The notion of essential locality for nonlocalizable fields

J. Bümmerstede and W. Lücke

Institut für Theoretische Physik der Technischen Universität Clausthal, Clausthal-Zellerfeld, Germany
(Received 2 December 1974)

The classification of extension of the field commutator outside the light cone suggested by Constantinescu and Taylor is analyzed and shown to be to a large extent mathematically equivalent to the notion of essential locality, introduced in a recent paper by the present authors. Simple model fields are constructed which disprove the interpretation given by Constantinescu and Taylor. Essential locality is shown to hold for the two-point function of every scalar Hermitian field, including the massless case. It is, moreover, shown to be weaker than locality and independent of the other Wightman axioms. Unfortunately, essential locality turns out to be unstable under limits. In order to indicate the possibility that there are essentially local fields which do not fall into Jaffe's class and the commutator of which is concentrated in the closed light cone, Jaffe's concept of strict localizability is generalized. As a by-product it is indicated that local fields (in the generalized sense) may have extreme high energy behavior.

1. INTRODUCTION

There are good reasons[1] to expect that certain physically relevant field theories must be formulated on test spaces of entire functions, i.e., the field operators are nonlocalizable in the sense of Jaffe.[2] As nobody was able to give a consistent definition of *support* for generalized functions on such test spaces til now, the standard formulation of microcausality,[3] also called *locality*, has no meaning in that case. As a consequence, most of the important achievements of Wightman field theory such as Haag–Ruelle scattering, *PCT*, spin, and statistics do not *a priori* apply to the nonlocalizable case.

In a recent paper[4] we suggested a new attempt of circumventing this problem, which already proved successful for a derivation of the Haag–Ruelle–Hepp scattering formalism.[5] As a substitute for the axiom of locality we introduced an axiom of "essential locality" which demands some local continuity property of the field commutator on the light cone (see Sec. 2 of the present paper).

Independently Constantinescu and Taylor[6] introduced an "*order of extension of the commutator bracket outside the light cone*" for the nonlocalizable case. The main purpose of the present paper is to compare this concept with the notion of essential locality.

In Sec. 2 we briefly review the notion of essential locality and stress its Lorentz invariance. In Sec. 3 we show that essential locality is no restriction on the two-point function of a scalar field. The mathematical connection between essential locality and the above mentioned order of extension will be clarified in Sec. 4. In Sec. 5 we construct fields, the commutator bracket of which has finite order γ of extension outside the light cone but does not decrease as fast as predicted by Constantinescu and Taylor. In order to indicate the possibility that even for finite order γ the field commutator may not extend outside the light cone, we extend Jaffe's definition of strict localizability in Sec. 6 and construct corresponding c-number examples. Section 7, finally, is devoted to a short discussion of the results.

In order to keep the amount of technical considerations at a minimum, we shall not work out the most general versions of our results.

2. ESSENTIAL LOCALITY REVIEWED

For convenience we use the well-known Gel'fand spaces[7]

$$S^s(R^n)=S^{\beta_1,\cdots,\beta_n},\quad \beta_1=\beta_2=\cdots=\beta_n=s,\quad 0\leq s<1.$$

We adopt the usual notation:

$$\|\chi\|=\left(\sum_{j=1}^{n}(\chi^j)^2\right)^{1/2},\quad \chi=(\chi^1,\cdots,\chi^n),$$

$$Z_+^n=\{\alpha=(\alpha^1,\cdots,\alpha^n):\alpha^j \text{ nonnegative integer}\}$$

$$|\alpha|=\sum_{j=1}^{n}\alpha^j,\quad \alpha!=\alpha^1!\cdots\alpha^n!,\quad \chi^n=\prod_{j=1}^{n}(\chi^j)^{n^j},$$

$$\varphi^{(\alpha)}(\chi)=\left(\frac{\partial}{\partial\chi^1}\right)^{\alpha^1}\cdots\left(\frac{\partial}{\partial\chi^n}\right)^{\alpha^n}\varphi(\chi^1,\cdots,\chi^n),$$

$$U_\epsilon(M)=\{\chi\in R^n:\|\chi-\chi'\|<\epsilon \text{ for at least one } \chi'\in M\}.$$

In order to stress the Lorentz invariance of essential locality (see Definition 3 below), we use the modified

Definition 1: Let M be a closed subset of R^n and S a subset of $S^s(R^n)$. Then S is called *locally bounded* on M in $S^s(R^n)$ iff there is a positive constant A such that for every nonnegative integer N

$$\sup_{\varphi\in S}\sup_{\chi\in M}\sup_{\alpha\in Z_+^n}A^{-|\alpha|}\alpha^{-s\alpha}\|\chi\|^N|\varphi^{(\alpha)}(\chi)|<\infty.$$

In the original version of Definition 1 we used $\sup_{\chi\in U_\epsilon(M)}$ instead of $\sup_{\chi\in M}$ with suitable $\epsilon>0$ depending on S. Since, in the present paper, we restrict to $s<1$ both versions are equivalent:

Lemma 1: Let M be some subset of R^n and $A, C_0, C_1, \cdots$ some sequence of positive constants. Moreover, let the entire function φ over R^n fulfil the inequalities:

$$\sup_{\chi\in M}\sup_{\alpha\in Z_+^n}A^{-|\alpha|}\alpha^{-s\alpha}\|\chi\|^N|\varphi^{(\alpha)}(\chi)|<C_N,\quad N\in Z_+.$$

Then there is a sequence of positive constants

$B, D_0, D_1, \cdots$ independent of φ such that also

$$\sup_{\chi\in U_\epsilon(M)} \sup_{\alpha\in Z_+^n} (2A)^{-|\alpha|}\,\alpha^{-s\alpha}\|\chi\|^N\,|\varphi^{(\alpha)}(\chi)|$$

$$< D_N \exp(B\epsilon^{1/(1-s)})$$

holds for all $N\in Z_+$ and $\epsilon>0$.

Proof: Since φ is entire, we may write

$$\varphi^{(\alpha)}(\chi+\chi') = \sum_{\beta\in Z_+^n} \frac{1}{\beta!}\chi'^{\beta}\varphi^{(\alpha+\beta)}(\chi).$$

Hence for $\chi\in M$ and $\chi'\in R^n$ with $\|\chi'\|<\epsilon$ we have

$$\|\chi+\chi'\|^N\,|\varphi^{(\alpha)}(\chi+\chi')|$$

$$< (1+\epsilon)^N \sum_{\beta\in Z_+^n} \frac{\epsilon^{|\beta|}}{\beta!}(1+\|\chi\|)^N\,|\varphi^{(\alpha+\beta)}(\chi)|$$

$$< 2^{N+n}\ (C_0+C_N)(1+\epsilon)^N \sup_{\beta\in Z_+^n} \frac{\beta^{s\beta}}{\beta!}(4\epsilon A)^{|\beta|}(2A)^{|\alpha|}\alpha^{s\alpha},$$

which directly implies the statement of Lemma 1. ∎

The notion of essential locality is based on the following

Definition 2: Let M be a closed subset of R^n and F a generalized function on $S^s(R^n)$. Then, for nonnegative $s_0\le s$, F is called *locally continuous* on M with respect to $S^{s_0}(R^n)$ iff $\sup_{\varphi\in S}|F(\varphi)|$ is finite for every $S\subset S^{s_0}(R^n)$ that is locally bounded on M in $S^{s_0}(R^n)$.

Now we can formulate "essential locality":

Definition 3: Let $A(x)$ be a scalar Hermitian quantum field defined on the dense invariant subset D of the Hilbert space of states H which satisfies all the Wightman axioms,[3] except locality, where the Schwartz space $\mathcal{S}(R^4)$ is substituted by the Gel'fand space $S^s(R^4)$, $s\ge 0$. Then $A(x)$ is called *essentially local*, iff for arbitrary $\Phi, \Psi\in D$ the expectation value of the field commutator $\langle\Phi|[A(x),A(y)]|\Psi\rangle$, uniquely defined[4] as a generalized function over $S^s(R^8)$, is locally continuous on $V_8\equiv\{(x,y)\in R^8 : x-y\in\bar V\}$ with respect to $S^s(R^8)$.

Nonlocalizable power series of the (massive) free field, for example, proved to be essentially local with respect to every space $S^s(R^4)$ on which they can be defined at all.[8]

As a by-product, Lemma 1 exhibits a basic difficulty connected with local continuity: Let $F_1, F_2, \cdots$ be a sequence of generalized functions on $S^s(R^n)$, weakly convergent to F in $S^s(R^n)'$. If all the F_j are locally continuous on the same closed subset M of R^n, then also F is so as long as $s>1$. For $s<1$ this does not hold in general, as can be immediately seen by choosing F_j to be the characteristic function of $U_j(0)\subset R^n$. Then, by Lemma 1, all F_j are locally continuous on $\{0\}$ with respect to $S^s(R^n)$, but F is not. Even worse, since every generalized function over $S^s(R^n)$ that is locally continuous on $\{0\}$ with respect to $S^s(R^n)$ can be approximated by finite linear combinations of derivatives of the δ function,[9] F may be also represented as a limit of Schwartz distributions with supports equal to $\{0\}$. This indicates that proofs of essential locality for nonlocalizable fields will by no means be trivial.

3. ESSENTIAL LOCALITY OF ARBITRARY TWO-POINT FUNCTIONS

Given a strictly localizable scalar field $A(x)$, then the vacuum expectation value of the field commutator $\langle\Omega|[A(x),A(y)]|\Omega\rangle$ is well known to vanish in the region $(x-y)^2<0$, whether $A(x)$ fulfils locality or not. In this section we shall see that a similar result holds for the nonlocalizable case as well.

Let us introduce the notation

$$G_\epsilon \equiv \{p\in R^4 : \min\{|p^0|, \|\mathbf{p}\|\}<\epsilon\}$$

for $\epsilon>0$ and denote by $\mathcal{S}(G_\epsilon)$ the set $\{\varphi\in\mathcal{S}(R^4) : \operatorname{supp}\varphi\subset G_\epsilon\}$ endowed with the trace topology relative to $\mathcal{S}(R^4)$. Then the essential point is to realize the following fact:

Lemma 2: There is a partition of the unity $\{\tilde k_r\}\subset\mathcal{D}(R^4)$ and there are proper Lorentz transformations Λ_r for which

$$\tilde g_\varphi(p) \equiv \sum_r \tilde k_r(\Lambda_r^{-1}p)\tilde\varphi(\Lambda_r^{-1}p)$$

converges in $\mathcal{S}(G_2)$ and depends continuously on $\tilde\varphi\in\mathcal{S}(R^4)$.

Proof: For every $p\in R^4$ one may choose a proper Lorentz transformation Λ_p with

(i) $\Lambda_p p\in G_{3/2}$,

(ii) $|(\Lambda_p^{-1})^\nu_\mu|\le 16(1+\|p\|)$,

(iii) $\Lambda_p q\in G_2$ if $\|p-q\|<16^{-2}(1+\|p\|)^{-1}$.

By standard techniques one may construct a partition of the unity $\{\tilde k_r\}$ fulfilling the conditions

$$\|p-q\|<16^{-2}(1+\|p\|)^{-1} \quad \text{for } p,q\in\operatorname{supp}\tilde k_r$$

and

$$\sum_r \sup_{n\le\|p\|<n+1} |D_p^\alpha \tilde k_r(p)| < P_\alpha(n) \quad \text{for } n\in Z_+,$$

$\alpha\in Z_+^4$, where the P_α are suitable polynomials. Thus, if we choose an arbitrary sequence of vectors $p_r\in\operatorname{supp}\tilde k_r$ and for every p_r a proper Lorentz transformation $\Lambda_r\equiv\Lambda_{p_r}$ according to (i)—(iii), then $\sum_r\tilde k_r(\Lambda_r^{-1}p)\tilde\varphi(\Lambda_r^{-1}p)$ has all required properties. ∎

As a direct consequence we have the following:

Corollary 1: A Lorentz invariant Schwartz distribution $\tilde F$ is tempered if $\tilde F\in\mathcal{S}(G_2)'$.

Finally, we can prove:

Theorem 1: Every Lorentz invariant odd generalized function F on $S^s(R^4)$, $0\le s<1$, is locally continuous on the closed light cone $\bar V$ with respect to $S^s(R^4)$.

Proof: We essentially use the technique of Ref. 8. Let us first consider the case $s>0$. Since $\operatorname{supp}\tilde F\subset\bar V$, we may multiply $\tilde F(p)$ by $\exp[-(1+p^2)^{1/(2s)}]$ and thus, by Corollary 1, get a tempered distribution. Therefore, there are $N', C'\in Z_+$ with $\|\tilde F\|^s_{N'}<C'$, where

$$\|\tilde\varphi\|^s_{N'} \equiv \sup_{p\in U_1(\bar V)} \max_{|\alpha|\le N'} \{(1+\|p\|^{N'})$$

$$\times |D_p^\alpha(\tilde{\psi}(p)\exp[(1+p^2)^{1/(2s)}])|\}$$

$(=\infty$ possible).

Moreover, there is a $N'' \in Z_+$ such that

$$\|\tilde{\psi}\|_{N'}^s \leq C''\|\tilde{\psi}\|_{N''} \sup_{p\in U_1(\bar{V})\cap \mathrm{supp}\tilde{\psi}} \exp[(1+p^2)^{1/(2s)}]$$

for some suitable finite constant C'', where

$$\|\tilde{\varphi}\|_{N''} \equiv \max_{p\in R^4} \max_{|\alpha|\leq N''} (1+\|p\|^{2N''})\,|\tilde{\varphi}^{(\alpha)}(p)|$$

for $\tilde{\varphi} \in \mathcal{S}(R^4)$.

Hence, if we choose a partition of the unity $\{\tilde{h}_r\} \subset \mathcal{D}(R^1)$ with $\mathrm{supp}\,\tilde{h}_r \subset (r-1, r+1)$, $\tilde{h}_r(t) = \tilde{h}_0(t-r)$ and define tempered Schwartz distributions

$$\tilde{F}_{r,k}(p) \equiv (p^2)^{-k} h_r(p^2)\tilde{F}(p)$$

for either $k, r-1 \in Z_+$ or $k = r = 0$, we have

$$\|\tilde{F}_{r,k}(p)\|_N < C(1+k^N)(\max\{1, r-1\})^{-k} \times \exp[(r+2)^{1/(2s)}] \tag{3.1}$$

for suitable $N, C \in Z_+$ not depending on r, k. Keeping in mind that the $F_{r,k}(x)$, being tempered Lorentz-invariant odd distributions, have supports contained in $\bar{V}$, choose a multiplier $g \in \mathcal{O}_M(R^4)$ fulfilling the conditions

$$g(x) = \begin{cases} 1 & \text{for } x \in V \\ 0 & \text{for } x \notin U_1(V), \end{cases}$$

$$\sup_{x\in R^4} \max_{|\alpha|\leq 2N} |D_x^\alpha g(x)| < \infty.$$

Introducing the notation

$$\tilde{\psi}_k(p) \equiv (2\pi)^2 \int dp'(p'^2)^k \tilde{\psi}(p')\tilde{g}(p-p'),$$

we then get

$$|\int dx F(x)\psi(x)| \leq \|\tilde{F}_{0,0}\|_N \|\tilde{\psi}_0\|_N + \sum_{r=1}^{\infty} \|\tilde{F}_{r,k(r)}\|_N \|\tilde{\psi}_{k(r)}\|_N, \tag{3.2}$$

$$\|\tilde{\psi}_k\|_N \leq \max_{|\alpha|\leq N} \int dx\,|(1+\Delta_x^N)[x^\alpha g(x)\Box_x^k \psi(x)]|$$

$$(\Delta_x \equiv \partial_0^2 + \cdots + \partial_3^2,\ \Box_x \equiv \partial_0^2 - \partial_1^2 - \partial_2^2 - \partial_3^2).$$

With the definition

$$B_\epsilon \equiv \{\psi \in S^s(R^4) : \max_{x\in U_1(\bar{V})} \max_{\alpha\in Z_+^4} \epsilon^{-|\alpha|}\alpha^{-s\alpha} \times (1+\|x\|^{N+5})\,|D^\alpha\psi(x)| < 1\}$$

we have for suitable constants $A, C_{\epsilon,N}$

$$\sup_{\psi\in B_\epsilon} \|\tilde{\psi}_k\|_N < C_{\epsilon,N}(\epsilon A)^{2k} k^{2sk}. \tag{3.3}$$

Thus, choosing ϵ sufficiently small and arranging $k(r)$ such that

$$r^{1/(2s)} < k(r) \leq 1 + r^{1/(2s)},$$

we conclude with (3.1) and (3.2) that

$$\sup_{\psi\in B_\epsilon} |\int dx\, F(x)\psi(x)| < \infty. \tag{3.4}$$

Summarizing: For small $\epsilon > 0$ [not depending on $F(x)$] (3.4) holds for arbitrary $F(x)$ fulfilling the requirements of Theorem 1. Since, for arbitrary $\lambda > 0$, also $F(\lambda x)$ fulfills these requirements if $F(x)$ does, (3.4) must be valid for arbitrary $\epsilon > 0$. Hence, by Lemma 1, $F(x)$ is essentially local on $\bar{V}$.

For $s = 0$ we have to modify the above proof in the following way. Equation (3.1) has to be replaced by

$$\|\tilde{F}_{r,k}(p)\|_{N(r)} < C_{r,N(r)}(1+k^{N(r)})(\max\{1, r-1\})^{-k}. \tag{3.1'}$$

Equation (3.2) still holds, with the only modification that N depends on r now. This r dependence makes us substitute for B_ϵ an arbitrary set S that is locally bounded on $\bar{V}$ in $S^s(R^4)$; i.e., there are numbers $C_{S,N(r)}$, A_S such that

$$\sup_{\psi\in S} \|\tilde{\psi}_k\|_{N(r)} < C_{S,N(r)} A_S^k. \tag{3.3'}$$

Thus, choosing $k = k(r)$ of sufficiently rapid increase for $r \to \infty$, we see that

$$\sup_{\psi\in S} |\int dx\, F(x)\psi(x)| < \infty \tag{3.4'}$$

for every set S that is locally bounded on $\bar{V}$ in $S^s(R^4)$; i.e., $F(x)$ is locally continuous on $\bar{V}$ with respect to $S^s(R^4)$ in the case $s = 0$, too. ∎

Theorem 1 shows that essential locality does not impose any restriction on the two-point function of a scalar field $A(x)$ on $S^s(R^4)$, $0 \leq s < 1$. A direct consequence is essential locality of arbitrary generalized free fields (compare Ref. 6). We conclude this section with the remark that Theorem 1 in connection with Ref. 8 indicates a way how to overcome Constantinescu and Taylor's difficulties concerning power series of the *massless* free field.

4. CONSTANTINESCU AND TAYLOR'S CLASSIFICATION OF EXTENSION OF THE COMMUTATOR OF A NONLOCALIZABLE FIELD OUTSIDE THE LIGHT CONE

Constantinescu and Taylor[6] introduced spaces $T_{n,j}(s)$ of entire functions over $R^{4(n-1)}$ which, by Lemma 1, may be described as follows:

For $M \subset R^n$ define the topological vector space S_M^s of entire functions over R^n by

(i) $\varphi \in S_M^s$ iff there is a positive constant A such that

$$\|\varphi\|_{M;s}^{A,N} \equiv \sup_{\chi\in M} \sup_{\alpha\in Z_+^n} A^{-|\alpha|}\alpha^{-s\alpha}\|\chi\|^N\,|\varphi^{(\alpha)}(\chi)|$$

is finite for every $N \in Z_+$.

(ii) $\varphi_\lambda \to \varphi$ in the topology of S_M^s iff there is a positive constant A such that

$$\|\varphi_\lambda - \varphi\|_{M;s}^{A,N} \to 0 \text{ for every } N \in Z_+.$$

Then $T_{n+1,j}(s) = S^s{}_{\{\hat{\xi}\in R^{4n};\, \xi_j \in \bar{V}\}}$.

Given a field $A(x)$ as in Definition 3, Constantinescu and Taylor[6] defined the supremum of all $\gamma \geq 1$ such that for arbitrary $n-1 \in Z_+$ and $1 \leq j \leq n$ the generalized

function[10]

$$C_{n+1,j}(\xi_1, \cdots, \xi_n) \equiv \langle \Omega | A(x_1) \cdots A(x_{j-1})[A(x_j), A(x_{j+1})]A(x_{j+2}) \cdots \times A(x_{n+1}) | \Omega \rangle,$$

$$\xi_l \equiv x_l - x_{l+1} \text{ for } 1 \leq l \leq n,$$

can be extended from $S^0(R^{4n})$ to $T_{n+1,j}(1 - 1/\gamma)$ as the *order of extension of the commutator bracket outside the light cone.*

In order to compare this classification with our definition of essential locality, let us first recall the following technical result:

Lemma 3: Let B be a positive constant and $s \in (0, 1)$. Then there is a nontrivial nonnegative entire function ψ_B over R^1 and a positive constant A such that

$$\sup_{t \in R^1} \sup_{j \in Z_+} A^{-j} j^{-sj} \exp(B |t|^{1/(1-s)}) |\psi_B^{(j)}(t)| < \infty. \tag{4.1}$$

Proof: See Ref. 7, p. 192. ■

Corollary 2: Let M be a closed subset of R^n and $s \in (0, 1)$. Then $S^s(R^n)$ is dense in S_M^s. Hence a generalized function F on $S^s(R^n)$ is locally continuous on M with respect to $S^s(R^n)$ iff it can be extended to S_M^s.

Proof: Without loss of generality we assume $0 \in M$. Let $\varphi \in S_M^s$. Then there is a positive constant A such that

$$\|\varphi\|_{M;s}^{A;N} \text{ is finite for every } N \in Z_+.$$

Hence, by Lemma 1, there are positive constants $B, D_0, D_1, \cdots$ with

$$\sup_{\alpha \in Z_+^n} (2A)^{-|\alpha|} \alpha^{-s\alpha} \|\chi\|^N |\varphi^{(\alpha)}(\chi)| < D_N \exp\left(B \sum_{j=1}^n |\chi^j|^{1/(1-s)}\right).$$

Choosing ψ_B according to Lemma 3 and defining

$$h_R(\chi) \equiv \prod_{j=1}^n \left(\int_{-R}^R dt \psi_B(\chi^j - t)\right) \text{ for } R > 0,$$

we get $h_R \varphi \in S^s(R^n)$ and, for A' big enough,

$$\lim_{R \to \infty} \|(1 - h_R)\varphi\|_{M;s}^{A';N} = 0 \text{ for } N \in Z_+.$$

Since $\varphi \in S_M^s$ was arbitrary, this directly implies the statements of Corollary 2. ■

Corollary 3: Let $A(x)$ be a scalar Hermitian field as in Definition 3, which is not neccessarily essentially local, however. Suppose D to be the set of all vectors of the form

$$\sum_{n=0}^N \int d\hat{x}\, \varphi_n(\hat{x}) A(x_1) \cdots A(x_n)\Omega,$$

where $\varphi_n \in S^s(R^{4n})$, $N \in Z_+$. Then $A(x)$ is essentially local iff $C_{n+1,j}(\xi_1, \cdots, \xi_n)$ can be extended to $T_{n+1,j}(s)$ for every $n \in Z_+$, $1 \leq j \leq n$.

Proof: With our special choice for D the field $A(x)$ is essentially local iff[4]

$$\langle \Omega | A(x_1) \cdots A(x_{j-1})[A(x_j), A(x_{j+1})] \times A(x_{j+2}) \cdots A(x_{n+1}) | \Omega \rangle \quad (n, j \in Z_+,\ 1 \leq j \leq n)$$

is locally continuous on $\{\hat{x} \in R^{4(n+1)} : x_j - x_{j+1} \in \bar{V}\}$ with respect to $S^s(R^{4(n+1)})$. This holds iff $C_{n+1,j}$ is locally continuous on $\{\hat{\xi} \in R^{4n} : \xi^j \in \bar{V}\}$ with respect to $S^s(R^{4n})$. Finally, by Corollary 2, the latter holds iff $C_{n+1,j}$ can be extended to $T_{n+1,j}(s)$. ■

Corollary 3 means that the Constantinescu–Taylor order is $\geq (1 - s)^{-1}$ iff the field $A(x)$ on $S^s(R^4)$ is essentially local. Hence the interpretation of this order given by Constantinescu and Taylor also applies to essential locality. For finite order γ they claim that the field commutator neccessarily extends outside the light cone and that optimal bounds on spacelike cones are given by $C \exp(-A\|x\|^\gamma)$, where C, A are suitable positive constants depending on the cone. This interpretation is wrong, as will be shown in the subsequent section.

5. NONLOCAL TEMPERED FIELDS

Let $A_0(x)$ be the scalar Hermitian free field for mass $m > 0$. Let $\mathcal{H}$ be its Fock space and Ω the no-particle state. Consider the tempered field

$$A_g(x) \equiv A_0(x) + [h * : A_0^2 :](x),$$

$$\tilde{h}(p) = \tilde{g}(p^2), \quad \tilde{g} \in \mathcal{D}_M(R^1)$$

on the dense invariant domain $D \subset \mathcal{H}$ consisting of all vectors Φ of the form

$$\Phi = \sum_{n=0}^N \int d\hat{x} A_g(x_1) \cdots A_g(x_n) \varphi_n(\hat{x})\Omega,$$

where $\varphi_n \in \mathcal{S}(R^{4n})$, $N \in Z_+$. If there are positive constants C, A with

$$0 < \tilde{g}(t) < C \exp(-A|t|^{1/s_0}), \quad t \in R^1, \tag{5.1}$$

for some $s_0 \in (0, 1)$, then we can easily prove that $A_g(x)$, while fulfilling all the other Wightman axioms, is not local.

To this end let us check the expectation value

$$E(x_1, x_2) \equiv \langle \Phi_{\hat{f}} | [A_g(x_1), A_g(x_2)] | \Omega \rangle \in \mathcal{S}(R^8)',$$

where $\Phi_{\hat{f}}$ is the two-particle state corresponding to the wavefunction $\hat{f}_2(\mathbf{p}_1, \mathbf{p}_2) = \omega_{\mathbf{p}_1}\omega_{\mathbf{p}_2}\hat{f}(\mathbf{p}_1, \mathbf{p}_2) \in \mathcal{D}(R^6)$. An elementary calculation using Wick's theorem shows that

$$k(y) \equiv \int dx_1\, dx_2\, E(x_1, x_2)\varphi(x_1 - y)\psi(x_2)$$

$$= \text{const} \int_{p_j^0 = \omega_{\mathbf{p}_j}} d\mathbf{p}_1\, d\mathbf{p}_2 \hat{f}^*(\mathbf{p}_1, \mathbf{p}_2) \int dq\, \delta(q^2 - m^2)\epsilon(q^0) \times \tilde{g}((q - p_1)^2)\tilde{g}((q + p_2)^2)\tilde{\varphi}(p_1 - q)\tilde{\psi}(p_2 + q) \exp[iy(p_1 - q)]$$

for $\varphi, \psi \in \mathcal{S}(R^4)$ and $y \in R^4$ (* denotes the complex conjugate). By (5.1) and since

$$|(p \pm q)^2| > 2(p^0 - \|\mathbf{p}\|)\|\mathbf{q}\| - 2m^2$$

for p, q on the (positive) mass shell, $k(y)$ is an entire function of order[11] $\leq (1 - s_0)^{-1}$. Assume $A_g(x)$ to be local. Then, since for arbitrary $\varphi, \psi \in \mathcal{D}(R^4)$, $\hat{f} \in \mathcal{D}(R^6)$ the entire function $k(y)$ must vanish for all y outside some sufficiently large neighborhood of the light cone (depending on φ and ψ), it must be identically zero. But, since $\tilde{g} \neq 0$, this cannot hold for all $\varphi, \psi \in \mathcal{D}(R^4)$. Hence the assumption is wrong; $A_g(x)$ is not local.

$A_g(x)$ is not even essentially local with respect to

$S^s(R^4)$, if $s > s_0$, at least for $s_0 \leq \frac{1}{2}$. This may be demonstrated as follows: Assume $A_g(x)$ to be essentially local. Taking

$$\varphi(x) \equiv \psi(-x) \equiv \prod_{j=0}^{3} \psi_B(x^j), \quad B > 0,$$

with not neccessarily positive ψ_B fulfilling (4.1), we get bounds of the form $C \exp(-A\|y\|^{1/(1-s)})$ for $k(y)$ on spacelike cones. Hence, again, we can conclude[12] that k is identically zero. Since the set of all ψ_B, $0 < B$, fulfilling (4.1) is dense in $\mathcal{S}(R^1)$, this leads to the same contradiction as above.

Given $s_0 \in (0, 1)$, there is (Ref. 7, p. 192) an entire function

$$\tilde{g}(t) = \sum_{r=0}^{\infty} c_r (-t)^r$$

of order $\leq s_0^{-1}$ fulfilling (5.1). For such $\tilde{g}$ the nonlocal tempered field $A_g(x)$ is the $N \to \infty$ limit of the local tempered fields

$$A_0(x) + \sum_{r=0}^{N} c_r \Box^r : A_0^2 : (x)$$

in the topology of $S^{s_-}(R^4)'$ for $s_- \in (0, s_0/2)$. This also shows that $A_g(x)$ is essentially local with respect to $S^{s_-}(R^4)$. Thus, by Corollary 3, the order γ of $A_g(x)$ fulfills the inequality

$$(1 - s_0)^{-1} \geq \gamma \geq (1 - s_0/2)^{-1}.$$

Assume the interpretation of γ given by Constantinescu and Taylor to be correct. Then inequalities of the form

$$|\langle \Omega | A(x_1) \cdots [A(x_j), A(x_{j+1})] \cdots A(x_n) | \Omega \rangle |$$
$$< C_{j,n} \exp(-A\|x_j - x_{j+1}\|^{\gamma})$$

hold for $j < n$ in all points $(x_1, \cdots, x_j, x_{j+1}, \cdots, x_n)$ which are totally spacelike[3] simultaneously with $(x_1, \cdots, x_{j+1}, x_j, x_{j+2}, \cdots, x_n)$. Since $\gamma > 1$, a well-known result by Borchers and Pohlmeyer[13] directly implies locality of $A_g(x)$. This contradiction shows that the bound for the extension of the field commutator outside the light cone predicted by Constantinescu and Taylor—which would be extremely powerful—unfortunately does not hold in general.

6. STRICTLY LOCALIZABLE GENERALIZED FUNCTIONS

Although it is obvious that finiteness of the order γ of a field $A(x)$ on a Jaffe space[2] $\mathcal{C}_\varepsilon$ implies definite extension of the field commutator outside the light cone, we doubt whether this holds for all fields on $S^s(R^4)$, $0 \leq s < 1$, as well. Unfortunately, one does not have an interpretation of "extension of the commutator bracket outside the light cone" in terms of familiar notions for arbitrary nonlocalizable fields. This problem does not arise, however, if the "nonlocalizable" field $A(x)$ on $S^s(R^4)$, $0 \leq s < 1$, has an extension to a field $A'(x)$ on some larger test space $\mathcal{C}$ which allows the definition of "support" of a generalized function on $\mathcal{C}$. Definite extension of $[A(x), A(y)]$ outside V_8 would then be naturally interpreted as $\operatorname{supp}[A(x), A(y)] \not\subset V_8$.

Therefore, we shall introduce a generalization of Jaffe's definition of strict localizability. Presumably the extended class of strictly localizable fields contains fields with finite order γ and $\operatorname{supp}[A(x), A(y)] \subset V_8$. As a preliminary step in this direction we shall give corresponding c-number examples in Lemma 5.

In the spirit of Jaffe[2] a field theory is called strictly localizable, if the field operators are defined on a test space $\mathcal{C}$ containing "sufficiently many" test functions with compact support. We call such spaces "local" and generalized functions on such spaces *strictly localizable*:

Definition 4: Let $\mathcal{C}$ be a test space over R^n. $\mathcal{C}$ is called *local* iff the following conditions are fulfilled:

(i) $\mathcal{C}(\cup_{j \in I} \mathcal{O}_j)$ is[14] contained in the closed linear hull of $\cup_{j \in I} \mathcal{C}(\mathcal{O}_j)$ for every family of open sets $\mathcal{O}_j \subset R^n$, $j \in I$ (I an arbitrary index set).

(ii) $\mathcal{C}(\mathcal{O})$ is a dense subset of $L^2(\mathcal{O})$ for every open subset $\mathcal{O} \subset R^n$.

The characterization of these spaces as "local" may be justified by the following:

Remarks: (1) As usual, the *support* of a generalized function F on $\mathcal{C}$ can be defined as the complement of the set of all points having a neighborhood, where $F = 0$.

(2) Then $F(\varphi) = 0$ for every $\varphi \in \mathcal{C}$ with $\operatorname{supp}\varphi \cap \operatorname{supp} F = \emptyset$.

(3) F is completely known, if it is known on some neighborhood of every point $\chi \in \operatorname{supp} F$.

Moreover, local test spaces allow the following:

Definition 5: Let $\mathcal{C}_1$, $\mathcal{C}_2$ be test spaces over R^n and assume $\mathcal{C}_1$ to be local. Let $\mathcal{O}$ be an open subset of R^n, and let $/\!\!\eta$ be a mapping from $\mathcal{C}_1(\mathcal{O})$ into $\mathcal{C}_2$. Then we say that $/\!\!\eta$ can be extended to a mapping from $\{F \in \mathcal{C}_1' : \operatorname{supp} F \subset \mathcal{O}\}$ into $\mathcal{C}_2'$ iff there is a linear continuous mapping $/\!\!\eta'$ from $\mathcal{C}_2$ into $\mathcal{C}_1$ with

$$\int d\chi (/\!\!\eta \varphi_1) \varphi_2 = \int d\chi \varphi_1 /\!\!\eta' \varphi_2$$

for all $\varphi_1 \in \mathcal{C}_1(\mathcal{O})$, $\varphi_2 \in \mathcal{C}_2$.

The *extension of* $/\!\!\eta$ is defined by

$$\int d\chi (/\!\!\eta F)(\chi) \varphi(\chi)$$
$$\equiv \int d\chi F(\chi) /\!\!\eta' \varphi(\chi) \quad \text{for } \varphi \in \mathcal{C}_2.$$

Remark: Definition 5 is allowed since, by condition (ii) of Definition 4, $/\!\!\eta' \varphi(\chi)$ is unique for $\chi \in \mathcal{O}$.

A useful criterion for condition (i) of Definition 4 is given by

Lemma 4: Let $\mathcal{C}$ be a test space over R^n fulfilling the following conditions:

(i) for every pair of precompact open sets $\mathcal{O}_1$, $\mathcal{O}_2$ with $\bar{\mathcal{O}}_1 \subset \mathcal{O}_2 \subset R^n$ there is a multiplier k in $\mathcal{C}$ with

$$k(\chi) = \begin{cases} 1 & \text{for } \chi \in \mathcal{O}_1 \\ 0 & \text{for } \chi \notin \mathcal{O}_2; \end{cases}$$

(ii) for every open set $\mathcal{O} \subset R^n$, $\mathcal{C}(\mathcal{O}) \cap \mathcal{D}$ is dense in $\mathcal{C}(\mathcal{O})$. Then condition (i) of Definition 4 is fulfilled.

Proof: See Appendix A. ■

The Gel'fand spaces[7] S^s, $s > 1$, the Schwartz spaces[15] $\mathcal{S}$, $\mathcal{D}$ and the Jaffe spaces[2] $\mathcal{C}_g$ are well-known examples for local test spaces. The class of local test spaces as specified by Definition 4 is considerably larger, however. This may be illustrated by the following example: For $s \in (0, 1)$, $n \in Z_+$, let $\mathcal{C}^s(R^n)$ be the set of all C^∞-functions φ fulfilling

$$\|\varphi\|^s_{N,\epsilon,A} \equiv \sup_{\substack{\alpha,\beta \in Z_+^n \\ |\beta| \leq N}} \sup_{\substack{\chi \in R^n \\ \|\chi\| \geq \epsilon|\alpha|}} A^{-|\alpha|}\alpha^{-s\alpha}\|\chi\|^N \left|\varphi^{(\alpha+\beta)}(\chi)\right|$$

for arbitrary positive A, ϵ, N, endowed with the natural topology given by the family of norms $\|\varphi\|^s_{N,\epsilon,A}$.

Corollary 4: The test space $\mathcal{C}^s(R^n)$, $0 < s < 1$, is local. The test spaces $\mathcal{D}(R^n)$ and $S^{s_0}(R^n)$, $0 \leq s_0 < s$, are contained in $\mathcal{C}^s(R^n)$ and their topology is finer than the trace topology relative to $\mathcal{C}^s(R^n)$.

Proof: See Appendix B. ■

The spaces $\mathcal{C}^s(R^4)$ are well suited for field theory, since all the usual operations on generalized functions on $\mathcal{C}^s$ such as partial differentiation, multiplication with polynomials, coordinate transformations, etc., are defined by Definition 5 in the familiar way. Operator products $A(x_1)\cdots A(x_n)$, finally, have a natural definition on $\mathcal{C}^s(R^4)\otimes_\pi \cdots \otimes_\pi \mathcal{C}^s(R^4)$ (compare Ref. 4, Appendix).

The motivation for introducing these spaces is given by the following:

Lemma 5: Let $s \in (\frac{1}{2}, 1)$. Then there is a generalized function F on $\mathcal{C}^s(R^4)$ with $\operatorname{supp} F \subset \bar{V}$, which cannot be extended to a generalized function on $S^s(R^4)$ and the Fourier transform of which is an entire function which cannot be bounded over R^4 by $C \exp\|p\|^{1/s}$ for any $C > 0$.

Proof: See Appendix C. ■

7. CONCLUSIONS

We have seen that essential locality is a Lorentz-invariant concept and that it does not impose any restriction on the two-point function of a scalar Hermitian field, even in the massless case. The Constantinescu–Taylor order γ of "extension of the commutator bracket outside the light cone" was briefly reviewed and proved (under some mild restriction) to be $\geq 1/(1-s)$ iff the field is essentially local with respect to $S^s(R^4)$, $s \in (0, 1)$. This caused our special interest in the Constantinescu–Taylor interpretation of γ.

For arbitrary $s_0 \in (0, \frac{1}{2}]$ we constructed nonlocal tempered fields which are not even essentially local with respect to any $S^s(R^4)$ with $s > s_0$, but which are essentially local and limits of local tempered fields with respect to every $S^{s_-}(R^4)$, $s_- \in (0, s_0/2)$. These examples were used to disprove the bound for the extension of the commutator bracket outside the light cone predicted by Constantinescu and Taylor. In addition, these examples show for $s > 0$

(i) Even for strictly localizable fields locality is a definitely stronger condition than essential locality with respect to some $S^s(R^4)$, $s < \frac{1}{4}$. Therefore, we believe the extension of the Haag–Ruelle–Hepp scattering theory to essentially local fields, given in Ref. 4, to be really nontrivial.

(ii) Essential locality, just as locality in the localizable case, is independent of the other Wightman axioms.

(iii) Unfortunately, essential locality is not stable under limits: Even limits of local fields need not be essentially local with respect to $S^s(R^4)$.

Finally, we presented a generalization of Jaffe's concept of strict localizability in order to indicate the possibility that even for finite order γ the field commutator may not extend outside the light cone, as indicated by c-number examples. This generalization does not seem to exclude the possibility that local fields, strictly localizable in the generalized sense, may increase stronger than exponentially in momentum space.

We believe that essential locality, despite its nonstability under limits, is a useful concept. This has been demonstrated in Ref. 4, and we hope to be able to prove analyticity properties of the S matrix for essentially local fields in a forthcoming paper.

APPENDIX A: PROOF OF LEMMA 4

By condition (ii) of the lemma we may assume without loss of generality the $\mathcal{O}_j$ to be precompact and I to be finite, say $I = \{1, \cdots, N\}$, $N - 1 \in Z_+$. We prove Lemma 4 for $N = 2$ only. For arbitrary N the statement follows by simple induction, then.

Let $\varphi \in \mathcal{C}(\mathcal{O}_1 \cup \mathcal{O}_2)$. Then there is a precompact open set $\mathcal{O} \subset R^n$ with $\bar{\mathcal{O}} \subset \mathcal{O}_2$ and $\operatorname{supp}\varphi \subset \mathcal{O}_1 \cup \mathcal{O}$. By condition (i) of the lemma we may choose a multiplier k in $\mathcal{C}(R^4)$ with $\operatorname{supp} k \subset \mathcal{O}_2$ and $k(\chi) = 1$ for $\chi \in \mathcal{O}$. With the definitions

$$\varphi_1(\chi) \equiv [1 - k(\chi)]\varphi(\chi), \quad \varphi_2(\chi) \equiv k(\chi)\varphi(\chi)$$

we then have $\varphi = \varphi_1 + \varphi_2$ and $\varphi_j \in \mathcal{C}(\mathcal{O}_j)$, $j = 1, 2$. Hence every $\varphi \in \mathcal{C}(\mathcal{O}_1 \cup \mathcal{O}_2)$ is in the linear hull of $\mathcal{C}(\mathcal{O}_1) \cup \mathcal{C}(\mathcal{O}_2)$.

APPENDIX B: PROOF OF COROLLARY 4

We only have to prove that $\mathcal{C}^s(R^n)$ is local, because the second part of the corollary is obvious. As $\mathcal{D}(\mathcal{O})$ is contained in $\mathcal{C}^s(\mathcal{O})$ and dense in $L^2(\mathcal{O})$ for every open set $\mathcal{O} \subset R^n$, condition (ii) of Definition 4 is fulfilled. To prove that condition (i) of the definition is fulfilled as well, we use Lemma 4. Since we can use suitable multipliers in $\mathcal{D} \subset \mathcal{C}^s$ in order to fulfil condition (i) of Lemma 4, we just have to prove that condition (ii) of Lemma 4 holds:

Choose a function $d(\chi) \in S^{1+s/2}(R^n)$ with

$$d(\chi) = \begin{cases} 1 & \text{for } \|\chi\| < \frac{1}{2} \\ 0 & \text{for } \|\chi\| > 1 \end{cases}$$

and define $d_r(\chi) \equiv d(\chi/r)$ for positive $r \in Z_+$. These functions d_r fulfill the inequality

$$\sup_{r-1 \in Z_+} \sup_{\substack{\alpha,\beta \in Z_+^n \\ |\beta| < N}} \sup_{\substack{\chi \in R^n \\ \|\chi\| \geq \epsilon|\alpha|}} A^{-|\alpha|}\alpha^{-s\alpha}\left|d_r^{(\alpha+\beta)}(\chi)\right|$$

$$\leq \sup_{\substack{\alpha,\beta \in Z_+^n \\ |\beta| < N}} \sup_{\chi \in R^n} A^{-|\alpha|}(\epsilon|\alpha|)^{-|\alpha+\beta|}\alpha^{-s\alpha}\left|d^{(\alpha+\beta)}(\chi)\right| < \infty$$

for arbitrary $A, \epsilon, N > 0$. Now, let $\varphi \in \mathcal{C}^s(R^n)$. Then we have

$$\sup_{r-1 \in Z_+} \|d_r \varphi\|^s_{A,\epsilon,N} < \infty$$

for arbitrary $A, \epsilon, N > 0$. Since, on the other hand, every derivative of $d_r(\chi)\varphi(\chi)$ converges uniformly to that of $\varphi(\chi)$ on every compact subset of R^n for $r \to \infty$, we conclude: $d_r(\chi)\varphi(\chi) \underset{r\to\infty}{\to} \varphi(\chi)$ in the topology of $\mathcal{C}^s(\mathcal{O})$ for every $\varphi \in \mathcal{C}^s(\mathcal{O})$. Since $d_r \in \mathcal{D}(R^n)$, this means that condition (ii) of Lemma 4 is fulfilled.

APPENDIX C: PROOF OF LEMMA 5

Let us define the entire function ($\tau \in R^1$)

$$\tilde{f}_s(\tau) \equiv \prod_{j=1}^{\infty} [1 + j^{-2s}\tau^2 \exp(2i\tau)]$$
$$= \sum_{j=0}^{\infty} c_j (-i\tau)^{2j} \exp(2ij\tau).$$

Since there are positive C, A such that[16]

$$|c_j| < CA^j j^{-2sj}$$

the entire function $\tilde{f}_s(\tau)$ is the Fourier transform of the generalized function

$$f_s(t) = \sum_{j=0}^{\infty} c_j \delta^{(2j)}(t - 2j)$$

on $\mathcal{C}^s(R^1)$. Therefore,

$$F_s(x) \equiv \sum_{j=0}^{\infty} c_j \delta^{(2j)}(x^0 - 2j)\delta(\mathbf{x})$$

fulfills all the requirements of Lemma 5 if $\tilde{f}_s(\tau)$ cannot be extended to a generalized function on $S^s(R^1)$. Choose a nonnegative function $\tilde{\varphi} \in \mathcal{D}(R^1)$ with

$$\tilde{\varphi}(\tau) = \begin{cases} 1 & \text{for } |\tau| < \frac{1}{2} \\ 0 & \text{for } |\tau| > 1 \end{cases}$$

and define

$$\tilde{\varphi}_{n,\lambda}(\tau) \equiv \tilde{\varphi}(\lambda n^2(\tau - n\pi)) \left(\prod_{j=1}^{\infty} (1 - j^{-2s}(\tau/\lambda)^2) \right)^2$$

for $n \in Z_+$, $\lambda > 0$. Clearly,[16] $\sum_{n=1}^{\infty} \tilde{\varphi}_{n,\lambda}$ converges in $\tilde{S}^s(R^1)$ for every $\lambda > 0$. Therefore, it is sufficient to prove that

$$\sum_{n=1}^{\infty} \int d\tau \tilde{f}_s(\tau)\tilde{\varphi}_{n,\lambda}(\tau) \qquad \text{(C1)}$$

does not converge for some $\lambda > 0$:

For $\tau \in R^1$ with $\cos 2\tau \geq 0$ we have the inequality

$$|\tilde{f}_s(\tau) - f_s(n\pi)| = \left| \int_{n\pi}^{\tau} d\rho \tilde{f}_s'(\rho) \right|$$
$$= \left| \int_{n\pi}^{\tau} d\rho \tilde{f}_s(\rho) \sum_{j=1}^{\infty} [1 + j^{-2s}\tau^2 \exp(2i\tau)]^{-1} \right.$$
$$\left. \times 2j^{-2s}(\tau + i\tau^2) \exp(2i\tau) \right|$$
$$< |\tau - n\pi| \max_{\substack{\rho \in R^1 \\ |\rho - n\pi| \leq |\tau - n\pi|}} [|\tilde{f}_s(n\pi)| + |\tilde{f}_s(n\pi) - f_s(\rho)|]$$
$$\times 4(1 + \tau^2) \sum_{j=1}^{\infty} j^{-2s}.$$

For sufficiently large $\lambda > 0$, consequently, $\lambda n^2 |\tau - n\pi| < 1$ implies $|\tilde{f}_s(\tau) - \tilde{f}_s(n\pi)| < \frac{1}{2}\tilde{f}_s(n\pi)$. Thus for sufficiently large $\lambda > 0$ there are positive constants C, A with

$$\mathrm{Re} \int d\tau \tilde{f}_s(\tau)\tilde{\varphi}_{n,\lambda}(\tau) > \tfrac{1}{2}\tilde{f}_s(n\pi) \int d\tau \tilde{\varphi}_{n,\lambda}(\tau) > \tfrac{1}{2}\tilde{f}_s(n\pi)\lambda^{-1}n^{-2}$$
$$\times \min_{\substack{\tau \in R^1 \\ \lambda n^2 |\tau - n\pi| < 1/2}} \prod_{j=1}^{\infty} [1 - j^{-2s}(\tau/\lambda)^2]^2 > C \exp(An^{1/s})$$

and (C1) does not converge.

[1] O. Steinmann, Commun. Math. Phys. 18, 179 (1970).
[2] A. Jaffe, Phys. Rev. 158, 1454 (1967).
[3] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[4] J. Bümmerstede and W. Lücke, Comm. Math. Phys. 37, 121 (1974).
[5] We are indepted to Professor R. F. Streater for pointing out to us an error in Ref. 4: In order to prove the correct transformation properties of the asymptotic states, one needs more detailed estimates of type given for the localizable case by R. F. Streater, J. Math. Phys. 8, 1685 (1967).
[6] F. Constantinescu and J. G. Taylor, J. Math. Phys. 15, 824 (1974).
[7] I. M. Gel'fand and G. E. Schilow, *Verallgemeinerte Funktionen* (VEB Deutsch. Verl. Wissen., Berlin, 1962), Vol. II.
[8] W. Lücke, J. Phys. A 7, 2258 (1974).
[9] J. Bümmerstede, thesis, in preparation.
[10] A. Rieckers, Int. J. Theor. Phys. 4, 55 (1971).
[11] B. A. Fuks, *Analytic Functions of Several Complex Variables* (Am. Math. Soc., Providence, R.I., 1963).
[12] R. P. Boas, *Entire Functions* (Academic, New York, 1954), Corollary 5.1.14.
[13] H.-J. Borchers and K. Pohlmeyer, Commun. Math. Phys. 8, 269 (1968).
[14] Generalizing the notation $\mathcal{S}(G_\epsilon)$ of Sec. 3, we define $\mathcal{C}(\mathcal{O})$ to be the set $\{\varphi \in \mathcal{C}: \mathrm{supp}\varphi \subset \mathcal{O}\}$ endowed with the trace topology relative to $\mathcal{C}$.
[15] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966).
[16] Ref. 7, Chap. IV, Sec. 8, No. 2.

The separating topology for the Lorentz group L

P. G. Vroegindeweij

Department of Mathematics, Technological University, Eindhoven, The Netherlands
(Received 7 June 1974)

Some properties of the Lorentz group L are presented if it is endowed with a topology induced by one of the topologies for the Minkowski space M, proposed by E. C. Zeeman.

1. PRELIMINARIES

Let M denote Minkowski space, the four-dimensional real vector space $\mathbf{R}^4$, provided with the indefinite quadratic form

$$Q(x)=x_0^2-x_1^2-x_2^2-x_3^2,$$

where $x=(x_0,x_1,x_2,x_3)\in M$. The vectors x of M are called *timelike* if $Q(x)>0$, *lightlike* (or *isotropic*) if $Q(x)=0$, and *spacelike* if $Q(x)<0$.

L is the full Lorentz group (all linear maps leaving Q invariant). L' is the orthochronous Lorentz group that is the subgroup of L whose elements preserve the sign of the first coordinate. $L_+^\uparrow$ is the subgroup of L' whose elements l have the property $\det l=+1$.

Using the canonical basis of $\mathbf{R}^4$ we introduce the *parity* p by

$$p=(p_{ij}),\quad 0\leq i,\ j\leq 3,$$

$$p_{00}=1,\quad p_{ii}=-1,\quad 1\leq i\leq 3,$$

and $p_{ij}=0$ for all $i\neq j$. We shall also use the *time reversal* $l=-p$.

Notice that $L/L_+^\uparrow\cong V_4$ where V_4 denotes Klein's four-group. By O_3^+ we mean the centralizer of p in $L_+^\uparrow$ that is to say the subgroup of $L_+^\uparrow$ whose elements r have the property $prp^{-1}=r$. The elements of 0_3^+ are called *pure rotations*.

Z is the subgroup of $L_+^\uparrow$ whose elements z have the form

$$z=\begin{pmatrix}\cosh\alpha & \sinh\alpha & 0 & 0\\ \sinh\alpha & \cosh\alpha & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}.$$

We introduce furthermore HS, being that subset of $L_+^\uparrow$ whose elements h have the property $php^{-1}=h^{-1}$. Notice that $h={}^th$ (th is the transposed of h); h is called *hyperbolic screw*. $L_+^\uparrow$ has no proper invariant subgroups, cf. Ref. 1.

Let $SL(2,\mathbf{C})$ be the group of unimodular 2×2 matrices over the complex numbers. As is known, there is a surjective homomorphism φ which induces an isomorphism

$$SL(2,\mathbf{C})/Z_2\cong L_+^\uparrow$$

where Z_2 is the set

$$\left\{\begin{pmatrix}1&0\\0&1\end{pmatrix};\ \begin{pmatrix}-1&0\\0&-1\end{pmatrix}\right\}.$$

The homomorphism φ can be described in the following way (cf. Ref. 2): Let $x=(x_0,x_1,x_2,x_3)\in\mathbf{R}^4$ and let $\tilde{x}$ denote the Hermitian matrix

$$\begin{pmatrix}x_0+x_1 & x_2-ix_3\\ x_2+ix_3 & x_0-x_1\end{pmatrix}\in M(2,\mathbf{C}).$$

Consider the bijection $f:\mathbf{R}^4\to M(2,\mathbf{C})$, given by $f(x)=\tilde{x}$. Let $l\in L_+^\uparrow$ and $s=\varphi^{-1}(l)\in SL(2,\mathbf{C})$. We have the relation $l\tilde{x}=s\tilde{x}s^*$, where $\tilde{x}\to l\tilde{x}$ is a Hermitian map.

Using matrix language, we may write

$$\begin{pmatrix}y_0+y_1 & y_2-iy_3\\ y_2+iy_3 & y_0-y_1\end{pmatrix}=\begin{pmatrix}\alpha&\beta\\ \gamma&\delta\end{pmatrix}\begin{pmatrix}x_0+x_1 & x_2-ix_3\\ x_2+ix_3 & x_0-x_1\end{pmatrix}\begin{pmatrix}\bar{\alpha}&\bar{\gamma}\\ \bar{\beta}&\bar{\delta}\end{pmatrix}$$

where $y=lx=(y_0,y_1,y_2,y_3)$ and $s=\begin{pmatrix}\alpha&\beta\\ \gamma&\delta\end{pmatrix}$.

We shall also use the group G, that is the group generated by L, the group T of translations of M and the group of multiplications by a positive scalar of the vectors of M. G' is the subgroup of G that we obtain by considering L' instead of L.

There is a partial order $\ll$ on M given by $x\ll y$ if and only if $Q(y-x)>0$ and $x_0<y_0$. Another partial order $<$ on M is given by $x<y$ if and only if $Q(y-x)\geq 0$ and $x_0\leq y_0$. We still need the relation $<\cdot$, given on M by $x<\cdot y$ if and only if $Q(y-x)=0$ and $x_0\leq y_0$. We introduce furthermore the sets:

$$C(x)=\{y\,|\,Q(y-x)=0\},$$
$$S(x)=\{y\,|\,Q(y-x)<0\},$$
$$I(x)=\{y\,|\,Q(y-x)>0\},$$
$$I^+(x)=\{y\,|\,x\ll y\},$$
$$I^-(x)=\{y\,|\,y\ll x\}.$$

C is the group of bijections of M, preserving the relation $\ll$.

Zeeman[3] proved that C and G' coincide. Zeeman's theorem has been generalized in several ways, cf. Refs. 4–9.

2. THE SEPARATING TOPOLOGY FOR MINKOWSKI SPACE M

Usually M is endowed with the Euclidean topology, but one can argue (Zeeman[3,10]) that this is objectionable for physical reasons. On the other hand, it is impossible to define a topology for M by means of the indefinite quadratic form Q in a way similar to the Euclidean topology by means of the definite quadratic form. In

Ref. 10 Zeeman has proposed several non-Euclidean topologies for M related to the Lorentz group L. Nanda[11–13] investigated them and added some more of this kind of topologies. All these topologies have the property that the corresponding group of autohomeomorphisms of M coincides with G and for that reason they seem to be physically significant. Unfortunately, they are very complicated from a topological point of view; for instance, they fail to satisfy the normal property and hence they are not metrizable. In this section we shall deal with that one of the topologies, proposed by Zeeman, that seems to be the most suitable for physics, cf. Ref. 9. We call it the *separating topology*. Similar topologies are also proposed by Cole[14] and Cel'nik.[15]

Let $d(x,y)$ denote the Euclidean metric

$$d(x,y)=\{(x_0-y_0)^2+(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2\}^{1/2}.$$

Given $x\in M$ and $\epsilon>0$, let $N_E^\epsilon(x)$ denote the Euclidean ϵ-neighbourhood of x, given by

$$N_E^\epsilon(x)=\{y\,|\,d(x,y)<\epsilon\}.$$

We introduce

$$N_s^\epsilon(x)=N_E^\epsilon(x)\cap(C(x)\setminus\{x\})^*,\quad x\in M$$

(by V^* we mean the complement of a set V).

Definition: The separating topology for M is the topology, given by the basis of open sets $N_s^\epsilon(x)$, $x\in M$.

We use the notations M_s for M with the separating topology and M_E for M with the Euclidean topology.

Remark: It is also possible to define our topology by using only the relations $\ll$, $<$, and $<\cdot$. That offers the possibility of introducing the separating topology in more general causal spaces, cf. Refs. 9, 16.

Let $x,\ y,\ z\in M$; $y\ll x\ll z$ and let us write

$$O_x(y,z)=I^+(y)\cap I^-(z)\cap(C(x)\setminus\{x\})^*.$$

Clearly the topology for M with basic open sets $O_x(y,z)$ is equivalent with the topology with basic open sets $N_s^\epsilon(x)$. Notice that M_s is a Hausdorff space; it satisfies the first axiom of countability and it is a separable space but it does not have a countable basis. However M_s is locally connected and pathwise connected it is not locally compact. From a physical point of view it seems to be interesting that on lightlike lines the discrete topology is induced and that on timelike lines and spacelike hyperplanes the Euclidean topology is induced, cf. Ref. 10.

Comparing M_E and M_s we still note the following properties:

(1) The set O is open in M_s and not in M_E if and only if for all $x\in O$ there is an $\epsilon>0$ such that $N_s^\epsilon(x)\subset O$ and there is an $x\in O$ with the property $(C(x)\setminus\{x\})\cap N_E^\epsilon(x)\cap O^*\neq\phi$ for all $\epsilon>0$.

(2) The subset X of M_s is compact in M_s if and only if X is compact in M_E and all $x\in X$ are isolated in $X\cap C(x)$ (with respect to M_E).

(3) The group of autohomeomorphisms of M_s is G.

For details we refer to Ref. 9.

3. THE SEPARATING TOPOLOGY FOR LORENTZ GROUP L

This is the main part of our paper; we shall investigate the topology for L induced by the separating topology for M. As is to be expected, M_s induces a topology for L, deviating from the usual Lie group topology, such as we obtain by considering L as a six-dimensional manifold in $\mathbf{R}$.[9] There are several ways to topologize a set of maps. In this section we shall deal with the topology of pointwise convergence. See e.g., Ref. 17.

A. Introduction

For each $x\in M_s$ and for every open set $O\subset M_s$ we define

$$(x,O)=\{l\in L_+^\uparrow\,|\,lx\in O\}.$$

Let L_s denote $L_+^\uparrow$, endowed with the topology that has the family of all sets (x,O) as a subbasis, and let L_E denote $L_+^\uparrow$, endowed with the topology, defined in a similar way as for L_s, but coming from M_E instead of M_s. The family of intersections of sets of the form (x,O) is a basis for the topological space L_s, each number of this basis having the form $\cap_{i=1}^n(x_i,O_i)$, where $x_i\in M_s$ and O_i is open in M_s. Notice that L_s is finer than L_E, for M_s is finer than M_E. As we shall show below, L_s is strictly finer. Notice furthermore that L_s is a Hausdorff space, for M_s has that property.

It is also possible to describe our topology by means of convergence of nets (see, e.g., Ref. 17, p. 77). To that end one can define: The net of Lorentz transformations (l_ν) converges to l in L_s if and only if $(l_\nu x)$ converges to lx for all $x\in M_s$. We shall say that a set $O\subset L_s$ is open if and only if every net (l_ν), converging to an element $l\in O$, is eventually in O. Remark that, if the net (l_ν) does not converge to l in L_E, it does not converge to l in L_s. As we shall show below, the converse is also true if we restrict ourselves to timelike vectors.

B. Properties of L_s

L_s is strictly finer than L_E. Example.

$$l_n=\begin{pmatrix}\cosh\frac{1}{n} & \sinh\frac{1}{n} & 0 & 0\\ \sinh\frac{1}{n} & \cosh\frac{1}{n} & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix},\quad x=\begin{pmatrix}1\\1\\0\\0\end{pmatrix},$$

l is the unit element of $L_+^\uparrow$. In L_E we find that (l_n) converges to l if $n\to\infty$, but $l_nx=e^{1/n}x$, and therefore (l_n) does not converge in L_s, for $l_n(x)\notin N_s^\epsilon(x)$, even for all n. Also in the case of spacelike vectors, there are nets converging in L_E, but not in L_s. The same sequence (l_n) as above, but applied to the spacelike vector $y=(1,1,0,1)$, gives us $l_ny\notin N_s^\epsilon(y)$ for all n.

Theorem 1: L_s and L_E induce the same topology on the subgroup O_3^+.

Proof: It suffices to prove that a net of pure rotations (r_ν), converging in L_E, also converges in L_s (with the same limit). Suppose that (r_ν) converges to r in L_E. Then we have for all x that eventually $\{(r_\nu x)\}\subset N_E^\epsilon(rx)$.

On the other hand, we know that all $r_\nu x$ are situated in the same spacelike hyperplane through rx and therefore

$$\{(r_\nu x)\} \cap (C(rx) \setminus \{rx\}) = \phi$$

and, consequently,

$$\{(r_\nu x)\} \cap N_s^\epsilon(rx) = \{(r_\nu x)\} \cap N_E^\epsilon(rx).$$

This means that $(r_\nu x)$ eventually belongs to $N_s^\epsilon(rx)$. In other words, (r_ν) converges to r in L_s. □

Corollary: L_s induces the same topology as L_E on every compact subgroup of $L_+^\uparrow$, because 0_3^+ is a maximal compact subgroup of L_E and consequently of L_s.

A *semitopological group* G is a topological space, provided with a group structure such that the product map $G \times G \to G$, given by $(a, b) \to ab$, $(a, b \in G)$, is separately continuous. See, e.g., Ref. 18.

Theorem 2: L_s is a semitopological group.

Proof: Suppose that (l_ν) converges to l, i.e., $(l_\nu x)$ converges to lx, $x \in M_s$. In particular, if we consider $l'x$ instead of x, then $(l_\nu l'x)$ converges to $ll'x$. Therefore, for all neighborhoods $O_{ll'}$ of ll' there is a neighborhood O_l of l such that $O_l \cdot l' \subset O_{ll'}$. On the other hand, we know that the elements of L_s are homeomorphisms of M_s and therefore it follows from $(l_\nu x)$ converges to lx that $(l' l_\nu x)$ converges to $l'lx$ for all $l' \in L$, i.e., for all $O_{ll'}$ there is a neighborhood O_l of l such that $l' O_l \subset O_{l'l}$. □

C. The main theorem

The definition of L_s uses the action of $L_+^\uparrow$ on M and the topology of M_s. Now we want to give an intrinsic definition of L_s, by comparing it with L_E. In Sec. 1 we have seen that $L_+^\uparrow$ is very close to $SL(2, \mathbb{C})$.

Lemma 1: For timelike vectors x, $(l_\nu x)$ converges to lx in M_E if and only if $(l_\nu x)$ converges to lx in M_s.

Proof: Obviously, convergence in M_s implies convergence in M_E. To prove the converse, we remark that the nets, converging in L_E and not in L_s, are exactly those having the property that there is an x such that eventually

$$(l_\nu x - lx,\ l_\nu x - lx) = 0 \quad \text{and} \quad l_\nu x \neq lx,$$

i.e.,

$$(l^{-1} l_\nu x - x,\ l^{-1} l_\nu x - x) = 0 \quad \text{and} \quad l^{-1} l_\nu x \neq x.$$

It is sufficient to consider only one timelike vector. We choose $x' = (a, 0, 0, 0)$ and note that it is possible to transform all timelike vectors, situated on the same hypersurface $(x, x) = a^2$, into $(a, 0, 0, 0)$ by a suitable Lorentz transformation $(a \neq 0)$. The intersection of $\{x \mid (x, x) = a^2\}$ and the light cone $C(x')$ consists only of the vertex x' of the cone. Therefore, the relations

$$(l^{-1} l_\nu x - x,\ l^{-1} l_\nu x - x) = 0$$

and $l^{-1} l_\nu x' \neq x'$ do not hold together. In other words, $(l_\nu x)$ converges to lx in M_E implies that $(l_\nu x)$ converges to lx in M_s. □

Let φ denote the surjective homomorphism of $SL(2, \mathbb{C})$ onto $L_+^\uparrow$ (as introduced in Sec. 1) and let B denote the image under φ of the set of upper triangular matrices of the form $\left[\begin{smallmatrix} \alpha & \beta \\ 0 & \alpha^{-1} \end{smallmatrix}\right]$ with $|\alpha| \neq 1$.

Lemma 2: Let x be an isotropic vector and let (l_ν) be a net of Lorentz transformations. Then $(l_\nu x)$ converges to lx in M_s if and only if

(i) $l_\nu x$ converges to lx in M_E,

(ii) no $\bar{l} \in L$ exists such that eventually $\bar{l}^{-1} l^{-1} l_\nu \bar{l} \in B$.

Proof: Similarly, as in the proof of Lemma 1, it suffices to consider only one isotropic vector. We choose $x' = (1, 1, 0, 0)$ and (compare Sec. 1) the relation

$$\widetilde{lx} = s \tilde{x} s^*,$$

written out and applied to our situation, becomes

$$\begin{pmatrix} y_0 + y_1 & y_2 - iy_3 \\ y_2 + iy_3 & y_0 - y_1 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \bar{\alpha} & \bar{\gamma} \\ \bar{\beta} & \bar{\delta} \end{pmatrix} = 2 \begin{pmatrix} |\alpha|^2 & \alpha\bar{\gamma} \\ \bar{\alpha}\gamma & |\gamma|^2 \end{pmatrix}.$$

Again, we have to exclude nets (l_ν) with the property that eventually $(l^{-1} l_\nu x - x,\ l^{-1} l_\nu x - x) = 0$ and $l^{-1} l_\nu x \neq x$. The intersection of $\{x \mid (x, x) = 0\}$ and the light cone $C(x')$ only consists of the line $\lambda(1, 1, 0, 0)$, $\lambda \in \mathbf{R}$, and therefore we must look for l_ν with $l^{-1} l_\nu x' = \lambda_\nu x'$ $(\lambda_\nu \neq 1)$.

Let

$$l_\nu = \varphi \begin{pmatrix} \alpha_\nu & \beta_\nu \\ \gamma_\nu & \delta_\nu \end{pmatrix};$$

then for such l_ν we have

$$|\alpha_\nu|^2 \neq 1, \quad \alpha_\nu \bar{\gamma}_\nu = 0, \quad \bar{\alpha}_\nu \gamma_\nu = 0, \quad |\bar{\gamma}_\nu|^2 = 0;$$

in other words,

$$|\alpha_\nu|^2 \neq 1 \quad \text{and} \quad \gamma_\nu = 0.$$

Consequently, the 2×2 matrices in question correspond with elements $b \in B$; i.e., $l^{-1} l_\nu = b$ or $l_\nu = lb$. The Lorentz transformations, leaving invariant the other one-dimensional isotropic subspaces, have the form $\bar{l} b \bar{l}^{-1}$, where $\bar{l}$ is a suitable Lorentz transformation. Summarizing, we have to exclude l_ν, eventually satisfying the relation $l^{-1} l_\nu = \bar{l} b \bar{l}^{-1}$ or $\bar{l}^{-1} l^{-1} l_\nu \bar{l} = b$. Now the proof is complete. □

Corollary: L_s induces the discrete topology on the subgroup Z and on its conjugates.

Proof: As is known the elements of $\varphi^{-1}(z)$ have the form $\left[\begin{smallmatrix} t & 0 \\ 0 & t^{-1} \end{smallmatrix}\right]$ with $t \in \mathbf{R}_+^\times$, being a subset of B. □

Notice that in the case of Z there are two isotropic eigenvectors, viz., $(1, \pm 1, 0, 0)$ but in the case of B there is only the isotropic eigenvector $(1, 1, 0, 0)$.

Let C be the image under φ of the matrices $\left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right]$ of $SL(2, \mathbb{C})$ with properties:

(i) $|\alpha|^2 - |\beta|^2 - |\gamma|^2 + |\delta|^2 = 2$,

(ii) $|\alpha|^2 - |\beta|^2 \neq 1$.

Lemma 3: Let x be a spacelike vector and let (l_ν) be a net of Lorentz transformations. Then $(l_\nu x)$ converges to lx in M_s, if and only if:

(i) $(l_\nu x)$ converges to lx in M_E,

(ii) no $\bar{l} \in L$ exists such that eventually $\bar{l}^{-1} l^{-1} l_\nu \bar{l} \in C$.

Proof: Again we only need one spacelike vector to start with and we choose $x' = (0, a, 0, 0)$, situated on the

hypersurface $(x,x)=-a^2$ $(a\neq 0)$. Similarly, as for Lemma 2, we find

$$\begin{pmatrix} y_0+y_1 & y_2-iy_3 \\ y_2+iy_3 & y_0-y_1 \end{pmatrix}=\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}\begin{pmatrix} \bar{\alpha} & \bar{\gamma} \\ \bar{\beta} & \bar{\delta} \end{pmatrix}$$
$$=\begin{pmatrix} |\alpha|^2-|\beta|^2 & \alpha\bar{\gamma}-\beta\bar{\delta} \\ \bar{\alpha}\gamma-\bar{\beta}\delta & |\gamma|^2-|\delta|^2 \end{pmatrix}.$$

Now the intersection of $\{x|(x,x)=-a^2\}$ and $C(x')$ is situated in the hyperplane $x_1=a$ and therefore we have to look for the elements of $L_+^\uparrow$, transforming $(0,a,0,0)$ into $(v,\ a,\ v\cos u,\ v\sin u)$, where $v\neq 0$.

This means that

$$\begin{pmatrix} v+a & ve^{-iu} \\ ve^{iu} & v-a \end{pmatrix}=a\begin{pmatrix} |\alpha|^2-|\beta|^2 & \alpha\bar{\gamma}-\beta\bar{\delta} \\ \bar{\alpha}\gamma-\bar{\beta}\delta & |\gamma|^2-|\delta|^2 \end{pmatrix}.$$

or

$$v+a=a(|\alpha|^2-|\beta|^2),$$
$$v-a=a(|\gamma|^2-|\delta|^2),$$
$$ve^{iu}=a(\bar{\alpha}\gamma-\bar{\beta}\delta),$$

and these relations are equivalent with the conditions:

(i) $|\alpha|^2-|\beta|^2-|\gamma|^2+|\delta|^2=2$,

(ii) $|\alpha|^2-|\beta|^2\neq 1$ $(v\neq 0)$,

(iii) $(|\alpha|^2-|\beta|^2-1)^2=|\bar{\alpha}\gamma-\bar{\beta}\delta|^2$;

but condition (iii) is superfluous for it is implied by (i) and $\alpha\delta-\beta\gamma=1$. Similarly, as for Lemma 2, it turns out that in this case we must exclude the nets (l_ν) with the property that eventually $\bar{l}^{-1}l^{-1}l_\nu\bar{l}=c$, with $c\in C$. □

Now we are able to state:

Theorem: Let (l_ν) be a net of Lorentz transformations. Then (l_ν) converges to l in L_s if and only if:

(i) (l_ν) converges to l in L_E,

(ii) no $\bar{l}\in L$ exists such that eventually $\bar{l}^{-1}l^{-1}l_\nu\bar{l}\in B\cup C$ (B and C as defined above).

Proof: The theorem follows immediately from the Lemmas 1, 2, and 3. □

Remarks:

1. The theorem gives us an intrinsic definition of the topology of L_s by means of convergence of nets.
2. The topology of M_s can be recovered from L_s.
3. The Lorentz transformations that we have excluded for convergence, are exactly those leaving not only (x,x) invariant but also the intersections of the hypersurfaces $(x,x)=p$ and the light cones in the points of contact. In Lemma 1 this intersection only consists of one point and in Lemma 2 we found the one-dimensional subvarieties.
4. Probably the condition $|\alpha|^2-|\beta|^2-|\gamma|^2+|\delta|^2=2$ has to do with the roots of the equation $s\tilde{x}s^*=\lambda\tilde{x}$.
5. The set $B\cup C$ has the property that $l\in B\cup C$ implies $l^{-1}\in B\cup C$ and therefore L_s is a T_1-group (see Ref. 19, p. 27). As is known (cf. Ref. 1) L_s has no proper invariant subgroups and hence L_s is connected (see Ref. 19, p. 28).
6. Probably L_s has any representations that are not representations of L_E; these representations might lead to new invariants of physics. I did not succeed in finding examples of these new representations until now.

[1] I.M. Gel'fand, R.A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, Oxford, 1963).
[2] H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1970).
[3] E.C. Zeeman. J. Math. Phys. **5**, 490 (1964).
[4] J.L. Alonso and F.Y. Yndurain, Commun. Math. Phys. 4, 349 (1967).
[5] G. Barucchi, Nuovo Cimento A **55**, 385 (1968).
[6] G. Barucchi and G. Teppati, Nuovo Cimento A **52**, 50 (1967).
[7] O.S. Rothaus, Proc. Am. Math. Soc. **17**, 1284 (1966).
[8] G. Teppati, Nuovo Cimento **54**, 800 (1968).
[9] P.G. Vroegindewey "Some algebraic and topological investigations on space-time," Thesis, Utrecht, 1973.
[10] E.C. Zeeman, Topology **6**, 161 (1967).
[11] S. Nanda, "Topologies for Minkowski Space," Ph.D. thesis, Queens University, Kingston, 1969.
[12] S. Nanda, J. Math. Phys. **12**, 394 (1971).
[13] S. Nanda, J. Math. Phys. **13**, 12 (1972).
[14] M. Cole, Int. J. Theor. Phys. 2, 1 (1969).
[15] F.A. Cel'nik, Sov. Math. Dokl. 9, 1151 (1968).
[16] E.H. Kronheimer and R. Penrose, Proc. Camb. Phil. Soc. 63, 481 (1967).
[17] S. Willard, *General Topology* (Addison-Wesley, Reading, Mass., 1970).
[18] T. Husain, *Introduction to Topological Groups* (Saunders, Philadelphia, 1966).
[19] I. Kaplansky, *An Introduction to Differential Algebra* (Hermann, Paris, 1957).

Physical applications of multiplicative stochastic processes. III. Nonequilibrium entropy

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 12 July 1974; revised manuscript received 7 August 1974)

It is argued that the expression $-K_B$ Trace $[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle]$, which appears in a stochastic treatment of the dynamics of the density matrix is indeed the nonequilibrium entropy. The reasoning involves consideration of the time evolution of the free energy for a relaxing magnetic moment in a fluctuating magnetic environment. It is shown that the H theorem, and the monotonic decrease of the free energy, as described by Pauli's master equation can be generalized to the full density matrix, at least for the case of magnetic relaxation, which requires the presence of off-diagonal density matrix elements.

INTRODUCTION

In this paper, we shall argue that the expression

$$-K_B \,\mathrm{Trace}[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle] \tag{1}$$

is indeed the entropy for nonequilibrium behavior in quantum mechanical systems. Previously, it has been shown that $H(t)$, which is defined by the above expression without the $-K_B$, is a monotonically decreasing function of time.[1] However, it was not argued that $H(t)$ leads directly to the entropy. Such a connection requires a consideration of the physical quantity heat and its relation with the above expression. In this paper, it will be shown, in the special case of magnetic relaxation, how experimental verification of such a connection could be achieved.

The contention that thermodynamics quantities can have meaning in nonequilibrium situations requires support. Here, we will explicitly write out expressions for internal energy, entropy, and free energy which are assumed to be valid for nonequilibrium situations. We shall proceed by establishing the close relationship between these expressions and expressions found in the earlier literature on this subject. It will be seen that such expressions can be reasonably interpreted as nonequilibrium thermodynamical quantities.

The mathematical context for these ideas is given by the stochastic Schrödinger equation[2]

$$i\frac{d}{dt}C_\alpha(t) = M_{\alpha\alpha'}C_{\alpha'}(t) + \tilde{M}_{\alpha\alpha'}(t)C_{\alpha'}(t) \tag{2}$$

in which $M_{\alpha\alpha'} = M^*_{\alpha'\alpha}$ and $\tilde{M}_{\alpha\alpha'}(t) = \tilde{M}^*_{\alpha'\alpha}(t)$ and $\tilde{M}_{\alpha\alpha'}(t)$ is a purely random Gaussian stochastic Hamiltonian.

The corresponding density matrix equation is

$$\frac{d}{dt}\rho_{\alpha\beta}(t) = -iL_{\alpha\beta\alpha'\beta'}\rho_{\alpha'\beta'}(t) - i\tilde{L}_{\alpha\beta\alpha'\beta'}(t)\rho_{\alpha'\beta'}(t), \tag{3}$$

in which $\rho_{\alpha\beta}(t) \equiv C^*_\alpha(t)C_\beta(t)$, $L_{\alpha\beta\alpha'\beta'} \equiv \delta_{\alpha\alpha'}M_{\beta\beta'} - \delta_{\beta\beta'}M^*_{\alpha\alpha'}$, and $\tilde{L}_{\alpha\beta\alpha'\beta'}(t) = \delta_{\alpha\alpha'}\tilde{M}_{\beta\beta'}(t) - \delta_{\beta\beta'}\tilde{M}^*_{\alpha\alpha'}(t)$. The stochastic average of (3) is

$$\frac{d}{dt}\langle\rho_{\alpha\beta}(t)\rangle = -iL_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle - R_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle, \tag{4}$$

in which $R_{\alpha\beta\alpha'\beta'}$ is given by

$$R_{\alpha\beta\alpha'\beta'} = \delta_{\alpha\alpha'}Q_{\beta\theta\theta\beta'} + \delta_{\beta\beta'}Q_{\theta\alpha\alpha'\theta} - 2Q_{\beta\beta'\alpha'\alpha}, \tag{5}$$

wherein $Q_{\alpha\beta\alpha'\beta'}$ is determined by the second moments of $\tilde{M}_{\alpha\alpha'}(t)$,

$$\langle\tilde{M}_{\alpha\beta}(t)\tilde{M}_{\alpha'\beta'}(s)\rangle = 2Q_{\alpha\beta\alpha'\beta'}\delta(t-s). \tag{6}$$

The $\langle\rho_{\alpha\beta}(t)\rangle$ in (4) has been used in

$$H(t) \equiv \mathrm{Trace}[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle] \tag{7}$$

to prove that

$$\frac{d}{dt}H(t) \leq 0, \tag{8}$$

which is an H theorem for quantum mechanical systems.[1] This H theorem differs from previously published H theorems for quantum mechanical systems which instead show that if, at $t = 0$, $\langle\rho_{\alpha\beta}(0)\rangle$ has no off-diagonal matrix elements, then $H(t) < H(0)$ for $t > 0$.[3] However, for $s > 0$ as well, it can not be decided whether or not $H(s) > H(t)$ when $t > s$. Therefore, the H theorem given by (8) is much stronger, and closely related in form to Boltzmann's original H theorem for classical gases.[4]

If we now shift attention to consideration of a subsystem in contact with a heat reservoir, we can get an equation for the time evolution of the subsystem alone, which is[5]

$$\frac{d}{dt}\langle\rho_{ij}(t)\rangle = -i(E_i - E_j)\langle\rho_{ij}(t)\rangle - T_{iji'j'}\langle\rho_{i'j'}(t)\rangle, \tag{9}$$

wherein $T_{iji'j'}$ satisfies a detailed balancing condition given by

$$T_{iji'j'} = T^*_{i'j'ij}\exp[-(1/2K_BT)(E_i + E_j - E_{i'} - E_{j'})]. \tag{10}$$

In (10), T is the temperature of the heat reservoir. The interaction between the sybsystem and the reservoir has been assumed to be an energy conserving, stochastic interaction which leaves the reservoir always in equilibrium while the subsystem relaxes. The $T_{iji'j'}$ comes from the average over the stochastic interaction followed by a trace over all reservoir states. Equation (9) has been used to show, for the case of magnetic relaxation by a spin $\frac{1}{2}$ magnetic moment in a fluctuating magnetic environment, how the Block equations are rigorously constructed.[6] In general, Eq. (9) with (10) leads to the canonical equilibrium density matrix asymptotically in time.

EARLIER THEORIES

The Pauli master equation

$$\frac{d}{dt}P_\alpha(t)=\sum_\beta[W_{\alpha\beta}P_\beta(t)-W_{\beta\alpha}P_\alpha(t)], \tag{11}$$

in which $W_{\alpha\beta}=W_{\beta\alpha}$ and $W_{\alpha\beta}>0$ for $\alpha\neq\beta$ has been used as a model for nonequilibrium thermodynamics.[7] The $P_\alpha(t)$'s correspond with the diagonal density matrix elements: $P_\alpha(t)=\langle\rho_{\alpha\alpha}(t)\rangle$. Pauli's equation does not include off-diagonal density matrix elements. Its original appeal stems from the fact that

$$H(t)\equiv\sum_\alpha P_\alpha(t)\ln P_\alpha(t) \tag{12}$$

is montonically decreasing, as is readily seen from (11) by

$$\begin{aligned}\frac{d}{dt}H(t)&=\sum_\alpha\frac{d}{dt}P_\alpha(t)\ln P_\alpha(t)\\&=\sum_\alpha\sum_\beta[W_{\alpha\beta}P_\beta(t)-W_{\beta\alpha}P_\alpha(t)]\ln P_\alpha(t)\\&=\sum_\alpha\sum_\beta W_{\alpha\beta}[P_\beta(t)-P_\alpha(t)]\ln P_\alpha(t)\\&=\tfrac{1}{2}\sum_\alpha\sum_\beta W_{\alpha\beta}[P_\beta(t)-P_\alpha(t)]\ln\frac{P_\alpha(t)}{P_\beta(t)}\\&\leq 0.\end{aligned} \tag{13}$$

To prove (13), we used $W_{\alpha\beta}=W_{\beta\alpha}$ and $\sum_\alpha P_\alpha(t)=1$ for all t. The identification of $-K_BH(t)$ in this case with the entropy is very natural because the microcanonical equilibrium distribution for the probabilities, P_α^{eq}, as determined by equilibrium statistical mechanics, is

$$P_\alpha^{eq}=1/N \quad \text{when } \alpha=1,2,\ldots,N. \tag{14}$$

Consequently,

$$-K_B\sum_\alpha P_\alpha^{eq}\ln P_\alpha^{eq}=K_B\sum_{\alpha=1}^{N}\frac{1}{N}\ln N \tag{15}$$

or

$$S-S_0=K_B\ln N. \tag{16}$$

This is the Boltzmann–Planck formula for the entropy, $S-S_0$, when there are N states, each equally probable. Thus, Pauli's formula for $-K_BH(t)$ surely gives the correct nonequilibrium expression for the entropy.

