\pm 1/2}|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{3}{4},n\rangle=0, \tag{9.6}$$

$$R_{1,\pm 1/2}|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{3}{4},n\rangle$$
$$=\pm(1/\sqrt{2})(4n\pm 1)^{1/2}|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{1}{4},n\pm\tfrac{1}{2}\rangle, \tag{9.7}$$

$$R_{0,\pm 1/2}|\tfrac{1}{2},m,-\tfrac{3}{4},n\rangle$$
$$=\mp m(4n\pm 1)^{1/2}|\tfrac{1}{2},m,-\tfrac{1}{4},n\pm\tfrac{1}{2}\rangle \quad (m=\pm\tfrac{1}{2}), \tag{9.8}$$

$$R_{-1,\pm 1/2}|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{3}{4},n\rangle = \mp(1/\sqrt{2})(4n\pm 1)^{1/2}|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{4},n\pm\tfrac{1}{2}\rangle, \quad (9.9)$$

$$R_{-1,\pm 1/2}|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{3}{4},n\rangle = 0. \quad (9.10)$$

The actions of the subalgebra generators are

$$s_0|\tfrac{1}{2},m,t,n\rangle = m|\tfrac{1}{2},m,t,n\rangle, \quad (9.11)$$

$$s_\pm|\tfrac{1}{2},m,t,n\rangle = [(\tfrac{1}{2}\mp m)(\tfrac{3}{2}\pm m)]^{1/2}|\tfrac{1}{2},m\pm 1,t,n\rangle, \quad (9.12)$$

$$t_0|\tfrac{1}{2},m,t,n\rangle = n|\tfrac{1}{2},m,t,n\rangle, \quad (9.13)$$

$$t_\pm|\tfrac{1}{2},m,t,n\rangle = \pm[-(t\mp n)(t\pm n+1)]^{1/2}|\tfrac{1}{2},m,t,n\pm 1\rangle, \quad (9.14)$$

where $t=-\frac{1}{4}$ or $t=-\frac{3}{4}$.

The metaplectic representation of SU(1,1) has been studied by Sternberg and Wolf,[10] and by Hughes[7] in the case of $\mathrm{Osp}(1,2)\supset \mathrm{SU}(1,1)$. It can be realized in terms of the space $\mathscr{H}(\mathbb{C},\mathbb{C})$ of all holomorphic functions $f:\mathbb{C}\to\mathbb{C}$ such that

$$\int |f(z)|^2 \exp(-|z|^2) d\lambda(z) < \infty, \quad (9.15)$$

where λ is the Lebesgue measure on $\mathbb{C}$. In the case of osp(3,2), the irrep under consideration decomposes into two metaplectic representations of su(1,1), and therefore it will be possible to realize the basis states as elements of $\mathscr{H}(\mathbb{C},\mathbb{C}^2)$, the space of holomorphic functions $f:\mathbb{C}\to\mathbb{C}^2$ which satisfy

$$\int (|f_1(z)|^2+|f_2(z)|^2)\exp(-|z|^2) d\lambda(z) < \infty, \quad (9.16)$$

where f_1 and f_2 are the components of f. We shall denote f by $\binom{f_1}{f_2}$. The generators of osp(3,2) are then operators acting in the space $\mathscr{H}(\mathbb{C},\mathbb{C}^2)$. If z is a complex variable, then the osp(3,2) generators can be realized as

$$s_+ = \begin{pmatrix}0&0\\1&0\end{pmatrix},\quad s_- = \begin{pmatrix}0&1\\0&0\end{pmatrix},\quad s_0 = \frac{1}{2}\begin{pmatrix}-1&0\\0&1\end{pmatrix},$$

$$t_+ = \frac{i}{2}\begin{pmatrix}z^2&0\\0&z^2\end{pmatrix},\quad t_- = \frac{i}{2}\begin{pmatrix}d^2/dz^2&0\\0&d^2/dz^2\end{pmatrix},$$

$$t_0 = \begin{pmatrix}\frac{1}{2}z\frac{d}{dz}+\frac{1}{4}&0\\0&\frac{1}{2}z\frac{d}{dz}+\frac{1}{4}\end{pmatrix},$$

$$R_{1,1/2} = e^{-3\pi i/4}\begin{pmatrix}0&0\\z&0\end{pmatrix},$$

$$R_{1,-1/2} = e^{-\pi i/4}\begin{pmatrix}0&0\\d/dz&0\end{pmatrix}, \quad (9.17)$$

$$R_{0,1/2} = -\frac{1}{\sqrt{2}}e^{-3\pi i/4}\begin{pmatrix}-z&0\\0&z\end{pmatrix},$$

$$R_{0,-1/2} = -\frac{1}{\sqrt{2}}e^{-\pi i/4}\begin{pmatrix}-d/dz&0\\0&d/dz\end{pmatrix},$$

$$R_{-1,1/2} = -e^{-3\pi i/4}\begin{pmatrix}0&z\\0&0\end{pmatrix},$$

$$R_{-1,-1/2} = -e^{-\pi i/4}\begin{pmatrix}0&d/dz\\0&0\end{pmatrix}.$$

It can be immediately verified that

$$S^2 = \begin{pmatrix}3/4&0\\0&3/4\end{pmatrix},$$

$$T^2 = \begin{pmatrix}-3/16&0\\0&-3/16\end{pmatrix},$$

$$I_2 = I_4 = \begin{pmatrix}0&0\\0&0\end{pmatrix}, \quad (9.18)$$

hence only the metaplectic representation can be realized in this way.

The states of the representation are the following elements of $\mathscr{H}(\mathbb{C},\mathbb{C}^2)$ $(n\in\mathbb{N})$:

$$|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = e^{i(\pi/2)(n+3/2)}\begin{pmatrix}z^{2n}/[(2n)!]^{1/2}\\0\end{pmatrix},$$

$$|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = e^{i(\pi/2)(n+3/2)}\begin{pmatrix}0\\z^{2n}/[(2n)!]^{1/2}\end{pmatrix}, \quad (9.19)$$

$$|\tfrac{1}{2},-\tfrac{1}{2},-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = e^{i(\pi/2)n}\begin{pmatrix}z^{2n+1}/[(2n+1)!]^{1/2}\\0\end{pmatrix},$$

$$|\tfrac{1}{2},\tfrac{1}{2},-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = e^{i(\pi/2)n}\begin{pmatrix}0\\z^{2n+1}/[(2n+1)!]^{1/2}\end{pmatrix}.$$

The normalization is chosen to correspond with the actions of the operators (9.17) as given in (9.1)–(9.14). The star conditions (6.5), with $\epsilon=+1$, reduce to the Fock condition $(d/dz)^\dagger = z$. If

$$|a\rangle = \begin{pmatrix}\sum_{j=0}^{\infty} a_{1,j}z^j\\ \sum_{k=0}^{\infty} a_{2,k}z^k\end{pmatrix} \quad \text{and} \quad (9.20)$$

$$|b\rangle = \begin{pmatrix}\sum_{j=0}^{\infty} b_{1,j}z^j\\ \sum_{k=0}^{\infty} b_{2,k}z^k\end{pmatrix}$$

are general states of the space spanned by (9.19), then the inner product is given by

$$\langle a|b\rangle = \sum_{j=1}^{2}\sum_{k=0}^{\infty} k!\, a^*_{j,k}b_{j,k}. \quad (9.21)$$

The expressions (9.19), considered as functions from $\mathbb{C}$ to $\mathbb{C}^2$, form a complete orthonormal set in $\mathscr{H}(\mathbb{C},\mathbb{C}^2)$, and the inner product which corresponds to (9.21) can be expressed as

$$\langle f|g\rangle = \frac{1}{\pi}\int f(z)^\dagger g(z)\exp(-|z|^2)d\lambda(z).$$

$$= \frac{1}{\pi}\int [f_1(z)^* g_1(z)+f_2(z)^* g_2(z)]\exp(-|z|^2)d\lambda(z). \quad (9.22)$$

In the case of osp(1,2), one of the infinite-dimensional dispin representations, investigated by Hughes,[7] could be realized as a metaplectic representation. It is interesting that we find a similar result for a particular infinite-dimensional osp(3,2) representation. This extends the general result of Sternberg and Wolf, who show that the metaplectic representation of $\mathrm{Sp}(2m)$ can be considered as an irreducible representation of $\mathrm{Osp}(1,2m)$.[10]

X. THE HARMONIC OSCILLATOR WITH SPIN ½ STATES

The Hamiltonian H of a one-dimensional harmonic oscillator may be written as (see, for instance, Ref. 13)

$$H = a^+a+\tfrac{1}{2}, \quad (10.1)$$

where a^+ and a are the boson creation and annihilation operators which satisfy

$$[a,a^+] = 1. \tag{10.2}$$

Suppose we also have the disposal of a pair of fermion creation and annihilation operators (in this section, anticommutation is written as $\{\ \}$):

$$\begin{aligned} &\{b,b^+\} = 1, \quad \{c,c^+\} = 1, \\ &b^2 = (b^+)^2 = c^2 = (c^+)^2 = 0, \\ &\{b^{(+)},c^{(+)}\} = 0, \end{aligned} \tag{10.3}$$

which commute with the boson operators

$$[a^{(+)},b^{(+)}] = [a^{(+)},c^{(+)}] = 0. \tag{10.4}$$

The fermion operators give rise to an SU(2) algebra

$$s_+ = b^+c, \quad s_- = c^+b, \quad s_0 = \tfrac{1}{2}(b^+b - c^+c). \tag{10.5}$$

It follows that b^+ and c^+ (resp., b and c) can be interpreted as creation (resp., annihilation) operators of spin $+\frac{1}{2}$ and $-\frac{1}{2}$ states.