There is more appeal to Pauli's equation then this. If we again consider a subsystem interacting with a heat reservoir, Pauli's equation becomes

$$\frac{d}{dt}P_\alpha(t)=\sum_\beta[T_{\alpha\beta}P_\beta(t)-T_{\beta\alpha}P_\alpha(t)] \tag{17}$$

in which $T_{\alpha\beta}=T_{\beta\alpha}\exp[-(1/K_BT)(E_\alpha-E_\beta)]$ is the detailed balancing condition.[8] It is therefore natural to take for the free energy the expression

$$F(t)=\sum_\alpha P_\alpha(t)E_\alpha+TK_B\sum_\alpha P_\alpha(t)\ln P_\alpha(t) \tag{18}$$

since the first sum is surely the internal energy, even in nonequilibrium situations, and the second sum is the negative of T, the temperature, times the nonequilibrium entropy. Using (17) with the detailed balancing condition, we get

$$\begin{aligned}\frac{d}{dt}F(t)&=\sum_\alpha\frac{d}{dt}P_\alpha(t)E_\alpha+TK_B\sum_\alpha\frac{d}{dt}P_\alpha(t)\ln P_\alpha(t)\\&=\sum_\alpha\sum_\beta[T_{\alpha\beta}P_\beta(t)-T_{\beta\alpha}P_\alpha(t)]E_\alpha\\&\quad+TK_B\sum_\alpha\sum_\beta[T_{\alpha\beta}P_\beta(t)-T_{\beta\alpha}P_\alpha(t)]\ln P_\alpha(t)\\&=\tfrac{1}{2}\sum_\alpha\sum_\beta[T_{\alpha\beta}P_\beta(t)-T_{\beta\alpha}P_\alpha(t)](E_\alpha-E_\beta)\\&\quad+TK_B\tfrac{1}{2}\sum_\alpha\sum_\beta[T_{\alpha\beta}P_\beta(t)-T_{\beta\alpha}P_\alpha(t)]\ln\frac{P_\alpha(t)}{P_\beta(t)}\\&=\frac{TK_B}{2}\sum_\alpha\sum_\beta[T_{\alpha\beta}P_\beta(t)-T_{\beta\alpha}P_\alpha(t)]\ln\frac{T_{\beta\alpha}}{T_{\alpha\beta}}+\ln\frac{P_\alpha(t)}{P_\beta(t)}\\&=\frac{TK_B}{2}\sum_\alpha\sum_\beta[T_{\alpha\beta}P_\beta(t)-T_{\beta\alpha}P_\alpha(t)]\ln\frac{T_{\beta\alpha}P_\alpha(t)}{T_{\alpha\beta}P_\beta(t)}\\&\leq 0.\end{aligned} \tag{19}$$

Thus, the free energy monotonically decreases to equilibrium, as it should.

The identification of $-K_BH(t)$, in this case, with the entropy is again very natural because the canonical equilibrium distribution for the probabilities, P_α^{eq}, as determined by equilibrium statistical mechanics, is

$$P_\alpha^{eq}=(1/Q)\exp(-E_\alpha/K_BT) \tag{20}$$

where $Q=\sum_\alpha\exp(-E_\alpha/K_BT)$. Now, note that $\ln P_\alpha^{eq}=-\ln Q-(1/K_BT)E_\alpha$. Consequently,

$$\begin{aligned}-K_B\sum_\alpha P_\alpha^{eq}\ln P_\alpha^{eq}&=K_B\sum_\alpha P_\alpha^{eq}\ln Q+(1/T)\sum_\alpha P_\alpha^{eq}E_\alpha\\&=K_B\ln Q+(1/T)U,\end{aligned} \tag{21}$$

where U is the internal energy. In addition $K_B\ln Q$ is $-(1/T)F$, where F is the Helmholtz free energy. So we have

$$-K_B\sum_\alpha P_\alpha^{eq}\ln P_\alpha^{eq}=-(1/T)F+(1/T)U \tag{22}$$

or

$$S=-K_B\sum_\alpha P_\alpha^{eq}\ln P_\alpha^{eq},$$

where S is the entropy. Thus, Pauli's formula for $-K_BH(t)$ surely gives the correct nonequilibrium for the entropy.

The shortcoming of this approach is that restriction to diagonal density matrix elements only is physically unrealistic. In the case of magnetic relaxation for the spin $\frac{1}{2}$ magnetic moment, we need the full density matrix because the off-diagonal density matrix elements give rise to the transverse relaxation of the magnetization.[6] The diagonal density matrix elements only contribute to the longitudinal relaxation. In general, even though the equilibrium density matrix will be diagonal only, the nonequilibrium density matrix will possess off-diagonal density matrix elements. Equations (9) and (10) are the generalization of Pauli's equation (17).

FULL DENSITY MATRIX TREATMENT

We argue that the natural analog of Pauli's $H(t)$ is

$$\text{Trace}[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle], \tag{23}$$

wherein the logarithm of a Hamiltonian matrix is defined through its diagonal representation.[1] Equation (23) reduces to (22) when $\langle\rho(t)\rangle$ takes on its equilibrium value

$$\langle\rho_{ij}(t)\rangle \underset{t\to\infty}{\longrightarrow} (1/Q)\delta_{ij}\exp(-E_i/K_BT) \tag{24}$$

in the case of a subsystem in contact with a heat reservoir. It has already been proved that $H(t)$ as given by (23) for an isolated system is monotone.[1] We shall now show, for the case of magnetic relaxation, that

$$F(t) \equiv \sum_i \langle\rho_{ii}(t)\rangle E_i + TK_B\,\text{Trace}[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle] \tag{25}$$

is also monotone when $\langle\rho_{ij}(t)\rangle$ satisfies (9) with (10). This theorem appears to be difficult in general, so we shall give a proof for the magnetic relaxation case only. Equation (25) is clearly the natural generalization (18).

MAGNETIC RELAXATION AND MONOTONICITY

For a spin $\frac{1}{2}$ magnetic moment, i and j in $\langle\rho_{ij}(t)\rangle$ take on only two values. Therefore, $\langle\rho_{ij}(t)\rangle$ is a 2×2 Hermitian positive definite matrix. The values for $T_{iji'j'}$ can be worked out and are[6]

$$\begin{aligned}
&T_{ijij} = 2Q^{x,y} + 2Q^z, \quad i \neq j,\\
&T_{ijji} = 0, \quad i \neq j,\\
&T_{iiii} = 4Q^{x,y}[1 - (1/Q_R)\exp(-E_i/K_BT)],\\
&T_{iijj} = -4Q^{x,y}(1/Q_R)\exp(-E_i/K_BT), \quad i \neq j
\end{aligned} \tag{26}$$

where $Q_R = \exp(-E_1/K_BT) + \exp(-E_2/K_BT)$ and $Q^{x,y}$ and Q^z are the second moment correlation strengths for transverse and longitudinal magnetic fluctuations in the magnetic environment. To get the time derivative of $F(t)$ in (25), we proceed in the following way.

$\langle\rho(t)\rangle$ has two eigenvalues given by

$$\lambda_\pm = \tfrac{1}{2} \pm \tfrac{1}{2}\sqrt{1 - 4\det\langle\rho(t)\rangle}. \tag{27}$$

The positive definiteness of $\langle\rho(t)\rangle$ restricts the determinant so that $\frac{1}{4} > \det\langle\rho(t)\rangle$. Equivalently, $\lambda_\pm$ are both positive eigenvalues, and it is clear that $\lambda_+ + \lambda_- = 1$ as should be the case for the eigenvalues of a density matrix. The entropy term in (25) can be written as

$$TK_B\,\text{Trace}[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle] = TK_B(\lambda_+\ln\lambda_+ + \lambda_-\ln\lambda_-) \tag{28}$$

since the trace is invariant under a unitary similarity transformation. Moreover, $\lambda_+ + \lambda_- = 1$ for all t implies that $(d/dt)\lambda_+ = -(d/dt)\lambda_-$. Therefore,

$$\frac{d}{dt}F(t) = E_1\frac{d}{dt}\langle\rho_{11}(t)\rangle + E_2\frac{d}{dt}\langle\rho_{22}(t)\rangle + TK_B\left(\frac{d}{dt}\lambda_+\right)\ln\frac{\lambda_+}{\lambda_-}. \tag{29}$$

Similarly, $\langle\rho_{11}(t)\rangle + \langle\rho_{22}(t)\rangle = 1$ for all t so that $(d/dt)\langle\rho_{11}(t)\rangle = -(d/dt)\langle\rho_{22}(t)\rangle$. Therefore,

$$\frac{d}{dt}F(t) = (E_1 - E_2)\frac{d}{dt}\langle\rho_{11}(t)\rangle + K_BT\left(\frac{d}{dt}\lambda_+(t)\right)\ln\frac{\lambda_+(t)}{\lambda_-(t)}. \tag{30}$$

To get $(d/dt)\lambda_+(t)$, we find

$$\frac{d}{dt}\lambda_+(t) = \tfrac{1}{4}[1 - 4\det\langle\rho(t)\rangle]^{-1/2}\left[-4\frac{d}{dt}\det\langle\rho(t)\rangle\right]. \tag{31}$$

Using (9) and (26), one gets

$$\begin{aligned}
&\frac{d}{dt}\det\langle\rho(t)\rangle\\
&= \langle\dot\rho_{11}\rangle\langle\rho_{22}\rangle + \langle\rho_{11}\rangle\langle\dot\rho_{22}\rangle - \langle\dot\rho_{12}\rangle\langle\rho_{21}\rangle - \langle\rho_{12}\rangle\langle\dot\rho_{21}\rangle\\
&= -4Q^{x,y}(1/Q_R)\exp(-E_2/K_BT)\langle\rho_{11}\rangle\langle\rho_{22}\rangle\\
&\quad + 4Q^{x,y}(1/Q_R)\exp(-E_1/K_BT)\langle\rho_{22}\rangle^2\\
&\quad - 4Q^{x,y}(1/Q_R)\exp(-E_1/K_BT)\langle\rho_{11}\rangle\langle\rho_{22}\rangle\\
&\quad + 4Q^{x,y}(1/Q_R)\exp(-E_2/K_BT)\langle\rho_{11}\rangle^2\\
&\quad + (2Q^{x,y} + 2Q^z)(\langle\rho_{12}\rangle\langle\rho_{21}\rangle + \langle\rho_{12}\rangle\langle\rho_{21}\rangle)\\
&= 4(Q^{x,y} + Q^z)\langle\rho_{12}\rangle\langle\rho_{21}\rangle - 4Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle\\
&\quad + 4Q^{x,y}[(1/Q_R)\exp(-E_1/K_BT)\langle\rho_{22}\rangle^2\\
&\quad + (1/Q_R)\exp(-E_2/K_BT)\langle\rho_{11}\rangle^2]\\
&= 4(Q^{x,y} + Q^z)\langle\rho_{12}\rangle\langle\rho_{21}\rangle - 8Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle + 4Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle\\
&\quad + 4Q^{x,y}[(1/Q_R)\exp(-E_1/K_BT)\langle\rho_{22}\rangle^2\\
&\quad + (1/Q_R)\exp(-E_2/K_BT)\langle\rho_{11}\rangle^2]\\
&= 4(Q^{x,y} + Q^z)\langle\rho_{12}\rangle\langle\rho_{21}\rangle - 8Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle\\
&\quad + 4Q^{x,y}(1/Q_R)\exp(-E_1/K_BT)\langle\rho_{22}\rangle\\
&\quad + 4Q^{x,y}(1/Q_R)\exp(-E_2/K_BT)\langle\rho_{11}\rangle
\end{aligned} \tag{32}$$

since $(1/Q_R)\exp(-E_1/K_BT) + (1/Q_R)\exp(-E_2/K_BT) = 1$ and $\langle\rho_{11}\rangle + \langle\rho_{22}\rangle = 1$.

Again using (9) and (26), we get

$$\begin{aligned}
&(E_1 - E_2)\frac{d}{dt}\langle\rho_{11}(t)\rangle\\
&= (E_1 - E_2)4Q^{x,y}\frac{1}{Q_R}\left[\exp\left(-\frac{E_1}{K_BT}\right)\langle\rho_{22}\rangle - \exp\left(-\frac{E_2}{K_BT}\right)\langle\rho_{11}\rangle\right].
\end{aligned} \tag{33}$$

Therefore,

$$\begin{aligned}
\frac{d}{dt}F(t) &= (E_1 - E_2)4Q^{x,y}\frac{1}{Q_R}\left[\exp\left(-\frac{E_1}{K_BT}\right)\langle\rho_{22}\rangle - \exp\left(-\frac{E_2}{K_BT}\right)\langle\rho_{11}\rangle\right]\\
&\quad - K_BT\frac{\{4(Q^{x,y} + Q^z)\langle\rho_{12}\rangle\langle\rho_{21}\rangle - 8Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle + 4Q^{x,y}(1/Q_R)[\exp(-E_1/K_BT)\langle\rho_{22}\rangle + \exp(-E_2/K_BT)\langle\rho_{11}\rangle]\}}{\sqrt{1 - 4\det\langle\rho(t)\rangle}}\ln\frac{\lambda_+}{\lambda_-}.
\end{aligned} \tag{34}$$

To see monotonicity in (34) we must express various of the terms in different form, using $\beta = 1/K_BT$:

$$\frac{1}{Q_R}[\exp(-\beta E_2)\langle\rho_{11}\rangle - \exp(-\beta E_1)\langle\rho_{22}\rangle] = \frac{(1/Q_R)[\exp(-\beta E_2)\langle\rho_{11}\rangle - \exp(-\beta E_1)\langle\rho_{22}\rangle](\langle\rho_{11}\rangle - \langle\rho_{22}\rangle)}{(\langle\rho_{11}\rangle - \langle\rho_{22}\rangle)}$$

$$= \frac{(1/Q_R)[\exp(-\beta E_2)\langle\rho_{11}\rangle^2 + \exp(-\beta E_1)\langle\rho_{22}\rangle^2] - \langle\rho_{11}\rangle\langle\rho_{22}\rangle}{\langle\rho_{11}\rangle - \langle\rho_{22}\rangle}$$

$$= \frac{(1/Q_R)[\exp(-\beta E_2)(\langle\rho_{11}\rangle^2 + \langle\rho_{11}\rangle\langle\rho_{22}\rangle) + \exp(-\beta E_1)(\langle\rho_{22}\rangle^2 + \langle\rho_{22}\rangle\langle\rho_{11}\rangle)] - 2\langle\rho_{11}\rangle\langle\rho_{22}\rangle}{\langle\rho_{11}\rangle - \langle\rho_{22}\rangle}$$

$$= \frac{(1/Q_R)[\exp(-\beta E_2)\langle\rho_{11}\rangle + \exp(-\beta E_1)\langle\rho_{22}\rangle] - 2\langle\rho_{11}\rangle\langle\rho_{22}\rangle}{\langle\rho_{11}\rangle - \langle\rho_{22}\rangle}. \tag{35}$$

Therefore,

$$(E_1 - E_2)4Q^{x,y}(1/Q_R)[\exp(-\beta E_1)\langle\rho_{22}\rangle - \exp(-\beta E_2)\langle\rho_{11}\rangle]$$
$$= K_B T \frac{[4Q^{x,y}(1/Q_R)(\exp(-\beta E_2)\langle\rho_{11}\rangle + \exp(-\beta E_1)\langle\rho_{22}\rangle) - 8Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle]}{\langle\rho_{11}\rangle - \langle\rho_{22}\rangle} \ln\exp[-\beta(E_1 - E_2)]. \tag{36}$$

Consequently,

$$\frac{d}{dt}F(t) = K_B T\left(4Q^{x,y}\frac{1}{Q_R}[\exp(-\beta E_2)\langle\rho_{11}\rangle + \exp(-\beta E_1)\langle\rho_{22}\rangle) - 8Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle\right)\left(\frac{1}{\langle\rho_{11}\rangle - \langle\rho_{22}\rangle}\ln\left(\frac{\exp(-\beta E_1)}{\exp(-\beta E_2)}\right) + \frac{1}{\sqrt{1 - 4\det\langle\rho(t)\rangle}}\ln\frac{\lambda_-}{\lambda_+}\right)$$
$$+ K_B T 4(Q^{x,y} + Q^z)\langle\rho_{12}\rangle\langle\rho_{21}\rangle\frac{1}{\sqrt{1 - 4\det\langle\rho(t)\rangle}}\ln\frac{\lambda_-}{\lambda_+}. \tag{37}$$

From (27) it is seen that

$$\sqrt{1 - 4\det\langle\rho(t)\rangle} = \lambda_+ - \lambda_-. \tag{38}$$

The second summand in (37) is, therefore,

$$K_B T 4(Q^{x,y} + Q^z)\langle\rho_{12}\rangle\langle\rho_{21}\rangle[1/(\lambda_+ - \lambda_-)]\ln(\lambda_-/\lambda_+) \leq 0. \tag{39}$$

If the first summand in (37) is also nonpositive, then we will have monotonicity. To see that this is so, we first note that

$$\frac{1}{\lambda_+ - \lambda_-}\ln\frac{\lambda_-}{\lambda_+} \leq \frac{1}{\langle\rho_{11}\rangle - \langle\rho_{22}\rangle}\ln\frac{\langle\rho_{22}\rangle}{\langle\rho_{11}\rangle}. \tag{40}$$

The proof of (40) requires writing $\langle\rho_{11}\rangle$ and $\langle\rho_{22}\rangle$ as

$$\langle\rho_{11}\rangle = \tfrac{1}{2} + r \quad \text{and} \quad \langle\rho_{22}\rangle = \tfrac{1}{2} - r, \tag{41}$$

where r is positive. This means, of course, that $\langle\rho_{11}\rangle \geq \langle\rho_{22}\rangle$. This does not limit the generality of our proof because the situation $\langle\rho_{22}\rangle > \langle\rho_{11}\rangle$ can be treated in a similar manner, using on the right-hand side of (40) the expression $[1/(\langle\rho_{22}\rangle - \langle\rho_{11}\rangle)]\ln(\langle\rho_{11}\rangle/\langle\rho_{22}\rangle)$, which is in fact equal to the right-hand side of (40). Therefore, we proceed with (41) with $\frac{1}{2} \geq r \geq 0$. Now, from (27) we see that

$$\lambda_\pm = \tfrac{1}{2} \pm \tfrac{1}{2}\sqrt{1 - 4(\langle\rho_{11}\rangle\langle\rho_{22}\rangle - \langle\rho_{12}\rangle\langle\rho_{21}\rangle)}$$
$$= \tfrac{1}{2} \pm \tfrac{1}{2}[1 - 4(\tfrac{1}{4} - r^2 - \langle\rho_{12}\rangle\langle\rho_{21}\rangle)]^{1/2}$$
$$= \tfrac{1}{2} \pm r(1 + \langle\rho_{12}\rangle\langle\rho_{21}\rangle/r^2)^{1/2}. \tag{42}$$

Consequently,

$$\lambda_+ - \lambda_- = 2r(1 + \langle\rho_{12}\rangle\langle\rho_{21}\rangle/r^2)^{1/2} \geq 2r = \langle\rho_{11}\rangle - \langle\rho_{22}\rangle. \tag{43}$$

Equivalently,

$$1/(\lambda_+ - \lambda_-) \leq 1/(\langle\rho_{11}\rangle - \langle\rho_{22}\rangle) \tag{44}$$

Additionally,

$$\frac{\lambda_-}{\lambda_+} = \frac{\frac{1}{2} - r(1 + \langle\rho_{12}\rangle\langle\rho_{21}\rangle/r^2)^{1/2}}{\frac{1}{2} + r(1 + \langle\rho_{12}\rangle\langle\rho_{21}\rangle/r^2)^{1/2}} \leq \frac{\frac{1}{2} - r}{\frac{1}{2} + r} = \frac{\langle\rho_{22}\rangle}{\langle\rho_{11}\rangle} \tag{45}$$

since the numerators satisfy $\frac{1}{2} - r(1 + \langle\rho_{12}\rangle\langle\rho_{21}\rangle/r^2)^{1/2} \leq \frac{1}{2} - r$ while the denominators satisfy $\frac{1}{2} + r(1 + \langle\rho_{12}\rangle\langle\rho_{21}\rangle/r^2)^{1/2} \geq \frac{1}{2} + r$. Consequently, we get

$$\ln\frac{\lambda_-}{\lambda_+} \leq \ln\frac{\langle\rho_{22}\rangle}{\langle\rho_{11}\rangle} \tag{46}$$

since the logarithm is a monotone function. Together, (44) and (46) prove (40).

In (37) the first factor of the first summand can be written as

$$K_B T\{4Q^{x,y}(1/Q_R)[\exp(-\beta E_2)\langle\rho_{11}\rangle + \exp(-\beta E_1)\langle\rho_{22}\rangle] - 8Q^{x,y}\langle\rho_{11}\rangle\langle\rho_{22}\rangle\}$$
$$= K_B T 4Q^{x,y}(1/Q_R)[\exp(-\beta E_2)\langle\rho_{11}\rangle - \exp(-\beta E_1)\langle\rho_{22}\rangle](\langle\rho_{11}\rangle - \langle\rho_{22}\rangle). \tag{47}$$

as is seen from (35). Therefore, for the first summand in (37) we get

$$K_B T 4Q^{x,y}\frac{1}{Q_R}[\exp(-\beta E_2)\langle\rho_{11}\rangle - \exp(-\beta E_1)\langle\rho_{22}\rangle](\langle\rho_{11}\rangle\langle\rho_{22}\rangle)$$
$$\times\left[\frac{1}{\langle\rho_{11}\rangle - \langle\rho_{22}\rangle}\ln\left(\frac{\exp(-\beta E_1)}{\exp(-\beta E_2)}\right) + \frac{1}{\lambda_+ - \lambda_-}\ln\frac{\lambda_-}{\lambda_+}\right]$$
$$\leq K_B T 4Q^{x,y}\frac{1}{Q_R}[\exp(-\beta E_2)\langle\rho_{11}\rangle - \exp(-\beta E_1)\langle\rho_{22}\rangle]$$
$$\times\left[\ln\left(\frac{\exp(-\beta E_1)}{\exp(-\beta E_2)}\right) + \ln\frac{\langle\rho_{22}\rangle}{\langle\rho_{11}\rangle}\right]$$
$$= K_B T 4Q^{x,y}\frac{1}{Q_R}[\exp(-\beta E_2)\langle\rho_{11}\rangle - \exp(-\beta E_1)\langle\rho_{22}\rangle]$$
$$\times\left(\ln\left[\frac{\exp(-\beta E_1)\langle\rho_{22}\rangle}{\exp(-\beta E_2)\langle\rho_{11}\rangle}\right]\right) \leq 0. \tag{48}$$

The first inequality in (48) follows from (40) while the second follows from its form, as was the case in (39). Together, (39) and (48) prove

$$\frac{d}{dt}F(t) \leq 0. \tag{49}$$

With the full density matrix for this magnetic relaxation problem we have been able to parallel the Pauli master equation picture described by (19). One can then begin to believe that $F(t)$ as given by (25) is indeed the nonequilibrium free energy.

EXPERIMENTAL CONFIRMATION

The principal question is whether or not

$$-K_B \,\mathrm{Trace}[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle] \tag{50}$$

is indeed the nonequilibrium entropy. In (25) the term $\sum_i E_i\langle\rho_{ii}(t)\rangle$ is surely the nonequilibrium internal energy, so that only (50) remains to be considered. The advantage in using this picture of a subsystem in contact with a heat reservoir is that the temperature of the reservoir is held constant. Consequently, one can use

$$-TK_B \,\mathrm{Trace}[\langle\rho(t_2)\rangle \ln\langle\rho(t_2)\rangle] + TK_B \,\mathrm{Trace}[\langle\rho(t_1)\rangle \ln\langle\rho(t_1)\rangle] = \Delta Q \tag{51}$$

as the heat exchanged during the time interval from t_1 to t_2. Equation (51) parallels the quasistatic relationship $dQ = Tds$ of thermodynamics, and appears plausible here because the temperature is constant. Experimentally, one must attempt to measure the heat exchanged with the reservoir by the subsystem magnetic moment and see if the time course of heat exchange follows the time course of

$$-TK_B \,\mathrm{Trace}[\langle\rho(t)\rangle \ln\langle\rho(t)\rangle]$$

as determined by (9). Agreement would confirm all of the physical assignments for the mathematical expressions suggested in this paper.

Such an experimental test is perhaps quite difficult in the case of magnetic relaxation as described here. Two considerations must be made to properly test this theory. First of all, we have assumed that fluctuation correlations are very short lived compared with the relaxation times. This is evidenced in Eq. (6). For longer-lived correlations there are reasons for believing that the initial stages of relaxation do not show monotonic free energy changes until times of the order of the correlation time have elapsed, after which times the relaxation would become monotonic.[9] Second of all, we have treated an isolated magnetic moment in a stochastic magnetic environment. In reality this separation of subsystem and reservoir may not be so clean, or easily achieved experimentally. Instead, the magnetic moment may have nearby neighbors with which there is systematic interaction as well as having a stochastic reservoir interaction. These nearby neighbors can introduce some oscillatory behavior in addition to the relaxation and can renormalize transverse frequencies. A more sophisticated calculation along the lines given here for the case of neighboring magnetic moment interactions, and for correlation times of greater length for the reservoir magnetic fluctuations may be required before experimental confirmation is possible.

Within the context described by Eqs. (9) and (10), it is also desired to find general proof that the free energy as defined by (25) is monotonically decreasing, without specializing considerations to the magnetic relaxation case.

ACKNOWLEDGMENT

This work was supported by NSF Grant GP 39309. *This paper is dedicated to Mark Kac on the occasion of his 60th birthday.*

[1] R.F. Fox, J. Math. Phys. **13**, 1726 (1972); K_B is Boltzmann's constant, and $\langle \cdot\cdot\cdot \rangle$ denotes stochastic averaging.

[2] R.F. Fox, J. Math. Phys. **13**, 1196 (1972); $\hbar = 1$ is assumed throughout.

[3] R.C. Tolman, *The Principles of Statistical Mechanics* (Oxford U.P., London, 1968), Sec. 106.

[4] K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), Chap. 4.

[5] R.F. Fox, J. Math. Phys. **14**, 20 (1973).

[6] R.F. Fox, J. Math. Phys. **15**, 217 (1974).

[7] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965), Chap. 15, Sec. 1.

[8] See Ref. 7, Chap. 15, Sec. 2.

[9] If the correlation function is not a delta function, the monotonicity of $F(t)$ depends upon monotonicity of the correlation function. An exponentially decaying correlation function can still lead to a monotone $F(t)$, but a correlation function which is nonmonotonically decaying may not.

Solution of potential problems near the corner of a conductor

L. C. Davis and J. R. Reitz

Scientific Research Staff, Ford Motor Company, Dearborn, Michigan 48121
(Received 17 January 1975)

The Green's function for a space defined (in cylindrical coordinates) by the intersection of two half-planes S_1 ($\varphi=0$) and S_2 ($\varphi=\theta$ where $0 < \theta \leq 2\pi$) is found by a technique due to Sommerfeld. The Green's function (or its normal derivative) is required to vanish on the surface $S_1 + S_2$ as well as at infinity. When $\theta = m\pi/k$ where k and m are integers, the solution can be written in terms of the Green's function u_m for a Riemann space of m windings (in φ). For $m = 1$ and 2, u_m can be expressed in terms of elementary functions. For $m = 3$, we find u_m to be given in terms of complete elliptic integrals. Application to some simple electrostatic and magnetostatic problems is made, particularly for $\theta = 3\pi/2$.

1. INTRODUCTION

The use of images for solving boundary-value problems in electrostatics and magnetostatics is well known. It is not as well known, however, that Sommerfeld[1] extended the technique to a space consisting of several Riemann windings. As a consequence, he was able to find the potential of a point charge when the potential on a nearby semi-infinite sheet is required to vanish. To satisfy the boundary conditions, the images are placed in the second winding. Sommerfeld's procedure has also been applied to the problem of a charge near a conducting disk.[2-4] (For a more recent review, see Ref. 5.)

A method of finding the potential for the case of a charge near two conducting planes intersecting at an arbitrary angle θ (see Fig. 1) was also described by Sommerfeld.[1] In Sec. 2 we prove that his prescription is correct. Then in Sec. 3 we consider some examples, and specifically evaluate the case $\theta=3\pi/2$. In Sec. 4 we apply our results to some simple problems in electrostatics and magnetostatics. Our conclusions are stated briefly in Sec. 5.

The Sommerfeld technique is of interest now because of its use in certain magnetostatic problems associated with magnetic levitation of high-speed ground vehicles.[6,7]

2. SOMMERFELD'S METHOD OF CONSTRUCTING THE GREEN'S FUNCTION

Let us find the Green's function $G(\mathbf{x}, \mathbf{x}')$ for the problem shown in Fig. 1. We use cylindrical coordinates: for the field point $\mathbf{x}=(r, \varphi, z)$ and for the source point $\mathbf{x}'=(r', \varphi', z')$. The boundary $S=S_1+S_2$ is formed by the planes $\varphi=0$ and $\varphi=\theta$. The z axis is along the line of intersection of the planes S_1 and S_2. On S we require either $G(\mathbf{x}, \mathbf{x}')$ or $\partial G(\mathbf{x}, \mathbf{x}')/\partial n$ to vanish. Also $G(\mathbf{x}, \mathbf{x}')$ must vanish as $|\mathbf{x}| \to \infty$, $0<\varphi<\theta$. Hence we require the solution of

$$\nabla^2 G(\mathbf{x}, \mathbf{x}') = -4\pi\,\delta(\mathbf{x}-\mathbf{x}') \tag{1}$$

in the region V_0 defined by $0<\varphi<\theta$ for all r and z (subject to the boundary conditions).

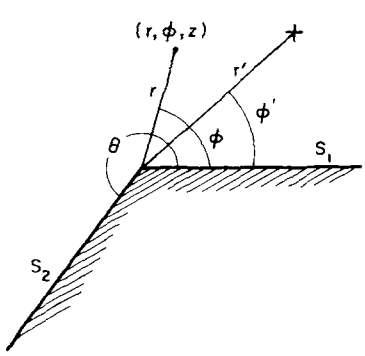


FIG. 1. Cylindrical coordinates of the field point (r, φ, z) and of the source (r', φ', z'). The z axis is perpendicular to the figure and is formed by the intersection of the planes S_1 and S_2.

The distance between the field and source points is

$$R = |\mathbf{x}-\mathbf{x}'| = [r^2 + r'^2 - 2rr'\cos(\varphi-\varphi') + (z-z')^2]^{1/2}. \tag{2}$$

The Green's function when there is no boundary S is simply $1/R$. It can be written as

$$\frac{1}{R} = \frac{1}{2\pi i}\oint_C \frac{f(\alpha)\,d\alpha}{R_\alpha}, \tag{3}$$

where R_α is found by replacing φ' by the complex variable α in (2). The function $f(\alpha)$ has a simple pole at $\alpha=\varphi'$ and C is a contour about φ' (see Fig. 2). Sommerfeld[1] showed that if we take

$$f(\alpha)=f_m(\alpha)=\frac{i}{m}\,\frac{\exp(i\alpha/m)}{\exp(i\alpha/m)-\exp(i\varphi'/m)}, \tag{4}$$

we can find the Green's function for a Riemann space of m windings in φ (m is an integer) as follows. The contour C can be deformed to C'. The points $\alpha=\varphi\pm i\alpha_1 + (\nu-1)2\pi$, $\nu=1, 2, \ldots, m$, are branch points of the integrand where α_1 is defined by (α_1 real and positive)

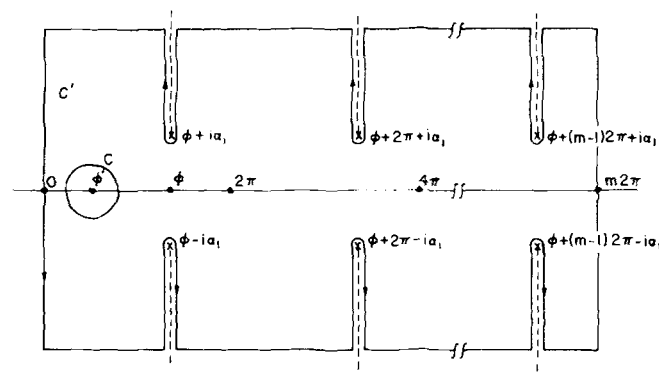


FIG. 2. Integration contours in the complex α plane. Branch cuts are denoted by dashed lines. The original contour C encloses the pole at φ'. The deformed contour C' is expanded around the branch cuts.

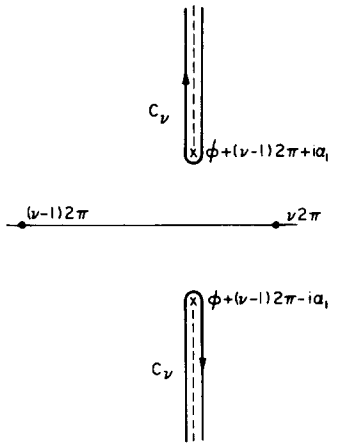


FIG. 3. Definition of the contour C_ν in the complex α plane. This contour is formed by going around the two branch cuts lying between $\mathrm{Re}\alpha=(\nu-1)2\pi$ and $\mathrm{Re}\alpha=2\nu\pi$.

$$\cos i\alpha_1=\cosh\alpha_1=[r^2+r'^2+(z-z')^2]/2rr'. \tag{5}$$

The contribution to (3) from the portions of the contour C' at $\mathrm{Re}\alpha=0$ and $\mathrm{Re}\alpha=m2\pi$ cancel due to the periodicity of the integrand. The horizontal portions give a vanishing contribution when taken at $\mathrm{Im}\alpha\to\pm\infty$. Therefore, we are left with the contributions from around the branch cuts, i.e.,

$$\frac{1}{R}=\sum_{\nu=1}^{m}\frac{1}{2\pi i}\int_{C_\nu}\frac{f_m(\alpha)\,d\alpha}{R_\alpha}\,. \tag{6}$$

The contour C_ν is around the two branch cuts defined by $\alpha=\varphi\pm i\alpha_1+(\nu-1)2\pi$ (see Fig. 3).

Let us define

$$u_m(r,\varphi,z;r',\varphi',z')=\frac{1}{2\pi i}\int_{C_1}\frac{f_m(\alpha)\,d\alpha}{R_\alpha} \tag{7a}$$

$$=\frac{1}{2m\pi(2rr')^{1/2}}\int_{C_1}\frac{d\alpha}{[\cosh\alpha_1-\cos(\varphi-\alpha)]^{1/2}}$$

$$\times\frac{1}{1-\exp[(i/m)(\varphi'-\alpha)]}\,. \tag{7b}$$

(7b) is obtained by using (5) to write R_α as

$$R_\alpha=(2rr')^{1/2}\,[\cosh\alpha_1-\cos(\varphi-\alpha)]^{1/2}. \tag{8}$$

By setting $\alpha=\alpha'+(\nu-1)2\pi$, we find

$$\frac{1}{2\pi i}\int_{C_\nu}\frac{f_m(\alpha)\,d\alpha}{R_\alpha}=\frac{1}{2m\pi(2rr')^{1/2}}$$

$$\times\int_{C_\nu}\frac{d\alpha}{[\cosh\alpha_1-\cos(\varphi-\alpha)]^{1/2}}$$

$$\times\frac{1}{1-\exp[(i/m)(\varphi'-\alpha)]}$$

$$=\frac{1}{2m\pi(2rr')^{1/2}}\int_{C_1}\frac{d\alpha'}{[\cosh\alpha_1-\cos(\varphi-\alpha')]^{1/2}}$$

$$\times\frac{1}{1-\exp\{(i/m)[\varphi'-(\nu-1)2\pi-\alpha']\}}$$

$$=u_m[r,\varphi,z;r',\varphi'-(\nu-1)2\pi,z'], \tag{9}$$

so that

$$1/R=\sum_{\nu=1}^{m}u_m[r,\varphi,z;r',\varphi'-(\nu-1)2\pi,z']. \tag{10}$$

It is straightforward to show that[1,3]

$$u_m(r,\varphi,z;r',\varphi',z')$$

$$=\frac{1}{m\pi(2rr')^{1/2}}\int_{\alpha_1}^{\infty}\frac{d\beta}{(\cosh\beta-\cosh\alpha_1)^{1/2}}$$

$$\times\frac{\sinh(\beta/m)}{\cosh(\beta/m)-\cos[(1/m)(\varphi'-\varphi)]}\,. \tag{11}$$

As noted above, u_m is the Green's function for a Riemann space of m windings since it

(i) satisfies Laplace's equation (by construction) except when $(r,\varphi,z)\to(r',\varphi',z')$ where it goes as $1/R$,[8]

(ii) vanishes at infinity,

(iii) is continuous, and

(iv) is periodic in φ and φ' with period $2m\pi$ [as can be seen from (11)].

An alternative derivation of u_m is given in the Appendix.

We define $u_\gamma(r,\varphi,z;r',\varphi',z')$ for arbitrary γ (not necessarily equal to an integer) by replacing m by γ in (11) [or, equivalently, in (7b)]:

$$u_\gamma=\frac{1}{\gamma\pi(2rr')^{1/2}}\int_{\alpha_1}^{\infty}\frac{d\beta}{(\cosh\beta-\cosh\alpha_1)^{1/2}}$$

$$\times\frac{\sinh(\beta/\gamma)}{\cosh(\beta/\gamma)-\cos[(1/\gamma)(\varphi'-\varphi)]}\,. \tag{11'}$$

Now for $\theta=\gamma\pi$ $(0<\gamma\leq 2)$, the solution of (1) in region V_0 which satisfies the boundary condition $G=0$ on S and vanishes as $|\mathbf{x}|\to\infty$ $(0<\varphi<\theta)$ is

$$G(\mathbf{x},\mathbf{x}')=u_\gamma(r,\varphi,z;r',\varphi',z')-u_\gamma(r,\varphi,z;r',-\varphi',z'). \tag{12}$$

The proof goes as follows. Clearly $u_\gamma(r,\varphi,z;r',\varphi',z')$ satisfies (i), (ii), and (iii) above. Likewise $u_\gamma(r,\varphi,z;r',-\varphi',z')$ must also satisfy (i), (ii), and (iii), the only difference being that the $1/R$ divergence is not encountered because φ and φ' are restricted to $0<\varphi,\varphi'<\theta$. Hence G as written in (12) is a solution to (1) which vanishes at infinity. We need only show that it vanishes on S.

The variables φ and φ' enter u_γ only in the denominator of the last term in (11'). Hence we must examine the quantity

$$1/\{\cosh(\beta/\gamma)-\cos[(1/\gamma)(\varphi'-\varphi)]\}$$
$$-1/\{\cosh(\beta/\gamma)-\cos[(1/\gamma)(\varphi'+\varphi)]\} \tag{13}$$

to determine the behavior of G on S_1 and S_2. On S_1, $\varphi=0$ so that (13) obviously vanishes on this surface. On

S_2, $\varphi = \theta = \gamma\pi$ so that $\cos[(1/\gamma)(\varphi' \mp \varphi)] = -\cos(\varphi'/\gamma)$. Therefore, (13) must also vanish on S_2, which means that $G = 0$ on S. This completes the proof since G satisfies (1) and the boundary conditions.

If we require the normal derivative $\partial G/\partial n$ to vanish on S instead of G itself, then we must change the sign of the second term in (12) to +. The proof proceeds in a manner similar to that already given and, consequently, is omitted. In summary then, the Green's function for the region V_0 of Fig. 1 is for $\theta = \gamma\pi$

$$G(\mathbf{x}, \mathbf{x}') = u_\gamma(r, \varphi, z; r', \varphi', z') \mp u_\gamma(r, \varphi, z; r', -\varphi', z'). \tag{14}$$

The (−) sign is for $G = 0$ on S and (+) is for $\partial G/\partial n = 0$ on S.

We can relate (14) to the Green's function for a Riemann space of m windings (u_m) when γ is the ratio of integers m/k. Consider

$$u_{m/k}(r, \varphi, z; r', \varphi', z') = \frac{k}{m\pi(2rr')^{1/2}} \int_{\alpha_1}^{\infty} \frac{d\beta}{(\cosh\beta - \cosh\alpha_1)^{1/2}} \times \frac{\sinh(k\beta/m)}{\cosh(k\beta/m) - \cos[(k/m)(\varphi' - \varphi)]}. \tag{15}$$

Let

$$S(x) = \frac{\sinh kx}{\cosh kx - \cos k\Delta}, \tag{16}$$

where $x = \beta/m$ and $\Delta = (\varphi' - \varphi)/m$. It is straightforward to show that the denominator of (16) is

$$\cosh kx - \cos k\Delta = 2^{k-1} \prod_{\nu=1}^{k} (\cosh x - \cos\Delta_\nu), \tag{17}$$

where $\Delta_\nu = \Delta + 2\pi(\nu - 1)/k$. Now

$$S(x) = \frac{1}{k} \frac{\partial}{\partial x} \ln(\cosh kx - \cos k\Delta). \tag{18}$$

Substituting (17) into (18) and differentiating with respect to x, we find

$$S(x) = \frac{\sinh x}{k} \sum_{\nu=1}^{k} (\cosh x - \cos\Delta_\nu)^{-1}. \tag{19}$$

Equation (19) enables us to write (15) as

$$u_{m/k}(r, \varphi, z; r', \varphi', z') = \frac{k}{m\pi(2rr')^{1/2}} \int_{\alpha_1}^{\infty} \frac{d\beta}{(\cosh\beta - \cosh\alpha_1)^{1/2}} \frac{\sinh(\beta/m)}{k} \times \sum_{\nu=1}^{k} (\cosh(\beta/m) - \cos\{(1/m)[\varphi' - \varphi + 2\pi m(\nu - 1)/k]\})^{-1}. \tag{20}$$

By definition (11), then

$$u_{m/k}(r, \varphi, z; r', \varphi', z') = \sum_{\nu=1}^{k} u_m[r, \varphi, z; r', \varphi' + 2\pi m(\nu - 1)/k, z']. \tag{21}$$

Hence, the Green's function for $\theta = m\pi/k$ is

$$G(\mathbf{x}, \mathbf{x}') = \sum_{\nu=1}^{k} \{u_m[r, \varphi, z; r', \varphi' + 2\pi m(\nu - 1)/k, z'] \mp u_m(r, \varphi, z; r', -\varphi' + 2\pi m\nu/k, z')\}. \tag{22}$$

The form of the second term follows from the fact that u_m is periodic in φ' with period $2\pi m$ so that $-\varphi'$ is equivalent to $2\pi m - \varphi'$.

A simple interpretation of (22) can be given. The Green's function for the region V_0 (defined by $0 < \varphi < \theta = m\pi/k$) is found by constructing a Riemann space of m windings. The real source is at φ' in V_0 (which is in the first winding) and the $(2k - 1)$ images are in the other windings and/or regions of the first winding not in V_0 (i.e., $\theta = m\pi/k < \varphi < 2\pi$). This is the result Sommerfeld described. Some examples are given in the next section.

3. EXAMPLES

In this section we consider some specific examples of spaces bounded by two intersecting, conducting planes, and we evaluate the appropriate Green's functions.

A. $\theta = \pi/n$ (n = integer)

In this case $m = 1$ and $k = n$. From (11),

$$u_1(r, \varphi, z; r', \varphi', z') = \frac{1}{\pi(2rr')^{1/2}} \int_{\alpha_1}^{\infty} \frac{d\beta}{(\cosh\beta - \cosh\alpha_1)^{1/2}} \times \frac{\sinh\beta}{\cosh\beta - \cos(\varphi' - \varphi)}. \tag{23}$$

Setting $\xi = \cosh\beta$, $\sigma = \cosh\alpha_1$, and $\tau = \cos(\varphi' - \varphi)$, we have

$$u_1(r, \varphi, z; r', \varphi', z') = \frac{1}{\pi(2rr')^{1/2}} \int_{\sigma}^{\infty} \frac{d\xi}{(\xi - \sigma)^{1/2}} \frac{1}{\xi - \tau} = \frac{1}{\pi(2rr')^{1/2}} \frac{1}{(\sigma - \tau)^{1/2}} = \frac{1}{R}. \tag{24}$$

Hence, we obtain the well-known result that

$$G(\mathbf{x}, \mathbf{x}') = \sum_{\nu=1}^{n} (1/R_\nu \mp 1/R'_\nu), \tag{25a}$$

where

$$R_\nu = \{r^2 + r'^2 - 2rr'\cos[\varphi - \varphi' - 2\pi(\nu - 1)/n] + (z - z')^2\}^{1/2} \tag{25b}$$

and R'_ν is obtained from R_ν by replacing φ' by $-\varphi'$. Here the Riemann space has only one winding (corresponding to real space).

B. $\theta = 2\pi$ (semi-infinite sheet)

In this case $m = 2$ and $k = 1$. It has been worked out in detail previously.[1,5] It was found that

$$u_2(r, \varphi, z; r', \varphi', z') = (2/\pi R)\tan^{-1}[(\sigma + \tau)/(\sigma - \tau)]^{1/2}, \tag{26}$$

where here we define $\sigma = \cosh(\alpha_1/2)$ and $\tau = \cos\frac{1}{2}(\varphi - \varphi')$. In (26) $\tan^{-1}$ is restricted to the range 0 to $\pi/2$.

The Green's function is

$$G(\mathbf{x},\mathbf{x}') = (2/\pi R)\tan^{-1}[(\sigma+\tau)/(\sigma-\tau)]^{1/2} \mp (2/\pi R')\tan^{-1}[(\sigma+\tau')/(\sigma-\tau')]^{1/2}, \quad (27)$$

where R' and τ' are found from R and τ by replacing φ' with $-\varphi'$.

C. $\theta = 3\pi/2$ (edge of a thick, conducting plate)

In this case $m=3$ and $k=2$. From (11) we have

$$u_3(r,\varphi,z;r',\varphi',z') = \frac{1}{3\pi(2rr')^{1/2}}\int_{\alpha_1}^{\infty}\frac{d\beta}{(\cosh\beta-\cosh\alpha_1)^{1/2}} \times \frac{\sinh(\beta/3)}{\cosh(\beta/3)-\cos\frac{1}{3}(\varphi'-\varphi)}. \quad (28)$$

We now make the substitutions

$$\xi = \cosh(\beta/3), \quad (29a)$$

$$\sigma = \cosh(\alpha_1/3), \quad (29b)$$

$$\tau = \cos\tfrac{1}{3}(\varphi'-\varphi), \quad (29c)$$

so that

$$u_3(r,\varphi,z;r',\varphi',z') = \frac{1}{\pi(2rr')^{1/2}}\int_{\sigma}^{\infty}\frac{d\xi}{[4(\xi^3-\sigma^3)-3(\xi-\sigma)]^{1/2}}\frac{1}{(\xi-\tau)}. \quad (30)$$

Following Byrd and Friedman,[9] we make a change of variables:

$$\xi = \sigma + A(1+x)/(1-x), \quad -1<x<1, \quad (31a)$$

where

$$A = (3\sigma^2 - 3/4)^{1/2}. \quad (31b)$$

Equation (30) becomes

$$u_3(r,\varphi,z;r',\varphi',z') = B/[2\pi(2rr'A)^{1/2}(A+\sigma-\tau)], \quad (32a)$$

where

$$B = \int_{-1}^{1} dx\left(\frac{1-x}{1+x}\right)^{1/2}\frac{1}{1+\eta x}\frac{1}{[1-k^2(1-x^2)]^{1/2}}, \quad (32b)$$

$$\eta = (A-\sigma+\tau)/(A+\sigma-\tau), \quad (32c)$$

and

$$k^2 = \tfrac{1}{2}(1-3\sigma/2A). \quad (32d)$$

Equation (32b) can be rewritten as

$$B = [2(\eta+1)/\eta)] \times \int_0^1 dx/\{(1-x^2)^{1/2}[1-k^2(1-x^2)]^{1/2}(1-\eta^2x^2)\} - (2/\eta)\int_0^1 dx/\{(1-x^2)^{1/2}[1-k^2(1-x^2)]^{1/2}\}. \quad (33)$$

Letting $x=\cos\Theta$ gives the standard forms for the complete elliptic integrals[10]:

$$B = (2/\eta)(1-\eta)^{-1}\Pi(n,k) - (2/\eta)K(k), \quad (34a)$$

where

$$n = \eta^2/(1-\eta^2). \quad (34b)$$

Substituting (34a) into (32a), we obtain

$$u_3(r,\varphi,z;r',\varphi',z') = [\pi(2rr'A)^{1/2}(A-\sigma+\tau)]^{-1}\,[(1-\eta)^{-1}\Pi(n,k)-K(k)] \quad (35)$$

The Green's function for $\theta = 3\pi/2$ is therefore

$$G(\mathbf{x},\mathbf{x}') = [\pi(2rr'A)^{1/2}]^{-1}\sum_{\nu=1}^{2}\left(\frac{(1-\eta_\nu)^{-1}\Pi(n_\nu,k)-K(k)}{A-\sigma+\tau_\nu} \mp \frac{(1-\eta_\nu')^{-1}\Pi(n_\nu',k)-K(k)}{A-\sigma+\tau_\nu'}\right), \quad (36a)$$

where

$$n_\nu = \eta_\nu^2/(1-\eta_\nu^2), \quad (36b)$$

$$\eta_\nu = (A-\sigma+\tau_\nu)/(A+\sigma-\tau_\nu), \quad (36c)$$

and

$$\tau_\nu = \cos\tfrac{1}{3}[\varphi'+3\pi(\nu-1)-\varphi]. \quad (36d)$$

The quantities n_ν', η_ν', and τ_ν' are found by replacing φ' by $-\varphi'$ in n_ν, η_ν, and τ_ν, respectively. As before, A is defined by (31b), σ by (29b), and k by (32d). Equation (5) defines α_1.

4. APPLICATION TO ELECTROSTATICS AND MAGNETOSTATICS

A. Electrostatic surface charge and potential mapping

The electrostatic potential V of a point charge q at (r',φ',z') near a grounded conductor of the shape shown in Fig. 1 is, according to (14) (in *mks* units),

$$V(r,\varphi,z) = (q/4\pi\epsilon_0)[u_\gamma(r,\varphi,z;r',\varphi',z') - u_\gamma(r,\varphi,z;r',-\varphi',z')] \quad (37)$$

for $\theta=\gamma\pi$. The surface charge density σ_s on the conductor is

$$\sigma_s = -\epsilon_0\frac{1}{r}\frac{\partial V}{\partial\varphi}, \quad \varphi=0 \quad \text{(surface } S_1\text{)}, \quad (38a)$$

$$= +\epsilon_0\frac{1}{r}\frac{\partial V}{\partial\varphi}, \quad \varphi=\gamma\pi \quad \text{(surface } S_2\text{)}. \quad (38b)$$

Let us consider the behavior of σ_s near the corner (i.e., as $r\to 0$, $\varphi=0$ or $\gamma\pi$). Sommerfeld[1] has shown that, for small r, u_γ can be approximated by

$$u_\gamma(r,\varphi,z;r',\varphi',z') \approx \frac{1}{\gamma R_0}\left[1+C_\gamma\cos\frac{1}{\gamma}(\varphi-\varphi')\left(\frac{rr'}{R_0^2}\right)^{1/\gamma}\right], \quad (39a)$$

where

$$R_0 = [r'^2+(z-z')^2]^{1/2}. \quad (39b)$$

We have evaluated C_γ and find it to be

$$C_\gamma = \frac{4}{\pi}\int_0^\infty\frac{dt}{(t^2+1)^{\gamma'}}, \quad \gamma' = (\gamma+1)/\gamma. \quad (40)$$

C_γ depends only on γ. Equations (39) and (40) are obtained by noting that as $r\to 0$

$$\cosh\alpha_1 \approx \exp(\alpha_1)/2 \approx [r'^2+(z-z')^2]/2rr' \to \infty, \quad (41a)$$

$$(\cosh\beta-\cosh\alpha_1)^{1/2} \approx (e^\beta - e^{\alpha_1})^{1/2}/\sqrt{2}, \quad (41b)$$

and

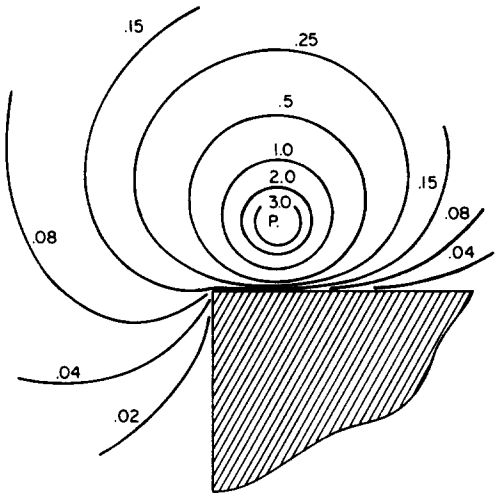


FIG. 4. Traces of equipotential surfaces surrounding a point charge q located near the edge of a thick conducting plate ($\theta = 3\pi/2$). The point charge is located at point P ($r' = 1$, $\varphi' = \pi/4$).

$$\sinh(\beta/\gamma)/\{\cosh(\beta/\gamma) - \cos[(1/\gamma)(\varphi - \varphi')]\}$$
$$\approx 1 + 2\exp(-\beta/\gamma)\cos[(1/\gamma)(\varphi - \varphi')]. \tag{41c}$$

From (37)—(41) we find (for either $\varphi = 0$ or $\varphi = \gamma\pi$)

$$\sigma_s \approx \frac{-qC_\gamma \sin(\varphi'/\gamma)(r')^{1/\gamma}}{2\gamma^2\pi r^{1-1/\gamma}R_0^{1+2/\gamma}} \qquad (r \to 0). \tag{42}$$

For $\theta < \pi$ ($\gamma < 1$), σ_s vanishes as $r^{1/\gamma - 1}$, whereas, for $\theta > \pi$ ($\gamma > 1$), σ_s diverges as $1/r^{1-1/\gamma}$. For $\theta = \pi$ ($\gamma = 1$), σ_s reduces to

$$\sigma_s = -qr' \sin\varphi'/2\pi R_0^3 \quad (r = 0), \tag{43}$$

a result easily obtained by simple image techniques.

Figure 4 shows a set of equipotential surfaces surrounding a point charge q located near the edge of a thick conducting plate ($\theta = 3\pi/2$). The point charge is located at $r' = 1$, $\varphi' = \pi/4$. The equipotentials were evaluated numerically from Eq. (36a). The complete elliptic integral $K(k)$ was calculated from the Gauss hypergeometric series[11] in k^2; $\Pi(n, k)$ was obtained by numerical integration of its defining formula.

To find the force on q, we must first subtract the $1/R$ potential due to q from the total potential. Hence we are interested in

$$V' = V - q/4\pi\epsilon_0 R. \tag{44}$$

We now restrict γ to the ratio of integers m/k. The subtraction of $1/R$ is done most conveniently by using (10). Hence, $4\pi\epsilon_0 V'/q$ may be expressed as the difference between Eqs. (22) and (10).

From $m = 1$, the required force can be obtained easily and need not be written down here. For $m = 2$, $k = 1$, the forces have been obtained previously[5] and will not be discussed here either. Let us consider then $m = 3$, $k = 2$ ($\theta = 3\pi/2$). First we find

$$f_r(r', \varphi, \varphi_1) = -\frac{\partial u_3(r, \varphi, z; r', \varphi_1, z')}{\partial r} \tag{45a}$$

and

$$f_\varphi(r', \varphi, \varphi_1) = -\frac{1}{r}\frac{\partial u_3(r, \varphi, z; r', \varphi_1, z')}{\partial \varphi} \tag{45b}$$

for $r = r'$, $z = z'$, φ and φ_1 arbitrary. Later we set $\varphi = \varphi'$ and $\varphi_1 = \varphi' + 2\pi m(\nu - 1)/k$, etc. Making use of the well-known expressions for $\partial K/\partial k$, $\partial \Pi/\partial k$ and $\partial \Pi/\partial n$,[9] we find after a somewhat tedious but straightforward calculation

$$f_r(r', \varphi, \varphi_1) = \frac{1}{4r'^2}\,\frac{1}{\sqrt{1-\tau}\,(\sqrt{1-\tau} + \sqrt{3/2}\,)}, \tag{46a}$$

and

$$f_\varphi(r', \varphi, \varphi_1) = -\frac{\sin\frac{1}{3}(\varphi - \varphi_1)}{12\sqrt{3}\,r'^2}\left(\frac{2\sqrt{1-\tau} + \sqrt{3/2}}{(1-\tau)^{3/2}(\sqrt{1-\tau} + \sqrt{3/2}\,)^2}\right), \tag{46b}$$

where

$$\tau = \cos\tfrac{1}{3}(\varphi - \varphi_1). \tag{47}$$

The force on q is $\mathbf{F} = -q\nabla V'$ and is given by

$$F_j(r', \varphi') = (q^2/4\pi\epsilon_0)[f_j(r', \varphi', \varphi' + 3\pi) - f_j(r', \varphi', 3\pi - \varphi')$$
$$- f_j(r', \varphi', 6\pi - \varphi') - f_j(r', \varphi', \varphi' + 2\pi)$$
$$- f_j(r', \varphi', \varphi' + 4\pi)], \qquad j = r, \varphi. \tag{48}$$

Clearly there is no z component to the force. The periodicity of u_3 has been used in (48) to put all angles φ_1 into the region $0 < \varphi_1 < 6\pi$. From (46) and (48) we find that (dropping primes)

$$F_r = \frac{q^2}{16\pi\epsilon_0 r^2}\left(\frac{4}{3} - \sqrt{3}\right.$$
$$- \frac{1}{2\cos(\varphi/3)[\cos(\varphi/3) + \sqrt{3/2}\,]}$$
$$\left. - \frac{1}{2\sin(\varphi/3)[\sin(\varphi/3) + \sqrt{3/2}]}\right) \tag{49a}$$

and

$$F_\varphi = \frac{q^2\sin(2\varphi/3)}{96\sqrt{3}\,\pi\epsilon_0 r^2}\left(\frac{\cos(\varphi/3) + \sqrt{3}/4}{[\cos(\varphi/3)]^3[\cos(\varphi/3) + \sqrt{3/2}]^2}\right.$$
$$\left. - \frac{\sin(\varphi/3) + \sqrt{3}/4}{[\sin(\varphi/3)]^3[\sin(\varphi/3) + \sqrt{3/2}]^2}\right). \tag{49b}$$

B. Magnetostatic current circulation function and surface currents

The scalar potential Ω of a monopole q at (r', φ', z') near a perfectly diamagnetic conductor of the shape shown in Fig. 1 is, according to (14) (in mks units),

$$\Omega(r, \varphi, z) = (\mu_0 q/4\pi)[u_\gamma(r, \varphi, z; r', \varphi', z')$$
$$+ u_\gamma(r, \varphi, z; r', -\varphi', z')] \tag{50}$$

for $\theta = \gamma\pi$. (The boundary condition is $\partial\Omega/\partial n = 0$ on S.) It is convenient to define a current circulation function (or stream function)

$$\Phi(r, z) = \mu_0^{-1}\Omega(r, 0, z) \quad \text{on } S_1 \tag{51a}$$

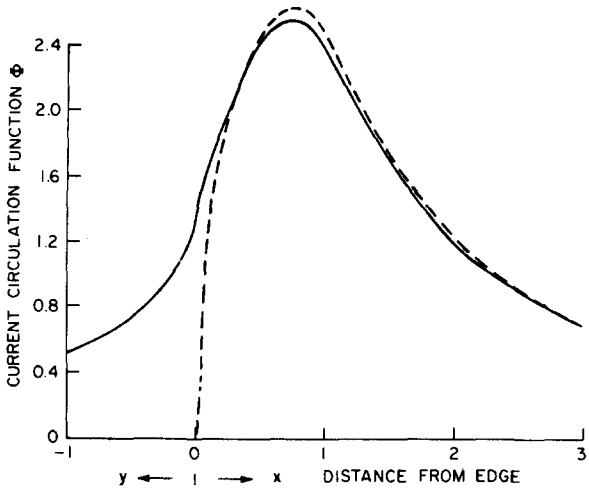


FIG. 5. The solid curve depicts the current circulation function for the top and side surfaces of a thick conducting plate ($\theta = 3\pi/2$), under the influence of a point pole at $r' = 1$, $\varphi' = \pi/4$. The top surface of the plate ($\varphi = 0$) is denoted by x, whereas the vertical surface ($\varphi = 3\pi/2$) is denoted by y. The dashed curve depicts the corresponding current circulation function for the semi-infinite sheet ($\theta = 2\pi$).

$$= \mu_0^{-1}\Omega(r, \gamma\pi, z) \text{ on } S_2. \tag{51b}$$

From Ampere's circuital law and the vanishing of B $= -\nabla\Omega$ inside the diamagnetic conductor, we find that the surface currents are given by

$$i_z = \partial\Phi/\partial r, \quad i_r = -\partial\Phi/\partial z \quad \text{on } S_1 \tag{52a}$$

and

$$i_z = -\partial\Phi/\partial r, \quad i_r = \partial\Phi/\partial z \quad \text{on } S_2. \tag{52b}$$

By reasoning similar to that in Sec. 4A, we find that, at $r = 0$, Φ is continuous and

$$\Phi(0, z) = q/2\pi\gamma R_0. \tag{53}$$

The current flowing on S_1 away from the corner is [from (52) and (53)]

$$i_r = q(z - z')/2\pi\gamma R_0 \quad (r = 0). \tag{54}$$

The current flowing on S_2 into the corner can be shown to be given by (54) also. Therefore, current flows around the corner and is conserved. The current flowing along the edge is found to be [from (39), (40), (50)—(52)]

$$i_z = \frac{q\, C_\gamma \cos(\varphi'/\gamma)(r')^{1/\gamma}}{2\gamma^2\pi R_0^{1+2/\gamma} r^{1-1/\gamma}} \qquad (r \to 0) \tag{55}$$

for both S_1 and S_2. From (55) we see that i_z vanishes as $r^{1/\gamma - 1}$ for $\theta < \pi$ ($\gamma < 1$), diverges as $1/r^{1-1/\gamma}$ for $\theta > \pi$ ($\gamma > 1$), and reduces to

$$i_z = qr' \cos\varphi'/2\pi R_0^3 \quad (r = 0) \tag{56}$$

for $\theta = \pi$ ($\gamma = 1$). Equation (56) also follows from simple image theory.

When $\theta = 2\pi$ ($\gamma = 2$), it is more convenient to define the current circulation function as

$$\Phi = \mu_0^{-1}[\Omega(r, \varphi = 0, z) - \Omega(r, \varphi = 2\pi, z)]. \tag{57}$$

In this definition the surface currents $i_z = \partial\Phi/\partial r$ and $i_r = -\partial\Phi/\partial z$ correspond to the sum of the currents associated with S_1 and S_2. This is useful since the two current sheets coincide.

Figure 5 shows the current circulation function Φ for the top and side surfaces of the thick conducting plate ($\theta = 3\pi/2$) along with the corresponding current circulation function for the semi-infinite sheet ($\theta = 2\pi$). The point pole is located at $r' = 1$, $\varphi' = \pi/4$. i_z is infinite at the edge in both cases, but for the thick plate i_z is proportional to $r^{-1/3}$ at the edge whereas for the semi-infinite sheet i_z is proportional to $r^{-1/2}$.

The force on the monopole q is $\mathbf{F} = -q\nabla\Omega'$, where $\Omega' = \Omega - \mu_0 q/4\pi R$. The calculation of $\mathbf{F}$ is the same as for the electrostatic case (Sec. 4A) except for the sign of certain terms. The result for the case $\theta = 3\pi/2$ ($\gamma = 3/2$) is (dropping primes)

$$F_r = \frac{\mu_0 q^2}{16\pi r^2}\left(\frac{4}{3} - \sqrt{3} + \frac{1}{2\cos(\varphi/3)[\cos(\varphi/3) + \sqrt{3}/2]}\right.$$

$$\left. + \frac{1}{2\sin(\varphi/3)[\sin(\varphi/3) + \sqrt{3}/2]}\right) \tag{58a}$$

and

$$F_\varphi = -\frac{\mu_0 q^2 \sin(2\varphi/3)}{96\sqrt{3}\,\pi r^2}\left(\frac{\cos(\varphi/3) + \sqrt{3}/4}{[\cos(\varphi/3)]^3[\cos(\varphi/3) + \sqrt{3}/2]^2}\right.$$

$$\left. - \frac{\sin(\varphi/3) + \sqrt{3}/4}{[\sin(\varphi/3)]^3[\sin(\varphi/3) + \sqrt{3}/2]^2}\right). \tag{58b}$$

5. CONCLUSIONS

We have derived the Green's function for the problem shown in Fig. 1, i.e., for the region $0 < \varphi < \theta$ and for all r and z. When $\theta = m\pi/k$ (m and k integers), the result can be expressed in terms of the Green's function u_m for a Riemann space of m windings (in φ) as first suggested by Sommerfeld.[1] For $m = 1$ and $m = 2$, u_m can be written in terms of elementary functions. We have evaluated u_3 here and find it involves complete elliptic integrals.

Our results have been applied to some simple electrostatics and magnetostatics problems. In particular we have investigated the behavior of the surface charge density and the surface current density near the corner formed by two intersecting, conducting planes. The forces on a point charge and on a monopole have also been calculated.

One unsolved problem of interest for applications is that of the strip conductor. Although Sommerfeld stated

that a transformation to bipolar coordinates would give the solution, we found that it gives the Green's function for a disk instead.[5]

APPENDIX

An alternative derivation of $u_m(r, \varphi, z; r', \varphi', z')$ can be found by defining $\delta(\varphi - \varphi')$ [in (A1) below] appropriately for a Riemann space of m windings. Let us solve

$$\nabla^2 u = -(4\pi/r)\,\delta(r - r')\,\delta(z - z')\,\delta(\varphi - \varphi') \tag{A1}$$

subject to the boundary condition that u vanishes as r or $z \to \infty$. The standard solution proceeds by putting[12]

$$\delta(z - z') = \pi^{-1} \int_0^\infty dk\, \cos k(z - z') \tag{A2}$$

and

$$\delta(\varphi - \varphi') = \pi^{-1} \sum_{n=0}^{\infty} \epsilon_n \cos n(\varphi - \varphi'), \tag{A3}$$

where

$$\epsilon_n = \tfrac{1}{2}, \quad n = 0,$$
$$= 1, \quad n \geq 1. \tag{A4}$$

The result is[12]

$$u_1 = 1/R$$
$$= (4/\pi) \sum_{n=0}^{\infty} \epsilon_n \cos n(\varphi - \varphi')$$
$$\times \int_0^\infty dk \cos k(z - z') I_n(kr_<) K_n(kr_>), \tag{A5}$$

where $I_n(x)$ and $K_n(x)$ are modified Bessel functions of order n. The symbol $r_>$ ($r_<$) denotes the greater (lesser) of r and r'.

If, instead of (A3), we let

$$\delta(\varphi - \varphi') = (m\pi)^{-1} \sum_{n=0}^{\infty} \epsilon_n \cos[(n/m)(\varphi - \varphi')] \tag{A6}$$

in (A1), we will obtain u_m (m = integer). Definition (A6) is appropriate to a space of m windings since $\delta(\varphi - \varphi')$ has period $2m\pi$. Let

$$u = (m\pi^2)^{-1} \sum_{n=0}^{\infty} \epsilon_n \cos \frac{n}{m}(\varphi - \varphi')$$
$$\times \int_0^\infty dk \cos k(z - z')\, v_n(k, r, r'). \tag{A7}$$

In cylindrical coordinates

$$\nabla^2 u = \frac{1}{r} \frac{\partial}{\partial r}\left(r \frac{\partial u}{\partial r}\right) + \frac{1}{r^2} \frac{\partial^2 u}{\partial \varphi^2} + \frac{\partial^2 u}{\partial z^2}, \tag{A8}$$

so that

$$\frac{1}{r} \frac{\partial}{\partial r}\left(r \frac{\partial v_n}{\partial r}\right) - \left(\frac{n}{m}\right)^2 \frac{v_n}{r^2} - k^2 v_n = -\frac{4\pi}{r}\,\delta(r - r'). \tag{A9}$$

For $r < r'$, $v_n = A I_{n/m}(kr)$, and, for $r > r'$, $v_n = B K_{n/m}(kr)$. By integrating (A9) from $r' - 0$ to $r' + 0$ we obtain

$$\left.\frac{\partial v_n}{\partial r}\right|_{r'-0}^{r'+0} = -\frac{4\pi}{r'}. \tag{A10}$$

Also v_n is continuous at $r = r'$. By making use of the value of the Wronskian, $W[I_\nu(x), K_\nu(x)] = -1/x$, we find that $A = 4\pi K_{n/m}(kr')$ and $B = 4\pi I_{n/m}(kr')$. Hence,

$$u = \frac{4}{m\pi} \sum_{n=0}^{\infty} \epsilon_n \cos \frac{n}{m}(\varphi - \varphi')$$
$$\times \int_0^\infty dk \cos k(z - z') I_{n/m}(kr_<) K_{n/m}(kr_>). \tag{A11}$$

Clearly u is a solution of Laplace's equation except when $\mathbf{x} \to \mathbf{x}'$ where it goes as $1/R$, satisfies the boundary conditions, and is periodic in φ with period $2m\pi$. Since Sommerfeld[1] showed that u_m is unique, u must equal u_m.

We can demonstrate the equivalence directly. It is known that[13]

$$\int_0^\infty dk \cos k(z - z') I_{n/m}(kr_<) K_{n/m}(kr_>)$$
$$= \frac{1}{2(rr')^{1/2}}\, Q_{n/m-1/2}(\cosh\alpha), \tag{A12a}$$

$$\cosh\alpha = [r^2 + r'^2 + (z - z')^2]/2rr', \tag{A12b}$$

where $Q_\nu(x)$ is a Legendre function of the second kind. An integral representation of $Q_\nu(x)$ is ($\alpha > 0$, $\mathrm{Re}\,\nu > -1$)[14]

$$Q_\nu(\cosh\alpha) = \frac{1}{\sqrt{2}} \int_\alpha^\infty \frac{\exp[-(\nu + 1/2)t]\,dt}{(\cosh t - \cosh\alpha)^{1/2}}. \tag{A13}$$

Substituting into (A11), we find

$$u = \frac{2}{m\pi\sqrt{rr'}} \sum_{n=0}^{\infty} \epsilon_n\, Q_{n/m-1/2}(\cosh\alpha) \cos \frac{n}{m}(\varphi - \varphi')$$
$$= \frac{\sqrt{2}}{m\pi\sqrt{rr'}} \sum_{n=0}^{\infty} \epsilon_n \cos \frac{n}{m}(\varphi - \varphi') \int_\alpha^\infty \frac{dt \exp(-nt/m)}{(\cosh t - \cosh\alpha)^{1/2}}$$
$$= \frac{\sqrt{2}}{m\pi\sqrt{rr'}} \int_\alpha^\infty \frac{dt}{(\cosh t - \cosh\alpha)^{1/2}}$$
$$\times \sum_{n=0}^{\infty} \epsilon_n \exp(-nt/m) \cos \frac{n}{m}(\varphi - \varphi'),$$
$$= \frac{1}{m\pi\sqrt{rr'}} \int_\alpha^\infty \frac{dt}{(\cosh t - \cosh\alpha)^{1/2}}$$
$$\times \frac{\sinh(t/m)}{\cosh(t/m) - \cos[(\varphi - \varphi')/m]}. \tag{A14}$$

Equation (A14) agrees with our previous definition of u_m [Eq. (11)].

[1]A. Sommerfeld, Proc. London Math. Soc. 28, 395 (1897); 30, 161 (1899).
[2]E. W. Hobson, Cambridge Phil. Trans. 18, 277 (1900).
[3]S. F. Neustadter, Univ. Calif. Publ. Math. N.S. 1, 397 (1951).
[4]H. Bateman, *Partial Differential Equations of Mathematical Physics* (Dover, New York, 1944), Chap. X.
[5]L. C. Davis and J. R. Reitz, Am. J. Phys. 39, 1255 (1971).
[6]L. C. Davis and J. R. Reitz, J. Appl. Phys. 42, 4119 (1971).
[7]R. H. Borcherts and L. C. Davis, J. Appl. Phys. 43, 2418 (1972).
[8]This is, of course, equivalent to being a solution of Eq. (1). In addition, although u_m is continuous at the branch line $r=0$, it is not a solution of Laplace's equation there.
[9]P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, New York, 1971), 2nd ed.
[10]For the complete elliptic integral of the third kind we use the common definition

$$\Pi(n,k)=\int_0^{\pi/2}\frac{d\Theta}{(1+n\sin^2\Theta)(1-k^2\sin^2\Theta)^{1/2}}$$

whereas in Ref. 9 n is replaced by $-\alpha^2$. The complete elliptic integral of the first kind is standard, however:

$$K(k)=\int_0^{\pi/2}\frac{d\Theta}{(1-k^2\sin^2\Theta)^{1/2}}$$

[11]A. Erdelyi, *Higher Transcendental Functions*, Bateman Manuscript Project (McGraw-Hill, New York, 1953), Vol. II, p. 318.
[12]See, for example, J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962), Chap. 3.
[13]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. prepared by Yu. V. Geronimus and M. Yu. Tseytlin (Academic, New York, 1965), p. 732.
[14]See Ref. 13, p. 1002.

Cross sections in quantum mechanics

Enrique Campesino-Romeo* and John R. Taylor

Department of Physics and Astrophysics, University of Colorado, Boulder, Colorado 80302
(Received 12 December 1974)

The definition of scattering cross sections requires an averaging over wavepackets with random impact parameters ρ; this leads to an integral of the scattering probability over all ρ in a plane perpendicular to the incident beam. We show that, for scattering off a potential which is $O(1/r^{\beta})$ as $r\to\infty$, the scattering probability is $O(1/\rho^{2\beta-4})$ as $\rho\to\infty$. Thus for any $\beta > 3$, the integral over impact parameters is well-defined and convergent.

1. INTRODUCTION

The quantum theory of scattering has developed considerably in the sixteen years since the beautiful papers of Jauch.[1,2] In particular, great progress has been made towards proving that, in single-channel scattering off "reasonable" potentials, the ideas proposed by Jauch are correct—that the asymptotic condition is satisfied, and that the S operator is unitary. Most of the results of this extensive work are summarized in the book of Simon.[3]

However, one part of scattering theory—the definition of the observable cross section—has received comparatively little attention. A realistic definition of the cross section must be given, and this definition must be shown to lead to the well-known result

$$\frac{d\sigma}{d\Omega} = |f(\mathbf{p} \leftarrow \mathbf{p}_0)|^2.$$

(Here, and throughout this paper, we consider just single-channel scattering and follow the notation of Ref. 4.) Newton and Shtokhamer[5] have emphasized that there are two essential steps in defining and calculating the cross section. First, the measured cross section is related to the probability that a scattered particle be observed in a given cone in *position space*; however, the quantity that is theoretically accessible is the probability for a scattered particle to be found with its *momentum* in the corresponding cone in *momentum space*. Dollard[6] has shown that, as one would certainly expect, these two probabilities are equal. Thus one can take as the starting point of the theoretical discussion the momentum-space probability

$$w(C \leftarrow \phi) = \int_C d^3p\, |(S\phi)(\mathbf{p})|^2. \tag{1.1}$$

Here $w(C \leftarrow \phi)$ is the probability that the incident packet ϕ be observed to scatter into the cone C with apex at the origin; and $S\phi$ is the out state corresponding to the in state ϕ.

The second main idea is that in a real experiment one uses many different wavepackets which are randomly distributed over a wide beam.[7] Thus we must use a succession of packets ϕ_ρ, where each ϕ_ρ is a packet obtained from some definite ϕ by a rigid displacement $\boldsymbol{\rho}$; that is,

$$\phi_\rho(\mathbf{p}) = \exp(-i\boldsymbol{\rho}\cdot\mathbf{p})\phi(\mathbf{p})$$

where the vector $\boldsymbol{\rho}$, which can be called the impact parameter, takes on values in the plane perpendicular to the incident mean momentum $\mathbf{p}_0$. If the $\boldsymbol{\rho}$ are distributed uniformly with density $n_{\rm inc}$ over the cross section of the incident beam, then the total number of scatterings[8] into the cone C is

$$N_{\rm SC}(C) = n_{\rm inc}\int_{\rho<R} d^2\rho\, w(C \leftarrow \phi_\rho) \tag{1.2}$$

where R is the radius of the incident beam (which we take to have circular cross section). The importance of this integration over impact parameters is clear in Jauch's original paper[1]; it was emphasized in a subsequent paper by Wichmann[9]; and has since been described in at least two texts (Ref. 4 Chap. 3, and Ref. 10, Chap. 5).

We now come to the step that is the subject of this paper. One certainly expects the scattering probability to be very small for impact parameters ρ that are large compared to the interaction radius. Since the beam size R is certainly large compared to the interaction radius, this would mean that one can replace the integral (1.2) with $\rho < R$ by an integral over *all* ρ:

$$\int_{\rho<R} d^2\rho\, w(C \leftarrow \phi_\rho) \approx \int_{\infty} d^2\rho\, w(C \leftarrow \phi_\rho). \tag{1.3}$$

With this replacement one can quickly derive the result (see Ref. 4, pp. 49–51)

$$N_{\rm SC}(C) = n_{\rm inc}\sigma(C)$$

where $\sigma(C)$ is the cross section for scattering into the cone C:

$$\sigma(C) = \int_C d\Omega_p\, |f(\mathbf{p} \leftarrow \mathbf{p}_0)|^2.$$

By choosing C to be a small cone of solid angle $d\Omega$ we then obtain the differential cross section $d\sigma/d\Omega = |f|^2$.

That the probability $w(C \leftarrow \phi_\rho)$ goes rapidly to zero for large impact parameters ρ, and hence that one can make the replacement (1.3), is certainly a very natural assumption. However, we are unaware of any published proof of the result, and the result is obviously very important. If the integral over ρ with $\rho < R$ does not converge as $R\to\infty$, we would have the absurd situation where the scattering cross section depends on the size R of the beam, however large we make R.

In this paper we consider the scattering of a particle by a fixed potential and prove that, for suitable potentials, the probability $w(C \leftarrow \phi_\rho)$ *does* go to zero as $\rho\to\infty$, and that it does so sufficiently fast to justify the convergence of its integral over all ρ. We shall state our pre-

cise assumptions and results in the next section. However, the essential points can be briefly summarized below.

We shall show that if $V(\mathbf{r})$ falls off like $1/r^\beta$ at large r,

$$|V(\mathbf{r})| < K/r^\beta \quad (r \text{ sufficiently large}),$$

then the outgoing wave $(S\phi_\rho)(\mathbf{p})$ falls off at least as fast as $1/\rho^{\beta-2}$ for large impact parameters; that is (for suitsuitable ϕ),

$$|(S\phi_\rho)(\mathbf{p})| < K'/\rho^{\beta-2} \quad (\rho \text{ sufficiently large}). \tag{1.4}$$

This result shows, just as one would expect, that the rate of decrease of $(S\phi_\rho)(\mathbf{p})$ with impact parameter ρ depends on how fast $V(\mathbf{r})$ falls off with r.

Since the scattering probability is given by (1.1) as

$$w(C \leftarrow \phi_\rho) = \int_C d^3p\, |(S\phi_\rho)(\mathbf{p})|^2, \tag{1.5}$$

our result implies that

$$w(C \leftarrow \phi_\rho) < K''/\rho^{2\beta-4}.$$

In particular, provided $\beta > 3$ (that is, V falls off like $1/r^{3+\epsilon}$ for some $\epsilon > 0$), $w(C \leftarrow \phi_\rho)$ falls off like $1/\rho^{2+\delta}(\delta > 0)$, and the integral $\int d^2\rho\, w(C \leftarrow \phi_\rho)$ is convergent as required.

Naturally, we shall need certain additional (and quite mild) assumptions on the smoothness of $V(\mathbf{r})$ and on the incident wavefunction $\phi(\mathbf{p})$. These details will be discussed in Sec. 2.

2. ASSUMPTIONS, DEFINITIONS, AND RESULTS

The potential

Concerning the potential $V(\mathbf{r})$ we make the following assumptions:

(i) For some $\beta > 3$ and all $r >$ some R_0

$$|V(\mathbf{r})| < K/r^\beta; \tag{2.1}$$

(ii) $V \in L^2(R^3)$;

(iii) except at a finite number of point singularities, $V(\mathbf{r})$ is locally Hölder continuous.

As mentioned above, assumption (i) is essential for our proof. Assumptions (ii) and (iii) are more technical and we use them simply to guarantee the validity of some standard results from time-independent scattering theory. Specifically, Ikebe has shown that conditions (i)–(iii) imply the standard relations (2.2) and (2.4) below and the boundedness of the stationary wavefunctions $\psi^-_\rho(\mathbf{r})$. [See Ref. 11, Eqs. (2.2), (1.1), and (1.2).] These results can be proved under a variety of different assumptions, some weaker than ours in some respects, as discussed by Simon.[3]

The incident wave packet

We consider an incident wavepacket given by a momentum-space wavefunction $\phi(\mathbf{p})$ which is infinitely differentiable and of compact support,

$$\phi \in C_0^\infty(R^3).$$

Further, if $\mathbf{p}_0$ denotes the incident mean momentum (i.e., $\mathbf{p}_0 = \langle \mathbf{p} \rangle_\phi \neq 0$), then we require that the support of ϕ lie in the half-space $\mathbf{p} \cdot \mathbf{p}_0 > 0$. This means simply that all components of ϕ are moving forwards.

We define α to be the angular half-width of the incident packet; more precisely,

$$\alpha = \sup\{\text{angle between } \mathbf{p} \text{ and } \mathbf{p}_0 : \mathbf{p} \in \text{supp}\phi\}.$$

(See Fig. 1.) Clearly $0 < \alpha < \pi/2$, and in practice α is of the order of a degree or less. We define the *forward cone* C_ϕ

$$C_\phi = \{\mathbf{p}: (\text{angle between } \mathbf{p} \text{ and } \mathbf{p}_0) \leq \alpha\};$$

that is, C_ϕ is the smallest circular cone with axis $\mathbf{p}_0$ that contains suppϕ. (See Fig. 1.) Finally, we define $p_{\max}$ and $p_{\min}$ to be the largest and smallest values of $|\mathbf{p}|$ in suppϕ.

The cone of observation

We discuss the probability that the scattered particle be observed in a circular cone C with apex at the origin 0. In order to define a cross section we must require that C not overlap the forward cone C_ϕ: that is,

$$C \cap C_\phi = \{0\}.$$

This requirement makes precise the well-known restriction that one cannot measure (directly) a forward elastic cross section.

The argument and results

The outgoing packet corresponding to the incident packet $\phi(\mathbf{p})$ is

$$(S\phi)(\mathbf{p}) = \phi(\mathbf{p}) + \frac{ip}{2\pi} \int d\Omega_u f(\mathbf{p} \leftarrow p\mathbf{u})\phi(p\mathbf{u}) \tag{2.2}$$

where $\mathbf{u}$ is a unit vector. (See, for example, Ref. 4, p.49.) Since we require $(S\phi)(\mathbf{p})$ for $\mathbf{p}$ inside the cone of observation C, it follows that $\phi(\mathbf{p})$ is zero. Thus in our case we have (replacing ϕ by the displaced ϕ_ρ)

$$(S\phi_\rho)(\mathbf{p}) = \frac{ip}{2\pi} \int d\Omega_u f(\mathbf{p} \leftarrow p\mathbf{u}) \exp(-ip\rho \cdot p\mathbf{u})\phi(p\mathbf{u}) \tag{2.3}$$

with $\mathbf{p}$ in the cone C. The amplitude $f(\mathbf{p} \leftarrow \mathbf{p}')$ is given by the well-known expression (see, for example, Ref. 4, p.169)

$$f(\mathbf{p} \leftarrow \mathbf{p}') = -(2\pi)^2 m \langle \mathbf{p} - |V|\mathbf{p}' \rangle. \tag{2.4}$$

Thus we can write[12]

$$\frac{ip}{2\pi} f(\mathbf{p} \leftarrow p\mathbf{u}) = -(2\pi)^{-1/2} mp \int d^3r\, \psi^-_{\mathbf{p}}(\mathbf{r})^* V(\mathbf{r}) \exp(ip\mathbf{u} \cdot \mathbf{r})$$
$$= \int d^3r\, U_{\mathbf{p}}(\mathbf{r}) \exp(ip\mathbf{u} \cdot \mathbf{r}) \tag{2.5}$$

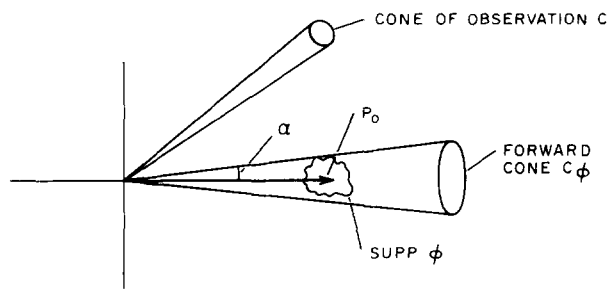


FIG. 1. Incident packet ϕ, the forward cone C_ϕ, and the cone of observation C.

where we have defined

$$U_p(\mathbf{r}) = -(2\pi)^{-1/2} mp\psi_p^-(\mathbf{r})^* V(\mathbf{r}). \tag{2.6}$$

The important point about the function $U_p(\mathbf{r})$ is that its rate of decrease as $r \to \infty$ is the same as that of the potential. This is because the stationary scattering wave $\psi_p^-(\mathbf{r})$ is bounded (and also continuous) for all $\mathbf{r}$ and all $\mathbf{p}$ in any compact domain not containing 0. [See Ikebe,[11] Eq. (1.2).] Since we are concerned only with $\mathbf{p}$ that lie in the cone C and satisfy $p_{\min} \le p \le p_{\max}$, it follows that

$$|U_p(\mathbf{r})| \le K_1 |V(\mathbf{r})| \tag{2.7}$$

for some constant K_1, for all $\mathbf{r}$, and all $\mathbf{p}$ of interest.

Substitution of (2.5) into (2.3) and a change of order of integration (which is certainly justified) yield

$$\begin{aligned}(S\phi_p)(\mathbf{p}) &= \int d^3r\, U_p(\mathbf{r}) \int d\Omega_u \exp[ip\mathbf{u}\cdot(\mathbf{r}-\rho)]\phi(p\mathbf{u}) \\ &= \int d^3r\, U_p(\mathbf{r}) g_p(\rho - \mathbf{r})\end{aligned} \tag{2.8}$$

where

$$g_p(\xi) = \int d\Omega_u \exp(-ip\mathbf{u}\cdot\xi)\phi(p\mathbf{u}). \tag{2.9}$$

The result (2.8) expresses the outgoing wavefunction $S\phi_p$ as a convolution of $U_p(\mathbf{r})$ and $g_p(\mathbf{r})$. Therefore, the behavior of $S\phi_p$ for large ρ is determined by the behavior of these two functions. We have already seen that $U_p(\mathbf{r})$ has essentially the same bound as $V(\mathbf{r})$—see (2.7). Thus it remains to examine the function $g_p(\xi)$ given by (2.9). This function is nearly (but not exactly) the Fourier transform of the incident wavefunction $\phi(\mathbf{p})$. If it *were* the Fourier transform, then $g_p(\xi)$ would fall off faster than any inverse power of ξ as $\xi \to \infty$. Unfortunately, the integral in (2.9) is over angles only (not over all $\mathbf{p}$) and the situation is more complicated. We shall prove in the next section that

$$|g_p(\xi)| < K_2/\xi \quad (\text{all } \xi) \tag{2.10}$$

where K_2 is some constant independent of p and ξ; and also that, provided ξ is outside both the forward and backward cones $\pm C_\phi$,

$$|g_p(\xi)| < K_3(n)/\xi^n \quad (\text{any } n,\ \xi \notin \pm C_\phi) \tag{2.11}$$

where $K_3(n)$ is a constant depending on the integer n but not p or ξ.

Returning to the convolution (2.8) for the outgoing wavefunction, we find that

$$|(S\phi_p)(\mathbf{p})| \le \int d^3r |U_p(\mathbf{r})|\,|g_p(\rho - \mathbf{r})|. \tag{2.12}$$

We shall show in detail in Sec. 3 that if this integral is split into two parts, $r \lesssim \rho\cos\alpha$, then in the first region we can apply the bound (2.11) while in the second we can use (2.10). Using the bound $|U_p(\mathbf{r})| < K_4/r^\beta$ for large r, we obtain

$$|(S\phi_p)(\mathbf{p})| \le K_5(n)/\rho^n + K_6/\rho^{\beta-2} \tag{2.13}$$

for ρ sufficiently large, and for all $\mathbf{p}$ of interest (i.e., all $\mathbf{p}$ in the cone C with $p_{\min} \le p \le p_{\max}$). Since n is arbitrary, (2.13) implies that

$$|(S\phi_p)(\mathbf{p})| \le K_7/\rho^{\beta-2},$$

as promised in the introduction.

The scattering probability $w(C \leftarrow \phi_p)$ is given by Eq. (1.5) and can now be bounded as follows:

$$\begin{aligned}w(C \leftarrow \phi_p) &= \int_C d^3p\, |(S\phi_p)(\mathbf{p})|^2 \\ &\le \frac{K_7^2}{\rho^{2(\beta-2)}} \int_C d\Omega_p \int_{p_{\min}}^{p_{\max}} p^2\, dp \\ &= K_8/\rho^{2(\beta-2)}.\end{aligned}$$

Thus, as anticipated, we see that $w(C \leftarrow \phi_p)$ falls off at a rate determined by the decrease of the potential $(1/r^\beta)$. In particular, since $\beta > 3$, $w(C \leftarrow \phi_p)$ falls off like some power better than $1/\rho^2$ and the integral of $w(C \leftarrow \phi_p)$ over all impact parameters is convergent.

3. DETAILED PROOFS

Bound on $g_p(\xi)$

We have to justify the bounds (2.10) and (2.11) on the function

$$g_p(\xi) = \int d\Omega_u \exp(-ip\mathbf{u}\cdot\xi)\phi(p\mathbf{u}). \tag{3.1}$$

If we choose spherical coordinates with polar axis along ξ and label the direction of $\mathbf{u}$ by (θ, φ), we can rewrite (3.1) as

$$g_p(\xi) = \int_{-1}^{1} dz \exp(-ip\xi z)\Phi(p,z) \tag{3.2}$$

where $z = \cos\theta$ and

$$\Phi(p,z) = \int_0^{2\pi} d\varphi\, \phi(p\mathbf{u}). \tag{3.3}$$

Now it can be shown that, if $\phi \in C_0^\infty(R^3)$, then $\Phi(p,z)$ is infinitely differentiable with respect to z on the closed interval $[-1,1]$ and that its nth derivative satisfies

$$\left|\frac{\partial^n \Phi}{\partial z^n}(p,z)\right| \le K_9(n)$$

where $K_9(n)$ is a constant depending on n but independent of p, z, and ξ.

If we integrate (3.2) by parts, then we obtain

$$g_p(\xi) = \frac{i}{p\xi}\left([\exp(-ip\xi z)\Phi(p,z)]_{-1}^{1} - \int_{-1}^{1} dz \exp(-ip\xi z)\frac{\partial\Phi}{\partial z}(p,z)\right). \tag{3.4}$$

Thus, since $p \ge p_{\min}$,

$$|g_p(\xi)| \le K_{10}/\xi \tag{3.5}$$

which is the bound (2.10).

If the polar axis—that is, the direction of ξ—does not lie in the forward or backward cones $\pm C_\phi$, then the endpoint term in (3.4) is zero. In this case we can integrate by parts as often as we please and obtain

$$|g_p(\xi)| \le K_{11}(n)/\xi^n \quad (n = 1, 2, \cdots), \tag{3.6}$$

which is (2.11).

Bound on $(S\phi_p)(\mathbf{p})$

The scattered wave is given by (2.8) as

$$(S\phi_p)(\mathbf{p}) = \int d^3r\, U_p(\mathbf{r}) g_p(\rho - \mathbf{r}).$$

We now split this integral into two parts, $r \lesseqgtr \rho\cos\alpha$, where α is the half-angle of the forward cone C_ϕ. (See Fig. 1.) If $r < \rho\cos\alpha$, then it is easily seen that $\mathbf{r}-\rho$ does not lie in either $\pm C_\phi$. Thus for this part of the integral we can use the bound (3.6). For the other part we use (3.5) and obtain

$$|(S\phi_\rho)(\mathbf{p})| \leq K_{11}(n)\int_{r<\rho\cos\alpha} d^3r \frac{|U_\rho(\mathbf{r})|}{|\mathbf{r}-\rho|^n} + K_{10}\int_{r>\rho\cos\alpha} d^3r \frac{|U_\rho(\mathbf{r})|}{|\mathbf{r}-\rho|}.$$

In the first term we can make the replacement $|\mathbf{r}-\boldsymbol{\rho}| > \rho(1-\cos\alpha)$, while in the second we can use the bound $|U_\rho(\mathbf{r})| \leq K_4/r^\beta$ (for ρ sufficiently large). Thus

$$|(S\phi_\rho)(\mathbf{p})| \leq \frac{K_{11}(n)}{\rho^n(1-\cos\alpha)^n}\int d^3r|U_\rho(\mathbf{r})| + K_{12}\int_{r>\rho\cos\alpha} d^3r \frac{1}{r^\beta|\mathbf{r}-\rho|}.$$

The first integral is finite,[13] while the second integral can be performed explicitly and has the form $K_{13}/\rho^{\beta-2}$. Therefore,

$$|(S\phi_\rho)(\mathbf{p})| \leq \frac{K_{14}(n)}{\rho^n} + \frac{K_{13}}{\rho^{\beta-2}} \quad [\rho \text{ sufficiently large}],$$

which is the bound (2.13).

ACKNOWLEDGMENTS

We would like to thank Professor T. Kibble for hospitality at Imperial College, where most of this work was done, and Dr. Misra, David Goodmanson, and Mark Semon for helpful discussions. One of use (J.R.T.) thanks the University of Colorado for the award of a faculty fellowship for the academic year 1973-4.

*Supported by CONACYT, Mexico.

[1]J.M. Jauch, Helv. Phys. Acta 31, 127 (1958).
[2]J.M. Jauch, Helv. Phys. Acta 31, 661 (1958).
[3]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U.P., Princeton, N.J., 1971).
[4]J.R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[5]R.G. Newton and R. Shtokhamer, "*N*-Particle Scattering Rates" in *Physical Reality and Mathematical Description*, edited by Enz and Mehra (Reidel, Boston, 1974), p. 286.
[6]J.D. Dollard, Commun. Math. Phys. 12, 193 (1969).
[7]In some experiments a broad beam impinges on a localized target; in others, the beam is narrow and the target broad. These two situations are equivalent because of translational invariance. To be definite, we shall discuss the former.
[8]We consider a finite beam with a finite total number of particles. Thus, n_{inc} is the *number* per unit area, and N_{sc} the actual number of counts. We could, of course, convert both into *rates* by dividing by the duration T of the experiment, but there is no need to do so.
[9]E.H. Wichmann, Am. J. Phys. 33, 20 (1965).
[10]E. Prugovecki, *Quantum Mechanics in Hilbert Space*, (Academic, New York, 1971).
[11]T. Ikebe, Pacific J. Math. 15, 511 (1965).
[12]We use $\psi_{\mathbf{p}}^-(\mathbf{r})$ for the ingoing stationary wave function $\langle \mathbf{r}|\mathbf{p}-\rangle$. This seems to be the most popular notation in text books. [See, for example, E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), Chap. 19.] Ikebe[11] uses $\varphi(x,-k^*)$.
[13]It is easily checked that if V satisfies the assumptions of Sec. 2, then it is in $L^1(R^3)$.

Harmonic oscillator Green's function from a BCH formula

G. Berendt and E. Weimar

Mathematisches Institut, Freie Universität Berlin, West Berlin, West Germany

R. Gilmore

Physics Department, University of South Florida, Tampa, Florida 33620
(Received 20 January 1975)

An integral operator expression is formulated for the n-dimensional harmonic oscillator by exploiting the $U(n)$ symmetry group of the oscillator Hamiltonian. The operator expression is disentangled using a Baker–Campbell–Hausdorff formula appropriate to the dynamical group $Sp(4;\mathbb{R})$ of the Green's function. The BCH formula is computed in a faithful matrix representation of $Sp(4;\mathbb{R})$. It is sufficient to compute the disentangling theorem for the more restricted dynamical group $SO(2,2)$.

In a previous correspondence[1] the Green's function of the n-dimensional harmonic oscillator was computed. The computation exploited the $U(n)$ symmetry of the oscillator Hamiltonian and employed an algebraic derivation of the Green's function matrix elements. We return to this problem to show how these matrix elements can be computed by exploiting a Baker–Campbell–Hausdorff formula for a dynamical group $Sp(4;\mathbb{R})$ associated with the oscillator.

The Hamiltonian for the n-dimensional oscillator can be written

$$H=\frac{1}{2m}\,\mathbf{p}\cdot\mathbf{p}+\frac{m\omega_0^2}{2}\,\mathbf{x}\cdot\mathbf{x}$$

$$=\hbar\omega_0\sum_{k=1}^{n}(a_k^\dagger a_k+\tfrac{1}{2}), \tag{1}$$

where

$$\mathbf{x}=(x_1,\ldots,x_n),\quad \mathbf{p}=\frac{\hbar}{i}\nabla_x=\left(\frac{\hbar}{i}\frac{\partial}{\partial x_1},\ldots,\frac{\hbar}{i}\frac{\partial}{\partial x_n}\right),$$

$$a_k=\frac{1}{\sqrt{2\alpha}}(\alpha x_k+\nabla_k),\quad \alpha=\frac{m\omega_0}{\hbar},$$

$$a_k^\dagger=\frac{1}{\sqrt{2\alpha}}(\alpha x_k-\nabla_k),\quad [a_i,a_k^\dagger]=\delta_{ik}. \tag{2}$$

The n^2 bilinear operator products $a_i^\dagger a_k$ $(1\le i,\ k\le n)$ commute with the Hamiltonian (1) and therefore generate its symmetry algebra, $u(n)$.

It is useful, for dynamical purposes, to introduce the Green's function

$$S(\lambda)=\langle\mathbf{y}\,|\,1/(H-\lambda I)\,|\,\mathbf{x}\rangle\qquad(\lambda\neq\text{eigenvalue of }H) \tag{3}$$

It is convenient to express this in dimensionless coordinates $\mathbf{u}=\sqrt{\alpha}\,\mathbf{x}$, $\mathbf{v}=\sqrt{\alpha}\,\mathbf{y}$:

$$S(\lambda;\mathbf{v},\mathbf{u})=\left(\frac{\alpha}{\pi}\right)^{n/2}\sum_{\nu=0}^{\infty}\frac{1}{2^\nu\nu!}\,\frac{1}{E_\nu-\lambda}\,\{(\mathbf{v}-\nabla_v)\cdot(\mathbf{u}-\nabla_u)\}^\nu$$

$$\times\exp[-(\mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v})/2], \tag{4}$$

where $E_\nu=\hbar\omega_0(\nu+\frac{1}{2}n)$. A useful integral representation for the Green's function is

$$S(\lambda;\mathbf{v},\mathbf{u})=\left(\frac{\alpha}{\pi}\right)^{n/2}\frac{1}{\hbar\omega_0}\int d\xi\;\xi^{n/2-\mu-1}\exp[\tfrac{1}{2}\xi(\mathbf{u}-\nabla_u)\cdot(\mathbf{v}-\nabla_v)]$$

$$\times\exp[-\tfrac{1}{2}(\mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v})], \tag{5}$$

where we have introduced the dimensionless parameter $\mu=\lambda/\hbar\omega_0$ and $\mathrm{Re}(n/2-\mu)>0$. It remains to evaluate the operator expression under the integral.

In order to carry this computation out, we will try to rewrite the exponential operator product as a product of exponential operators in which all the differential operators are ordered to the right and all the operators of the form $\mathbf{u}\cdot\mathbf{u}$, $\mathbf{v}\cdot\mathbf{v}$, $\mathbf{u}\cdot\mathbf{v}$ are ordered to the left. Such operator reorderings can be carried out in closed form whenever the arguments of the exponential functions are elements in a Lie algebra.

The operators appearing in the exponentials in (5), namely $\mathbf{u}\circ\mathbf{v}$, $\mathbf{u}\circ\nabla_v$, $\mathbf{v}\circ\nabla_u$, $\nabla_u\cdot\nabla_v$, $\mathbf{u}\circ\mathbf{u}$, $\mathbf{v}\cdot\mathbf{v}$, do not close under commutation. However, their commutators give rise to four additional operators: $\frac{1}{2}(\mathbf{u}\cdot\nabla_u+\nabla_u\cdot\mathbf{u})=\mathbf{u}\cdot\nabla_u+n/2$, $\frac{1}{2}(\mathbf{v}\cdot\nabla_v+\nabla_v\cdot\mathbf{v})=\mathbf{v}\cdot\nabla_v+n/2$, $\nabla_u\cdot\nabla_u$, $\nabla_v\cdot\nabla_v$. These ten operators close under commutation to give a realization of the real form $sp(4;\mathbb{R})$ of the Lie algebra C_2. As a result, standard[2] matrix techniques can be used to disentangle the product of exponential operators appearing in (5).

To construct the desired BCH formula, we must find a faithful finite-dimensional matrix representation of the ten operators spanning $sp(4;\mathbb{R})$. The smallest such representation clearly consists of 4×4 matrices:

$$\begin{aligned}&A_{11}(\mathbf{u}\cdot\nabla_u+\tfrac{1}{2}n)\\&+A_{22}(\mathbf{v}\cdot\nabla_v+\tfrac{1}{2}n)\\&+A_{12}\,\mathbf{u}\cdot\nabla_v+A_{21}\,\mathbf{v}\circ\nabla_u\end{aligned}\longrightarrow\left[\begin{array}{cc|cc}A_{22}&A_{21}&&\\A_{12}&A_{11}&\multicolumn{2}{c}{0}\\\hline&&-A_{11}&-A_{21}\\\multicolumn{2}{c|}{0}&-A_{12}&-A_{22}\end{array}\right], \tag{6D}$$

$$\begin{aligned}&B_{11}\,\mathbf{u}\cdot\mathbf{u}+B_{22}\,\mathbf{v}\cdot\mathbf{v}\\&+B_{12}\,\mathbf{u}\cdot\mathbf{v}\end{aligned}\longrightarrow\beta\left[\begin{array}{cc|cc}&&B_{12}&2B_{22}\\\multicolumn{2}{c|}{0}&2B_{11}&B_{12}\\\hline&&\multicolumn{2}{c}{}\\\multicolumn{2}{c|}{0}&\multicolumn{2}{c}{0}\end{array}\right], \tag{6U}$$

$$C_{11}\nabla_u\cdot\nabla_u + C_{22}\nabla_v\cdot\nabla_v + C_{12}\,\nabla_u\cdot\nabla_v \xrightarrow{\gamma} \left[\begin{array}{cc|c} 0 & & 0 \\ & & \\ \hline C_{12} & 2C_{11} & \\ 2C_{22} & C_{12} & 0 \end{array}\right] \tag{6L}$$

Here β and γ are scale factors and $\beta\gamma = -1$. The Lie algebra $sp(4;\mathbb{R})$ is generated by all real 4×4 matrices A obeying $JA + A^TJ = 0$, where

$$J = \left[\begin{array}{c|c} 0 & \sigma_x \\ \hline -\sigma_x & 0 \end{array}\right] = \left[\begin{array}{cc|cc} & & 0 & 1 \\ & 0 & 1 & 0 \\ \hline 0 & -1 & & \\ -1 & 0 & & 0 \end{array}\right]. \tag{7}$$

In this 4×4 representation, the product of the exponential operators appearing in (5) is represented by

$$\exp\tfrac{1}{2}\xi\left[\begin{array}{c|c} -\sigma_x & I_2 \\ \hline -I_2 & \sigma_x \end{array}\right] \quad \exp\left[\begin{array}{c|c} 0 & -\sigma_x \\ \hline 0 & 0 \end{array}\right], \tag{8}$$

where $\sigma_x = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and we have chosen $\beta = +1$, $\gamma = -1$. The exponentials appearing in (8) are group elements in $Sp(4;\mathbb{R})$. Moreover, both exponentials are simple to carry out, since both involve exponentials of nilpotent matrices whose squares are zero. The exponential product appearing in (5) is an abstract group element in $Sp(4;\mathbb{R})$ whose matrix representative is obtained simply from (8) by matrix multiplication:

$$\left[\begin{array}{c|c} I_2 - \frac{1}{2}\xi\sigma_x & \frac{1}{2}\xi I_2 \\ \hline -\frac{1}{2}\xi I_2 & I_2 + \frac{1}{2}\xi\sigma_x \end{array}\right] \times \left[\begin{array}{c|c} I_2 & -\sigma_x \\ \hline 0 & I_2 \end{array}\right]$$

$$= \left[\begin{array}{c|c} I_2 - \frac{1}{2}\xi\sigma_x & \xi I_2 - \sigma_x \\ \hline -\frac{1}{2}\xi I_2 & I_2 + \xi\sigma_x \end{array}\right]. \tag{9}$$

We now demand that the group element (9) be written as the product of three group elements, one representing differential operators and ordered on the right, another representing operators of the form $\exp(\mathbf{u}\cdot\mathbf{v} + \text{etc.})$ and ordered on the left, and a third representing exponentials of mixed operators ($\mathbf{u}\cdot\nabla_v$, etc.):

$$\exp\left[\begin{array}{c|c} 0 & R \\ \hline 0 & 0 \end{array}\right] \exp\left[\begin{array}{c|c} \ln M & 0 \\ \hline 0 & \ln N \end{array}\right] \exp\left[\begin{array}{c|c} 0 & 0 \\ \hline L & 0 \end{array}\right]$$

$$= \left[\begin{array}{c|c} I_2 & R \\ \hline 0 & I_2 \end{array}\right] \left[\begin{array}{c|c} M & 0 \\ \hline 0 & N \end{array}\right] \left[\begin{array}{c|c} I_2 & 0 \\ \hline L & I_2 \end{array}\right] = \left[\begin{array}{c|c} M + RNL & RN \\ \hline NL & N \end{array}\right] \tag{10}$$

Equating the group elements (9) and (10) provides four equations which may be solved for the matrices N, R, L, M:

$$N = I_2 + \xi\sigma_x = \exp(-rI_2 - s\sigma_x), \tag{11a}$$

$$R = \frac{2\xi}{1-\xi^2} I_2 - \frac{1+\xi^2}{1-\xi^2}\,\sigma_x, \tag{11b}$$

$$L = \frac{-\frac{1}{2}\xi}{1-\xi^2}\,(I_2 - \xi\sigma_x), \tag{11c}$$

$$M = \frac{1}{1-\xi^2}\,(I_2 - \xi\sigma_x) = \exp(rI_2 + s\sigma_x), \tag{11d}$$

$$e^{-2r} = 1 - \xi^2, \quad \tanh s = -\xi. \tag{12}$$

The matrix BCH formula, valid in $Sp(4:\mathbb{R})$, is

$$\exp\tfrac{1}{2}\xi\left[\begin{array}{c|c} -\sigma_x & I_2 \\ \hline -I_2 & \sigma_x \end{array}\right] \quad \exp\left[\begin{array}{c|c} 0 & -\sigma_x \\ \hline 0 & 0 \end{array}\right]$$

$$= \exp\left[\begin{array}{c|c} 0 & [1/(1-\xi^2)][2\xi I_2 - (1+\xi^2)\sigma_x] \\ \hline 0 & 0 \end{array}\right]$$

$$\times\exp\left[\begin{array}{c|c} rI_2 + s\sigma_x & 0 \\ \hline 0 & -rI_2 - s\sigma_x \end{array}\right] \tag{13}$$

$$\times\exp\left[\begin{array}{c|c} 0 & 0 \\ \hline [-\frac{1}{2}\xi/(1-\xi^2)](I_2 - \xi\sigma_x) & 0 \end{array}\right].$$

The desired operator BCH formula is now obtained by replacing the matrices in (13) by the operators they represent:

$$\exp[\tfrac{1}{2}\xi(\mathbf{u} - \nabla_u)\cdot(\mathbf{v} - \nabla_v)]\,\exp[-\tfrac{1}{2}(\mathbf{u}\cdot\mathbf{u} + \mathbf{v}\cdot\mathbf{v})] = \mathcal{O}(1)\mathcal{O}(2)\mathcal{O}(3) \tag{14}$$

$$\mathcal{O}(1) = \exp\left(\frac{2\xi}{1-\xi^2}\,\mathbf{u}\cdot\mathbf{v} - \frac{1}{2}\,\frac{1+\xi^2}{1-\xi^2}\,(\mathbf{u}\cdot\mathbf{u} + \mathbf{v}\cdot\mathbf{v})\right), \tag{15a}$$

$$\mathcal{O}(2) = \exp[r(\mathbf{u}\cdot\nabla_u + \mathbf{v}\cdot\nabla_v + n) + s(\mathbf{u}\cdot\nabla_v + \mathbf{v}\cdot\nabla_u)], \tag{15b}$$

$$\mathcal{O}(3) = \exp\left(\frac{\frac{1}{2}\xi}{1-\xi^2}\,[\nabla_u\cdot\nabla_v - \tfrac{1}{2}\xi(\nabla_u\cdot\nabla_u + \nabla_v\cdot\nabla_v)]\right). \tag{15c}$$

The matrix elements of the Green's function $S(\lambda;\mathbf{v},\mathbf{u})$ are now obtained by applying the disentangled operator on the right-hand side of (14) to the constant function 1. Clearly,

$$\mathcal{O}(3)\cdot 1 = 1, \tag{16a}$$

$$\mathcal{O}(2)\cdot 1 = (e^{2r})^{n/2} = (1-\xi^2)^{-n/2}. \tag{16b}$$

Thus, only the operator $\mathcal{O}(1)$ remains explicitly in the integral expression for the Green's function:

$$S(\lambda; v, u) = \left(\frac{\alpha}{\pi}\right)^{n/2} \frac{1}{\hbar\omega_0} \int_0^1 d\xi\ \xi^{n/2-\mu-1}(1-\xi^2)^{-n/2}$$

$$\times \exp\left(\frac{2\xi}{1-\xi^2}\ u\cdot v - \frac{1}{2}\,\frac{1+\xi^2}{1-\xi^2}\,(u\cdot u + v\cdot v)\right). \tag{17}$$

Remark: In the final Baker–Campbell–Hausdorff operator formula only six operators appear explicitly: $u\cdot v$, $\frac{1}{2}(u\cdot u + v\cdot v)$; $(u\cdot\nabla_v + v\cdot\nabla_u)$, $(u\cdot\nabla_u + v\cdot\nabla_v + n)$; $\nabla_u\cdot\nabla_v$, $\frac{1}{2}(\nabla_u\cdot\nabla_u + \nabla_v\cdot\nabla_v)$. These six operators span the subalgebra $so(2,2)$ of $sp(4;\mathbb{R})$, corresponding to the Lie algebra restriction $C_2 \downarrow D_2$. Therefore, it would have been sufficient to compute the disentangling theorem for this subgroup only. If we define the generators J_j of the symmetric $so(2,1)$ subalgebra of $so(2,2)$ according to

$$J_1 = \tfrac{1}{2}(u\cdot v - \nabla_u\cdot\nabla_v),$$

$$J_2 = \tfrac{1}{2}(u\cdot\nabla_u + v\cdot\nabla_v + n),$$

$$J_3 = \tfrac{1}{2}(u\cdot v + \nabla_u\cdot\nabla_v) \qquad (J_3 \text{ compact}) \tag{18a}$$

and the remaining generators K_i according to

$$K_1 = \tfrac{1}{4}(u\cdot u + v\cdot v) - \tfrac{1}{4}(\nabla_u\cdot\nabla_u + \nabla_v\cdot\nabla_v),$$

$$K_2 = \tfrac{1}{2}(u\cdot\nabla_v + v\cdot\nabla_u),$$

$$K_3 = \tfrac{1}{4}(u\cdot u + v\cdot v) + \tfrac{1}{4}(\nabla_u\cdot\nabla_u + \nabla_v\cdot\nabla_v) \quad (K_3 \text{ compact}), \tag{18b}$$

we can rewrite (14) as a BCH formula for $so(2,2)$. Since the linear combinations $\frac{1}{2}(J_i + K_i)$ and $\frac{1}{2}(J_i - K_i)$ $(i = 1, 2, 3)$ span mutually commuting subalgebras of $so(2,2)$ corresponding to the Lie algebra decomposition $D_2 = A_1 \oplus A_1$, we finally get the following two BCH formulas for $so(2,1)$, responsible for (14):

$$\exp[\mp\xi(L_2 \mp L_3)]\cdot\exp[\mp(L_1 + L_3)]$$

$$= \exp\left(\mp\frac{1\mp\xi}{1\pm\xi}(L_1 + L_3)\right)\cdot\exp[2(\gamma\pm s)L_2]$$

$$\cdot\exp\left(-\frac{1}{2}\,\frac{\xi}{1\pm\xi}(L_1 - L_3)\right), \tag{19}$$

with $L_i = \frac{1}{2}(J_i + K_i)$ or $\frac{1}{2}(J_i - K_i)$, L_3 compact.

[1]G. Berendt and E. Weimar, Lett. Nuovo Cimento 5, 613 (1972).
[2]R. Gilmore, J. Math. Phys. 15, 2090 (1974).

The de Donder coordinate condition and minimal class 1 space–time

Jorge Krause

Instituto de Física, Universidad Católica de Chile, Casilla 114-D-Santiago, Chile
(Received 11 December 1974)

By means of immersion techniques a set of "adapted coordinates" are introduced as preferred coordinates for class 1 space–time. It is proved that the necessary and sufficient condition for the adapted coordinates to be harmonic coordinates is that class 1 space–time be a minimal variety. Some interesting features of the embedding approach to curved space–time are also shown in terms adapted coordinates.

1. INTRODUCTION

Some arguments on behalf of the embedding approach to general relativity have been presented by the author in a previous paper.[1] In the present note we prove that if class 1 space–time[2] is a minimal variety[3] there exist harmonic coordinates, which represent a kind of preferred coordinates according to a well-defined geometric feature. This property will be proved and briefly discussed by means of immersion techniques.

Einstein's theory of gravitation has an enormous gauge freedom, notwithstanding the fact that the principle of general covariance, by itself, is devoid of physical content, for, indeed, every physical theory may be written in a general covariant manner.[4,5] On the other hand, it is clear that the geometry of some generic space–times may admit a class of preferred coordinates and that the group of general covariance is therefore unnecessarily broad for handling the dynamics of such particular geometries. This point of view has been stressed by Fock[6] in connection with the existence of harmonic coordinates as the preferred ones for some kinds of gravitational problems.

The most favored approach to the issue of preferred coordinates in general relativity consists in fixing the gauge to some extent by imposing a set of *ad hoc* coordinate conditions directly upon the metric tensor.[7] The question thus arises, quite naturally, whether there exist some variational principles leading us to the desired coordinate conditions,[8] i.e., allowing us to obtain a set of "optimal" coordinates defined by a stationary integral property. The study of some extremal behavior of the coordinates, however, is barren so long as we lack the geometric meaning of the variational principle involved.[8] In this paper we show that such a principle exists for fitting harmonic coordinates in curved space–time, at least when the space–time belongs to the simplest embedding class. As is well known, the concept of minimal variety is arrived at by generalizing the definition of minimal surface. Thus (according to the result to be shown in this note), for curved space–time embeddable in five dimensions, the variational principle leading to the harmonic coordinate condition[9] states that the four-dimensional volume integral must be stationary; i.e.,

$$\delta\int d^4x\sqrt{-g(x)}=0. \tag{1.1}$$

It is interesting to observe that if we handle this problem directly, the corresponding Euler–Lagrange equation collapses to a useless identity; namely, we get

$$(w-1/2)[(-g)^w]_{,\mu}\equiv 0, \tag{1.2}$$

for $w=1/2$ in this case.[10] Therefore, a different approach must be followed in order to relate (1.1) with the de Donder coordinate condition.

The present note deals only with space–time geometry. The dynamical content of class 1 space–time will be discussed elsewhere. In Sec. 2 we prepare the five-dimensional immersion scaffolding. Covariant derivatives afforded by the embedding formalism are briefly presented in Sec. 3, while in Sec. 4 we analyze the second fundamental form of the embedded space–time. Finally, in Sec. 5 the connection between minimal class 1 space–time and the existence of harmonic coordinates is proved.

2. THE EMBEDDING SCAFFOLDING AND THE ADAPTED COORDINATES

Let us consider the class 1 relativistic embeddings from a synthetic point of view; namely, we consider the curved space–time manifold as a given four-dimensional (normal hyperbolic) hypersurface $E_{(4)}$, already embedded in a five-dimensional flat space $M_{(5)}$. In $E_{(4)}$ we adopt signature (-2); therefore, according to the local isometric embedding theorem,[11] the pseudo-Euclidean metric tensor belonging to the host space is given by $\eta_{AB}=\mathrm{diag}(+---\pm)$, in terms of a system $\{X^A\}$ of Cartesian rectangular coordinates.[12] We explicitly decompose the fundamental form of $M_{(5)}$ in the following way:

$$ds^2=\eta_{AB}\,dX^A\,dX^B=\eta_{\mu\nu}dX^\mu\,dX^\nu+\eta(dX^4)^2, \tag{2.1}$$

where, clearly, $\eta_{\mu\nu}=\mathrm{diag}(+---)$ denotes the usual four-dimensional Minkowski metric, and $\eta=\eta_{44}=\pm 1$. We now assume that the embedded $E_{(4)}$ is defined by introducing a coordinate relation in $M_{(5)}$ of the form

$$E(X^A)=E(X,X^4)=e, \tag{2.2}$$

say, where e is a constant. Since we want $E_{(4)}$ to be a space–time-like hypersurface, somehow leaning smoothly on the (X^0,X^1,X^2,X^3) hyperplane (i.e., Minkowski space–time $M_{(4)}$), we require

$$\eta\eta^{AB}E_{,A}(X)E_{,B}(X)\big|_{X\in E_{(4)}}>0. \tag{2.3}$$

Hence the unit 5-vector N_A normal to $E_{(4)}$, at points on

$E_{(4)}$, in terms of the $\{X^A\}$ coordinates, obtains:

$$N_A(X)=(\eta\eta^{BC}E_{,B}E_{,C})^{-1/2}E_{,A}(X)\big|_{X\in E_{(4)}}. \tag{2.4}$$

Next we introduce new curvilinear coordinates $\{X^A\}$ in the embedding space. These we choose as the following *adapted coordinates*[13]:

$$\begin{aligned} x^\mu &= X^\mu, \\ x^4 &= E(X^\nu, X^4), \end{aligned} \tag{2.5}$$

so that in terms of these coordinates $E_{(4)}$ is simply given by the equation $x^4=e$ (hence the name). The Jacobian of this transformation is equal to $\partial E/\partial X^4 = E_{,4}$, which we obviously assume to be different from zero. Then for the inverse transformation of (2.5) we write, say,

$$\begin{aligned} X^\mu &= x^\mu, \\ X^4 &= F(x^\nu, x^4). \end{aligned} \tag{2.6}$$

If we now define the function $\phi(x)=F(x^\nu,e)$, we observe, from (2.6), that the following parametric equations hold as a definition of $E_{(4)}$:

$$\begin{aligned} X^\mu &= x^\mu, \\ X^4 &= \phi(x). \end{aligned} \tag{2.7}$$

Hence, for a local isometric embedding we have, as usual,[14]

$$g_{\mu\nu}(x)=X^A{}_{,\mu}X^B{}_{,\nu}\eta_{AB}, \tag{2.8}$$

and thus we get

$$g_{\mu\nu}(x)=\eta_{\mu\nu}+\eta\phi_{,\mu}(x)\phi_{,\nu}(x). \tag{2.9}$$

Henceforth we use $\{X^\mu\}$, i.e., the first four adapted coordinates, as a set of internal curvilinear coordinates for $E_{(4)}$, while retaining the old Cartesian coordinates $\{X^A\}$ in $M_{(5)}$.

The decomposition (2.9) of the metric $g_{\mu\nu}$ holds locally, at least, over that coordinate patch on which representation (2.6), with $x^4=e$, is valid. Furthermore, the stated decomposition of the curved metric has general tensorial character.[15] The adapted coordinates $\{x^\mu\}$, however, are preferred coordinates, as are the Galilean coordinates in spacial relativity, for they bring the *flat part* of the curved metric $g_{\mu\nu}$ to the canonical Minkowskian form $\eta_{\mu\nu}$. In paper I we have shown, for the general local embedding scheme, that the introduction of this special kind of preferred coordinates reduces the general covariance of the theory, investing Einstein's theory of gravitation with a new restricted covariance under a group of transformations which represents an enlargement of the Poincaré group.[15] For class 1 space–time this group corresponds to five-dimensional rotations and translations in $M_{(5)}$.

3. COVARIANT DERIVATIVES IN THE EMBEDDING FORMALISM

In this section we present some useful formulas which will be needed in the following discussion. First we observe that (2.3) means that the determinant $g(x)$ of the space–time metric tensor $g_{\mu\nu}(x)$ has the property

$$-g(x)=1+\eta\eta^{\mu\nu}\phi_{,\mu}(x)\phi_{,\nu}(x)>0, \tag{3.1}$$

everywhere on the embedded patch. The expression for the contravariant metric tensor in $E_{(4)}$, in terms of the adapted coordinates, is

$$g^{\mu\nu}(x)=\eta^{\mu\nu}-\frac{\eta\phi^{,\mu}\phi^{,\nu}}{1+\eta\phi_{,\lambda}\phi^{,\lambda}}, \tag{3.2}$$

where

$$\phi^{,\mu}=\eta^{\mu\nu}\phi_{,\nu}=\frac{\phi^{;\mu}}{1-\eta\phi_{;\lambda}\phi^{;\lambda}}, \tag{3.3}$$

and also

$$\phi^{;\mu}=g^{\mu\nu}\phi_{,\nu}=\frac{\phi^{,\mu}}{1+\eta\phi_{,\lambda}\phi^{,\lambda}}. \tag{3.4}$$

These relations hold because the "fundamental potential" $\phi(x)$ behaves as a scalar field on both (curved and flat) space–times.[16] Furthermore, it is interesting to observe that the fields $g_{\mu\nu}, \eta_{\mu\nu}, \phi_{,\mu}, \phi^{;\mu}$, have tensorial character in both space–times.[17]

Finally, for the Christoffel symbols, in terms of the adapted set $\{x^\mu\}$, we get the expressions

$$\left\{{\lambda \atop \mu\nu}\right\}=\frac{\eta\phi_{,\mu\nu}\phi^{,\lambda}}{1+\eta\phi_{,\sigma}\phi^{,\sigma}}=\frac{\eta\phi_{;\mu;\nu}\phi^{;\lambda}}{1-\eta\phi_{;\sigma}\phi^{;\sigma}} \overset{\text{def}}{=}: \Gamma^\lambda_{\mu\nu}. \tag{3.5}$$

These expressions manifestly define a tensor field $[\Gamma^\lambda_{\mu\nu}(x)$, say] belonging both in $E_{(4)}$ and $M_{(4)}$. Indeed, they represent a space–time tensor whose components, once a set of adapted coordinates is introduced, become identical with the components of the affine connection. Of course, this result is a "virtue" of the adapted coordinates only, for, clearly, we have to transform these quantities differently (as a tensor and as the affine connection) while going to a general set of space–time coordinates. In effect, (3.5) shows that $\Gamma^\lambda_{\mu\nu}$ is that *part* of $\left\{{\lambda \atop \mu\nu}\right\}$ which always transforms as a tensor while using general coordinates; i.e., this decomposition of the affine connection preserves its geometric character under a general transformation of coordinates, since the transformation law obeyed by the Christoffel symbols will not mix up the two parts of the affine connection. This result strongly suggests, as does the concomitant decomposition of the metric, the very special character of the adapted coordinates.

4. THE SECOND FUNDAMENTAL FORM OF CLASS 1 SPACE-TIME

The components of the unit normal to $E_{(4)}$, at points on $E_{(4)}$, in terms of the $\{X^A\}$ coordinates, can be written explicitly as functions of the internal (adapted) coordinates $\{x^\mu\}$. One finds that

$$N_A(X)=\begin{cases} N_\mu=\dfrac{\sigma\phi_{,\mu}}{(1+\eta\phi_{,\nu}\phi^{,\nu})^{1/2}} \\ N^4=\dfrac{\sigma}{(1+\eta\phi_{,\nu}\phi^{,\nu})^{1/2}}, \end{cases} \tag{4.1}$$

where we define

$$\sigma=\frac{E_{,4}(x,\phi(x))}{|E_{,4}(x,\phi(x))|}=\pm 1. \tag{4.2}$$

It is well known from the Riemannian geometry of subspaces that the $X^A{}_{,\mu}$ [cf. Eq. (2.7)] are the components of the unit vectors tangent to the x^μ-parametric lines,

in terms of the $\{X^A\}$ coordinates. These are vectors in $M_{(5)}$ tangent to $E_{(4)}$. As vectors in $E_{(4)}$, their covariant derivatives with respect to g-differentiation are

$$X^A{}_{,\mu;\nu} = \begin{cases} X^\lambda{}_{,\mu;\nu} = -\dfrac{\eta\phi_{,\mu\nu}\phi^{,\lambda}}{1+\eta\phi_{,\sigma}\phi^{,\sigma}} \\ X^4{}_{,\mu;\nu} = \dfrac{\phi_{,\mu\nu}}{1+\eta\phi_{,\sigma}\phi^{,\sigma}}. \end{cases} \tag{4.3}$$

We now calculate the expression for the components of of the tensor which gives us the second fundamental form of the embedded space–time (we call it the *Gauss tensor*), namely,

$$\psi = \Omega_{\mu\nu}\,dx^\mu\,dx^\nu. \tag{4.4}$$

It is well known that, since the $\{X^A\}$ is a Cartesian set, the Gauss tensor is given by[18]

$$\Omega_{\mu\nu} = N_A X^A{}_{,\mu;\nu}. \tag{4.5}$$

Therefore, using (4.1) and (4.3), we get

$$\Omega_{\mu\nu} = \frac{\sigma\phi_{,\mu\nu}}{\sqrt{1+\eta\phi_{,\lambda}\phi^{,\lambda}}} = \frac{\sigma\phi_{;\mu;\nu}}{\sqrt{1-\eta\phi_{;\lambda}\phi^{;\lambda}}}. \tag{4.6}$$

The trace of the space–time Gauss tensor is, thus,

$$\Omega = g^{\mu\nu}\Omega_{\mu\nu} = \frac{\sigma\phi^{;\mu}{}_{;\mu}}{\sqrt{1-\eta\phi_{;\lambda}\phi^{;\lambda}}}. \tag{4.7}$$

5. CONCLUSION; THE ADAPTED COORDINATES AS HARMONIC COORDINATES

We are now in position to prove the connection, stated in the Introduction, between our adapted coordinates and the harmonic coordinates. Indeed, the de Donder condition for harmonic coordinates is

$$(\sqrt{-g}\,g^{\mu\nu})_{,\nu} \equiv -\sqrt{-g}\,g^{\nu\lambda}\left\{{\mu \atop \nu\lambda}\right\} = 0. \tag{5.1}$$

Therefore, according to our previous results, cf. Eqs. (3.2), (3.3), (3.5), and (4.7), we get

$$(\sqrt{-g}\,g^{\mu\nu})_{,\nu} = -\sigma\eta\phi^{,\mu}\Omega. \tag{5.2}$$

Let us recall that the necessary and sufficient condition for a curved space–time (immersed in a five-dimensional space) to be a minimal variety is that the Gauss tensor be traceless. Thus we conclude: For class 1 space–time, the necessary and sufficient condition for the adapted coordinates to be harmonic coordinates is that space–time be a minimal variety. (Clearly so, since $\phi_{,\mu} = 0$ corresponds to Minkowski space–time and affords a trivial model.) Incidentally, this fact tells us that for a minimal class 1 space–time the "fundamental potential" has to satisfy the equation $\phi^{;\mu}{}_{;\mu} = 0$, i.e., the general covariant homogeneous wave equation in the curved space–time manifold generated by the potential itself! We wish to note this fact here, although this paper does not explicitly touch on dynamical questions.

Since the de Donder coordinate condition forms a convenient mathematical tool for treating some problems of general relativity, it is certainly interesting to have a variational principle, with a clear geometric meaning, related to the existence of harmonic coordinates. We have shown, for those space–time metrics which admit a five-dimensional embedding, that the de Donder condition is essentially equivalent to the requirement that curved space–time be a minimal variety. In other words, this means that the underlying variational principle related to harmonic coordinates is that (class 1) space–time must be a solution of the corresponding four-dimensional Plateau's problem.[19] How far we can push this principle into physics we do not know. Let us remark, however, that a hopeful analogy between the soap film minimal surface and the geometry involved in Einstein's field equations was suggested by Wheeler some years ago.[19]

To conclude, we wish to mention here that the detailed study of class 1 space–time deserves some interest in itself for, as is well known, many cosmological solutions to the Einstein field equations belong to this class.[20]

ACKNOWLEDGMENT

This work was strongly influenced by several ideas stated by J.A. Wheeler in a private conversation.

[1] J. Krause, Nuovo Cimento B 18, 302 (1973); hereafter referred to as Paper I.

[2] Let us recall that, in general, an n-dimensional Riemannian metric $g_{\mu\nu}$ is of class p if the minimal flat embedding space of $g_{\mu\nu}$ is of dimensions $N = n+p$, with $n \le N \le (n/2)(n+1)$. This definition of p-embeddability alludes to local embeddings. See A. Friedman, J. Math. Mech. 10, 625 (1961).

[3] L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N.J., 1926), p. 176.

[4] E. Kretchmann, Ann. Phys. 53, 575 (1917).

[5] A. Einstein, Ann. Phys. 55, 241 (1918).

[6] V. Fock, Rev. Mod. Phys. 29, 325 (1957); also V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon, New York, 1964), 2nd. ed.

[7] See, for instance, P.G. Bergmann and A. Komar, Int. J. Theor. Phys. 5, 15 (1972).

[8] J. Plebanski and J. Rytten, J. Math. Phys. 2, 677 (1961).

[9] T. de Donder, *La gravifique Eisteinienne* (Gauthier-Villars, Paris, 1921).

[10] Cf. J. Plebanski and J. Rytten, Ref. 8, Eqs. (2.4) and (2.6) of that paper.

[11] A. Friedman, Ref. 2.

[12] Capital italic suffixes denote tensor indices in $M_{(5)}$; thus $A,B,C,\cdots = 0,1,2,3,4$. Greek suffixes $\mu,\nu,\lambda,\cdots = 0,1,2,3$, have their usual meaning as space–time tensor indices.

[13] See Paper I.

[14] L. P. Eisenhart, Ref. 3, p. 143.

[15] See Paper I for a more comprehensive discussion of the issue.

[16] Cf. Paper I.

[17] The possible role of the flat metric field in curved space-time has been discussed by several authors; cf. W.R. Davis, Nuovo Cimento, B 43, 200 (1966). Also, N. Rosen, Ann. Phys. (N.Y.) 84, 455 (1974).

[18] Cf. L. P. Eisenhart, Ref. 3.

[19] J. A. Wheeler, *Geometrodynamics* (Academic, New York, 1962).

[20] J. Rosen, Rev. Mod. Phys. 37, 204 (1965).

Evaluation of lattice sums. IV. A five-dimensional sum

M. L. Glasser

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Battelle Memorial Institute, Columbus, Ohio 43201

(Received 7 February 1975)

An approach for evaluating lattice sums is presented, which requires the use of basic hypergeometric functions. The sum $\Sigma(x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1 + x_2x_5)^{-s}$ is given as an example.

The motivation for this series of papers[1] has been the eventual "exact" evaluation of physically important three-dimensional lattice sums, principally of the madelung type. Two, not entirely unrelated, procedures were given in Papers I and II based on the use of theta functions and representation theory for integer quadratic forms, respectively. As a result the evaluation of several classes of two-dimensional sums and one three-dimensional sum was achieved. Since then this work has been greatly extended by Zucker,[2,3] who provided the evaluation of a large number of such sums of high even dimension and an exhaustive survey of the simpler sums in two dimensions. In addition, since several of his sums relate to quadratic forms which lie outside of the class considered in II, his work holds promise for yielding new information of value in the theory of numbers. The pessimism, then, that Zucker describes relating to the evaluation of new odd-dimensional sums is discouraging with respect to achieving the initial aims of this research. The motivation of this paper was to reexamine our previous approaches to the problem in order to find a method not subject to the limitation of known theta—or number—theoretic results. We feel that some limited progress has been made and describe here an approach which utilizes the properties of basic hypergeometric series. The method will be illustrated by evaluating a simple five-dimensional lattice sum.

Both the theta function method and the use of number theory provide nothing more than procedures for constructing q-series identities, examples of which are given in I. Another way of systematically obtaining such identities is by means of the basic hypergeometric series. Due to the presence of a number of independent parameters, these functions appear to offer great potentialities. The basic hypergeometric functions along with a large number of applications have recently been described at length in an excellent review by Andrews,[4] and the reader is referred there for details.

Let $(a)_n = (1-a)(1-aq)\cdots(1-aq^{n-1})$. The basic hypergeometric function is

$$ {}_m\phi_n\left[\begin{matrix} a_1,\ldots,a_m \\ b_1,\ldots,b_n \end{matrix}; q, z\right] = \sum_{k=0}^{\infty} \frac{(a_1)_k(a_2)_k\cdots(a_m)_k z^k}{(b_1)_k(b_2)_k\cdots(b_n)_k(q)_k}. $$

By extending Andrews' proof of Heine's theorem [Eq. (2.5) of Ref. 4] to one additional parameter (and in addition using Andrews' Theorem 2.6) we obtain the identity

$$ {}_3\phi_2\left[\begin{matrix} a, b, c \\ d, e \end{matrix}; q, z\right] = \frac{(c)_\infty(az)_\infty(bz)_\infty}{(d)_\infty(e)_\infty(z)_\infty}\, {}_3\phi_2\left[\begin{matrix} d/c, e/c, z \\ az, bz \end{matrix}; q, c\right]. \tag{1} $$

Equation (1) is a generalization of the Thomae formula[5] for the ordinary hypergeometric function ${}_3F_2$ to the basic function. As a particular case we take

$$ z = a = q, \quad b = c = \xi, \quad d = e = q\xi $$

so that

$$ {}_3\phi_2\left[\begin{matrix} q, \xi, \xi \\ q\xi, q\xi \end{matrix}; q, q\right] = \frac{(\xi)_\infty(q^2)_\infty}{(q)_\infty(q\xi)_\infty}\, {}_3\phi_2\left[\begin{matrix} q, q, q \\ q^2, q\xi \end{matrix}; q, \xi\right]. \tag{2} $$

After writing out the sums explicitly and some simplification, (2) becomes

$$ \sum_{n=0}^{\infty}\left[\frac{(\xi)_n}{(q\xi)_n}\right]^2 q^n = \frac{1-\xi}{1-q}\sum_{n=0}^{\infty}\frac{[(q)_n]^2}{(q^2)_n(q\xi)_n}\,\xi^n. \tag{3} $$

However,

$$ \frac{(\xi)_n}{(q\xi)_n} = \frac{1-\xi}{1-\xi q^n} $$

so that (3) reduces to

$$ (1-\xi)\sum_{n=0}^{\infty}\frac{q^n}{(1-\xi q^n)^2} = \sum_{n=0}^{\infty}\frac{(q)_n}{(q\xi)_n}\,\frac{\xi^n}{1-q^{n+1}}; \tag{4} $$

after expanding the denominator on the right-hand side of (4) by the binomial theorem and interchanging the order of summation, we have

$$ \sum_{k=1}^{\infty}\frac{k\xi^{k-1}}{1-q^k} = \frac{1}{1-\xi}\sum_{n=0}^{\infty}\frac{(q)_n}{(q\xi)_n}\,\frac{\xi^n}{1-q^{n+1}}. \tag{5} $$

Finally, we operate on both sides of (5) with $(d/d\xi)\xi$ and take the limit as $\xi \to q$. As a result, we find

$$ \sum_{k=1}^{\infty}\frac{k^2q^k}{1-q^k} = \sum_{k=1}^{\infty}\frac{q^k}{(1-q^k)^2}\left(k + \frac{q}{1-q} + \frac{q^2}{1-q^2} + \cdots + \frac{q^k}{1-q^k}\right) \tag{6} $$

(both sides have also been multiplied by q to render the result more symmetric). Equation (6) is a q-series that apparently has little relationship to a theta function identity.

The two q-series

$$ \eta_1 = \sum_{k=1}^{\infty}\frac{k^{ak}}{(1-q^k)^2}, \quad \eta_2 = \sum_{k=1}^{\infty}\frac{k^2q^k}{(1-q^k)} $$

are familiar in number theory. Indeed, we have

$$ \eta_1 = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} klq^{kl} = \sum_{n=1}^{\infty} n\left(\sum_{d\mid n} 1\right) q^n = \sum_{n=1}^{\infty}\sigma_0(n)\,q^n, $$

where $\sigma_k(n)$ denotes the sum of the kth powers of the divisors of n. Similarly,

$$ \eta_2 = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} k^2q^{kl} = \sum_{n=1}^{\infty}\left(\sum_{d\mid n} d^2\right)q^n = \sum_{n=1}^{\infty}\sigma_2(n)\,q^n $$

is the generating function for the sum of the squares of the divisors of n.

Now we have from (6)

$$S \equiv \sum_{k=1}^{\infty} \frac{q^k}{(1-q^k)^2}\left(\frac{q}{1-q} + \cdots + \frac{q^k}{1-q^k}\right) = \sum_{n=1}^{\infty} [\sigma_2(n) - n\sigma_0(n)]\, q^n. \tag{7}$$

If we now expand the left-hand side of (7) in powers of q, we find

$$S = \sum_{k=1}^{\infty} \sum_{l=1}^{k} \frac{q^k q^l}{(1-q^k)^2(1-q^l)} = \sum_{k=1}^{\infty} \sum_{l=1}^{k} \sum_{s=1}^{\infty} \frac{q^{k+sl}}{(1-q^k)^2} = \sum_{k=1}^{\infty} \sum_{s=1}^{\infty} \sum_{t=1}^{\infty} \sum_{l=1}^{k} t q^{kt+sl}. \tag{8}$$

Next, we note that t = no. of ways of writing $t = u + v$, $u \geq 0$, $v > 0$, so that we have

$$S = \sum_{k=1}^{\infty} \sum_{s=1}^{\infty} \sum_{u=0}^{\infty} \sum_{v=1}^{\infty} \sum_{l=1}^{k} q^{k(u+v)+sl}. \tag{9}$$

Finally, to eliminate the restriction on the l-sum, we write $k = l + r$ to obtain

$$S = \sum_{l=1}^{\infty} \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \sum_{u=0}^{\infty} \sum_{v=1}^{\infty} q^{(l+r)(u+v)+sl}. \tag{10}$$

The sum in (10) is precisely

$$S = \sum_{n=1}^{\infty} R(n)\, q^n, \tag{11}$$

where $R(n)$ is the number of ways of expressing n in the form

$$x_1x_2 + x_1x_4 + x_2x_3 + x_3x_4 + x_2x_5, \quad x_1 \geq 0, x_2, \ldots, x_5 \geq 1$$

As a result of (7) and (11), we see that

$$R(n) = \sigma_2(n) - n\sigma_0(n).$$

Therefore, we have evaluated the five-dimensional lattice sum:

$$\sum_{x_1=0}^{\infty} \sum_{x_2,\ldots,x_5=1}^{\infty} (x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1 + x_2x_5)^{-s} = \sum \frac{\sigma_2(n) - n\sigma_0(n)}{n^s} = \sum_{k=1}^{\infty} \sum_{d=1}^{\infty} \left(\frac{1}{k^s d^{s-2}} - \frac{1}{k^{s-1}d^{s-1}}\right)$$
$$= \zeta(s)\zeta(s-2) - \zeta^2(s-1), \tag{12}$$

where ζ denotes the Riemann zeta function.

Although we have described an alternative approach to the evaluation of lattice sums, we are far from providing an algorithm. Starting from a given sum, it requires a good deal of experimentation to determine a promising starting point, and even then there is an element of luck in the process. The result (12) was found by tracing backwards through (8) to determine the q-series involved and then relating this to the proper basic hypergeometric function. It then required some guesswork before the identity (1) was conjectured. However, our main point is that odd-dimensional lattice sums can be obtained independently of our previous results and that a systematic exploration with the present procedure may provide new three-dimensional results of physical interest.

Note added in proof: Professor G. Andrews has informed me that a special case of (1) which includes (3) has been proven by N. Hall [J. Lond. Math. Soc. **11**, 276 (1936)] and that (6) was a conjecture of E.T. Bell.

ACKNOWLEDGMENTS

The author is grateful to Professor George Andrews for the suggestion to look at basic series and to Professor I.J. Zucker for continuing correspondence.

[1]M.L. Glasser, J. Math. Phys. **14**, 409, 701 (1973); **15**, 188 (1974). The first two will be referred to as I and II, resp.
[2]I.J. Zucker, J. Math. Phys. **15**, 187 (1974).
[3]I.J. Zucker, J. Phys. A **7**, 1568 (1974), and to be published.
[4]G.E. Andrews, SIAM Review **16**, 441 (1974).
[5]W.N. Bailey, *Generalized Hypergeometric Series*, Cambridge Mathematical Tract No. 32 (Cambridge U.P., Cambridge, 1935).

Eigenvalues of the invariant operators of the orthogonal and symplectic groups

M. K. F. Wong

Fairfield University, Fairfield, Connecticut 06430

Hsin-Yang Yeh

Moorhead State College, Moorhead, Minnesota 56560
(Received 2 December 1974)

Eigenvalues of the invariant operators of the orthogonal and symplectic groups have been obtained in closed form. All semisimple Lie groups, the unitary, orthogonal, and symplectic groups, are treated in a systematic way by modifying Perelomov and Popov's method. The eigenvalues of the invariant operators for the orthogonal and symplectic groups are then calculated with reference to the unitary group.

1. INTRODUCTION

The unitary, orthogonal, and symplectic groups have played an important role in the development of molecular, atomic, nuclear, and elementary particle physics. Especially essential are the eigenvalues of the invariant (Casimir) operators of these groups which prescribe their irreducible representations.

The eigenvalues of the invariant operators (C_p) of any degree p for all Lie groups have been extensively studied by many authors, and the literature in this area is extensive. A number of authors obtained formulas for eigenvalues of the invariant operators of low degrees. Gel'fand[1] discussed a system of invariant operators. Racah[2] gave the explicit eigenvalues C_2 for any Lie groups. Umezawa[3] treated C_p up to $p=5$ for $U(n)$, Micu[4] C_4 of $Sp(4)$, and Wong[5] has given recurrence relations of C_4 for $O(n)$. Perelomov and Popov[6,7] first gave the explicit expressions of the eigenvalues of C_p for $U(n)$ in closed form, and Louck and Louck and Biedenharn[8] treated the same subject with a different method. Recently Hudson[9] has obtained a similar expression for $U(n)$. For $O(n)$ Louck[10] has obtained a recurrence relation for the eigenvalues of the invariant operators. Proceeding in the same manner as the unitary group, Perelomov and Popov[6b] have also developed a tensor method to calculate the eigenvalues of the invariant operators for $O(n)$ and $Sp(2n)$. Although with their method the eigenvalues C_p can be calculated in principle by reducing the problem to calculating a known matrix (denoted by A) raised to the pth power, to our knowledge no explicit expressions of C_p in closed form have been obtained.

Gilmore[11] has treated briefly the spectrum of Casimir invariants for the simple classical Lie groups in a unified manner. The equation he uses is

$$C(\mathbf{M}^h) = f(\mathbf{M}^h + \mathbf{R}) - f(\mathbf{R}), \tag{1.1}$$

where $f(x)$ are those terms in the irreducible polynomial invariant which depend on the diagonal group generators H_i only and $\mathbf{R}$ is half the sum of all positive roots of the algebra. This equation was first quoted from Racah's work by Baird and Biedenharn[12] in 1964. Baird and Biedenharn demonstrated that the invariant operators I_2 and I_3 for $SU(3)$ can be obtained from (1.1), and said further that "Racah has asserted that the result we have just demonstrated for $SU(3)$ is true in general." At first sight, therefore, it would seem that Eq. (1.1) gives the complete solution to the eigenvalue problem of the invariants of all classical groups. However, it was pointed out by Partensky[13] in 1972 that this is not so. Partensky has explicitly shown that Eq. (1.1) is only true for I_2 and and I_3 invariants of $SU(n)$ and breaks down for higher order invariants. Furthermore, it is also easy to show that Eq. (1.1) breaks down for higher order invariants of the orthogonal and symplectic groups. For example, it does not work for the eigenvalues I_4 of $O(5)$ and $Sp(4)$. Therefore, Eq. (1.1) is not the most general solution for the eigenvalues of the invariant operators of the classical groups.

The main purpose of this paper is to present explicit expressions of the eigenvalues of the invariant operators for $O(n)$ and $Sp(2n)$ in closed form. We modify Perelomov and Popov's method, treat these groups with reference to $U(n)$, and add correction terms to the A matrix of $U(n)$.

In Sec. 2 the generators and invariant operators of $O(N)$ and $Sp(2n)$ are defined. Our method treats the A matrix for each $O(N)$ and $Sp(2n)$ with reference to $U(N)$ by adding correction terms. In order to do this, the A matrix of $U(N)$ has to be diagonalized. This is carried out in Sec. 3. We then proceed to modify the A matrix of $O(N)$ and $Sp(2n)$ in Sec. 4 by adding (or subtracting) another matrix (called F) from the A matrix of $U(N)$. We then expand $(A \pm F)^p$ in matrix powers of F. The explicit expressions of the eigenvalues of the invariant operators are obtained as a sum over four indices. Examples of $O(4)$ and $Sp(4)$ are given. In Sec. 5 an alternative method of calculating the eigenvalues is shown by directly diagonalizing the A matrices of $O(N)$ and $Sp(2n)$. This method only applies to cases when the value of n is small. Examples are displayed for the cases of $O(3)$ and $O(4)$. The results of the present paper are summarized in Sec. 6.

2. GENERATORS AND INVARIANT OPERATORS OF $O(N)$ AND $Sp(2n)$

We use the same generators and invariants as defined by Perelomov and Popov (Ref. 6b):

$$C_p = X_{i_2}{}^{i_1} X_{i_3}{}^{i_2} \cdots X_{i_1}{}^{i_p}, \tag{2.1}$$

where summation over repeated indices is understood.

The generators are subject to the commutation relations

$$[X_j^i, X_l^k] = \delta_j^k X_l^i - \delta_l^i X_j^k + \begin{cases} \delta_j^{-l} X_{-i}^k - \delta_{-i}^k X_l^{-j} & \text{for } O(N) \\ \epsilon_i \epsilon_j \delta_j^{-l} X_{-i}^k - \epsilon_j \epsilon_k \delta_{-i}^k X_l^{-j} & \text{for } Sp(2n). \end{cases} \tag{2.2}$$

For $O(N)$, the generators X_j^i are related to the X_{ij} defined by Racah in Ref. 2b as follows:

$$X_j^i = X_{l,-j}. \tag{2.3}$$

X_{ij} is directly related to the root vectors. They in turn are connected to the J_{ij} defined by Wong[14] as follows:

$$\begin{aligned} &X_{i,-i} = H_i = J_{2i,2i-1}, \\ &X_{p+q} = -A_p^q, \quad X_{-p-q} = B_p^q, \\ &X_{p-q} = C_p^q \quad (p > q), \\ &X_{q-p} = D_p^q, \quad X_{0+p} = -iE_{2k+1}^p, \quad X_{0-p} = -iF_{2k+1}^p, \end{aligned} \tag{2.4}$$

The invariants of $O(N)$ are usually written in terms of J_{ij}. (See, for example, Racah,[2] Gruber and O'Raifeartaigh,[15] and Wong.[5]) Therefore, at least in appearance they are different from the invariants given by (2.1). However, it turns out that the eigenvalues are the same for both expressions.

The generators of $Sp(2n)$ are those used by Perelomov and Popov. The relationship between these generators and the ones defined by Racah (Ref. 2b) has been explained by Perelomov and Popov in the footnote of their paper.[6]

For the group $O(2n)$, we also introduce the invariant operator C_n' in order to distinguish between the two nonequivalent spinor representations Δ_+ and Δ_-:

$$C_n' = \sum \epsilon_{i_1 j_1 \cdots i_n j_n} X^{i_1 j_1} \cdots X^{i_n j_n}, \tag{2.5}$$

where $\epsilon_{i_1 j_1 \cdots i_n j_n}$ is the totally antisymmetric tensor. The eigenvalues of C_n' have been obtained by Perelomov and Popov already; so they will not be mentioned further.

3. DIAGONALIZATION OF THE A MATRIX OF $U(n)$ AND THE EIGENVALUES C_p FOR $U(n)$

We start from the A matrix of $U(n)$ obtained by Perelomov and Popov:

$$A = (A_{ij}) = \begin{pmatrix} \lambda_1 & -1 & -1 & \cdots & -1 \\ & \lambda_2 & -1 & \cdots & -1 \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ & & & & \cdot \\ 0 & & & & \lambda_n \end{pmatrix}, \tag{3.1}$$

which is an upper triangular matrix with λ_i in the main diagonal and -1 at all other positions above the diagonal, with

$$\lambda_i = m_i + n - i.$$

In order to apply our method to $O(N)$ and $Sp(2n)$, we have to diagonalize this matrix. This is carried out in the following manner. First, we introduce the characteristic matrix $g(\lambda)$ defined by

$$g(\lambda) = (\lambda I - A), \tag{3.2}$$

where I is the unit matrix of order n and its adjoint matrix defined by

$$G(\lambda) = (G_{ij}(\lambda)), \tag{3.3}$$

where $G_{ij}(\lambda)$ is the cofactor of g_{ij} in g. Secondly, $G_{ij}(\lambda_i)$ for a specific value of λ is written in the form of a product of one column matrix (u_{ik}) and one row matrix (v_{ij}) for $k, j = 1, \ldots, n$, i.e.,

$$G(\lambda_i) = \begin{pmatrix} u_{i1} \\ u_{i2} \\ u_{i3} \\ \cdot \\ \cdot \\ \cdot \\ u_{in} \end{pmatrix} (v_{i1} v_{i2} v_{i3} \cdots v_{in}). \tag{3.4}$$

Next, we use the vectors (u_{ik}) and (v_{ik}), obtained from all values of λ, to construct the u and v matrices:

$$u = (u_{ik}), \quad v = (v_{ik}). \tag{3.5}$$

With these matrices and their inverse matrices u^{-1} and v^{-1}, we can show that A can be diagonalized as

$$A = uDu^{-1} = v^{-1}Dv \tag{3.6}$$

where D is the diagonal matrix with the elements $D_{ij} = \lambda_i \delta_{ij}$.

The matrices u and v are both upper triangular whose explicit expressions are

$$\begin{aligned} u_{kh} &= \begin{cases} \prod_{h \geq j > k} (\lambda_h - \lambda_j - 1) \prod_{k > i} (\lambda_h - \lambda_i) & \text{for } h \geq k \\ 0 & \text{for } h < k, \end{cases} \\ v_{kh} &= \begin{cases} \prod_{h \geq j > k} (\lambda_k - \lambda_j - 1) \prod_{i > k} (\lambda_k - \lambda_i) & \text{for } h \geq k \\ 0 & \text{for } h < k, \end{cases} \end{aligned} \tag{3.7}$$

where $\prod_{h \geq j > k}$ (or $\prod_{h > j \geq k}$) $= 1$ if $h \geq j > k$ (or $h > j \geq k$) does not hold, or $h = k$. The inverse matrices u^{-1} and v^{-1} are easily obtained from (3.7) and are also upper triangular matrices:

$$\begin{aligned} u_{kh}^{-1} &= \begin{cases} \dfrac{\prod_{h > j \geq k} (\lambda_k - \lambda_j - 1)}{\prod_{h \geq j \neq k} (\lambda_k - \lambda_j)} & \text{for } h \geq k \\ 0 & \text{for } h < k \end{cases} \\ v_{kh}^{-1} &= \begin{cases} \dfrac{\prod_{h \geq j > k} (\lambda_h - \lambda_j - 1)}{\prod_{h \neq j \geq k} (\lambda_h - \lambda_j)} & \text{for } h \geq k \\ 0 & \text{for } h < k \end{cases} \end{aligned} \tag{3.8}$$

where $\prod_{h > j \geq k}$ (or $\prod_{h \geq j > k}$) $= 1$ for $h = k$. By means of the matrix E ($E_{ij} = 1$ for all i and j) introduced by Perelomov and Popov, the eigenvalues of the invariant operators of $U(n)$ can be calculated by evaluating $\mathrm{Tr}(A^p E)$, giving

$$\begin{aligned} C_p &= \mathrm{Tr}(A^p E) = \mathrm{Tr}(uD^p u^{-1} E) \\ &= \sum_{i=1}^{n} \lambda_i^p \prod_{j \neq i} \frac{\lambda_i - \lambda_j - 1}{\lambda_i - \lambda_j}. \end{aligned} \tag{3.9}$$

This result was first obtained by Popov and Perelomov.[7]

4. EIGENVALUES OF THE INVARIANT OPERATORS FOR $O(N)$ AND $Sp(2n)$

The A matrices for the orthogonal and symplectic groups of any dimension have been obtained by Perelomov and Popov. We start by writing down explicitly the $A_{O(\text{or}S)}$ matrix (O or S refers to the orthogonal or symplectic group respectively) as a sum (or difference) of A, given by (3.1) in the form of $U(N)$, and a correction matrix F.

For $O(2n)$ or $SP(2n)$:

$$A_{O(\text{or}S)} = A \pm F, \tag{4.1}$$

where the sign + or − refers to $O(2n)$ or $Sp(2n)$ respectively and

$$F = \begin{array}{c|cc|cc} & 1 & \cdots & n & n+1 & \cdots & 2n \\ \hline 1 & & & & 1 & & \\ \vdots & & 0 & & & \ddots & \\ n & & & & 1 & & \\ \hline n+1 & & & & & & \\ \vdots & & 0 & & & 0 & \\ 2n & & & & & & \end{array} . \tag{4.2}$$

For $O(2n)$ and $Sp(2n)$ our λ_i is related to λ_i of Ref. 6b (denoted by $\lambda\!\!\!^{\wedge}_i$) by

$$\lambda_i = \lambda\!\!\!^{\wedge}_i \delta_{(\epsilon_{-j}-\epsilon_j)/4+1/2+n+j,\,i}$$

where

$$\epsilon_i = \begin{cases} 1 & \text{for } i>0 \\ 0 & \text{for } i=0 \\ -1 & \text{for } i<0. \end{cases} \tag{4.3}$$

For $O(2n+1)$: $A_O = A + F$ where A is given by Eq. (3.1) with $\lambda_i = \lambda\!\!\!^{\wedge}_j \delta_{i,j+n+1}$

$$\begin{array}{c|ccc|ccc} & 1 & \cdots & n & n+1 & n+2 & \cdots & 2n+1 \\ \hline 1 & & & & & 1 & & \\ \vdots & & 0 & & & & \ddots & \\ n & & & & & 1 & & \\ \hline n+1 & & & & & 0 & & \\ n+2 & & & & & & & \\ n+3 & & & & & & & \\ \vdots & & 0 & & & 0 & & \\ 2n+1 & & & & & & & \end{array} . \tag{4.4}$$

We proceed to calculate the eigenvalue of C_p by evaluating

$$\mathrm{Tr}(A^p_{O(\text{or}S)}E) = \begin{cases} \mathrm{Tr}[(A+F)^pE] & \text{for } O(N) \\ \mathrm{Tr}[(A-F)^pE] & \text{for } Sp(2n). \end{cases} \tag{4.5}$$

In what follows we shall make use of the first expression of (3.6) only. Expanding $(A+F)^p$ into matrix powers of F, we get

$$(A \pm F)^p = A^p \pm \sum_{i=0}^{p-1} A^i F A^{p-i-1}$$

+terms involving products of two or more F's

It is shown in the Appendix that terms involving a product of two or more F's are all zero, so that

$$(A \pm F)^p = A^p \pm \sum_{i=0}^{p-1} A^i F A^{p-i-1}. \tag{4.6}$$

Making use of (4.5), we have for the eigenvalues of C_p:

$$\begin{aligned} \mathrm{Tr}[(A \pm F)^pE] &= \mathrm{Tr}\Big((uDu^{-1})^pE \\ &\quad \pm \sum_{i=0}^{p-1} (uDu^{-1})^i F (uDu^{-1})^{p-i-1}E\Big) \\ &= \mathrm{Tr}(uD^pu^{-1}E) \pm \mathrm{Tr}\Big(u\sum_{i=0}^{p-1} D^iu^{-1}FuD^{p-i-1}u^{-1}E\Big). \end{aligned} \tag{4.7}$$

The first term on the right-hand side of (4.7) is equal to

$$\sum_{i=1}^{N} \lambda_i^p \prod_{j\neq i} \frac{\lambda_i - \lambda_j - 1}{\lambda_i - \lambda_j},$$

which is similar to $U(N)$ except for the values of λ_i, where $N=2n$ for $Sp(2n)$. The second term is the correction term to $U(N)$, so that the eigenvalues of the invariant operators for $O(N)$ and $Sp(2n)$ can be obtained by evaluating this term.

Calculation of the correction term

It is convenient to treat $O(2n)$ and $Sp(2n)$ together, and $O(2n+1)$ separately. The quantities for $O(2n+1)$ will be denoted by the bracket $[+1]$. The calculation will be divided into four steps by evaluating separately (1) $K = u^{-1}E$, (2) $B = \sum_{i=0}^{p-1} D^iu^{-1}FuD^{p-i-1}$, (3) BK, and finally (4) $\mathrm{Tr}(uBK)$.

(1) $K = u^{-1}E$

From the explicit expression of u_{kh} given by (3.7), it follows that

$$(u^{-1}E)_{lj} = \prod_{2n[+1]\geq j>l} (\lambda_l - \lambda_j - 1) \Big/ \prod_{2n[+1]\geq j\neq l} (\lambda_l - \lambda_j) = K_l, \tag{4.8}$$

which is independent of the second index j, i.e., each column has the same elements.

(2) $u^{-1}Fu$

Making use of Eqs. (3.7), (4.2), and (4.4), we obtain

$$(u^{-1}Fu)_{t,n+k[+1]} = \sum_{j=1}^{k} u^{-1}_{t,n-j+1} u_{n+j,n+k[+1]} \tag{4.9}$$

for $t \leq n$ and $1 \leq k \leq n$.

(3) BK

With the help of the results (4.8) and (4.9), we have

$$(BK)_{tk} = \sum_{i=0}^{p-1} \lambda_t^i$$

$$\times\sum_{k=1}^{n}\sum_{j=1}^{k}u^{-1}_{t,n-j+1}u_{n+j,n+k[+1]}K_{n+k[+1]}\lambda^{p-i-1}_{n+k[+1]}$$

$$=\sum_{k=1}^{n}\sum_{j=1}^{k}u^{-1}_{t,n-j+1}u_{n+j,n+k[+1]}K_{n+k[+1]}\frac{\lambda^{p}_{n+k[+1]}-\lambda^{p}_{t}}{\lambda_{n+k[+1]}-\lambda_{t}} \tag{4.10}$$

for $1\leq(t,k)\leq n$ where we have used the relation

$$\sum_{i=0}^{p-1}\lambda^{i}_{t}\lambda^{p-i-1}_{n+k[+1]}=\frac{\lambda^{p}_{n+k[+1]}-\lambda^{p}_{t}}{\lambda_{n+k[+1]}-\lambda_{t}}. \tag{4.11}$$

(4) Tr(uBK)

The correction term becomes

$$\mathrm{Tr}(uBK)=\sum_{r,k,t=1}^{n}u_{rt}\sum_{j=1}^{k}u^{-1}_{t,n-j+1}u_{n+j,n+k[+1]}\frac{\lambda^{p}_{n+k[+1]}-\lambda^{p}_{t}}{\lambda_{n+k[+1]}-\lambda_{t}}$$

$$\times\frac{\prod_{2n[+1]\geq j>n+k[+1]}(\lambda_{n+k[+1]}-\lambda_{j}-1)}{\prod_{2n[+1]>j\neq n+k[+1]}(\lambda_{n+k[+1]}-\lambda_{j})}. \tag{4.12}$$

With the correction term expressed in terms of the λ_i's we obtain from (4.5) and (4.12) the eigenvalues of the invariant operators of any degree p for $O(2n)$, $Sp(2n)$, and $O(2n+1)$:

$$C_{p}=\sum_{i=1}^{2n[+1]}\lambda^{p}_{i}\prod_{j\neq i}\frac{\lambda_{i}-\lambda_{j}-1}{\lambda_{i}-\lambda_{j}}\pm\sum_{r,k,t=1}^{n}u_{rt}$$

$$\times\sum_{j=1}^{k}u^{-1}_{t,n-j+1}u_{n+j,n+k[+1]}\frac{\lambda^{p}_{n+k[+1]}-\lambda^{p}_{t}}{\lambda_{n+k[+1]}-\lambda_{t}}$$

$$\times\frac{\prod_{2n[+1]\geq j>n+k[+1]}(\lambda_{n+k[+1]}-\lambda_{j}-1)}{\prod_{2n[+1]\geq j\neq n+k[+1]}(\lambda_{n+k[+1]}-\lambda_{j})}, \tag{4.13}$$

where u_{rt} is given by (3.7). The eigenvalues of the invariant operators of any degree p have thus been cast in closed form in terms of the λ_i's. Although in (4.13) each term is a fractional function of the λ_i's, the C_p is eventually expressible as a polynomial of the λ_i's.

As an illustration of the expression (4.13) let us evaluate the eigenvalues C_4 ($p=4$) for $O(4)$ and $Sp(4)$. In this case ($n=2$) we choose the + (or −) sign for $O(4)$ [or $Sp(4)$] in (4.13) without the brackets.