Let us define a vacuum state by means of

$$a|0\rangle = b|0\rangle = c|0\rangle = 0. \tag{10.6}$$

We are interested in those states which are simultaneously eigenstates of the Hamiltonian H and of the total spin $S^2 = s_+s_- + s_0^2 - s_0$ with eigenvalues $s(s+1) = \frac{3}{4}$ (i.e., the states with nonvanishing total spin). We find

$$(a^+)^n b^+|0\rangle \text{ and } (a^+)^n c^+|0\rangle \quad (n = 0,1,2,\ldots). \tag{10.7}$$

But this is precisely the metaplectic representation, considered in Sec. IX. Indeed, the osp(3,2) algebra is realized by (10.5) and

$$t_+ = \tfrac{1}{2}(a^+)^2 N, \quad t_- = -\tfrac{1}{2}a^2 N, \quad t_0 = \tfrac{1}{2}(a^+a + \tfrac{1}{2})N, \tag{10.8}$$

where

$$N = \tfrac{4}{3}S^2 = b^+b + c^+c + 2b^+c^+bc,$$

and

$$\begin{aligned} &R_{1,1/2} = -a^+b^+c, \quad R_{-1,-1/2} = -ac^+b, \\ &R_{0,1/2} = \frac{1}{\sqrt{2}}a^+(b^+b - c^+c), \\ &R_{0,-1/2} = -\frac{1}{\sqrt{2}}a(b^+b - c^+c), \\ &R_{-1,1/2} = a^+c^+b, \quad R_{1,-1/2} = ab^+c. \end{aligned} \tag{10.9}$$

The identification with the states of Sec. IX is given by

$$\begin{aligned} &\left|\frac{1}{2},\frac{1}{2},-\frac{1}{4},\frac{1}{4}+n\right\rangle = \frac{(a^+)^{2n}}{\sqrt{(2n)!}}b^+|0\rangle, \\ &\left|\frac{1}{2},-\frac{1}{2},-\frac{1}{4},\frac{1}{4}+n\right\rangle = \frac{(a^+)^{2n}}{\sqrt{(2n)!}}c^+|0\rangle, \\ &\left|\frac{1}{2},\frac{1}{2},-\frac{3}{4},\frac{3}{4}+n\right\rangle = -\frac{(a^+)^{2n+1}}{\sqrt{(2n+1)!}}b^+|0\rangle, \\ &\left|\frac{1}{2},-\frac{1}{2},-\frac{3}{4},\frac{3}{4}+n\right\rangle = -\frac{(a^+)^{2n+1}}{\sqrt{(2n+1)!}}c^+|0\rangle, \end{aligned} \tag{10.10}$$

The energy and spin eigenvalues are determined by

$$\begin{aligned} &H|\tfrac{1}{2},\pm\tfrac{1}{2},-\tfrac{1}{4},\tfrac{1}{4}+n\rangle = (2n+\tfrac{1}{2})|\tfrac{1}{2},\pm\tfrac{1}{2},-\tfrac{1}{4},\tfrac{1}{4}+n\rangle, \\ &H|\tfrac{1}{2},\pm\tfrac{1}{2},-\tfrac{3}{4},\tfrac{3}{4}+n\rangle = (2n+\tfrac{3}{2})|\tfrac{1}{2},\pm\tfrac{1}{2},-\tfrac{3}{4},-\tfrac{3}{4}+n\rangle, \end{aligned} \tag{10.11}$$

$$\begin{aligned} &S^2|\tfrac{1}{2},\pm\tfrac{1}{2},t,-t+n\rangle = \tfrac{3}{4}|\tfrac{1}{2},\pm\tfrac{1}{2},t,-t+n\rangle, \\ &s_0|\tfrac{1}{2},\pm\tfrac{1}{2},t,-t+n\rangle = \pm\tfrac{1}{2}|\tfrac{1}{2},\pm\tfrac{1}{2},t,-t+n\rangle, \\ &(t = -\tfrac{1}{4},-\tfrac{3}{4})(n = 0,1,2,\ldots). \end{aligned}$$

Consequently, osp(3,2) is the spectrum generating algebra for the one-dimensional harmonic oscillator with spin $\frac{1}{2}$ states.

XI. FINAL REMARKS

We think it is useful, finally, to illustrate some properties, which are well known by pure mathematicians, in the case study of osp(3,2). In this section, $\langle V\rangle$ denotes the linear span over $\mathbb{C}$ of the subset V of a linear space W.

We recall the definition of a Jordan superalgebra.[15] Let $J = J_{\bar{0}} \oplus J_{\bar{1}}$ be a $\mathbb{Z}_2$-graded algebra with bilinear product $x\circ y$, which satisfies

$$x\circ y = (-1)^{\xi\eta} y\circ x \tag{11.1}$$

for $x\in J_\xi$, $y\in J_\eta$. Then J is a commutative superalgebra. Let L_a be the element of End(J), defined by

$$\begin{aligned} &L_a : J \to J \\ &\quad x \to L_a(x) = a\circ x. \end{aligned} \tag{11.2}$$

Then J is called a Jordan superalgebra if

$$\begin{aligned} &(-1)^{\alpha\gamma}[L_{a\circ b},L_c] \\ &\quad + (-a)^{\beta\alpha}[L_{b\circ c},L_a] + (-1)^{\gamma\beta}[L_{c\circ a},L_b] = 0, \end{aligned} \tag{11.3}$$

for all $a\in J_\alpha$, $b\in J_\beta$, and $c\in J_\gamma$.

We give an example of Jordan superalgebra by means of the following: Suppose that e is a basis for $J_{\bar{0}}$ and $\{a,b\}$ is a basis for $J_{\bar{1}}$. We define a nondegenerate bilinear form $(\ ,\ )$ on J by (i) $(e,e) = 1$; (ii) the restriction of $(\ ,\ )$ to $J_{\bar{1}}$ is skew symmetric, with $(a,b) = 1$; (iii) $(J_{\bar{0}},J_{\bar{1}}) = 0$.
The multiplication on J is then defined as

$$x\circ y = (e,x)y + (e,y)x - (x,y)e, \quad \forall x,y\in J. \tag{11.4}$$

This implies the following multiplication table for the basis elements of J:

$\circ$	e	a	b
e	e	a	b
a	a	0	$-e$
b	b	e	0

(11.5)

This example is a Jordan superalgebra of type D (see Ref. 15).

Let us now return to the case of L = osp(3,2). It is easy to verify that this Lie superalgebra is a $\mathbb{Z}$-graded Lie superalgebra[15,16]

$$L = L^{-1} \oplus L^0 \oplus L^{+1}, \tag{11.6}$$

where

$$L^0 = \langle t_0,t_+,t_-,R_{0,1/2},R_{0,-1/2},s_0\rangle, \tag{11.7}$$

$$L^{+1} = \langle s_+,R_{1,1/2},R_{1,-1/2}\rangle, \tag{11.8}$$

$$L^{-1} = \langle s_-,R_{-1,1/2},R_{-1,-1/2}\rangle. \tag{11.9}$$

Note that the subsuperalgebra L^0 is in fact osp(1,2) $\oplus$ so(2). The $\mathbb{Z}$-grading in (11.5)–(11.6) is obviously not consistent with the $\mathbb{Z}_2$-grading of L.

Kac has proven[15] that for a $\mathbb{Z}$-graded Lie superalgebra $L = L^{-1} \oplus L^0 \oplus L^1$, the space $L^{-1} = L_{\bar{0}}^{-1} \oplus L_{\bar{1}}^{-1}$ is a Jordan superalgebra if we let

$$x\circ y = [[p,x],y], \quad x,y\in L^{-1}, \tag{11.10}$$

where $p\in L_{\bar{0}}^{+1}$. In our case it is easy to verify that L^{-1} is the

Jordan superalgebra of type D defined by (11.5). If we take $p = is_+/2$, then the relation between the basis (11.5) and the basis (11.9) is given by

$$e = is_-, \qquad (11.11)$$
$$a = \sqrt{2}R_{-1,1/2}, \quad b = \sqrt{2}R_{-1,-1/2}.$$

More generally, the Lie superalgebra osp(3,2) is in fact that image of the Jordan superalgebra (11.5) under the Kantor factor Kan, which gives an isomorphism of the category of finite-dimensional Jordan superalgebras onto the category of finite-dimensional admissible Z-graded Lie superalgebras.[15]

ACKNOWLEDGMENTS

This work would not have been realized without the valuable discussions and helpful criticisms of Dr. J. W. B. Hughes (Applied Mathematics Department, Queen Mary College, London), whom I gratefully acknowledge.