The first term of (4.13) is readily evaluated, giving

$$\prod_{i=1}^{4}\lambda^{4}_{i}\prod_{j\neq i}\left(1-\frac{1}{\lambda_{i}-\lambda_{j}}\right)$$

$$=\lambda^{4}_{1}\left(1-\frac{1}{\lambda_{1}-\lambda_{2}}\right)\left(1-\frac{1}{\lambda_{1}-\lambda_{3}}\right)\left(1-\frac{1}{\lambda_{1}-\lambda_{4}}\right)$$

$$+\lambda^{4}_{2}\left(1-\frac{1}{\lambda_{2}-\lambda_{1}}\right)\left(1-\frac{1}{\lambda_{2}-\lambda_{3}}\right)\left(1-\frac{1}{\lambda_{2}-\lambda_{4}}\right)$$

$$+\lambda^{4}_{3}\left(1-\frac{1}{\lambda_{3}-\lambda_{1}}\right)\left(1-\frac{1}{\lambda_{3}-\lambda_{2}}\right)\left(1-\frac{1}{\lambda_{3}-\lambda_{4}}\right)$$

$$+\lambda^{4}_{4}\left(1-\frac{1}{\lambda_{4}-\lambda_{1}}\right)\left(1-\frac{1}{\lambda_{4}-\lambda_{2}}\right)\left(1-\frac{1}{\lambda_{4}-\lambda_{3}}\right). \tag{4.14}$$

It is convenient to use the l_i's, rather than the m_i's. The λ_i's are related to the l_i's as follows:

For $O(4)$:

$$\lambda_1=-l_2+1,\quad \lambda_2=-l_1+1,\quad \lambda_3=l_1+1,\quad \lambda_4=l_2+1;$$

For $Sp(4)$:

$$\lambda_1=-l_2+2,\quad \lambda_2=-l_1+2,\quad \lambda_3=l_1+2,\quad \lambda_4=l_2+2.$$

Substituting the above values of the λ_i's into Eq. (4.14) and rearranging terms, we have the first term of (4.13):

$$\begin{aligned}&2l^{4}_{1}+2l^{4}_{2}-4l^{2}_{1}-4l^{2}_{2} && \text{for } O(4)\\ &2l^{4}_{1}+2l^{4}_{2}+12l^{2}_{1}+12l^{2}_{2}-40 && \text{for } Sp(4).\end{aligned} \tag{4.15}$$

The correction term in (4.13) becomes

$$\mathrm{Tr}(uBK)=u_{11}u^{-1}_{11}u_{44}K_{4}\frac{\lambda^{4}_{4}-\lambda^{4}_{1}}{\lambda_{4}-\lambda_{1}}$$

$$+u_{11}u^{-1}_{12}u_{34}K_{4}\frac{\lambda^{4}_{4}-\lambda^{4}_{1}}{\lambda_{4}-\lambda_{1}}+u_{11}u^{-1}_{12}u_{33}K_{3}\frac{\lambda^{4}_{3}-\lambda^{4}_{1}}{\lambda_{3}-\lambda_{1}}$$

$$+u_{12}u^{-1}_{22}u_{33}K_{3}\frac{\lambda^{4}_{3}-\lambda^{4}_{2}}{\lambda_{3}-\lambda_{2}}+u_{12}u^{-1}_{22}u_{34}K_{4}\frac{\lambda^{4}_{4}-\lambda^{4}_{2}}{\lambda_{4}-\lambda_{2}}$$

$$+u_{22}u^{-1}_{22}u_{33}K_{3}\frac{\lambda^{4}_{3}-\lambda^{4}_{2}}{\lambda_{3}-\lambda_{2}}+u_{12}u^{-1}_{22}u_{34}K_{4}\frac{\lambda^{4}_{4}-\lambda^{4}_{2}}{\lambda_{4}-\lambda_{2}}. \tag{4.16}$$

From (4.8) the column matrix K is given by

$$(K_{i})=\begin{pmatrix}\dfrac{(\lambda_{1}-\lambda_{2}-1)(\lambda_{1}-\lambda_{3}-1)(\lambda_{1}-\lambda_{4}-1)}{(\lambda_{1}-\lambda_{2})(\lambda_{1}-\lambda_{3})(\lambda_{1}-\lambda_{4})}\\ \dfrac{(\lambda_{2}-\lambda_{3}-1)(\lambda_{2}-\lambda_{4}-1)}{(\lambda_{2}-\lambda_{1})(\lambda_{2}-\lambda_{3})(\lambda_{2}-\lambda_{4})}\\ \dfrac{(\lambda_{3}-\lambda_{4}-1)}{(\lambda_{3}-\lambda_{1})(\lambda_{3}-\lambda_{2})(\lambda_{3}-\lambda_{4})}\\ \dfrac{1}{(\lambda_{4}-\lambda_{1})(\lambda_{4}-\lambda_{2})(\lambda_{4}-\lambda_{3})}\end{pmatrix}. \tag{4.17}$$

With the help of the substitutions of (3.7), (3.8), and (4.17) into (4.16), the correction term can easily be calculated in terms of the λ_i's:

$$\mathrm{Tr}(uBK)=\frac{\lambda^{4}_{4}-\lambda^{4}_{1}}{\lambda_{4}-\lambda_{1}}\left(1-\frac{1}{(\lambda_{4}-\lambda_{3})(\lambda_{2}-\lambda_{1})}\right)$$

$$+\frac{\lambda^{4}_{3}-\lambda^{4}_{1}}{\lambda_{3}-\lambda_{1}}\frac{\lambda_{3}-\lambda_{4}-1}{(\lambda_{2}-\lambda_{1})(\lambda_{3}-\lambda_{4})}$$

$$+\frac{\lambda^{4}_{3}-\lambda^{4}_{2}}{\lambda_{3}-\lambda_{2}}\left(1-\frac{1}{(\lambda_{4}-\lambda_{3})(\lambda_{2}-\lambda_{1})}\right)$$

$$+\frac{(\lambda^{4}_{4}-\lambda^{4}_{2})(\lambda_{2}-\lambda_{1}-1)}{(\lambda_{4}-\lambda_{2})(\lambda_{3}-\lambda_{4})(\lambda_{2}-\lambda_{1})}. \tag{4.18}$$

When expressed in terms of the l_i's, the correction terms become

$$\begin{aligned}&2l^{2}_{1}+2l^{2}_{2} && \text{for } O(4)\\ &6l^{2}_{1}+6l^{2}_{2}+24 && \text{for } Sp(4).\end{aligned} \tag{4.19}$$

Substituting the results of (4.15) and (4.19) into (4.13), we obtain the eigenvalues C_4:

$$C_{4}=\begin{cases}2l^{4}_{1}+2l^{4}_{2}-2l^{2}_{1}-2l^{2}_{2} & \text{for } O(4)\\ 2l^{4}_{1}+2l^{4}_{2}+6l^{2}_{1}+6l^{2}_{2}-64 & \text{for } Sp(4)\end{cases} \tag{4.20}$$

The above results agree with those given by Perelomov and Popov [Ref. 6b, Eqs. (20) and (28)].

5. AN ALTERNATIVE METHOD FOR SMALL n

For $O(n)$ and $Sp(2n)$ with small values of n, we may

directly diagonalize the matrix $A_{O(\text{or}S)}$. Although this method cannot be easily generalized to any dimension, it still merits to be displayed. We shall consider the examples of $O(3)$ and $O(4)$ and give only the results for the u' and u'^{-1} matrices which diagonalize the A_O matrix.

For $O(3)$:

$$A_O=\begin{pmatrix}-J & -1 & 0\\ 0 & 1 & -1\\ 0 & 0 & J+1\end{pmatrix}=u'Du'^{-1}, \tag{5.1}$$

where

$$\begin{pmatrix}(J+1)(2J+1) & 2J+1 & 1\\ 0 & J & 1\\ 0 & 0 & 1\end{pmatrix}, \tag{5.2}$$

$$u'^{-1}=\frac{1}{J(J+1)(2J+1)}\times\begin{pmatrix}J & -(2J+1) & J+1\\ 0 & (J+1)(2J+1) & -(J+1)(2J+1)\\ 0 & 0 & J(J+1)(2J+1)\end{pmatrix}. \tag{5.3}$$

For $O(4)$:

$$A_O=\begin{pmatrix}-m_2+1 & -1 & -1 & 0\\ 0 & -m_1 & 0 & -1\\ 0 & 0 & m_1+2 & -1\\ 0 & 0 & 0 & m_2+1\end{pmatrix}, \tag{5.4}$$

where

$$u'=\begin{pmatrix}2m_2ab & -2m_2a & -2m_2b & -2m_2\\ 0 & -(2m_1+2)a & 0 & -(2m_1+2)\\ 0 & 0 & -(2m_1+2)b & 2m_1+2\\ 0 & 0 & 0 & 2m_2\end{pmatrix}, \tag{5.5}$$

$$u'^{-1}=\begin{pmatrix}1 & 1 & 1 & 1\\ 0 & b & 0 & b\\ 0 & 0 & -a & -a\\ 0 & 0 & 0 & -ab\end{pmatrix}. \tag{5.6}$$

Here we have used the abbreviations $a=m_1+m_2+1$, $b=m_1-m_2+1$.

Since the $A_{O(\text{or}S)}$ has been diagonalized, the eigenvalues are given by

$$C_p=\mathrm{Tr}(u'D^pu'^{-1}E).$$

The problem reduces to the evaluation of this trace.

6. CONCLUSION

We have modified Perelomov and Popov's method to calculate the eigenvalues of the invariant operators for $O(N)$ and $Sp(2n)$. The $A_{O(\text{or}S)}$ matrix of $O(N)$ or $Sp(2n)$ are diagonalized with reference to that of $U(N)$ and the correction terms are shown to include only those terms linear in F. By calculating the correction terms, together with the results of $U(N)$, the eigenvalues of the invariant operators for the orthogonal and symplectic groups of any degree p have been obtained in closed form in (4.13) as a sum over four indices. Examples of $O(4)$ and $Sp(4)$ are explicitly shown. An alternative method for calculating the eigenvalues of C_p for small values of n have also been presented. Examples are given for $O(3)$ and $O(4)$.

Since the eigenvalues C_p of $SO(n,1)$ may be evaluated as a special case of $SO(n+1)$,[16] the present method also provides the eigenvalues of the invariant operators for $SO(n,1)$.

APPENDIX

In this Appendix we show that any term in the expansion $(A\pm F)^p$ involving two or more F's must be zero. First A is an upper triangular matrix; therefore, A^n is also an upper triangular matrix, where n is any positive integer. Next F is a matrix which has nonzero elements only in the "first quadrant." We shall denote a matrix with nonzero matrix elements only in the first quadrant a Q matrix. Thus

$$FA=Q_1, \tag{A1}$$

$$AF=Q_2, \tag{A2}$$

$$FA^n=Q_3, \tag{A3}$$

$$A^nF=Q_4, \tag{A4}$$

where the Q_i are all Q matrices. But

$$Q_iQ_j=0 \tag{A5}$$

Hence, the proof; since any term in the expansion of $(A\pm F)^p$ involving two or more F's must be of the form Q_iQ_j, which is a zero matrix.

[1]I.M. Gel'fand, Mat. Sb. **26**, 103 (1950).
[2]G. Racah, (a) Atti Accad. Nazl. Lincei, Rend., Classe Sci. Fis., Mat. Nat. **8**, 108 (1950); (b) Ergeb. exact. Naturw. 37, 28 (1965).
[3]M Umezawa, Nucl. Phys. **48**, 111 (1963); **53**, 54 (1964); **57**, 65 (1964).
[4]M. Micu, Nuclear Phys. **60**, 353 (1964).
[5]M.K.F. Wong, J. Math. Phys. **12**, 1530 (1971).
[6]A.M. Perelomov and V.S. Popov, Yad. Fiz. **3**, (a) 924, (b) 1127 (1966) [Sov. J. Nucl. Phys. **3**, (a) 676, (b) 819 (1966)].
[7]V.S. Popov and A.M. Perelomov, Yad. Fiz. **5**, 693 (1967) [Sov. J. Nucl. Phys. **5**, 489 (1967)].
[8]J.D. Louck, Am.J. Phys. **38**, 3 (1970); J.D. Louck and L.C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[9]R. L. Hudson, J. Math. Phys. **15**, 1067 (1974).
[10]J.D. Louck, Los Alamos Scientific Laboratory, Report LA 2451, 1960.
[11]R. Gilmore, J. Math. Phys. **11**, 1855 (1970).
[12]G.E. Baird and L.C. Biedenharn, J. Math. Phys. **5**, 1723 (1964).
[13]A. Partensky, J. Math. Phys. **13**, 621 (1972).
[14]M.K.F. Wong, J. Math. Phys. **8**, 1899 (1967).
[15]B. Gruber and L. O'Raifeartaigh, J. Math. Phys. **5**, 1796 (1964).
[16]M.K.F. Wong, J. Math. Phys. **15**, 25 (1974).

A Goursat problem for the fourth moment equation*

M. N. Oğuztöreli

Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada T6G 2G1
(Received 15 July 1974)

The solution of a Goursat problem for the pseudoparabolic equation satisfied by the fourth statistical moment of an initially plane wave propagating in a random medium is presented, using an integro–differential equation technique. Two-dimensional propagations are considered.

I. FORMULATION OF THE PROBLEM

Since 1960, a number of contributions have been devoted to the study of the fourth statistical moment of a scalar wave propagating in a random medium (cf. Refs. 1–18).

Following Brown,[6e] we shall denote by I the fourth moment of an initially plane wave. Let ζ, ξ_α, and ξ_β be the dimensionless variables, and γ the parameter which are introduced by Brown. He has shown that $I=I(\xi_\alpha, \xi_\beta, \zeta; \gamma)$ satisfies the pseudoparabolic partial differential equation

$$\frac{\partial I}{\partial \zeta} = i\nabla \xi_\alpha \xi_\beta I - FI \tag{1}$$

where $i=\sqrt{-1}$, ∇_t denotes the gradient with respect to ξ, and $F=F(\xi_\alpha, \xi_\beta; \gamma)$ is a known function defined by Eqs. (26)–(27) below.

In two dimensions, Eq. (1) becomes

$$\frac{\partial I}{\partial \zeta} = i\frac{\partial^2 I}{\partial \xi_\alpha \partial \xi_\beta} - F(\xi_\alpha, \xi_\beta)I \tag{2}$$

where, for simplicity, we do not indicate the obvious dependence of F and I on γ. In this case, the "basic problem" for the fourth moment equation consists of the finding a function $I(\xi_\alpha, \xi_\beta, \zeta)$ which satisfies Eq. (2) in the region $-\infty < \xi_\alpha, \xi_\beta < \infty$, $\zeta > 0$, the "initial condition"

$$I(\xi_\alpha, \xi_\beta, 0) = 1 \tag{3}$$

and the conditions

$$I(\xi_\alpha, \xi_\beta, \zeta) \to \begin{cases} \exp[-\zeta D(\xi_\beta)], & \xi_\alpha \to \infty, \\ \exp[-\zeta D(\xi_\alpha)], & \xi_\beta \to \infty, \end{cases} \tag{4}$$

where D is a given function defined by Eq. (27) below. It is very unfortunate that the treatments in the Refs. 1–18 for the solution of the basic problem are not very satisfactory in the analytical point of view.

Brown,[6e] solving the basic problem (2)–(4) numerically, has plotted the functions

$$\sigma_I^2 = I(0,0,\zeta) - 1, \quad C_I = [I(\xi_\alpha, 0, \zeta) - 1]/\sigma_I^2, \tag{5}$$

where σ_I^2 is the variance of irradiance scintillations and C_I is the covariance function of irradiance normalized by variance. Clearly, Brown's method can also be used for the computation of the covariance function

$$\tilde{C}_I = [I(0, \xi_\beta, \zeta) - 1]/\sigma_I^2. \tag{6}$$

The irradiance scintillations σ_I^2 and the covariance functions of the irradiance scintillations, C_I and $\tilde{C}_I$, are measurable. In this work, using these functions, we made an attempt to find the fourth moment $I(\xi_\alpha, \xi_\beta, \zeta)$ provided that they are measured with sufficient accuracy in the ζ direction. Thus, instead of dealing with the solution of the basic problem for the fourth moment equation (2), we investigate the solution of the associated Goursat problem which consists of the finding of a solution of Eq. (2) satisfying the conditions (5) and (6) for given sufficiently smooth σ_I^2, C_I, and $\tilde{C}_I$. To simplify our presentation, we reformulate the same Goursat problem in the following more general form using different notation:

Find a solution of the partial differential equation

$$\frac{\partial u}{\partial z} = i\frac{\partial^2 u}{\partial x \partial y} - fu \tag{7}$$

for $(x, y, z) \in \mathcal{D}$, satisfying the conditions

$$u|_{x=0} = p, \quad u|_{y=0} = q, \quad u|_{x=y=0} = r, \tag{8}$$

where

$u = u(x, y, z)$ is the unknown function,

$\mathcal{D} = \{(x, y, z) \mid |x| \le a, \ |y| \le b, \ z \ge 0\}$,

$f = f(x, y)$ is a given function continuous in $= \{(x, y) \mid |x| \le a, \ |y| \le b\}$,

$p = p(y, z)$, $q = q(x, z)$, $r = r(z)$ are given functions continuous in $\mathcal{D}$.

We assume that $p(0, z) = q(0, z) = r(z)$ and the function

$$g = g(x, y, z) \equiv p(y, z) + q(x, z) - r(z) \tag{9}$$

admits a continuous partial derivative $\partial^2 g/\partial x \partial y$, and is analytic with respect to z, such that

$$\left|\frac{\partial^k g}{\partial z^k}\right| \le M \quad (k = 0, 1, 2, \cdots) \tag{10}$$

for $(x, y, z) \in \mathcal{D}$, where M is a positive constant.

For example, if $0 \le \lambda_n \le 1$ and if the series $\sum_{n=0}^{\infty} |g_n(x, y)|$ $(\le M)$ and $\sum_{n=0}^{\infty} |\partial^2 g_n/\partial x \partial y|$ are uniformly convergent in $\mathcal{R}$, then the function

$$g(x, y, z) = \sum_{n=0}^{\infty} \exp(-\lambda_n z) g_n(x, y) \tag{11}$$

satisfies the above conditions.

Note that, by virtue of Brown's results on σ_I^2, C_I, and $\tilde{C}_I$, Eqs. (5) and (6), and by Eqs. (3) and (8), we

have the relationships

$$\sigma_I^2 = r(z) - 1, \quad C_I = \frac{q(x,z)-1}{r(z)-1}, \quad \tilde{C}_I = \frac{p(y,z)-1}{r(z)-1}, \tag{12}$$
$$h(0) = 1, \quad \lim_{x\to\infty} q(x,z) = \lim_{y\to\infty} p(y,z) = 1, \quad \lim_{z\to\infty} r(z) = 2.$$

II. SOLUTION

Now, supposing that the Goursat problem (7)—(8) admits a solution $u = u(x,y,z)$, let us integrate Eq. (7) with respect to x from 0 to x, and with respect to y from 0 to y. Then putting

$$(\mathcal{T}u)(x,y,z) = -i\int_0^x\int_0^y \left(f(\xi,\eta)u(\xi,\eta,z) + \frac{\partial u(\xi,\eta,z)}{\partial z}\right) d\xi\, d\eta \tag{13}$$

and using Eqs. (8) and (9), we find the integro-differential equation

$$u = g + \mathcal{T}u. \tag{14}$$

Conversely, we can easily verify that any solution of Eq. (14), which is continuous in $\mathcal{D}$ with a continuous partial derivative with respect to z, is also a solution of Eqs. (7)—(8); i. e., the solution of the initial value problem (7)—(8) is equivalent to the solution of Eq. (14).

The formal solution of Eq. (14) is of the form

$$u = \sum_{n=0}^{\infty} \mathcal{T}^n g. \tag{15}$$

We now show that this formal solution is the actual solution of Eq. (14). Indeed, since

$$\frac{\partial^k}{\partial z^k}[(\mathcal{T}g)(x,y,z)]$$
$$= -i\int_0^x\int_0^y \left(f(\xi,\eta)\frac{\partial^k g(\xi,\eta,z)}{\partial z^k} + \frac{\partial^{k+1} g(\xi,\eta,z)}{\partial z^{k+1}}\right) d\xi\, d\eta,$$

we have, by virtue of the inequalities (10),

$$\left| \frac{\partial^k \mathcal{T}g}{\partial z^k} \right| \le M\frac{(\theta xy)}{(2!)^2} \quad (k = 0,1,2,\cdots), \tag{16}$$

where

$$\theta = 1 + \max_{R} |f(x,y)|. \tag{17}$$

Further, we have

$$\frac{\partial^k}{\partial z^k}[(\mathcal{T}^n g)(x,y,z)]$$
$$= -i\int_0^x\int_0^y \left(f(\xi,\eta)\frac{\partial^k}{\partial z^k}[(\mathcal{T}^{n-1}g)(\xi,\eta,z)] + \frac{\partial^{k+1}}{\partial z^{k+1}}[(\mathcal{T}^{n-1}g)(\xi,\eta,z)]\right) d\xi\, d\eta.$$

We can easily prove, by the mathematical induction, that

$$\left| \frac{\partial^k(\mathcal{T}^n g)}{\partial z^k} \right| \le M\frac{(\theta xy)^n}{(n!)^2} \quad (k,n = 0,1,2,\cdots). \tag{18}$$

Since the series in (15) is dominated by the series

$$M\sum_{n=0}^{\infty}\frac{(\theta xy)^n}{(n!)^2} = MJ_0(2i\sqrt{\theta xy}), \tag{19}$$

where J_0 is the first kind Bessel function of order zero, the series in (15) is absolutely and uniformly convergent in $\mathcal{D}$, and its sum is the actual solution of Eq. (14), as asserted. The asymptotic relationship

$$J_0(2i\sqrt{2\theta xy}) \approx \frac{1}{2\sqrt{\pi}}\,\frac{\exp(2\sqrt{\theta xy})}{\sqrt[4]{\theta xy}} \tag{20}$$

for large $|xy|$, gives an important limitation for the growth of the solution of Eq. (14).

Now, to compute the successive iterates of g more efficiently, we put

$$\mathcal{T}^n g = (-i)^n \sum_{i=1}^{n}\int_0^x\int_0^y M_{n,k}(x,y;\xi,\eta)$$
$$\times[f(\xi,\eta)G_{k-1}(\xi,\eta,z) + G_k(\xi,\eta,z)]\,d\xi\,d\eta, \tag{21}$$

where

$$G_k(x,y,z) = \frac{\partial^k g(x,y,z)}{\partial z^k} \quad (k = 0,1,2,\cdots). \tag{22}$$

Then, we easily verify that

$$M_{1,1}(x,y;\xi,\eta) = 1, \quad M_{2,1}(x,y;\xi,\eta) = \int_\xi^x\int_\eta^y f(\xi',\eta')\,d\xi'\,d\eta',$$
$$M_{2,2}(x,y;\xi,\eta) = \frac{x-\xi}{1!}\,\frac{y-\eta}{1!} = \int_\xi^x\int_\eta^y M_{1,1}(\xi',\eta';\xi,\eta)\,d\xi'\,d\eta',$$
$$M_{3,1}(x,y;\xi,\eta) = \int_\xi^x\int_\eta^y f(\xi',\eta')M_{2,1}(\xi',\eta';\xi,\eta)\,d\xi'\,d\eta',$$
$$M_{3,2}(x,y;\xi,\eta) = \int_\xi^x\int_\eta^y [f(\xi',\eta')M_{2,2}(\xi',\eta';\xi,\eta) + M_{2,1}(\xi',\eta';\xi,\eta)]\,d\xi'\,d\eta',$$
$$M_{3,3}(x,y;\xi,\eta) = \frac{(x-\xi)^2}{2!}\,\frac{(y-\eta)^2}{2!} = \int_\xi^x\int_\eta^y M_{2,2}(\xi',\eta';\xi,\eta)\,d\xi'\,d\eta'.$$

In general, we have the recurrence relations:

$$M_{n+1,1}(x,y;\xi,\eta) = \int_\xi^x\int_\eta^y f(\xi',\eta')M_{n,1}(\xi',\eta';\xi,\eta)\,d\xi'\,d\eta',$$
$$M_{n+1,k}(x,y;\xi,\eta) = \int_\xi^x\int_\eta^y [f(\xi',\eta')M_{n,k}(\xi',\eta';\xi,\eta) + M_{n,k-1}(\xi',\eta';\xi,\eta)\,d\xi'\,d\eta', \quad (k = 2,3,\ldots,n), \tag{23}$$
$$M_{n+1,n+1}(x,y;\xi,\eta) = \int_\xi^x\int_\eta^y M_{n,n}(\xi',\eta';\xi,\eta)\,d\xi'\,d\eta',$$

which can be demonstrated by the mathematical induction. Clearly, all $M_{n,k}$'s are real. Accordingly, we have

$$u(x,y,z) = \sum_{n=0}^{\infty}(\mathcal{T}^n g)(x,y,z)$$
$$= G_0(x,y,z) + \sum_{k=1}^{\infty}\sum_{n=k}^{\infty}(-i)^n\int_\xi^x\int_\eta^y M_{n,k}(x,y;\xi,\eta)$$
$$\times[f(\xi,\eta)G_{k-1}(\xi,\eta,z) + G_k(\xi,\eta,z)\,d\xi\,d\eta$$

or, putting

$$\Gamma_k(x,y;\xi,\eta) = \sum_{n=k}^{\infty}(-i)^n M_{n,k}(x,y;\xi,\eta) \quad (k = 1,2,3,\cdots), \tag{24}$$

we obtain the formula

$$u(x,y,z) = G_0(x,y,z) + \sum_{k=1}^{\infty}\int_0^x\int_0^y \Gamma_k(x,y;\xi,\eta)$$
$$\times[f(\xi,\eta)G_{k-1}(\xi,\eta,z) + G_k(\xi,\eta,z)]\,d\xi\,d\eta. \tag{25}$$

Thus, if the function $g(x,y,z)$ satisfies the conditions

(10), then the solution of the initial value problem (7)–(8) is given by the formula (25).

III. DISCUSSION

According to Brown,[6e] the function F in Eq. (1) is of the form

$$F(\xi_\alpha, \xi_\beta, \gamma) = D_S(\xi_\alpha, \gamma) + D_S(\xi_\beta, \gamma) - 0.5[D_S(\xi_\alpha + \xi_\beta, \gamma) + D_S(\xi_\alpha - \xi_\beta, \gamma)], \tag{26}$$

where

$$D_S(\xi, \gamma) = \begin{cases} 2.91\xi^{5/3} & \text{if } L_0 \text{ is infinite,} \\ 2.61\xi^{5/3}\left(\frac{3}{5} - \frac{(\gamma\xi)^{5/6}K_{5/6}(\gamma\xi)}{2^{5/6}\Gamma_{(11/6)}}\right) & \text{if } L_0 \text{ is finite.} \end{cases} \tag{27}$$

Here L_0 is the outer scale of turbulence and $K_{5/6}$ is the modified Bessel function of order $5/6$.

Clearly F [and consequently f in Eq. (7)] presents singularities for $\xi_\alpha = 0$, $\xi_\beta = 0$, $\xi_\alpha \mp \xi_\beta = 0$ $(x = 0, y = 0, x \mp y = 0)$, and for $\xi_\alpha = \xi_\beta = \infty$ $(x = y = \infty)$. Note that F and f are continuous in any compact domain.

Although f is continuous and bounded in the domain $\mathcal{R}$, its successive derivatives are not continuous. For this reason we avoided from the successive derivatives of f in our analysis.

Further, let us remark that the constant M in (10) can be replaced by a function of z, say $M = M(z)$, continous and bounded in D, without affecting absolute and uniform convergence of the series (15).

In a forthcoming paper we present the solution of the corresponding initial value problem for the more general equation (1).

*Work partially supported by the National Research Council of Canada Grant NRC-A4345 through the University of Alberta and by the Canada Council Leave Fellowship W730016.

[1]M.J. Beran:(a) J. Opt. Soc. Amer. **56**, 1475 (1966); (b) J. Opt. Soc. Amer. **60**, 518 (1970).
[2]M.J. Beran and T.L. Ho: (a) J. Opt. Soc. Amer. **58**, 1335 (1968); (b) J. Opt. Soc. Amer. **59**, 1134 (1969).
[3]M.J. Beran and G.B. Parrent, Jr., *Theory of Partial Coherence*, (Prentice-Hall, Englewood Cliffs, N.J., 1964).
[4]M.J. Beran and A.M. Whitman: (a) J. Opt. Soc. Amer. **60**, 1595 (1970); (b) J. Opt. Soc. Amer. **61**, 1044 (1971).
[5]R.C. Bourret, Nuovo Cimento **26**, 3833 (1962).
[6]W.P. Brown, Jr.: (a) J. Opt. Soc. Amer. **56**, 1045 (1966); (b) J. Opt. Soc. Amer. **57**, 1539 (1967); (c) IEEE Trans. AP-15, 81 (1967); (d) J. Opt. Soc. Amer. **61**, 1051 (1971); (e) J. Opt. Soc. Amer. **62**, 966 (1972).
[7]L.E. Chernov, *Wave Propagation in a Random Medium* (McGraw-Hill, New York, 1960).
[8]I.M. Dagkesamanskaya and V.I. Shishov, Izv. Vuz. Radiofiz. **13**, 16 (1970).
[9]D.L. Fried: (a) J. Opt. Soc. Amer. **57**, 159 (1967); (b) J. Opt. Soc. Amer. 57, 268 (1967); (c) Proc. IEEE **55**, 57 (1967).
[10]M.E. Gracheva, A.S.Gurvich, and M.A. Kallistratova, Izv. Vuz. Radiofiz. **13**, 56 (1970).
[11]T.L. Ho, J. Opt. Soc. Amer. **60**, 667 (1970).
[12]V.I. Klyatskin, Zh. Eksp. Teor. Fiz. **60**, 1300 (1971) [Sov. Phys. JETP **33**, 703 (1971)].
[13]J.E. Molyneux: (a) J. Opt. Soc. Amer. **58**, 951 (1968); (b) J. Opt. Soc. Amer. **61**, 248 (1971); (c) J. Opt. Soc. Amer. **61**, 369 (1971).
[14]C. Roddier and F. Roddier, J. Opt. Soc. Amer. **63**, 661 (1973).
[15]V.I. Shishov, Izv. Vuz. Radiofiz. **11**, 866 (1968).
[16]V.I. Tatarski: (a) *Wave Propagation in a Turbent Medium* (McGraw-Hill, New York, 1961); (b) Zh. Eksp.Teor. Fiz. **46**, 1399 (1964) [Sov. Phys. JETP **19**, 946 (1964)]; (c) Zh. Eksp. Teor. Fiz. **56**, 2106 (1969) [Sov. Phys. JETP **29**, 1133 (1969)].
[17]V.I. Tatarski and V.I. Klyatskin, Zh. Eksp. Teor. Fiz. **58**, 624 (1970). [Sov. Phys. JETP **31**, 335 (1970).
[18]D.A. de Wolf: (a) J. Opt. Soc. Amer. **55**, 812 (1965); (b) J. Opt. Soc. Amer. **63**, 171 (1973).

Two applications of the Racah coefficients of the Poincaré group

W. H. Klink

Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242
(Received 6 February 1975)

Racah coefficients of the Poincaré group are defined and compared with the rotation group. It is then shown how they arise in a natural manner in the inelastic unitarity equations and in crossing multiparticle amplitudes.

I. INTRODUCTION

While the use of Racah coefficients is well known in problems in atomic and nuclear physics,[1] the corresponding coefficients for relativistic systems arising from the Poincaré group are not so well known,[2] in part because they are more complicated and in part because the need has not arisen. The Racah coefficients for the rotation group arise when angular momenta are coupled together in different sequences to form an overall angular momentum state. Corresponding coefficients for the Poincaré group can be defined in a similar fashion. In this paper we will discuss another way of defining the Racah coefficients of the Poincaré group and show how this definition can be used in analyzing the inelastic unitarity equations,[3,4] and in crossing multiparticle amplitudes.[5] We make use of the fact that it is possible to construct multiparticle states of the Poincaré group which are not formed through a sequential or stepwise process in which one state after another is tacked on to the previously coupled states.[3,6] How this is done is discussed in Sec. II. But given such a possibility it is clear that one can single out these "symmetrically" coupled multiparticle states as standards, defining the Racah coefficients as those coefficients which carry the standard states to the sequential states and vice versa. Products of such coefficients then give the conventionally defined Racah coefficients of the Poincaré group.

Now while the possibility of defining such coefficients may be interesting mathematically, there is no intrinsic reason to prefer one type of Racah coefficient over another. The type of coefficient that one uses is always related to the physical problem at hand. For example, if one is analyzing cascade decay processes the sequential or stepwise type of coupling would presumably be preferred. Therefore, after having discussed the mathematical properties of Racah coefficients of the Poincaré group in Sec. II, Sec. III will show how the symmetrical type coefficients arise naturally in the inelastic unitarity equations, while Sec. IV will show how they are used in crossing multiparticle amplitudes.

It should be pointed out that the natural mathematical context in which to discuss these coefficients involves distribution theory;[7] distribution theory is needed because the Poincaré group is noncompact and its basis states nonnormalizable so that the attendant Clebsch–Gordan coefficients and their generalizations are in general distributions. Nevertheless, there are several reasons why distribution theory will not be used in this paper. Firstly, the machinery of distribution theory is sufficiently cumbersome as to hide the simple steps that are needed to derive the Racah coefficients. More importantly, the final result involves only Wigner $O(3)$ functions, which are well-defined functions and not distributions. We thus leave undone the exercise of expressing all relevant quantities in their proper mathematical context.

II. RACAH COEFFICIENTS OF THE POINCARÉ GROUP

To contrast the way in which the "symmetric" Racah coefficients of the Poincaré group will be defined with the more familiar definition, it is useful to begin with a brief discussion of the rotation group.[1] Labeling states of the rotation group as $|[J]m\rangle$, where J is the angular momentum and m the spin projection, we consider the tensor product of three representations $J_1 \otimes J_2 \otimes J_3$. This can be written

$$|[J_1]m_1;\,[J_2]m_2;\,[J_3]m_3\rangle = \sum_{Jm\eta} \begin{bmatrix} J & J_1 & J_2 & J_3 \\ m & m_1 & m_2 & m_3 \\ \eta & & & \end{bmatrix} |[J]m;\eta J_1 J_2 J_3\rangle. \tag{II.1}$$

To determine the degeneracy parameter η, one couples these states in a stepwise fashion, one possibility being

$$|[J_1]m_1;[J_2]m_2\rangle = \sum_{j_{12}m_{12}} \langle j_{12}m_{12}|J_1m_1J_2m_2\rangle\,|[j_{12}]m_{12};J_1J_2\rangle,$$

$$|[J_1]m_1;[J_2]m_2;[J_3]m_3\rangle = \sum_{j_{12}m_{12}} \langle j_{12}m_{12}|J_1m_1J_2m_2\rangle \times |[j_{12}]m_{12};J_1J_2\rangle\,|[J_3]m_3\rangle$$

$$= \sum_{j_{12}m_{12}} \sum_{Jm} \langle j_{12}m_{12}|J_1m_1J_2m_2\rangle \times \langle Jm\,|\,j_{12}m_{12}J_3m_3\rangle\,|[J]m;j_{12}J_1J_2J_3\rangle. \tag{II.2}$$

Comparing Eqs. (II.1) and (II.2) we see that the degeneracy parameter η is j_{12} and

$$\begin{bmatrix} J & J_1 & J_2 & J_3 \\ m & m_1 & m_2 & m_3 \\ j_{12} & & & \end{bmatrix} = \sum_{m_{12}} \langle j_{12}m_{12}|J_1m_1J_2m_2\rangle\langle Jm\,|\,j_{12}m_{12}J_3m_3\rangle. \tag{II.3}$$

But there is no reason that, for example, states 2 and 3 could not have been coupled first, and then 1; Eqs. (II.1) and (II.2) would then yield

$$\begin{bmatrix} J & J_1 & J_2 & J_3 \\ m & m_1 & m_2 & m_3 \\ j_{23} & & & \end{bmatrix} = \sum_{m_{23}} \langle j_{23} m_{23} | J_2 m_2 J_3 m_3 \rangle \langle J m | j_{23} m J_1 m_1 \rangle . \tag{II.4}$$

Since in the tensor product space one basis is equivalent to another, there must be operators which transform one coupling scheme to another. That is,

$$|[J]m; j_{12} J_1 J_2 J_3\rangle = \sum_{j_{23}} \begin{Bmatrix} J_1 & J_2 & j_{12} \\ J_3 & J & j_{23} \end{Bmatrix} |[J]m; j_{23} J_1 J_2 J_3\rangle$$

where $\{\}$ are the Racah coefficients.[1]

A similar procedure has been carried out for the Poincaré group.[2] Instead of states $|[J]m\rangle$ we now have states $|[MJ]\mathbf{p}\sigma\rangle$, where M and J are the mass and spin of a particle and $\mathbf{p}$ and σ are the momentum and spin projection. For simplicity we will assume that the spins of all particles are zero. This makes the analysis much easier to follow; further, the generalization to arbitrary spin is not difficult and is carried out in Ref. 3. Proceeding as was done for the rotation group, one considers the tensor product of states $|[M_1]\mathbf{p}_1\rangle |[M_2]\mathbf{p}_2\rangle |[M_3]\mathbf{p}_3\rangle$ and writes

$$|[M_1]\mathbf{p}_1; [M_2]\mathbf{p}_2; [M_3]\mathbf{p}_3\rangle = \sum\!\!\!\!\!\!\int_{\substack{M J \eta \\ \mathbf{p}\sigma}} \begin{bmatrix} MJ & M_1 & M_2 & M_3 \\ \mathbf{p}\sigma & \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \\ \eta & & & \end{bmatrix} \times |[MJ]\mathbf{p}\sigma; \eta M_1 M_2 M_3\rangle \tag{II.5}$$

where η again refers to the degeneracy parameters. The symbol $\sum\!\!\!\!\!\!\int$ allows for the possibility of a summation or integration over the relevant variables. Now momentum conservation dictates that $\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3$, while energy conservation says $\sqrt{s} = \sqrt{(p_1 + p_2 + p_2)^2}$, where $p_i = (E_i, \mathbf{p}_i)$ with $p_i^2 = M_i^2$ is a 4-vector and J, σ are the spin and spin projections of the multiparticle. Thus there is really a summation or integration only over the parameters J, σ, and η in Eq. (II.5). The degeneracy parameters can be computed in exactly the same way as was done for the angular momentum states; first it is necessary to couple two single particle states:

$$|[M_1]\mathbf{p}_1; [M_2]\mathbf{p}_2\rangle = \sum_{j_{12}\sigma_{12}} (2j_{12}+1)^{1/2} \mathcal{D}^{j_{12}}_{\sigma_{12}0} R[\hat{p}_1(12\mathrm{CM})] \times |[s_{12} j_{12}]\mathbf{p}_1 + \mathbf{p}_2, \sigma_{12}; M_1 M_2\rangle \tag{II.6}$$

where $R[\hat{p}_1(12\mathrm{CM})] \equiv R(\varphi, \theta, 0)$ designates the azimuthal and polar angles of particle 1 in the 1-2 CM frame. Equation (II.6) can be derived in many ways.[3,6] Perhaps the simplest way is to introduce the projection operator $\int dR \mathcal{D}^{J*}_{mm'}(R)\, U(R)$, [where $U(R)$ is the unitary operator representing the rotation R] for it will be used later when three-particle states are discussed. Thus, consider the two-particle state in its rest frame, which is the 1-2 CM frame:

$$|[s_{12} j_{12}]\mathbf{p} = 0, \sigma_{12}; M_1 M_2\rangle \equiv \int dR[\hat{p}_1(12\ \mathrm{CM})] \times \mathcal{D}^{j*}_{\sigma_{12}0}\{R[\hat{p}_1(12\,\mathrm{CM})]\} \times U\{R[\hat{p}_1(12\,\mathrm{CM})]\}\, |\mathbf{p}_1|\hat{z}\rangle ; -|\mathbf{p}_1|\hat{z}\rangle . \tag{II.7}$$

Here the single-particle states are oriented along the z axis; the rotation $R[\hat{p}_1(12\,\mathrm{CM})]$ carries them to the configuration $\mathbf{p}_1$ and $-\mathbf{p}_1$. Then using the orthogonality properties of the Wigner $\mathcal{D}$ functions gives Eq. (II.6).

By making use of Eq. (II.6) a second time it is possible to compute the stepwise coupled three-particle state which gives the analog of Eq. (II.2):

$$\begin{aligned} &|[M_1]\mathbf{p}_1; [M_2]\mathbf{p}_2; [M_3]\mathbf{p}_3\rangle \\ &= \sum_{j_{12}\sigma_{12}} (2j_{12}+1)^{1/2} \mathcal{D}^{j_{12}}_{\sigma_{12}0}\{R[\hat{p}_1(12\,\mathrm{CM})]\} \\ &\quad \times |[s_{12} j_{12}]\mathbf{p}_1 + \mathbf{p}_2, \sigma_{12}; M_1 M_2\rangle\, |[M_3]\mathbf{p}_3\rangle \\ &= \sum_{\substack{j_{12}\sigma_{12} \\ j\sigma}} (2j_{12}+1)^{1/2} (2J+1)^{1/2} \\ &\quad \times \mathcal{D}^{j_{12}}_{\sigma_{12}0}\{R[\hat{p}_1(12\,\mathrm{CM})]\} \mathcal{D}^{J}_{\sigma,\sigma'_{12}}[R(\hat{p}_3)] \\ &\quad \times \mathcal{D}^{j_{12}}_{\sigma'_{12}\sigma_{12}}\{R^{-1}[\hat{p}_3(12\,\mathrm{CM})]\} \\ &\quad \times |[sJ]\mathbf{p}, \sigma; j_{12}\sigma'_{12} s_{12}, M_1 M_2 M_3\rangle \end{aligned} \tag{II.8}$$

where $s = (p_1 + p_2 + p_3)^2$ is the invariant mass and $R(\hat{p}_3)$ the rotation associated with the polar and aximuthal angles of particle 3 in the overall CM frame.

From Eq. (II.8) it is seen that $\eta = \{\sigma'_{12}, j_{12}, s_{12}\}$ so that

$$\begin{bmatrix} [sJ] & M_1 & M_2 & M_3 \\ \mathbf{p}\sigma & \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \\ \eta & & & \end{bmatrix} = \sum_{j_{12}\sigma'_{12}} (2j_{12}+1)^{1/2} (2J+1)^{1/2} \mathcal{D}^{J}_{\sigma,\sigma'_{12}}[R(\hat{p}_3)] \times \mathcal{D}^{j_{12}}_{\sigma'_{12}0}\{R^{-1}[\hat{p}_3(12\,\mathrm{CM})] R[\hat{p}_1(12\,\mathrm{CM})]\} . \tag{II.9}$$

$\sqrt{s_{12}} = \sqrt{(p_1 + p_2)^2}$ is seen to be the invariant mass of the two-particle system, while j_{12} and σ'_{12} are the spin and spin projection, that is the intermediate angular momentum, of the 1 – 2 system. It is to be noted that though the coupling coefficients Eq. (II.9) have been evaluated in the overall CM frame, the same result holds in any other frame.[3]

Finally, as with the angular momentum states, there is no reason for having coupled particles 1 and 2 first; if particles 2 and 3 had been coupled first a result analogous to Eq. (II.9) would result with suitably permuted labels. To compute the Racah coefficients one finally writes

$$|[sJ]\mathbf{p}\sigma; j_{12}\sigma_{12} s_{12}\rangle = \sum\!\!\!\!\!\!\int_{j_{23}\sigma_{23} M_{23}} \begin{Bmatrix} M_1 & M_2 & (j_{12}\sigma_{12} s_{12}) \\ M_3 & (sJ) & (j_{23}\sigma_{23} s_{23}) \end{Bmatrix} \times |[sJ]\mathbf{p}\sigma; j_{23}\sigma_{23} s_{23}\rangle . \tag{II.10}$$

The actual coefficients $\{\}$ contain products of $\mathcal{D}$ functions involving rotations in the 12, 23, and overall CM frames; it can be shown that all these rotations are (in general complicated) functions of the subenergies s_{12}, s_{23}, and s_{13}.

Now while it is seen that the coefficients defined in Eq. (II.10) are indeed the analog of the Racah coefficients of the rotation group, it is nevertheless desirable to find a set of variables which are more closely linked to quantities accessible to experiment. Except for cascade decay processes one usually does not detect the intermediate orbital angular momentum in a reaction, so it is desirable to eliminate such quantities (and their attendant spin projections).

To do so one introduces the generalization of the Omnes variables,[6,8] in which the three degeneracy parameters of a three-particle state are chosen to be two subenergies and a spin projection along a body-fixed axis specified by the three particles. Such a set of variables is most simply obtained by applying the projection operator defined earlier on a three-particle state in its CM frame:

$$|[sJ]\mathbf{p}=0,\sigma;Ms_{12}s_{23}\rangle=\int dR\mathcal{D}^{J*}_{\sigma,M}(R)\times|[M_1]\mathbf{p}_1;[M_2]\mathbf{p}_2;[M_3]\mathbf{p}_3\rangle \tag{II.11}$$

with inverse

$$|[M_1]\mathbf{p}_1;[M_2]\mathbf{p}_2;[M_3]\mathbf{p}_3\rangle=\sum(2J+1)^{1/2}\mathcal{D}^{J}_{\sigma,M}(R)\,|[sJ]\mathbf{p}=0,\sigma;Ms_{23}s_{12}\rangle; \tag{II.12}$$

here R is a rotation from an observer's (or space fixed or double coset) frame to the body-fixed frame (bf), fixed by the three particles. Notice that Eq. (II.11) defines a three-particle state that is symmetric with respect to all three particles and hence contains no intermediate angular momentum labels. One may object that s_{12} and s_{23} single out a choice of particles, but this is not the case since s_{13} is determined from s, s_{12}, and s_{23} by energy–momentum conservation, that is, $s_{12}+s_{23}+s_{13}=s+M_1^2+M_2^2+M_3^2$.

Given Eqs. (II.11) and (II.12), we are in a position to define a new type of Racah coefficient for the Poincaré group, one which singles out one type of coupling (the symmetric coupling) and refers all other types of couplings to it as a standard. For starting with Eq. (II.11), we make use of the stepwise coupled scheme of Eq. (II.8) and the orthogonality relations of the Wigner $\mathcal{D}$ functions:

$$\begin{aligned}
&|[sJ]\mathbf{p}\sigma;Ms_{12}s_{23}\rangle\\
&\quad=\int dR\mathcal{D}^{J*}_{\sigma,M}(R)\,|[M_1]\mathbf{p}_1;[M_2]\mathbf{p}_2;[M_3]\mathbf{p}_3\rangle\\
&\quad=\int dR\mathcal{D}^{J*}_{\sigma,M}(R)\sum_{\substack{j_{12}\sigma_{12}\\ J'\sigma'}}(2j_{12}+1)^{1/2}(2J'+1)^{1/2}\\
&\qquad\times\mathcal{D}^{J'}_{\sigma'\sigma_{12}}[R(\hat{p}_3)]\mathcal{D}^{j_{12}}_{\sigma_{12}0}\{R^{-1}[\hat{p}_3(12\,\mathrm{CM})]R[\hat{p}_1(12\,\mathrm{CM})]\}\\
&\qquad\times|[sJ]\mathbf{p}\sigma';\sigma_{12}j_{12}s_{12}\rangle\\
&\quad=\sum_{\substack{j_{12}\sigma_{12}\\ J'\sigma'\sigma''}}\int dR\mathcal{D}^{J}_{\sigma M}(R)\mathcal{D}^{J'}_{\sigma'\sigma''}(R)\mathcal{D}^{J}_{\sigma''\sigma_{12}}\{R[\hat{p}_3(\mathrm{bf})]\}\\
&\qquad\times(2j_{12}+1)^{1/2}(2J'+1)^{1/2}\mathcal{D}^{j_{12}}_{\sigma_{12}0}\{R^{-1}[\hat{p}_3(12\,\mathrm{CM})]\\
&\qquad\times R[\hat{p}_1(12\,\mathrm{CM})]\}\,|[sJ]\mathbf{p}\sigma;j_{12}\sigma_{12}s_{12}\rangle\\
&\quad=\sum_{j_{12}\sigma_{12}}(2j_{12}+1)\mathcal{D}^{J}_{m,\sigma_{12}}\{R[\hat{p}_3(\mathrm{bf})]\}\\
&\qquad\times\mathcal{D}^{j_{12}}_{\sigma_{12}0}\{R^{-1}[\hat{p}_3(12\,\mathrm{CM})]R[\hat{p}_1(12\,\mathrm{CM})]\}\\
&\qquad\times|[sJ]\mathbf{p}\sigma;\sigma_{12}j_{12}s_{12}\rangle.
\end{aligned} \tag{II.13}$$

$R[\hat{p}_3(\mathrm{bf})]$ is the rotation associated with the polar and azimuthal angles of particle 3 in the body-fixed frame. Use has been made of the group properties of rotations to write $R(\hat{p}_3)=RR[\hat{p}_3(\mathrm{bf})]$. Further, $R^{-1}[\hat{p}_3(12\,\mathrm{CM})]\times R[\hat{p}_1(12\,\mathrm{CM})]$ can be written as $R_zR_y[\theta_{13(12)}]R'_z$, where $\theta_{13(12)}$ is the angle between particles 1–3 in the 12 CM system and R_z, R'_z are z axis rotations that all cancel out in the final result. Quite generally the notation used is to specify polar angles as $\theta_{ij(kl)}$, which means the angle between vectors i and j evaluated in the frame where $\mathbf{p}_k+\mathbf{p}_l=0$. Equation (II.13) becomes

$$|[sJ]\mathbf{p}\sigma;Ms_{12}s_{23}\rangle=\sum_{\substack{j_{12}\\ \sigma_{12}}}(2j_{12}+1)^{1/2}\mathcal{D}^{J}_{M,\sigma_{12}}\{R[\hat{p}_3(\mathrm{bf})]\}\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{13(12)}]\,|[sJ]\mathbf{p}\sigma;\sigma_{12}j_{12}s_{12}\rangle. \tag{II.14}$$

What remains is to compute the inverse of the Racah coefficient, Eq. (II.14), for then any stepwise coupled scheme can be transformed to any other stepwise coupled scheme by using the symmetrically coupled state as an intermediary. To compute the inverse it is convenient to change variables from s_{12}, s_{23} to s_{12}, $\cos\theta_{13(12)}$ since this angle appears explicitly in the $d^{j_{12}}$ function of Eq. (II.14). Now $s_{13}=M_1^2+M_2^2+2(E_1E_3-p_1p_3\cos\theta_{13})$, so it is clear, when evaluated in the (12) frame, that kinematics alone determines the change of variables. Note that E_1 and E_3, when evaluated in the (12) frame, are functions of only s_{12} and s. Then if both sides of Eq. (II.14) are multiplied by the inverse of $\mathcal{D}^{J}_{M,\sigma_{12}}[R(\hat{p}_3)]$ and if $d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{13(12)}]$ is integrated over $\cos\theta_{13(12)}$, making use of the orthogonality properties of the d^J functions, we obtain the final result:

$$\begin{aligned}
&|[sJ]\mathbf{p}\sigma;Ms_{12}\cos\theta_{13(12)}\rangle\\
&\quad=\sum_{\substack{j_{12}\\ \sigma_{12}}}(2j_{12}+1)^{1/2}\mathcal{D}^{J}_{M,\sigma_{12}}[\hat{p}_3(\mathrm{bf})]\\
&\qquad\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{13(12)}]\,|[sJ]\mathbf{p}\sigma;\sigma_{12}j_{12}s_{12}\rangle,\\
&|[sJ]\mathbf{p}\sigma;\sigma_{12}j_{12}s_{12}\rangle=\frac{1}{2}\int_{-1}^{+1}d\cos\theta_{13(12)}\sum_M\mathcal{D}^{J}_{\sigma_{12},M}[\hat{p}_3(\mathrm{bf})]^{-1}\\
&\qquad\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{13(12)}]\\
&\qquad\times|[sJ]\mathbf{p}\sigma;Ms_{12}\cos\theta_{13(12)}\rangle.
\end{aligned} \tag{II.15}$$

III. APPLICATION OF THE RACAH COEFFICIENTS TO THE INELASTIC UNITARY EQUATIONS

It is clear that the Racah coefficients obtained in Eqs. (II.15) will be useful when one wishes to convert from a stepwise coupled scheme to a symmetric scheme and vice versa. Such a situation arises in the inelastic unitarity equations. This can be most readily seen by writing the unitarity equation, $2\mathrm{Im}\mathcal{T}=\mathcal{T}^\dagger\mathcal{T}$, where $\mathcal{T}$ is the reaction operator, in terms of bubble diagrams.[9] For simplicity we will consider only three-particle initial and final states. Then one term in the unitarity equation comes from a three-particle intermediate state:

$$\begin{aligned}
&\textstyle\sum_{\mathrm{inter}}\langle\mathrm{final}\,|\mathcal{T}^\dagger|\,\mathrm{inter}\rangle\langle\mathrm{inter}\,|\mathcal{T}\,|\,\mathrm{initial}\rangle\\
&\quad=\int\frac{d^3p_1}{E_1}\frac{d^3p_2}{E_2}\frac{d^3p_3}{E_3}\langle1''2''3''\,|\mathcal{T}^\dagger|123\rangle\langle123\,|\mathcal{T}\,|1'2'3'\rangle\\
&\quad=\int d^4p\,dR\,\mathcal{T}(s_{12},s_{13})\,ds_{12}\,ds_{13}\langle1''2''3''\,|\mathcal{T}^\dagger|123\rangle\\
&\qquad\times\langle123\,|\mathcal{T}\,|1'2'3'\rangle.
\end{aligned} \tag{III.1}$$

Here the three initial particles are labeled by primes, the intermediate particles are unprimed, while the final particles are labeled with "double primes." When the integration $\int dR$ and energy–momentum delta functions are extracted there results two partial wave amplitudes, suitably integrated and summed over intermediate particle labels[3]:

$$\sum_M \int \mathcal{J}(s_{12}, s_{13})\, ds_{12}\, ds_{13} \mathcal{A}^{\text{final}\to\text{inter}*}(sJ; s_{12}s_{13}M; s''_{12}s''_{13}M'') \times \mathcal{A}^{\text{initial}\to\text{inter}}(sJ; s_{12}s_{13}M; s'_{12}s'_{13}M'). \quad \text{(III.2)}$$

$\mathcal{J}()$ is a relativistically invariant Jacobian resulting from the change of variables in Eq. (III.1) and $\mathcal{A}()$ is the partial wave amplitude of the appropriate reaction.

Now in (III.2) all particles are treated on an equal footing. This is to be contrasted with another term that occurs in the inelastic unitarity equation, in which particle 3′ does not interact with 1′ and 2′, but interacts with intermediate particles 1 and 2. Such a term results from a disconnected diagram in one amplitude and is written

$$\sum\!\!\!\!\!\!\int_{\text{inter}} \langle \text{final}|T^\dagger|\text{inter}\rangle\langle\text{inter}|T_{12}|\text{initial}\rangle = \int \frac{d^3p_1}{E_1}\frac{d^3p_2}{E_2}\frac{d^3p_3}{E_3}\langle 1''2''3''|T^\dagger|123\rangle\langle 12|T_{12}|1'2'\rangle\langle 3|3'\rangle \quad \text{(III.3)}$$

Here all the intermediate particles hooking on to the $T^\dagger$ matrix element are again coupled together in a symmetric fashion; but in the other matrix element, labeled T_{12} meaning the only particles 1 and 2 interact, the third particle is not symmetrically coupled to the other two. $\langle 3|3'\rangle$ gives a 3-momentum delta function which guarantees that particle 3′ remains unchanged in direction and energy in going from the initial to the intermediate state. By again making a change of variables, from $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ to R, p_μ, and s_{12}, s_{13}, with measure $\int dR\, d^4p\, \mathcal{J}\, ds_{12}\, ds_{23}$, and then extracting the energy–momentum delta function and rotation to give partial wave amplitudes, one obtains

$$\sum_M \int \mathcal{J}\, ds_{12}\, ds_{13} \mathcal{A}^{\text{final}\to\text{inter}*}(sJ; s_{12}s_{13}M; s''_{12}s''_{13}M'') \times \langle [sJ]\mathbf{p}\sigma; Ms_{12}s_{13}|T_{12}|[sJ]\mathbf{p}\sigma; M's'_{12}s'_{13}\rangle. \quad \text{(III.4)}$$

But T_{12} operates only on the two-particle subspace of the total three-particle space. Therefore, we use the Racah coefficients, Eqs. (II.15), to insert three-particle states in which particles 1,2 and 1′,2′ are coupled first, and then coupled to particles 3 and 3′, respectively. Thus

$$\begin{aligned}\langle [sJ]\mathbf{p}\sigma; Ms_{12}s_{13}|T_{12}|[sJ]\mathbf{p}\sigma; M's'_{12}s'_{13}\rangle &= \sum_{\substack{j_{12}j'_{12}\\ \sigma_{12}\sigma'_{12}}} \{(2j_{12}+1)(2j'_{12}+1)\}^{1/2} \mathcal{D}^{J*}_{M',\sigma'_{12}}[\hat{p}_3(\text{bf})] \\ &\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{13(12)}] \\ &\times \langle [sJ]\mathbf{p}\sigma; \sigma_{12}j_{12}s_{12}|T_{12}|[sJ]\mathbf{p}\sigma; \sigma'_{12}j'_{12}s'_{12}\rangle \\ &\times \mathcal{D}^{J}_{M,\sigma_{12}}[\hat{p}_3(\text{bf})]\, d^{j'_{12}}_{\sigma'_{12}0}[\cos\theta_{1'3'(1'2')}] \quad \text{(III.5)} \\ &= \sum_{\substack{j_{12}\\ \sigma_{12}}} (2j_{12}+1)\mathcal{D}^{J}_{M',\sigma_{12}}[\hat{p}_3(\text{bf})]\mathcal{D}^{J*}_{M,\sigma_{12}}[\hat{p}_3(\text{bf})] \\ &\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{1'3'(1'2')}] \\ &\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{13(12)}]\mathcal{A}^{1'+2'\to 1+2}(s_{12}, j_{12})\delta(s_{12}-s'_{12}) \quad \text{(III.5)}\end{aligned}$$

where $\mathcal{A}^{1'+2'\to 1+2}(s_{12}j_{12})$ is the partial wave amplitude for the $1'+2'\to 1+2$ reaction, with s_{12} the energy of this reaction and j_{12} the angular momentum. When (III.5) is combined with (III.4) the expression for this term in the inelastic unitarity equation becomes

$$\begin{aligned}&\sum_M \int \mathcal{J}\, ds_{12}\, ds_{13} \mathcal{A}^{\text{final}\to\text{inter}*}(sJ; s_{12}s_{13}M; s''_{12}s''_{13}M'') \\ &\times \sum_{\substack{j_{12}\\ \sigma_{12}}} (2j_{12}+1)\mathcal{D}^{J}_{M',\sigma_{12}}[\hat{p}'_3(\text{bf})]\mathcal{D}^{J}_{M,\sigma_{12}}[\hat{p}_3(\text{bf})] \\ &\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{1'3'(1'2')}] \\ &\times d^{j_{12}}_{\sigma_{12}0}[\cos\theta_{13(12)}]\mathcal{A}^{1'2'\to 12}(s_{12}, j_{12})\delta(s_{12}-s'_{12}). \quad \text{(III.6)}\end{aligned}$$

If the body-fixed (bf) frame is chosen so that particle 3 (particle 3′) points along the z axis, then both of the $\mathcal{D}^J$ functions collapse to Kronecker deltas. Of course such a choice is only possible for one of the three disconconected terms. There will in general be three terms like (III.6) in the unitarity equation, corresponding to particles 1,2, or 3 not interacting with the other two particles. For particle 3 (particle 3′) along the z axis (z' axis) (III.6) simplifies to

$$\begin{aligned}&\sum_M \int \mathcal{J}\, ds_{12}\, ds_{13} \mathcal{A}^{\text{final}\to\text{inter}*}(sJ; s_{12}s_{13}M; s''_{12}s''_{13}M'') \\ &\times \sum_{j_{12}} (2j_{12}+1)\, d^{j_{12}}_{M',0}[\cos\theta_{13(12)}]\mathcal{A}^{1'+2'\to 1+2}(s_{12}, j_{12}) \\ &\times d^{j_{12}}_{M',0}[\cos\theta_{1'3'(1'2')}]\delta(s_{12}-s'_{12})\delta_{MM'}.\end{aligned}$$

IV. RACAH COEFFICIENTS AS APPLIED TO CROSSING MULTIPARTICLE AMPLITUDES

In the previous section it was shown how Racah coefficients could be used to handle disconnected diagrams in the inelastic unitarity equations. In this section we continue the analysis of multiparticle partial wave amplitudes by considering their properties under crossing.

To begin we review previous work[5] in which it is shown how the crossing of any multiparticle amplitude involves analytic continuation in only one variable; such a result is valid only if a canonical set of variables is used and this set of variables involves precisely the variables entering the Racah coefficients. Here it should be pointed out that crossing means crossing one particle at a time in a multiparticle reaction. The more conventional use of the term crossing, in which one incoming and one outgoing particle are crossed, then results from two sequential one-particle crossings; the goal of this section is to show that the conventional crossing of two particles involves the Racah coefficients in a crucial way.

To make the following discussion as straightforward as possible we consider only the simplest multiparticle amplitude, in which two particles react to produce three particles; that is, we consider the reaction $1'+2' \to 1+2+3$. The generalization to arbitrary multiparticle

amplitudes is straightforward because of the structure of the Racah coefficients for N-particle states.

The amplitude for $1'+2' \to 1+2+3$ is proportional to $\langle 123|T|1'2'\rangle$:

$$\bar{A}^{1'+2'\to 1+2+3}[s_{1'2'}, \cos\theta_{1'3(1'2')}\varphi_{1'3-31}, s_{12}, s_{13}] \propto \langle 123|T|1'2'\rangle$$

$$= \sum_{\substack{J_{1'2'}\\ M}} \mathcal{D}^{J_{1'2'}}_{M,0}[R(\hat{p}_{1'})]\bar{\mathcal{A}}^{1'+2'\to 1+2+3}(s_{1'2}J_{1'2}, Ms_{12}s_{23}) \quad \text{(IV.1)}$$

where $\theta_{1'3(12)}$ and $\varphi_{1'3-31}$ are polar and azimuthal angles relating the line formed by the two incoming particles to the plane formed by the three outgoing particles; the angles are easily related to the relativistic invariants $t_{1'3}=(p_{1'}-p_3)^2$ and $t_{1'1}=(p_{1'}-p_1)^2$. The other two variables arise as the two subenergies of the symmetrically coupled three-particle state. $\bar{\mathcal{A}}^{1'+2'\to 1+2+3}$ is the partial wave amplitude labeled by the total angular momentum $J_{1'2'}$ and spin projection M.

To change to canonical variables for crossing the amplitude (IV.1) it is necessary to choose which particle is to be crossed. We eventually want to cross to the reaction $2'+\bar{3}\to\bar{1}'+1+2$ so we begin the one-particle crossing by crossing particle 3 to the reaction $1'+2'+\bar{3}\to 1+2$. Then the amplitude for $1'+2'\to 1+2+3$ should contain variables in which 1 and 2 are first coupled together, and then the two-particle state coupled to 3, as was done in Eq. (II.8). A simple change of variables, from s_{13} to $\cos\theta_{13(12)}$ [see Eq. (II.14)], gives

$$A^{1'+2'\to 1+2+3}(s_{1'2'}, \cos\theta_{1'3(12)}, \varphi_{1'3-31}, s_{12}, \cos\theta_{13(12)})$$

$$= \sum_{\substack{J_{1'2'}\\ M}} \mathcal{D}^{J_{1'2'}}_{M,0}[R(\hat{p}_1)]\mathcal{A}^{1'+2'\to 1+2+3}[s_{1'2'}J_{1'2'}, Ms_{12}\cos\theta_{13(12)}]$$

$$= \sum_{\substack{J_{1'2'}\\ J_{12}\\ M}} d^{J_{1'2'}}_{M,0}[\cos\theta_{1'3(12)}]\exp(-iM\varphi_{1'3-31})d^{J_{12}}_{M,0}[\cos\theta_{13(12)}]$$

$$\times \mathcal{A}^{1'+2'\to 1+2+3}(s_{1'2'}, J_{1'2'}, M, s_{12}, J_{12}). \quad \text{(IV.2)}$$

Notice that $d^{J_{12}}_{M,0}[\cos\theta_{13(12)}]$ is part of the kernel carrying a stepwise coupled state to a symmetric state as seen in Eq. (II.14). In Ref. 5 it is shown that the variables in the amplitude of Eq. (IV.2) are well suited for crossing to the $1'+2'+3'\to 1+2$ reaction, for when the substitution rule is used in which $p_3\to -p_{\bar{3}}$, then $\cos\theta_{1'3(12)}\to -\cos\theta_{1'\bar{3}(12)}$, $\cos\theta_{13(12)}\to -\cos\theta_{1\bar{3}(12)}$, and $\varphi_{1'3-31}$ goes into itself. Thus the substitution rule generates a transformation in which the two polar angles stay in their physical region, but change their physical meaning. Further, if $\sqrt{s_{12}}$ is chosen to be greater than $M_{1'}+M_{2'}+M_3$, then under crossing (in which s_{12} changes from a subenergy to the total energy) it also stays in its physical region and there is no analytic continuation of $A^{1'+2'\to 1+2+3}$ in terms of s_{12}. Analytic continuation of $A^{1'+2'\to 1+2+3}$ is required only with respect to $s_{1'2'}$ as this variable is the total energy in the direct channel, but a subenergy in the crossed channel. The sense in which $A^{1'+2'\to 1+2+3}$ can be analytically continued to give the crossed amplitude is discussed in Ref. 5; here we merely write

$$A^{1'+2'+\bar{3}\to 1+2}[s_{12}, \cos\theta_{1\bar{3}(12)}, \varphi_{1\bar{3}-\bar{3}1'}, s_{1'2'}, \cos\theta_{1'\bar{3}(1'2')}]$$

$$= A^{1'+2'\to 1+2+3}_{\text{cont}}[s_{1'2'}-\cos\theta_{1'3(1'2')}, \varphi_{1'3-31}, s_{12}, -\cos\theta_{13(12)}] \quad \text{(IV.3)}$$

where $A^{1'+2'\to 1+2+3}_{\text{cont}}[\,]$ means the analytic continuation with respect to $s_{1'2'}$ from the physical region where $s_{1'2'}$ is the total energy to the physical region where it is a subenergy. The crossing condition for the partial wave amplitude is even simpler:

$$\mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}J_{12}Ms_{1'2'}J_{1'2'}) = (-1)^{J_{12}+J_{1'2'}}$$

$$\times \mathcal{A}^{1'+2'\to 1+2+3}_{\text{cont}}(s_{1'2'}J_{1'2'}Ms_{12}J_{12}); \quad \text{(IV.4)}$$

that is, one merely interchanges to role of J_{12} and $J_{1'2'}$ (though, of course, still analytically continuing in $s_{1'2'}$). It is to be noticed that crossing one particle rather than two as is conventionally done means it is possible to cross partial wave amplitudes rather than just amplitudes. This is of importance when combining crossing with the inelastic unitarity equations, terms of which were given in Sec. III, for the content of unitarity is best expressed in terms of partial wave amplitudes.

It is clear that the canonical variables for crossing one particle depend on which particle is to be crossed and arise as the variables of the Racah coefficients. This suggests that when crossing two particles the Racah coefficients are needed just to move from the coupling scheme involving one set of canonical variables to the coupling scheme involving another canical set. This is seen most readily by noting that the canonical variables for the reaction $1'+2'+\bar{3}\to 1+2$, in which $\bar{3}$ has just been crossed, are $(s_{12}, J_{12}, M_{\bar{3}}, s_{1'\bar{2}'}, J_{1'2'})$, while the canonical variables for $1'+2'+\bar{3}\to\bar{1}+2$, in which $1'$ is to be crossed to the reaction $2'+\bar{3}\to\bar{1}'+1+2$, are $(s_{12}, J_{12}, M_1', s_{2'\bar{3}}, J_{2'\bar{3}})$.

More concretely, we start with the partial wave amplitude $\mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M_{\bar{3}}s_{1'2'}, J_{1'2'})$ of Eq. (IV.4) and use the Racah coefficients to get the amplitude $\mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M_1, s_{2'\bar{3}}, J_{2'\bar{3}})$:

$$\mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M_{1'}, s_{2'\bar{3}}, J_{2'\bar{3}}) = \langle 12|S_{s_{12}J_{12}}|1'(2'\bar{3})\rangle$$

$$= \sum_M \int d\theta_{\bar{3}1'(2'\bar{3})}\mathcal{D}^{J_{12}}_{M_{1'},M}(\hat{p}_{1'})^{-1} d^{J_{2'\bar{3}}}_{M_{1'},0}[\cos\theta_{\bar{3}1'(2'\bar{3})}]$$

$$\times \mathcal{A}^{1'+2'+\bar{3}\to 1+2}[s_{12}, J_{12}, M, s_{2'\bar{3}}, \theta_{\bar{3}1'(2'\bar{3})}]. \quad \text{(IV.5)}$$

Now make a change of variable in the subenergies, from $s_{2'\bar{3}}, \cos\theta_{\bar{3}1'(2'\bar{3})} \to s_{2'\bar{3}}, s_{1'2'} \to \cos\theta_{1'\bar{3}(1'2')}, s_{1'2'}$ using the fact that

$$s_{ij} = (p_i+p_j)^2 = M_i^2+M_j^2+2(E_iE_j - p_ip_j\cos\theta_{ij}) \quad \text{(IV.6)}$$

and E_i, E_j, p_i, and p_j are functions only of s_{ij} and s_{12} when evaluated in the frame given in the parentheses after the angle. Thus

$$\mathcal{A}^{1'+2'+\bar{3}\to 1+2}[s_{12}, J_{12}, M, s_{2'\bar{3}}, \cos\theta_{1'\bar{3}(2'\bar{3})}]$$

$$= \mathcal{A}^{1'+2'+\bar{3}\to 1+2}[s_{12}, J_{12}, M, s_{1'2'}, \cos\theta_{1'\bar{3}(1'2')}] \quad \text{(IV.7)}$$

and the inverse Racah coefficient, Eq. (II.15), can be used to write $\mathcal{A}^{1'+2'+3\to 1+2}$ in those canonical variables needed to cross particle $\bar{3}$:

$$\mathcal{A}^{1'+2'+\bar{3}\to 1+2}[s_{12}, J_{12}, M, s_{1'2'}, \cos\theta_{1'\bar{3}(1'2')}]$$
$$= \sum_{\substack{J_{1'2'} \\ M_{\bar{3}}}} \mathcal{D}^{J_{12}}_{M,M_{\bar{3}}}(p_{\bar{3}})\, d^{J}_{M_{\bar{3}},0}[\cos\theta_{1'\bar{3}(1'2')}]$$
$$\times \mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M_{\bar{3}}, s_{1'2'}, J_{1'2'}). \qquad \text{(IV.8)}$$

Equations (IV.5)–(IV.8) can be combined to give

$$\mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M_{\bar{3}}, s_{1'2'}, J_{1'2'})$$
$$\to \mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M'_1, s_{2'\bar{3}}, J_{2'\bar{3}}) \qquad \text{(IV.9)}$$

where the arrow denotes the sequence of operations involving the Racah coefficient, the change of variable, and the inverse Racah coefficient.

To finally cross $1'+2' \to 1+2+3$ to $2'+\bar{3} \to \bar{1'}+1+2$ we cross $\mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M_{1'}, s_{2'\bar{3}}, J_{2'\bar{3}})$ as was done in Eq. (IV.4) to get

$$\mathcal{A}^{2'+\bar{3}\to \bar{1'}+1+2}(s_{2'\bar{3}}, J_{2'\bar{3}}, M_{1'}, s_{12}, J_{12})$$
$$= (-1)^{J_{12}+J_{2'\bar{3}}} \mathcal{A}^{1'+2'+\bar{3}\to 1+2}_{\text{cont}}(s_{12}, J_{12}, M_{1'}, s_{2'\bar{3}}, J_{2'\bar{3}}). \qquad \text{(IV.10)}$$

Putting all these steps together then gives

$$\mathcal{A}^{1'+2'\to 1+2+3}(s_{1'2'}, J_{1'2'}, M_{\bar{3}}, s_{12}, J_{12})$$
$$\to \mathcal{A}^{1'+2'+\bar{3}\to 1+2}(s_{12}, J_{12}, M_{\bar{3}}, s_{1'2'}, J_{1'2'})$$
(analytic continuation in $s_{1'2'}$)
$$\to \mathcal{A}^{1'+2'+\bar{3}\to 1+2}_{\text{cont}}(s_{12}, J_{12}, M_{1'}, s_{2'\bar{3}}, J_{2'\bar{3}})$$
(use of Racah coefficients)
$$\to \mathcal{A}^{2'+\bar{3}\to \bar{1'}+1+2}_{\text{cont}}(s_{2'\bar{3}}, J_{2'\bar{3}}, M_{1'}, s_{12}, J_{12})$$
(analytic continuation in s_{12}) $\qquad$ (IV.11)

and shows, assuming that analytic continuation is possible, that the Racah coefficients provide the transformations between the correct canonical variables. Note that the angular momentum label is not preserved under the double crossing Eq. (IV.11) as it is in the single crossing Eq. (IV.10), which is of course why one conventionally crosses amplitudes rather than partial wave amplitudes.

V. CONCLUSION

It has been shown how one may define Racah coefficients for the Poincaré group by relating all basis transformations to a standard preferred basis, the symmetrically coupled basis. With only a slight amount of complication it is possible to calculate the coefficients for particles with arbitrary spin and boost.[3]

Further, one can consider going from three to four and, in general, to N particles, corresponding to the $3J$, $6J$, and $9J$, $\cdots$ symbols.[1] Since it is possible to construct symmetrically coupled N-particle states, the entire procedure outlined in Sec. II quite easily generalizes to N particles. For example for $N=4$ one could couple 1 to 2, then 1 – 2 to 3, and finally 1 – 2 – 3 to 4; the Racah coefficient relating such a stepwise coupled state to the symmetrically coupled 4-particle state involves two Wigner $\mathcal{D}$ functions, one with an argument evaluated in the 1 – 2 CM frame, and the other in the 1 – 2 – 3 CM frame. Again the argument of these $\mathcal{D}$ functions can be shown to involve relativistically invariant subenergies only. And, in general, for N particles, $N-1$ $\mathcal{D}$ functions are involved, each $\mathcal{D}$ function labeling the intermediate angular momentum of the previously coupled particles. The only complication which arises in this procedure is that the set of subenergies formed from momenta such as $(p_1+p_4)^2$, $(p_2+p_3+p_5)^2$, etc. no longer fixes the momenta of all the particles in the body-fixed frame; it is also necessary to introduce invariants such as $\epsilon_{\alpha\beta\gamma\delta} p^\alpha p_1^\beta p_2^\gamma p_i^\delta$. But even this complication is handled by the group theory. It can be shown that the manifold corresponding to an N-particle phase space is given as the double coset manifold

$$\underbrace{SU(2)\otimes SU(2)\otimes\cdots\otimes SU(2)}_{N \text{ times}}/$$
$$\underbrace{SL(2,C)\otimes SL(2,C)\otimes\cdots\otimes SL(2,C)}_{N \text{ times}}/SL(2,C).$$

Written in this form one sees that any choice of relativistic invariants formed from four momenta provides a set of coordinates for the double coset space. This topic will be discussed in more detail in a subsequent paper dealing with particles having intrinsic spin and the relation between Racah coefficients and spin crossing matrices.

The general conclusion to be drawn is that dealing with symmetrically coupled N-particle states is no more complicated than dealing with three-particle states; this result arises solely from the induced representation structure of the Poincaré group and is to be contrasted with the rotation group and the complexities arising from $9J$, $12J$, etc., symbols.

[1] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957); A. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N.J. 1957).
[2] A. McKerrel, Nuovo Cimento **34**, 1289 (1964); G.C. Wick, Ann. Phys. (N.Y.) **18**, 65 (1963).
[3] W.H. Klink, Phys. Rev. D **4**, 2260 (1971).
[4] W.H. Klink, Nucl. Phys. B **77**, 56 (1974).
[5] W.H. Klink, "Crossing Multiparticle Amplitudes, I. Pole Singularities," University of Iowa Preprint 74-38.
[6] J. Werle, *Relativistic Theory of Reactions* (Wiley, New York, 1966).
[7] G. Rideau, Ann. Inst. Henri Poincaré 3, 339 (1965).
[8] R. L. Omnes, Phys. Rev. **134**, B1358 (1964).
[9] P.D.B. Collins and E.J. Squires, *Springer Tracts in Modern Physics* (Springer-Verlag, Berlin, 1968), Vol. 45.

General chiral $SU_2 \times SU_2$ pion Lagrangian and the generators of $O(4,1)$ group

Abdul Ebrahim

Physics Department, Middle East Technical University, Ankara, Turkey
(Received 26 December 1974)

It is shown that the general chiral $SU_2 \times SU_2$ invariant pion Lagrangian in the form given by Gürsey can be used to obtain a general and convenient parametrization of an infinite set of generators of the noncompact $O(4,1)$ group. Various special cases of the general form of the generators are given, and one particular form is shown to coincide with the generators of the $O(4,1)$ group used in the literature.

I. INTRODUCTION

We consider here the problem of representing the generators of the noncompact $O(4,1)$ group using a pair of canonically conjugate operators. For the purpose of parametrizing the generators in a convenient way, we use the general form of the chiral $SU_2 \times SU_2$ invariant pion Lagrangian given by Gürsey.[1] The pair of canonically conjugate operators then becomes the isovector pion field φ and its canonically conjugate momentum Π defined through the chiral invariant pion Lagrangian. Since the most general chiral $SU_2 \times SU_2$ invariant pion Lagrangian contains an arbitrary function of the pion field, we thus define an infinite set of operators Π that are canonically conjugate to φ. Using this infinite set of canonically conjugate operators, we construct an infinite set of generators of $O(4,1)$ group. For the $O(4) \sim SU_2 \times SU_2$ subgroup of the $O(4,1)$ group, we take the usual vector and axial vector operators of the chiral $SU_2 \times SU_2$ group. To these six operators, we adjoin a set of four operators that form a chiral 4-vector and thus transform as the $(\frac{1}{2}, \frac{1}{2})$ representation of the chiral $SU_2 \times SU_2$ group. The resulting set of ten operators is shown to generate the noncompact $O(4,1)$ group. Several special cases of the general form of the generators are considered, and one particular form is found to coincide with the generators of $O(4,1)$ group used by Gürsey.[2]

II. CHIRAL $SU_2 \times SU_2$ PION LAGRANGIAN

In this section we define the chiral $SU_2 \times SU_2$ invariant pion Lagrangian as given by Gürsey[1] and cast it in a form that will fascilitate the discussion of this paper.

In the method initiated by Gürsey, the $SU_2 \times SU_2$ invariant pion Lagrangian is expressed in terms of a 2×2 unitary, unimodular pion matrix U, which is a nonlinear function of the dimensionless operator if $\vec{\tau} \cdot \vec{\varphi}$, where f is a real coupling parameter with the dimension of length, and τ are the usual 2×2 Pauli matrices. A general and convenient parametrization of the pion matrix is given by

$$U(if\boldsymbol{\tau} \cdot \boldsymbol{\varphi}) = \sigma(f^2\varphi^2) + 2if\boldsymbol{\tau} \cdot \boldsymbol{\varphi}\, \rho(f^2\varphi^2), \tag{1}$$

where σ and ρ are real, scalar, isoscalar functions of the dimensionless operator $f^2\varphi^2$. Without any loss of generality, we choose σ and ρ such that

$$\sigma(0) = \rho(0) = 1. \tag{2}$$

The requirement that U be unitary implies that

$$\sigma^2 + 4f^2\rho^2\varphi^2 = 1, \tag{3}$$

and hence, from Eqs. (2) and (3), we get

$$\sigma' = -2 + O(f^2), \tag{4}$$

where the prime denotes a derivative with respect to $f^2\varphi^2$. Thus, in general, σ and ρ are given by

$$\sigma(f^2\varphi^2) = 1 - 2f^2\varphi^2 + O(f^4), \tag{5a}$$

$$\rho(f^2\varphi^2) = 1 + O(f^2). \tag{5b}$$

The matrix U is assumed to transform linearly under the chiral $SU_2 \times SU_2$ group as the $(\frac{1}{2}, \frac{1}{2})$ representation, so that its transformation law is

$$U \to \exp[i\boldsymbol{\tau} \cdot (\omega - \nu)/2]\, U \exp[-i\boldsymbol{\tau} \cdot (\omega + \nu)/2], \tag{6}$$

where ω and ν are the space–time independent isospin and chiral parameters, respectively.

In terms of the pion matrix U, the chiral $SU_2 \times SU_2$ invariant pion Lagrangian can be written as

$$\mathcal{L}_\pi = -(1/16f^2)\,\mathrm{Tr}[\partial_\mu U \partial_\mu U^\dagger], \tag{7}$$

where the normalization has been chosen such that the expansion of $\mathcal{L}_\pi$ in powers of f gives the free pion kinetic energy term in the lowest order. By using the explicit form of U given in Eq. (1), the pion Lagrangian (7) can also be written as

$$\mathcal{L}_\pi = -\tfrac{1}{2}[\partial_\mu(\rho\boldsymbol{\varphi}) \cdot \partial_\mu(\rho\boldsymbol{\varphi}) + (1/4f^2)\,\partial_\mu\sigma\,\partial_\mu\sigma]. \tag{8}$$

We note that the unitarity condition (3) gives a relation between σ and ρ, and hence only one of these pion functions is independent. Until a specific choice is made for this independent pion function, the $SU_2 \times SU_2$ invariant pion Lagrangian given above is quite general.

We now proceed to cast the chiral invariant pion Lagrangian (7) or (8) in a form that utilizes a metric in three-dimensional curved isospin space. To this end, we write the Lagrangian given in Eq. (7) as

$$\mathcal{L}_\pi = -(1/16f^2)\,\mathrm{Tr}[\partial_\mu U \partial_\mu U^\dagger] = -\tfrac{1}{2}G^{ij}\partial_\mu\varphi^i\partial_\mu\varphi^j, \tag{9}$$

where $i, j = 1, 2, 3$, and

$$G^{ij} = \frac{1}{8f^2}\,\mathrm{Tr}\left(\frac{\partial U}{\partial\varphi^i}\,\frac{\partial U^\dagger}{\partial\varphi^j}\right). \tag{10a}$$

By using the explicit form for U, computing the trace, and using several identities that follow from the unitarity relation (3), the expression for the metric G^{ij} can be simplified to

$$G^{ij}=\rho^2\left[\delta^{ij}+f^2\varphi^i\varphi^j\left(\frac{\sigma'^2-4\rho^4}{\rho^2(1-\sigma^2)}\right)\right]. \tag{10b}$$

We note that the metric form of the $SU_2\times SU_2$ pion Lagrangian given here provides a generalization of the pion Lagrangian written previously in this form by Meetz[3] who considered the special case when $\rho=1$.

Using the pion Lagrangian (9), we define the four-momentum $\Pi_\mu{}^i$ canonically conjugate to φ^i as

$$\Pi_\mu{}^i=-\frac{\delta\mathcal{L}_\pi}{\delta(\partial_\mu\varphi^i)}=G^{ij}\partial_\mu\varphi^j, \tag{11}$$

and, inverting this equation, we can express the ordinary derivative of the pion field in terms of its canonically conjugate momentum as

$$\partial_\mu\varphi^i=(G^{-1})^{ij}\Pi_\mu{}^j, \tag{12}$$

where the inverse metric $(G^{-1})^{ij}$ is given by

$$(G^{-1})^{ij}=\frac{1}{\rho^2}\left[\delta^{ij}-f^2\varphi^i\varphi^j\left(\frac{4\rho^2(\sigma'^2-4\rho^4)}{\sigma'^2(1-\sigma^2)}\right)\right]. \tag{13}$$

In concluding this section, we observe that the method used here to obtain the explicit form of the metric for a chiral invariant meson Lagrangian can be extended to the case of chiral $SU_3\times SU_3$ symmetry.