I also thank Dr. R. C. King and Prof. M. Englefield (Southampton University, U.K.) for very useful discussions.

The British Council is acknowledged for a maintenance grant. This research project was partially supported by SERC-grant GR/C/33147.

[1]V. G. Kac, Adv. Math. **26**, 8–96 (1977).
[2]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626–39 (1976).
[3]M. Scheunert, *Lecture Notes in Mathematics* (Springer, Berlin, 1979), Vol. 716.
[4]V. G. Kac, *Springer Lecture Notes in Mathematics* (Springer, Berlin, 1978), Vol. 676, pp. 597–626.
[5]R. J. Farmer and P. D. Jarvis, J. Phys. A: Math. Gen. **16**, 473–487 (1983).
[6]J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068–78 (1978).
[7]J. W. B. Hughes, J. Math. Phys. **22**, 245–250 (1981).
[8]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146–154 (1977).
[9]M. De Crombrugghe and V. Rittenberg, Ann. Phys. **151**, 99–126 (1983).
[10]S. Sternberg and J. A. Wolf, Trans. Am. Math. Soc. **238**, 1–43 (1978).
[11]J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, Ann. Phys. **147** (1), 85–139 (1983).
[12]A. R. Edmonds, *Angular momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957).
[13]B. G. Wybourne, *Classical groups for physicists* (Wiley, NY, 1974), pp. 195–206.
[14]This is not in contradiction with the results in Ref. 8, where the authors conclude that the osp(m,2p) ($m \neq 2$) algebras have two classes of finite-dimensional grade star representations with a graded Hilbert space. Indeed, in their last paragraph they stress that their conditions are only necessary, and that the classes can actually be empty.
[15]V. G. Kac, Comm. Algebra **5** (13), 1375–1400 (1977).
[16]I. Bars and M. Günaydin, Commun. Math. Phys. **91**, 31–51 (1983).

Some integrals involving three Bessel functions when their arguments satisfy the triangle inequalities[a)]

A. Gervois[b)] and H. Navelet
Service de Physique Théorique, CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 3 January 1984; accepted for publication 13 April 1984)

We calculate definite integrals involving three Bessel functions of the form $\int_0^\infty x J_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx$ and $\int_0^\infty x Y_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx$, which appeared in some absorption calculations. Results are given for all relative positions of lengths a, b, c. They complete formulas outside the triangle case and correct some misprints given in tables.

PACS numbers: 12.40.Pp, 02.30.Bi

I. INTRODUCTION

For a while, the phenomenology of high-energy physics relied heavily on the Regge formalism. To perform analytically the absorption-type calculation[1,2] the integral $\int_0^\infty r\,dr\,J_\alpha(r\sqrt{-t})J_\beta(r\sqrt{-t'})J_{\alpha+\beta}(r\sqrt{-t''})$ was needed when the three real variables $\sqrt{-t}$, $\sqrt{-t'}$, $\sqrt{-t''}$ obey the triangle inequalities. Unfortunately such integrals were known only in particular cases,[3] for $\alpha = 0$ say. On the other hand all the papers and books about the Bessel functions did give the values of these integrals but in the case $\sqrt{-t''} > \sqrt{-t'} + \sqrt{-t}$, where the value is zero.

For all these reasons, it was interesting to compute analytically these integrals. To do that we have been led to compute other integrals involving two or three Bessel functions of the form

$$\int_0^\infty xJ_\mu(ax)J_{\pm\nu}(bx)dx,$$
$$\int_0^\infty xJ_\mu(ax)\begin{Bmatrix} J_\nu & (bx) \\ Y_\nu & (bx) \end{Bmatrix} J_{\mu+\nu}(cx)dx. \tag{1.1}$$

All these integrals are given in Tables I and II.

The second part of this paper is devoted to the particular case when the three orders α, β, and $\gamma = \alpha + \beta$ are integers and we give a very simple interpretation of the result so obtained by comparing it with the sum[2,4]

$$\sum_j (2j+1)d^j_{m_1m_2}(\theta_3)d^j_{m_1m_3}(\theta_2)d^j_{m_3m_2}(\theta_1).$$

In the third part, we treat the general case by extending the result about the normalization of the Bessel functions of integer order[5] to real order and get the announced result about integrals of type (1.1). In the last part we check known results compiled in Refs. 3 and 6 where particular cases are examined. A short summary of our new results is contained in the conclusion.

II. INTEGRALS INVOLVING THREE BESSEL FUNCTIONS OF INTEGER ORDER FOR $|c-b| < a < c+b$

Following Graf,[7,8] we write

$$e^{i\mu\phi_b}J_\mu(xa) = \sum_{n=-\infty}^{+\infty} J_{\mu+n}(xc)J_n(xb)e^{in\phi_a}, \tag{2.1}$$

where the angles ϕ_i ($i = a,b$, or c, $\Sigma_i \phi_i = \pi$) are defined in Fig. 1 and x is a real parameter, n is an integer, and μ is real.

This yields at once

$$J_{\mu+n}(xc)J_n(xb) = \frac{1}{2\pi}\int_0^{2\pi} e^{i(\mu\phi_b - n\phi_{a'})}J_\mu(xa')d\phi_{a'}, \tag{2.2}$$

where

$$a'^2 = c^2 + b^2 - 2cb\cos\phi_{a'}$$

and

$$\int_0^\infty x\,dx\,J_\mu(ax)J_n(bx)J_{\mu+n}(cx)$$
$$= \frac{1}{2\pi}\int_0^\infty x\,dx\,J_\mu(xa)\int_0^{2\pi} e^{i(\mu\phi_b - n\phi_{a'})}\times J_\mu(xa')d\phi_{a'}.$$

By inverting the order of the two integrations of the right-hand side and taking advantage of the orthogonality relation for Bessel functions of integer order[4] namely,

$$\int_0^\infty x\,dx\,J_m(xa)J_m(xa') = \frac{\delta(a-a')}{(aa')^{1/2}} \tag{2.3}$$

we get for $\mu = m$, m integer

$$\int_0^\infty x\,dx\,J_m(ax)J_n(bx)J_{m+n}(cx)$$
$$= \frac{2}{\pi}\int_0^{2\pi} d\phi_{a'}\cos(m\phi_b - n\phi_{a'})\frac{\delta(a-a')}{(aa')^{1/2}},$$

where the factor 2 comes from the symmetry of the integrand.

Now

$$\frac{\delta(a-a')}{(aa')^{1/2}} = \frac{\delta(\phi_{a'}-\phi_a)}{(aa')^{1/2}|\partial a'/\partial\phi_{a'}|} = \frac{\delta(\phi_{a'}-\phi_a)}{bc\sin\phi_a}.$$

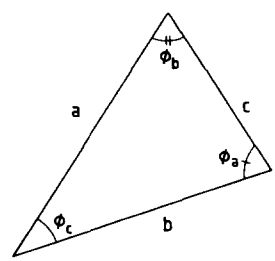


FIG. 1. Convention for the angles of the triangle.

[a)] Dedicted to the memory of Friedrich W. Bessel on his 200th birthday (1784–1846).
[b)] Chercheur au CNRS.

 0022-2488/84/113350-07$02.50

The final result reads

$$\int_0^\infty x\,dx\,J_m(ax)J_n(bx)J_{m+n}(cx)=\frac{\cos(m\phi_b-n\phi_a)}{\pi ab\sin\phi_c}. \tag{2.4}$$

(In a triangle $bc\sin\phi_a=ab\sin\phi_c=ac\sin\phi_b$.)

At this stage a comment is in order. By exactly the same kind of trick it is possible to get very easily the sum

$$f=\sum_j(2j+1)d^j_{m_1m_2}(\theta_3)d^j_{m_1m_3}(\theta_2)d^j_{m_3m_2}(\theta_1),$$

where the three unit vectors $\hat q_1,\hat q_2,\hat q_3$ have polar angles (0,0), $(\theta_3,0)$, and (θ_2,φ_1), respectively, the angles between $\hat q_i$ and $\hat q_j$ being θ_k $(i,j,k=1,2,3)$.

Let $\Omega_{\hat q_2}$ be a rotation which brings the unit vector $\hat z$ onto $\hat q_2$ such that $\Omega_{\hat q_1}^{-1}\Omega_{\hat q_2}$ is a rotation with Euler angle $(0,\theta_3,0)$,

$\Omega_{\hat q_1}^{-1}\Omega_{\hat q_3}$ is a rotation with Euler angles $(\varphi_1,\theta_2,0)$ and

$\Omega_{\hat q_3}^{-1}\Omega_{\hat q_2}$ is a rotation with Euler angles $(\varphi_3,\theta_1,\varphi_2)$.

From

$$\mathscr{D}^j_{m_1m_2}(\Omega_{\hat q_1}^{-1}\Omega_{\hat q_3}\Omega_{\hat q_3}^{-1}\Omega_{\hat q_2})=\sum_{m_3}\mathscr{D}^j_{m_1m_3}(\Omega_{\hat q_1}^{-1}\Omega_{\hat q_3})\mathscr{D}^j_{m_3m_2}(\Omega_{\hat q_3}^{-1}\Omega_{\hat q_2}),$$

we deduce at once

$$d^j_{m_1m_2}(\theta_3)=\sum_{m_3}\exp[-i(m_1\varphi_1+m_2\varphi_2+m_3\varphi_3)]\times d^j_{m_1m_3}(\theta_2)d^j_{m_3m_2}(\theta_1)$$

or

$$d^j_{m_1m_3}(\theta_2)\,d^j_{m_3m_2}(\theta_1)=2\times\frac{1}{2\pi}\int_0^{2\pi}\cos(m_1\varphi_1+m_2\varphi_2+m_3\varphi'_3)d^j_{m_1m_2}(\theta_3)d\varphi'_3$$

and

$$f=\frac{1}{\pi}\int_0^{2\pi}d\varphi'_3\cos(m_1\varphi_1+m_2\varphi_2+m_3\varphi'_3)\times\sum_j(2j+1)d^j_{m_1m_2}(\theta'_3)d^j_{m_1m_2}(\theta_3)$$
$$=\frac{2}{\pi}\int_0^{2\pi}d\varphi'_3\cos(m_1\varphi_1+m_2\varphi_2+m_3\varphi_3)\times\delta(\cos\theta_3-\cos\theta'_3).$$

Noting that $\cos\theta_3=\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2\cos\varphi_3$ and

$$\left|\frac{\partial\cos\theta_3}{\partial\varphi_3}\right|=\sin\theta_1\sin\theta_2\sin\varphi_3=|\hat q_1\cdot(\hat q_2\times\hat q_3)|,$$

we get

$$\sum_j(2j+1)d^j_{m_1m_2}(\theta_3)d^j_{m_1m_3}(\theta_2)d^j_{m_3m_2}(\theta_1)=\frac{2}{\pi}\frac{\cos(m_1\varphi_1+m_3\varphi_3+m_2\varphi_2)}{|\hat q_1\cdot(\hat q_2\times\hat q_3)|}\Theta(\Delta), \tag{2.5}$$

where Θ is the Heaviside step function

$$\sin\theta_1\sin\theta_3e^{-i\varphi_2}=\cos\theta_2-\cos\theta_1\cos\theta_3-i\hat q_1\cdot(\hat q_2\times\hat q_3),$$
$$\sin\theta_2\sin\theta_3e^{-i\varphi_1}=\cos\theta_1-\cos\theta_2\cos\theta_3-i\hat q_1\cdot(\hat q_2\times\hat q_3),$$
$$\sin\theta_1\sin\theta_2e^{-i\varphi_3}=-[\cos\theta_3-\cos\theta_2\cos\theta_1-i\hat q_1\cdot(\hat q_2\times\hat q_3)],$$

and

$$|q_1\cdot(\hat q_2\times\hat q_3)|^2=1-\cos^2\theta_1-\cos^2\theta_2-\cos^2\theta_3+2\cos\theta_1\cos\theta_2\cos\theta_3=\Delta.$$

As expected, result (2.5) is strikingly similar to the one obtained for the three Bessel functions. Indeed for large j and small θ, $d^j_{\alpha\beta}(\theta)\simeq J_{|\alpha-\beta|}(j\theta)$ and the sum over integer value of j becomes an integral

$$2\int_0^\infty j\,dj\,J_{|m_1-m_2|}(j\theta_3)J_{|m_1-m_3|}(j\theta_2)J_{|m_2-m_3|}(j\theta_1).$$

For small θ_i's

$$\Delta\cong-\tfrac14\{\theta_1^4+\theta_2^4+\theta_3^4-2\theta_1^2\theta_2^2-2\theta_1^2\theta_3^2-2\theta_3^2\theta_2^2\}=-\delta/4,$$
$$\delta^{1/2}=\sqrt{[\theta_3^2-(\theta_1-\theta_2)^2][(\theta_1+\theta_2)^2-\theta_3^2]},$$
$$2\theta_1\theta_3\cos\varphi_2\cong\theta_1^2+\theta_3^2-\theta_2^2,\quad 2\theta_1\theta_3\sin\varphi_2\simeq\delta^{1/2},$$
$$2\theta_2\theta_3\cos\varphi_1\cong\theta_2^2+\theta_3^2-\theta_1^2,\quad 2\theta_2\theta_3\sin\varphi_1\simeq\delta^{1/2},$$
$$2\theta_1\theta_2\cos\varphi_3\cong+(\theta_3^2-\theta_2^2-\theta_1^2),\quad 2\theta_1\theta_2\sin\varphi_3\simeq-\delta^{1/2},$$
$$\varphi_1+\varphi_2+\varphi_3\simeq0$$

and

$$\int j\,dj\,J_{|m_1-m_2|}(j\theta_3)J_{|m_1-m_3|}(j\theta_2)J_{|m_2-m_3|}(j\theta_1)\simeq\frac{1}{\pi}\frac{\cos[(m_1-m_3)\varphi_1-(m_3-m_2)\varphi_2]}{\theta_2\theta_1\sin\varphi_3}\Theta(-\delta). \tag{2.6}$$

We get at once that the integral is zero if $\theta_1,\theta_2,\theta_3$ do not fulfill the triangle inequalities.

In the next part, we show that it is possible to generalize the formula (2.4) in the case where both m and m' are real and for the case were J_λ is replaced by the Neumann function Y_λ.

III. DERIVATION OF THE INTEGRALS

A. Some preliminary results

One of the main problems we shall encounter is the estimation of the integral

$$\int_0^\infty xJ_\lambda(cx)J_\rho(\tilde cx)dx=\lim_{R\to\infty}\int_0^R xJ_\lambda(cx)J_\rho(\tilde cx)dx,$$

where $c,\tilde c$ are positive numbers. Provided $\mathrm{Re}(\lambda+\rho)>-2$, the behavior at $x=0$ is regular. For large R, the result involves very rapidly oscillating functions whose limits are actually the distribution $\delta(\tilde c-c)$ and P.P./$(\tilde c^2-c^2)$ (P.P. stands for principal part).

Starting with the indefinite integral[9]

$$(\tilde c^2-c^2)\int xJ_\lambda(cx)J_\rho(\tilde cx)dx+(\lambda^2-\rho^2)\int\frac{J_\lambda(cx)J_\rho(\tilde cx)}{x}dx=cxJ_\rho(\tilde cx)J_{\lambda-1}(cx)-\tilde cxJ_{\rho-1}(\tilde cx)J_\lambda(cx)+(\rho-\lambda)J_\lambda(cx)J_\rho(\tilde cx), \tag{3.1}$$

and using the well-known behavior of Bessel functions for small and large (real) variables, we get

$$\int_0^R xJ_\lambda(cx)J_\rho(\tilde{c}x)dx + \frac{(\lambda^2-\rho^2)}{\tilde{c}^2-c^2}\int_0^R \frac{J_\lambda(cx)J_\rho(\tilde{c}x)}{x}dx$$
$$= -\frac{1}{\pi(c\tilde{c})^{1/2}}\left\{\frac{\cos[(c+\tilde{c})R-(\lambda+\rho)\pi/2]}{c+\tilde{c}} - \frac{\sin[(\tilde{c}-c)R-(\rho-\lambda)\pi/2]}{\tilde{c}-c}\right\}+O\left(\frac{1}{R}\right),$$

where we have assumed $\mathrm{Re}(\lambda+\rho)>0$ (behavior at $x=0$).

Now[10]

$$\int_0^\infty \frac{J_\lambda(cx)J_\rho(\tilde{c}x)}{x}dx = \frac{1}{2\pi}\left(\frac{\tilde{c}}{c}\right)^\rho \frac{\Gamma((\lambda+\rho)/2)\Gamma((\rho-\lambda)/2)}{\Gamma(\rho+1)} \times\sin\left[(\rho-\lambda)\frac{\pi}{2}\right]\times {}_2F_1\left(\frac{\rho+\lambda}{2},\frac{\rho-\lambda}{2};\rho+1;\frac{\tilde{c}^2}{c^2}\right)$$

with $\tilde{c}<c$, and a similar result holds for $\tilde{c}>c$ by the exchanges $\lambda\leftrightarrow\rho$, $\tilde{c}\leftrightarrow c$.