III. GENERATORS OF $O(4,1)$ GROUP

In this section we construct the generators of $O(4,1)$ group in their general form in terms of the pion field operator φ^i, its canonically conjugate momentum $\Pi^i\equiv\Pi_0{}^i$, and isoscalar functions of φ^i and Π^i. Our basic assumption is that the operators φ^i and Π^i satisfy the equal-time canonical commutation relation

$$[\varphi^i(x),\ \Pi^j(y)]\delta(x_0-y_0)=i\delta^{ij}\delta(\mathbf{x}-\mathbf{y}). \tag{14}$$

The generators of the $O(4)\sim SU_2\times SU_2$ subgroup of the $O(4,1)$ group are taken to be the time components of the usual vector and the axial vector currents of the chiral $SU_2\times SU_2$ group defined according to

$$\mathbf{J}_\mu=(1/8if^2)\,\mathrm{Tr}[(U^\dagger\partial_\mu U+U\partial_\mu U^\dagger)\boldsymbol{\tau}/2]=\rho^2\boldsymbol{\varphi}\times\partial_\mu\boldsymbol{\varphi}, \tag{15}$$

$$\mathbf{J}_{5\mu}=(1/8if^2)\,\mathrm{Tr}[(U^\dagger\partial_\mu U-U\partial_\mu U^\dagger)\boldsymbol{\tau}/2]$$
$$=(1/2f)[\sigma\partial_\mu(\rho\boldsymbol{\varphi})-(\rho\boldsymbol{\varphi})\partial_\mu\sigma]. \tag{16}$$

To these six operators, we adjoin a set of four current operators that transform as a 4-vector under the chiral $SU_2\times SU_2$ group, and are defined as

$$\mathbf{F}_{5\mu}=(1/2f)\,\partial_\mu(\rho\boldsymbol{\varphi}), \tag{17}$$

$$S_\mu=(1/4f^2)\partial_\mu\sigma. \tag{18}$$

Then using the relation between $\partial_0\varphi^i$ and $\Pi_0{}^i$ given by Eq. (12), the time components of the operators $\mathbf{J}_\mu$, $\mathbf{J}_{5\mu}$, $\mathbf{F}_{5\mu}$, and S_μ can be expressed in terms of φ^i and Π^i according to

$$\mathbf{J}_0=\boldsymbol{\varphi}\times\boldsymbol{\Pi}, \tag{19}$$

$$\mathbf{J}_{50}=\frac{1}{2f}\left[\left(\frac{\sigma}{\rho}\right)\boldsymbol{\Pi}+4f^2\left(\frac{\rho'}{\sigma'}\right)\boldsymbol{\varphi}\boldsymbol{\varphi}\cdot\boldsymbol{\Pi}\right], \tag{20}$$

$$\mathbf{F}_{50}=\frac{1}{2f}\left[\left(\frac{1}{\rho}\right)\boldsymbol{\Pi}+\frac{4f^2}{\sigma'}(\sigma\rho'-\rho\sigma')\boldsymbol{\varphi}\boldsymbol{\varphi}\cdot\boldsymbol{\Pi}\right], \tag{21}$$

$$S_0=\frac{2\rho^2}{\sigma'}\,\boldsymbol{\varphi}\cdot\boldsymbol{\Pi}. \tag{22}$$

If we now define the charges corresponding to these ten operators as

$$Q^i=\int J_0{}^i(x)\,dx^3,\quad Q_5^i=\int J_{50}{}^i(x)\,d^3x, \tag{23}$$

$$R_5{}^i=\int F_{50}{}^i(x)\,d^3x,\quad S=\int S_0(x)\,d^3x, \tag{24}$$

then, using the canonical commuation relation (14), we find that the ten operators Q^i, $Q_5{}^i$, $R_5{}^i$, and S obey the commutation relations

$$[Q^i,Q^j]=[Q_5^i,Q_5{}^j]=i\epsilon^{ijk}Q^k, \tag{25}$$

$$[Q^i,Q_5{}^j]=i\epsilon^{ijk}Q_5{}^k, \tag{26}$$

$$[Q^i,R_5{}^j]=i\epsilon^{ijk}R_5{}^k,\quad [Q^i,S]=0, \tag{27}$$

$$[Q_5{}^i,R_5{}^j]=i\delta^{ij}S,\quad [Q_5{}^i,S]=-iR_5{}^i, \tag{28}$$

$$[R_5{}^i,S]=-iQ_5{}^i,\quad [R_5{}^i,R_5{}^j]=-i\epsilon^{ijk}Q_5{}^k, \tag{29}$$

which are recognized as the commutation relations of the noncompact $O(4,1)$ group.

We note that the commutation relations (25)—(29) hold in general for arbitrary choices of σ and ρ consistent with the unitarity constraint (3). Any particular choice of σ and ρ will give a special parametrization of the generators of $O(4,1)$ group. We consider some special cases in the next section.

IV. SPECIAL MODELS

In this section we give several special forms of the generators of the $O(4,1)$ group. These special forms result from specific choices for the pion functions σ and ρ. The cases we consider correspond to various models of the chiral $SU_2\times SU_2$ invariant pion—pion interaction that have been studied by various authors.

A. Square-root model[4]

In this model,

$$\sigma=(1-4f^2\varphi^2)^{1/2},\quad \rho=1, \tag{30}$$

and hence

$$\mathbf{J}_0=\boldsymbol{\varphi}\times\boldsymbol{\Pi},\quad \mathbf{J}_{50}=(1/2f)\,(1-4f^2\varphi^2)^{1/2}\,\boldsymbol{\Pi}, \tag{31}$$

$$\mathbf{F}_{50}=(1/2f)\,(\boldsymbol{\Pi}-4f^2\boldsymbol{\varphi}\,\boldsymbol{\varphi}\cdot\boldsymbol{\Pi}),$$

$$S_0=-(1-4f^2\varphi^2)^{1/2}\,\boldsymbol{\varphi}\cdot\boldsymbol{\Pi}. \tag{32}$$

B. Inverse square-root model[5]

For this case,

$$\sigma=\rho=(1+4f^2\varphi^2)^{-1/2}, \tag{33}$$

and

$$\mathbf{J}_0=\boldsymbol{\varphi}\times\boldsymbol{\Pi},\quad \mathbf{J}_{50}=\frac{1}{2f}\,(\boldsymbol{\Pi}+4f^2\boldsymbol{\varphi}\,\boldsymbol{\varphi}\cdot\boldsymbol{\Pi}), \tag{34}$$

$$\mathbf{F}_{50} = (1/2f)\,(1 + 4f^2\varphi^2)^{1/2}\,\Pi,$$
$$S_0 = -\,(1 + 4f^2\varphi^2)^{1/2}\varphi\cdot\Pi. \tag{35}$$

C. Exponential model[1]

This model is defined by

$$\sigma = \cos(4f^2\varphi^2)^{1/2},$$
$$\rho = (4f^2\varphi^2)^{-1/2}\sin(4f^2\varphi^2)^{1/2}, \tag{36}$$

and hence the generators are given by

$$\mathbf{J}_0 = \varphi \times \Pi, \tag{37}$$
$$\mathbf{J}_{50} = (1/2f)\,\{(4f^2\varphi^2)^{1/2}\cot(4f^2\varphi^2)^{1/2}\,\Pi - 4f^2[(4f^2\varphi^2)^{-1/2}\cot(4f^2\varphi^2)^{1/2} - (4f^2\varphi^2)^{-1}]\,\varphi\,\varphi\cdot\Pi\}, \tag{38}$$
$$\mathbf{F}_{50} = (1/2f)\,\{(4f^2\varphi^2)^{1/2}\csc(4f^2\varphi^2)^{1/2}\,\Pi - 4f^2[(4f^2\varphi^2)^{-1/2}\csc(4f^2\varphi^2)^{1/2} - (4f^2\varphi^2)^{-1}\cos(4f^2\varphi^2)^{1/2}]\,\varphi\,\varphi\cdot\Pi\}, \tag{39}$$
$$S_0 = -\,(4f^2\varphi^2)^{-1/2}\sin(4f^2\varphi^2)^{1/2}\;\varphi\cdot\Pi. \tag{40}$$

D. Linear model[6]

For this model,

$$\sigma = 1 - 2f^2\varphi^2, \quad \rho = (1 - f^2\varphi^2)^{1/2}, \tag{41}$$

and, hence,

$$\mathbf{J}_0 = \varphi \times \Pi, \tag{42}$$
$$\mathbf{J}_{50} = \frac{1}{2f}\,[(1 - f^2\varphi^2)^{-1/2}(1 - 2f^2\varphi^2)\,\Pi + f^2(1 - f^2\varphi^2)^{-1/2} \times \varphi\,\varphi\cdot\Pi\,, \tag{43}$$
$$\mathbf{F}_{50} = \frac{1}{2f}\,[(1 - f^2\varphi^2)^{-1/2}\,\Pi + f^2(1 - f^2\varphi^2)^{-1/2}(2f^2\varphi^2 - 3)\varphi\,\varphi\cdot\Pi] \tag{44}$$
$$S_0 = -\,(1 - f^2\varphi^2)\,\varphi\cdot\Pi. \tag{45}$$

E. Rational model[7]

Here

$$\sigma = (1 - f^2\varphi^2)(1 + f^2\varphi^2)^{-1}, \quad \rho = (1 + f^2\varphi^2)^{-1}, \tag{46}$$

and hence the generators are

$$\mathbf{J}_0 = \varphi \times \Pi, \tag{47}$$
$$\mathbf{J}_{50} = (1/2f)\,[(1 - f^2\varphi^2)\,\Pi + 2f^2\varphi\,\varphi\cdot\Pi], \tag{48}$$
$$\mathbf{F}_{50} = (1/2f)[(1 + f^2\varphi^2)\Pi - 2f^2\varphi\,\varphi\cdot\Pi], \tag{49}$$
$$S_0 = -\,\varphi\cdot\Pi. \tag{50}$$

We note that this particular form of the generators of the $O(4, 1)$ group has been used previously by Gürsey[2] and by Barut.[8] This form has also been found useful for the description of chiral $SU_2 \times SU_2$ representation mixing for the pseudoscalar mesons.[9]

In conclusion, we remark that the method used here for constructing the general form of the generators of the $O(4, 1)$ group can be extended to the case of the conformal group $O(4, 2)$.

ACKNOWLEDGMENTS

I am indebted to Professor Feza Gürsey for guidance and valuable suggestions.

[1] F. Gürsey, Nuovo Cimento **16**, 230 (1960); Ann. Phys. (N.Y.) **12**, 91 (1961); Acta Phys. Austr. **5**, 186 (1968).
[2] F. Gürsey, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964), p. 365.
[3] K. Meetz, J. Math Phys. 589 (1969). See also C.J. Isham, Nuovo Cimento A **61**, 188 (1969).
[4] M. Gell-Mann and M. Levy, Nuovo Cimento **16**, 705 (1960); S. Weinberg, Phys. Rev. Lett. **18**, 188 (1967); L.S. Brown, Phys. Rev. **163**, 1802 (1967).
[5] I. Bars, Phys. Rev. D **2**, 1630 (1970).
[6] F. Gürsey, Lectures at Yale University (1969), unpublished.
[7] G. Kramer, H. Rollnick, and B. Steck, Z. Phys. **154**, 564 (1959); J. Schwinger, Phys. Lett. B **24**, 473 (1967); S. Weinberg, Phys. Rev. **166**, 1568 (1968); J. Wess and B. Zumino, Phys. Rev. **163**, 1727 (1967).
[8] A.O. Barut and G.L. Bornzin, Lett. Nuovo Cimento **6**, 177 (1973).
[9] A. Ebrahim and F. Gürsey, Lett. Nuovo Cimento **9**, 9, 716E (1974).

Discrete Coulomb gas in one dimension: Correlation functions

M. L. Mehta

Service de Physique Theorique, Centre d'Etudes Nucleaires de Saclay, B.P. n°2-91190, Gif-sur-Yvette, France

G. C. Mehta*

State University of New York at Stony Brook, L.I., New York 11790
(Received 25 April 1974)

In this short note we calculate the correlation and cluster functions of the discrete Coulomb gas in one dimension considered recently by Gaudin. We consider also the discrete versions of the Gaussian ensembles previously studied extensively.

Recently Gaudin[1] considered a discrete version of the circular ensembles of Dyson.[2] The positions which a unit charge can occupy on the circumference of the unit circle are restricted to N equidistant points $\exp(i\theta_j)$ $\theta_j = 2\pi j/N$, $1 \leq j \leq N$. One considers only three values of the inverse temperature β, $\beta = 1, 2$, and 4. The joint probability density for n unit charges to occupy positions $j_1, \ldots, j_n$ is taken as

$$P_\beta(j_1, \ldots, j_n) = C_{Nn\beta}^{-1} \frac{1}{N^n} \exp(-\beta W), \tag{1}$$

where

$$W \equiv W(j_1, \ldots, j_n) = -\sum_{1 \leq l < m \leq n} \ln \left| e^{i\theta_l} - e^{i\theta_m} \right|, \tag{2}$$

$$\theta_l = 2\pi j_l / N, \tag{3}$$

and $C_{Nn\beta}$ is the normalisation constant

$$C_{Nn1} = n! N^{-[n/2]} \prod_{j=1}^{[n/2]} \cot(n - 2j + 1)\frac{\pi}{2N}, \tag{4}$$

$$C_{Nn2} = n! \tag{5}$$

$$C_{Nn4} = \frac{(2\nu)!}{2^\nu \nu!} N^{n-\nu} n!, \tag{6}$$

where $[x]$ denotes the integral part of x and

$$\nu = \tfrac{1}{2}N - \left|\tfrac{1}{2}N - n\right|. \tag{7}$$

The values of $C_{Nn\beta}$ given above are due to Gaudin. He proceeds then to calculate other thermodynamic quantities begining with the free energy.

We want to say that one can calculate the correlation and cluster functions for this discrete system. The simplest case is $\beta = 2$ because the orthogonality property

$$\frac{1}{2\pi} \int_0^{2\pi} e^{im\theta} d\theta = \delta_{m0}, \quad m \text{ an integer}, \tag{8}$$

persists when we replace the integration by a sum over N discrete points

$$\frac{1}{N} \sum_{j=1}^{N} e^{im(2\pi/N)j} = \sum_{l=-\infty}^{\infty} \delta_{m, lN}, \quad m \text{ and } l \text{ integers}. \tag{9}$$

If p and q take the values $l - n - \frac{1}{2}$, $l = 1, 2, \ldots, 2n$, then

$$\frac{1}{N} \sum_{j=1}^{N} e^{i(p-q)(2\pi/N)j} = \delta_{pq} \tag{10}$$

provided $2n \leq N$. Hence for $\beta = 4$, $2n \leq N$, one can calculate all correlation and cluster functions by Dyson's method[3] of quaternion determinants. If $2n > N$ the right-hand side of Eq. (10) may contain additional terms $\delta_{p, q \pm N}$. Also if p takes the values $l - \frac{1}{2}(n+1)$, $l = 1, 2, \ldots, n$ and q the values $\pm(l + \frac{1}{2}(n-1))$, $l = 1, 2, \ldots, \infty$, then the sum in Eq. (10) is not longer zero. Thus the method of quaternion determinants fails for $\beta = 4$, $2n > N$ or for the case $\beta = 1$. However, in this later case one may still calculate the two point correlation function by the method of summation over alternate variables.[4]

The results are as follows. With

$$S_n(\theta) = \frac{1}{N} \sum_p e^{ip\theta} = \frac{1}{N} \frac{\sin(\frac{1}{2}n\theta)}{\sin(\frac{1}{2}\theta)}, \tag{11}$$

$$D_{2n}(\theta) = \frac{1}{N} \sum_q iq e^{iq\theta}, \tag{12}$$

$$I_{2n}(\theta) = \frac{1}{N} \sum_q (iq)^{-1} e^{iq\theta}, \tag{13}$$

where p and q take, respectively, the values

$$p = -\frac{n-1}{2}, -\frac{n-3}{2}, \ldots, \frac{n-1}{2}, \tag{14}$$

$$q = \pm 1/2, \pm 3/2, \ldots, \pm(n - (1/2)), \tag{15}$$

we define

$$\sigma_2(\theta) = S_n(\theta), \tag{16}$$

and

$$\sigma_4(\theta) = \frac{1}{2} \begin{bmatrix} S_{2n}(\theta) & D_{2n}(\theta) \\ I_{2n}(\theta) & S_{2n}(\theta) \end{bmatrix}. \tag{17}$$

The m point correlation function defined by

$$R_{m\beta}(j_1, \ldots, j_m) = \frac{n!}{(n-m)!} \sum_{j_{m+1}=1}^{N} \cdots \sum_{j_n=1}^{N} P_\beta(j_1, \ldots, j_n) \tag{18}$$

is given for $\beta = 2$ and $\beta = 4$ by

$$R_{m\beta}(j_1,\ldots,j_m)=\{\det[\sigma_\beta(\theta_k-\theta_l)]_{k,l=j_1,\ldots,j_m}\}^\lambda, \tag{19}$$

where

$$\lambda=1 \text{ for } \beta=2 \text{ and } \lambda=1/2 \text{ for } \beta=4.$$

Moreover, if $\beta=4$, Eq. (19) is valid only for $N\geq 2n$. The determinant in the above equation is of order $m\times m$ if $\beta=2$ and of order $2m\times 2m$ if $\beta=4$. The m point cluster function defined[4] by

$$T_{m\beta}(j_1,\ldots,j_m)=\sum_G (-1)^{m-l}(l-1)!\prod_{i=1}^{l} R_{h_i\beta}(\theta_k, \text{ with } k \text{ in } G_i), \tag{20}$$

where G denotes any division of the indices $j_1,j_2,\ldots,j_m$ into unordered subsets $G_1,\ldots,G_l$ and h_i is the number of indices in G_i, is given for $\beta=2$ and $\beta=4$ by

$$T_{m\beta}(j_1,\ldots,j_m)=\lambda\ \mathrm{Tr}\sum_P\{\sigma_\beta(\theta_{j_1}-\theta_{j_2})\sigma_\beta(\theta_{j_2}-\theta_{j_3})\ldots\sigma_\beta(\theta_{j_m}-\theta_{j_1})\} \tag{21}$$

with $\lambda=1$ for $\beta=2$ and $\lambda=1/2$ for $\beta=4$ and where $\sum_P$ denotes a sum over the $(m-1)!$ distinct cyclic permutations of the indices $1,2,\ldots,m$. In case $\beta=4$, we again suppose $N\geq 2n$.

Except that $(2\pi)^{-1}\sum_p$ is replaced by $N^{-1}\sum_p$ in the definitions of S_n, D_n, and I_n above, the correlation and cluster functions are exactly the same as in the continuous case of Dyson's circular ensembles; therefore, they coincide when $N\to\infty$, as they should.

Also one may consider discrete versions of the Gaussian ensembles and in general ensembles related to other classical orthogonal polynomials.[5] For this the positions which a unit charge can occupy is restricted to the set of zeros x_j, $1\leq j\leq N$, of $H_N(x)$, where $H_N(x)$ is the Nth Hermite polynomial for Gaussian ensembles or it is the Nth degree polynomial of the orthogonal set in general. The joint probability density for the n unit charges to be at positions $x_1,\ldots,x_n$ is taken as

$$P_\beta(x_1,\ldots,x_n)=G_{n\beta}^{-1}\exp(-\beta W), \tag{22}$$

where

$$W\equiv W(x_1,\ldots,x_n)=-\frac{1}{\beta}\sum_{i=1}^{n} w_i-\sum_{1\leq i<j\leq n}\ln|x_i-x_j|, \tag{23}$$

w_i are the weights of the Gaussian quadrature formula[6]

$$w_i=\left[\sum_{j=0}^{N-1}\{H_j(x_i)\}^2\right]^{-1} \tag{24}$$

related to the set of orthogonal polynomials used and the normalization constant $G_{n\beta}$ does not depend on N; for the Gaussian case,

$$G_{n\beta}=\pi^{n/2}2^{-n(n-1)/4}\beta^{n(n-1)(1-\beta)/4}\{\Gamma(1+\tfrac{1}{2}\beta)\}^{-n}\cdot\prod_{j=1}^{n}\Gamma(1+\tfrac{1}{2}\beta j). \tag{25}$$

The presence of β in Eq. (23) above is very uncomfortable, but we do not know how to get rid of it.

We give below the results only for the Gaussian case. Because of the formula

$$\sum_{i=1}^{N} w_i f(x_i)=\int_{-\infty}^{\infty}\exp(-x^2)f(x)\,dx, \tag{26}$$

where $f(x)$ is a polynomial of degree $<2N$, the working is almost the same as in Ref. 7 with minor changes and we will not repeat it here.

Denoting the normalized harmonic oscillator wavefunctions by

$$\varphi_j(x)=(\pi^{1/2}2^j j!)^{-1/2}\exp(x^2/2)\left(-\frac{d}{dx}\right)^j\exp(-x^2), \tag{27}$$

we set

$$f_{ij}=\tfrac{1}{2}\int_{-\infty\leq y\leq x\leq\infty}dx\,dy\,\{\varphi_i(x)\varphi_j(y)-\varphi_j(x)\varphi_i(y)\}, \tag{28}$$

$$\epsilon(x,y)=\sum_{i,j=0}^{N-1} f_{ij}\,\varphi_i(x)\varphi_j(y), \tag{29}$$

$$S_n(x,y)=\sum_{i=0}^{n-1}\varphi_i(x)\varphi_i(y)+(\tfrac{1}{2}n)^{1/2}\varphi_{n-1}(x)\int_{-\infty}^{\infty}\epsilon(y,t)\varphi_n(t)\,dt, \tag{30}$$

$$I_n(x,y)=\int_{-\infty}^{\infty}\epsilon(x,t)S_n(t,y)\,dt, \tag{31}$$

$$D_n(x,y)=\sum_{i=0}^{n-2}\left(\frac{i+1}{2}\right)^{1/2}\{\varphi_i(x)\varphi_{i+1}(y)-\varphi_{i+1}(x)\varphi_i(y)\}, \tag{32}$$

$$J_n(x,y)=I_n(x,y)-\epsilon(x,y), \tag{33}$$

$$\alpha(x)=\begin{cases}0 & \text{if } n \text{ is even},\\ \varphi_{n-1}(x)/\int_{-\infty}^{\infty}\varphi_{n-1}(t)\,dt & \text{if } n \text{ is odd},\end{cases} \tag{34}$$

$$\sigma_1(x,y)=\begin{bmatrix}S_n(x,y)+\alpha(x) & D_n(x,y)\\ J_n(x,y) & S_n(y,x)+\alpha(y)\end{bmatrix}, \tag{35}$$

$$\sigma_2(x,y)=\sum_{i=0}^{n-1}\varphi_i(x)\varphi_i(y), \tag{36}$$

and

$$\sigma_4(x,y)=\frac{1}{2}\begin{bmatrix}S_{2n+1}(x,y) & D_{2n+1}(x,y)\\ I_{2n+1}(x,y) & S_{2n+1}(y,x)\end{bmatrix}. \tag{37}$$

Then one has

$$R_{m\beta}\equiv\frac{n!}{(n-m)!}\sum_{j_{m+1}=1}^{N}\cdots\sum_{j_n=1}^{N}P_\beta(x_{j_1},\ldots,x_{j_n})=\{\det[\sigma_\beta(x_i,x_k)]_{i,k=j_1,\ldots,j_m}\}^\lambda \tag{38}$$

where $\lambda=1/2$ for $\beta=1$ and $\beta=4$ and $\lambda=1$ for $\beta=2$. As for the continuous case, the determinant in Eq. (38) is of order $m\times m$ for $\beta=2$ and of order $2m\times 2m$ for $\beta=1$ and $\beta=4$.

It might seem strange[8] that the result (19) was not stated for the case $\beta=1$ and it was stated for $\beta=4$ only when $2n\leq N$, while the corresponding result (38) for the Gaussian model is valid without these restrictions. The reason why Dyson's method works in the later case under all circumstances is that the functions ϵ, S_n, I_n, D_n, and J_n defined by Eqs. (29)—(33) satisfy the relations

$$\epsilon * S_n=I_n,\quad \epsilon * D_n=D_n*\epsilon=S_n,\quad J_n*S_n=J_n*D_n=0, \tag{39}$$

$$S_n*S_n=D_n*I_n=I_n*D_n=S_n,\qquad S_n*I_n=I_n,\quad D_n*S_n=D_n, \tag{40}$$

with the notation

$$F * G \equiv \sum_{j=1}^{N} F(x, z_j)\, G(z_j, y), \tag{41}$$

z_j being the zeros of the Hermite polynomial $H_N(x)$. The method would work for the circular model as well if one could define functions satisfying the above relations (39)—(40), with the notation

$$F * G \equiv \sum_{j=1}^{N} F\left(\theta, \frac{2\pi}{N} j\right) G\left(\frac{2\pi}{N} j, \Phi\right). \tag{42}$$

However, because of the reasons stated just after Eq. (10), relations (40) are not valid for the case $\beta = 4$, $2n > N$, and we could not define an ϵ with the required properties.

One of us (M. L. M.) thanks Dr. J. des Cloizeaux for helpful discussions.

[1]M. Gaudin, J. Phys. **34**, 511 (1973).
[2]F. J. Dyson, J. Math. Phys. **3**, 140, 157, 166 (1962).
[3]F. J. Dyson, Commun. Math. Phys. **19**, 235 (1970).
[4]M. L. Mehta, *Random Matrices* (Academic, New York, 1967), Chap. 5.
[5]D. Fox and P. B. Kahn, Phys. Rev. B **134**, 1151 (1964); H. S. Leff, thesis, State University of Iowa, SUI-63-23 (1963); B. V. Bronk, J. Math. Phys. **5**, 1661 (1964); **6**, 228 (1965).
[6]See, for example, H. S. Wilf, *Mathematics for Physical Sciences* (Wiley, New York, 1962), Par. 2.9.
[7]M. L. Mehta, Commun. Math. Phys. **20**, 245 (1971).
[8]Referee's remarks.

Two-level radiative systems and perturbation theory

F. T. Hioe and E. W. Montroll

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627*

(Received 25 November 1974; revised manuscript received 6 February 1975)

We obtain expressions for the radiative level shifts of a two-level system in terms of (i) recurrence relations, (ii) ratios of determinants, (iii) continued fractions, and (iv) a Lidstone expansion. These expressions are shown to be very useful for numerical computations. It is pointed out that perturbation series in powers of the coupling constant are not the most appropriate way of representing the solutions of the problem, but if they are to be used, different series should be employed depending on the relative value of the frequency of the field to the frequency of the two-level system. The significance of these results in the general theory of perturbation is discussed.

1. INTRODUCTION

This paper contains a discussion of a simple model of the interaction of radiation characterized by a frequency ω with a two level atom whose energy levels are postulated to be $\pm\frac{1}{2}\omega_0$ (in units with $\hbar=1$). The Hamiltonian frequently chosen to represent the photon—atom interaction is

$$H=\omega_0 S^z+\omega a^\dagger a+\lambda[(a^\dagger S^-+aS^+)+(a^\dagger S^+ +aS^-)], \tag{1.1}$$

where $a^\dagger$ and a are the usual photon creation and annihilation operators and S^z, S^-, and S^+ are the commonly used spin operators. In particular S^z for a spin $\frac{1}{2}$ system, which is equivalent to our two level atom, has the characteristic values $\pm\frac{1}{2}$. The parameter λ is a measure of the strength of the photon—atom interaction. This model interaction of only one field mode with a two level atom is a simple model and does not represent adequately many features of the more realistic interaction involving an infinite number of modes. The many mode case is being examined and will be published in a later paper.

We will employ the Bargmann representation displayed in Eq. (2.2) of the next section to express H as a differential operator. In this representation the unperturbed wavefunctions have an especially simple form being merely a product of powers of three complex variables. By expanding the wavefunctions of the full Hamiltonian as a linear combination of the unperturbed wave functions, the coefficients of the various products of powers will be shown to satisfy certain recurrence formulas which can be investigated in detail through the introduction of appropriate generating functions. From these generating functions, it is found that the required energy levels of (1.1) are eigenvalues of a secular equation which is essentially composed as the sum of two continued fractions. In the rotating wave approximation in which the second interaction term in the square bracket of (1.1) is neglected, our generating functions immediately yield the well-known energy levels for the coupled system as first obtained by Jaynes and Cummings.

The continued fraction formulation of the secular equation for the energy levels of (1.1) is similar to one derived by Swain by a quite different type of analysis. The advantage of the continued fraction formulation is that a number of different kinds of expansions can be made, depending on the regime of interest of the basic parameters $(\lambda, \omega, \omega_0)$. When λ is small, power series in λ equivalent to those which would follow from standard perturbation theory are obtained. In the large quantum number limit they are the same as expansions derived by Shirley by semiclassical methods.

The various formulas which are equivalent to results of standard perturbation theory have resonance denominators proportional to $(\omega_0-\omega)$, $(\omega_0-3\omega)$, $(\omega_0-5\omega)\cdots$. Hence it would seem that the expansions diverge as field and atom modes become resonant. However, if one sets $\omega=\omega_0$ in the basic continued fractions before calculations are made, new series expansion in λ without resonance terms are obtained. There are also alternative expansions when $\omega_0=3\omega$, etc. The resonance difficulty can be traced to invalid and resolved by making valid series expansions. In its most primitive form the problem is clear if one remembers the two expansions for $(a+b)^{-1}$,

$$(a+b)^{-1}=\begin{cases} a^{-1}[1-(b/a)+\cdots] & \text{if } a>b \quad (1.2a)\\ b^{-1}[1-(a/b)+\cdots] & \text{if } b>a. \quad (1.2b)\end{cases}$$

In our basic continued fraction a term analogous to b is proportional to λ while one analogous to a is proportional to $(\omega-\omega_0)$. The standard perturbation series would correspond to (1.2a) while the appropriate expansion at or near resonance is (1.2b). The proper expansion of the continued fraction in this regime yields energy levels which are power series in $(\omega-\omega_0)$. Analogous series are also obtained near the resonance $3\omega=\omega_0$, etc.

We will also discuss a novel perturbation theory which takes advantage of the fact that energy levels are known exactly when the rotating wave approximation is made [i.e., when $(a^\dagger S^+ + aS^-)$ is omitted] as well as when the term $(a^\dagger S^- + aS^+)$ is omitted. Hence, if we write the interaction term in (1.1) as

$$2\lambda[\gamma(a^\dagger S^- + aS^+)+(1-\gamma)(a^\dagger S^+ + aS^-)], \tag{1.3}$$

we know the properties of the system exactly when $\gamma=1$ and when $\gamma=0$. We require the properties when $\gamma=\frac{1}{2}$. As a first approximation a simple interpolation between the two exact results would be chosen. Hence if we were concerned with a function $A(2\lambda,\gamma)$ which depended on the full Hamiltonian $H(2\lambda,\gamma)$, then

$$A(2\lambda,\gamma)\approx(1-\gamma)A(2\lambda,0)+\gamma A(2\lambda,1), \tag{1.4}$$

the $A(2\lambda,0)$ and $A(2\lambda,1)$ being known exactly from the

rotating and "counter rotating" wave approximation. Equation (1.4) is then systematically improved by making an expansion (the Lidstone expansion) of $A(2\lambda, \gamma)$ about the two points $\gamma = 0$ and $\gamma = 1$.

This paper is the first in what we hope will develop into a series on models of increasing complexity which involve Hamiltonians which can be expressed in terms of creation, annihilation, and spin operators. The style will be to use the Bargmann representation and to then find appropriate generating functions which will allow one to find a clear formulation of the secular equation which must be solved. A richer variety of series expansions, some of which do not involve divergences will then be possible, than is available from standard perturbation theory. Investigation of singularities which are inherent in problems might be made directly from the differential equations obtained by using the Bargmann representation.

2. THE HAMILTONIAN AND THE BASIC RECURRENCE RELATIONS

The problem of calculating the stimulated and spontaneous radiative frequency shifts in a two-level atom, and the mathematically equivalent problem of determining the so-called Bloch–Siegert shifts in a spin $\frac{1}{2}$ system in a magnetic field, are of considerable theoretical and experimental interest and have been the subject of many studies in recent years.[1] The problem involves, in principle, the determination of the eigenvalues of the following Hamiltonian (in units of $\hbar = 1$):

$$H = \omega_0 S^z + \omega a^\dagger a + \lambda[(a^\dagger S^- + aS^+) + (a^\dagger S^+ + aS^-)] \equiv H_0 + \lambda H_1 \tag{2.1}$$

where $a^\dagger$ and a are the photon creation and annihilation operators and the S's are the usual spin operators. We begin by casting the eigenvalue equations $H|E\rangle = E|E\rangle$ into the form of a differential equation using the Bargmann representation[2] for the field and spin operators

$$a^\dagger \to z, \quad a \to \frac{\partial}{\partial z}, \quad S^+ \to u\frac{\partial}{\partial v}, \quad S^- \to v\frac{\partial}{\partial u}, \quad S^z \to \frac{1}{2}\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right). \tag{2.2}$$

In terms of this representation, the eigenvalue equation becomes

$$\left\{\tfrac{1}{2}\omega_0\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right) + \omega z\frac{\partial}{\partial z} + \lambda\left[\left(zv\frac{\partial}{\partial u} + \frac{\partial}{\partial z}\,u\frac{\partial}{\partial v}\right) + \left(zu\frac{\partial}{\partial v} + \frac{\partial}{\partial z}\,v\frac{\partial}{\partial u}\right)\right]\right\} \times f(\lambda; u, v, z) = E(\lambda) f(\lambda; u, v, z). \tag{2.3}$$

The unperturbed eigenfunction $f^{S, m, n}(\lambda; u, v, z)$ which corresponds to the quantum numbers S, m (for the total and z-component of the spin), and n (for the number of photons) is given by

$$f^{S, m, n}(0; u, v, z) = u^m v^{2S-m} z^n. \tag{2.4}$$

The associated unperturbed eigenvalue is given by

$$E^{S, m, n}(0) = (-S + m)\omega_0 + n\omega. \tag{2.5}$$

For the spin $\frac{1}{2}$ system, we write the unperturbed eigenfunctions and eigenvalues as $u^{(1+\sigma)/2} v^{(1-\sigma)/2} z^n$ and $\frac{1}{2}\sigma\omega_0 + n\omega$, where $\sigma = -1$ (lower level) and 1 (upper level) and $n = 0, 1, 2, \cdots$. In the presence of interaction, let the energy and eigenfunction of Eq. (2.3) corresponding to the quantum numbers σ and n be expressed by

$$E^{\sigma, n}(\lambda) = \tfrac{1}{2}\sigma\omega_0 + n\omega + \sum_{p=1}^{\infty} A_p^{\sigma, n}\lambda^p \tag{2.6}$$

and

$$f^{\sigma, n}(\lambda; u, v, z) = u^{(1+\sigma)/2} v^{(1-\sigma)/2} z^n + u^{(1+\sigma)/2} v^{(1-\sigma)/2} z^n \sum_{p=1}^{\infty} B_p^{\sigma, n}(u, v, z)\lambda^p. \tag{2.7}$$

We now make a crucial step by letting[3]

$$B_p^{\sigma, n}(u, v, z) = \sum_{k=-p}^{p} \sum_{i=-p}^{p} b_{p; i, k}^{\sigma, n} u^i v^{-i} z^k \tag{2.8}$$

with $b_{p; 0, 0}^{\sigma, n} = \delta_{p, 0}$. Then substitution of Eqs. (2.6) and (2.7) into the eigenvalue equation (2.3) and comparison of the coefficients of like powers of u, v, and z lead us, as shown in Appendix A, to the following recurrence relation involving A's and b's. This enables us to determine these quantities recurrsively in a *consistent* manner (omitting the superscripts σ and n on the A's and b's for convenience):

$$\begin{aligned}(i\omega_0 + k\omega) b_{p; i, k} &+ \{[\tfrac{1}{2}(1+\sigma) + i + 1] b_{p-1; i+1, k-1} \\ &+ [-\tfrac{1}{2}(1+\sigma) - i + 2](n + k + 1) b_{p-1; i-1, k+1}\} \\ &+ \{[-\tfrac{1}{2}(1+\sigma) - i + 2] b_{p-1; i-1, k-1} \\ &+ [\tfrac{1}{2}(1+\sigma) + i + 1](n + k + 1) b_{p-1; i+1, k+1}\} \\ &= \sum_{q=0}^{p-1} A_{p-q} b_{q; i, k}\end{aligned} \tag{2.9}$$

with $b_{p; 0, 0} = \delta_{p, 0}$ and $b_{p; i, k} = 0$ if $|i|$ or $|k|$ or both is $> p$ [see Eq. (2.8)]. The step given by Eq. (2.8) is crucial because if the lower or upper limits of the summations were chosen differently, the resulting recurrence relation would lead to either an inconsistency or redundancy.[3] We note that the eigenfunction in (2.7) should be expressible in the form

$$\sum_{k=0}^{\infty} (c_k u + d_k v) z^k, \tag{2.10}$$

where c_k and d_k are some constants, but this appears to be inconsistent with the form obtained by substituting (2.8) into (2.7). Actually, however, the recurrence relation (2.9) automatically gives, as we show in Appendix B, $b_{p, i, k}^{\pm, n} = 0$ for all $k < n$ and $b_{p; i, k}^{+, n} = 0$ unless $i = 0$ or -1 and $b_{p; i, k}^{-, n} = 0$ unless $i = 0$ or 1, and thus we have the correct form (2.10) for the eigenfunctions. We also show in Appendix B that the only nonzero $b_{p; i, k}$ for $i = 0$ are those for which k is even and the only nonzero $b_{p; i, k}$ for $i = 1$ or -1 are those for which k is odd. All these results hold no matter whether $i\omega_0 + k\omega \neq 0$ or $= 0$ for certain i and k. In the case $i\omega_0 + k\omega \neq 0$, the nonzero b's are

$b_{p; 0, k}$, p even, k even

and

$b_{p; \pm 1, k}$, p odd, k odd.

The recursion scheme for the case $n = 0$ (spontaneous

$p\downarrow$	1	2	3	4	5	6	7	8
1	x							
2		x						
3	x		x					
4		x		x				
5	x		x		x			
6		x		x		x		
7	x		x		x		x	
8		x		x		x		x

$k\rightarrow$

FIG. 1.

emission) in the nonresonance case, for example, is given in Fig. 1, where p denotes, as usual, the order of the perturbation and the crosses represent the nonzero b's which need to be calculated in order to obtain the coefficient A in Eq. (2.6) of that order. As can be seen from Eq. (2.9), the $b_{p;i,k}$ are expressed in terms of the A's and b's of the previous orders, and they can be determined *individually* as there is only one b of order p appearing in each equation (2.9). Thus, although the number of b's which need to be determined increases as the order of the perturbation term required increases, the labor involved increases only linearly, and the recursion scheme is seen to clearly provide a very powerful and very efficient method for numerical computation.

For a finite value of n, e.g., $n=4$, the recursion scheme is shown in Fig. 2. For very large value of n, the values of b become symmetrical about $k=0$ and the scheme is effectively reduced to that shown in Fig. 1.

It will be noted from (2.9) that if $\omega=\omega_0$, or $3\omega=\omega_0$, etc., the recurrence relation (2.9) can still be used, by rearranging terms on the left- and right-hand sides, to give $b_{p;i,k}$. These resonance cases are conveniently dealt with by expressing A and b in a different way, as we shall discuss in the following sections.

3. ENERGY EXPRESSIONS IN TERMS OF DETERMINANTS AND IN TERMS OF CONTINUED FRACTIONS

We now show that the recurrence relation (2.9) given in the preceding section can be expressed in terms of certain determinants which in turn satisfy certain recurrence relations.

Let us consider only the case $\sigma=-1$ because the result for the case $\sigma=1$ can be immediately deduced by a simple symmetry consideration (changing ω_0 into $-\omega_0$, etc.). We also omit the superscripts σ and n on the A's and b's for convenience. By using the following more compact notations,

$$\mu_p(2k)\equiv b_{p;0,2k}$$

and

$$\nu_p(2k-1)\equiv b_{p;1,2k-1}, \tag{3.1}$$

the recurrence relation (2.9), for $\sigma=-1$, can be written as two simultaneous equations:

$$2k\omega\mu_p(2k)+(n+2k+1)\nu_{p-1}(2k+1)+\nu_{p-1}(2k-1)=\sum_{q=0}^{p-1}A_{p-q}\mu_p(2k), \tag{3.2}$$

$$[(2k-1)\omega+\omega_0]\nu_{p-1}(2k-1)+(n+2k)\mu_{p-2}(2k)+\mu_{p-2}(2k-2)=\sum_{q=0}^{p-1}A_{p-q}\nu_p(2k-1)$$

with the conditions $\mu_p(0)=\delta_{p,0}$ and $\mu_p(2k)=\nu_{p-1}(2k-1)=0$ if $|2k|>p$. We define the generating functions

$$U_{2k}(\lambda)\equiv\sum_{p=1}^{\infty}\mu_p(2k)\lambda^p, \tag{3.3}$$

$$V_{2k-1}(\lambda)\equiv\sum_{p=1}^{\infty}\nu_p(2k-1)\lambda^{p+1}, \tag{3.4}$$

and

$$A(\lambda)\equiv\sum_{p=1}^{\infty}A_p\lambda^p. \tag{3.5}$$

From (3.2), we get

$$[2k\omega-A(\lambda)]U_{2k}(\lambda)+V_{2k-1}(\lambda)+(n+2k+1)V_{2k+1}(\lambda)=A(\lambda)\delta_{0,2k}, \tag{3.6}$$

$$[(2k-1)\omega+\omega_0-A(\lambda)]V_{2k-1}(\lambda)+\lambda^2[U_{2k-2}(\lambda)+(n+2k)U_{2k}(\lambda)]=-\lambda^2[\delta_{0,2k-2}+(n+2k)\delta_{0,2k}].$$

Finally by abbreviating

$$a_{2k}\equiv 2k\omega-A(\lambda),$$
$$a_{2k-1}\equiv[(2k-1)\omega+\omega_0-A(\lambda)]\lambda^{-2}, \tag{3.7}$$
$$c_k\equiv n+k,$$

Eq. (3.6) may be expressed in the tridiagonal matrix form by

$$\begin{bmatrix}
\cdot & & & & & & & \\
\cdot & \cdot & \cdot & & & & & \\
\cdot & \cdot & \cdot & & & & & \\
-1 & a_{-2} & -c_{-1} & & & & & \\
 & -1 & a_{-1} & -c_0 & & & & \\
 & & -1 & a_0 & -c_1 & & & \\
 & & & -1 & a_1 & -c_2 & & \\
 & & & & -1 & a_2 & -c_3 & \\
 & & & & & \cdot & \cdot & \cdot \\
 & & & & & & \cdot & \cdot \\
 & & & & & & & \cdot
\end{bmatrix}$$

$$\times\begin{bmatrix}\cdot\\ \cdot\\ \cdot\\ U_{-2}(\lambda)\\ -V_{-1}(\lambda)\\ U_0(\lambda)\\ -V_1(\lambda)\\ U_2(\lambda)\\ \cdot\\ \cdot\\ \cdot\end{bmatrix}=\begin{bmatrix}\cdot\\ \cdot\\ \cdot\\ 0\\ n\\ A(\lambda)\\ 1\\ 0\\ \cdot\\ \cdot\\ \cdot\end{bmatrix}. \tag{3.8}$$

Let us define the tridiagonal determinants

$p\downarrow$	$k\rightarrow$ −4	−3	−2	−1	0	1	2	3	4	5	6	7	8
1				x		x							
2			x				x						
3		x		x		x		x					
4	x		x				x		x				
5		x		x		x		x		x			
6	x		x				x		x		x		
7		x		x		x		x		x		x	
8	x		x				x		x		x		x

FIG. 2.

$$X_{m-k} \equiv \begin{vmatrix} a_{-m} & -c_{-m+1} & & & \\ -1 & a_{-m+1} & -c_{-m+2} & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \\ & & & -1 & a_{-1-k} \end{vmatrix}, \quad k = 0, 1, \ldots, m-1, \tag{3.9}$$

$$Y_{m+k} \equiv \begin{vmatrix} a_{k-1} & -c_k & & & \\ -1 & a_k & -c_{k+1} & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot \\ & & & -1 & a_m \end{vmatrix}, \quad k = 2, 3, \ldots, m+1. \tag{3.10}$$

The elements $g^{(-1)}_{m+1,m}$, $g^{(-1)}_{m+1,m+1}$, and $g^{(-1)}_{m+1,m+2}$ of the inverse of the tridiagonal matrix

$$G = \begin{bmatrix} a_{-m} & -c_{-m+1} & & & & & & & \\ -1 & a_{-m+1} & -c_{-m+2} & & & & & & \\ & \cdot & \cdot & \cdot & & & & & \\ & & \cdot & \cdot & \cdot & & & & \\ & & & -1 & a_{-1} & -c_0 & & & \\ & & & & -1 & a_0 & -c_1 & & \\ & & & & & -1 & a_1 & -c_2 & \\ & & & & & & \cdot & \cdot & \cdot \\ & & & & & & & \cdot & \cdot \\ & & & & & & & -1 & a_m \end{bmatrix} \tag{3.11}$$

are given by[4]

$$g^{(-1)}_{m+1,m} = Y_{m+2}X_{m-1}/D, \tag{3.12}$$

$$g^{(-1)}_{m+1,m+1} = X_m Y_{m+2}/D, \tag{3.13}$$

and

$$g^{(-1)}_{m+1,m+2} = c_1 X_m Y_{m+3}/D, \tag{3.14}$$

where $D = \det G$. From Eq. (3.8), we obtain

$$U_0(\lambda) = (1/D)[nY_{m+2}X_{m-1} + A(\lambda)X_m Y_{m+2} + (n+1)X_m Y_{m+3}]. \tag{3.15}$$

Since $U_0(\lambda) = 0$, we find

$$A(\lambda) = -\lim_{m\to\infty}[(nX_{m-1}/X_m) + (n+1)(Y_{m+3}/Y_{m+2})]. \tag{3.16}$$

Expression (3.16) represents a solution for the perturbation part of the energy $E(\lambda)$. However, it is not an explicit one because the determinants X's and Y's contain $A(\lambda)$ [see Eq. (3.7)]; but by letting $A(\lambda) = A_1\lambda + A_2\lambda^2 + \cdots$ the coefficients $A_1, A_2, \cdots$ can be readily obtained by iteration. The determinants $\{X_k\}$ and $\{Y_k\}$ have the convenient properties that they satisfy the following recurrence relations:

$$X_k = a_{k-(m+1)}X_{k-1} - c_{k-(m+1)}X_{k-2}, \quad k = 2, 3, \ldots, m, \tag{3.17}$$

and

$$Y_k = a_{k-(m+1)}Y_{k+1} - c_{k-m}Y_{k+2}, \quad k = m+2, m+3, \ldots, 2m, \tag{3.18}$$

where X_0 and Y_{2m+2} are defined to be equal to 1. As far as numerical computation is concerned, no particular advantage is gained in using (3.16) rather than the recurrence relation (2.9). However, (3.16) is a compact expression which, incidentally, is suggestive of the Padé approximent technique. It is also useful in giving an over-all view of the nature of the perturbation solution, as will become evident as we proceed to a continued fraction representation.

In the interim, we note that if the Hamiltonian (2.1) consists of only the rotating terms, i.e., if

$$H = \omega_0 S^z + \omega a^\dagger a + \lambda(a^\dagger S^- + aS^+), \tag{3.19}$$

the recurrence relation corresponding to (3.2) turns out to consist of only two equations

$$\begin{aligned} &0 + 0 + \nu_{p-1}(-1) = A_p, \\ &(-\omega + \omega_0)\nu_{p-1}(-1) + n\mu_{p-2}(0) + 0 = \sum_{q=0}^{p-1} A_{p-q}\nu_q(-1), \end{aligned} \tag{3.20}$$

the other $\nu_{p-1}(2k-1)$ and $\mu_p(2k)$ for $k \neq 0$ being equal to zero. Thus by defining as before

$$V_{-1}(\lambda) = \sum_{p=1}^{\infty} \nu_p(-1)\lambda^{p+1} \tag{3.21}$$

and

$$A(\lambda) = \sum_{p=1}^{\infty} A_p\lambda^p, \tag{3.22}$$

we get from (3.20)

$$\begin{aligned} &V_{-1}(\lambda) = A(\lambda), \\ &-(\omega - \omega_0)V_{-1}(\lambda) + n\lambda^2 = A(\lambda)V_{-1}(\lambda), \end{aligned} \tag{3.23}$$

or

$$A(\lambda)^2 + (\omega - \omega_0)A(\lambda) - n\lambda^2 = 0, \tag{3.24}$$

i.e.,

$$A(\lambda) = -\tfrac{1}{2}(\omega - \omega_0) \pm \tfrac{1}{2}[(\omega - \omega_0)^2 + 4n\lambda^2]^{1/2}. \tag{3.25}$$

Since $A(\lambda) = 0$ for $\lambda = 0$, only the root with the plus sign before the square root should be taken and

$$E(\lambda) = (n - \tfrac{1}{2})\omega + \tfrac{1}{2}[(\omega - \omega_0)^2 + 4n\lambda^2]^{1/2} \tag{3.26}$$

which is a well-known result.[5]

There are two series expansions for $E(\lambda)$. If

$$4n\lambda^2/(\omega - \omega_0)^2 < 1, \tag{3.26a}$$

then

$$E(\lambda) = (n - \tfrac{1}{2})\omega + \tfrac{1}{2}(\omega - \omega_0)[1 + 2\lambda^2 n/(\omega - \omega_0)^2 - 8\lambda^4 n^2/(\omega - \omega_0)^4 + \cdots]. \tag{3.26b}$$

On the other hand, if

$$4n\lambda^2/(\omega - \omega_0)^2 > 1, \tag{3.26c}$$

$$E(\lambda) = (n - \tfrac{1}{2})\omega + n^{1/2}\lambda[1 + (\omega - \omega_0)^2/8n\lambda^2 - (\omega - \omega_0)^4/128n^2\lambda^4 + \cdots]. \tag{3.26d}$$

One might be concerned about the fact that (3.26d) does not yield the unperturbed energy levels as $\lambda \to 0$. This, however, is irrelevent because (3.26d) is not valid for $\lambda = 0$ in view of (3.26c). Normal perturbation theory would yield only (3.26b). It would imply the existence of a divergence as $\omega \to \omega_0$. It is clear from (3.26) that it is not a characteristic of the model but only of the application of a series expansion outside of its range of convergence. This suggests that if we develop approximation technique for the discussion of the energy levels of the full Hamiltonian (2.1), we should be alert to the need for different kinds of expansions for different regimes in the space of the basic parameters of the problem, ω, ω_0, n, and λ.

Let us return to expression (3.16). Factorizing the matrices corresponding to the determinants X and Y into product of lower and upper triangular matrices, we write for X_m:

$$\begin{bmatrix} a_{-m} & -c_{-m+1} & & & \\ -1 & a_{-m+1} & -c_{-m+2} & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & a_{-1} \end{bmatrix} = \begin{bmatrix} 1 & & & \\ l_{21} & 1 & & \\ & l_{32} & 1 & \\ & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & & \\ & r_{22} & r_{23} & \\ & & \ddots & \ddots \\ & & & \ddots \end{bmatrix}.$$

Equating both sides, we get

$$r_{11} = a_{-m}$$

and, for $k > 1$,

(i) $r_{kk} = a_{-m+k-1} - l_{k,k-1}r_{k-1,k}$,

(ii) $l_{k,k-1}r_{k-1,k-1} = -1$,

(iii) $r_{k-1,k} = -c_{-m+k-1}$.

Combining (i), (ii), and (iii), we get

$$r_{kk} = a_{-m+k-1} - c_{-m+k-1}/r_{k-1,k-1} \quad \text{for } k > 1.$$

Since

$$X_{m-1} = r_{11}r_{22}\cdots r_{m-1,m-1},$$

$$X_m = r_{11}r_{22}\cdots r_{mm},$$

we get

$$\frac{X_{m-1}}{X_m} = \frac{1}{r_{mm}} = \cfrac{1}{a_{-1} - \cfrac{c_{-1}}{r_{m-1,m-1}}} = \cfrac{1}{a_{-1} - \cfrac{c_{-1}}{a_{-2} - \cfrac{c_{-2}}{a_{-3} - \cfrac{c_{-3}}{\cdots - \cfrac{c_{-m+1}}{a_{-m}}}}}}.$$

Similarly,

$$\frac{Y_{m+3}}{Y_{m+2}} = \cfrac{1}{a_{-1} - \cfrac{c_2}{a_2 - \cfrac{c_3}{a_3 - \cfrac{c_4}{\cdots - \cfrac{c_m}{a_m}}}}}.$$

Substituting (3.7) into the above expressions, we find that $A(\lambda)$ is given by

$$A(\lambda) = -n\lambda^2 \cfrac{1}{\beta_{-1} - A(\lambda) + \cfrac{-(n-1)\lambda^2}{\beta_{-2} - A(\lambda) + \cfrac{-(n-2)\lambda^2}{\beta_{-3} - A(\lambda) + \cfrac{-(n-3)\lambda^2}{\cdots \; \beta_{-m+1} - A(\lambda) + \cfrac{-(n-m+1)\lambda^2}{\beta_{-m} - A(\lambda)}}}}}$$

$$- (n+1)\lambda^2 \cfrac{1}{\beta_1 - A(\lambda) + \cfrac{-(n+2)\lambda^2}{\beta_2 - A(\lambda) + \cfrac{-(n+3)\lambda^2}{\beta_3 - A(\lambda) + \cfrac{-(n+4)\lambda^2}{\cdots \; \beta_{m-1} - A(\lambda) + \cfrac{-(n+m)\lambda^2}{\beta_m - A(\lambda)}}}}}, \quad m \to \infty, \tag{3.27}$$

where $\beta_{2k-1} = (2k-1)\omega + \omega_0$, $\beta_{2k} = 2k\omega$, $k = 0, \pm 1, \pm 2, \cdots$, and where the first continued fraction in (3.27) terminates at $m = n$. A similar expression was given by Swain[6] using a different method. Continued fraction expressions were also used for various purposes by Autler and Towns,[7] Schweber,[8] Stenholm and Lamb,[9] and many others.[10] As with the determinental expression (3.16), $A(\lambda)$ can be solved iteratively from (3.27) by substituting $A(\lambda) = A_1\lambda + A_2\lambda^2 + \cdots$ into it and comparing coefficients of like powers of λ on both sides. The remarkably clear layout of expressions in (3.27), however, has the advantage of providing one with a clearer understanding of the structure of the perturbation series in various regimes of interest, as we shall discuss in the next section.

We want to point out that the results of this section, expressions (3.16) and (3.27) among them, can be derived *without* assuming the eigenvalues $E^{\sigma,n}(\lambda)$ and

eigenfunctions $f^{\sigma,n}(\lambda;u,v,z)$ to be *power series* in λ. In fact the recursion relations (3.6) can be derived without using the results of the previous section. Skipping Eqs. (3.1)—(3.5) in the beginning of this section, we let

$$E^{\sigma,n}(\lambda)=\tfrac{1}{2}\sigma\omega_0+n\omega+A^{\sigma,n}(\lambda) \tag{3.28}$$

and

$$f^{\sigma,n}(\lambda;u,v,z)=u^{(1+\sigma)/2}v^{(1-\sigma)/2}z^nB^{\sigma,n}(\lambda;u,v,z). \tag{3.29}$$

For $\sigma=-1$,

$$E(\lambda)=-\tfrac{1}{2}\omega_0+n\omega+A(\lambda) \tag{3.30}$$

and, in accordance with (2.10),

$$\begin{aligned}f(\lambda;u,v,z)&=vz^n+vz^n\sum_{k=-n}^{\infty}[U_k(\lambda)+uv^{-1}V_k(\lambda)]z^k\\&=vz^n\sum_{k=-n}^{\infty}\{[\delta_{k,0}+U_k(\lambda)]+uv^{-1}V_k(\lambda)\}z^k,\end{aligned} \tag{3.31}$$

where $U_0(\lambda)$ is set equal to zero to satisfy the initial condition. Substituting (3.30) and (3.31) into the eigenvalue equation (2.3), employing a similar method used in Appendix A for cancelling out the factors vz^n on both sides by making use of the relations given by (A7), and then comparing coefficients of z^k and $uv^{-1}z^k$ on both sides of the equation gives us two recursion relations relating the $U_k(\lambda)$'s and $V_k(\lambda)$'s. The same argument as that given in Appendix B now gives $U_k(\lambda)=0$ if k is odd and $V_k(\lambda)=0$ if k is even. The recursion relations (3.6) immediately follow.

The difference between the recurrence relations (3.6) and the recurrence relations (2.9) becomes clear if we compare Eq. (3.8) with the recursion scheme shown in Fig. B1 in Appendix B. The assumption of *power series* in λ for the energies and the eigenfunctions fixes a starting point (the zeroth order in λ), so to speak, from which one can proceed to calculate the relevant quantities order by order recursively. The recurrence relations (3.8), on the other hand, do not have this property and must be solved in a different manner.

4. PERTURBATION SERIES FOR $E(\lambda)$

We have seen that the energy eigenvalues corresponding to the quantum numbers $\sigma=-1$ and arbitrary n are

$$E(\lambda)=-\tfrac{1}{2}\omega_0+n\omega+A(\lambda), \tag{4.1}$$

where $A(\lambda)$ is given by (3.27). As noted earlier, $A(\lambda)$ can be solved iteratively. Let us examine this idea more carefully. By letting $A(\lambda)=A_1\lambda+A_2\lambda^2+\cdots$, the second continued fraction in (3.27), expanded in the normal way in which $A(\lambda)$ is assumed to be sufficiently small, always gives an expansion consisting of only even powers of λ (except in the trivial case $\omega=\omega_0=0$). The same is true with the first continued fraction in (3.27) if $\beta_{-1}\neq 0$ (i.e., $\omega\neq\omega_0$). Thus if $\omega\neq\omega_0$ and provided that $A(\lambda)$ is sufficiently small, we shall get an expansion of $A(\lambda)$ which consists of only even powers of λ:

$$A(\lambda)=A_2\lambda^2+A_4\lambda^4+A_6\lambda^6+\cdots, \tag{4.2}$$

where

$$A_2=\frac{-\omega+(2n+1)\omega_0}{\omega^2-\omega_0^2}, \tag{4.3}$$

$$A_4=-\frac{n}{(\omega-\omega_0)^2}\left(A_2+\frac{n-1}{-2\omega}\right)-\frac{n+1}{(\omega+\omega_0)^2}\left(A_2+\frac{n+2}{2\omega}\right), \tag{4.4}$$

$$\begin{aligned}A_6=&-\frac{n}{(\omega-\omega_0)^2}\left[A_4+\frac{n-1}{4\omega^2}\left(A_2+\frac{n-2}{-3\omega+\omega_0}\right)\right.\\&\left.+\frac{1}{-\omega+\omega_0}\left(A_2+\frac{n-1}{-2\omega}\right)^2\right]+\frac{n+1}{(\omega+\omega_0)^2}\left[A_4+\frac{n+2}{4\omega^2}\right.\\&\left.\times\left(A_2+\frac{n+3}{3\omega+\omega_0}\right)+\frac{1}{\omega+\omega_0}\left(A_2+\frac{n+2}{2\omega}\right)^2\right],\end{aligned} \tag{4.4'}$$

etc. As with the appearance of the resonance factors $(\omega-\omega_0)$ in the denominators of the coefficients A_2, A_4, and A_6, it is easy to see that third order resonance factors $(3\omega-\omega_0)$ begin to appear in the denominators of A_6 and the fifth order resonance factors $(5\omega-\omega_0)$ begin to appear in the denominators of A_{10}, and so on. The presence of such factors in a perturbation series implies trouble which normally compels one to devise "tricks," such as summing a certain subset of terms of the series to avoid the divergence caused by the possibility of $\omega=\omega_0$ or $3\omega=\omega_0$, etc.; the use of these tricks is forced because one attempts to apply the same perturbation series to handle all cases. We now wish to demonstrate that prevention, namely, getting the correct expansion from the start, rather than curing, namely, employing tricks, is the preferred policy. If $\omega=\omega_0$, one should, of course, set $\beta_{-1}=0$ in (3.27) before expanding. Then letting $A(\lambda)=A_1\lambda+A_2\lambda^2+\cdots$, we see that the first continued fraction now gives an expansion in *all* powers of λ. By comparing the coefficients of like powers of λ on both sides of the resulting equation, we obtain an expansion[11]

$$\begin{aligned}A(\lambda)=&n^{1/2}\lambda-\frac{1}{2\omega}\lambda^2-\frac{n^{3/2}}{8\omega^2}\lambda^3-\frac{n}{4\omega^3}\lambda^4\\&-\frac{13n^{5/2}\lambda^5}{128\omega^4}\left(1+\frac{20}{13n^2}\right)-\cdots.\end{aligned} \tag{4.5}$$

Thus, in a proper expansion, factors $(\omega-\omega_0)$ never appear in denominators. Furthermore, (4.5) turns out to be different from the series (4.2) in its structural form.

If ω is not exactly equal to ω_0 but is very close to it, a perturbation expansion for $A(\lambda)$ can also be obtained by iteration from (3.27) by expanding the right-hand side of it in the right way. We let, in this case, $A(\lambda)$ be a series of powers of β_{-1}:

$$A(\lambda)=A_0(\lambda)+B_1(\lambda)\beta_{-1}+B_2(\lambda)\beta_{-1}^2+\cdots. \tag{4.6}$$

Substituting this into the left- and right-hand sides of (3.27), and expanding the right-hand side *in powers of* β_{-1}, then equating the coefficients of like powers of β_{-1} enables us to determine $A_0(\lambda)$, $B_1(\lambda)$, $B_2(\lambda),\cdots$ successively. Thus we find that $A_0(\lambda)$ is given by

$$\begin{aligned}A_0(\lambda)=&-\frac{n\lambda^2}{-A_0(\lambda)+\dfrac{-(n-1)\lambda^2}{\beta_{-2}-A_0(\lambda)+\dfrac{-(n-2)\lambda^2}{\cdots}}}\\&-\frac{(n+1)\lambda^2}{\beta_1-A_0(\lambda)+\dfrac{-(n+2)\lambda^2}{\beta_2-A_0(\lambda)+\dfrac{-(n+3)\lambda^2}{\cdots}}}.\end{aligned} \tag{4.7}$$

Substituting

$$A_0(\lambda) = A_1\lambda + A_2\lambda^2 + A_3\lambda^3 + A_4\lambda^4 + \cdots \tag{4.8}$$

into both sides of (4.7) and comparing the coefficients of like powers of λ enable us to determine $A_1, A_2, A_3, A_4, \cdots$ successively. It is clear from (4.7) that $A_0(\lambda)$ reduces to the $A(\lambda)$ given by (4.5) if ω is set equal to ω_0. The first few coefficients of $A_0(\lambda)$ are found to be

$$\begin{aligned}
A_1 &= n^{1/2},\\
A_2 &= \frac{n}{2\omega} - \frac{n}{\omega+\omega_0}\left(1+\frac{1}{n}\right),\\
A_3 &= \frac{n^{3/2}}{8\omega^2}\left(1+\frac{2}{n}\right) - \frac{n^{3/2}}{(\omega+\omega_0)^2}\left(1+\frac{1}{n}\right),\\
A_4 &= \frac{n^2}{8\omega^3}\left(1+\frac{4}{n}-\frac{1}{n^2}\right) + \frac{n^2}{4\omega^2(3\omega-\omega_0)}\left(1-\frac{3}{n}+\frac{2}{n^2}\right)\\
&\quad - \frac{n^2}{2\omega(\omega+\omega_0)^2}\left(1+\frac{2}{n}+\frac{1}{n^2}\right) - \frac{n^2}{(\omega+\omega_0)^3}\left(1+\frac{1}{n}\right),
\end{aligned} \tag{4.9}$$

which, as can be easily verified, reduce to the corresponding coefficients in (4.5) on setting $\omega = \omega_0$. The coefficient $B_1(\lambda)$ of β_{-1} in (4.6) is found to be given by the following equation:

$$\begin{aligned}
B_1(\lambda) = &-\frac{n\lambda^2}{Y_{-1}^2}\left\{B_1(\lambda) - 1 + \frac{(n-1)\lambda^2}{Y_{-2}^2}\right.\\
&\times\left.\left[B_1(\lambda) + \frac{(n-2)\lambda^2}{Y_{-3}^2}\left(B_1(\lambda) + \frac{(n-3)\lambda^2}{Y_{-4}^2}(\cdots)\right)\right]\right\}\\
&-\frac{(n+1)\lambda^2}{Y_1^2}\left\{B_1(\lambda) + \frac{(n+2)\lambda^2}{Y_2^2}\right.\\
&\times\left.\left[B_1(\lambda) + \frac{(n+3)\lambda^2}{Y_3^2}\left(B_1(\lambda) + \frac{(n+4)\lambda^2}{Y_4^2}(\cdots)\right)\right]\right\},
\end{aligned} \tag{4.10}$$

where the Y's are defined by

$$Y_{-i} \equiv X_{-i} + \cfrac{-(n-i)\lambda^2}{X_{-i-1} + \cfrac{-(n-i-1)\lambda^2}{X_{-i-2} + \cfrac{-(n-i-2)\lambda^2}{\cdots}}}, \tag{4.11}$$

$$Y_i = X_i + \cfrac{-(n+i+1)\lambda^2}{X_{i+1} + \cfrac{-(n+i+2)\lambda^2}{X_{i+2} + \cfrac{-(n+i+3)\lambda^2}{\cdots}}}, \tag{4.12}$$

$i = 1, 2, 3, \cdots$, and the X's being given by

$$X_i = \beta_i - A_0(\lambda) \tag{4.13}$$

except

$$X_{-1} = -A_0(\lambda). \tag{4.14}$$

Again by substituting

$$B_1(\lambda) = p_0 + p_1\lambda + p_2\lambda^2 + \cdots \tag{4.15}$$

into both sides of (4.10) and comparing coefficients of like powers of λ we are able to determine $p_0, p_1, p_2, \cdots$ successively. We find

$$\begin{aligned}
p_0 &= \tfrac{1}{2},\\
p_1 &= -\frac{1}{4\omega n^{1/2}} + \frac{n^{1/2}}{2(\omega+\omega_0)}\left(1+\frac{1}{n}\right),\\
p_2 &= -\frac{n}{4\omega^2}\left(1-\frac{1}{4n}-\frac{1}{4n^2}\right) - \frac{n}{4\omega(\omega+\omega_0)}\left(\frac{1}{n}+\frac{1}{n^2}\right)\\
&\quad + \frac{n}{2(\omega+\omega_0)}\left(1+\frac{3}{2n}+\frac{1}{2n^2}\right).
\end{aligned} \tag{4.16}$$

The equation for determining $B_2(\lambda)$, the coefficient of β_{-1}^2 in (4.6), is also not difficult to find, and is given by the following:

$$\begin{aligned}
B_2(\lambda) = &-\frac{n\lambda^2}{Y_{-1}^2}\left\{B_2(\lambda) + \frac{(n-1)\lambda^2}{Y_{-2}^2}\left[B_2(\lambda) + \frac{(n-2)\lambda^2}{Y_{-3}^2}(\cdots)\right.\right.\\
&\left.\left. + \frac{(n-2)\lambda^2}{Y_{-3}^3}\left(B_1(\lambda) + \frac{(n-3)\lambda^2}{Y_{-4}^2}(\cdots)\right)^2\right]\right.\\
&\left. + \frac{(n-1)\lambda^2}{Y_{-2}^3}\left[B_1(\lambda) + \frac{(n-2)\lambda^2}{Y_{-3}^2}\left(B_1(\lambda) + \frac{(n-3)\lambda^3}{Y_{-4}^2}(\cdots)\right)\right]^2\right.\\
&\left. - \frac{n\lambda^2}{Y_{-1}^3}\left\{B_1(\lambda) - 1 + \frac{(n-1)\lambda^2}{Y_{-2}^2}\left[B_1(\lambda) + \frac{(n-2)\lambda^2}{Y_{-3}^2}\right.\right.\right.\\
&\left.\left.\left.\times\left(B_1(\lambda) + \frac{(n-3)\lambda^2}{Y_{-4}^2}(\cdots)\right)\right]\right\}^2 - \frac{(n+1)\lambda^2}{Y_1^2}\right.\\
&\times\left\{B_2(\lambda) + \frac{(n+2)\lambda^2}{Y_2^2}\left[B_2(\lambda) + \frac{(n+3)\lambda^2}{Y_3^2}(\cdots)\right.\right.\\
&\left.\left. + \frac{(n+3)\lambda^2}{Y_3^3}\left(B_1(\lambda) + \frac{(n+4)\lambda^2}{Y_4^2}(\cdots)\right)^2\right]\right.\\
&\left. + \frac{(n+2)\lambda^2}{Y_2^3}\left[B_1(\lambda) + \frac{(n+3)\lambda^2}{Y_3^2}\right.\right.\\
&\left.\left.\times\left(B_1(\lambda) + \frac{(n+4)\lambda^2}{Y_4^2}(\cdots)\right)\right]^2 - \frac{(n+1)\lambda^2}{Y_1^3}\right.\\
&\left.\times\left\{B_1(\lambda) + \frac{(n+2)\lambda^2}{Y_2^2}\left[B_1(\lambda) + \frac{(n+3)\lambda^2}{Y_3^2}\right.\right.\right.\\
&\left.\left.\left.\times\left(B_1(\lambda) + \frac{(n+4)\lambda^2}{Y_4^2}(\cdots)\right)\right]\right\}^2.
\end{aligned} \tag{4.17}$$

We find that in terms of power series in λ,

$$B_2(\lambda) = q_{-1}/\lambda + q_0 + q_1\lambda + \cdots, \tag{4.18}$$

where

$$\begin{aligned}
q_{-1} &= \frac{1}{8n^{1/2}}, \quad q_0 = 0,\\
q_1 &= -\frac{3n^{1/2}}{64\omega^2}\left(1-\frac{1}{n}+\frac{1}{n^2}\right) - \frac{n^{1/2}}{8(\omega+\omega_0)^2}\left(1+\frac{5}{2n}+\frac{3}{2n^2}\right)\\
&\quad + \frac{3n^{1/2}}{16\omega(\omega+\omega_0)}\left(\frac{1}{n}+\frac{1}{n^2}\right).
\end{aligned} \tag{4.19}$$

It becomes clear that for the case $\omega \approx \omega_0$, the general form of the perturbation expansion for $A(\lambda)$ is

$$A(\lambda) = A_0(\lambda) + B(\lambda)\beta_{-1} + C(\lambda)\beta_{-1}^2/\lambda + D(\lambda)\beta_{-1}^3/\lambda^2 + \cdots, \tag{4.20}$$

where $B(\lambda), C(\lambda), D(\lambda), \cdots$ are series in powers of λ beginning with some constants, namely, the expansion of $A(\lambda)$ contains all powers as well as all inverse powers of λ, although it should be remembered that the expansion parameter here is β_{-1} and not λ. The interesting fact is that here we have another form of the perturbation expansion of $A(\lambda)$ which is structurally different from those given by (4.2) and (4.5). All these expansions were derived, as we have seen, from the same expression (3.27). The expansions were carried out correctly from the beginning depending on the regime of interest. Numerically therefore, (4.2), (4.5),

and (4.20) are expected to give a continuous curve as ω is varied around the value of ω_0, provided, of course, that these series are convergent.

The perturbation series for the higher resonance cases ($3\omega=\omega_0$, $5\omega=\omega_0$, etc.) are equally interesting. From (3.27) we see that if $\beta_{-1}\neq 0$, i.e., if $\omega\neq\omega_0$, the expansion of $A(\lambda)$ will consist of *only even powers* of λ irrespective of whether β_{-3} or β_{-5}, etc., is or is not equal to zero [assuming that $A(\lambda)$ is sufficiently small]. If $3\omega=\omega_0$, then, of course, we put $\beta_{-3}=0$ in (3.27) before expanding, and hence no $(3\omega-\omega_0)$ factors will appear in the denominators of the coefficients of λ^6 and higher powers of λ. Since the factors $(3\omega-\omega_0)$ first appear in the coefficient of λ^6 in the nonresonant series (4.2), it may be asked whether the coefficients A_2 and A_4 in the correctly expanded series for the case $3\omega=\omega_0$ are equal to ones given by setting $3\omega=\omega_0$ in (4.3) and (4.4). For A_2 the answer is yes, but for A_4 the answer is no. This can be easily seen from (3.27) as we observe that putting $\beta_{-3}=0$ affects not only the coefficients of λ^6 and higher powers of λ but also the coefficient of λ^4 (but none of the earlier ones) in the resulting expansion. The expansion $A(\lambda)$ for the case $3\omega=\omega_0$ is found to be

$$A(\lambda)=A_2\lambda^2+A_4\lambda^4+A_6\lambda^6+\cdots, \tag{4.21}$$

where

$$A_2=-(3n+1)/4\omega \tag{4.22}$$

and

$$A_4=-\frac{n}{4\omega^2}\left(A_2+\frac{n-1}{-2\omega+\dfrac{n-2}{A_2-\dfrac{n-3}{4\omega}}}\right)-\frac{n+1}{16\omega^2}\left(A_2+\frac{n+2}{2\omega}\right). \tag{4.23}$$

It is seen that setting $3\omega=\omega_0$ in (4.3) and (4.4) gives the A_2 in (4.22) but not the A_4 in (4.23). The number

$$\frac{n-2}{A_2-\dfrac{n-3}{4\omega}}$$

in (4.23) which is not present in (4.4) is the contribution resulting from putting $\beta_{-3}=0$ *before* expanding. The expression for A_6 is found to be

$$\begin{aligned}A_6=&-\frac{n}{4\omega^2}\left\{A_4+\frac{n-1}{\left(-2\omega+\dfrac{n-2}{A_2-\dfrac{n-3}{4\omega}}\right)^2}\right.\\&\times\left[A_2+\frac{n-2}{\left(A_2-\dfrac{n-3}{4\omega}\right)^2}\left(A_4-\frac{n-3}{4\omega}\left(A_2-\frac{n-4}{2\omega}\right)\right)\right]\\&+\frac{1}{2\omega}\left(A_2+\frac{n-1}{-2\omega+\dfrac{n-2}{A_2-\dfrac{n-3}{4\omega}}}\right)^2\\&\left.-\frac{n+1}{16\omega^2}\left[A_4+\frac{n+2}{4\omega^2}\left(A_2+\frac{n+3}{6\omega}\right)+\frac{1}{4\omega}\left(A_2+\frac{n+2}{2\omega}\right)^2\right]\right\}\end{aligned} \tag{4.23'}$$

which may be compared with the expression for A_6 for the case $3\omega\neq\omega_0$ given in (4.4'). Similarly in the case $5\omega=\omega_0$, the correctly expanded perturbation series for $A(\lambda)$ will consist of only even powers of λ without any singular factors $1/(5\omega-\omega_0)$ in it, and the coefficients A_2, A_4, and A_6, but not A_8 and higher order coefficients, will be equal to ones given by setting $5\omega=\omega_0$ in the nonresonant series.

For the case that 3ω is not exactly equal to but is very close to the value of ω_0, the perturbation expansion for $A(\lambda)$ can be obtained by iteration in a similar way as for the case $\omega\approx\omega_0$. Thus we let

$$A(\lambda)=A_0(\lambda)+B_1(\lambda)\beta_{-3}+B_2(\lambda)\beta_{-3}^2+\cdots \tag{4.24}$$

and substitute this into the left- and right-hand sides of (3.27), then expanding out the right-hand side as power series in β_{-3} and equating the coefficients of like powers of β_{-3} enable us to determine $A_0(\lambda)$, $B_1(\lambda)$, $B_2(\lambda)$, successively. We find that $A_0(\lambda)$ in this case is given by the following equation:

$$\begin{aligned}A_0(\lambda)=&-\cfrac{n\lambda^2}{\beta_{-1}-A_0(\lambda)+\cfrac{-(n-1)\lambda^2}{\beta_{-2}-A_0(\lambda)+\cfrac{-(n-2)\lambda^2}{-A_0(\lambda)+\cfrac{-(n-3)\lambda^2}{\cdots}}}}\\&-\cfrac{(n+1)\lambda^2}{\beta_1-A_0(\lambda)+\cfrac{-(n+2)\lambda^2}{\beta_2-A_0(\lambda)+\cfrac{-(n+3)\lambda^2}{\beta_3-A_0(\lambda)+\cfrac{-(n+4)\lambda^2}{\cdots}}}}.\end{aligned} \tag{4.25}$$

In terms of power series in λ, it is easy to see that $A_0(\lambda)$ is a series in even powers of λ beginning with a term in λ^2 and that it reduces to the series given by (4.21) on setting $3\omega=\omega_0$. The coefficient $B_1(\lambda)$ of β_{-3} in (4.24) is given by

$$\begin{aligned}B_1(\lambda)=&-\frac{n\lambda^2}{Y_{-1}^2}\left\{B_1(\lambda)+\frac{(n-1)\lambda^2}{Y_{-2}^2}\right.\\&\left.\times\left[B_1(\lambda)+\frac{(n-2)\lambda^2}{Y_{-3}^2}\left(B_1(\lambda)-1+\frac{(n-3)\lambda^2}{Y_{-4}^2}(\cdots)\right)\right]\right\}\\&-\frac{(n+1)\lambda^2}{Y_1^2}\left\{B_1(\lambda)+\frac{(n+2)\lambda^2}{Y_2^2}\left[B_1(\lambda)+\frac{(n+3)\lambda^2}{Y_3^2}\right.\right.\\&\left.\left.\times\left(B_1(\lambda)+\frac{(n+4)\lambda^2}{Y_4^2}(\cdots)\right)\right]\right\},\end{aligned} \tag{4.26}$$

where the Y's are defined as in (4.11) and (4.12) but the exception stated by Eq. (4.14) is here replaced by

$$X_{-3}=-A_0(\lambda). \tag{4.27}$$

Notice also the difference in the places where the -1 appears in (4.10) and (4.26). The equation for determining $B_2(\lambda)$ in (4.24) is also similar to (4.17) but with X_{-3} defined by (4.27) and with the -1 appearing after the factor $(n-2)\lambda^2/Y_{-3}^3$ instead of after the factor $n\lambda^2/Y_{-1}^3$. Noting that $A_0(\lambda)$ and hence Y_{-3} are series in even powers of λ beginning with terms in λ^2 while the rest of the Y's are series in even powers of λ beginning with some constant terms, it is not difficult to see that the expansion of $A(\lambda)$ for the case $3\omega\approx\omega_0$ is generally of the form

$$A(\lambda)=A_0(\lambda)+B(\lambda)\lambda^2\beta_{-3}+C(\lambda)\beta_{-3}^2+D(\lambda)\beta_{-3}^3/\lambda^2+\cdots, \tag{4.28}$$

where $B(\lambda), C(\lambda), D(\lambda), \cdots$ are series in even powers of λ beginning with some constant terms. Similarly for the case $5\omega \approx \omega_0$, the perturbation expansion for $A(\lambda)$ is generally of the form

$$A(\lambda) = A_0(\lambda) + B(\lambda)\lambda^6\beta_{-5} + C(\lambda)\lambda^4\beta_{-5}^2 + D(\lambda)\lambda^2\beta_{-5}^3 + E(\lambda)\beta_{-5}^4 + F(\lambda)\beta_{-5}^5/\lambda^2 + \cdots . \tag{4.29}$$

We thus see that for cases of near resonance of third and higher orders, the perturbation expansions of $A(\lambda)$ contain terms in even and inverse even powers of λ.

We believe a great deal has been learned here which will have many useful and far reaching implications on the general theory of perturbation expansions. The cause of many complications encountered in perturbation theory can be traced to our insistence on getting an expansion in power series of λ and our not realizing the fact that this power series may possibly take on different forms in various regimes of interest. We see that (3.27) is a perfectly good representation of the required solution and that if we do not insist on getting an expansion of the form $A(\lambda) = A_1\lambda + A_2\lambda^2 + \cdots$, then the calculation of $A(\lambda)$ for any given λ (sufficiently small), ω, and ω_0 can be done in the following straightforward manner from (3.27): Truncate the continued fractions at certain point, put in the values of λ, ω, and ω_0 and then some trial value of $A(\lambda)$, and perform iterations until both sides are equal, the standard method being used for choosing the next trial value of $A(\lambda)$ each time. We then repeat the same operations as we let the continued fractions extend, and hopefully the values of $A(\lambda)$ so calculated will quickly converge. The complications involving resonances would not even come up. If we insist on getting an expansion in power series of λ, then we should be prepared to accept and use possibly different forms of that series for various regimes of interest. Such flexibility in our thought is clearly useful even in dealing with problems when, unlike the two-level model discussed in this paper, we do not know how to make the correct expansion right from the start.

We end this section by mentioning a "trick" used by Shirley[12] for the removal of the singular terms in the Bloch–Siegert shifts. For very large n, denoting $n^{1/2}\lambda$ by b, Shirley found from a semiclassical theory that $E(b)$ is given by

$$E(b) = (n - \tfrac{1}{2})\omega + \frac{2\omega_0}{\omega^2 - \omega_0^2}b^2 - \frac{2\omega_0(\omega^2 + 3\omega_0^2)}{(\omega^2 - \omega_0^2)^3}b^4 + \left(\frac{32\omega_0^3(\omega^2 + \omega_0^2)}{(\omega^2 - \omega_0^2)^5} + \frac{8\omega_0}{(\omega^2 - \omega_0^2)(9\omega^2 - \omega_0^2)}\right)b^6 - \cdots , \tag{4.30}$$

which, incidentally, agrees with our quantum mechanical result given in (4.2).[13] It was noticed by Shirley that the square of $[\frac{1}{2}(\omega - \omega_0) + A(b)]$, where $A(b)$ is the perturbation part of $E(b)$ above, is nonsingular at $\omega = \omega_0$, namely,

$$[\tfrac{1}{2}(\omega - \omega_0) + A(b)]^2 = \tfrac{1}{4}(\omega - \omega_0)^2 + \frac{2\omega_0}{\omega + \omega_0}b^2 - \frac{2\omega_0}{(\omega + \omega_0)^3}b^4 + \frac{8\omega_0(\omega^2 - 5\omega\omega_0 - 2\omega_0^2)}{(\omega + \omega_0)^5(9\omega^2 - \omega_0^2)}b^6 - \cdots , \tag{4.31}$$

and thus $A(b)$ can actually be evaluated at $\omega = \omega_0$, giving

$$A(b) = b(1 - b^2/4\omega^2 - 3b^4/16\omega^4 - \cdots)^{1/2} = b - (1/8\omega^2)b^3 - (13/128\omega^4)b^5 - \cdots . \tag{4.32}$$

It is seen that this expression is a special case of our expression (4.5) for general n. As far as we know, no simple trick like that of Shirley will turn a nonresonant series (4.2) for general n into the resonant series (4.5), not to mention tricks which will turn (4.2) into the higher order resonant series. Thus again the importance of starting the expansion correctly (if one is really needed!) should be emphasized.

5. A LIDSTONE EXPANSION FOR THE PERTURBATION ENERGY

In this section we present a novel form of perturbation theory which is not an expansion in any of the basic variables of our Hamiltonian but rather one in terms of a new parameter γ which is chosen in such a manner that when $\gamma = 0$ our Hamiltonian contains the rotating terms only [cf. (3.19)] and when $\gamma = 1$ it contains only the "counter rotating" terms $(a^\dagger S^+ + aS^-)$. We thus choose it to be

$$H(\gamma) = \omega_0 S^z + \omega a^\dagger a + 2\lambda[\gamma(a^\dagger S^- + aS^+) + (1 - \gamma)(a^\dagger S^+ + aS^-)]. \tag{5.1}$$

Our basic Hamiltonian (2.1) is then $H(\frac{1}{2})$. The motivation for the introduction of $H(\gamma)$ is that we know the correction $A(2\lambda, 1)$ to the unperturbed levels (for $\sigma = -1$) $-\frac{1}{2}\omega_0 + n\omega$ for $\gamma = 1$

$$A(2\lambda, 1) = -\tfrac{1}{2}(\omega - \omega_0) + \tfrac{1}{2}[(\omega - \omega_0)^2 + 16n\lambda^2]^{1/2} \tag{5.2}$$

as well as the correction $A(2\lambda, 0)$ for $\gamma = 0$,

$$A(2\lambda, 0) = \tfrac{1}{2}(\omega + \omega_0) - \tfrac{1}{2}[(\omega + \omega_0)^2 + 16(n+1)\lambda^2]^{1/2}. \tag{5.3}$$

A primitive zero order guess for the perturbation correction to the unperturbed levels $-\frac{1}{2}\omega_0 + n\omega$ of $H(\gamma)$ would then be the linear interpolation formula

$$A(2\lambda, \gamma) = (1 - \gamma)A(2\lambda, 0) + \gamma A(2\lambda, 1). \tag{5.4}$$

Without further investigation there would be no reason to assume that this approximation is more or less valid for any special regime of ω, ω_0, or λ. It contains no singularities or divergences.