For $c=\tilde{c}$, both expressions have the same limit[10]

$$\frac{2}{\pi}\frac{\sin[(\rho-\lambda)\pi/2]}{\rho^2-\lambda^2}.$$

As $c+\tilde{c}>0$,

$$\frac{\cos[c+\tilde{c})R]}{c+\tilde{c}} \quad\text{and}\quad \frac{\sin[(c+\tilde{c})R]}{c+\tilde{c}}$$

are bounded and rapidly oscillate as $R\to\infty$; in further integrations, they give a vanishing contribution. Terms like

$$\frac{\sin(\tilde{c}-c)R}{\tilde{c}-c} \quad\text{and}\quad \frac{\cos[(\tilde{c}-c)R]}{\tilde{c}-c}$$

may give a large (of the order of R) contribution in a region $\tilde{c}-c=O(1/R)$; when further integrated, the larger-order terms give a finite, nonzero contribution. Thus

$$\int_x^R xJ_\lambda(cx)J_\rho(\tilde{c}x)dx$$

behaves like

$$-\frac{1}{\pi\sqrt{c\tilde{c}}}\left\{\frac{\cos[(c+\tilde{c})R-(\lambda+\rho)\pi/2]}{c+\tilde{c}} - \frac{\sin[(\tilde{c}-c)R-(\rho-\lambda)\pi/2]}{\tilde{c}-c}\right\} - \frac{(\lambda^2-\rho^2)}{\tilde{c}^2-c^2}\int_0^\infty \frac{J_\lambda(cx)J_\rho(\tilde{c}x)dx}{x}$$

or, using the notation of distribution

$$\int_0^\infty xJ_\lambda(cx)J_\rho(\tilde{c}x)dx = \frac{\cos[(\rho-\lambda)\pi/2]}{(c\tilde{c})^{1/2}}\delta(\tilde{c}-c) - \frac{(\lambda^2-\rho^2)}{2\pi} \times\frac{\Gamma((\rho+\lambda)/2)\Gamma((\rho-\lambda)/2)\sin[(\rho-\lambda)\pi/2]}{\Gamma(\rho+1)} \times\left(\frac{\tilde{c}}{c}\right)^\rho {}_2F_1\left(\frac{\rho+\lambda}{2},\frac{\rho-\lambda}{2};\rho+1;\frac{\tilde{c}^2}{c^2}\right)\frac{\mathrm{P.P.}}{\tilde{c}^2-c^2}, \tag{3.2a}$$

with

$$\mathrm{Re}(\lambda+\rho)>0 \quad\text{and}\quad \tilde{c}>c. \tag{3.2b}$$

The singular behavior coming from the division by $\tilde{c}^2-c^2$ is killed by the oscillating contribution $\cos[(\tilde{c}-c)R]/(\tilde{c}-c)$ and is replaced by a principal part. Result (3.2a) for $x\neq y$ was already proposed by Schindler.[11]

When $\rho=\lambda$, the second term disappears and we have the closure relation

$$\int_0^\infty xJ_\lambda(cx)J_\lambda(\tilde{c}x)dx = \frac{1}{(c\tilde{c})^{1/2}}\delta(\tilde{c}-c), \tag{3.3}$$

which generalizes the well-known formula[4] when λ is an integer. The condition $\mathrm{Re}\,\lambda>0$ may be relaxed to $\mathrm{Re}\,\lambda>-1$ as the integral $\int(J_\lambda J_\lambda/x)\,dx$ does not come in. Expression (3.2) may be extended to the case $\lambda+\rho=0$ by starting again from the indefinite integral (3.1) as again $\int(J_\lambda J_{-\lambda}/x)\,dx$ must not be taken into account. Some care is needed to examine the behavior near $x=0$ of the right-hand-side. The final normalization condition reads

$$\int_0^\infty xJ_\lambda(cx)J_{-\lambda}(\tilde{c}x)dx = \frac{\cos\lambda\pi}{(c\tilde{c})^{1/2}}\delta(\tilde{c}-c) - \frac{2\sin\pi\lambda}{\pi}\left(\frac{c}{\tilde{c}}\right)^\lambda\frac{\mathrm{P.P.}}{\tilde{c}^2-c^2}, \quad \mathrm{Re}\,\lambda>-1 \tag{3.4a}$$

or more simply

$$\int_0^\infty xJ_\lambda(cx)Y_\lambda(\tilde{c}x)dx = \frac{2}{\pi}\left(\frac{c}{\tilde{c}}\right)^\lambda\frac{\mathrm{P.P.}}{\tilde{c}^2-c^2}, \quad \mathrm{Re}\,\lambda>-1, \tag{3.4b}$$

where the condition $\tilde{c}<c$ may be removed, and Y_λ is the Bessel function of the second kind.[12] For $\lambda=l$ integer, result (3.4a) is obvious since $J_{-l}=(-)^lJ_l$. For $\lambda\neq l$, we have an extra term as expected from Carlson's theorem.

B. Calculation of $\int_0^\infty xJ_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx=I$

The basic formula is the expression of $J_\mu(Z)J_\nu(z)$ for real Z,z as an integral[13]

$$J_\mu(Z)J_\nu(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\nu\theta}\left(\frac{Z+ze^{i\theta}}{Z+ze^{-i\theta}}\right)^{(\mu+\nu)/2}J_{\mu+\nu}(\bar{\omega}_1)d\theta - \frac{\sin\nu\pi}{\pi}\int_1^{Z/z}\rho^{-\nu-1}\left\{\frac{Z-z\rho}{Z-z/\rho}\right\}^{(\mu+\nu)/2}J_{\mu+\nu}(\bar{A})d\rho,$$

where $z<Z$, $\bar{\omega}_1^2=Z^2+z^2+2Zz\cos\theta$, $\bar{A}^2=Z^2+z^2-Zz(\rho+1/\rho)$. There exists a similar formula for $Z<z$ by exchanging μ and ν and Z and z. For $Z=z$, we get the well-known formula for the isosceles triangle[14]

$$J_\mu(z)J_\nu(z) = \frac{1}{\pi}\int_{-\pi/2}^{+\pi/2}\cos[(\mu-\nu)\theta]J_{\mu+\nu}(2z\cos\theta)d\theta.$$

With some changes in the notations ($Z\Rightarrow ax$, $z\Rightarrow bx$) and variables ($\theta\to-\theta\to\pi-\theta$, $\rho=e^u$) we rewrite it in a more symmetrical form

$$J_\mu(ax)J_\nu(bx) = \frac{1}{\pi}\mathrm{Re}\int_0^\pi e^{i\nu\pi}e^{-i\nu\theta}\left(\frac{a-be^{i\theta}}{a-be^{-i\theta}}\right)^{(\mu+\nu)/2} \times J_{\mu+\nu}(\omega_1 x)d\theta - \frac{\sin\nu\pi}{\pi}\int_0^{\ln a/b}e^{-\nu u} \times\left(\frac{a-be^u}{a-be^{-u}}\right)^{(\mu+\nu)/2}J_{\mu+\nu}(Ax)du, \tag{3.5a}$$

with

$$b<a, \quad \omega_1^2 = a^2 + b^2 - 2ba\cos\theta,$$
$$A^2 = a^2 + b^2 - 2ab\cosh u. \tag{3.5b}$$

Now,

$$\lim_{R\to\infty}\int_0^R xJ_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx$$
$$= \frac{1}{\pi}\,\mathrm{Re}\int_0^\pi e^{i\nu\pi}e^{-i\nu\theta}\left(\frac{a-be^{i\theta}}{a-be^{-i\theta}}\right)^{(\mu+\nu)/2}$$
$$\times \lim_{R\to\infty}\int_0^R xJ_{\mu+\nu}(cx)J_{\mu+\nu}(\omega_1 x)dx$$
$$- \frac{\sin\nu\pi}{\pi}\int_0^{\ln a/b} e^{-\nu u}\left(\frac{a-be^{u}}{a-be^{-u}}\right)^{(\mu+\nu)/2}$$
$$\times \lim_{R\to\infty}\int_0^R xJ_{\mu+\nu}(cx)J_{\mu+\nu}(Ax)dx,$$
$$\mathrm{Re}(\mu+\nu) > -1.$$

We may permute the integrations, as the intervals are finite and all integrals have a meaning. Using formula (3.3) for $\lambda=\rho=\mu+\nu$ and with $\tilde{c}=$ either ω_1 or A, we get

$$I = \frac{1}{\pi c}\int_0^\pi \delta(c-\omega_1)\mathrm{Re}\left\{e^{-i\nu\pi-i\nu\theta}\left(\frac{a-be^{i\theta}}{a-be^{-i\theta}}\right)^{(\mu+\nu)/2}d\theta\right\}$$
$$- \frac{\sin\nu\pi}{\pi c}\int_0^{\ln a/b}\delta(c-A)e^{-\nu u}\left(\frac{a-be^{u}}{a-be^{-u}}\right)^{(\mu+\nu)/2}du,$$
$$\mathrm{Re}(\mu+\nu) > -1,\ a<b. \tag{3.6}$$

Three cases may occur.

(i) either $c>a+b$, then $c>\omega_1$ for every $\theta\in[0,\pi]$ and $c>A$ for $u\in[0,\ln a/b]$. (3.7)

Hence, both terms are zero and $I=0$. (See formula 2.6 for integer indices.)

(ii) $a-b<c<a+b$ (triangular configuration). Then $c^2=a^2+b^2-2ab\cos\phi_c$, where ϕ_a,ϕ_b,ϕ_c $(0\leqslant\phi_i\leqslant\pi)$ are the angles related to sides a,b,c of the triangle (see Fig. 1), then $c>A$ for $u\in[0,\ln a/b]$. The second term in the right-hand side of (3.6) is zero; but there exists $\theta=\phi_c$ such that $\omega_1=c$ and the first term is not zero.

As

$$\delta(c-\omega_1) = 2c\delta(c^2-\omega_1^2)$$
$$= \frac{c}{ab}\delta(\cos\theta-\cos\phi_c)$$
$$= \frac{c}{ab\sin\phi_c}\delta(\theta-\phi_c)$$

we get for the first term

$$I = \frac{\cos(\nu\phi_a-\mu\phi_b)}{\pi ab\sin\phi_c}, \tag{3.8a}$$
$$\Delta = \tfrac{1}{2}ab\sin\phi_c = \tfrac{1}{2}bc\sin\phi_a = \tfrac{1}{2}ca\sin\phi_b, \tag{3.8b}$$

where Δ is the area of the triangle; we have explicitly used the relation $\pi=\phi_a+\phi_b+\phi_c$ to restore the symmetry ($\mu\leftrightarrow\nu$), ($a\leftrightarrow b$). Note that formula (3.8a) is the generalization of formula (2.4) obtained fro $m=\mu$ and $n=\nu$ integers.

(iii) $0<c<a-b$. We introduce the three positive arguments u_a,u_b,u_c which generalize the angles ϕ in the triangle case (ii) ($a^2=b^2+c^2+2bc\cosh u_a$, $b^2=a^2+c^2-2ac\cosh u_b$, $c^2=a^2+b^2-2ab\cosh u_c$). Only the last term in (3.6) gives a nonzero contribution. As $\delta(A-c)=c\delta(u-u_c)/(ab\sinh u_c)$, we get, after some manipulations, the nearly symmetric formula

$$I = -(\sin\nu\pi/2\pi\tilde{\Delta})e^{-\nu u_a-\mu u_b}, \tag{3.9a}$$

with

$$\tilde{\Delta} = \tfrac{1}{2}ab\sinh u_c = \tfrac{1}{2}bc\sinh u_a = \tfrac{1}{2}ca\sinh u_b. \tag{3.9b}$$

Alternative formulations may be derived by using the generalized triangle properties, mainly the relation $u_a=u_b+u_c$. Expressions (3.7)–(3.9) are listed on Table I (without the restriction $a>b$). For integer indices, we recover the result given in (2.6).

Result (3.7) was already known.[15] It is a particular case of a more general formula involving products of J functions (Ref. 3, pp. 691 and 694). Result (3.8) was calculated only in

TABLE I. Definite integrals involving three Bessel functions of the first kind.

$$I=\int_0^\infty xJ_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx \quad [\mathrm{Re}(\mu+\nu)>-1],$$

(i) $c>a+b$,	$I=0$,
(ii) $a+b>c>\lvert a-b\rvert$,	$I=\frac{1}{2\pi\Delta}\cos(\mu\phi_b-\nu\phi_a)$ (triangle case),
(iii) $\lvert a-b\rvert>c>0$	
if $a>b$	$I=-\frac{1}{2\pi\tilde{\Delta}}\sin\nu\pi e^{-\nu u_a-\mu u_b}$,
if $a<b$	$I=-\frac{\sin\mu\pi}{2\pi\tilde{\Delta}}e^{-\nu u_a-\mu u_b}$.

Notations: Cases (i) and (iii)
$a^2=b^2+c^2\pm 2bc\cosh u_a$, $b^2=c^2+a^2\pm 2ac\cosh u_b$, $c^2=a^2+b^2\pm 2ab\cosh u_c$,
$u_a,u_b,u_c>0$; + sign only for the largest length. If $a>b,c$ $u_a=u_b+u_c$.

Case (ii)
$a^2=b^2+c^2-2bc\cos\phi_a$, $b^2=c^2+a^2-2ac\cos\phi_b$, $c^2=a^2+b^2-2ab\cos\phi_c$,
$0<\phi_a,\phi_b,\phi_c<\pi$, $\phi_a+\phi_b+\phi_c=\pi$,
$2\Delta=bc\sin\phi_a=ca\sin\phi_b=ab\sin\phi_c$,
$2\tilde{\Delta}=bc\sinh u_a=ca\sinh u_b=ab\sinh u_c$.

particular cases and is new in this general form. As to result (3.9), it was found already, but in a very sophisticated form. As $b,c<a$, we may set $b=a\sin\varphi\cos\psi$, $c=a\cos\varphi\sin\psi$ $(0<\varphi,\psi<\pi/2)$, where $2c^2\tan^2\varphi$ (resp. $2b^2\tan^2\psi)=a^2-b^2-c^2\pm\sqrt{\delta}(\delta=a^4+b^4+c^4-2a^2b^2-2a^2c^2-2c^2b^2)$. Then, the formula of Ref. 15 reads[16]

$$\int_0^\infty xJ_\nu(bx)J_\mu(ax)J_{\mu+\nu}(cx)dx$$
$$=\int_0^\infty xJ_\nu(ax\sin\varphi\cos\psi)J_\mu(ax)$$
$$\times J_{\mu+\nu}(ax\sin\psi\cos\varphi)dx$$
$$=-\frac{2\sin\nu\pi}{\pi a^2}\left(\frac{\sin\varphi}{\cos\psi}\right)^\nu\left(\frac{\sin\psi}{\cos\varphi}\right)^{\mu+\nu}$$
$$\times\frac{1}{\cos(\varphi+\psi)\cos(\varphi-\psi)},$$

and after some manipulations we recover (3.9a). Nevertheless, the formula in Ref. 15 keeps some ambiguities; mainly it implies that $\cos(\phi+\psi)>0$ and it does not fix completely the determination of $\tan^2\varphi$ and $\tan^2\psi$ which must be the smallest of the two (positive) roots of each equation. Formula (3.9a) has no such ambiguity.

C. Integrals involving Bessel functions of the second kind

As Y_ρ is a linear combination of J_ρ and $J_{-\rho}$, the above results may be used to calculate integrals involving one Bessel function of the second kind, perhaps with further restrictions on the indices. Assuming always that $a>b$, we consider the two cases

$$J=\int_0^\infty xY_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx,$$
$$K=\int_0^\infty xJ_\mu(ax)Y_\nu(bx)J_{\mu+\nu}(cx)dx,$$

for all three positions of c: $a+b<c$, $a-b<c<a+b$, $0<c<a-b$.

Integral J involves as an intermediate step the integrals

$$\int xJ_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx$$

and

$$\int xJ_{-\mu}(ax)J_{\mu+\nu}(cx)J_\nu(bx)dx,$$

whence the integrals

$$\int xJ_{\mu+\nu}(\tilde{c}x)J_{\mu+\nu}(cx)dx$$

and

$$\int xJ_\nu(\tilde{b}x)J_\nu(bx)dx$$

for some $\tilde{c}$, $\tilde{b}$ which are of the same kind as the ω_1 and A of formula (3.5)–(3.6). Provided $\mathrm{Re}(\mu+\nu)>-1$ and $\mathrm{Re}\,\nu>-1$, all calculations may be completely carried out as we deal only with the distribution δ. The same holds for integral K provided $\mathrm{Re}(\mu+\nu)>-1$ and $\mathrm{Re}\,\nu>-1$. Results are listed on Table II and some particular cases which already appeared in the literature are checked in the next section. Most of the results are new.

TABLE II. Definite integrals involving two Bessel functions of the first kind and one of the second kind for $a>b$. Notations are the same as in Table I.

$(a>b)$

$$J=\int_0^\infty xY_\mu(ax)J_\nu(bx)J_{\mu+\nu}(cx)dx \quad (\mathrm{Re}\,\mu+\nu>-1,\ \mathrm{Re}\,\nu>-1),$$

$$K=\int_0^\infty xJ_\mu(ax)Y_\nu(bx)J_{\mu+\nu}(cx)dx \quad (\mathrm{Re}\,\mu+\nu>-1,\ \mathrm{Re}\,\mu>-1),$$

(i) $c>a+b$,

$$J=-\frac{1}{2\pi\tilde{\Delta}}e^{-\nu u_a+\mu u_b},$$
$$K=-\frac{1}{2\pi\tilde{\Delta}}e^{\nu u_a-\mu u_b},$$

(ii) $a+b>c>|a-b|$,

$$J=-K=\frac{1}{2\pi\Delta}\sin(\nu\phi_a-\mu\phi_b)\ \text{(triangle case)},$$

(iii) $|a-b|>c>0$,

$$J=-K=\frac{\cos\pi\nu}{2\pi\tilde{\Delta}}e^{-\mu u_b-\nu u_b}.$$

Finally, the last possibility

$$L=\int_0^\infty xJ_\mu(ax)J_\nu(bx)Y_{\mu+\nu}(cx)dx$$

it involves the intermediate integral

$$\int_0^\infty xJ_{\mu+\nu}(\tilde{c}x)Y_{\mu+\nu}(cx)dx,$$

i.e., a principal part. In that case the integration is hopeless, except in some particular cases which are already known.