It would seem natural to seek a systematic generalization of (5.4) which would be a symmetrical series expansion about two points $\gamma = 0$ and $\gamma = 1$. Since a Taylor expansion about a single point completely characterizes an analytic function, a two point expansion would seem to have some redundancy. An appropriate two point generalization of the Taylor expansion was proposed in 1929 by Lidstone.[14] It avoids redundancy by requiring a knowledge of only even derivatives at the two points about which the expansion is made, thus differing from the Taylor expansion which employs all positive integral derivatives at one point. The Lidstone series of a function $f(\gamma)$ about the two points $\gamma = 0$ and $\gamma = 1$ is

$$f(\gamma) = f(1)\Lambda_0(\gamma) + f(0)\Lambda_0(1 - \gamma) + f''(1)\Lambda_1(\gamma) + f''(0)\Lambda_1(1 - \gamma) + \cdots , \tag{5.5}$$

where

$$\Lambda_0(\gamma)=\gamma,\quad \Lambda_n''(\gamma)=\Lambda_{n-1}(\gamma),$$
$$\Lambda_n(0)=\Lambda_n(1)=0. \tag{5.6}$$

The first few $\Lambda_n(\gamma)$ are

$$\Lambda_0(\gamma)=\gamma,\quad \Lambda_1(\gamma)=\gamma(\gamma^2-1)/6,$$
$$\Lambda_2(\gamma)=\gamma(\gamma^2-1)(3\gamma^2-7)/360,\quad \text{etc.} \tag{5.7}$$

Note that when $\gamma=\frac{1}{2}$,

$$\Lambda_0(\tfrac{1}{2})=\tfrac{1}{2},\quad \Lambda_1(\tfrac{1}{2})=-\tfrac{1}{16},\quad \Lambda_2(\tfrac{1}{2})=5/768,\cdots \tag{5.8}$$

A detailed investigation of the general theory of Lidstone series has been made by J. M. Whittaker[15] and D. V. Widder.[16] We now proceed to calculate $A''(2\lambda,\gamma)$ and $A''''(2\lambda,\gamma)$ when $\gamma=0$ and $\gamma=1$ so that we can construct the first three terms in the expansion of $A(2\lambda,\gamma)$.

When the Hamiltonian is given by (5.1), the continued fraction expression for the perturbation part $A(2\lambda,\gamma)$ of the energy $E^{n}(2\lambda,\gamma)$ can be readily shown to be

$$A(2\lambda,\gamma)=-4\gamma^2 n\lambda^2\cfrac{1}{\beta_{-1}-A(2\lambda,\gamma)+\cfrac{-4\gamma'^2(n-1)\lambda^2}{\beta_{-2}-A(2\lambda,\gamma)+\cfrac{-4\gamma^2(n-2)\lambda^2}{\beta_{-3}-A(2\lambda,\gamma)+\cfrac{-4\gamma'^2(n-3)\lambda^2}{\cdots}}}}$$
$$-4\gamma'^2(n+1)\lambda^2\cfrac{1}{\beta_1-A(2\lambda,\gamma)+\cfrac{-4\gamma^2(n+2)\lambda^2}{\beta_2-A(2\lambda,\gamma)+\cfrac{-4\gamma'^2(n+3)\lambda^2}{\beta_3-A(2\lambda,\gamma)+\cfrac{-4\gamma^2(n+4)\lambda^2}{\cdots}}}}, \tag{5.9}$$

where $\gamma'=1-\gamma$. To obtain $A''(2\lambda,0)$ and $A''''(2\lambda,0)$, let us write

$$A(2\lambda,\epsilon)=A(2\lambda,0)+\epsilon B_1(\lambda)+\epsilon^2B_2(\lambda)+\epsilon^3B_3(\lambda)+\epsilon^4B_4(\lambda)+\cdots \tag{5.10}$$

Substituting (5.10) into (5.9) and putting $\gamma=\epsilon$ and $\gamma'=1-\epsilon$, we see that to obtain the correct expansion on the right-hand side up to ϵ^4, the first continued fractions should include fractions up to $-1(1-\epsilon)^2(n-3)\lambda^2/[\beta_{-4}-A(2\lambda,\epsilon)]$ and the second continued fraction should include fractions up to $-(1-\epsilon)(n+5)\lambda^2/[\beta_5-A(2\lambda,\epsilon)]$. Expanding out in powers of ϵ and comparing the coefficients of like powers of ϵ, we get, for the coefficients of ϵ^0,

$$A(2\lambda,0)=-4(n+1)\lambda^2/[\beta_1-A(2\lambda,0)], \tag{5.11}$$

which gives (5.3) after taking account of the condition that $A(2\lambda,0)=0$ when $\lambda=0$. Defining, for $i=1,2,3,\cdots$,

$$c_i\equiv n+i,\quad c_{-i}\equiv n-i+1,$$
$$f_i\equiv 1/[\beta_i-A(2\lambda,0)],\quad f_{-i}\equiv 1/[\beta_{-i}-A(2\lambda,0)], \tag{5.12}$$
$$g_i\equiv 1/(f_i^{-1}-4c_{i+1}\lambda^2f_{i+1}),\quad g_{-i}\equiv 1/(f_{-i}^{-1}-4c_{-i}\lambda^2f_{-i-1}),$$

comparisons of the coefficients of ϵ, ϵ^2, ϵ^3, and ϵ^4 give

$$B_1(\lambda)=8c_1\lambda^2f_1(1+4c_1\lambda^2f_1^2)^{-1}, \tag{5.13}$$

$$B_2(\lambda)=(1+4c_1\lambda^2f_1^2)^{-1}\{-4c_0\lambda^2g_{-1}-16c_1c_2\lambda^4f_1^2g_2-4c_1\lambda^2f_1[1-f_1B_1(\lambda)]^2\}, \tag{5.14}$$

$$B_3(\lambda)=(1+4c_1\lambda^2f_1^2)^{-1}\{-4c_0\lambda^2g_{-1}^2(1+4c_{-1}\lambda^2f_{-2}^2)B_1(\lambda)+32c_0c_{-1}\lambda^4f_{-2}g_{-1}^2-16c_1c_2\lambda^4f_1^2g_2^2(1+4c_3\lambda^2f_3^2)B_1(\lambda)+128c_1c_2c_3\lambda^6f_1^2f_3g_2^2+8c_1\lambda^2f_1^2[1-f_1B_1(\lambda)]\times[B_2(\lambda)+4c_2\lambda^2g_2]-4c_1\lambda^2f_1^2B_1(\lambda)[1-f_1B_1(\lambda)]^2\}, \tag{5.15}$$

$$B_4(\lambda)=(1+4c_1\lambda^2f_1^2)^{-1}(-4c_0\lambda^2f_{-1}^3[(1+4c_{-1}\lambda^2f_{-2}^2)B_1(\lambda)-8c_{-1}\lambda^2f_{-2}]^2-4c_0\lambda^2g_{-1}^2\{B_2(\lambda)+4c_{-1}\lambda^2f_{-2}[1-f_{-2}B_1(\lambda)]^2+4c_{-1}\lambda^2f_{-2}^2[B_2(\lambda)+4c_{-2}\lambda^2g_{-3}]\}-16c_1c_2\lambda^4f_1^2g_2^3\times[(1+4c_3\lambda^2f_3^2)B_1(\lambda)-8c_3\lambda^2f_3]^2-16c_1c_2\lambda^4f_1^2g_2^2\{B_2(\lambda)+4c_3\lambda^2f_3[1-f_3B_1(\lambda)]^2+4c_3\lambda^2f_3^2[B_2(\lambda)+4c_4\lambda^2g_4]\}+8c_1\lambda^2f_1^2[1-f_1B_1(\lambda)]\{B_3(\lambda)+4c_2\lambda^2g_2^2[(1+4c_3\lambda^2f_3^2)B_1(\lambda)-8c_3\lambda^2f_3]\}-4c_1\lambda^2f_1^2[1-f_1B_1(\lambda)][1-3f_1B_1(\lambda)]\times[B_2(\lambda)+4c_2\lambda^2g_2]-4c_1\lambda^2f_1^3[B_2(\lambda)+4c_2\lambda^2g_2]^2). \tag{5.16}$$

$A''(2\lambda,0)$ and $A''''(2\lambda,0)$ are given in terms of $B_2(\lambda)$ and $B_4(\lambda)$ by

$$A''(2\lambda,0)=2B_2(\lambda), \tag{5.17}$$
$$A''''(2\lambda,0)=24B_4(\lambda). \tag{5.18}$$

Similarly, to obtain $A''(2\lambda,1)$ and $A''''(2\lambda,1)$, we write

$$A(2\lambda,1-\epsilon)=A(2\lambda,1)+\epsilon D_1(\lambda)+\epsilon^2D_2(\lambda)+\epsilon^3D_3(\lambda)+\epsilon^4D_4(\lambda)+\cdots. \tag{5.19}$$

Substituting (5.19) into both sides of (5.9) and putting $\gamma'=\epsilon$ and $\gamma=1-\epsilon$, we see that, to obtain the correct expansion on the right-hand side up to ϵ^4, the first continued fraction should include fractions up to $-(1-\epsilon)(n-4)\lambda^2/[\beta_{-5}-A(2\lambda,1-\epsilon)]$ and the second continued fraction should include fractions up to $-(1-\epsilon)(n+4)\lambda^2/[\beta_4-A(2\lambda,1-\epsilon)]$. Expanding out in powers of ϵ and comparing the coefficients of like powers of ϵ, we get, for the coefficient of ϵ^0,

$$A(2\lambda,1)=-4n\lambda^2/[\beta_{-1}-A(2\lambda,1)], \tag{5.20}$$

which gives (5.2) after noting the condition that $A(2\lambda,1)=0$ when $\lambda=0$. Comparisons of the coefficients of ϵ, ϵ^2, ϵ^3, and ϵ^4 give us the expressions for $D_1(\lambda)$, $D_2(\lambda)$, $D_3(\lambda)$, and $D_4(\lambda)$ in terms of which $A''(2\lambda,1)$ and $A''''(2\lambda,1)$ are given by

$$A''(2\lambda,1)=2D_2(\lambda), \tag{5.21}$$
$$A''''(2\lambda,1)=24D_4(\lambda). \tag{5.22}$$

The expressions for $D_1(\lambda)$, $D_2(\lambda)$, $D_3(\lambda)$, and $D_4(\lambda)$ are exactly the same as those given on the right-hand sides of (5.13), (5.14), (5.15), and (5.16) respectively, but with the c, f, and g now given by the following:

$$c_i \equiv n-i+1, \quad c_{-i} \equiv n+i$$
$$f_i \equiv 1/[\beta_{-i} - A(2\lambda,1)], \quad f_{-i} \equiv 1/[\beta_i - A(2\lambda,1)], \tag{5.23}$$
$$g_i \equiv 1/(f_i^{-1} - 4c_{i+1}\lambda^2 f_{i+1}), \quad g_{-i} \equiv 1/(f_{-i}^{-1} - 4c_{-i-1}\lambda^2 f_{-i-1})$$

for $i = 1, 2, 3, \cdots$. Thus we obtain

$$A(2\lambda, \tfrac{1}{2}) = \tfrac{1}{2}[A(2\lambda,0) + A(2\lambda,1)] - \tfrac{1}{16}[A''(2\lambda,0) + A''(2\lambda,1)] + \tfrac{5}{768}[A''''(2\lambda,0) + A''''(2\lambda,1)] + \cdots, \tag{5.24}$$

which provides a useful and instructive alternative to the perturbation expansions given in the previous sections. As far as we know, this is probably the first nontrivial application of the Lidstone expansion method in physics.

6. SUMMARY

We began our discussion of the two-level system with a differential equation (2.3) and obtained a recurrence relation (2.9) which we showed to provide a powerful technique for numerical computation of the energies of the system. We proceeded to obtain expressions in terms of ratios of determinants (3.16) and in terms of continued fractions (3.27) which we showed to clarify not only the nature of the perturbation solution for the two-level system in particular, but also of perturbation theory in general. We have also presented a novel and useful form of perturbation expansion (5.24) using an idea which may find applications in many other problems in physics.

APPENDIX A

In this Appendix, we give the derivation of the recurrence relation (2.9) in a little more detail. Let us rewrite (2.6) and (2.7) as

$$E^{\sigma,n}(\lambda) = \sum_{p=0}^{\infty} A_p^{\sigma,n}\lambda^p \tag{A1}$$

and

$$f(\lambda;u,v,z) = f_0^{\sigma,n}(u,v,z)\sum_{p=0}^{\infty} B_p^{\sigma,n}(u,v,z)\lambda^p, \tag{A2}$$

where $A_0^{\sigma,n} = (\sigma/2)\omega_0 + n\omega$ and $f_0^{\sigma,n}(u,v,z) = u^{(1+\sigma)/2}v^{(1-\sigma)/2}z^n$, and $B_0^{\sigma,n}(u,v,z)$ is defined to be equal to 1. Writing the Hamiltonian (2.1) as

$$H = H_0 + \lambda H_1, \tag{A3}$$

then substituting (A1) and (A2) into the eigenvalue equation for H,

$$\text{lhs} = \sum_{p=0}^{\infty}\lambda^p(H_0+\lambda H_1)f_0B_p = \sum_{p=0}^{\infty}\lambda^p[B_p(H_0f_0) + f_0(H_0B_p) + H_1f_0B_{p-1}] \tag{A4}$$

and

$$\text{rhs} = \left(\sum_{p=0}^{\infty}A_p\lambda^p\right)\left(f_0\sum_{q=0}^{\infty}B_q\lambda^q\right) = A_0f_0\sum_{q=0}^{\infty}B_q\lambda^q + f_0\left(\sum_{p=1}^{\infty}A_p\lambda^p\right)\left(\sum_{q=0}^{\infty}B_q\lambda^q\right) = \sum_{p=0}^{\infty}\lambda^p\left(B_pA_0f_0 + f_0\sum_{q=0}^{p-1}A_{p-q}B_q\right). \tag{A5}$$

Comparing coefficients of λ^p in (A4) and (A5) and noting that $H_0f_0 = A_0f_0$, we get

$$f_0(H_0B_p) + H_1f_0B_{p-1} = f_0\sum_{q=0}^{p-1}A_{p-q}B_q. \tag{A6}$$

f_0 can be factored out in $H_1f_0B_{p-1}$, for if we write $f_0 = u^av^bz^c$, where $a = \frac{1}{2}(1+\sigma)$, $b = \frac{1}{2}(1-\sigma)$, $c = n$, we have the following:

$$zv\frac{\partial}{\partial u}f_0B_{p-1} = f_0\left(au^{-1}vz + vz\frac{\partial}{\partial u}\right)B_{p-1},$$
$$\frac{\partial}{\partial z}u\frac{\partial}{\partial v}f_0B_{p-1} = f_0\left(bcuv^{-1}z^{-1} + cuz^{-1}\frac{\partial}{\partial v} + buv^{-1}\frac{\partial}{\partial z} + u\frac{\partial^2}{\partial v\,\partial z}\right)B_{p-1},$$
$$zu\frac{\partial}{\partial v}f_0B_{p-1} = f_0\left(buv^{-1}z + uz\frac{\partial}{\partial v}\right)B_{p-1},$$
$$\frac{\partial}{\partial z}v\frac{\partial}{\partial u}f_0B_{p-1} = f_0\left(acu^{-1}vz^{-1} + cvz^{-1}\frac{\partial}{\partial u} + au^{-1}v\frac{\partial}{\partial z} + v\frac{\partial^2}{\partial u\,\partial z}\right)B_{p-1}. \tag{A7}$$

Thus writing out the operators H_0 and H_1 in full, substituting (A7) into (A6), and cancelling f_0 on both sides, we get

$$\left[\omega_0\tfrac{1}{2}\left(u\frac{\partial}{\partial u} - v\frac{\partial}{\partial v}\right) + \omega z\frac{\partial}{\partial z}\right]B_p + \left[\left(au^{-1}vz + vz\frac{\partial}{\partial u}\right) + \left(bcuv^{-1}z^{-1} + cuz^{-1}\frac{\partial}{\partial v} + buv^{-1}\frac{\partial}{\partial z} + u\frac{\partial^2}{\partial v\,\partial z}\right) + \left(bcuv^{-1}z + uz\frac{\partial}{\partial v}\right) + \left(acu^{-1}vz^{-1} + cvz^{-1}\frac{\partial}{\partial u} + au^{-1}v\frac{\partial}{\partial z} + v\frac{\partial^2}{\partial u\,\partial z}\right)\right]B_{p-1} = \sum_{q=0}^{p-1}A_{p-q}B_q. \tag{A8}$$

Now substituting (2.8) into (A8) and comparing the coefficients of $u^iv^{-i}z^k$ on both sides, we obtain the recurrence relation (2.9).

APPENDIX B

In this appendix, we shall exhibit the zero and nonzero $b_{p;i,k}$ as given by Eq. (2.9). First let us show that $b_{p;i,k} = 0$ for $k < n$.

According to (2.8), the nonzero b's, which we denote by the x's, are shown in Fig. B1 as p and k increase. Consider first the case $i\omega_0 + k\omega \neq 0$. From (2.9), we see that the contributions to $b_{p;\cdot,k}$ come from $b_{p-1;\cdot,k-1}$, $b_{p-1;\cdot,k+1}$, and $b_{q;\cdot,k}$, $q = 1, 2, \ldots, p-1$. As p increases, consider the $b_{p;\cdot,k}$ for which the outer most k value on the left ($= -p$) becomes equal to $-(n+1)$. The contributions of this $b_{n+1;\cdot,-(n+1)}$ come only from the terms $b_{p-1;\cdot,k+1}$. But we see from (2.9) that those terms contain the factors $(n+k+1)$ which are equal to zero for $k = -(n+1)$ and hence $b_{n+1;\cdot,-(n+1)} = 0$. For $p = n+2$, $b_{n+2;\cdot,-(n+2)} = 0$ because $b_{n+1;\cdot,-(n+1)} = 0$, and again $b_{n+2;\cdot,-(n+1)} = 0$ because of $n+k+1 = 0$. It is then clear that all $b_{p;i,k}$ for which $k < n$ are equal to zero.

If $i\omega_0 + k\omega = 0$, then, by rearranging the terms in (2.9) and replacing $p-1$ by p, we see that the contributions to $b_{p;\cdot,k}$ come from $b_{p;\cdot,k-1}$, $b_{p;\cdot,k+1}$, and $b_{q;\cdot,k}$, $q = 1, 2, \ldots, p-1$. Then a similar consideration as above shows that $b_{p;\cdot,k} = 0$ for $k < n$.

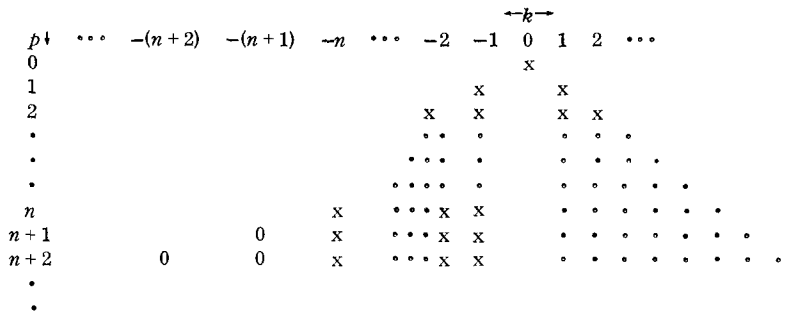


FIG. B1.

We now show in a similar way that $b_{p;i,k}=0$ unless $i=0$ or 1 for $\sigma=-1$, the crucial factors here being $[\frac{1}{2}(1+\sigma)+i+1]=i+1$ and $[-\frac{1}{2}(1+\sigma)-i+2]=i+2$. For $p=1$, $i=-1$, the contribution to $b_{p;i,\cdot}$ comes from $b_{p-1;i+1,\cdot}$ if $i\omega_0+k\omega\neq 0$, and from $b_{p;i+1,\cdot}$ if $i\omega_0+k\omega=0$. But the factor $(i+1)$ which is zero for $i=-1$ makes $b_{1;-1,\cdot}=0$. For $p=2$, $b_{2;-2,\cdot}=0$ because $b_{1;-1,0}=0$ or because $b_{2;-1,0}=0$ in the case $i\omega_0+k\omega=0$, and $b_{2;-1,0}=0$ because of the factor $(i+1)=0$ for $i=-1$. It readily follows that $b_{p;i,\cdot}=0$ for any negative values of i.

For positive values of i, there is a nonzero contribution to $b_{1;1,\cdot}$ from $b_{0;0,0}$ and hence $b_{1;1,\cdot}$ is nonzero. For $p=2$, $i=2$, however, the factor $(-i+2)=0$ makes the contribution from $b_{1;1,\cdot}$ zero, or in the case $i\omega_0+k\omega=0$, the factor $(-i+2)=0$ makes the contribution from $b_{2;1,\cdot}$ zero. For $p=3$, $b_{3;3,\cdot}=0$ because $b_{2;2,\cdot}=0$ or because $b_{3;2,\cdot}$ is zero in the case $i\omega_0+k\omega=0$, and $b_{3;2,\cdot}=0$ again because of $-i+2=0$. It follows that $b_{p;i,\cdot}=0$ for all $i>1$.

In a similar way, it can be shown that $b_{p;i,k}=0$ unless $i=0$ or -1 for the case $\sigma=1$.

We now go one step further and show that the nonzero $b_{p;i,k}$ for $i=1$ or -1 are those for which k is odd, and that this is so irrespective of whether $i\omega_0+k\omega\neq 0$ or $=0$. This result is a consequence of the nature of the interaction terms in the Hamiltonian and follows readily by noting that $b_{\cdot;i,k}$ is "connected" to $b_{\cdot;i\pm1,k\pm1}$ and $b_{\cdot;i,k}$ and that the initial condition is $b_{p;0,0}=\delta_{p,0}$. Thus, imagine that (i,k) are coordinates of lattice points in a square lattice; then only those lattice points which are separated from the origin by an even number of steps are related to the origin. Thus, for $i=0$, k must be $\pm 2m$, and, for $i=1$ or -1, k must be $\pm(2m-1)$, where $m=1,2,\cdots$.

*This work was partially supported by a National Science Foundation Grant No. DID71-04010 A02.

[1]For a review of this subject, see S. Stenholm, Phys. Rept. 6, 1 (1973), see also C. Cohen-Tanoudji, J. Dupont-Roc, and C. Fabre, J. Phys. B 6, L214 (1973); P. Hannaford, D. T. Pegg, and G. W. Series, J. Phys. B 6, L222 (1973); F. Ahmad and R. K. Bullough, J. Phys. B 7, L275 (1974); F. T. Hioe and J. H. Eberly, "Stimulated and Spontaneous Radiative Frequency Shifts of a Two-Level System," to appear in Phys. Rev. The original paper of F. Bloch and A. Siegert appeared in Phys. Rev. 57, 522 (1940) and the first significant extension of the Bloch—Siegert shifts to the sixth order term was made by J. H. Shirley, Phys. Rev. 138, B979 (1965) using a semiclassical theory.
[2]V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961); Rev. Mod. Phys. 34, 829 (1962).
[3]F. T. Hioe, J. Math. Phys. 15, 445 (1974).
[4]For a discussion of many properties of this type of determinants and other applications, see E. W. Montroll, Commun. Pure Appl. Math. 5, 415 (1952).
[5]M. Tavis and F. W. Cummings, Phys. Rev. 170, 379 (1968); E. T. Jaynes and F. W. Cummings, Proc. IEEE 51, 89 (1963).
[6]S. Swain, J. Phys. A 6, 1919 (1973).
[7]S. H. Autler and C. H. Townes, Phys. Rev. 100, 703 (1955).
[8]S. Schweber, Ann. Phys. 41, 205 (1967).
[9]S. Stenholm and W. E. Lamb, Jr., Phys. Rev. 181, 618 (1969); S. Stenholm, J. Phys. B 5, 878 (1972).
[10]See also P. Kusch and V. W. Hughes, in *Handbuch der Physik, Vol. XXXVII/I*, edited by S. Flügge (Springer-Verlag, Berlin, 1959), esp. p. 140. Y. K. Wang and W. E. Lamb, Jr., Phys. Rev. A 8, 866 (1973) and many other references given in the standard textbooks on finite difference equations.
[11]The terms up to λ^4 of this expansion were given by Swain (Ref. 6), but the coefficient of λ^4 was incorrectly given as $-n^2/4\omega^3$ there.
[12]J. H. Shirley, Phys. Rev. 138, B979 (1965).
[13]See also F. T. Hioe and J. H. Eberly, "Stimulated and Spontaneous Radiative Frequency Shifts of a Two-Level System," to appear in Phys. Rev.
[14]G. J. Lidstone, Proc. Edinburgh Math. Soc. 2, 16 (1929).
[15]J. M. Whittaker, Proc. London Math. Soc. 36, 451 (1934).
[16]D. V. Widder, Trans. Amer. Math. Soc. 51, 387 (1941).

The general relativistic fields of a charged rotating source*

Robert J. Finkelstein

Department of Physics, University of California, Los Angeles, California 90024
(Received 9 October 1974; revised manuscript received 13 January 1975)

The gravitational and electromagnetic fields of a charged, rotating source are obtained by an elementary algebraic method.

1. INTRODUCTION

It has been shown how the Kerr geometry may be obtained very simply as a complexified Schwarzschild geometry.[1,2] According to the work of Schiffer *et al.*, here referred to as (A), one may reach this result by first restricting the metric to the Kerr—Schild form; then the source-free field equations determine the null vector field l_μ that defines this metric. It turns out that l_μ is a simple functional of a harmonic function γ that in general may be complex. If γ is real, one has the Schwarzschild geometry; and if γ is complex, then one has the Kerr geometry. Its real part (α) may be identified with l_0^2 and may be regarded as a generalization of the Newtonian potential while its imaginary part (β) is proportional to the specific angular momentum of the source.

We shall here describe an elementary and straightforward extension of (A) to the case of a charged, rotating source. In this case, which corresponds to the Kerr—Newman geometry, the generalized Newtonian potential is still l_0^2, but it is not α any longer. Instead α and β are simply scalar potentials for the electric and magnetic fields, respectively. Thus β determines both the angular momentum density and the magnetic field.

To obtain this result one combines (A) with the requirement that the Kerr—Newman solution reduce to the Reissner—Nordstrom solution when the angular momentum vanishes. The analysis is then greatly simplified by the ansatz that $F_{\alpha\beta}l^\beta$, as well as l_α, is a null vector. Then it is only necessary to work with the Minkowski form of the Maxwell equations. Within this framework the role of the harmonic function γ in determining the electromagnetic field, as well as the gravitational field, is very clear.

2. KERR-SCHILD METRIC

Let l_α be a null vector whose contravariant components are defined with respect to the Minkowski metric $\eta^{\alpha\beta}$. Then

$$l^\alpha l_\alpha = 0 \tag{2.1}$$

where

$$l^\alpha = \eta^{\alpha\beta} l_\beta. \tag{2.2}$$

Define the matrix

$$L^\alpha{}_\beta = l^\alpha l_\beta. \tag{2.3}$$

Then

$$\mathrm{Tr}\mathbf{L} = 0 \tag{2.4}$$

and

$$\mathbf{L}^2 = 0. \tag{2.5}$$

The Kerr—Schild metric is

$$g_{\alpha\beta} = \eta_{\alpha\beta} - 2ml_\alpha l_\beta$$

$$= \eta_{\alpha\gamma}(\delta^\gamma{}_\beta - 2ml^\gamma l_\beta). \tag{2.6}$$

In matrix notation

$$\mathbf{g} = \eta(\mathbf{I} - 2m\mathbf{L}) \tag{2.7}$$

$$= \eta \exp(-2m\mathbf{L}). \tag{2.8}$$

Then

$$\det g = \det\eta \exp(-2m\,\mathrm{Tr}\mathbf{L}) \tag{2.9}$$

$$= \det\eta \tag{2.10}$$

by (2.4). Since $\sqrt{-g} = 1$, it follows that

$$\Gamma^\mu_{\sigma\mu} = \partial_\sigma \ln\sqrt{-g} = 0. \tag{2.11}$$

Also,

$$\mathbf{g}^{-1} = \exp(2m\mathbf{L})\eta^{-1}$$

$$= (1 + 2m\mathbf{L})\eta^{-1}.$$

Therefore,

$$g^{\alpha\beta} = \eta^{\alpha\beta} + 2ml^\alpha l^\beta \tag{2.12}$$

and

$$l^\alpha = g^{\alpha\beta} l_\beta. \tag{2.13}$$

The following relations may also be noted

$$l^\mu \partial_\alpha l_\mu = l^\mu l_{\mu|\alpha} = 0 \tag{2.14}$$

$$l^\mu{}_{|\mu} = \partial_\mu l^\mu \tag{2.15}$$

where the bar denotes the covariant derivative. Finally, by (2.11), the Einstein—Maxwell equations are

$$\partial_\sigma \Gamma^\sigma_{\alpha\beta} - \Gamma^\lambda_{\alpha\mu}\Gamma^\mu_{\beta\lambda} = -\kappa\theta_{\alpha\beta} \tag{2.16}$$

$$\partial_\sigma F^{\mu\sigma} = J^\mu. \tag{2.17}$$

These are coupled by

$$-\theta_{\alpha\beta} = \frac{1}{4\pi}\left(F_{\alpha\lambda}F_\beta{}^\lambda - \tfrac{1}{4} g_{\alpha\beta} F_{\mu\lambda}F^{\mu\lambda}\right) \tag{2.18}$$

corresponding to the signature $(+,-,-,-)$.

3. ROTATING UNCHARGED SOURCE

We discuss the Kerr solution first. Then one needs to consider only the source-free Einstein equations

$$R_{\alpha\beta} = 0. \tag{3.1}$$

Following reference (A) one may expand in powers of m. Then the order m equations are

$$\partial_\sigma(\eta^{\sigma\tau}[\alpha\beta,\tau]) = 0, \tag{3.2}$$

while the order m^3 equations are

$$(l^\sigma\partial_\sigma)l_\mu = -Al_\mu \tag{3.3}$$

where A is a scalar. The order m^4 equations are satisfied identically and the order m^2 equations are also satisfied if (3.2) and (3.3) are.[1] By combining (3.2) and (3.3) one finds

$$-\Box(l_\alpha l_\beta) = \partial_\alpha[(A+B)l_\beta] + \partial_\beta[(A+B)l_\alpha] \tag{3.4}$$

where

$$B = -l^\alpha_{\ |\alpha} = -\partial_\alpha l^\alpha. \tag{3.5}$$

Let us consider only stationary solutions

$$\frac{\partial l_\alpha}{\partial t} = 0. \tag{3.6}$$

Then the problem becomes three dimensional. Introduce the three-dimensional vector λ_i such that

$$l_\mu = (l_0, l_0\lambda_i) \tag{3.7a}$$

and

$$\lambda_i\lambda_i = 1. \tag{3.7b}$$

The six ij-components may be rewritten with the aid of the remaining four equations of (3.4) in the following form:

$$\partial_k\lambda_i\partial_k\lambda_j = p(\partial_i\lambda_j + \partial_j\lambda_i) \tag{3.8}$$

where

$$p = (A+B)/2l_0. \tag{3.8'}$$

By (3.7) one also has

$$\lambda_i\partial_k\lambda_i = 0 \tag{3.9a}$$

and by (3.8) and the preceding equation

$$\lambda_k\partial_k\lambda_i = 0. \tag{3.9b}$$

To solve the Eqs. (3.8) and (3.9) for the vector field λ_i we follow (A) by writing these equations in the following matrix form:

$$MM^T = p(M + M^T), \tag{3.10}$$

$$M\lambda = 0, \tag{3.11a}$$

$$M^T\lambda = 0, \tag{3.11b}$$

where

$$M_{ji} = \partial_i\lambda_j \tag{3.12}$$

where T means transpose.

4. LINEARIZATION OF THE MATRIX FIELD EQUATIONS

Equations (3.10) are nonlinear in M or in $\partial_i\lambda_k$. They may be linearized as follows. Let

$$R = 1 - M/p. \tag{4.1}$$

Then according to (3.10) and (3.11)

$$RR^T = 1, \tag{4.2}$$

$$R\lambda = R^T\lambda = \lambda. \tag{4.3}$$

By (4.2) and (4.3), R is a three-dimensional rotation matrix and λ is an eigenvector of R. R may therefore be written in the following way:

$$R = \exp(\theta\Lambda) \tag{4.4}$$

where Λ is a three rowed antisymmetric matrix

$$\Lambda^T = -\Lambda. \tag{4.5}$$

Let us put

$$\Lambda_{mn} = \epsilon_{mnp}\lambda_p. \tag{4.6}$$

Then R describes a rotation of θ about the unit vector λ_p. Define

$$P = -\Lambda^2. \tag{4.7}$$

Then P is the matrix

$$P_{mn} = \delta_{mn} - \lambda_m\lambda_n \tag{4.8}$$

and

$$P^2 = P, \tag{4.9}$$

$$\Lambda P = P\Lambda = \Lambda. \tag{4.10}$$

Therefore,

$$R = \exp(\theta\Lambda)$$

$$= 1 - P + P\cos\theta + \Lambda\sin\theta \tag{4.11}$$

and by (4.1)

$$M = p[1 - \exp(\theta\Lambda)] \tag{4.12}$$

$$= \alpha P + \beta\Lambda. \tag{4.13}$$

Also,

$$P\lambda = \Lambda\lambda = 0, \tag{4.14}$$

so that (3.11) or (4.3) is satisfied. Here α and β are two new functions related to p and θ by

$$\alpha = p(1 - \cos\theta), \tag{4.15a}$$

$$\beta = -p\sin\theta \tag{4.15b}$$

according to (4.11) and (4.13). Then (4.13) is the desired linearization of (3.10).

Let us rewrite (4.13) and (4.14) in terms of a new complex function γ and Hermitian matrix $\mathcal{M}$ defined as

$$\gamma = \alpha + i\beta, \tag{4.16a}$$

$$\mathcal{M} = P - i\Lambda. \tag{4.16b}$$

Then (4.13) and (4.14) become

$$M = \mathrm{Re}\gamma/\mathcal{M}, \tag{4.17}$$

$$\mathcal{M}\lambda = 0. \tag{4.18}$$

Notice the useful relations:

$$\Lambda/\mathcal{M} = i/\mathcal{M}, \tag{4.19}$$

$$P/\mathcal{M} = /\mathcal{M}, \tag{4.20}$$

$$\mathcal{M}^2 = 2/\mathcal{M}, \tag{4.21}$$

$$/\mathcal{M}/\mathcal{M}^T = /\mathcal{M}/\mathcal{M}^* = 0. \tag{4.22}$$

Equations (4.17) and (4.18) are equivalent to (3.8) and (3.9). They may be regarded as differential equations for the two independent components of the λ_i if the functions α and β are known, or alternatively as conditions

on the functions α and β if the vector function λ_i is known. In (A) it is shown how the vector functions λ_i may be eliminated altogether to give a simple differential equation for the complex function γ. We shall now show how the derivations of (A) may be slightly simplified by working directly with the complex function γ and the Hermitian matrix $\mathcal{M}$ rather than with the separate components of γ and the vector field λ_i.

5. DIFFERENTIAL EQUATIONS FOR THE COMPLEX POTENTIAL (γ)

The following relations are easily verified:

$$\partial_s M_{ik} = \partial_k M_{is}, \tag{5.1}$$

$$\lambda_s \partial_s \mathcal{M}_{ik} = 0, \tag{5.2}$$

$$\lambda_s \partial_k \mathcal{M}_{is} = -\gamma \mathcal{M}_{ik}, \tag{5.3}$$

$$\partial_k \mathcal{M}_{ks} = -2\gamma\lambda_s. \tag{5.4}$$

Contract (5.1) with λ_s. Then by (4.17) and (4.18) and by (5.2) and (5.3) one finds

$$\mathrm{Re}[\lambda_s \partial_s \gamma + \gamma^2]\mathcal{M}_{ik} = 0$$

and therefore

$$\lambda_s \partial_s \gamma + \gamma^2 = 0. \tag{5.5}$$

To obtain a second differential equation for γ, antisymmetrize (5.1) with respect to i and s. Then write in terms of $\mathcal{M}$ by (4.17) and contract with respect to i and k. One then obtains

$$\mathrm{Re}[2\partial_s \gamma - \gamma \partial_k \mathcal{M}_{ks} - \partial_s \gamma \mathcal{M}_{ks}] = 0 \tag{5.6}$$

Then by (5.4) and (5.5) one finds

$$\mathrm{Re}(\mathcal{M}_{sk} \partial_k \gamma) = 0. \tag{5.7}$$

By (4.19) one finally obtains

$$\mathcal{M}_{sk} \partial_k \gamma = 0. \tag{5.8}$$

Equation (5.8) is a concise statement of the following two vector equations of (A):

$$\nabla\alpha = (\beta^2 - \alpha^2)\lambda - \nabla\beta \times \lambda, \tag{A.4.20}$$

$$\nabla\beta = -2\alpha\beta\lambda + \nabla\alpha \times \lambda. \tag{A.4.22}$$

Note also the following useful relation

$$\mathcal{M}_{sk} \partial_k \gamma^* = 2P_{sk} \partial_k \gamma^*. \tag{5.8*}$$

Equations (5.5) and (5.8) together determine the two components of γ. Moreover, by differentiation of (5.8) one finds

$$\partial_s \mathcal{M}_{sk} \partial_k \gamma + \mathcal{M}_{sk} \partial_s \partial_k \gamma = 0.$$

This is a second order equation for γ. Using (5.4) and (5.5) one finds

$$\nabla^2 \gamma = 0. \tag{5.9}$$

Therefore, γ, and both α and β, are harmonic. Denote tha reciprocal of γ by ω. It satisfies (5.8) and also the eikonal equation:

$$\mathcal{M}_{sk} \partial_k \omega = 0, \tag{5.10}$$

$$(\partial_s \omega)(\partial_s \omega) = 1. \tag{5.11}$$

These same equations hold for both $\nabla\omega$ and λ. ω is not harmonic but satisfies

$$\nabla^2 \omega = 2\gamma \tag{5.12}$$

and therefore

$$\nabla^2 \cdot \nabla^2 \omega = 0. \tag{5.13}$$

Note also

$$\lambda_s \partial_s \omega = 1. \tag{5.14}$$

Equations (5.10) and (5.14) imply

$$\lambda_s = \partial_s \omega - i\Lambda_{sk} \partial_k \omega. \tag{5.15}$$

The solution of (5.15) for λ_s is

$$\lambda_s = \frac{\partial_s \omega + \partial_s \omega^* - i\epsilon_{smn}(\partial_m \omega)(\partial_n \omega^*)}{1 + (\partial_m \omega)(\partial_m \omega^*)}. \tag{5.16}$$

6. THE KERR SOLUTION

So far only six of the ten field equations (3.1) have been satisfied. The remaining four equations may be written in the following form[1]:

$$R_{00} = 0 \rightarrow \nabla^2 l_0^2 = 0, \tag{6.1}$$

$$R_{0s} = 0 \rightarrow \nabla^2(l_0^2 \lambda_s) = \partial_s[(A+B)l_0]. \tag{6.2}$$

These relations have in fact already been used in simplifying the six relations $R_{ij} = 0$ to give (3.8). To make the present note self-contained we shall now review the argument of (A) that the conditions (6.1) and (6.2) may be satisfied by the choice

$$l_0^2 = \alpha. \tag{6.3}$$

One may check that this choice is satisfactory by first noting that (6.3) satisfies (6.1) since α is harmonic. To check (6.2) observe that

$$\nabla^2 \lambda_j = \partial_k \mathrm{Re}\gamma \mathcal{M}_{jk} = -2|\gamma|^2 \lambda_j \tag{6.4}$$

by (5.4) and (5.8). In a similar way one finds by (5.8)

$$2\partial_s \gamma \partial_s \lambda_j = \mathcal{M}^*_{js} \partial_s |\gamma|^2, \tag{6.5}$$

so that

$$\nabla^2(\gamma\lambda_j) = -2|\gamma|^2 \gamma\lambda_j + \mathcal{M}^*_{js} \partial_s |\gamma|^2. \tag{6.6}$$

This result turns out to be useful not only in checking (6.2) but also in checking the field equations when there is an electromagnetic source (paragraph 10). The real part of (6.6) is by (5.5)

$$\nabla^2(\alpha\lambda_j) = \partial_j |\gamma|^2. \tag{6.7}$$

It now follows from (6.7) that (6.2) is also satisfied if

$$|\gamma|^2 = (A+B)l_0. \tag{6.8}$$

One already knows, however, by (3.8′) and (4.15) that

$$p = |\gamma|^2/2\alpha = (A+B)/2l_0. \tag{6.9}$$

Then (6.6), and therefore(6.2), is satisfied if $l_0^2 = \alpha$. At this point the field equations (3.1) are solved by (5.16) and (6.3) in terms of the still unknown function γ.

One obtains the Kerr solution by choosing the complex harmonic function γ to be

$$\gamma = [x^2 + y^2 + (z - ia)^2]^{-1/2}. \tag{6.10}$$

Then

$$\omega = [x^2 + y^2 + (z - ia)^2]^{1/2}. \tag{6.11}$$

Let

$$\omega = \rho + i\sigma. \tag{6.12}$$

Then

$$\rho^2 - \sigma^2 = r^2 - a^2, \tag{6.13a}$$

$$\rho\sigma = -za, \tag{6.13b}$$

$$\rho^4 - (r^2 - a^2)\rho^2 - z^2a^2 = 0, \tag{6.14}$$

where

$$r^2 = x^2 + y^2 + z^2. \tag{6.15}$$

The connection between (ρ, σ) and (α, β) is an inversion:

$$\alpha = \frac{\rho}{\rho^2 + \sigma^2}, \quad \beta = -\frac{\sigma}{\rho^2 + \sigma^2}, \tag{6.16}$$

$$\rho = \frac{\alpha}{\alpha^2 + \beta^2}, \quad \sigma = -\frac{\beta}{\alpha^2 + \beta^2}, \tag{6.17}$$

$$(\alpha^2 + \beta^2)(\rho^2 + \sigma^2) = 1. \tag{6.18}$$

If $a = 0$, one has the Schwarzschild solution. Then $\rho = r$ and λ_k is the unit vector in the radial direction; the two parts of γ are $\alpha = 1/r$ and $\beta = 0$.

In terms of ρ, which may be regarded as a rescaled distance, one finds, in the general case

$$l_0^2 = \alpha(\rho) = \rho^3/(\rho^4 + a^2z^2), \tag{6.19}$$

$$\lambda_1 = (\rho x + ay)/(\rho^2 + a^2), \tag{6.20a}$$

$$\lambda_2 = (\rho y - ax)/(\rho^2 + a^2), \tag{6.20b}$$

$$\lambda_3 = z/\rho. \tag{6.20c}$$

These equations completely determine the initial metric (2.6) and the Kerr geometry.

In general the following relations hold:

$$x\lambda_1 + y\lambda_2 + z\lambda_3 = \rho, \tag{6.21}$$

$$\nabla^2\rho = 2\alpha, \tag{6.22}$$

$$\frac{\partial\rho}{\partial x} = \alpha x, \tag{6.23a}$$

$$\frac{\partial\rho}{\partial y} = \alpha y, \tag{6.23b}$$

$$\frac{\partial\rho}{\partial z} = \alpha z + a\beta. \tag{6.23c}$$

These equations are all obtained in a straightforward way. For example, Eq. (6.22) follows from (5.12). These equations all have familiar limits when $a \to 0$ and $\rho \to r$. When $a \neq 0$, we may interpret ρ as a generalization of the radial coordinate r, and σ may be regarded as a generalization of the azimuthal variable: $\cos\theta = z/r$.

Each of the two potentials (α, β) depends on both the radial and azimuthal variable, as shown in (6.16). In addition, one has the Cauchy–Riemann equations[3]:

$$\frac{\partial\alpha}{\partial\rho} = \frac{\partial\beta}{\partial\sigma}, \tag{6.24a}$$

$$\frac{\partial\alpha}{\partial\sigma} = -\frac{\partial\beta}{\partial\rho}. \tag{6.24b}$$

7. THE KERR-NEWMAN GEOMETRY

By (2.6)

$$g_{00} = \eta_{00} - 2ml_0^2. \tag{7.1}$$

Then $2ml_0^2$ gives the Newtonian potential in the weak field limit and of course satisfies Laplace's equation in this limit. We have just seen, however, that $l_0^2 = \alpha$ is correct for all field strengths and therefore l_0^2 always satisfies Laplace's equation as long as the electromagnetic field vanishes. Therefore, l_0^2 may be regarded as a generalization of the Newtonian potential.

We now seek the correct generalization of l_0^2 in the case that the source is charged as well as rotating. Let us assume

$$l_0^2 = \psi(Q, a, \rho)\alpha(\rho) \tag{7.2}$$

where Q is the charged and $\psi(Q, a, \rho)$ is a new function to be determined. The case already considered corresponds to

$$\psi(0, a, \rho) = 1. \tag{7.3}$$

Guided by the Reissner–Nordstrom solution we try

$$l_0^2 = \left(1 - \frac{Q^2}{2mr}\right)\alpha(\rho) \tag{7.4}$$

when $a = 0$. When $a \neq 0$, ρ is the natural generalization of r. [see equations (6.21)–(6.23).] Therefore let us try

$$l_0^2 = \left(1 - \frac{Q^2}{2m\rho}\right)\alpha(\rho) \tag{7.5}$$

when $a \neq 0$. It turns out that this ansatz for l_0^2 together with the unchanged expression for the three-dimensional vector λ_i [Eqs. (6.20a)–(6.20c)] describe the exactly correct generalization of the Kerr metric to the problem of a charged rotating source. This metric defines the Kerr–Newman geometry. To support this statement one must of course show that the complete Einstein–Maxwell system (2.16)–(2.18) is satisfied.

8. THE ELECTROMAGNETIC VECTOR POTENTIAL

In order to investigate the complete field equations, it is necessary to solve the inhomogeneous Maxwell equations (2.17). The homogeneous set of Maxwell equations is satisfied identically by the usual representation of the covariant components of the electromagnetic field in terms of the vector potential:

$$F_{\beta\alpha} = \partial_\alpha A_\beta - \partial_\beta A_\alpha \tag{8.1}$$

The contravariant components are then related to $F_{\alpha\beta}$ by

$$F^{\alpha\beta} = g^{\alpha\lambda}g^{\beta\mu}F_{\lambda\mu} \tag{8.2}$$

$$= (\eta^{\alpha\lambda} + 2ml^\alpha l^\lambda)(\eta^{\beta\mu} + 2ml^\beta l^\mu)F_{\lambda\mu} \tag{8.3}$$

$$= \eta^{\alpha\lambda}\eta^{\beta\mu}F_{\lambda\mu} + 2mF_{\lambda\mu}(l^\alpha l^\lambda\eta^{\beta\mu} + \eta^{\alpha\lambda}l^\beta l^\mu). \tag{8.3'}$$

With our ansatz, namely,

$$l_\mu = (l_0, l_0\lambda_k), \tag{8.4}$$

$$l_0^2 = \left(1 - \frac{Q^2}{2m\rho}\right)\alpha(\rho), \tag{8.5}$$

it follows that the contravariant differ from the covariant components by terms of higher order in Q^2 unless

$$F_{\lambda\mu}(l^\lambda l^\alpha \eta^{\mu\beta} + \eta^{\lambda\alpha} l^\mu l^\beta) = 0. \tag{8.6}$$

According to (2.17), however, $\partial_\beta F^{\alpha\beta}$ must be linear in Q, and (8.6) therefore suggests itself as a possible ansatz. We shall then assume (8.6) and later verify that the $F_{\lambda\mu}$ actually do satisfy (8.6).

Therefore, by (8.3′), it is necessary only to solve the equations

$$\eta^{\alpha\lambda}\eta^{\beta\mu}\partial_\beta(\partial_\lambda A_\mu - \partial_\mu A_\lambda) = 0, \tag{8.7}$$

except for the δ-function source. One may therefore follow the usual analysis appropriate to flat space.

With the Lorentz gauge and the assumption of stationarity, we then have

$$\mathrm{div}\mathbf{A} = 0, \tag{8.8}$$

$$\nabla^2 A_0 = 0, \tag{8.9}$$

$$\nabla^2 \mathbf{A} = 0. \tag{8.10}$$

We are now assuming that l_μ does not enter the generally covariant Maxwell equations. On the other hand, the solution of (8.9) and (8.10) determines the source $\theta_{\alpha\beta}$ of the field equations that l_μ must satisfy, so that

$$R_{\alpha\beta}[l_\mu] = -8\pi\theta_{\alpha\beta}[A_\mu].$$

Therefore, the harmonic 4-vector A_μ must be a functional of l_μ, and hence of α, β, and λ_k. Consequently, to find the harmonic scalar and vector potentials shown in (8.9) and (8.10) it is natural to examine the two harmonic functions α and β. According to (6.13) and (6.16) in the limit of large ρ

$$\alpha = \frac{\rho^3}{\rho^4 + a^2 z^2} \to \frac{1}{\rho}, \tag{8.11}$$

$$\beta = \frac{za\rho}{\rho^4 + a^2 z^2} \to \frac{az}{\rho^3}. \tag{8.12}$$

Therefore, in the limit of large ρ it is clear that $Q\alpha$ behaves like the Coulomb potential while $Q\beta$ goes into the scalar potential of the dipole $(0, 0, Qa)$. This limiting behavior suggests that we put

$$A_0 = Q\alpha \tag{8.13}$$

and that we look for a vector potential that gives the same dipole field as the scalar potential $Q\beta$.

For this purpose note that β may be rewritten by (6.23c) and the nonlimiting forms of (8.11) and (8.12) in the following way:

$$\beta = \frac{a}{\rho^2 + a^2}\frac{\partial\rho}{\partial z}. \tag{8.14}$$

If one now defines a new scalar potential ϕ by

$$\frac{d\phi}{d\rho} = \frac{1}{\rho^2 + a^2}, \tag{8.15}$$

then

$$\beta = a\frac{d\phi}{d\rho}\frac{\partial\rho}{\partial z}. \tag{8.16}$$

This new scalar potential is

$$\phi = \frac{1}{a}\tan^{-1}\frac{\rho}{a}. \tag{8.17}$$

At large distances

$$\phi \to \frac{1}{a}\left(\frac{\pi}{2} - \frac{a}{\rho} + \cdots\right), \quad a/\rho \ll 1.$$

This function is also harmonic,

$$\nabla^2\phi = 0, \tag{8.18}$$

as one may show by direct computation with the aid of (6.14), (6.23), (6.22), and (8.15) which implies

$$\nabla\phi = \frac{1}{\rho^2 + a^2}\nabla\rho. \tag{8.19}$$

Next introduce the dipole moment vector

$$\mu = (0, 0, Qa). \tag{8.20}$$

Then

$$Q\beta = \mu\nabla\phi. \tag{8.21}$$

To find the equivalent vector potential we may make use of the familiar vector identity

$$\begin{aligned}\mathrm{curl}(\mu\times\nabla\phi) &= \mu\nabla^2\phi - (\mu\nabla)(\nabla\phi)\\ &= -(\mu\nabla)(\nabla\phi)\end{aligned} \tag{8.22}$$

since ϕ is harmonic. Then

$$\mathrm{curl}_k(\mu\times\nabla\phi) = -(\mu_s\partial_s)(\partial_k\phi) \tag{8.23}$$

$$= -\partial_k(\mu_s\partial_s\phi)$$

$$= -\partial_k(Q\beta). \tag{8.24}$$

It follows that the vector potential

$$\mathbf{A} = \mu\times\nabla\phi \tag{8.25}$$

is equivalent to the scalar potential $Q\beta$ in the sense of (8.24). In addition, $\mathbf{A}$ is also harmonic since ϕ is. Therefore, (8.25) is a possible solution of (8.10) and its curl yields a dipole field at large distances according to (8.24) and (8.12). The components of $\mathbf{A}$ are

$$A_x = -Qa\frac{\partial\phi}{\partial y} = -\frac{Qa}{\rho^2 + a^2}(\alpha y), \tag{8.26a}$$

$$A_y = Qa\frac{\partial\phi}{\partial x} = \frac{Qa}{\rho^2 + a^2}(\alpha x), \tag{8.26b}$$

$$A_z = 0. \tag{8.26c}$$

The complete four potential is then

$$A_\mu = Q\alpha\left(1, \frac{ay}{\rho^2 + a^2}, -\frac{ax}{\rho^2 + a^2}, 0\right). \tag{8.27}$$

The form of A_μ just given together with the earlier prescription (8.4) for l_μ completely describe the Kerr–Newman fields.[4] To verify that these forms are satisfactory, one must of course satisfy (2.16): the left side of (2.16) depends on l_μ alone and the right side depends on A_μ only. We may also verify that (8.6) is satisfied.

9. THE ELECTROMAGNETIC FIELD AND ENERGY-MOMENTUM TENSOR

The preceding work provides simple representations of the exact electric and magnetic fields. The electric field is derivable from the potential $Q\alpha$ by

$$E_k = -\partial_k(Q\alpha) \tag{9.1}$$

and the magnetic field is also derivable from a scalar potential $Q\beta$; since

$$H_k = -\partial_k(Q\beta) \tag{9.2}$$

by (8.24). Then

$$E_k + iH_k = -\partial_k(Q\gamma). \tag{9.3}$$

One may therefore obtain the electromagnetic field of a charged spinning source from the Coulomb field of a nonrotating source by simply shifting the origin in the complex z-direction and using (9.3) as noted by Newman.[5]

Then also

$$Q^2(\nabla\gamma)^2 = \mathbf{E}^2 - \mathbf{H}^2 + 2i\mathbf{EH}, \tag{9.4}$$

so that the real and imaginary parts of $Q^2(\nabla\gamma)^2$ give the scalar and pseudoscalar invariants.

The energy density, to the lowest order in m, is

$$Q^2\overset{\circ}{\theta}_{00} = (E^2 + H^2)/8\pi \tag{9.5a}$$

$$= (Q^2/8\pi)\nabla\gamma\nabla\gamma^* \tag{9.5b}$$

$$= (Q^2/16\pi)\nabla^2(\gamma\gamma^*). \tag{9.5c}$$

The Poynting vector is, to the same order,

$$\theta_k = -Q^2\overset{\circ}{\theta}_{0k} = \frac{1}{4\pi}\epsilon_{klm}E_lH_m \tag{9.6a}$$

or

$$\theta = (iQ^2/8\pi)\nabla\gamma\times\nabla\gamma^*. \tag{9.6b}$$

The Maxwell stresses are, for example when $k \neq l$,

$$-Q^2\theta_{kl} = (1/4\pi)(E_kE_l + H_kH_l) \tag{9.7a}$$

$$= (Q^2/4\pi)\,\mathrm{Re}(\partial_k\gamma)(\partial_l\gamma^*). \tag{9.7b}$$

Note that the basic relations (5.5) and (5.8) become

$$\lambda(\mathbf{E} + i\mathbf{H}) = Q\gamma^2 \tag{9.8}$$

and

$$\not{\eth}(\mathbf{E} + i\mathbf{H}) = 0. \tag{9.9}$$

In component form (9.8) is

$$\lambda\mathbf{E} = Q(\alpha^2 - \beta^2), \tag{9.10a}$$

$$\lambda\mathbf{H} = Q(2\alpha\beta), \tag{9.10b}$$

while (9.9) may also be written in component form with the aid of (9.10) as follows:

$$\mathbf{E} = Q(\alpha^2 - \beta^2)\lambda - \mathbf{H}\times\lambda, \tag{9.11a}$$

$$\mathbf{H} = Q(2\alpha\beta)\lambda + \mathbf{E}\times\lambda. \tag{9.11b}$$

The Poynting vector (θ) may be related to λ with the aid of (9.11) or directly by

$$\lambda_k = \frac{-2\,\mathrm{Re}\gamma^2\partial_k\gamma^* - i\epsilon_{kmn}\partial_m\gamma\partial_n\gamma^*}{(\gamma\gamma^*)^2 + (\partial_m\gamma)(\partial_m\gamma^*)}, \tag{9.12}$$

which in turn follows from (5.16). By (9.12) and (5.5)

$$-8\pi\theta\lambda = \mathbf{E}^2 + \mathbf{H}^2 - Q^2(\gamma\gamma^*)^2. \tag{9.13}$$

The following 4-vector is also important here:

$$n_\mu = F_{\mu\lambda}l^\lambda. \tag{9.14}$$

By definition it is orthogonal to l_μ:

$$n_\mu l^\mu = 0; \tag{9.15}$$

and by (9.10a) it is also proportional to l_μ:

$$n_\mu = (\mathbf{E}\lambda)l_\mu. \tag{9.16}$$

Since l is a null vector, n is also:

$$n_\mu n^\mu = 0. \tag{9.17}$$

In terms of n_μ the ansatz (8.6) becomes

$$n^\alpha l^\beta - n^\beta l^\alpha = 0. \tag{9.18}$$

By (9.16) it is clear that (9.18) is satisfied. By (9.17) or (9.18) one may also show that

$$F = F_{\alpha\beta}F^{\alpha\beta} = \eta^{\alpha\lambda}\eta^{\beta\mu}F_{\alpha\beta}F_{\lambda\mu}. \tag{9.19}$$

This invariant may then be computed directly from (9.11):

$$F = 2(H^2 - E^2)$$

$$= 2[(\mathbf{H}\lambda)^2 - (\mathbf{E}\lambda)^2] \tag{9.20a}$$

$$= 2Q^2[(2\alpha\beta)^2 - (\alpha^2 - \beta^2)^2]. \tag{9.20b}$$

10. FIELD EQUATIONS

The gravitational field equations are

$$R_{\alpha\beta} = -8\pi\theta_{\alpha\beta} \tag{10.1}$$

where

$$R_{\alpha\beta} = \overset{1}{R}_{\alpha\beta}m + \overset{2}{R}_{\alpha\beta}m^2 + \overset{3}{R}_{\alpha\beta}m^3 + \overset{4}{R}_{\alpha\beta}m^4 \tag{10.1'}$$

and

$$\theta_{\alpha\beta} = Q^2(\overset{\circ}{\theta}_{\alpha\beta} + \overset{1}{\theta}_{\alpha\beta}m + \overset{2}{\theta}_{\alpha\beta}m^2). \tag{10.1''}$$

Here $\overset{4}{R}_{\alpha\beta}$ vanishes identically. Then

$$\overset{1}{R}_{\alpha\beta} = -16\pi\epsilon\overset{\circ}{\theta}_{\alpha\beta}, \tag{10.2a}$$

$$\overset{2}{R}_{\alpha\beta} = -16\pi\epsilon\overset{1}{\theta}_{\alpha\beta}, \tag{10.2b}$$

$$\overset{3}{R}_{\alpha\beta} = -16\pi\epsilon\overset{2}{\theta}_{\alpha\beta}, \tag{10.2c}$$

where the parameter

$$\epsilon = Q^2/2m \tag{10.3}$$

is to be regarded as fixed.

Let us examine (10.2c) first. One finds

$$\overset{3}{R}_{\alpha\beta} = -4[(l_\alpha V_\beta + l_\beta V_\alpha)(Vl) + V^2 l_\alpha l_\beta] = 8\epsilon(n^2/Q^2)l_\alpha l_\beta \tag{10.4a}$$

$$= 0 \tag{10.4b}$$

where

$$V_\alpha = l^\sigma\partial_\sigma l_\alpha. \tag{10.5}$$

It follows from (10.4a) that

$$Vl = 0 \tag{10.6}$$

and from (10.4b) that

$$V^2 = 0. \tag{10.7}$$

Just as in the uncharged case V and l are two orthogonal null vectors.[6] They are therefore proportional:

$$V_\alpha = -Cl_\alpha \tag{10.8}$$

where C corresponds to the scalar A of the uncharged source. Let us also introduce the scalar D that corresponds to the B used earlier:

$$D = -\partial_\mu l^\mu. \tag{10.9}$$

Using the form of l_μ that holds when the source is charged, one finds

$$Cl_0 = \tfrac{1}{2}\gamma\gamma^* - \alpha l_0^2, \tag{10.10}$$

$$Dl_0 = \tfrac{1}{2}\gamma\gamma^* + \alpha l_0^2. \tag{10.11}$$

Proceeding now as in the uncharged case, one obtains

$$\overset{1}{R}_{\alpha\beta} = \nabla^2(l_\alpha l_\beta) - \partial_\alpha[(C+D)l_\beta] - \partial_\beta[(C+D)l_\alpha], \tag{10.12}$$

$$\overset{2}{R}_{\alpha\beta} = 2l_\alpha l_\beta\{2\partial_\mu(Cl^\mu) + \eta^{\lambda\mu}(\partial_\lambda l^\sigma)(\partial_\sigma l_\mu) - \eta^{\lambda\mu}(\partial_\lambda l^\sigma)(\partial_\mu l_\sigma) - C^2\}. \tag{10.13}$$

The energy–momentum tensor of the electromagnetic field contains terms of order m^0 and m.

$$-4\pi Q^2 \overset{\circ}{\theta}_{\alpha\beta} = F_{\alpha\lambda}F_{\beta\mu}\eta^{\lambda\mu} - \tfrac{1}{4}\eta_{\alpha\beta}F, \tag{10.14}$$

$$-4\pi Q^2 \overset{1}{\theta}_{\alpha\beta} = 2(n_\alpha n_\beta + \tfrac{1}{4}l_\alpha l_\beta F) \tag{10.15}$$

$$= [(E\lambda)^2 + (H\lambda)^2]l_\alpha l_\beta,$$

or

$$-4\pi \overset{1}{\theta}_{\alpha\beta} = (\gamma\gamma^*)^2 l_\alpha l_\beta. \tag{10.16}$$

To check the first order field equations note that

$$\overset{1}{R}_{00} = \nabla^2(l_0^2) = \nabla^2(\alpha - \epsilon\gamma\gamma^*)$$

$$= -\epsilon\nabla^2(\gamma\gamma^*) = -16\pi\epsilon\overset{\circ}{\theta}_{00}. \tag{10.17}$$

This result is in agreement with (10.2a) as required.

For $\overset{1}{R}_{0k}$ one finds

$$\nabla^2(l_0^2\lambda_k) - \partial_k[(C+D)l_0] = \nabla^2[(\alpha - \epsilon\gamma\gamma^*)\lambda_k] - \partial_k(\gamma\gamma^*)$$

$$= -\epsilon\nabla^2(\gamma\gamma^*\lambda_k)$$

by (10.10), (10.11), and (6.7). Then the $0k$-component of (10.2a) becomes

$$\nabla^2(\gamma\gamma^*\lambda_k) = 16\pi\overset{\circ}{\theta}_{0k}. \tag{10.18}$$

This equation may be checked by computing the left side with the aid of (6.5), (6.6), and (9.12) and comparing with (9.6). Alternatively, (10.18) expresses the Poynting vector very simply in terms of λ_k.

Finally, the jk-part of (10.2a) is

$$\nabla^2(l_0^2\lambda_j\lambda_k) - \partial_k[(C+D)l_0\lambda_j] - \partial_j[(C+D)l_0\lambda_k] = -16\pi\epsilon\overset{\circ}{\theta}_{jk}$$

or

$$\nabla^2(\alpha\lambda_j\lambda_k) - \partial_k[\gamma\gamma^*\lambda_j] - \partial_j[\gamma\gamma^*\lambda_k]$$

$$= -16\pi\epsilon\overset{\circ}{\theta}_{jk} + \epsilon\nabla^2(\gamma\gamma^*\lambda_j\lambda_k). \tag{10.19}$$

If $Q = 0$, then $\theta_{jk} = 0$ and the left-hand side of (10.19) vanishes. The λ_k are not altered, however, by the presence of charge and therefore the left-hand side always vanishes. We then obtain a simple expression for the Maxwell stresses in terms of λ_k, namely,

$$16\pi\overset{\circ}{\theta}_{jk} = \nabla^2(\gamma\gamma^*\lambda_j\lambda_k) \tag{10.20}$$

provided that the corresponding field equations are correct. Alternatively, to check the field equations one must check (10.20). For this purpose one may use the following evaluation of the right-hand side:

$$\nabla^2(\gamma\gamma^*\lambda_j\lambda_k) = 2\,\mathrm{Re}\gamma^2(\lambda_j\partial_k + \lambda_k\partial_j)\gamma^* + 2(\gamma\gamma^*)^2\delta_{jk}$$

$$- i(\lambda_j\epsilon_{kmn} + \lambda_k\epsilon_{jmn})\partial_m\gamma\partial_n\gamma^* \tag{10.21}$$

To obtain (10.21) one may use (6.5) and (6.6). The right-hand side of Eq. (10.21) may now be reduced to forms such as (9.7b). To check the second order field equations one may reduce $\overset{2}{R}_{\alpha\beta}$ to the following form:

$$\overset{2}{R}_{\alpha\beta} = [D^2 - C^2 - \partial_k[l_k(D+C)] - l_0^2(\partial_k\lambda_j)(\partial_k\lambda_j)]2l_\alpha l_\beta. \tag{10.22}$$

The first order equations require

$$\nabla^2(l_0^2\lambda_k\lambda_k) - 2\partial_k[(C+D)l_k] = -16\pi\epsilon\overset{\circ}{\theta}_{kk} \tag{10.23a}$$

or

$$-2\partial_k[(C+D)l_k] = -16\pi\epsilon\overset{\circ}{\theta}_{00} - \nabla^2(l_0) = 0 \tag{10.23b}$$

since $\eta^{\mu\lambda}\overset{\circ}{\theta}_{\mu\lambda}$ vanishes. Then

$$\overset{2}{R}_{\alpha\beta} = [D^2 - C^2 - 2l_0^2\gamma\gamma^*]2l_\alpha l_\beta \tag{10.24}$$

$$= 4\epsilon(\gamma\gamma^*)^2 l_\alpha l_\beta$$

$$= -16\pi\epsilon\overset{1}{\theta}_{\alpha\beta} \tag{10.25}$$

by (10.10), (10.11), and (10.16) in agreement with (10.2b).

*Work supported in part by the National Science Foundation.

[1]M.M. Schiffer, R.J. Adler, J. Mark, and C. Sheffield, J. Math. Phys. **14**, 52 (1973).

[2]E.T. Newman and A.I. Janis, J. Math. Phys. **6**, 915 (1965).

[3]Then both α and β are also harmonic in the plane of the "polar" variables (ρ,σ) as well as in the space of the rectangular variables (x, y, z).

[4]G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969). Signs of A_μ correspond to (8.1).

[5]E.T. Newman, J. Math. Phys. **14**, 102 (1973).

[6]We are concluding from (10.4a) that V is null because n is. This argument may be reversed by first showing that V is null in the following way. Let

$$V_0 = -Cl_0.$$

Then compute

$$V_k = l^\sigma\partial_\sigma l_k$$

$$= l^\sigma\partial_\sigma(l_0\lambda_k)$$

and find

$$V_k = -Cl_k$$

and therefore

$$V_\mu = -Cl_\mu.$$

It follows that V is null because l is. One may then reverse the argument of (10.4a) to conclude that n is also a null vector.

Petrie matrices and generalized inverses

S. Aronowitz and B. E. Eichinger

Department of Chemistry, University of Washington, Seattle, Washington 98195
(Received 20 December 1974)

The connection between Petrie matrices and a special group of generalized inverses deriving from an incidence matrix is established. These matrices are encountered in the theory of the excluded volume effect in polymers, but have wide applicability to other problems. As an example of one such application we present a greatly simplified derivation of Lagrange's theorem, which relates the inertial tensor of a mechanical system to the distances between all pairs of particles.

I. INTRODUCTION

Petrie matrices recently have been shown to be useful in graph theory with specific application to polymer problems.[1,2] These matrices are comprised of only zeros and ones, with the stipulation that any ones appearing in a column must occur consecutively. In another recent paper one of us obtained exact expressions for the distribution function of the radius of gyration of the Gaussian model of a polymer with excluded volume.[3] The method used there relied upon a special group of generalized inverses. The main purpose of this communication is to demonstrate the connection between this particular group of generalized inverses and Petrie matrices. An interesting aspect of the matrices considered here is that generalized inverses are intimately involved with statistical theory.[4,5,6] One application of generalized inverses, to the derivation of Lagrange's theorem, is made to illustrate the characteristics of the matrices.

II. GENERALIZED INVERSES

The unique generalized inverse $\mathbf{B}^{(-1)}$, also referred to as the Moore–Penrose inverse, of the singular or rectangular matrix B, satisfies the following conditions:

$$\begin{aligned} &\text{(i)} && \mathbf{B}\mathbf{B}^{(-1)}\mathbf{B}=\mathbf{B},\\ &\text{(ii)} && \mathbf{B}^{(-1)}\mathbf{B}\mathbf{B}^{(-1)}=\mathbf{B}^{(-1)},\\ &\text{(iii)} && \mathbf{B}\mathbf{B}^{(-1)}=(\mathbf{B}\mathbf{B}^{(-1)})',\\ &\text{(iv)} && \mathbf{B}^{(-1)}\mathbf{B}=(\mathbf{B}^{(-1)}\mathbf{B})'. \end{aligned} \tag{2.1}$$

Superscript primes indicate the transpose. Other types of generalized inverses satisfy only select combinations of the four conditions in Eq. (2.1). We follow Pringle and Rayner[6] and designate such generalized inverses as one-, two-, three-condition generalized inverses, which respectively satisfy

(i) or (ii): one-condition,

(i) + (ii): two-condition,

[(i) or (ii)] + (iii) + (iv): three-condition.

Let C be an $N\times N(N-1)/2$ matrix whose nonzero elements c_{ij} are given by

$$c_{i,i+l}=1,\quad c_{i+k,i+l}=-1, \tag{2.2}$$

where $i=1,\ldots,N-k$, $k=1,\ldots,N-1$, and $l=N(k-1)-k(k-1)/2=(N-k/2)(k-1)$. The matrix C is the incidence matrix of the directed star graph on N nodes.[7] This particular representation is chosen because it greatly simplifies the algebraic arguments presented here. Any matrix equivalent to C by permutations and/or sign changes of rows and columns would equally suffice. Let U be an $N\times N$ matrix, all elements of which are equal to 1. Let E denote the identity matrix of order N. If A is a matrix satisfying the conditions

$$\mathbf{A}\mathbf{U}=\mathbf{U}\mathbf{A}=\mathbf{O}, \tag{2.3}$$

and if $\mathbf{A}^{(-1)}$ is a generalized inverse of A satisfying

$$\begin{aligned} \mathbf{A}\mathbf{A}^{(-1)}&=\mathbf{E}+\mathbf{X}\mathbf{U},\\ \mathbf{A}^{(-1)}\mathbf{A}&=\mathbf{E}+\mathbf{U}\mathbf{Y}, \end{aligned} \tag{2.4}$$

where the matrices X and Y are arbitrary, then the matrix

$$\mathbf{G}^{(-1)}=\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{C},$$

is the unique generalized inverse of the matrix

$$\mathbf{G}=N^{-2}\mathbf{C}'\mathbf{A}\mathbf{C},$$

and both are members of an Abelian group. That $\mathbf{G}^{(-1)}$ is unique can be shown upon considering the products $\mathbf{C}'\mathbf{A}\mathbf{A}^{(-1)}\mathbf{C}$ and $\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{A}\mathbf{C}$. A column of C contains only two nonzero elements, namely, +1 and −1. The product UC is therefore the null matrix. Similarly, C′U is null. We arrive at the result, with reference to Eq. (2.4), that

$$\mathbf{C}'\mathbf{A}\mathbf{A}^{(-1)}\mathbf{C}=\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{A}\mathbf{C}=\mathbf{C}'\mathbf{C}. \tag{2.5}$$

It is shown in Appendix C that

$$\mathbf{C}\mathbf{C}'=N\mathbf{E}-\mathbf{U}. \tag{2.6}$$

The products $\mathbf{G}\mathbf{G}^{(-1)}$ and $\mathbf{G}^{(-1)}\mathbf{G}$ are evaluated using Eqs. (2.5) and (2.6) as

$$\begin{aligned} \mathbf{G}\mathbf{G}^{(-1)}&=(N^{-2}\mathbf{C}'\mathbf{A}\mathbf{C})(\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{C})=N^{-2}\mathbf{C}\,\mathbf{A}(\mathbf{C}\mathbf{C}')\mathbf{A}^{(-1)}\mathbf{C}'\\ &=N^{-1}\mathbf{C}'\mathbf{A}\mathbf{A}^{(-1)}\mathbf{C}=N^{-1}\mathbf{C}'\mathbf{C} \end{aligned} \tag{2.7a}$$

and

$$\begin{aligned} \mathbf{G}^{(-1)}\mathbf{G}&=N^{-2}\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{C}\mathbf{C}'\mathbf{A}\mathbf{C}\\ &=N^{-1}\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{A}\mathbf{C}=N^{-1}\mathbf{C}'\mathbf{C}. \end{aligned} \tag{2.7b}$$

From Eqs. (2.7a, b) it is easily seen that

$$\begin{aligned} \mathbf{G}^{(-1)}\mathbf{G}&=(\mathbf{G}^{(-1)}\mathbf{G})',\\ \mathbf{G}\mathbf{G}^{(-1)}&=(\mathbf{G}\mathbf{G}^{(-1)})'. \end{aligned} \tag{2.8}$$

Equations (2.6) and (2.7a, b) also yield

$$\begin{aligned} \mathbf{G}^{(-1)}\mathbf{G}\mathbf{G}^{(-1)}&=N^{-1}\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{C}\mathbf{C}'\mathbf{C}=\mathbf{C}'\mathbf{A}^{(-1)}\mathbf{C}\\ &=\mathbf{G}^{(-1)}, \end{aligned} \tag{2.9a}$$

$$\mathbf{G}\mathbf{G}^{(-1)}\mathbf{G} = N^{-3}\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{A}\mathbf{C} = N^{-2}\mathbf{C}'\mathbf{A}\mathbf{C}$$
$$= \mathbf{G}. \tag{2.9b}$$

Equations (2.8) and (2.9a, b) show that $\mathbf{G}^{(-1)}$ fulfills all the conditions of Eq. (2.1) and is therefore the unique generalized inverse of G.

We now show that G and $\mathbf{G}^{(-1)}$ are members of an Abelian group. First note that the matrix $N^{-1}\mathbf{C}'\mathbf{C}$, arising in Eqs. (2.7a, b) and used in Eqs. (2.9a, b), acts as a generalized identity. That $N^{-1}\mathbf{C}'\mathbf{C}$ acts as a generalized identity can be proved by considering the products $N^{-1}\mathbf{C}'\mathbf{C}\mathbf{G}^{(m)}$ and $\mathbf{G}^{(m)}$ $(N^{-1}\mathbf{C}'\mathbf{C})$, where

$$\mathbf{G}^{(m)} = N^{-m-1}\mathbf{C}'\mathbf{A}^{(m)}\mathbf{C},$$
$$N^{-1}\mathbf{C}'\mathbf{C}\mathbf{G}^{(m)} = N^{-m-2}\mathbf{C}'\mathbf{C}\mathbf{C}'\mathbf{A}^{(m)}\mathbf{C}$$
$$= N^{-m-1}\mathbf{C}'\mathbf{A}^{(m)}\mathbf{C}$$
$$= \mathbf{G}^{(m)},$$
$$\mathbf{G}^{(m)}(N^{-1}\mathbf{C}'\mathbf{C}) = N^{-m-2}\mathbf{C}'\mathbf{A}^{(m)}\mathbf{C}\mathbf{C}'\mathbf{C}$$
$$= N^{-m-1}\mathbf{C}'\mathbf{A}^{(m)}\mathbf{C}$$
$$= \mathbf{G}^{(m)}, \tag{2.10}$$

where use has been made of Eq. (2.6). It has been proven elsewhere[3] and is readily seen that

$$\mathbf{G}^{(k)}\mathbf{G}^{(l)} = \mathbf{G}^{(k+l)}. \tag{2.11}$$

Equations (2.10) and (2.11) suffice to establish group properties.

The immediate application of this section to a physical problem results when A is taken as the Rouse–Zimm matrix for a linear chain polymer. We show in the next section and in Appendix A that if **A** is the Rouse–Zimm matrix, three different types of $\mathbf{A}^{(-1)}$ can be constructed which satisfy the requirements in Eq. (2.4). Hence $\mathbf{G}^{(-1)}$ can be constructed from three types of generalized inverses of **A**. (The one-condition inverse is not considered.) The first type, which is shown in the next section to be a two-condition generalized inverse, leads directly to a representation for $\mathbf{G}^{(-1)}$ in terms of a Petrie matrix. The second type, constructed in Appendix A, is a three-condition generalized inverse, and the third is the unique generalized inverse.

The importance of $\mathbf{G}^{(-1)}$ is that the partition function and the mean square radius of gyration can be expressed as a particular expansion involving determinants of principal minors of $\mathbf{G}^{(-1)}$ of various orders.[3] Though the treatment in the following section is confined to the linear chain, the arguments are sufficiently general that they can be extended to other chain forms.

III. THE CONNECTION WITH PETRIE MATRICES

A. A particular representation of $\mathbf{G}^{(-1)}$

We approach the connection between $\mathbf{G}^{(-1)}$ and a Petrie matrix in a somewhat circumspect fashion. Let **H** denote the $(N-1)\times N$ Petrie matrix whose elements are

$$h_{ij} = 1, \quad j \le i,$$
$$h_{ij} = 0, \quad j > i. \tag{3.1}$$

Let **W** denote the product **H'H**, with elements w_{jk} given by

$$w_{jk} = \sum_{i=1}^{N-1} h_{ij}h_{ik}. \tag{3.2}$$

For $j \le k$, nonzero terms of the product are encountered only for $i \ge k$, and therefore

$$w_{jk} = N - k, \quad j \le k \tag{3.3}$$
$$w_{jk} = N - j, \quad j \ge k$$

where the last equation is obtained by symmetry.

We will now show that **W** is the two-condition generalized inverse of the Rouse–Zimm matrix **A** defined by Eq. (A1) of Appendix A. Consider the product $\mathbf{F} = \mathbf{A}\mathbf{W}$.

The three disinct cases which arise in evaluating matrix elements f_{lk} of **F** are:

(i) $l = 1$,

$$f_{1k} = \sum_i (w_{ik}\delta_{1i} - w_{ik}\delta_{1,i-1})$$
$$= w_{1k} - w_{2k}$$
$$= \delta_{1k}; \tag{3.4a}$$

(ii) $1 < l < N$,

$$f_{lk} = \sum_i (2w_{ik}\delta_{li} - w_{ik}\delta_{l,i+1} - w_{ik}\delta_{l,i-1})$$
$$= 2w_{lk} - w_{l-1,k} - w_{l+1,k};$$

when $l < k$,

$$f_{lk} = 2(N-k) - (N-k) - (N-k) = 0;$$

when $l = k$,

$$f_{lk} = 2(N-k) - (N-k) - (N-k-1) = 1;$$

and for $l > k$, symmetry demands

$$f_{lk} = 0; \tag{3.4b}$$

(iii) $l = N$,

$$f_{Nk} = \sum (w_{ik}\delta_{Ni} - w_{ik}\delta_{N,i+1}) \tag{3.4c}$$
$$= w_{Nk} - w_{N-1,k} = 0 - 1 = -1$$

when $k < N$. In more succinct notation

$$f_{lk} = \delta_{lk}, \quad l < N,$$
$$f_{Nk} = -1 + \delta_{Nk}.$$

Hence we may write F as

$$\mathbf{F} = \mathbf{E} - \mathbf{X}\mathbf{U}, \tag{3.5a}$$

where the matrix **X** possesses a single nonzero (i, j) element equal to unity in the (N, N) location. The matrices **A**, **W**, **E**, **U**, **X** are symmetric; therefore,

$$\mathbf{F}' = \mathbf{W}\mathbf{A} = \mathbf{E} - \mathbf{U}\mathbf{X}. \tag{3.5b}$$

The Rouse–Zimm matrix **A** satisfies

$$\mathbf{U}\mathbf{A} = \mathbf{A}\mathbf{U} = \mathbf{O}.$$

Use of Eqs. (3.5) allows one to show that **W** satisfies conditions (i) and (ii) of Eq. (2.1); however, since $\mathbf{F} \neq \mathbf{F}'$, conditions (iii) and (iv) are not fulfilled. The matrix **W** is a two-condition generalized inverse of **A**. Furthermore, Eqs. (3.5) establish that **W** satisfies the requirements of Eq. (2.4) and it may be concluded that

$$\mathbf{G}^{(-1)} = \mathbf{C}'\mathbf{W}\mathbf{C}$$

is the unique generalized inverse of **G**, where

$$\mathbf{G}=N^{-2}\mathbf{C}'\mathbf{A}\mathbf{C}.$$

The three-condition and unique generalized inverses of **A**, constructed in Appendix A, also fulfill Eq. (2.4); thus $\mathbf{G}^{(-1)}$ may be obtained by at least three different routes.

B. $\mathbf{G}^{(-1)}$ as a product of Petrie matrices

The connection between $\mathbf{G}^{(-1)}$ and a particular Petrie matrix may be established by evaluation of the product $\mathbf{D}=\mathbf{H}\mathbf{C}$, where **H** is defined by Eq. (3.1) and **C** is defined by Eq. (2.2). We find, upon referring to Eqs. (2.2), (3.1), that

$$\begin{aligned} d_{i,j+l} &= h_{ij}c_{j,j+l}+h_{i,j+k}c_{j+k,j+l} \\ &= h_{ij}-h_{i,j+k}, \end{aligned} \tag{3.6}$$

where $i=1,\ldots,N-1$, $j=1,\ldots,N-k$, $k=1,\ldots,N-1$, $l=(N-k/2)(k-1)$. Three classes of terms are encountered on evaluating Eq. (3.6). They are given by

(i) $j>i$,

$$d_{i,j+l}=0; \tag{3.7a}$$

(ii) $j\leq i<j+k$,

$$d_{i,j+l}=1; \tag{3.7b}$$

(iii) $j+k\leq i$,

$$d_{i,j+l}=0. \tag{3.7c}$$

The matrix **D** is thus one whose columns contain only zeros and ones. If **D** is to be a Petrie matrix, then the ones in each column must occur consecutively. The condition in Eq. (3.7b) shows that the ones in each column do occur consecutively since the $d_{i,j+l}=1$ for $j+k-1\geq i\geq j$ and vanish outside this interval. This observation allows us to cast Eq. (3.7b) into the form

$$d_{j+m-1,j+l}=1,$$

where $j=1,\ldots,N-k$, $k=1,\ldots,N-1$, $l=(N-k/2)(k-1)$, $m=1,\ldots,k$. All other elements of the $j+l$ column are equal to zero. We conclude that **D** is a Petrie matrix. Now,

$$\mathbf{G}^{(-1)}=\mathbf{C}'\mathbf{W}\mathbf{C}=\mathbf{C}'\mathbf{H}'\mathbf{H}\mathbf{C};$$

therefore,

$$\mathbf{G}^{(-1)}=\mathbf{D}'\mathbf{D}. \tag{3.10}$$

An example is provided in Appendix B which illuminates the structure of these various matrices.