We just set the general equation. From (3.3)–(3.5) we get

$$L=\frac{2}{\pi^2}\mathrm{Re}\int_0^\infty e^{i\nu\pi}e^{-i\nu\theta}\left(\frac{a-be^{i\theta}}{a-be^{-i\theta}}\right)^{(\mu+\nu)/2}$$
$$\times\left(\frac{\omega_1}{c}\right)^{\mu+\nu}\frac{\mathrm{P.P.}}{c^2-\omega_1^2}d\theta$$
$$-\frac{2\sin\nu\pi}{\pi^2}\int_0^{\ln a/b}e^{-\nu u}\left(\frac{a-be^u}{a-be^{-u}}\right)^{(\mu+\nu)/2}$$
$$\times\left(\frac{A}{c}\right)^{\mu+\nu}\frac{\mathrm{P.P.}}{c^2-A^2}du, \tag{3.10}$$

where

$$\omega_1^2=a^2+b^2-2ab\cos\theta=(a-be^{i\theta})(a-be^{-i\theta}),$$
$$A^2=a^2+b^2-2ab\cosh u=(a-be^u)(a-be^{-u})$$

[cf. Eq. (3.5b)]. The integrals can be calculated only in very special cases, for example $\nu=0$ (so that the second term on the right-hand side disappears).

IV. CHECKING PARTICULAR RESULTS

A. Integrals involving three J functions

(i) When $\mu=\nu=0$, we get

$$I=\int_0^\infty xJ_0(ax)J_0(bx)J_0(cx)dx$$
$$=\begin{cases}\dfrac{1}{2\pi\Delta} & \text{in the triangular case},\\ 0 & \text{otherwise},\end{cases}$$

where Δ is the area of the triangle (Ref. 3, p. 696, formula 9 with $\nu = 0$).

(ii) When $\mu = 0$ and $\nu \neq 0$,

$$I = \int_0^\infty x J_0(ax) J_\nu(bx) J_\nu(cx)\,dx$$

$$= \begin{cases} 0 & \text{if} \quad c > a+b \\ \dfrac{\cos(\nu\phi_a)}{\pi bc \sin\phi_a} & \text{if} \quad |a-b| < c < a+b \\ -\dfrac{\sin \nu\pi e^{-\nu u_a}}{\pi bc \sinh u_a} & \text{if} \quad 0 < c < |a-b| \quad b < a, \\ 0 & \qquad\qquad\qquad\qquad\quad b > a \end{cases}$$

(cf. Ref. 3, pg. 695, formula 8 with $\mu = 0$).[17]

For $\nu = 1/2$,

$$I = \frac{2}{\pi}\int_0^\infty J_0(ax)\sin bx \sin cx\,dx$$

$$= \frac{1}{\pi}\int_0^\infty J_0(ax)[\cos(b-c)x - \cos(b+c)x]\,dx$$

is the cosine Fourier transform of $J_0(ax)$ and can be checked with tables of integral transforms.[18]

(iii) When $a = b$ and $\mu = (\lambda + n)/2$, $\nu = (\lambda - n)/2$ (Ref. 3, p. 674, formula 11),

$$I = \int_0^\infty x J_{(\lambda+n)/2}(ax) J_{(\lambda-n)/2}(ax) J_\lambda(cx)\,dx$$

$$= \begin{cases} = \dfrac{\cos(n\phi_a)}{\pi a^2 \sin 2\phi_a} & \text{if } c < 2a \quad \text{(isosceles triangle)} \\ = 0 \quad \text{if } c > 2a. \end{cases}$$

As $\cos\phi_a = c/2a$, the result may be rewritten

$$I = \frac{2}{\pi b (4a^2 - c^2)^{1/2}} T_n\left(\frac{c}{2a}\right) \quad \text{if } c < 2a$$

$$= 0 \text{ otherwise}$$

(where T_n is the Tchebyscheff polynomial).

B. Integrals involving two J functions and one Y function

(i) When $\mu = 0$,

$$\int_0^\infty x Y_0(ax) J_\nu(bx) J_\nu(cx)\,dx$$

$$= \begin{cases} -\dfrac{1}{2\pi\tilde{\Delta}} e^{-\nu u_a} & \text{if} \quad c > a+b \\ \dfrac{1}{2\pi\Delta}\sin\nu\phi_a & \text{if} \quad |a-b| < c < a+b \\ \dfrac{sg(a-b)}{2\pi\Delta}\cos\pi\nu e^{-\nu u_a} & \text{if} \quad 0 < c < |a-b|, \end{cases}$$

where we have removed condition $a > b$

$$\left(sg(a-b) = \begin{cases} +1 & \text{if} \quad a > b \\ -1 & \text{if} \quad a < b \end{cases}\right).$$

At this stage, a comment is in order. In a recent preprint Askey, Koornwinder, and Rahman[19] give the result for the integral $\int_0^\infty Y_\nu(ax) J_\nu(bx) J_\nu(cx) x^{\nu+1}\,dx$, and they agree with us for $\nu = 0$. The result outside the triangle is reported in Ref. 3 (p. 695, formula 5 with $\mu = 0$) but in an incorrect manner. The largest length must be a $(a > b + c)$ or b $(b > a + c)$ but not c like in their hypothesis. It is in contradiction with their formula 13, p. 696 for the isosceles case. The above result may be tested for $\nu = \frac{1}{2}$ by using the Fourier cosine transform of $Y_0(x)$.[14]

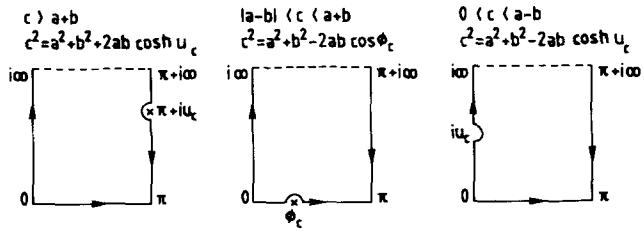


FIG. 2. Contours in the θ-plane for the calculation of L $(\nu = 0)$ for the three possible c's.

In the case of the isosceles triangle and for $\mu = \nu = 0$ we get two different integrals, listed in Ref. 3 (p. 695, formulas 12 and 13)

$$\int_0^\infty x Y_0(bx) J_0(ax) J_0(ax)\,dx$$

$$= 0 \qquad \text{if} \quad b < 2a$$

$$= \frac{2}{\pi b (b^2 - 4a^2)^{1/2}} \quad \text{if} \quad b > 2a,$$

$$\int_0^\infty x Y_0(ax) J_0(ax) J_0(bx)\,dx$$

$$= 0 \qquad \text{if} \quad b < 2a$$

$$= -\frac{2}{\pi b (b^2 - 4a^2)^{1/2}} \quad \text{if} \quad b > 2a,$$

(ii) When $\mu = \nu = \lambda/2$ (Ref. 3, p. 673, formula 10),

$$\int_0^\infty x J_{\lambda/2}(ax) Y_{\lambda/2}(ax) J_\lambda(cx)\,dx$$

$$= -\frac{2}{\pi c}\frac{1}{(c^2 - 4a^2)^{1/2}} \quad \text{if} \quad c > 2a$$

$$= 0 \quad \text{if} \quad c < 2a \quad \text{(isosceles triangle)}.$$

C. Calculation of some particular L integrals for $\nu = 0$

It was already calculated in (4.1)–(4.2) (with exchange of a and c). We may derive it directly. From (3.10) we have

$$L = \frac{2}{\pi^2}\,\mathrm{Re}\int_0^\pi \left(\frac{a - be^{i\theta}}{a - be^{-i\theta}}\right)^{\mu/2}\left(\frac{\omega_1}{c}\right)^\mu \frac{\text{P.P.}}{c^2 - \omega_1^2}\,d\theta.$$

This integral can be calculated by using a path in the θ plane which depends on the relative value of a,b,c (see Fig. 2). The two principal parts integrals along the imaginary axes do not contribute (pure imaginary result). The path at infinity gives zero and finally the integral is (to a $\pm$ sign) the real part of the only residue near the integration contour.

V. SUMMARY AND CONCLUSION

In this paper, we have shown that it is possible to get integrals of three Bessel functions when the arguments either satisfy the triangle inequalities or not. The key point of the proof is the use of the normalization relations between the

J_λ's which are well-defined distributions. Since the mathematical theory of the distributions is rather recent as compared to the theory of Bessel functions, it is not astonishing to discover new results by using a new technique. We hope that this will be an incentive for other Bessel buffs to study other properties unknown up to now.

ACKNOWLEDGMENTS

We thank our colleagues R. Lacaze and J. M. Normand who helped us during the elaboration of our work and M. L. Mehta for a careful reading of the manuscript and various discussions.

We are grateful to the referee for his interest and for many useful comments.

[1]G. Cohen-Tannoudji, A. Morel, and H. Navelet, Nuovo Cimento **48**, 1075 (1966).
[2]F. Henyey, G. L. Kane, J. Pumplin, and M. H. Ross, Phys. Rev. **182**, 1579 (1969).
[3]See for example, I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980).
[4]D. A. Varshalovich, A. N. Moskaliev, and V. K. Khersonskii, *Quantum Theory of Angular Momenta* (Science Academy USSR, Leningrad, 1975) (in Russian).
[5]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 943.
[6]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2.
[7]J. H. Graf, Math. Ann. **XLIII**, 142 (1883).
[8]G. N. Watson, *Theory of Bessel Functions* (Cambridge U.P., Cambridge, England, 1922).
[9]Ref. 6, p. 90, formula 8.
[10]G. Petiau, *La théorie des fonctions de Bessel* (CNRS, Paris, 1955) (in French), p. 204, formulas 90 (with $\rho = -1$) and 93.
[11]S. Schindler, SIAM J. Math. Anal. **4**, 367 (1973).
[12]Y_λ is the notation in Watson (Ref. 8) and Bateman (Ref. 6). In Gradshteyn and Ryzhik, the Bessel function of the second kind is denoted by N_λ.
[13]A. L. Dixon and W. L. Ferrar, Quart. J. Math. **4**, 193 (1933), formula (7.41).
[14]Ref. 10, p. 94, formula 137.
[15]W. N. Bailey, Proc. London Math. Soc. **40**, 37 (1935).
[16]This formula is reproduced in Ref. 3, p. 674, formula 14 with the wrong sign.
[17]The last result (ii) has a wrong sign in Ref. 3 as can be checked with $\nu = \frac{1}{2}$. The case $a < b$ (giving null result) has not been investigated.
[18]For example, see A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1953), Vol. 1, p. 43, and following.
[19]For the preprint, write either to T. Koornwinder centrum voor Wiskunde en Informata, Kruislaan 413, 1098 SJ Amsterdam, the Netherlands, or to R. Askey, Department of Mathematics, University of Wisconsin, Madison, Wisconsin 53706.

ERRATA

Erratum: Functional integrals as integrals on locally noncompact groups with generalized measures [J. Math. Phys. 25, 1412 (1984)]

N. N. Meiman
Maxim Gorky nab. 4/22, ap. 57, Moscow 11 3127, USSR

(Received 17 July 1984; accepted for publication 27 July 1984)

PACS numbers: 03.65. − w, 02.20. + b, 99.10. + g

The following are corrections to be made to the article cited above.

On page 1412 in the right-hand column, the fourth line from the bottom should read

$$d\sigma(g') = \exp[-(i/\hbar)S(g')]d\mu(g').$$

On page 1415, the first line after Eq. (1.13) should read $g \in G'$.

On page 1415, the second line of Eq. (1.16) should read

$$\exp\frac{i}{\hbar}[S_\mu(g') - S_\mu(g'_0)].$$

On page 1417, in the right-hand column, the fourth line from the top should read as follows: band being the curves $x_\mp(t)$.

On page 1417, in the right-hand column, the last line should read as follows: $\mu_H(\{x_-,x_+\}_{x^i}^{x^f})$ of a channel $\{x_-,x_+\}_{x^i}^{x^f}$ is the solution of.

On the page 1418, the third term on the right-hand side of Eq. (2.16) should read

$$\frac{1}{2}\left(i\hbar\frac{\partial A}{\partial x}\right).$$

On page 1420, the first of the right-hand sides of Eq. (2.37) should read

$$\int (x^{k+1} - x^k)^2 K_0(x^{k+1},t^{k+1}|x^k t^k)dx^k.$$

On page 1421, the limit on the second integral in Eq. (2.45) should read $(\gamma\delta^{-1}\hat{\gamma}^{-1})$.

On page 1421, the second γ on the line just below Eq. (2.45) should read $\hat{\gamma}$.

On page 1422, Eq. (3.2) should read $d\mu^{(1)}$ instead of $d\mu^{(4)}$.

On page 1422, the numerator of the fraction on the right-hand side of Eq. (3.13) should read $(x^f - x^i)^2$.

On page 1422, in the right-hand column, the fourth line from the bottom should read $d\mu_{H_1}(\gamma)$ instead of $d\delta\mu_{H_1}(\gamma)$.

On page 1423, the first integral on the right-hand side of Eq. (3.15) should read

$$\int_{t^i}^{t^f}\frac{1}{2}m\dot{\gamma}_y^2\,dt.$$

On page 1423, the second of Eqs. (3.27) should read

$$\lim\left|\frac{\mu_H(\{x_-,x_+\}_{x^i}^{x^f})}{\mu_{H_0}(\{x_-,x_+\}_{x^i}^{x^f})}\right| = 1.$$

On page 1424, the left-hand side of Eq. (3.34) should read $\sigma_\kappa\left[\{x_-,x_+\}_{x^i}^{x^f};\gamma',\bar{\gamma}_0\right]$.

On page 1424, Eq. (3.35) should read

$$\lim_{\to}\sum_{(\gamma')}\varphi(\gamma')\exp\left[\frac{i}{\hbar}S_H(\gamma') - \frac{i}{\hbar}S_H(\bar{\gamma}_0)\right]\mu_{H_0}\left[\{x_-,x_+\}_{x^i}^{x^f} - \gamma'\right].$$

On page 1424, Eq. (3.36) should read

$$\lim\sum_{(\gamma')}\exp\left[\frac{i}{\hbar}S_H(\gamma') - \frac{i}{\hbar}S_H(\bar{\gamma}_0)\right]\mu_{H_0}\left[\{x_-,x_+\}_{x^i}^{x^f} - \gamma'\right].$$

On page 1425 left-hand column, the beginning of the eighth line from the top should read $t^f \gtrless t^i$.

On page 1425, the left-hand side of Eq. (3.42) should read $K(x_1,t_1|H|x_2,t_2)$.

On page 1426, the first term on the right-hand side of the third of Eqs. (3.60) should read $K_{V_1}(0,t^f|0,t^i)$.

On page 1427, the integral on the right-hand side of Eq. (3.68) should read

$$\int_{U(\bar{\gamma}_k^{k+1},M)} d\mu_{H_0}(\gamma),$$

and the last term should be deleted.

On page 1427, Eq. (3.69) from the first exp on the right-hand side should read

$$\exp\left[-\frac{i}{\hbar}\int_{\bar{\gamma}_k^{k+1}} V\,dt\right]\int_{U(\bar{\gamma}_k^{k+1},N^\alpha)}\exp\left[\frac{i}{\hbar}\int_{\bar{\gamma}_k^{k+1}} V\,dt - \frac{i}{\hbar}\int_\gamma V\,dt\right]d\mu_{H_0}(\gamma).$$

On page 1429, the right-hand side of Eq. (4.9) should read

$$K_V^N(x^f t^f|\,x^i t^i) = \int\exp\left[\frac{i}{\hbar}\int_{H_V}(\bar{\gamma}_0^N)\right]d\sigma(\bar{\gamma}_0^N).$$

On page 1429, the left-hand side of Eq. (4.10) should read $d\sigma(\bar{\gamma}_0^N)$.

On page 1430, the line 18th from the top should read $2\Delta x^{k+1}$ instead of Δx^{k+1}.

 0022-2488/84/113357-01$02.50

Erratum: Inverse scattering for optical couplers. Exact solution of Marchenko equations [J. Math. Phys. 25, 1900 (1984)]

G. P. Bava and G. Ghione
Dipartimento di Elettronica, Politecnico di Torino, Torino, Italy

(Received 25 July 1984; accepted for publication 9 August 1984)

PACS numbers: 03.80. + r, 02.30.Rz, 42.80.Lt, 99.10. + g

Formula (2) must be written as

$$h_{mn} = k_{mn}(x)\exp[-(-1)^n i\beta_0 x] \quad (m \neq n). \tag{2}$$

In Sec. III B, point 3, instead of $r(k - k_0)$ one must read $r(k - k_0/2)$.

Formula (29) must be written as

$$R(x) = -2iCu(x)\frac{a_1}{a_1 \cosh 2a_1 x + \eta \sinh 2a_1 x}. \tag{29}$$

In Figs. 3–7 the scale of the vertical axis of the diagrams showing $|R|$ should be doubled (i.e., 2.0 instead of 1.0 and so on).

Erratum: Linear response theory revisited. IV. Applications [J. Math. Phys. 25, 1391 (1984)]

P. Vasilopoulos and C. M. Van Vliet
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Québec, H3C 3J7, Canada and Department of Electrical Engineering, University of Florida, Gainesville, Florida 32611

(Received 17 July 1983; accepted for publication 27 July 1984)

PACS numbers: 03.60. + w, 72.10. − d, 72.10.Di, 99.10. + g

(1) Using (4.2) it can be easily shown that the two terms in the brackets of Eq. (4.1) are equal. Thus, for all temperatures (4.1) becomes

$$\sigma_{yx}^{\mathrm{nd}}(0) \equiv \sigma_{yx} = \frac{e}{B\Omega}\sum_{k,N}(N+1)\langle n_N\rangle_{\mathrm{eq}} \times (1 - \langle n_{N+1}\rangle_{\mathrm{eq}})(1 - e^{-\beta\hbar\omega_0});$$

this, together with *only* $f_N(1 - f_N) \approx \beta^{-1}\delta(\varepsilon - \varepsilon_F')$ [cf. Eq. (4.3)] brings about the following changes: in Eqs. (4.6), (4.14), and (4.15), $2N + 1$ is replaced by $2(N + 1)$ and Eq. (4.16) is divided by 2; further, Eq. (4.18) is replaced by [the two terms in the brackets of Eq. (4.17) cancel each other]

$$\sigma_{xx}^{\mathrm{nd}}(0) \equiv \sigma_{xx} = 0,$$

for all temperatures and the comment following Eq. (4.18) does not apply.

(2) The ω_L/N_0^{-P} [third line after Eq. (3.32)] should be replaced by $\omega_L/\omega_0 - P$.

 0022-2488/84/113358-01$02.50