The rows of the matrix **H**, c.f. Eq. (3.1), may be thought of as representing the possible contact pair types, i.e., nearest-neighbor, next-nearest-neighbor and so on. The columns of the matrix **D**, on the other hand, represent all the possible combinations of pair contacts. The correspondence between the columns of **D** and the polymer chain is somewhat subtle. It rests upon the interpretation of a column of **D** as a representation of a series of catenated steps or consecutive bonds in the chain. Figure 1 is offered as an example of the preceeding discussion.

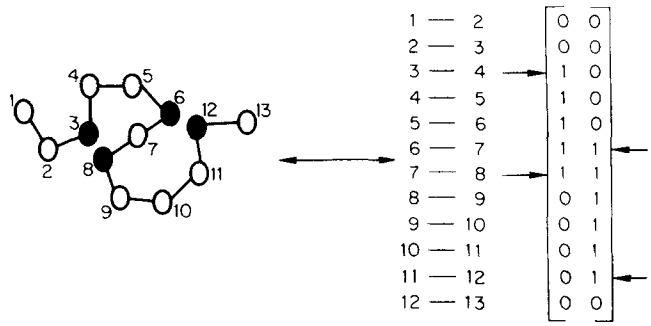


FIG. 1. One-to-one mapping of chains into Petrie matrices: The chain on the left has two contact pairs, (3,8) and (6,12). The labeling 1 – 2, 2 – 3,..., 12 – 13, indicates catenated steps or consecutive bonds on the chain.

IV. LAGRANGE'S THEOREM

In the most general form, Lagrange's theorem relates the inertial tensor of a mechanical system to the distances between all pairs of particles. In the standard form, the theorem relates the square radius of gyration of the mechanical system of identical particles, defined by

$$s^2=N^{-1}\sum_{i=1}^{N}\mathbf{r}_i^2, \tag{4.1}$$

where $\mathbf{r}_i$ is the vector from the center of gravity to the ith particle, to the interparticle distances $|\mathbf{r}_k-\mathbf{r}_l|$ by[8]

$$s^2=N^{-2}\sum_{k<l}\sum(\mathbf{r}_k-\mathbf{r}_l)^2. \tag{4.2}$$

That this transformation is correct is easily seen from the matrix representation. Let $\mathbf{r}'=[\mathbf{r}_1',\mathbf{r}_2',\ldots,\mathbf{r}_N']$ be a $1\times 3N$ row vector pertinent to a particular configuration. Then

$$s^2=N^{-1}\mathbf{r}'\mathbf{r}=N^{-1}\mathbf{r}'(\mathbf{E}_N\otimes\mathbf{E}_3)\mathbf{r}.$$

Substitution of Eq. (2.6) for $\mathbf{E}_N$ gives

$$\begin{aligned} s^2 &= N^{-2}\mathbf{r}'[(\mathbf{C}\mathbf{C}'+\mathbf{U})\otimes\mathbf{E}_3]\mathbf{r} \\ &= N^{-2}\mathbf{r}'[(\mathbf{C}\otimes\mathbf{E}_3)(\mathbf{C}'\otimes\mathbf{E}_3)]\mathbf{r} \end{aligned} \tag{4.3}$$

since $(\mathbf{U}\otimes\mathbf{E}_3)\mathbf{r}=\mathbf{O}$ by definition of the center of gravity. The last rendition, Eq. (4.3), is easily seen to be identical to Eq. (4.2).

The theorem is readily generalized to encompass the inertial tensor for a system of nonidentical particles. Let the vector from the center of mass to the ith particle in a mechanical system in a particular configuration, be denoted $\mathbf{r}_i$ as before. A constraint on the $\mathbf{r}_i$ arises from the definition of the center of mass according to

$$\sum_{i=1}^{N}m_i\mathbf{r}_i=\mathbf{O}, \tag{4.4}$$

where m_i denotes the mass of the ith particle. Now define the $N\times 3$ matrix **R** as

$$\mathbf{R}=[\mathbf{X}\,\mathbf{Y}\,\mathbf{Z}]$$

where **X**, **Y**, and **Z** are $N\times 1$ columns consisting of x, y, and z components of all the $\mathbf{r}_i$, respectively. (It is readily seen that the calculation is not restricted to three dimensional space.) Let **M** denote the $N\times N$ diagonal matrix with elements

$$m_{ij} = m_i \delta_{ij}$$

equal to the masses of the particles. Upon forming the sum of pre- and post-products of Eq. (2.6) with the matrix M, one obtains

$$\mathbf{M} = (2N)^{-1}[(\mathbf{MCC'} + \mathbf{CC'M}) + (\mathbf{MU} + \mathbf{UM})]. \tag{4.5}$$

The inertial tensor

$$S^{\alpha\beta} = \sum_{i=1}^{N} m_i x_i^\alpha x_i^\beta \Big/ \sum_{i=1}^{N} m_i$$

may be defined in matrix form as

$$\mathbf{S} = [\mathrm{Tr}(\mathbf{M})]^{-1} \mathbf{R'MR}. \tag{4.6}$$

Use of Eqs. (4.4) and (4.5) allows this equation to be written as

$$\mathbf{S} = [2N\,\mathrm{Tr}(\mathbf{M})]^{-1} \mathbf{R'(MCC' + CC'M)R} \tag{4.7}$$

since $\mathbf{UMR} = \mathbf{O}$. Upon reverting to indicial notation this is seen to be

$$S^{\alpha\beta} = \sum_{1 \le i \le j \le N} \sum (m_i x_i^\alpha - m_j x_j^\alpha)(x_i^\beta - x_j^\beta) + (x_i^\alpha - x_j^\alpha) \times (m_i x_i^\beta - m_j x_j^\beta) \Big/ 2N \sum_{i=1}^{N} m_i.$$

If the masses are identical, Eq. (4.7) reduces to Eq. (4.2) upon taking the trace of S.

V. DISCUSSION

The aim of this paper has been to establish a connection between two matrices which are intimately involved in the problem of the linear Gaussian chain with excluded volume. The connection between Petrie matrices and a special generalized inverse occurs at two levels. The generalized inverse $\mathbf{G}^{(-1)}$ is equal to the product of a Petrie matrix (D) and its transpose. This Petrie matrix is directly related to the specific pair interactions in the polymer chain. Another Petrie matrix (H) used in the construction of the generalized inverse contains information referring to the number of steps between interacting beads in the polymer. One is lead to speculate that matrices of the form $\mathbf{C'A}^{(-1)}\mathbf{C}$, where A and $\mathbf{A}^{(-1)}$ satisfy Eqs. (2.3) and (2.4), can always be written as the symmetric product of Petrie matrices for a connected chain. If this is so, then molecules of any arbitrary connectivity may be treated with Petrie matrices as well as with C.

Section IV is representative of a direct application of generalized matrices to a nontrivial problem. The usual derivation of Lagrange's theorem involves laborious algebra and demands careful choice of reference systems. That the derivation is so simple with use of C is illustrative of the comprehensive nature of this matrix.

ACKNOWLEDGMENT

The work was supported by the National Science Foundation.

APPENDIX A

The elements of the $N \times N$ Rouse–Zimm matrix A for a chain of N beads are given by

$$\begin{aligned} a_{ij} &= 2\delta_{ij} - \delta_{i,j+1} - \delta_{i,j-1}, \quad 1 < i < N, \\ a_{1j} &= \delta_{1j} - \delta_{1,j-1}, \\ a_{Nj} &= \delta_{Nj} - \delta_{N,j+1}. \end{aligned} \tag{A1}$$

Let V be a generalized inverse of A which satisfies Eq. (2.4) with

$$\mathbf{X} = \mathbf{Y} = -N^{-1}\mathbf{E},$$

where E is the identity matrix of order N. The elements of V are

$$\begin{aligned} v_{ij} &= (2N)^{-1}[i(i-1) + (N-j)(N-j+1)], \quad i \le j, \\ v_{ij} &= (2N)^{-1}[j(j-1) + (N-i)(N-i+1)], \quad i \ge j, \end{aligned} \tag{A2}$$

and V is a three-condition generalized inverse of A satisfying

$\mathbf{AV} = \mathbf{VA} = \mathbf{E} - N^{-1}\mathbf{U}$, which is to be proved. First, it is seen from Eq. (A2) that $\mathbf{V'} = \mathbf{V}$, i.e., V is symmetric. Consider the product AV. The elements of AV take the following forms:

$$(\mathbf{AV})_{ij} = -v_{i-1,j} + 2v_{ij} - v_{i+1,j}, \quad 1 < i < N,$$

which:

(a) when $i < j$,

$$(\mathbf{AV})_{ij} = -N^{-1};$$

(b) when $i = j$,

$$(\mathbf{AV})_{ij} = 1 - N^{-1};$$

(c) when $i > j$,

$$(\mathbf{AV})_{ij} = -N^{-1}.$$

(A3)

For elements of the first row of the product

$$(\mathbf{AV})_{1j} = v_{1j} - v_{2j},$$

which:

(a) when $j = 1$,

$$(\mathbf{AV})_{11} = 1 - N^{-1};$$

(b) when $j \neq 1$,

$$(\mathbf{AV})_{1j} = -N^{-1}.$$

(A4)

For the last row we have

$$(\mathbf{AV})_{Nj} = -v_{N-1,j} + v_{Nj},$$

which:

(a) when $j = N$,

$$(\mathbf{AV})_{NN} = 1 - N^{-1};$$

(b) when $j \neq N$,

$$(\mathbf{AV})_{Nj} = -N^{-1}.$$

(A5)

The explicit evaluation of AV along with the fact that A and V are symmetric matrices lead to the desired result that

$$\mathbf{AV} = \mathbf{VA} = \mathbf{E} - N^{-1}\mathbf{U}, \tag{A6}$$

where it is recalled that U is a matrix whose elements are all equal to 1. Equation (A6) also shows that V satisfies conditions (iii) and (iv) of Eq. (2.1). The

matrix **A** obeys Eq. (2.3). This is sufficient for the purposes of showing that **V** satisfies condition (i) of Eq. (2.1). Since **V** has only positive elements, the product **VU** cannot vanish. This fact suffices in showing that **V** cannot satisfy (ii) of Eq. (2.1). Therefore the matrix **V** is a three-condition generalized inverse of **A**.

Consider now the matrix **R** constructed from the diagonal matrix Λ_0 of nonvanishing eigenvalues of **A** according to

$$\mathbf{R}=\mathbf{T}_0'\,\Lambda_0^{-1}\mathbf{T}_0 \tag{A7}$$

where $\Lambda=\mathbf{TAT}'$; $\mathbf{T}^{-1}=\mathbf{T}'$; Λ_0 is an $(N-1)\times(N-1)$ diagonal matrix obtained from Λ by removal of the zero eigenvalue, while $\mathbf{T}_0$ is obtained from **T** by removing the constant column and is of dimension $(N-1)\times N$. It has been shown[3] that

$$\mathbf{AR}=(\mathbf{AR})'=\mathbf{RA}=\mathbf{E}-N^{-1}\mathbf{U}, \tag{A8}$$

which, together with the fact that **A** obeys Eq. (2.3), establishes that **R** satisfies conditions (i), (iii), and (iv) of Eq. (2.1). It also has been shown[3] that

$$\mathbf{T}_0\mathbf{U}=\mathbf{O},$$

where **O** is the null matrix. This relationship and Eq. (A8) suffice to establish that **R** satisfies condition (ii) of Eq. (2.1). Hence, **R** is the unique generalized inverse of **A**. Furthermore, Eq. (A8) shows that **R** satisfies Eq. (2.4) and thus is a third route to $\mathbf{G}^{(-1)}$.

APPENDIX B

Matrices for $N=5$ are given here to serve as examples. Reference to Eqs. (2.2), (3.1), (3.4a, b) of the text and Eq. (A2) of Appendix A allows us to write explicit expressions for **C**, **H**, **W**, **V**, respectively, as

$$\mathbf{C}=\begin{bmatrix}1&0&0&0&1&0&0&1&0&1\\-1&1&0&0&0&1&0&0&1&0\\0&-1&1&0&-1&0&1&0&0&0\\0&0&-1&1&0&-1&0&-1&0&0\\0&0&0&-1&0&0&-1&0&-1&-1\end{bmatrix}, \tag{B1a}$$

$$\mathbf{H}=\begin{bmatrix}1&0&0&0&0\\1&1&0&0&0\\1&1&1&0&0\\1&1&1&1&0\end{bmatrix}, \tag{B1b}$$

$$\mathbf{W}=\begin{bmatrix}4&3&2&1&0\\3&3&2&1&0\\2&2&2&1&0\\1&1&1&1&0\\0&0&0&0&0\end{bmatrix}, \tag{B1c}$$

$$\mathbf{V}=(1/5)\begin{bmatrix}10&6&3&1&0\\6&7&4&2&1\\3&4&5&4&3\\1&2&4&7&6\\0&1&3&6&10\end{bmatrix}. \tag{B1d}$$

Recalling that

$$\mathbf{G}^{(-1)}=\mathbf{C}'\mathbf{WC}=\mathbf{C}'\mathbf{VC},$$

one finds[9] that

$$\mathbf{G}^{(-1)}=\begin{bmatrix}1&0&0&0&1&0&0&1&0&1\\0&1&0&0&1&1&0&1&1&1\\0&0&1&0&0&1&1&1&1&1\\0&0&0&1&0&0&1&0&1&1\\1&1&0&0&2&1&0&2&1&2\\0&1&1&0&1&2&1&2&2&2\\0&0&1&1&0&1&2&1&2&2\\1&1&1&0&2&2&1&3&2&3\\0&1&1&1&1&2&2&2&3&3\\1&1&1&1&2&2&2&3&3&4\end{bmatrix}.$$

Alternatively, the expressions in Eqs. (B1a), (B1b) may be used to represent $\mathbf{G}^{(-1)}$ in another fashion. Since

$$\mathbf{D}=\mathbf{HC},$$

one finds that

$$\mathbf{D}=\begin{bmatrix}1&0&0&0&1&0&0&1&0&1\\0&1&0&0&1&1&0&1&1&1\\0&0&1&0&0&1&1&1&1&1\\0&0&0&1&0&0&1&0&1&1\end{bmatrix}.$$

It was shown in Sec. III of the text that

$$\mathbf{G}^{(-1)}=\mathbf{D}'\mathbf{D}.$$

Each column or combination of columns in **D** is a Petrie matrix. The matrix **D** represents the ten possible combinations of pair contacts, which conveys essentially the same information as **C**.

APPENDIX C

Given that **C** is the $N\times N(N-1)/2$ matrix defined in Eq. (2.2), then

$$\mathbf{CC}'=N\mathbf{E}-\mathbf{U}, \tag{C1}$$

where **C**′ is the transpose of **C**, **E** is the identity matrix of order N and **U** is a matrix of dimension $N\times N$ whose elements are all equal to unity. The relation in Eq. (C1) will be derived by explicit evaluation of **CC**′.

Consider the sum,

$$\sum_{m=1}^{N(N-1)/2} c_{im}c'_{mr}, \tag{C2}$$

when $i<r$. The only value of m for which c_{im} differs from zero, for a given i, occurs when $m=i+l$, where l is defined in Eq. (2.2). The only values of r for which $c'_{i+l,r}$ differs from zero occur for $r=i$ or $r=i+k$. By stipulation $i<r$, and

$$\sum_m c_{im}c'_{mr}=c_{i,i+l}c'_{i+l,i+k},$$
$$=(1)(-1)=-1. \tag{C3}$$

The matrix product $\mathbf{CC}'$ is symmetric, and therefore all off-diagonal elements are -1.

Consider now the case when $i=r$. The only values of c_{im} which differ from zero are those for

$$m=i+l, \tag{C4a}$$

or

$$i=j+k,$$
$$m=j+l. \tag{C4b}$$

The values of k which are permitted, for a given value of i in Eq. (C4a), are those for $1\leq k\leq N-i$ as Eq. (2.2) reveals. Equation (C4a) excludes the case $i=N$, since k must be greater than zero. Similarly, $i=1$ is excluded in Eq. (C4b) since $k\geq 1$ and $j\geq 1$. The values that j can assume in Eq. (C4b), for a given value of i, are $1\leq j\leq i-1$, since k must be less than i. Thus three cases arise in the calculation of the diagonal elements of $\mathbf{CC}'$.

(i) $i=1$,

$$\sum_m c_{1m}c'_{m1}=\sum_{k=1}^{N-1} c_{1,1+l}c'_{1+l,1},$$
$$=\sum_{k=1}^{N-1}(1)(1)=N-1; \tag{C5a}$$

(ii) $1<i<N$,

$$\sum_m c_{im}c'_{mi}=\sum_{k=1}^{N-i} c_{i,i+l}c'_{i+l,i}+\sum_{m=1}^{i-1} c_{i,j+l}c'_{j+l,i},$$
$$=N-i+\sum_{j=1}^{i-1}(-1)(-1)=N-1; \tag{C5b}$$

(iii) $i=N$, $j+k=N$,

$$\sum_m c_{Nm}c'_{mN}=\sum_{j=1}^{N-1} c_{j+k,j+l}c'_{j+l,j+k}$$
$$=\sum_{j=1}^{N-1}(-1)(-1)=N-1. \tag{C5c}$$

Thus, the diagonal elements are all equal to $N-1$. Upon combining the results in Eqs. (C3) and (C5), and using the definition of $\mathbf{U}$, we find that $\mathbf{CC}'=N\mathbf{E}-\mathbf{U}$. An exchange of rows or columns in $\mathbf{C}$ with a corresponding change in columns or rows in $\mathbf{C}'$ leaves the product $\mathbf{CC}'$ unchanged. If $\mathbf{P}$ represents a permutation matrix obeying

$$\mathbf{PP}'=\mathbf{P}'\mathbf{P}=\mathbf{E},$$

then

$$\mathbf{PCC}'\mathbf{P}'=\mathbf{P}(N\mathbf{E}-\mathbf{U})\mathbf{P}'=\mathbf{P}'\mathbf{CC}'\mathbf{P}=N\mathbf{E}-\mathbf{U}.$$

In closing, we note that $N^{-1}\mathbf{C}'$ is the unique generalized inverse of $\mathbf{C}$, since

$$(N^{-1}\mathbf{C}')\mathbf{C}(N^{-1}\mathbf{C}')=(N^{-1}\mathbf{C}')(\mathbf{E}-N^{-1}\mathbf{U})=N^{-1}\mathbf{C}',$$
$$\mathbf{C}(N^{-1}\mathbf{C}')\mathbf{C}=(\mathbf{E}-N^{-1}\mathbf{U})\mathbf{C}=\mathbf{C},$$
$$\mathbf{CC}'=(\mathbf{CC}')',$$
$$\mathbf{C}'\mathbf{C}=(\mathbf{C}'\mathbf{C})'.$$

The matrix $\mathbf{C}$ is associated with a directed graph. Its unique generalized inverse $N^{-1}\mathbf{C}'$ is also intimately connected with that graph.[10]

[1] M. Gordon and W.T. Tutte, Proc. Camb. Phil. Soc. 75, 155 (1974).
[2] M. Gordon, S.B. Ross-Murphy, and H. Suzuki, preprint.
[3] B.E. Eichinger, J. Chem. Phys. 59, 5787 (1973).
[4] C.R. Rao and S.K. Mitra, in *Generalized Inverse of Matrices and its Applications* (Wiley, New York, 1971).
[5] T. L. Boullion and P. L. Odell, in *Generalized Inverse Matrices* (Wiley, New York, 1971).
[6] R.M. Pringle and A.A. Rayner, "Generalized Inverse Matrices," Griffin's Statistical Monographs and Courses, No. 28 (Hafner, New York, 1971).
[7] Ref. 5, pp. 66—69.
[8] P.J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969), Appendix A, pp. 383—384.
[9] Ref. 3, Appendix C gives $\mathbf{V}$ (called $\mathbf{A}^{(-1)}$) and $\mathbf{G}^{(-1)}$ for the linear chain with $N=5$.
[10] See Ref. 5, p. 68. Upon recognizing $\mathbf{C}$, given by Eq. (2.2), as a *connected* incidence matrix and using the result, derived in Appendix C, that $N^{-1}\mathbf{C}'$ is the unique generalized inverse of $\mathbf{C}$, one may invoke Theorem 13 on p. 67. The theorem also leads to the result that $\mathbf{CC}'=N\mathbf{E}-\mathbf{U}$.

On the stability of equilibrium states of finite classical systems

Joel L. Lebowitz*

Belfer Graduate School of Science, Yeshiva University, New York, New York

Michael Aizemann†

Courant Institute, New York University, New York, New York

Sheldon Goldstein‡

Institute for Advanced Study, Princeton, New Jersey

(Received 2 January 1975)

The state of a system is characterized, in statistical mechanics, by a measure ω on Γ, the phase space of the system (i.e., by an ensemble). To represent an equilibrium state, the measure must be stationary under the time evolution induced by the systems Hamiltonian $H(x)$, $x \in \Gamma$. An example of such a measure is $\omega(dx) = f(H)dx$; dx is the Liouville (Lebesgue) measure and $f(H(x))$ is the ensemble density. For "nonergodic" systems there are also other stationary measures with ensemble densities, e.g., for integrable dynamical systems the density can be a function of any of the constants of the motion. We show, however, that the requirement that the equilibrium measure have a certain type of "stability" singles out, in the typical case, densities which depend only on H.

1. INTRODUCTION

The macroscopic description of a physical system is assumed in statistical mechanics to be given by a probability measure ω on the phase space Γ of the system[1,2]: If A is a region of the phase space, $A \subset \Gamma$, then $\omega(A)$ is the probability that the phase point of the system will be found in A. Equivalently, $\omega(A)$ is the fraction of systems in the ensemble in the region A. To describe a system in equilibrium the measure must be stationary under the time evolution. Since the energy (Hamiltonian) H of a finite system of particles is always a constant of the motion, a measure given by a function of the energy (times Lebesgue measure), $\omega(A) = \int_A f(H)\,dx$, will always be stationary. Conversely, if the time evolution is ergodic on all the energy surfaces S_E [specified by $H(x) = E$] equipped with their natural microcanonical measures, then every stationary measure ω given by a density ρ, i. e., $\omega(A) = \int_A \rho(x)\,dx$ so that $\omega(A) = 0$ if $\int_A dx = 0$, will be of this form.[2,3] If, in the other extreme, the system is integrable,[2,4] so that there are in addition to H other "smooth" constants of the motion, then there will also be stationary states whose densities are functions of these constants of the motion.

Consider, for example, an ideal gas consisting of n particles moving in a unit box with periodic boundary conditions—the unit torus T^3. The phase space of this system is $T^{3n} \times R^{3n}$ and the time evolution T_t, is given by

$$T_t(q_1, \ldots q_{3n}, p_1, \ldots p_{3n}) = T_t(\mathbf{q}, \mathbf{p}) = (\mathbf{q} + \mathbf{p}t, \mathbf{p}),$$

where $(\mathbf{q}, \mathbf{p}) = x \in T^{3n} \times R^{3n} = \Gamma$ and the addition is modulo 1. This evolution comes from the ideal gas Hamiltonian

$$H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{3n} \tfrac{1}{2} p_i^2.$$

The p_i, as well as H, are constants of the motion and thus any ensemble density which is a function of $\mathbf{p}$ only will be stationary under the time evolution.

Nevertheless, the equilibrium properties of finite systems, even those which are not ergodic, are usually assumed to be determined by a density which is a function of H only.[1,2] In this note we shall not discuss any specific form of this function but consider justifications of the assumption that the ensemble density is a function of H only even when there are also other constants of the motion present.[5] (It is known that all reasonable functions of H lead to the same results for local quantities in the thermodynamic limit.[1])

There may, of course, exist singular stationary measures (not given by a density) which are not "constant" on energy surfaces even if the system is ergodic on all such surfaces.[2] It may be argued, however, that these measures, which assign a finite probability to the system being found in a region of the phase space A which has zero (Liouville—Lebesgue) volume, i. e., $\int_A d\mathbf{q}\,d\mathbf{p} = 0$, should be irrelevant for explaining experimentally observed behavior. Experimental results depend on reproducibility and it seems plausible to assume that there will be a vanishing "probability" for "preparing" a physical system in such a region.[1a] We may then regard as physically reasonable only those measures which are absolutely continuous, i. e., have a density, with respect to Lebesgue measure. We shall adopt this attitude here and only worry about the justification of assuming $\rho(x)$ to depend on H only. [The microcanonical ensemble, at a fixed energy E, is itself singular with respect to Lebesgue measure $d\mathbf{q}\,d\mathbf{p}$; it may, however, be regarded[1,2] as the limit when $\Delta E \to 0$, of measures concentrated, with uniform density, on the energy shell $(E, E + \Delta E)$, i. e., $\rho(x) = f(H) = \text{const}$ for $E \leq H \leq E + \Delta E$, and is zero outside this shell. As already noted the results, for large systems, are independent of this limit.]

While we shall be concerned here exclusively with finite systems similar problems arise for infinite systems. In the case of infinite quantum systems, Haag, Kastler, and Trych-Pohlmeyer[6] (HKP) have shown that a condition of stability under local perturbations of the

time evolution is useful for the characterization of equilibrium states, i. e., under certain reasonable assumptions the only stable states are KMS states. Their argument may be adapted to prove a similar result for infinite classical systems.[7] In this note we wish to consider the extent to which "stability" may be useful for the characterization of equilibrium ensembles for finite systems.

2. FORMULATION OF PROBLEM

The notion of stability which we wish to use is similar to that used by HKP and may be described roughly as follows: Let ω be the stationary state given by the function $f=f(H)$. If we perturb H slightly to obtain a new Hamiltonian $H_\lambda = H + \lambda h$, we obtain a new time evolution $T_t^{\lambda h}$ ($\equiv T_t^\lambda$) such that there exists a measure $\omega^{\lambda h}$ ($\equiv \omega^\lambda$) [given by the function $f(H+\lambda h)$] which is (a) stationary under T_t^λ and (b) "close" to ω. We will say that a state ω stationary under T_t is stable if there exists such a family $\omega^{\lambda h}$ which is close to ω for all (sufficiently nice) perturbations h. A state ω which fails to be stable in this sense should not be regarded as "physical" because an arbitrarily small error in our knowledge of H could imply that ω does not even approximate a state stationary under the actual Hamiltonian time evolution.

To obtain a precise formulation of stability we must decide exactly how ω^λ is to be close to ω. Since the only use of the measure (or ensemble) is to obtain expectation values of physical observables, i. e., of functions $A(x)$, which (by the very nature of physical observations) may be assumed to be smooth functions of x, $x \in \Gamma$, closeness should refer to such expectation values. We shall write $\omega(A)$ and $\omega^\lambda(A)$ for the expectation value[8] of A, with respect to the measures ω and ω^λ, and will assume throughout that H and all perturbations are $\in C^2(\Gamma)$ and that h is bounded. Some possibilities are:

(i) $\omega^\lambda \to \omega$ in norm, i. e.,

$$|\omega^\lambda(A) - \omega(A)| \leq \epsilon(\lambda) \|A\|,$$

where $\lim_{\lambda\to 0}\epsilon(\lambda) = 0$, $A \in C(\Gamma)$, the bounded continuous functions on the phase space Γ of the finite system, and $\|A\| = \sup_{x\in\Gamma} |A(x)|$;

(ii) $\omega^\lambda \underset{\lambda\to 0}{\to} \omega$ weakly, i. e., $\omega^\lambda(A) \underset{\lambda\to 0}{\to} \omega(A)$ for all $A \in C(\Gamma)$.

Clearly, (i) implies (ii). It is also worth noting that there is a natural dynamical formulation of stability which is equivalent to (i).

(i′) $T_t^{\lambda h}\omega$ remains close (in norm) to ω uniformly in t, for any perturbation h, when λ is sufficiently small, i. e.,

$$|\omega(T_t^\lambda A) - \omega(A)| < \epsilon(\lambda)\|A\|$$

for all $A \in C(\Gamma)$ and all t.

To prove equivalence we note that (i′) follows from (i) because

$$|\omega(T_t^\lambda A) - \omega(A)| \leq |\omega(T_t^\lambda A) - \omega^\lambda(T_t^\lambda A)| + |\omega^\lambda(A) - \omega(A)| \leq 2\epsilon(\lambda)\|A\|,$$

since $\omega^\lambda(T_t^\lambda A) = \omega^\lambda(A)$ by the stationarity of ω^λ under the perturbed evolution and $\|T_t^\lambda A\| = \|A\|$. Conversely, if (i′) holds we may construct ω^λ norm close to ω as a weak limit point of the time averages $\bar{\omega}_T^\lambda$ of the measures

$$T_t^\lambda\omega \quad (\bar{\omega}_T^\lambda = 1/T \int_0^T dt\, T_t^\lambda\omega).$$

Condition (i′) may be called dynamical stability: Suppose a perturbation λh is added to H at some time, say $t=0$; then ω will change with time for $t>0$. If, however, ω satisfies (i′) and λ is small then the expectation values of physical observables will also be changed only slightly even after very long times. (This remains true also if the initial state is not exactly ω but some state ω' which is close to ω in norm.)

These conditions have quantum counterparts: one replaces $C(\Gamma)$ in the above by the C^*-algebra $B(\mathcal{H})$ of bounded operators on the Hilbert space $\mathcal{H}$ corresponding to the finite quantum system—of a finite number of particles in a finite volume. ω and ω^λ correspond to normal states on $B(\mathcal{H})$ [i. e., positive linear functionals $\tilde{\omega}$ of the form $A \to \mathrm{tr}(A\rho)$, $A \in B(\mathcal{H})$, where $\rho \in B(\mathcal{H})$ is positive and $\mathrm{tr}(\rho) = 1$] which are invariant under the one-parameter groups T_t and $T_t^{\lambda h}$ generated by the Hamiltonians H and $H+\lambda h$, $h \in B(\mathcal{H})$, respectively. For finite systems H has discrete spectrum and corresponding to states of the form $f(H)$ for classical systems one has the invariant states given by $\rho = f(H)$ (e. g., $\rho = e^{-\beta H}/\mathrm{Tr}e^{-\beta H}$) for quantum systems.

In both the classical and qunatum situations, a state given by a (reasonable) function $f(H)$ will satisfy (i) and (ii) and thus, also (i′). In the quantum case a state is stationary if and only if $[\rho, H]$ ($\equiv \rho H - H\rho$) $= 0$, so that if H has nondegenerate spectrum, ρ must clearly be of the desired form. Even if H is degenerate the restriction of ρ to each energy level must still be the identity if (ii) is to be satisfied, since any splitting of an energy level may be achieved by the appropriate choice of perturbation.[6] In the classical situation we need stronger conditions that (i) and (ii) to obtain a general result. Before introducing such a condition, in Sec. 4, we shall, in the next section, investigate some consequence which follow solely from the "weak stability" condition (ii).

3. SOME CONSEQUENCES OF WEAK STABILITY

Proposition 1: Let ω be weakly stable under the perturbation h as in (ii), i. e., there exists a collection $\omega^{\lambda h}$ of states invariant under the dynamics generated by $H+\lambda h$ which converge weakly to ω. Then $\omega^{\lambda h}(Q)$ is differentiable at $\lambda = 0$ on observables of the form $Q = \{H, B\}$ [the Poisson bracket (P. B.) of H with B] for some $B \in C_0^1(\Gamma)$ (C^1 functions of compact support)[9] and

$$\frac{d}{d\lambda}\,\omega^{\lambda h}(\{H, B\})_{\lambda=0} = -\,\omega(\{h, B\}). \tag{3.1}$$

In particular, if B is a constant of the motion $\{H, B\} = 0$, then

$$\omega(\{h, B\}) = 0. \tag{3.2}$$

Proof: For any $B \in C_0^1(\Gamma)$ the perturbed states satisfy

$$0 = \frac{d}{dt}\,\omega^{\lambda h}(T_t^\lambda B)_{t=0} = \omega^{\lambda h}(\{H+\lambda h, B\}),$$

or

$$\frac{1}{\lambda}\,\omega^{\lambda h}(\{H,B\}) = -\,\omega^{\lambda h}(\{h,B\}).$$

The weak continuity of $\omega^{\lambda h}$ at $\lambda = 0$ implies therefore the existence of the limit

$$\lim_{\lambda\to 0}\frac{1}{\lambda}\,\omega^{\lambda h}(\{H,B\}) = -\lim_{\lambda\to 0}\omega^{\lambda h}(\{h,B\}) = -\,\omega(\{h,B\}).$$

Since, by stationarity,

$$\omega(\{H,B\}) = 0,$$

the above limit is the weak derivative of $\omega^{\lambda h}$ on $Q = \{H,B\}$.

Proposition 2: If ω satisfies stability (ii) and is given by a $C^1(\Gamma)$ density ρ, then

$$\{\rho,B\} = 0 \tag{3.3}$$

for any $B \in C_0^1(\Gamma)$ such that $\{H,B\} = 0$.

Proof: By Proposition 1 $\{H,B\} = 0$ implies $\omega(\{h,B\}) = 0$ for any $h \in C_0^2$. In terms of ρ we thus have, using well-known properties of the P. B.,

$$\begin{aligned} 0 &= \int dx\,\rho\{h,B\} \\ &= \int dx\,\{\rho h,B\} - \int dx\,h\{\rho,B\} \\ &= -\int dx\,h\{\rho,B\}. \end{aligned}$$

Since h is arbitrary this implies (3.3.).

We have thus obtained a simple condition on ω, (3.2) and (3.3), necessary for stability (ii).

The above arguments can be reproduced for quantum systems, with the understanding that $\{\ ,\ \}$ stands for the commutator. According to (3.3) a state of a quantum system, given by a density operator ρ, is stable (ii) only if ρ commutes with all operators which commute with the Hamiltonian H. Since H has discrete spectrum it follows simply that ρ is a function of H.

No such general result can be expected for classical systems as may be seen by considering integrable systems for which the Kolmogorov—Arnold—Moser (KAM) theorem[2,4] is applicable. It can be shown, see remark at end of Sec. 4, that for such systems even the stronger stability condition (i) is not sufficient to insure the desired result $\rho = f(H)$.

The difference between classical and quantum systems appears to be due to the lack of a sufficient number of global constants of the motion in the classical case. This prevents fuller exploitation of Proposition 2 whose usefulness depends on the existence of an abundance of invariants. Even integrable systems, if they satisfy the conditions of the KAM theorem, have only a "limited" number of such constant (i.e., n constants when Γ is a $2n$-dimensional space). This shows up in the requirements for KAM theorem that the frequencies be incommensurable[4]; this reduces the number of smooth invariants; e.g., for two uncoupled oscillators there exists a function of the two phases which is a (smooth) invariant iff the frequencies are commensurable. Indeed, we shall now prove that in the extreme case of a periodic system weak stability alone implies that $\rho = f(H)$. We shall consider this case explicitly, despite its limited applicability, to illustrate the method used in the next section for more "typical" systems.

Proposition 3: Let ω be a state of a periodic system, given by a $C^1(\Gamma)$ density ρ. If ω is weakly stable [i. e., satisfies stability (ii)], then locally ρ is a function of H, i. e.,

$$\text{grad}\rho \text{ is parallel to } \text{grad}H. \tag{3.4}$$

Proof: Denote by τ the period of the system. Then, for any $A \in C_0^1$,

$$\bar{A}(x) = \int_0^\tau dt\,A(T_t x)$$

is a constant of the motion. Proposition 2 now implies that

$$\begin{aligned} 0 = \{\rho,\bar{A}\} &= \{\rho,\ \int_0^\tau dt\,T_t A\} = \int_0^\tau dt\,\{\rho,T_t A\} \\ &= \int_0^\tau dt\,T_t\{\rho,A\}, \end{aligned} \tag{3.5}$$

where we have used the invariance of ρ under T_t. Assume now that gradρ is not parallel to grad$H$ at some point x. One could then find on observable A, with support in a neighborhood of x, in which $\{\rho,A\} > 0$ along the orbit of x. This would contradict (3.5).

The typical (generic) integrable system is not periodic. Nevertheless its periodic points are dense in the phase space.[4] In the next section we show how to obtain a positive result for such systems at the price of imposing a somewhat stronger, and not so physical, requirement of stability on the equilibrium states.

4. A STRONGER STABILITY CONDITION

As we have seen in Propositions 1 and 2, the weak stability of a state ω enables one to define, for each smooth perturbation h of compact support, a functional L_h, whose domain are observables of the form $Q = \{H,B\}$, by

$$L_h(\{H,B\}) = -\,\omega(\{h,B\}).$$

L_h was shown there to be the weak derivative of the perturbed states $\omega^{\lambda h}$.

Definition: A state ω satisfies stability (iii) if it is weakly stable and if, for each $h \in C_0^2$, the functional L_h is given by a $C^2(\Gamma)$ function f_h, i. e.,

$$L_h(\{H,B\}) = \int dx\,f_h(x)\{H,B\}.$$

When ω has a density ρ

$$\int dx\,f_h\{H,B\} = -\int dx\,\rho\{h,B\}.$$

This gives after integration by parts, assuming $\rho \in C^1(\Gamma)$,

$$-\int dx\,B\{H,f_h\} = \int dx\,B\{h,\rho\}.$$

Since this holds for, essentially, any B it implies

$$-\{H,f_h\} = \{h,\rho\}. \tag{4.1}$$

Thus, for states given by a density, stability (iii) implies that for each perturbation h there exists a $C^1(\Gamma)$ function f_h which satisfies (4.1). This condition is satisfied by ρ of the desired form, i. e., $\rho = f(H)$, $f \in C^1$, since

$$\{h,\rho\}=\{h,f(H)\}=f'(H)\{h,H\}=\{f'(H)h,H\}$$

and one may choose $f_h=f'(H)h$.

We will now show that in the generic case, the converse of the above statement is also true.

Proposition 4: Let ω satisfy stability (iii) and be given by a C^1 density ρ. If periodic orbits (under T_t) are dense in Γ and if the energy surfaces S_E are connected then ρ is a function of H.

Proof: Let $y\in\Gamma$ be a periodic point with period τ. By stability (iii), there corresponds to each $h\in C_0^2$ a $C^1(\Gamma)$ function f_h such that

$$\{\rho,h\}=\{H,f_h\}.$$

Therefore, using the periodicity of the orbit through y, we obtain

$$\int_0^\tau du\,\{\rho,h\}(T_u y)=\int_0^\tau du\,\{H,f_h\}(T_u y)$$
$$=\int_0^\tau du\;\frac{d}{du}f_h(T_u y)\quad=f_h(T_u y)-f_h(y)=0$$

for any $h\in C_0^2$. By the same argument as in the proof of Proposition 3, we conclude that gradρ is parallel to grad$H$ at y.

Since the periodic points are dense the gradients of ρ and of H are parallel everywhere. The connectedness of energy surfaces now implies that ρ is a function of H.

Remark: The assumptions made in Proposition 4 cannot easily be weakened as may be seen by considering stability in integrable systems to which the KAM theorem is applicable.[4] (The ideal gas in a torus is such a system.) In these systems the phase space is decomposable into invariant (under T_t) tori "most" of which are stable under small (sufficiently smooth) perturbations h: That is, except for a family of tori of total measure $\epsilon(\lambda)$, there corresponds to each T_t-invariant torus M a uniformly close $T_t^{\lambda h}$-invariant torus M^λ (on which the $T_t^{\lambda h}$ time evolution uniformly approximates the T_t evolution on M). Here $\epsilon(\lambda)\to 0$ as $\lambda\to 0$ and M^λ is "differentiably close" to M. Hence for any T_t-stationary measure which is given by a smooth "function of the invariant tori" (i.e., a function of the "action variables" parameterizing the tori) one may use the correspondence $M\longleftrightarrow M^\lambda$ to construct a T_t^λ-stationary measure ω^λ which is norm close to ω and even differentiably close. Thus, unless the use of perturbations to which KAM does not apply is allowed—in our argument h could be arbitrarily smooth—the proposition will not hold if we replace in it stability (iii) by stability (ii) or even stability (i). Stability (iii), on the other hand, will rule out these cases because the derivative of ω^λ at $\lambda=0$ may fail to be even a function and will certainly not be C^1. A positive result may, however, be possible if the ω^λ are required to be given by smooth functions, since this is almost certainly not the case for the ω^λ which can be constructed by the use of the KAM theorem.

ACKNOWLEDGMENTS

We would like to thank Rudolf Haag for informing us of (and explaining to us) the results of Ref. 6 prior to publication, and for many very useful comments. We have also benefited greatly from discussions with Oscar Lanford. [The proof of the equivalence of (i) and (i′) is due to him.] Finally, we would like to thank the Battelle Foundation for the opportunity to participate in the Rencontres on Dynamical Systems (Seattle, 1974) where we had valuable discussions with Jurgen Moser and learned something about the KAM theorem.

*Supported in part by USAFOSR Grant #73-2430B.
†Supported in part by NSF Grant GP-37069X.
‡Supported in part by NSF Grant GP-16147A#1.

[1]See, for example, (a) O. Penrose, *Foundations of Statistical Mechanics* (Pergamon, Oxford, 1970); (b) Arthur Hobson, *Concepts in Statistical Mechanics* (Gordon and Breach, New York, 1971); (c) D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1968).

[2]J. L. Lebowitz and O. Penrose, Physics Today **26**, N. 2 (1973) and references cited there.

[3]We take our system to be confined to a box, or torus, with bounded energy surfaces.

[4]See V. I. Arnold and A. Avez, *Ergodic Problems of Statistical Mechanics* (Benjamin, New York, 1968).

[5]There are of course special situations; c.f. Ref. 1, where one considers explicitly equilibrium ensemble densities which depend also on the total linear and/or angular momentum of the system. We do not discuss these here.

[6]R. Haag, D. Kastler, and E. Trych-Pohlmeyer, Commun. Math. Phys. **38**, 173 (1974).

[7]M. Aizenman, G. Gallavotti, S. Goldstein, and J. L. Lebowitz (to be published).

[8]This is consistent with our definition of $\omega(A)$ in Sec. 1 if we think there of $A(x)$ as the characteristic function of the set $A\subset\Gamma$, $A(x)=1$ if $x\in A$, $A(x)=0$ for $x\notin A$.

[9]This collection includes observables of the form T_tA-A, for $t\in\mathbb{R}$ and $A\in C_0^1(\Gamma)$ [since $T_tA-A=\{H,\int_0^t T_u(A)\,du\}$] and is, therefore, dense in the orthogonal complement [in $L_2(\omega)$] of the constants of the motion.

Optical theorems for three-to-three processes

Joseph Coster*

Physics Department, Western Illinois University, Macomb, Illinois 61455

Henry P. Stapp†

Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720
(Received 11 December 1974)

Optical theorems for three-to-three processes are derived from S-matrix principles. These theorems express all single, double, and multiple discontinuities across all combinations of normal threshold cuts in terms of physical scattering amplitudes. The 2^{16} functions corresponding to all combinations of sides of the 16 normal threshold cuts are determined by analyticity requirements and the generalized Steinmann relations. These two conditions guarantee that these functions can be identified with the corresponding functions in the Regge discontinuity formulas of Weis. This identification provides for a possible enlargement of the scope of Regge–Mueller-type analyses of high-energy processes.

I. INTRODUCTION

A multiparticle optical theorem is used in the studies of high-energy processes stemming from the work of Mueller. Mueller[1] originally showed that important properties of inclusive cross sections follow from the assumption that certain matrix elements of current operators enjoy Regge behavior. Tan[2] then observed that Mueller's assumption about current operators could be replaced by the S-matrix assumption that the scattering function itself evaluated on various sides of normal threshold cuts enjoys Regge behavior. Tan's argument was based on a multiparticle generalization of the ordinary optical theorem, called the inclusive optical theorem, which was subsequently proved in Ref. 3. This theorem formed the basis of further developments of Mueller's ideas.

These further developments are generally formulated in the S-matrix framework. However, the proof in Ref. 3 of the underlying inclusive optical theorem is based on field theory, and depends on off-mass-shell continuations. The question thus arises whether this theorem can be proved in the mass-shell S-matrix framework.

The inclusive optical theorem is a special case of a discontinuity formula that had been proposed in early S-matrix works. Tan gave a heuristic S-matrix derivation based on crossing. However, many possible cuts stand in the way of on-mass-shell continuations to the cross channels and the effects of these cuts were not fully analyzed by Tan.

One purpose of the present work is to provide an S-matrix derivation of the inclusive optical theorem for three-to-three processes. More generally the purpose is to derive formulas for all of the single, double, and multiple discontinuities across all combinations of normal threshold cuts in three-to-three scattering functions. These discontinuities are expressed in terms of physical scattering functions, i.e., in terms of scattering functions evaluated at their physical boundary points.

The central problem in this endeavor is to determine the 2^{16} functions corresponding to all possible combinations of sides of the 16 normal threshold cuts. A basic requirement on these functions is that they continue around all normal threshold singularities in the appropriate manner. This requirement would seem at first insufficient to determine these functions, but it turns out to be very stringent, and it probably uniquely determines 26 018 of the $2^{16} = 65\,536$ functions. The remaining functions are then determined by the generalized Steinmann relations,[4] which must be satisfied if the functions are to be identified with corresponding functions occurring in the Regge discontinuity formulas obtained by Weis. If this identification is made then the optical-theorem expressions for the various discontinuities can be equated to the corresponding Regge formulas, thus opening the way to an enlargement of the scope of Regge—Mueller analyses.

Our general optical theorems will be described presently. First the inclusive optical theorem and certain related formulas are reviewed.

The inclusive optical theorem is essentially a formula for the discontinuity of the scattering function across a certain basic cut, evaluated on specified sides of each of the other basic cuts. These basic cuts are cuts in the channel energies that start at the lowest normal thresholds and extend to plus infinity. This formulation is based on field theory, and the use of channel energies as basic variables, rather than channel invariants, stems from the fact that the scattering function has simple cut-plane analyticity in the complex energy variables if the 3-momenta are all held fixed and real.[3] In particular, the singularities are confined to the union of the surfaces $\mathrm{Im}E_g = 0$, where E_g is the channel energy associated with channel g.

For the three-to-three case the complex energy space is five-dimensional, since one energy variable is fixed by energy conservation. There are sixteen channels g associated with basic cuts. These are the one direct or total-energy channel, the three initial subenergy channels, the three final subenergy channels, and the nine cross-energy channels that are defined by sets consisting of one initial particle and two final particles, or by the complementary sets consisting of two initial particles and one final particle.

The sixteen planes $\mathrm{Im}E_g = 0$ divide the five-dimension-

al space of imaginary energies into 2282 regions called zones. Each zone is a cone-shaped region with apex at the origin. The scattering function is analytic in each zone, and has, for each zone, a corresponding boundary value, which is defined by letting the complex point $p' = (p_1', \cdots, p_6')$ approach the real boundary point $p = (p_1, \cdots, p_6)$ from within that zone. The inclusive optical theorem, for the three-to-three case, is essentially a formula for the difference between two of these 2282 zone boundary values.

In the S-matrix framework the mass-shell constraints are retained. Hence the energies are real if the 3-momenta are real. Consequently, the zones defined above do not intersect the (mass-shell) domain of definition of the scattering function. It is therefore necessary to define the boundary values by a different procedure. This procedure uses the sixteen channel invariants s_g, instead of the sixteen energies E_g.

The sixteen channel invariants s_g are not independent variables: They are functions of the eight independent scalar invariants. Suppose for a moment, however, that these sixteen variables could be treated as independent variables, and that the scattering function had only normal threshold singularities, and hence enjoyed cut-plane analyticity in $(s_1, \cdots, s_{16})$ space, with the singularities confined to the union of the planes $\mathrm{Im}s_g = 0$. In this case there would be 2^{16} boundary values, one corresponding to each combination of sides of the 16 cuts $\mathrm{Im}s_g = 0$. Stated differently, for every one of the 2^{16} subsets G of the set E of sixteen indices g there would be a boundary value $M^G(s_1, \cdots, s_{16})$ obtained by approaching the real boundary point $(s_1, \cdots, s_{16})$ from the lower-half s_g plane for every g in G, and from the upper half s_g plane for every g in the complement $\bar{G} \equiv E - G$ of G.

Actually the sixteen variables s_g are not independent, and the singularities of the scattering function are not confined to the surfaces $\mathrm{Im}s_g = 0$. Nevertheless, there is a set of 2^{16} functions $M^G(p)$ that is analogous to the set of 2^{16} functions $M^G(s_1, \cdots, s_{16})$ defined above. A large number of these functions $M^G(p)$ satisfy the primary properties to be described next, and the rest satisfy a weakened version of these properties.

The primary properties of the functions $M^G(p)$ are as follows:

(1) $M^G(p)$ is analytic at all real p, except on certain Landau singularity surfaces.

(2) $M^G(p)$ is a single analytic function: It continues into itself around each Landau singularity surface by some infinitesimal detour.

(3) $M^G(p)$ continues into itself around each normal threshold singularity surface $s_g = (\sum m_j)^2$ by passing into the lower- or upper-half s_g plane according to whether g is contained in G or $\bar{G}$. Here $\sum m_j$ is some sum of physical particle masses.

(4) $M^G(p) - M^{Gg}(p)$ vanishes if $s_g(p)$ is less than its value on the leading g-channel normal threshold. Here $Gg \equiv G \cup g$.

(5) $M(p) = M^\Phi(p)$ is the physical scattering function, which is the connected part of the S matrix, divided by $(2\pi)^4\delta^4(\sum p_i - \sum p_f)$.

(6) $M^E(p) = -M^\dagger(p)$. This property is Hermitian analyticity: The scattering function evaluated below all the cuts is minus the Hermitian conjugate scattering function.

It might seem that these properties would be easy to satisfy: One might try to define the functions M^G by simply continuing the scattering function according to the prescribed rule around every normal threshold singularity surface, and continuing it according to any arbitrary rule around every other singularity surface. This procedure does not work. For the way in which the function is continued around a normal threshold singularity surface determines the way in which it must be continued around any nonnormal threshold singularity surfaces that emerge from it. And this determination fixes, in turn, the way in which the function must be continued around any normal threshold singularity surfaces into which this nonnormal threshold singularity surface merges. Consequently, it is not evident that a set of 2^{16} functions $M^G(p)$ satisfying the principal defining properties exists. And given that such a set exists it is not clear that these functions can be expressed in terms of *physical* scattering functions alone, as contrasted to unphysical boundary values of the analytic continuations of scattering functions.

A set of functions $M^G(p)$ that satisfy the primary defining properties is constructed in later sections by first constructing two sets of partial solutions, the T^G and the $\bar{T}^G$. The 2^{16} functions T^G satisfy properties (1), (4), (5), and (6), with T^G replacing M^G, and half of (3); the function T^G continues into itself around each normal threshold singularity $s_g = (\sum m_j)^2$ for $g \in G$ by passing into the lower-half s_g plane. Similarly, the 2^{16} functions $\bar{T}^G$ satisfy properties (1), (4), (5), and (6), with $\bar{T}^{E-G}$ replacing M^G, and the other half of (3); the function $\bar{T}^G$ continues into itself around each normal threshold singularity $s_g = (\sum m_j)^2$ for $g \in \bar{G}$ by passing into the upper-half s_g plane. For 26 018 of the $2^{16} = 65\,536$ possible sets G the identity $T^G = \bar{T}^{E-G}$ holds. For these values of G the functions M^G are defined by

$$M^G = T^G = \bar{T}^{E-G}.$$

Then the final condition, property (2), is proved. This is nontrivial, because each $M^G(p)$ is constructed as a sum of functions $M_H(p)$ only one of which, $M_\Phi(p) = M^\Phi(p) = M(p)$, is a single analytic function.

Each of the component functions M_H except $M_\Phi = M$ is a unitarity-type sum of products two or more physical scattering functions, or their complex conjugates. In a field theory framework off-mass-shell extensions of the scattering functions are introduced. Thus in that framework the functions M_H, and hence also the functions M^G, have off-mass-shell extensions. Near the off-mass-shell point $p_1 = \cdots = p_6 = 0$ all of the differences $M^G(p) - M^{Gg}(p)$ vanish, and hence all of the functions $M^G(p)$ are equal. Thus in this off-mass-shell framework the 26 018 functions M^G are all different boundary values of the analytically continued scattering function. They are, accordingly, called boundary values of the analytically continued scattering function, even though no *mass-shell* path of continuation that links each function M^G to the physical function M is constructed in the present work.

By virtue of property (3) the boundary value $M^G(p)$ is a boundary value from below the normal threshold cuts corresponding to channels $g \in G$ and from above the normal threshold cuts corresponding to channels $g \in \bar{G}$. For brevity this boundary value $M^G(p)$ is called the scattering function evaluated below the cuts $g \in G$ and above the cuts $g \in \bar{G}$. Similarly, the difference $M^G - M^{Gg} \equiv M_g{}^G$ is called the discontinuity across the cut g evaluated below the cuts $g' \in G$ and above the cuts $g' \in E - Gg$. The higher-order multiple discontinuities are defined similarly.

For the remaining $2^{16} - 26\,018$ values of G no functions $M^G(p)$ that strictly satisfy all of the primary properties have been found, and we believe that none exist. However, it is convenient, for reasons to be discussed later, to enlarge the set of 26 018 boundary values M^G into a full set of 2^{16} functions M^G by means of generalized Steinmann relations. This is discussed next.

The Steinmann relations can be formulated in terms of the notion of overlapping channels. Two channels are said to *overlap* if and only if neither of the two complementary subsets of particles that defines one channel is contained in either of the two complementary subsets that define the other. The Steinmann relations are equivalent to the following *Steinmann discontinuity property*: If two channels g and h overlap then the discontinuity across the cut g does not depend on whether it is evaluated above or below the cut h. The ordinary Steinmann relations assert that this property holds for all discontinuities formed from the 2282 zone boundary values, with the cuts g and h identified as two of the channel-energy cuts that separate these zones. The generalized Steinmann relations assert that the Steinmann discontinuity property holds for all discontinuities $M^G - M^{Gg} \equiv M_g{}^G$ formed from the 2^{16} functions M^G.

It is not obvious that there is a set of 2^{16} functions M^G that includes the 26 018 boundary values M^G defined above and that also satisfies the generalized Steinmann relations. However, there is such a set, and it is unique. In this set the remaining functions M^G are given by $M^G = T^G$ or $M^G = \bar{T}^{E-G}$ according to whether $\bar{G}$ or G contains the direct channel label $g = t$. This set of 2^{16} functions M^G satisfies the properties (1), (4), (5), and (6), and, for each G, half of property (3). The remaining half of (3) is disrupted by certain singularities that are associated with closed loop diagrams. Thus property (3) holds for the $2^{16} - 26\,018$ functions $M^G(p)$ in a tree-diagram approximation.

Property (2) does not hold in general for the $2^{16} - 26\,018$ functions M^G. Thus these functions cannot, in general, be identified as boundary values of the analytically continued scattering function. However, the singularities that block the continuation are also associated with closed loop diagrams. The situation is therefore this: For 26 018 values of G there are boundary values $M^G(p)$ of the analytically continued scattering function that satisfy all of the primary properties. This set of 26 018 boundary values is uniquely imbedded in a set of 2^{16} functions $M^G(p)$ that satisfy (1), (4), (5), and (6), together with the generalized Steinmann relations. These extra $2^{16} - 26\,018$ functions satisfy the two remaining properties, (2) and (3), in a tree-diagram approximation.

The 2^{16} functions $M^G(p)$ are connected in a natural way to the 2282 zone boundary values. Each of the 2282 zones lies below some set G of basic (channel energy) cuts, and above the rest. Thus each of the 2282 zone boundary values corresponds to one of the 2^{16} functions M^G, and in fact to one of the 26 018 boundary values M^G. It is shown in Ref. 4 that each of the zone boundary values is equal to the corresponding function M^G. Thus the full set of functions M^G is an extension of the set of 2282 zone boundary values to a set of 2^{16} functions that satisfies the generalized Steinmann relations. This extension is unique.[4] Moreover, each of the functions M^G can be expressed as a linear combination of the 2282 zone boundary values.[4] These results imply the uniqueness of the set of 2^{16} functions $M^G(p)$ insofar as one demands both their agreement with the zone boundary values and the validity of the generalized Steinmann relations.

A formula is given in Sec. II that compactly expresses in terms of physical scattering functions each of the 2^{16} functions M^G, and each of the single, double, and multiple discontinuities formed from the set of 2^{16} functions $M^G(p)$. The relevance of this formula to Regge theory is now discussed.

Regge behavior in its simplest form is simply a falloff property of the scattering amplitude itself in certain limits. However, under the impetus of Mueller's work the hypothesis of Regge behavior was extended to cover also discontinuities across the basic normal threshold cuts. To get Mueller's results, it is sufficient to assume merely that the particular discontinuity that occurs in the inclusive optical theorem enjoys Regge behavior. However, it then becomes natural to assume that the discontinuities across the other basic cuts also enjoy Regge behavior.

This expanded concept of Regge behavior was explored in detail and it was soon recognized that the Steinmann relations impose important conditions on the structure of the Regge vertices.[5] Ultimately, on the basis of many works, Weis[6] obtained a general formula for discontinuities that consolidates the tenets of Regge theory with the conditions imposed by the Steinmann relations.

A fundamental aspect of the Regge hypothesis for scattering functions is that the stipulated behavior holds for the actual scattering function itself, not merely for some part of the amplitude, or for some approximation to the amplitude. Similarly, the stipulated behavior of the discontinuities should hold for the actual discontinuities themselves, not merely for parts of the discontinuities, or for approximations to the discontinuities. Thus the question arises: What are the discontinuities to which the Weis formula applies? The problem is that this formula refers to discontinuities associated with the various normal threshold cuts, but it is not specified exactly how the functions on the various sides of the normal threshold cuts are to be defined. To the extent that the formulas are to be restricted to the discontinuities formed from the 2282 zone boundary values defined by the channel-energy cuts, the answer is clear.

However, the Regge considerations are formulated in an S-matrix framework, and there is no indication there that the formulas should be limited in this way.

Within the S-matrix framework it seems natural to define the discontinuities in question as the discontinuities formed from the 26 018 boundary values $M^G(p)$. These boundary values correspond to continuations around the normal threshold singularities in the prescribed fashions. Moreover, they are part of the *unique* extension of the set of 2282 zone boundary values to a set of 2^{16} functions $M^G(p)$ satisfying the generalized Steinmann relations. In the derivation of the Weis formulas there is no restriction to the 2282 zone boundary values, and hence the Steinmann relations used there are actually the generalized Steinmann relations. Thus the boundary values to which these formulas apply must evidently satisfy these generalized relations. This requirement uniquely determines the functions $M^G(p)$.

By the same argument the remaining $2^{16} - 26\,018$ functions from which the Weis discontinuities are formed must be the remaining functions $M^G(p)$. These remaining functions are not boundary values of the analytically continued scattering function. However, they must, as linear combinations of the 2282 zone boundary values, fall off in the Regge manner, if the zone boundary values do. Hence there seems to be no reason to restrict the Weis formulas to include only those discontinuities formed from the 26 018 boundary values. If the remaining discontinuities are to be defined at all as differences of well-defined functions, then these functions must be the $M^G(p)$.

It seems therefore reasonable to propose, within the general framework of contemporary Regge theory, that the discontinuities specified by the Weis formula are the discontinuities formed from the 2^{16} functions M^G. This hypothesis, together with the general discontinuity formula discussed earlier, adds substantial new conditions to Regge theory, for it allows the detailed Weis expressions in terms of Regge parameters to be identified with corresponding expressions in terms of physical scattering functions.

The first application of this hypothesis, which will be described in a later work, is the derivation of a unitarity type relation for particle–reggeon scattering amplitudes. This relation is identical in form to the unitarity equation for a two-particle scattering amplitude, except that one initial particle is replaced by a reggeon and one final particle is replaced by a reggeon.

The plan of the paper is as follows. In Sec. II the notation is introduced and the general formula that defines the 2^{16} functions M^G, and the discontinuities formed from them, is described. The rule that identifies the 26 018 boundary values M^G is given at the end of that section in Eq. (2.25). In Sec. III a preliminary discussion of the functions T^G and $\bar{T}^G$ is given. This discussion is heuristic, because it expresses each of the functions T^G and $\bar{T}^G$ as a formal infinite sum of bubble-diagram functions. Three properties of these formal expressions are identified in Sec. IV as the defining properties of the functions T^G and $\bar{T}^G$, and a fundamental analytic property of these functions is derived from their defining properties alone. In Sec. V the functions T^G and $\bar{T}^G$ are expressed in a well-defined way as sums that, like unitarity sums, reduce to finite sums of bubble-diagram functions on any finite region in p space. Then it is shown that $T^G = \bar{T}^{E-G}$ for the 26 018 sets G specified by Eq. (2.25).

In Sec. VI the functions $M^G = T^G = \bar{T}^{E-G}$ are shown to satisfy the six primary properties. The proof is divided into two parts. First a proof is given in a slightly enlarged theoretical framework in which it is assumed that the physical scattering functions, and hence the functions M^G, can be extended infinitesimally off the mass shell. Then the analytic continuations that connect the function M^G on different sides of the Landau surfaces can pass through slightly off-mass-shell regions. In this off-mass-shell framework the rule for continuing around any singularity surface can be expressed as a function of the Landau diagram D that corresponds to the surface, without specifying the particular point p on the surface near which the continuation takes place. In the second part of the proof the mass-shell constraints are rigorously maintained. In this mass-shell framework the continuation depends both on the diagram D and on the point p near which the continuation takes place.

The mass-shell continuation is blocked, however, by singularities lying on a certain well-defined set of exceptional Landau surfaces. The existence of these exceptional Landau surfaces causes no special difficulties in S-matrix theory or Regge theory. The point is simply that complex cuts emerge from the unphysical sides of the normal threshold cuts. Such cuts have been found in numerous other studies. The important conclusion to be drawn here is that the simple functions M^G relevant to Regge theory are not defined by simple on-mass-shell continuations.

II. NOTATION AND RESULTS

A *channel* g is defined by a separation of the complete set of incoming and outgoing particles of a reaction into two complementary disjoint subsets J_g and $\bar{J}_g$, each of which has at least two particles. For a $3 \to 3$ process there are twenty-five channels. These are the direct or total-energy channel t, the three initial subenergy channels i, the three final subenergy channels f, the nine cross-energy channels (if), and nine other cross-energy channels, which will remain unnamed. These channels are defined in Fig. 2.1. For definiteness the sets J_g and $\bar{J}_g$ are defined so that J_g contains at least two final particles.

Throughout this paper the index i stands for 1, 2, or 3. A line i is a line corresponding to one of the three initial particles; the channel i is the corresponding initial subenergy channel specified in Fig. 2.1b. The index f always stands for 4, 5, or 6, and is used to label final lines, and also the corresponding final subenergy channels, as indicated in Fig. 2.1c.

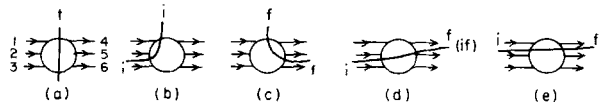


FIG. 2.1. Definition of the channels.

Stability conditions preclude the existence of normal threshold singularities in the nine channels of type (e) of Fig. 2.1, and these channels will henceforth be ignored. Thus the complete set E of channel labels g is the set of sixteen elements

$$E \equiv \{t, 1, \cdots, 6, (14), \cdots, (36)\}.$$

An arbitrary line will generally be represented by the letter j. The channel invariant s_g is the square of the sum of the momentum–energy vectors k_j of either one of the two sets J_g or $\bar{J}_g$:

$$s_g \equiv \Big(\sum_{j\in J_g} \epsilon_j p_j\Big)^2 = \Big(\sum_{j\in \bar{J}_g} \epsilon_j p_j\Big)^2.$$

The symbol ϵ_j is a sign that is plus or minus according to whether j is an f or an i. The real physical momentum–energy of particle j is denoted by p_j.

The set E has 2^{16} different subsets G. For each of these G there is a function M^G. In this section a formula is given that expresses each of these 2^{16} functions M^G, and every single, double, and higher-order multiple discontinuity formed from these M^G, in terms of physical scattering functions for other processes. Certain properties of these functions M^G are derived in later sections.

The function $M^G = M^G(p)$ is a function of the set of six real on-mass-shell energy–momentum conserving 4-vectors $p \equiv (p_1, \cdots, p_6)$. It is convenient to call these functions M^G by the names that would be appropriate if the sixteen channel invariants s_g were independent variables. Thus M^G is called the function evaluated below all the cuts g in G and above all the cuts g in the complement $\bar{G} \equiv E - G$ of G. Similarly, the difference

$$M_g \equiv M - M^g \tag{2.1}$$

is called the discontinuity across the cut g. And for any h in $\bar{G}$ the difference

$$M_h{}^G \equiv M^G - M^{Gh} \tag{2.2}$$

is called the discontinuity across the cut h evaluated below all the cuts g in G and above all the cuts g not in $Gh \equiv G \cup h$. The cut g means the cut lying in $\mathrm{Im} s_g = 0$.

There are, in addition to single discontinuities, also double discontinuities, and higher-order multiple discontinuities. The double discontinuity across a pair of different cuts h and k is the discontinuity across the cut h of the discontinuity across the cut k:

$$M_{hk} \equiv (M - M^k) - (M^h - M^{hk})$$

$$= M - M^k - M^h + M^{hk}. \tag{2.3}$$

Similarly, the multiple discontinuity across the set of cuts $H = (h_1, h_2, \cdots, h_m)$, with the h_i all different, is the discontinuity across h_1 of the discontinuity across $h_2 \cdots$ of the discontinuity across h_m. It is equal to

$$M_H \equiv \sum_{H'\subset H} (-1)^{n(H')} M^{H'}, \tag{2.4a}$$

where $M_\phi = M^\phi = M$ and the sum runs over all different subsets H' of H, including the empty set ϕ, and $n(H')$ is the number of elements of H'. Similarly, for $G \cap H = \phi$,

$$M_H{}^G \equiv \sum_{H'\subset H} (-1)^{n(H')} M^{GH'} \tag{2.4b}$$

is the multiple discontinuity across the set of cuts H evaluated below all the cuts g in G and above all the cuts g not in $GH \equiv G \cup H$.

Note that the multiple discontinuity $M_H{}^G$ is independent of the order of the elements in H. It is also independent of the order of the elements of G. These sets are regarded as unordered sets.

The formulas (2.4a) and (2.4b) can be inverted to give

$$M^G = \sum_{G'\subset G} (-1)^{n(G')} M_{G'} \tag{2.5a}$$

and, for $G \cap H = \phi$,

$$M_H{}^G = \sum_{G'\subset G} (-1)^{n(G')} M_{HG'}. \tag{2.5b}$$

Equation (2.5a) is just a special case of (2.5b). These formulas, which are derived in Appendix B, express all 2^{16} functions M^G, and also all the single, double, and higher-order multiple discontinuities formed from them, evaluated on all possible sides of all the remaining cuts, in terms of the various multiple discontinuities M_H.

Our general discontinuity formula is Eq. (2.5), together with explicit formulas for all of the functions M_H occurring on its right-hand side. Most of these functions M_H vanish by virtue of the generalized Steinmann relations.

The generalized Steinmann relations assert that the discontinuity $M_h{}^G$ across any cut h is independent of whether it is evaluated above or below any "crossed cut" g. Two cuts g and h are said to be crossed if and only if the corresponding channels overlap: i.e., if and only if each of the four sets $J_g \cap J_h$, $\bar{J}_g \cap J_h$, $J_g \cap \bar{J}_h$, and $\bar{J}_g \cap \bar{J}_h$ is nonempty. This requirement is equivalent to the condition that none of the four sets J_g, J_h, $\bar{J}_g$, $\bar{J}_h$ be a subset of any of the others.

Two different cuts $h \in E$ and $k \in E$ are crossed if and only if one of the following five conditions holds:

(a) both are initial subenergy cuts i,

(b) both are final subenergy cuts f,

(c) both are cross-energy cuts (if), (2.6)

(d) one is a cross-energy cut (if), and one is the total energy cut t,

(e) one is a cross energy cut (if) and the other is a subenergy cut that is neither i nor f.

It is convenient to speak of g as either a label, a channel, or a cut. Hence a set of g's can be called a set of labels, a set of channels, or a set of cuts.

A necessary and sufficient condition for the generalized Steinmann relation to hold is that

$$M_H = 0 \text{ if } H \text{ contains any pair of crossed cuts.} \tag{2.7}$$

It is immediately evident from (2.5) that this condition is sufficient: (2.7) and (2.5b) ensure that $M_h{}^G$ is independent of the presence in G of any cut g such that g and h are crossed. Conversely, if the M^G satisfy the generalized Steinmann relations, then (2.7) follows

from (2.4a). For if g and h are two crossed cuts in H then for every term $(-1)^{n(H')}M^{H'}$ in (2.4a) such that neither g nor h is in H' there are three other terms with H' replaced by gH', hH', and ghH', respectively, and the sum of these four terms will vanish, by virtue of the independence of M_h^G upon whether g is in G.

The conditions (2.6) entail that every set H of more than three cuts contains a pair of crossed cuts. In fact, the generalized Steinmann relations are equivalent to the following set of conditions:

$$M_H = 0 \quad \text{for } n(H) > 3, \tag{2.8a}$$

$$M_{ii'} = M_{ii'h} = 0, \tag{2.8b}$$

$$M_{ff'} = M_{ff'h} = 0, \tag{2.8c}$$

$$M_{(if)(i'f)} = M_{(if)(if')} = M_{(if)(i'f')}$$

$$= M_{(if)(i'f)h} = M_{(if)(if')h} = M_{(if)(i'f')h} = 0, \tag{2.8d}$$

$$M_{(if)t} = M_{(if)th} = 0, \tag{2.8e}$$

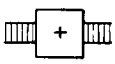
and

$$M_{i'(if)} = M_{(if)f'} = M_{i'(if)h} = M_{(if)f'h} = 0, \tag{2.8f}$$

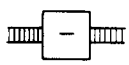
where i and i' are different elements of the set $\{1, 2, 3\}$, f and f' are different elements of the set $\{4, 5, 6\}$, h is an arbitrary element of E, and $Hh \equiv H \cup h$.

The conditions (2.8) reduce the number of nonzero M_H to 68. Moreover, all these are obtained from twelve basic forms by inserting particular integers for i and f. It is therefore feasible to exhibit explicitly all the nonzero M_H.

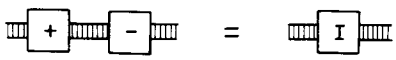
Because topological connections are of central importance it is convenient to represent M_H in a diagrammatic notation.[7] The S matrix is represented by a plus box:

$S =$  . (2.9a)

The inverse of the S matrix (or $S^\dagger$) is represented by a minus box:

$S^{-1} = S^\dagger =$  . (2.9b)

Thus unitarity says that

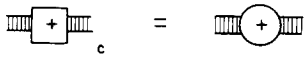
 (2.9c)

and

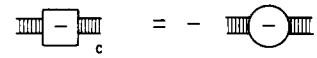
 (2.9d)

where the I-box represents the identity operator. The shaded strips represent arbitrary sets of lines, and there is an implied unitarity-type sum over all (mass-shell) values of all possible sets of intermediate-particle variables.[8]

The connected part of S is represented by a plus bubble:

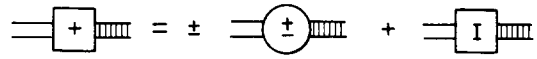
 (2.9e)

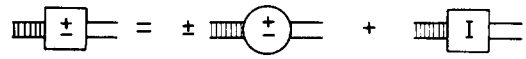
and the connected part of $S^\dagger = S^{-1}$ is represented by *minus* the minus bubble:

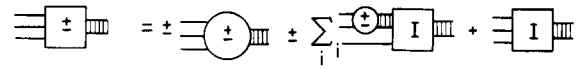
 (2.9f)

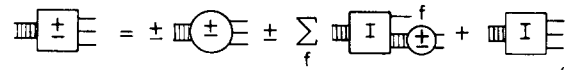
[The notation differs from that of Ref. 7 by the extra minus sign in Eq. (2.9f), which is introduced to make the minus bubble represent the continuation of the scattering function to below all the cuts. Also, in the present work the diagram are to be read from left to right, with initial lines coming first. Sums of diagrams represent the corresponding sums of functions.]

The cluster decomposition of the S matrix reads, in various special cases that are needed below,

 (2.9g)

 (2.9h)

 (2.9i)

 (2.9j)

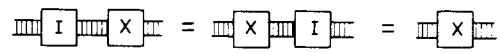
where ± is plus or minus throughout each equation. Two frequently used identities are

 (2.9k)

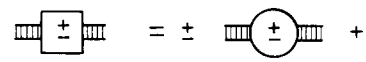
and

 . (2.9l)

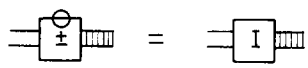
They follow immediately from unitarity and (2.9g) and (2.9h), respectively, together with the property of the I-box,[9]

 (2.9m)

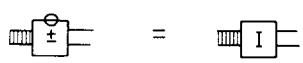
where the X box represents any combination of boxes and bubbles. Another frequently used symbol is defined by

 . (2.9n)

This equation combined with (2.9g) and (2.9h) gives

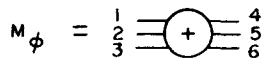
 (2.9o)

and

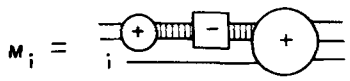
 . (2.9p)

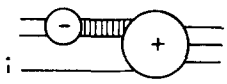
These two equations will be used later.

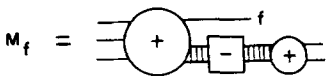
The function $M_\phi \equiv M^\phi \equiv M$ is the connected part of the physical scattering amplitude:

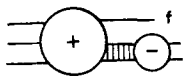
M_ϕ =  . (2.10)

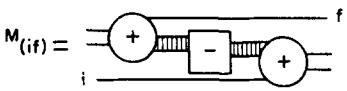
The sixteen single discontinuities M_g are

M_i = 

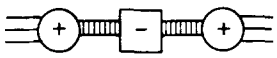
=  (2.11a)

M_f = 

=  (2.11b)

$M_{(if)}$ =  (2.11c)

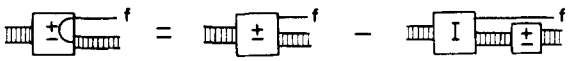
and

M_t =  . (2.11d)

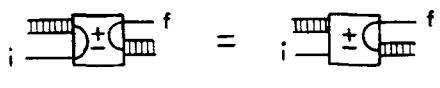
It is convenient to introduce special symbols to represent the sum of terms of S (or of $S^\dagger$) that have special connectedness properties. The symbol defined by

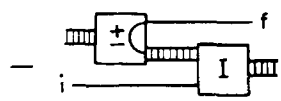
 (2.12a)

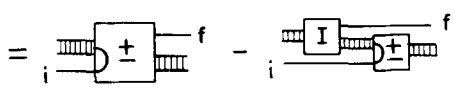
can be shown[10] to represent the sum of the terms of S (or $S^\dagger$) in which the initial line i is connected to some nontrivial bubble (a trivial bubble is a bubble that is connected to only two lines. It is usually represented by a dot); i.e., it represents the sum of terms in which the line i does not go straight through. Similarly, the symbol

 (2.12b)

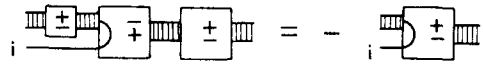
represents the sum of terms of S (or $S^\dagger$) in which the final line f does not go straight through. Finally, the symbol



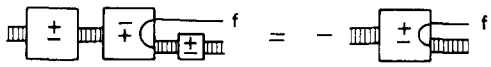
 (2.12c)

 (2.12d)

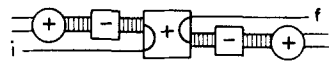
represents the sum of terms of S (or $S^\dagger$) in which neither i nor f go straight through. Two frequently used identities, which follow from (2.12a, b) and (2.9c, d, m), are

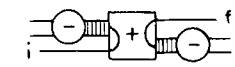
 (2.12e)

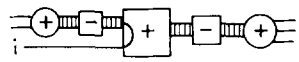
and

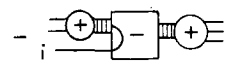
 (2.12f)

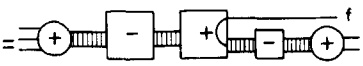
In terms of these symbols the nonvanishing M_{gh} are given by

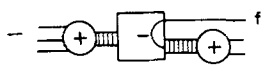
M_{if} = 

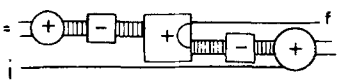
=  (2.13a)

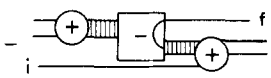
M_{it} = 

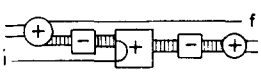
= −  (2.13b)

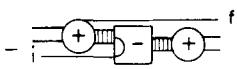
M_{tf} = 

= −  (2.13c)

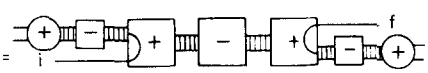
$M_{i(if)}$ = 

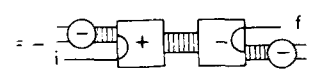
= −  (2.13d)

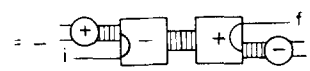
$M_{(if)f}$ = 

= −  (2.13e)

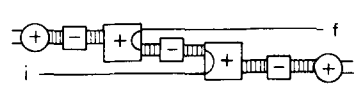
The nonvanishing functions M_{ghk} are

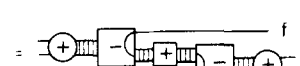
M_{itf} = 

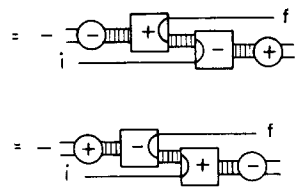
= − 

= −  (2.14a)

and

$M_{i(if)f}$ = 

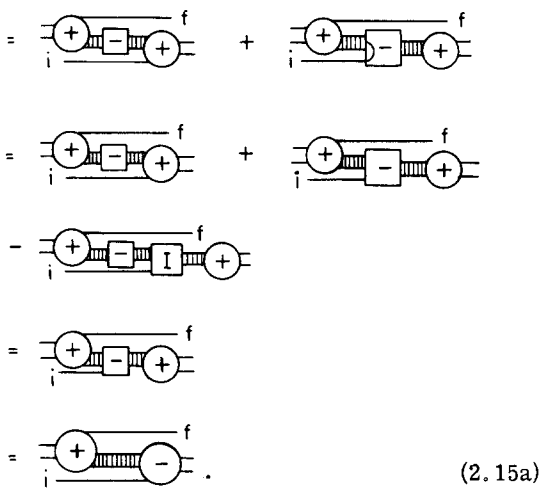
=  (2.14b)



The first form given for each of these functions M_H, although longer than the succeeding ones, exhibits a systematic rule: There is a minus box for each h in H, and these minus boxes occur between the parts of plus boxes that contain nontrivial bubbles on which the appropriate external lines terminate.

To show how these formulas work we calculate $M^f_{(if)}$, which is the discontinuity across the cut (if) evaluated below the cut f, but above all the other cuts. Using in order equations (2.5), (2.11c), (2.12a), (2.13e), (2.12a), (2.9m), and (2.9l), one obtains

$$M^f_{(if)} = M_{(if)} - M_{(if)f}$$



(2.15a)

In a similar way one obtains

$$M^i_{(if)} =$$

(2.15b)

These formulas (2.15) yield the inclusive optical theorem for the three-to-three case.

The formula (2.5a) expresses the 2^{16} functions M^G as the function $M_\phi = M^\phi = M$, which is the function evaluated above all the cuts, plus the sum of discontinuities that shift the point of evaluation, to *below* the set of cuts G. There is an equivalent formula that expresses M^G as the function $\bar{M} \equiv -M^\dagger$, which is the function evaluated below all the cuts, plus the sum of discontinuities that moves the point of evaluation to *above* the complementary set of cuts $\bar{G} \equiv E - G$.

To exhibit these other formulas we introduce for any function F formed from boxes and bubbles the notation

$$\bar{F} = -F^\dagger, \tag{2.16a}$$

where dagger represents Hermitian conjugate. In particular, for any function F represented by a single diagram one has

$$\bar{F} = -(-1)^{N_b} F(+ \rightarrow -), \tag{2.16b}$$

where N_b is the number of explicitly appearing bubbles in the diagram, and $F(+ \rightarrow -)$ is the function represented by the diagram obtained from the one representing F by reversing the sign inside each bubble, box, and modified box [such as occurs in (2.12) and (2.9n)].

The functions M^G all satisfy the important property [see (5.58)]

$$M^G = \bar{M}^{\bar{G}}. \tag{2.17}$$

But application of $F \to \bar{F}$ and $G \to \bar{G}$ to (2.5a) gives

$$M^G = \bar{M}^{\bar{G}} = \sum_{G' \subset \bar{G}} (-1)^{n(G')} \bar{M}_{G'}. \tag{2.18a}$$

Since the sums in (2.5a) and (2.18a) are over complementary sets, one of these formulas for M^G may have fewer terms than the other.

Similarly, the application of $F \to \bar{F}$ and $G \to \bar{G}$ to (2.5b) gives, for $\bar{G} \cap H = \phi$,

$$\bar{M}_H{}^{\bar{G}} = \sum_{G' \subset \bar{G}} (-1)^{n(G')} \bar{M}_{HG'}. \tag{2.18b}$$

These functions $\bar{M}_H{}^{\bar{G}}$ satisfy [see (2.4b)]

$$\begin{aligned} \bar{M}_H{}^{\bar{G}} &= \sum_{H' \subset H} (-1)^{n(H')} \bar{M}^{\bar{G}H'} \\ &= \sum_{H'' \subset H} (-1)^{n(H-H'')} \bar{M}^{\bar{G}(H-H'')} \\ &= (-1)^{n(H)} \sum_{H'' \subset H} (-1)^{n(H'')} \bar{M}^{\overline{(G-H)}H''} \\ &= (-1)^{n(H)} M_H^{G-H}, \end{aligned} \tag{2.19}$$

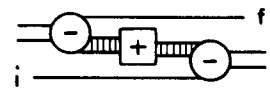
or, equivalently, for $G \cap H = \phi$

$$M_H{}^G = (-1)^{n(H)} \bar{M}_H{}^{\overline{GH}}. \tag{2.20}$$

The function $\bar{M}_H{}^{\bar{G}}$ (for $H \cap \bar{G} = \phi$) is the multiple discontinuity across the set of cuts H evaluated above the set of cuts $g \in \bar{G}$ and below the set of cuts $g \in G - H$. The extra factor $(-1)^{n(H)}$ in (2.19) and (2.20) reflects the fact that the multiple discontinuities $\bar{M}_H{}^{\bar{G}}$ are calculated by the rule "function below the cut minus function above the cut." The sets of cuts referred to in (2.19) are shown in Fig. 2.2.

As an example of these alternative formulas note that the discontinuity across the (if) cut evaluated below all other cuts is given by (2.20), (2.16), and (2.11c) as

$$M^{\overline{(if)}}_{(if)} = -\bar{M}_{(if)} =$$



(2.21)

This same discontinuity is given by (2.5), (2.8), and the definitions (2.11c), (2.13d), (2.13e), and (2.14b), as

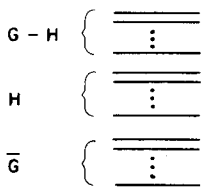
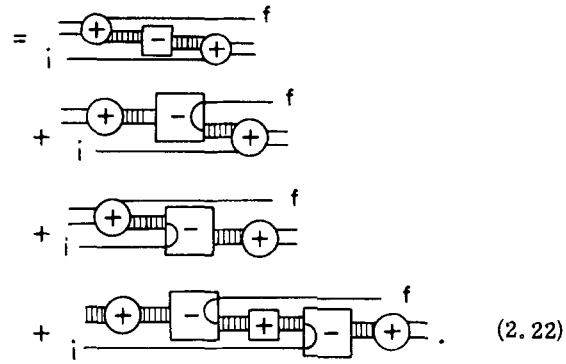


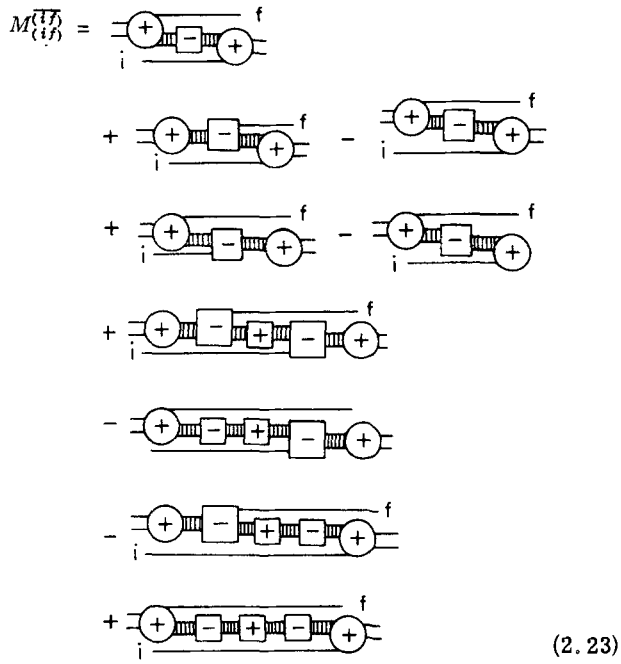
FIG. 2.2. The function $\bar{M}_H{}^{\bar{G}}$ represents the multiple discontinuity across the set of cuts H (calculated by the rule below minus above) evaluated above the set of cuts $\bar{G}$, and below the set of cuts G-H.

$$M^{(\overline{if})}_{(if)} = M_{(if)} - M_{i(if)} - M_{(if)f} + M_{i(if)f}$$



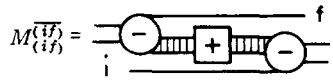
(2.22)

Introducing the definitions (2.12), and the identity property (2.9m) of the I-box, one obtains



(2.23)

which, by virtue of the unitarity equation (2.9d) and the identities (2.9k, l, m), becomes



(2.24)

which agrees with (2.21).

These results, together with several other single discontinuities that can be derived in similar ways are summarized in Fig. 2.3. Other valid formulas follow from these by a reflection about a vertical axis together with the substitution $i \leftrightarrow f$. Still others follow from the uniform substitution $+ \leftrightarrow -$ for all signs σ_g and all signs inside bubbles, boxes, and modified boxes. The $\pm$ sign standing outside the bubble on the left-hand side is not to be changed: It signifies that the discontinuity on the left-hand side is defined to be the function above the cut minus the function below the cut. That is, it is the difference $M(\sigma_g = +) - M(\sigma_g = -)$.

The formulas described above define the 2^{16} functions M^G, and all the discontinuities formed from them. In the following sections it will be shown that 26 018 of these functions have nice analyticity properties, and, in particular, continue in a well-defined way around each real p singularity surface. The 26 018 functions M^G that have this property can be identified in the following way: For any set G let the set of signs η_g be defined by the condition that η_g is plus if g lies in G and minus if g lies in $\bar{G}$. Then M^G is one of the 26 018 boundary values if and only if there is no pair (i,f) such that the following conditions are all satisfied:

$$\eta_{(if)} = \eta_i = -\eta_i = -\eta_f. \tag{2.25}$$

III. HEURISTICS

The properties of the functions M^G will be derived in Sec. VI from the properties of a similar set of functions T^G. These functions T^G are heuristically defined in this section as formal infinite sums of bubble diagram functions. Three properties of the functions T^G will be identified in Sec. IV as their defining properties, and finite expressions for them will be obtained in Sec. V.

If one introduces the definition

$$R = S - I \tag{3.1}$$

then unitarity takes the form

$$R = -R^\dagger - R^\dagger R.$$

Iteration gives, formally,

$$R = \sum_{n=1}^{\infty} (-R^\dagger)^n. \tag{3.2}$$

This is an infinite series expansion for R in terms of $R^\dagger$.

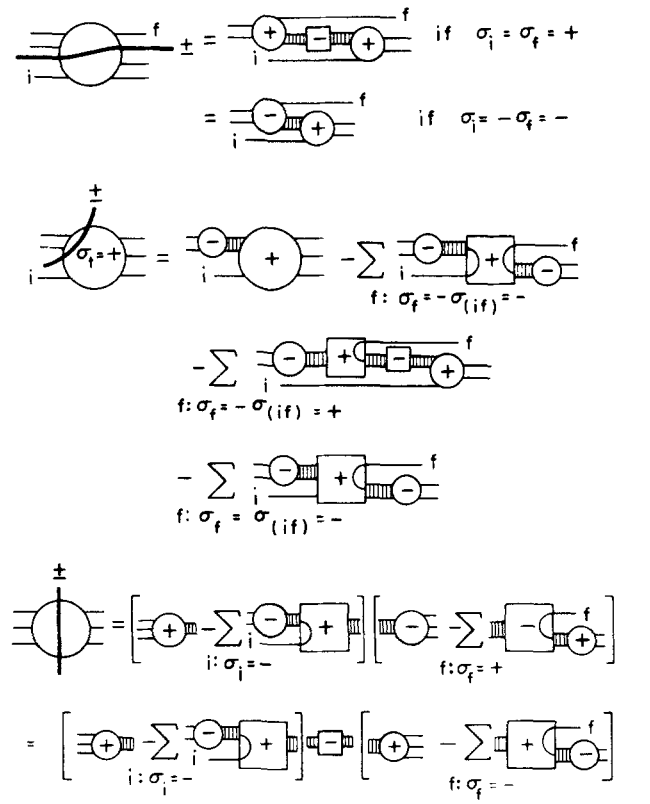
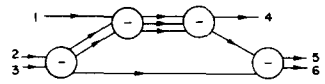


FIG. 2.3. The single discontinuity formulas.

The connected part of (3.2) can be expressed in the form[11]

$$R_c = S_c \equiv M = \sum_{B^-} F^{B^-} \equiv T. \tag{3.3}$$

Here, as throughout this paper, the symbol B^- represents a bubble diagram every bubble b of which is a (nontrivial) minus bubble. The sum in (3.3) is over all bubble diagrams B^- with the appropriate external lines, and F^{B^-} is the bubble diagram function corresponding to the bubble diagram B^-. A typical bubble diagram B^- is

$B^- =$  (3.4)

and the corresponding function F^{B^-} is the product of the four indicated functions $\bar{M} = -S_c^\dagger$, integrated over the physical values of the variables associated with the six sets of intermediate lines. Further details can be found in Ref. 11.

By definition a bubble diagram B is required to have its bubbles partially ordered by the condition that every line that connects two bubbles of B runs from the right-hand side of one bubble to the left-hand side of another bubble that stands completely to the right of the bubble from which the line came. This means that all lines can be drawn as directed lines that point from left to right. A diagram such as the one in Fig. 3.1 is not a bubble diagram.

Consider an arbitrary channel g. It is defined by a separation of the set $(1, \cdots, 6)$ into two complementary disjoint sets J_g and $\bar{J}_g$. The set β^- of all B^- (with initial lines 1, 2, 3 and final lines 4, 5, 6) can be separated into two sets β_g and β^g by the following rule: B^- belongs to β_g or β^g according to whether B^- has or does not have an explicit g-channel cut set.

Definitjon: An *explicit g-channel cut set of B* is a set of internal lines of B which if cut separates B into two connected bubble diagrams $B(J_g)$ and $B(\bar{J}_g)$, where $B(J_g)$ contains all the external lines $j \in J_g$, and $B(\bar{J}_g)$ contains all the external lines $j \in \bar{J}_g$. Moreover, each line L_j of the cut set must be directed from $B(\bar{J}_g)$ to $B(J_g)$.

Thus, for example, the diagram B^- of (3.4) belongs to β_g for $g = 1, 4$, and t, and to β^g for all other g in E.

This decomposition of β^- induces a corresponding decomposition

$$T = T_g + T^g \tag{3.5a}$$

of the inifinite series expansion T of M given in (3.3). Here

$$T_g = \sum_{B^- \in \beta_g} F^{B^-} \tag{3.5b}$$

and

$$T^g = \sum_{B^- \in \beta^g} F^{B^-}. \tag{3.5c}$$

For each channel g one can make this decomposition $T = T_g + T^g$. Moreover, for any sum F of bubble diagram functions F^B one can use (3.3) to give an expansion of F in terms of bubble diagram functions F^{B^-}, and then use the same procedure to define a separation of F into two parts,

$$F = F_g + F^g, \tag{3.6}$$

where F_g corresponds to a sum of B^- each of which has an explicit g-channel cut set, and F^g corresponds to a sum of B^- none of which has an explicit g-channel cut set. In particular, one can write

$$T_h = T_{hg} + T_h^g \tag{3.7}$$

and

$$T^h = T^h_g + T^{hg}. \tag{3.8}$$

These definitions entail that $T_{gh} = T_{hg}$: Both symbols represent the sum of F^{B^-} over those B^- that have both an explicit g-channel cut set and explicit h-channel cut set. In a similar way one can define T_{ghk}, T_{ghkm}, etc., all of which are independent of the order of their subscripts. From this symmetry property it follows that all of the various functions $T_H{}^G$ with both upper and lower indices are independent of the order of their indices. For example, (3.7) gives

$$T_h{}^g = T_h - T_{hg} \tag{3.9a}$$

while (3.6) gives

$$T^g{}_h = (T - T_g)_h = T_h - T_{gh}, \tag{3.9b}$$

so that $T_h{}^g = T^g{}_h$.

An expansion formally similar to (2.5) follows directly from (3.6). For example,

$$\begin{aligned} T^{gh} &= (T - T_g)^h = T^h - T_g{}^h \\ &= T - T_h - T_g + T_{gh}. \end{aligned} \tag{3.10}$$

More generally, one finds

$$T^G = \sum_{H \subset G} (-1)^{n(H)} T_H. \tag{3.11a}$$

The functions $\bar{T}^G$ and $\bar{T}_H$ defined by $\bar{T}^G = -(T^G)^\dagger$ and $\bar{T}_H = -(T_H)^\dagger$ satisfy

$$\bar{T}^G = \sum_{H \subset G} (-1)^{n(H)} \bar{T}_H. \tag{3.11b}$$

IV. ANALYTIC PROPERTIES OF T^G AND $\bar{T}^G$

A. The structure theorem

The proofs of analytic properties will be based on a theorem that specifies the analytic properties of an arbitrary bubble-diagram function F^B. This theorem has been described in detail in Ref. 12. A resumé of its main content is given here.

The main assumption of the theorem is a set of physical-region analyticity properties called the normal analytic structure. This analytic structure is equivalent to the S-matrix macrocausality condition.[13] It is also a

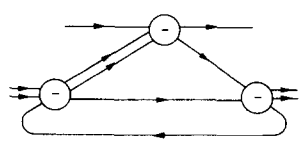


FIG. 3.1. A diagram that is not a bubble diagram.

formal property of the perturbation theory expression for any scattering function. Roughly it is the property that the singularities of the physical scattering function are confined to positive-α Landau surfaces, and that in some real neighborhood of almost any point p on any one of these surfaces the physical scattering function is the limit of an analytic function from directions lying in a certain cone. The rules specifying these directions are called the plus $i\epsilon$ rules.

The Landau surfaces are surfaces associated with diagrams called Landau diagrams. A Landau diagram is a topological diagram consisting of a set of directed line segments L_j and a set of point vertices V_r. The topological structure of the diagram is specified by a set of structure coefficients ϵ_{jr} defined as follows:

$$\epsilon_{jr}=\begin{cases}+1 & \text{if } L_j \text{ terminates at } V_r\\ -1 & \text{if } L_j \text{ originates at } V_r\\ 0 & \text{otherwise.}\end{cases} \tag{4.1}$$

The lines of the diagram are classified as incoming, outgoing, or internal according to the rule:

$$L_j \text{ is } \begin{cases}\text{incoming if } \epsilon_{jr}\geqslant 0 \text{ for all } r\\ \text{outgoing if } \epsilon_{jr}\leqslant 0 \text{ for all } r\\ \text{internal otherwise.}\end{cases}$$

The incoming and outgoing lines are collectively called external lines.

Each internal and external line L_j of a Landau diagram D is associated with a physical particle (of positive mass $m_j>0$), and with a momentum–energy 4-vector p_j. Each internal line L_j is associated also with a scalar parameter α_j. The Landau equations corresponding to D are the mass-shell constraints

$$p_j^2-m_j^2=0 \quad (\text{all } j), \tag{4.2a}$$

the conservation-law constraints

$$\sum_j p_j\epsilon_{jr}=0 \quad (\text{all } r), \tag{4.2b}$$

the Landau loop equations

$$\sum_{j\in \text{Int}} \alpha_j p_j \eta_{jl}=0 \quad (\text{all } l\in L), \tag{4.2c}$$

and the nontriviality condition

$$\sum_{j\in \text{Int}} \alpha_j^2-1=0. \tag{4.2d}$$

The set Int is the set of indices that label the internal lines of D, L is the set of indices that label the closed loops that can be constructed on the internal lines of D, and η_{jl} is the number of times loop l passes along line L_j in the positive direction minus the number of times loop l passes along line L_j in the negative direction. (It is sufficient to consider a set of linearly independent loops; then the η_{jl} can be restricted to ± 1, and zero.)

Let p represent the set of momentum–energy vectors p_j corresponding to the *external* lines of D. Then the complex Landau surface $L(D)$ is the set of complex points p such that for some choice of the complex p_j and α_j associated with the internal lines L_j of D the Landau equations corresponding to D can all be satisfied.

To describe the real Landau surface corresponding to D let a sign σ_j be introduced for each internal line L_j of D. The set of signs σ_j is denoted by σ. Then the real Landau surface $L(D^\sigma)$ is the set of points p that satisfy the above Landau equations and also the conditions

$$\text{Im}p_j=0, \quad \text{all } j, \tag{4.2e}$$

$$\text{Im}\alpha_j=0, \quad \text{all } j\in \text{Int}, \tag{4.2f}$$

$$p_j^0>0, \quad \text{all } j, \tag{4.2g}$$

and

$$\sigma_j\alpha_j>0, \quad \text{all } j\in \text{Int}. \tag{4.2h}$$

Equations (4.2a) through (4.2h) are called the Landau equations corresponding to D^σ.

The sign σ_j is allowed to be $\pm$, in which case the corresponding equation (4.2h) is eliminated from the equations that define $L(D^\sigma)$.

The symbol D^+ represents a D^σ with all $\sigma_j=+$. The corresponding Landau surface $L(D^+)$ is called a positive-α Landau surface. The union of all $L(D^+)$ is called L^+.

The Landau equations corresponding to D^+ have a simple physical interpretation.[14,15] The significance of the mass-shell and conservation-law constraints is obvious. The significance of the Landau loop equations is this: they ensure that the "displacement" vectors

$$\Delta_i=\alpha_i p_i \tag{4.3}$$

fit together to form a geometric diagram in a four-dimensional space. This geometric diagram has the topological structure specified by D^+, and it can be interpreted as a space–time diagram representing a possible classical multiple-scattering process in which point particles scatter at point vertices. The conditions (4.2e) through (4.2h) ensure that positive energy is carried forward in time on each leg of this multiple-scattering process. The parameter α_j is the proper time associated with L_j, divided by the mass m_j. This geometrical interpretation of the solutions of the positive-α Landau equations makes it clear that $L(D^+)$ can be nonempty only if the vertices of D^+ can be partially ordered by the condition that each internal line L_j of D^+ point from left to right.

These space–time diagrams are called *space–time representations of* D^σ. For every solution of the Landau equations corresponding to D^σ there is an associated space–time representation of D^σ. The vectors from an arbitrary origin to the vertices V_r of the space–time representation are denoted by ω_r. The line L_i is represented by the four-vector

$$\Delta_i=\sum_r \epsilon_{ir}\omega_r. \tag{4.4}$$

One of the conclusions of the structure theorem is an $i\epsilon$ rule of continuation for bubble diagram functions. This rule is a simple generalization of the plus $i\epsilon$ rule, and it is most easily described by first describing the plus $i\epsilon$ rule itself.

Let $\omega=(\omega_1,\cdots,\omega_m)$ represent the set of ω_r corresponding to a space–time representation of some D^σ. Let D^+ be some positive-α diagram, and let p be some

point on $L(D^+)$. Let $\Omega(D^+,p)$ be the union of all ω that correspond to space–time representations of D^+ having external lines specified by the set of (external) variables p. Let $\Omega^+(p)$ be the union of the sets $\Omega(D^+,p)$ over all positive-α diagrams.

The plus $i\epsilon$ rules can be stated in terms of a cone $C^+(p)$ that is closely related to $\Omega^+(p)$. Let $q \equiv (q_1, q_2, \cdots, q_6)$ represent the imaginary part of the complexification of $p \equiv (p_1, p_2, \cdots, p_6)$. Then

$$C^+(p) \equiv \left\{ q : -\sum_{\substack{j\in \mathrm{Ext} \\ r\in \mathrm{Ver}}} \epsilon_{jr} q_j \cdot \omega_r > 0 \ \text{ for all } \omega \in \Omega^+(p) \right\}. \quad (4.5a)$$

Here Ext is the set of indices j that label the components p_j of $p = (p_1, \cdots, p_6)$, and Ver is the set of indices r of the components ω_r of ω. By virtue of the Landau equations the cone $C^+(p)$ can be written in the alternative form[16]

$$C^+(p) = \left\{ q : \sum_{j\in \mathrm{Int}} q_j \cdot \Delta_j(\omega) > 0 \ \text{ for all } \omega \in \Omega^+(p) \right\}, \quad (4.5b)$$

where $\Delta_j(\omega)$ is defined by (4.4), Int is the set of indices labelling internal lines of the diagram specified by ω, and the q_j for $j \in$ Int are *any* set of 4-vectors that satisfy for every r the momentum–energy conservation law constraints $\sum q_j \epsilon_{jr} = 0$, where the external q_j are fixed by q.

The plus $i\epsilon$ rule says this: Let p be any point of $\mathcal{M}$, which is the real mass shell restricted by momentum–energy conservation. Let $C(p)$ be any cone that is closed apart from its missing apex $q = 0$, and that is contained in $C^+(p')$ for all points p' in some real neighborhood of p. Let $\mathcal{M}_c$ be the complex mass shell, restricted by momentum–energy conservation, and suppose the intersection of $\mathcal{M}_c$ with $\{q \in C(p)\}$ has p on its boundary. Then p has a real neighborhood $\mathcal{N}(p) \subset \mathbf{R}^{4n}$ such that the physical scattering function in $\mathcal{N}(p) \cap \mathcal{M}$ is the boundary value (in a distribution sense) of a function that is analytic[15] in the set

$$\mathcal{M}_c \cap \{p' \in \mathcal{N}(p)\} \cap \{q \in C(p)\} \cap \{q \in N\}$$

where $N \subset \mathbf{R}^{4n}$ is some neighborhood of $q = 0$. ($\mathbf{R}^{4n}$ is the real $4n$-dimensional space, and n is the number of particles, which is six in the case under consideration.) This statement of the $i\epsilon$ rule will be used presently.

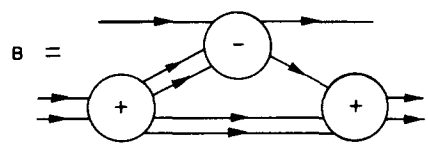
The first main conclusion of the structure theorem[17] is this*:

(1) Let B be any bubble diagram, and let $F^B(p)$ be the corresponding bubble diagram function. Then $F^B(p)$ is analytic at all real (mass-shell) points p not lying on

$$L(B) \equiv \bigcup_{D^\sigma \subset B} L(D^\sigma), \quad (4.6)$$

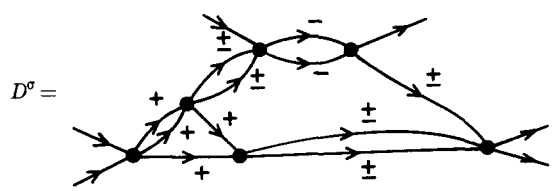
where D^σ is *contained* in B (i.e., $D^\sigma \subset B$) if and only if D^σ can be constructed by replacing each plus bubble b of B by either a point vertex V_b or a connected positive-α Landau diagram D_b^+ such that $L(D_b^+) \neq \phi$, and by replacing each minus bubble b of B by either a point vertex V_b or a connected negative-α Landau diagram D_b^- such that $L(D_b^-) \neq \phi$. The initial and final lines of b are to match the incoming and outgoing lines of D_b. Thus each line L_j of $D^\sigma \subset B$ is either an internal line of a D_b^+ or D_b^-, in which case it carries the sign $\sigma_j = +$ or $\sigma_j = -$, respectively, or it is a line L_j of the original bubble diagram B. In this latter case this line is assigned the sign $\sigma_j = \pm$, which means that the corresponding parameter α_j can be positive, negative, or zero. These lines of B itself are sometimes called *explicit* lines.

An example of a $D^\sigma \subset B$ is given by



(4.7a)

and



(4.7b)

The internal lines of each D_b^+ and D_b^- are drawn so as to lie inside the corresponding bubble b.

The set $\Omega(D^\sigma,p)$ is defined, in analogy to $\Omega(D^+,p)$, as the set of all ω that correspond to space–time representations of D^σ that correspond to solutions at p of the Landau equations corresponding to D^σ. And $\Omega^B(p)$ is defined in analogy to $\Omega^+(p)$ as

$$\Omega^B(p) \equiv \bigcup_{D^\sigma \subset B} \Omega(D^\sigma, p). \quad (4.8)$$

Finally, $C^B(p)$ is defined, in analogy to $C^+(p)$, as

$$C^B(p) \equiv \left\{ q : -\sum_{\substack{j\in \mathrm{Ext} \\ r\in \mathrm{Ver}}} \epsilon_{jr} q_j \cdot \omega_r > 0 \ \text{ for all } \omega \text{ in } \Omega^B(p) \right\} \quad (4.9a)$$

$$= \left\{ q : \sum_{j\in \mathrm{Int}} q_j \cdot \Delta_j(\omega) > 0 \ \text{ for all } \omega \text{ in } \Omega^B(p) \right\}. \quad (4.9b)$$

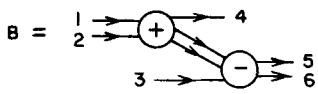
The second main consequence of the structure theorem is this:

(a) The functions $F^B(p)$ satisfy an $i\epsilon$ rule that is the same as the plus $i\epsilon$ rule described above, except that $C^B(p)$ replaces $C^+(p)$.

If $C^B(p)$ is empty then the $i\epsilon$ rule is devoid of content: No assertion about analyticity properties of F^B at p is made.

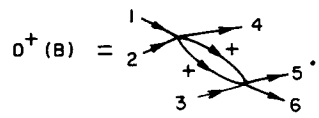
An important case where $C^B(p)$ is empty is the case in which p lies on $L(D^+(B))$, where $D^+(B)$ is the positive-α diagram obtained by contracting all the bubbles of B to point vertices. For example, if

*The contraction condition occurring in the statement of the theorem given in Ref. 17 is here replaced by the rule that condition (4.2h) is relaxed for lines L_j that are explicit lines of B itself. This new version is slightly stronger than the original version, but follows from essentially the same argument.



(4.10a)

then



(4.10b)

To see that $C^B(p)$ is empty in this case note that if p lies on $L(D^+(B))$, then it also lies on $L(D^-(B))$, because the Landau equations are invariant under the transformation $\alpha_j \to -\alpha_j$, $\sigma_j \to -\sigma_j$. But the change $\alpha_j \to -\alpha_j$ reverses all vectors Δ_j, $j \in \mathrm{Int}$. But then the $C^B(p)$ defined in (4.9b) is empty, since both $D^+(B) \subset B$ and $D^-(B) \subset B$. Thus no analyticity properties are asserted for F^B at p on $L(D^+(B))$.

It is in fact well known[18] that F^B is identically zero on one side of the surface $L(D^+(B))$, but not (in general) on the other.

The example just given illustrates a simple way in which the analytic continuation of a sum F of bubble diagram functions can be blocked: For some Landau diagram D^σ the function F has a Landau singularity surface corresponding to both D^σ and $D^{-\sigma}$, where $D^{-\sigma}$ is D^σ with all the signs σ_j reversed. The $i\epsilon$ rules associated with $D^{-\sigma}$ are opposite to those associated with D^σ, and hence the structure theorem provides no way to continue the sum F past the surface $L(D^\sigma) = L(D^{-\sigma})$.

This situation in which a function can have singularity surfaces associated with both a diagram D^σ and also the associated diagram $D^{-\sigma}$ is the canonical situation in which continuation is blocked. However, for a full proof that continuation is never blocked one must also rule out the possibility that surfaces corresponding to topologically different diagrams conspire to block the continuation. This will be done in Sec. VI.

The proofs of Sec. VI depend on certain analyticity properties of the functions T^G and $\bar{T}^G$. These properties are derived in subsection D by combining the results described in this subsection with the properties of the functions T^G and $\bar{T}^G$ described in the next subsection.

B. Properties of the T^G and $\bar{T}^G$

In this subsection three properties of the functions T^G are described. These properties are satisfied by the formal expressions for these functions given in Sec. III, and can be considered to be the defining properties of these functions as will be discussed in the next subsection. The analytic properties of the functions T^G and $\bar{T}^G$ that are derived in Sec. IV.D follow from these properties alone, and hence apply, in particular, to the well-defined expressions for these functions given in Sec. V.

The three properties of the T^G are now described.

Property 1: Each T^G can be written in the form

$$T^G = \sum_{H \subset G} (-1)^{n(H)} T_H \tag{4.11a}$$

where the T_H are a set of 2^{16} functions that can be written in the form

$$T_H = \sum_{B \in \mathcal{B}_H} F^B \tag{4.11b}$$

where every B in $\mathcal{B}_H$ has an explicit h-channel cut set for every h in H. [These cut sets are defined above (3.5).] Property 1 follows, for the infinite series expressions, from (3.11a).

Property 2: Each T^G can be converted solely by means of the unitarity and cluster decomposition properties of S from the form given in (4.11) to the form

$$T^G = \sum_{B \in \mathcal{B}^G} F^B, \tag{4.12}$$

where no $D^\sigma \subset \mathcal{B}^G$ has a positive-α g-channel cut set for any g in G. Here the following definitions are used:

Definition: A positive-α g-channel cut set of D^σ is a set of internal lines of D^σ such that D^σ is separated by the cutting precisely once of every line of this set into two connected diagrams $D^\sigma(J_g)$ and $D^\sigma(\bar{J}_g)$ such that $D^\sigma(J_g)$ contains all the external lines $j \in J_g$ and $D^\sigma(\bar{J}_g)$ contains all the external lines $j \in \bar{J}_g$. Moreover, each line L_j of the cut set must be directed from $D^\sigma(\bar{J}_g)$ to $D^\sigma(J_g)$ and have a sign $\sigma_j = +$ or $\pm$.

Definition: A $D^\sigma \subset \mathcal{B}$ is a D^σ such that for some B the conditions $D^\sigma \subset B \in \mathcal{B}$ hold.

Remark: For any collection $\mathcal{B}$ of diagrams B^- having only minus bubbles the statement no $D^\sigma \subset \mathcal{B}$ has a positive-α g-channel cut set is equivalent to the statement no $D^\sigma \subset \mathcal{B}$ has an explicit g-channel cut set. This is because only explicit lines of B^- can belong to a positive-α g-channel cut set. Thus Property 2 follows from the meaning given in (3.5) and (3.6) of superscripts.

Property 3:

$$T_\phi = T^\phi = T = M. \tag{4.13}$$

Remark: In equations such as (4.11a) and (4.12) involving sums of functions F^B over sets $B \in \mathcal{B}$ it is to be understood that the diagrams B of $\mathcal{B}$ can eventually have signs or other numerical coefficients, and that the functions F^B inherit these coefficients.

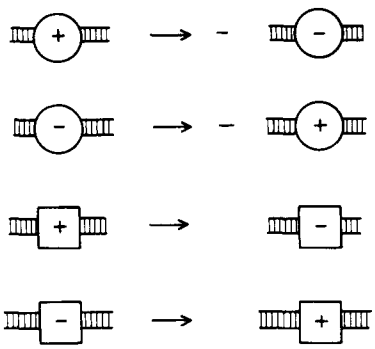
The functions $\bar{T}^G$ and $\bar{T}_H$ are defined, in accordance with (2.16), by

$$\bar{T}^G = -(T^G)^\dagger = \sum_{B \in \bar{\mathcal{B}}^G} F^B \tag{4.14a}$$

and

$$\bar{T}_H = -(T_H)^\dagger = \sum_{B \in \bar{\mathcal{B}}_H} F^B \tag{4.14b}$$

where $\bar{\mathcal{B}}^G$ and $\bar{\mathcal{B}}_H$ are obtained from $\mathcal{B}^G$ and $\mathcal{B}_H$ by the mapping



(4.14c)

followed by an overall sign change. The last two mappings in (4.14c) apply also to the special boxes defined by (2.9n), (2.12a), (2.12b), and (2.12c).

Negative-α g-channel cut sets are defined in the same way as positive-α g-channel cut sets [see (4.12)] except that "positive-α" is replaced by "negative-α," and $\sigma_j = +$ or $\pm$ is replaced by $\sigma_j = -$ or $\pm$. It is thus clear that no $D^{\sigma} \subset \bar{\beta}^G$ has a negative-α g-channel cut set for any g in G.

C. Uniqueness of the T^G

In Sec. V a set of 2^{17} well-defined functions T^G and T_H satisfying properties (1), (2), and (3) is constructed. The question of uniqueness arises: Can there be two different sets of 2^{17} functions T^G and T_H satisfying these three properties?

In this section it is shown that the functions T^G and T_H are unique in the following sense: Let T^G and T_H be members of any set of 2^{17} functions that satisfy properties (1), (2), and (3). Let T_-^G and T_H^- be the infinite series expressions obtained by introducing for each plus bubble of the expressions (4.12) and (4.11b) for T^G and T_H, respectively, the expansion (3.3), and then combining together the different terms that are multiples of each of the distinct possible minus-bubble diagram functions F^{B^-}. Then T_-^G and T_H^- are precisely the infinite sums that were represented in Sec. III by the symbols T^G, T_H. This result justifies the use of the same symbols T^G and T_H to represent, on the one hand, the infinite series expressions defined in Sec. III, and, on the other hand, the finite expressions that are obtained in Sec. V.

For a complete proof of uniqueness one should, strictly speaking, show that two different well-defined expressions in terms of bubble-diagram functions that have the same expression in terms of minus bubble diagram functions are in fact equal. This can probably be done. However, it is not necessary for our purposes. All that we need is *some* set of well-defined functions that satisfy the properties (1), (2), and (3), and these functions will be taken to be the functions defined in Sec. V. We doubt that others exist, but the uniqueness of the M^G is based in any case on the generalized Steinmann relations.

Because uniqueness is not really essential to our argument the proof of it will be simply sketched, rather than presented in full detail.

If the infinite series expansion (3.3) obtained from unitarity and the cluster decomposition is introduced back into the unitarity equation one finds that it is identically satisfied; i.e., if (3.3) is introduced into $SS^{\dagger} - I = 0$, then one obtains the result

$$\sum_{B^- \in \beta_{\phi}^-} F^{B^-} = 0 \tag{4.15}$$

where β_{ϕ}^- is empty. This means that if any function

$$F = \sum_{B \in \beta^F} F^B \tag{4.16a}$$

is converted by means of (3.3) to an infinite series

$$F = \sum_{B^- \in \beta^{-F}} F^{B^-}, \tag{4.16b}$$

then β^{-F} is invariant under a change in form of F generated by the application of unitarity; i.e., if F and F' are equal by virtue of unitarity and the cluster decomposition of S, then

$$\beta^{-F} = \beta^{-F'}. \tag{4.16c}$$

This is true because (4.15) implies that any identity among bubble diagram functions that follows from unitarity and cluster properties is identically satisfied when all the components are expanded in terms of F^{B^-} functions; i.e., unitarity acts as the identity in the minus-bubble representation.

Let T^G and T_H be members of a set of 2^{17} functions that satisfy (1), (2), and (3). The formula (4.11a) for T^G entails (see Appendix B) that the functions $T_H{}^G$ defined for $G \cap H = \phi$ by

$$T_H{}^G = \sum_{G' \subset G} (-1)^{n(G')} T_{HG'} \tag{4.17a}$$

satisfy

$$T_H{}^G = \sum_{H' \subset H} (-1)^{n(H')} T^{GH'} \tag{4.17b}$$

and also

$$T_H{}^G = \sum_{\substack{K', K'' \\ K' \cup K'' = K \\ K' \cap K'' = \phi}} T_{HK''}^{GK'} \tag{4.17c}$$

where in (4.17c) the sum is over all nonintersecting sets K' and K'' whose union is any fixed set K that does not intersect $G \cup H$.

A special case of (4.17c) is

$$T^G = \sum_{K' \subset \bar{G}} T_{GK'}^{GK'}. \tag{4.18}$$

This gives T^G as a sum over terms of the form $T_{\bar{K}}{}^K$. Equations (4.17a) and (4.17b) give

$$T_{\bar{K}}{}^K = \sum_{G' \subset K} (-1)^{n(G')} T_{\bar{K}G'} \tag{4.19a}$$

and

$$T_{\bar{K}}{}^K = \sum_{H' \subset \bar{K}} (-1)^{n(H')} T^{KH'}. \tag{4.19b}$$

The equality of the right-hand sides of (4.19a) and

(4.19b) follows from (4.11a) alone: It is identically satisfied if (4.11a) is introduced into (4.19b). According to Property (2) the equality of the expressions (4.11) and (4.12) for T^G follows just from unitarity and the cluster decomposition of S. Thus, according to (4.16c), Eq. (4.11a) with T^G and T_H replaced by $T_-{}^G$ and T_H^-, respectively, is identically satisfied: Both sides have the same minus-bubble representation. But then the replacements of the T^G and T_H on the right-hand sides of (4.19b) and (4.19a) by $T_-{}^G$ and T_H^-, respectively, must yield the same answer: Both procedures must give the same formal expression

$$T_{\bar{K}}^{K} = \sum_{B^- \in \beta_{\bar{K}}^{-K}} F^{B^-}, \tag{4.20}$$

where $\beta_{\bar{K}}^{-K}$ is some well-defined set of B^-.

Let β^{-G} be the set of B^- obtained by the expansion of the $B \in \beta^G$ in terms of B^-'s. And let β_H^- be the set of B^- obtained from the series expansion of β_H in terms of B^-'s. The characteristic properties of β^G and β_H are not destroyed by the series expansion. That is, no $D^\sigma \subset \beta^{-G}$ can contain a positive-α g-channel cut set for any g in G, and every $B \in \beta_H^-$ contains an explicit h-channel cut set for every h in H. Moreover, the condition that no $D^\sigma \subset \beta^{-G}$ contains a positive-α g-channel cut set is equivalent to the condition no $B \in \beta^{-G}$ contains an explicit g-channel cut set, since all lines L_j of any B^- with $\sigma_j = +$ or $\pm$ are explicit lines of B^-. Thus, by virtue of (4.19), $\beta_{\bar{K}}^{-K}$ contains a B^- only if B^- has no explicit g-channel cut set for any g in K, and has an explicit h-channel cut set for every h in $\bar{K}$. But then any given B^- can be contained in one and only of the sets $\beta_{\bar{K}}^{-K}$, namely the one such that $\bar{K}$ is the set of all g such that B^- contains an explicit g-channel cut set.

Each B^- must in fact be contained exactly once (with coefficient plus one) in the union of all $\beta_{\bar{K}}^{-K}$. This follows from (3.3), (4.13), and (4.18) for the special case $G = \phi$. Thus each $\beta_{\bar{K}}^{-K}$ is, by virtue of (4.11)–(4.13), exactly the set of all B^- such that B^- contains an explicit g-channel cut set for every g in $\bar{K}$, and contains no explicit g-channel cut set for any g in K. That is, each $\beta_{\bar{K}}^{-K}$ is uniquely defined by the conditions (4.11)–(4.13). Thus, by virtue of (4.18) and (4.20), every $T_-{}^G$ is also uniquely defined by the defining properties (1), (2), and (3) of T^G. The T_H^- are uniquely defined by (4.20) and

$$T_H = \sum_{K' \subset \bar{H}} T_{HK'}^{\bar{H}K'}. \tag{4.21}$$

D. Analytic properties of T^G and $\bar{T}^G$

For any g in G consider the Landau surface $L(D_g^+) = L(D_g^-)$ corresponding to the pair (D_g^+, D_g^-) of g-channel normal-threshold diagrams. (See Fig. 4.1.) The diagram D_g^+ has a positive-α g-channel cut set. Hence property (4.12) implies that no D_g^+ satisfies $D_g^+ \subset \beta^G$. If only normal threshold diagrams need to be considered, then the generalized $i\epsilon$ rule stated below (4.9) says that the function T^G continues into itself around $L(D_g^+) = L(D_g^-)$ by the rule associated with the negative-α diagram D_g^-. This rule is the minus $i\epsilon$ rule, which prescribes a detour into the lower-half s_g plane.

This argument extends immediately to a large class of singularities: Property (4.12) excludes from T^G all singularities associated with diagrams D^σ that can be contracted to any positive-α diagram D_g^+, for any g in G.

By a similar argument $\bar{T}^G$ can have no singularities associated with diagrams that can be contracted to any D_g^- for any g in G. If $T^G = \bar{T}^{E-G}$, then this function has no singularities associated with any diagram D^σ having a positive-α g-channel cut set for any g in G, or having a negative-α g-channel cut set for any g in $E-G$.

This property is the basis of the proof in Sec. VI of the analyticity properties of the 26 018 boundary values M^G. Before giving that proof we shall, in the following section, construct finite representations for the functions T^G and $\bar{T}^G$, and show that $T^G = \bar{T}^{E-G}$ for the values of G defined by (2.25).

V. CONSTRUCTION OF THE T^G

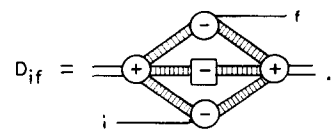
In this section a set of functions T^G is exhibited that satisfies the three properties listed in Sec. IV. The procedure is to make an ansatz for the functions T_H, and then to show that these three properties hold. The ansatz is that

$$T_H = M_H \tag{5.1a}$$

in all cases except those given by the formulas

$$T_{(if)t} = T_{i(if)t} = T_{(if)ft} = T_{i(if)ft} = D_{if}. \tag{5.1b}$$

The functions M_H are defined by (2.7)–(2.14), and the functions D_{if} are defined by



(5.1c)

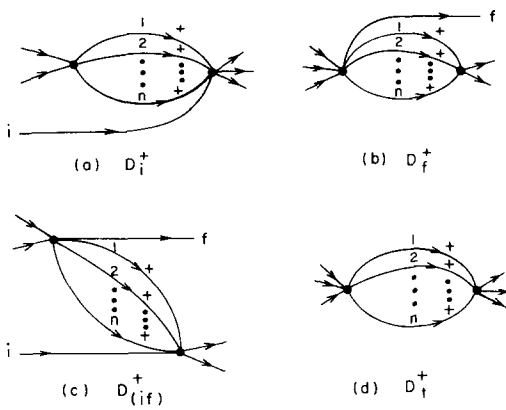


FIG. 4.1. All connected positive-α Landau diagrams with 3 incoming lines, 3 outgoing lines, and 2 vertices are shown. Line i is any one of the initial lines 1, 2, 3, and line f is any one of the final lines 4, 5, 6. The plus signs on the internal lines indicate that the corresponding Landau α's are positive. The number of the internal lines n is an arbitrary positive integer. The 6 diagrams of (a) and (b), the 9 diagrams of (c), and the diagram of (d) are called subenergy diagrams, cross-energy diagrams, and total-energy diagram, respectively. The connnected negative-α g-channel diagram D_g^- is obtained from D_g^+ by simply changing all the plus signs σ_j to minus signs.

Inspection of (5.1) and (2.11)–(2.14) shows that (4.11b) is satisfied. Thus if T^G is defined by (4.11a) then property (1) is satisfied. Property (3), i.e., Eq. (4.13), follows from (4.11) and (5.1). Thus it only remains to prove Property (2), which is that the expression for T^G given in (4.11) can be converted solely by means of unitarity and cluster properties to the form (4.12).

The conditions imposed by (4.12) can be compactly stated with the aid of the following:

Definition: A function F is said to belong to $\mathcal{R}^G$ if and only if F can be expressed, using only unitarity and cluster properties, in the form

$$F=\sum_{B\in\mathcal{B}^F} F^B \tag{5.2}$$

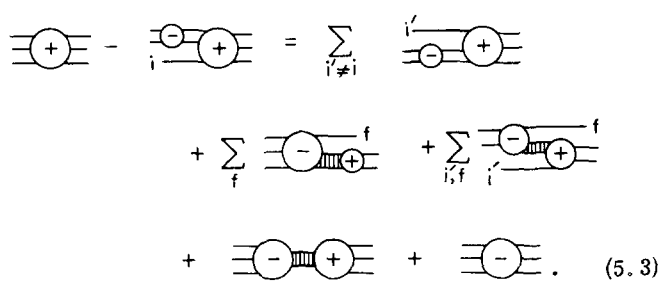
where no $D^\sigma\subset\mathcal{B}^F$ has a positive-α g-channel cut set for any g in G.

In terms of this definition (4.12) is the requirement that for every G the function T^G belongs to $\mathcal{R}^G$.

This property (4.12) must be proved for each of the 2^{16} possible sets G. The sixteen special cases in which G consists of a single element $G\in E$ are covered by

Proposition 5.1: $T^g\equiv T-T_g$ belongs to $\mathcal{R}^g$.

Proof: *Case* 1 $g=i$. Unitarity (2.9d), and the cluster decomposition formulas (2.9i), (2.9j), and (2.9m) give

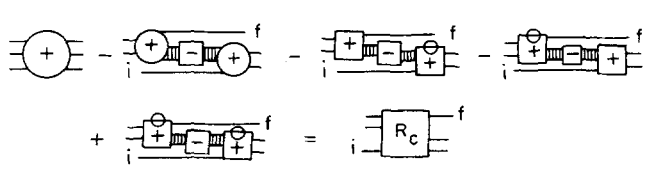


(5.3)

the left-hand side of (5.3) is $T-T_i$. The right-hand side is the required expression for T^i, for it is clear by inspection† that every term in this expression belongs to $\mathcal{R}^i$.

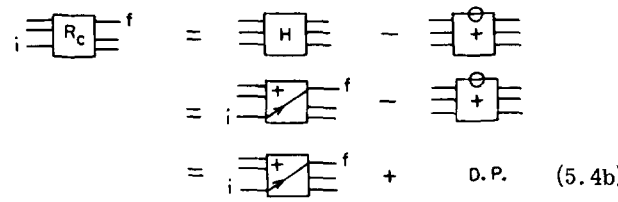
Case 2 $g=f$. The proof is essentially the same as for Case 1: One merely uses the alternative form (2.9c) of unitarity, in which the minus box appears on the right-hand side instead of the left-hand side.

Case 3 $g=(if)$. Equation (5.9) of Ref. 9, specialized to the present case, reads



where the function represented by the R_c box appearing on the right-hand side belongs, as explained below, to $\mathcal{R}^{(if)}$. The last three terms on the left-hand side of (5.4a) are disconnected, by virtue of (2.9o) and (2.9p), and hence cancel the disconnected terms on the right-hand side. This gives the required result.

The fact that the function represented by the R_c box in (5.4a) belongs to $\mathcal{R}^{(if)}$ follows from results of Ref. 9. Comparison of Eqs. (5.6), (5.8), and (5.9) of Ref. 9 shows that



(5.4b)

where D.P. is a sum of disconnected parts and H is a sum of F^B over a set $\mathcal{B}$ of B with the property that for any $D^\sigma\subset\mathcal{B}$ the line i can be connected to line f by a directed path from i to f that consists of a sequence of L_j having the property that every one of these L_j with $\sigma_j=+$ or $\pm$ points in the direction of this path from i to f. This directed path is indicated in the second line of (5.4b). (The lines L_j with $\sigma_j=-$ can be ignored.) The existence of such a path is implied by (5.7) of Ref. 9. It ensures that no $D^\sigma\subset\mathcal{B}$ has a positive-α (if)-channel cut set.

Case 4. $g=t$: Let A, B, and C denote certain bubble-diagram functions, and let A and C each be decomposed into a sum of two bubble-diagram functions so that

$$A=A'+A'' \tag{5.5a}$$

and

$$C=C'+C''. \tag{5.5b}$$

Then one finds, trivially, that

$$ABC=A'BC+ABC'-A'BC'+A''BC''. \tag{5.6}$$

Here the product form indicates the usual product of bubble-diagram functions so that, for example, in ABC the outgoing lines of the bubble diagram corresponding to A are identical with the ingoing lines of the bubble diagram corresponding to B.

Consider the special case of (5.6) where A and C each represent the plus (minus) box and where B represents the minus (plus) box. Unitarity [Eqs. (2.9c, d)] takes the form

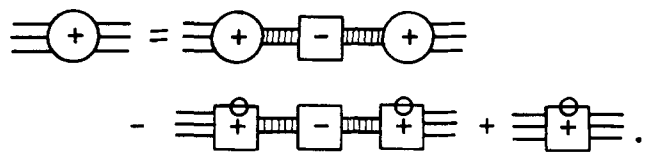
$$AB=BC=I. \tag{5.7}$$

Substituting (5.7) into (5.6) one obtains

$$A=A'+C'-A'BC'+A''BC''. \tag{5.8}$$

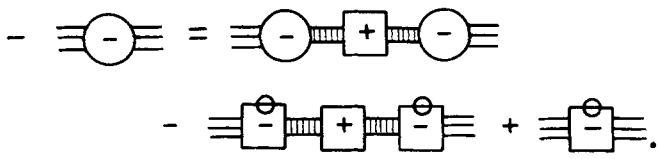
Suppose, specifically, that in (5.8) A and C represent the plus box and that B represents the minus box. Also let A' and C' each denote the circled plus box and let A'' and C'' each denote the plus bubble [so that Eqs. (5.5a) and (5.5b) are a form of (2.9n).] Then (5.8) takes the form

†No $D^\sigma\subset B$ can have a positive-α g-channel cut set if any minus bubble of B touches external lines from both J_g and $\bar{J}_g$. And no $D^\sigma\subset B$ can have a positive-α g-channel cut set if B is not a connected diagram. This remark will be referred to repeatedly in the arguments that follow by the use of the dagger symbol.



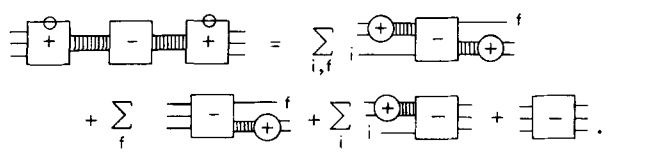
(5.9a)

Similarly, one finds



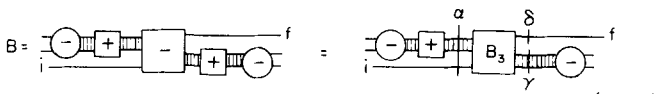
(5.9b)

Equations (2.9n), (2.9i), and (2.9j) show that



(5.10)

It is clear by inspection† that the last three terms of (5.10) belong to $\mathcal{R}^t$. The first term can be written with the aid of (2.9k, l) in the form



(5.11)

where the vertical lines α, γ, and δ cut through the sets of lines that are to be identified with the sets of lines labelled by α, γ, and δ in (6.1) of Ref. 9, respectively, and β is identified with the empty set.

Suppose that B is replaced by an equivalent B' as is explained in Corollary 6.1 of Ref. 9. Let D^σ be any $D^\sigma \subset B'$ and suppose that D^σ has a positive-α t-channel cut set. Then the right end point of every line of γ must lie in $D^\sigma(J_t)$. This is true because no line of a positive-α cut set of D^σ can be an outgoing line of the right minus bubble or an internal line of a negative-α Landau diagram D_b^- corresponding to this minus bubble. Similarly, line δ is in $D^\sigma(T_t)$. Hence, all the conditions required by the last sentence of Corollary 6.1 of Ref. 9 are met. [To convert the notation of Ref. 9 to that of Case 4 of the present paper use the correspondence: $\omega = \omega_1 \to \{1, 2, 3\}$, $\omega' = \omega_2' \to \{4, 5, 6\}$, connected $\bar{X}(C, \omega') \to D^\sigma(J_t)$, connected $\bar{X}(C, \omega) \to D^\sigma(\bar{J}_t)$, simple positive (ω, ω') cut set with $\bar{X}(C, \omega')$ and $\bar{X}(C, \omega)$ connected − positive-α t-channel cut set.] Thus, if that part of D^σ that corresponds to B_3' (B_3' is equivalent to B_3) is denoted by $D_3{}^\sigma$, then all points of $D_3{}^\sigma - \alpha$ lie in $D^\sigma(J_t)$. This means that the part of D^σ that corresponds to the part of B' lying to the right of the set of lines α must lie in $D^\sigma(J_t)$. But then the remaining part of D^σ must be disconnected, and the existence of a positive-α t-channel cut set is precluded.

Proposition 5.1 is a special case of

Proposition 5.2: The function

$$T^G = \sum_{H \subset G} (-1)^{n(H)} T_H \tag{5.12}$$

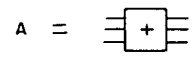
can be converted by means of the unitarity and cluster decomposition properties of S to the form

$$T^G = \sum_{B \in \beta^G} F^B, \tag{5.13}$$

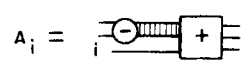
where no $D^\sigma \subset \beta^G$ has a positive-α g-channel cut set for any g in G.

That is, the T^G defined by (5.12), with the T_H defined by (5.1), belongs to $\mathcal{R}^G$.

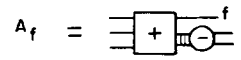
Proof: Several quantities that occur often in the proof are defined as follows:



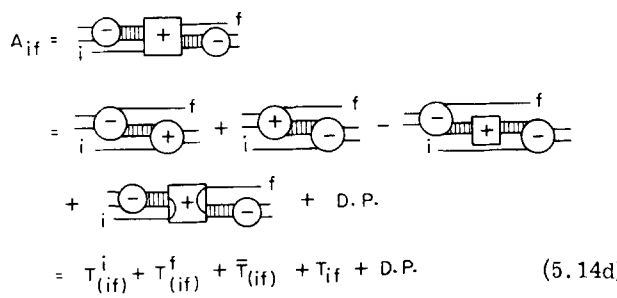
(5.14a)



(5.14b)



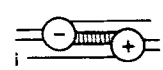
(5.14c)



(5.14d)

where D. P. stands, in general, for any sum of disconnected parts. The first two quantities appearing in the last line of (5.14d) are defined by

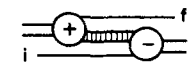
$$\begin{aligned} T^i_{(if)} &\equiv T_{(if)} - T_{i(if)} \\ &\equiv M_{(if)} - M_{i(if)} \\ &\equiv M^i_{(if)} = \end{aligned}$$



(5.14e)

where the calculation in (2.15b) is used, and

$$\begin{aligned} T^f_{(if)} &\equiv T_{(if)} - T_{(if)f} \\ &\equiv M_{(if)} - M_{(if)f} \\ &\equiv M^f_{(if)} = \end{aligned}$$



(5.14f)

where the calculation in (2.15a) is used. The second and third quantities in the last line of (5.14d) are defined by (5.1), (2.11c), (2.13a), and the definition

$$\begin{aligned} \bar{F}^B &= -(F^B)^\dagger \\ &= -(-1)^{N_b} F^B(+ - - -), \end{aligned} \tag{5.14g}$$

where N_b is the number of (explicitly appearing) bubbles in B and $F^B(+ - - -)$ is the bubble-diagram function for the bubble diagram $B(+ - - -)$ obtained from B by changing the sign inside each bubble, box, and modified box.

Each of the above equations can be converted to another one by the application of bars to each term. For example, from (5.14d) one obtains

$$\bar{A}_{if} = \bar{T}^i_{(if)} + \bar{T}^f_{(if)} + T_{(if)} + \bar{T}_{if} + \text{D. P.}. \tag{5.14h}$$

Using (5.14), (4.13), (2.10), and (5.1a), one may write (5.9a) and (5.10) in the form

$$T = T_t + \bar{A} - \sum_i \bar{A}_i - \sum_f \bar{A}_f + \sum_{i,f} \bar{A}_{if} + \text{D.P.}. \tag{5.15a}$$

From (5.9b) one obtains

$$\bar{T} = \bar{T}_t + A - \sum_i A_i - \sum_f A_f + \sum_{i,f} A_{if} + \text{D.P.}. \tag{5.15b}$$

For any G in E this equation can be written as

$$A - \sum_{i\in G} A_i - \sum_{f\in G} A_f + \sum_{i\in G, f\in G} A_{if} = F_1 \tag{5.16a}$$

where

$$F_1 \equiv \bar{T} - \bar{T}_t + \sum_{i\in\bar{G}} A_i + \sum_{f\in\bar{G}} A_f - \sum_{i\in\bar{G},f} A_{if} - \sum_{f\in\bar{G},i} A_{if} + \sum_{i\in\bar{G},f\in\bar{G}} A_{if} + \text{D.P.}. \tag{5.16b}$$

The unrestricted sums over i or f are sums over all three values of i or f. As mentioned earlier, the index i is always restricted in this paper to the values 1, 2, and 3, and the index f is restricted to 4, 5, and 6. A sum over $i\in G$ is a sum over those indices $i=g\in G$ that label initial subenergy channels, and the sum over $i\in\bar{G}$ is the sum over the remaining indices in the set (1, 2, 3).

From (5.14b), the cluster expansion (2.9j), and (5.14e), one obtains

$$A_i = \sum_f T^i_{(if)} + T_i + \text{D.P.}. \tag{5.17a}$$

Under the substitution $F \to \bar{F}$ this equation becomes

$$\bar{A}_i = -\sum_f T^f_{(if)} + \bar{T}_i + \text{D.P.}, \tag{5.17b}$$

where we have used (5.14e) and (5.14f). Similarly,

$$A_f = \sum_i T^f_{(if)} + T_f + \text{D.P.} \tag{5.18a}$$

and

$$\bar{A}_f = -\sum_i T^i_{(if)} + \bar{T}_f + \text{D.P.}. \tag{5.18b}$$

Substituting the right-hand sides of (5.17a), (5.18a), and (5.14d) into the second, third, and fourth terms on the left-hand side of (5.16a), respectively, one obtains

$$T - \sum_{i\in G} T_i - \sum_{f\in G} T_f + \sum_{i\in G,f\in G} T_{if} - \sum_{i\in G,f\in\bar{G}} T^i_{(if)} - \sum_{i\in\bar{G},f\in G} T^f_{(if)} + \sum_{i\in G,f\in G} \bar{T}_{(if)} = F_1. \tag{5.19}$$

The equation

$$-\bar{T}_{(if)} = T_{(if)} - T_{i(if)} - T_{(if)f} + T_{i(if)f} \tag{5.20}$$

was proved in Sec. II [see (2.21) and (2.22)]. This equation yields trivially

$$-\sum_{i\in G,f\in G} \bar{T}_{(if)} = \sum_{i\in G,f\in G,(if)\in G} [T_{(if)} - T_{i(if)} - T_{(if)f} + T_{i(if)f}] - \sum_{i\in G,f\in G,(if)\in\bar{G}} \bar{T}_{(if)}, \tag{5.21}$$

where the condition $(if)\in G$ or $(if)\in\bar{G}$ under the summation sign means that there is a sum over the pairs (i,f), and that this sum is to be restricted both by any other appearing conditions on i or f, and also by the condition that (if) be an element of G of $\bar{G}$, respectively. One also finds trivially from (5.14e) and (5.14f) that

$$\sum_{i\in G,f\in\bar{G}} T^i_{(if)} = \sum_{i\in G,f\in\bar{G},(if)\in G} [T_{(if)} - T_{i(if)}] + \sum_{i\in G,f\in\bar{G},(if)\in\bar{G}} T^i_{(if)}, \tag{5.22}$$

$$\sum_{i\in\bar{G},f\in G} T^f_{(if)} = \sum_{i\in\bar{G},f\in G,(if)\in G} [T_{(if)} - T_{(if)f}] + \sum_{i\in\bar{G},f\in G,(if)\in\bar{G}} T^f_{(if)}. \tag{5.23}$$

Substituting (5.21), (5.22), and (5.23) into (5.19), one obtains

$$T + F_2 = F_3 \tag{5.24}$$

where

$$F_2 \equiv -\sum_{i\in G} T_i - \sum_{f\in G} T_f + \sum_{i\in G,f\in G} T_{if} - \sum_{(if)\in G} T_{(if)} + \sum_{i\in G,(if)\in G} T_{i(if)} + \sum_{f\in G,(if)\in G} T_{(if)f} - \sum_{i\in G,f\in G,(if)\in G} T_{i(if)f} \tag{5.25a}$$

and

$$F_3 \equiv F_1 - \sum_{i\in\bar{G},f\in\bar{G},(if)\in G} T_{(if)} - \sum_{i\in G,f\in G,(if)\in\bar{G}} \bar{T}_{(if)} + \sum_{i\in G,f\in G,(if)\in\bar{G}} T^i_{(if)} + \sum_{i\in\bar{G},f\in G,(if)\in\bar{G}} T^f_{(if)}. \tag{5.25b}$$

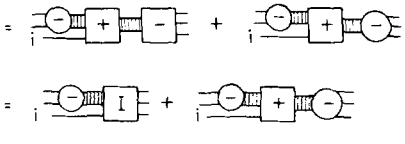
At this point it is convenient to consider separately two different cases

Case 1: $(t\in\bar{G})$. Suppose G does not contain t. Then (5.24) becomes

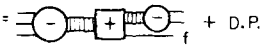
$$\sum_{H\subset G} (-1)^{n(H)} T_H \equiv T^G = F_3. \tag{5.26}$$

Indeed, the terms appearing on the left-hand side of (5.24) are identical to those appearing on the left-hand side of (5.26), by virtue of the fact that all the $T_H = M_H$ for $H\subset G$ not explicitly listed in (2.10), (2.11), (2.13), or (2.14) vanish. Thus the proof for Case 1 can be completed by showing that F_3 belongs to $\mathcal{R}^G$.

The cluster property (2.9j) and unitarity give

$A_j - \sum_f A_{jf}$ =  (5.27)

Similarly,

$A_f - \sum_i A_{if}$ =  (5.28)

After substituting the right-hand sides of (5.27) and (5.28) into (5.16b), one sees† that

$$F_1 - \sum_{i\in\bar{G},f\in\bar{G}} A_{if} \tag{5.29}$$

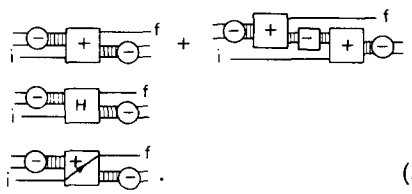
belongs to $\mathcal{R}^G$.

After replacing the plus bubbles in $T^f_{(if)}$ and in $T^f_{(if)}$ by the left-hand sides of (2.9k) and (2.9l) one sees† that the last three terms on the right-hand side of (5.25b) belong to $\mathcal{R}^G$. Hence, F_3 belongs to $\mathcal{R}^G$ if

$$\sum_{i\in\bar{G},f\in\bar{G}} A_{if} - \sum_{i\in\bar{G},f\in\bar{G},(if)\in G} T_{(if)} = \sum_{i\in\bar{G},f\in\bar{G},(if)\in\bar{G}} A_{if} + \sum_{i\in\bar{G},f\in\bar{G},(if)\in G} (A_{if} - T_{(if)}) \tag{5.30}$$

belongs to $\mathcal{R}^G$. One sees† from (5.14d) that this is true for the first term on the right-hand side of (5.30).

From (5.6) of Ref. 9 one obtains

$A_{if} - T_{(if)} =$ 

(5.31)

The H-box is the expression given in (5.7) of Ref. 9. It is a sum over a set β of bubble diagrams B such that any $D^\sigma \subset B \in \beta$ contains a path P from i to f that consists of lines L_j with the following property: Each line L_j with sign $\sigma_j = +$ or $\pm$ points in the direction of P, i.e., from i to f. This property is indicated by the internal line in the box appearing in the last line of (5.31). This line precludes the existence of a positive-α (if)-channel cut set in the $A_{if} - T_{(if)}$. Thus one sees† from (5.31) that the second term on the right-hand side of (5.30) belongs to $\mathcal{R}^G$. This completes the proof for Case 1.

Case 2: $(t \in G)$. Suppose G contains t. The equation

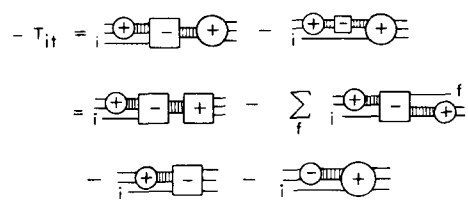
$$-\bar{T}_t = T_t - \sum_i T_{it} - \sum_f T_{ft} + \sum_{i,f} T_{itf} \tag{5.32}$$

is now needed.

Proof of (5.32): Equations (5.15a) and (5.15b) give

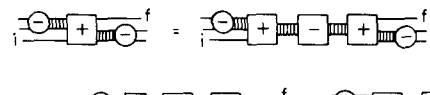
$$-\bar{T}_t = T_t - \sum_i \bar{A}_i - \sum_f \bar{A}_f + \sum_{i,f} \bar{A}_{if} - \sum_i A_i - \sum_f A_f + \sum_{i,f} A_{if} + \text{D. P.}. \tag{5.33}$$

Using (5.1), (2.13b), (2.12a), (2.9j), (2.9k), (5.14), and (2.11a), one obtains

$-T_{it} =$ 

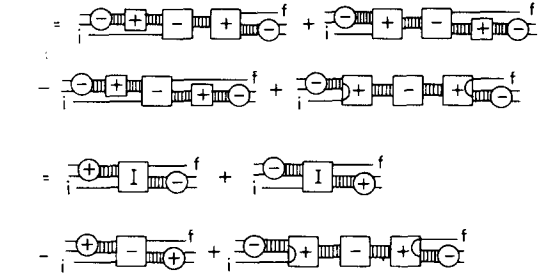
$$= \sum_f \bar{A}_{if} - \bar{A}_i - T_i + \text{D. P.} \tag{5.34}$$

Similarly,

$-T_{ft} =$ 

$$= \sum_i \bar{A}_{if} - \bar{A}_f - T_f + \text{D. P.} \tag{5.35}$$

By virtue of (2.9c), (5.6), (2.12a), (2.12b), (2.9k), and (2.9l) one has



(5.36)

so that, by (5.14), (5.1), and (2.14a),

$$T_{itf} = -T^i_{(if)} - T^f_{(if)} + A_{if} - \bar{A}_{if} + \text{D. P.}. \tag{5.37}$$

Combining (5.33), (5.34), (5.35), (5.37), and using (5.17a) and (5.18a) one obtains (5.32).

By virtue of (5.34), (5.35), and (5.37) one may write (5.32) in the form

$$-\bar{T}_t = -F_4 + \sum_{i\in\bar{G},f} \bar{A}_{if} - \sum_{i\in\bar{G}} \bar{A}_i - \sum_{i\in\bar{G}} T_i + \sum_{i,f\in\bar{G}} A_{if} - \sum_{f\in\bar{G}} \bar{A}_f - \sum_{f\in\bar{G}} T_f + \left(\sum_{i\in\bar{G},f} + \sum_{i,f\in\bar{G}} - \sum_{i\in\bar{G},f\in\bar{G}}\right)(-T^i_{(if)} - T^f_{(if)} + A_{if} - \bar{A}_{if}) + \text{D. P.} \tag{5.38}$$

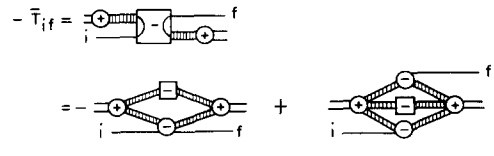
where

$$F_4 = -T_t + \sum_{i\in G} T_{it} + \sum_{f\in G} T_{ft} - \sum_{i\in G,f\in G} T_{itf}. \tag{5.39}$$

Consider the term $\sum_{i\in\bar{G},f\in\bar{G}} \bar{A}_{if}$ appearing in (5.38). Equations (5.14e, f, h) allow it to be written as

$$\sum_{i\in G,f\in\bar{G}} \bar{A}_{if} = \sum_{i\in\bar{G},f\in\bar{G},(if)\in\bar{G}} \bar{A}_{if} - \sum_{i\in\bar{G},f\in\bar{G},(if)\in G} (T^i_{(if)} + T^f_{(if)} - T_{(if)} - \bar{T}_{if}) + \text{D. P.}. \tag{5.40}$$

The term $-\bar{T}_{if}$ can be decomposed into two parts:

$-\bar{T}_{if} =$ 

$$= C_{if} + D_{if}. \tag{5.41}$$

The term C_{if} is the part of $-\bar{T}_{if}$ in which the two lines i and f touch a single minus bubble, and D_{if} is the part in which the two lines i and f do not touch a single minus bubble.

Substituting (5.40) into (5.38) gives an expression for $-\bar{T}_t$ that can be introduced into the expression (5.16b) for F_1. Substituting this new expression for F_1 into (5.25b) and using (5.17a), (5.17b), (5.18a), (5.18b), and (5.41) one finds that Eq. (5.24) takes the form

$$T + F_2 + F_4 + F_5 = F_6 \tag{5.42}$$

where

$$F_5 \equiv \sum_{i\in\bar{G}, f\in\bar{G}, (if)\in G} D_{if} \equiv D^G \tag{5.43}$$

and

$$\begin{aligned} F_6 \equiv \bar{T} &- \sum_{i\in\bar{G}} \bar{T}_i - \sum_{f\in\bar{G}} \bar{T}_f + \sum_{f\in\bar{G}, (if)\in\bar{G}} T^i_{(if)} \\ &+ \sum_{i\in\bar{G}, (if)\in\bar{G}} T^f_{(if)} - \sum_{i\in G, f\in G, (if)\in\bar{G}} \bar{T}_{(if)} \\ &+ \sum_{i\in\bar{G}, f\in\bar{G}, (if)\in\bar{G}} \bar{A}_{if} - \sum_{i\in\bar{G}, f\in\bar{G}, (if)\in G} C_{if} \\ &+ \text{D. P.}_{\circ} \end{aligned} \tag{5.44}$$

[The function D^G is defined by (5.43) only for the present case $t \in G$. See (5.49).]

According to (5.1)

$$T_{(if)t} = T_{i(if)t} = T_{(if)ft} = T_{i(if)ft} = D_{if}. \tag{5.45}$$

A trivial consequence of (5.45) and (5.43) (for the present $t \in G$) is that

$$\begin{aligned} D^G = &\sum_{(if)\in G} T_{(if)t} - \sum_{i\in G, (if)\in G} T_{i(if)t} \\ &- \sum_{f\in G, (if)\in G} T_{(if)ft} + \sum_{i\in G, f\in G, (if)\in G} T_{i(if)ft}. \end{aligned} \tag{5.46}$$

If one uses the representation (5.46) of F_5, then, for G containing t,

$$T + F_2 + F_4 + F_5 = \sum_{H\subset G} (-1)^{n(H)} T_H. \tag{5.47}$$

Thus, the proof of Proposition 5.2 for Case 2 can be completed by showing that F_6 belongs to $\mathcal{R}^G$.

Inspection† of the bubble diagrams B that correspond to the bubble-diagram functions F^B of F_6 [see (5.14), (2.11), and (5.41)] shows that all terms belong to $\mathcal{R}^G$ except possibly

$$\sum_{i\in\bar{G}, f\in\bar{G}, (if)\in\bar{G}} \bar{A}_{if}. \tag{5.48}$$

The function $\bar{A}_{if}$ is defined in (5.14d, g). Introducing the expression (5.11), and making use of the argument of Case 4 of Proposition 5.1, one sees that this term is also in $\mathcal{R}^G$. This completes the proof of Proposition 5.2.

Proposition 5.3:

$$T^G - D^G = \bar{T}^{\bar{G}} - \bar{D}^{\bar{G}} \tag{5.49}$$

where D^G is defined by (5.46) if G contains t, and by $D^G = 0$ if G does not contain t. The quantities $\bar{T}^{\bar{G}}$ and $\bar{D}^{\bar{G}}$ are defined for all sets $\bar{G}$ by

$$\bar{T}^{\bar{G}} \equiv -(T^{\bar{G}})^{\dagger} \tag{5.50a}$$

and

$$\bar{D}^{\bar{G}} \equiv -(D^{\bar{G}})^{\dagger}. \tag{5.50b}$$

Proof: It is sufficient to prove (5.49) for the case in which G contains t. For the application of Hermitian conjugation and a sign change to (5.49) gives

$$\bar{T}^G - \bar{D}^G = T^{\bar{G}} - D^{\bar{G}},$$

which is (5.49) with G replaced by $\bar{G}$.

Applying this same transformation to (5.20), and performing a summation, one obtains

$$\begin{aligned} &- \sum_{i\in\bar{G}, f\in\bar{G}, (if)\in\bar{G}} T_{(if)} \\ &= \sum_{i\in\bar{G}, f\in\bar{G}, (if)\in\bar{G}} (\bar{T}_{(if)} - \bar{T}_{i(if)} - \bar{T}_{(if)f} + \bar{T}_{i(if)f})_{\circ} \end{aligned} \tag{5.51}$$

Applying this same transformation to (5.14e) and (5.14f) one obtains

$$-T^f_{(if)} = \bar{T}_{(if)} - \bar{T}_{i(if)} \tag{5.52}$$

and

$$-T^i_{(if)} = \bar{T}_{(if)} - \bar{T}_{(if)f}. \tag{5.53}$$

Consider the case in which G contains t. If on the right-hand side of (5.44) one substitutes for $\bar{A}_{if}$ the right-hand side of (5.14h) and then uses (5.51), (5.52), (5.53), and (5.41), one finds that

$$\begin{aligned} F_6 = \bar{T} &- \sum_{i\in\bar{G}} \bar{T}_i - \sum_{f\in\bar{G}} \bar{T}_f - \sum_{(if)\in\bar{G}} \bar{T}_{(if)} + \sum_{i\in\bar{G}, f\in\bar{G}} \bar{T}_{if} \\ &+ \sum_{i\in\bar{G}, (if)\in\bar{G}} \bar{T}_{i(if)} + \sum_{f\in\bar{G}, (if)\in\bar{G}} \bar{T}_{(if)f} \\ &- \sum_{i\in G, f\in\bar{G}, (if)\in\bar{G}} \bar{T}_{i(if)f} + \sum_{i\in\bar{G}, f\in\bar{G}, (if)\in G} D_{if} \\ &= \bar{T}^{\bar{G}} + D^G. \end{aligned} \tag{5.54}$$

Inserting (5.54) into (5.42) and using (5.47) one obtains the required (5.49).

The Functions M^G: The function T^G was defined by (5.12) as

$$T^G \equiv \sum_{H\subset G} (-1)^{n(H)} T_H. \tag{5.55}$$

The function M^G was defined by (2.5a) as

$$M^G \equiv \sum_{K\subset G} (-1)^{n(K)} M_K. \tag{5.56}$$

According to (5.1) $T_H = M_H$ except when H contains, for some (i, f), one of the four sets $\{(if), t\}$, $\{i, (if), t\}$, $\{(if), t, f\}$, $\{i, (if), f, t\}$. Hence, by (5.55) and (5.56) $T^G = M^G$ if G does not include t. If G does include t, then (5.47) together with the definitions of F_2 [see (5.25a)], F_4 [see (5.39)], and F_5 [see (5.43) and (5.46)] show that

$$T^G = M^G + F_5.$$

Since $D^G = 0$ if G does not include t and is equal to F_5 if G does include t, the above result can be summarized in the equation

$$T^G = M^G + D^G. \tag{5.57}$$

Then (5.49) gives, for all G,

$$M^G = \bar{M}^{\bar{G}} = -(M^{\bar{G}})^{\dagger}, \tag{5.58}$$

which is (2.17).

Remark: Let the G in (5.58) be the complete set E. Then (5.58) says that M^E, the function evaluated below all the cuts, is

$$M^E = \bar{M}^\Phi \equiv \bar{M} \equiv -M^\dagger. \tag{5.59}$$

This result is called the Hermitian analyticity property of the scattering function. The more general property (5.58) says that the function M^G that corresponds to M plus the discontinuities that take the function to its value below the set of cuts g in G, is equal to the function $\bar{M}$ plus the discontinuities that take the function to above the set of cuts g in $\bar{G}$. The analogous result with T in place of M does not hold in general, as is shown by (5.49).

The function D^G is defined to be zero if t lies in $\bar{G}$. For t in G it can be expressed in the form (5.43). Thus for the set of 26 018 values of G defined by (2.25) one finds

$$D^G = \bar{D}^{\bar{G}} = 0. \tag{5.60a}$$

In these cases one has, by virtue of (5.57) and (5.58),

$$M^G = T^G = \bar{T}^{\bar{G}}. \tag{5.60b}$$

The analytic properties of these functions M^G are discussed in Sec. VI. It is already evident that these functions M^G satisfy the properties (4), (5), and (6) described in the introduction.

VI. ANALYTIC PROPERTIES OF THE M^G

A. Geometric representations of Landau diagrams

The proofs of analyticity properties given in this section are based on the existence of two different geometric representations of Landau diagrams. These two representations are discussed in this subsection.

Each internal line L_j of a Landau diagram has a well-defined direction: L_j is directed from the vertex V_r with $\epsilon_{jr} = -1$ to the vertex V_r with $\epsilon_{jr} = +1$. This direction is the direction of flow of positive energy. An arrow is often placed on L_j to indicate this direction, and L_j is said to *point* in the direction of this arrow, i.e., from the vertex V_r with $\epsilon_{jr} = -1$ to the vertex V_r with $\epsilon_{jr} = +1$. The external lines L_j are also directed: Each incoming line is directed toward a vertex V_r with $\epsilon_{jr} = +1$; each outgoing line L_j is directed away from a vertex V_r with $\epsilon_{jr} = -1$. (One can introduce *trivial* two-line vertices to take care of the trivial cases in which a line goes straight through the diagram without touching any nontrivial vertex.)

The first geometric representation of a Landau diagram D^σ is the space–time representation discussed in Sec. IV. Each space–time representation of D^σ represents a particular solution of the Landau equations corresponding to D^σ, and corresponds to some particular point p on $L(D^\sigma)$. In this representation each internal line L_j of D^σ is represented by a space–time 4-vector $\Delta_j = \alpha_j p_j$. The Landau equation $p_j{}^0 > 0$ entails that Δ_j point in the direction of increasing time if σ_j is plus, and in the direction of decreasing time if σ_j is minus. That is, the vertex V_r with $\epsilon_{jr} = +1$ lies later than the vertex V_r with $\epsilon_{jr} = -1$ if $\sigma_j = +$, but lies earlier if $\sigma_j = -$.

These conditions on the directions of the 4-vectors Δ_j impose a partial ordering requirement on the vertices of the Landau diagram D^σ. In particular, for any positive-α diagram D^+ with nonempty $L(D^+)$ the vertices must satisfy the partial ordering condition that the diagram can be drawn so that each internal line segment L_j points from left to right. Likewise, for any negative-α diagram D^- with nonempty $L(D^-)$ the vertices must also satisfy the partial ordering condition that the diagram can be drawn so that each internal line segment L_j points from left to right. To see this, one simply orders the vertices of the Landau diagram D^+ from left to right in accordance with the increasing time of the vertices of the space–time representations of D^+, and orders the vertices of D^- from left to right in accordance with decreasing time of the vertices of the space–time representations of D^-.

Consider now any bubble diagram B. By definition this diagram can be drawn so that every (explicit) internal line runs from the right-hand side of one bubble to the left-hand side of a bubble that stands completely to the right of the first bubble. Thus each line of B can be drawn as a line that points from left to right.

Consider next any $D^\sigma \subset B$. If one orders the bubbles of B in the way just described, so that all explicit lines of B point from left to right, and then draw each D_b^+ and D_b^- of $D^\sigma \subset B$ as a small diagram lying completely inside the corresponding bubble b, with all of its internal lines pointing from left to right, then all the lines of $D^\sigma \subset B$ will point from left to right [see (4.7b)].

The representation of a $D^\sigma \subset B$ as a diagram in which every line L_j points from left to right is called a flow diagram: Positive energy flows always from left to right in a flow diagram. This uniformity of directions of the lines of a flow diagram is to be contrasted with the nonuniformity of directions in the space–time representations of D^σ. In the space–time representations the vectors Δ_j point in the direction of increasing time if σ_j is plus, and in the direction of decreasing time if σ_j is minus. For any $D^\sigma \subset B$ it must be possible to draw both a flow diagram representation of D^σ and also a space–time representation of D^σ, if $L(D^\sigma)$ is nonempty.

The constraints imposed on D^σ by the existence of the flow diagram representation can be expressed in terms of the concept of a flow line.

Definition 6.1: A *flow line* is an ordered sequence of internal line segments L_j of a Landau diagram such that the leading end point ($\epsilon_{jr} = +1$) of any L_j in the sequence except the last one is the trailing end point ($\epsilon_{jr} = -1$) of the next one in the sequence. Thus positive energy flows always in the same direction along a flow line: It flows from the trailing vertex of the first L_j of the sequence to the leading end point of the final L_j. These two vertices are called the initial and final vertices of the flow line. In the flow diagram representation of D^σ the final vertex of any flow line stands to the right of the initial vertex of that flow line.

Definition 6.2: A flow line $V_r \to V_s$ is a flow line with initial vertex V_r and final vertex V_s.

Definition 6.3: A *maximal flow line* is a flow line that is not a proper subsequence of any other flow line.

B. Space–time ordering theorem

The proof of the analytic properties of M^G is based on a theorem proved in this subsection. This theorem depends on two lemmas.

Lemma 6.1: Let V_r and V_s be two distinct vertices of a connected Landau diagram D. Suppose X is a set of lines of D such that the cutting precisely once of each line of X separates D into two disjoint diagrams D_r and D_s, where V_r lies in D_r and V_s lies in D_s. Then there is a subset X^c of X such that the cutting precisely once of each line of X^c separates D into two diagrams D_r^c and D_s^c such that D_r^c is a connected diagram that contains V_r, and D_s^c is a connected diagram that contains V_s, and the two diagrams D_r^c and D_s^c exhaust D.

Proof: The cutting of the lines of the set X separates D into a number of connected parts. Let $D_s^{c\prime}$ be the connected part containing V_s. Let X_s be the set consisting of the lines of X that lie partly in $D_s^{c\prime}$. Let X^c be the subset of X_s that consists of each line L_j such that the cut in L_j can be reached from V_r by a path in D that is not cut by the set of cuts in the lines of X_s. This set X^c is the desired set: Cutting precisely once each line of X^c separates D into the two connected parts D_r^c and D_s^c.

It is clear that cutting the lines of X^c disconnects the part of D that is connected to V_r from the part that is connected to V_s, for any path in D from V_r to V_s would have to enter $D_s^{c\prime}$ at the cut on some line of X_s. But the first such cut reached by this path must be a cut on a line of X^c. Thus the cutting of the lines of X^c definitely separates D into at least two connected parts D_r^c and D_s^c. What must now be shown is that these two parts exhaust D.

Consider the diagram D cut on the lines of X^c. Suppose there is a point x that is not connected in this cut diagram to either V_r or V_s. This point x is connected in the original connected diagram D to the point V_r by some path P. Since x is, by assumption, not connected to V_r in the cut diagram, the path P must pass through the cut in at least one line of X^c. Let C be the first cut in X^c reached on the path P from x to V_r. Then x must be connected in the cut diagram to one side or the other of the cut C. But each cut in the lines of X^c is connected in the cut diagram on one side to V_s, since X^c is a subset of X_s and on the other side to V_r, since each cut of X^c can be connected to V_r by a path not cut by any cut in X_s. Thus x must be connected in the cut diagram to either V_r or V_s, contrary to the original assumption about x. Thus no such x can exist. This means that the two connected parts D_r^c and D_s^c of the cut diagram exhaust it.

Lemma 6.2: Let $V_r - V_s$ be a flow line of a flow diagram D^σ. Let X be the set of lines of D^σ that are cut by a plane T that lies perpendicular to the flow axis, that lies between V_r and V_s, and that touches no vertices of D^σ. Then the subset X^c of X of Lemma 6.1 is such that each L_j in X^c has its leading end point in D_s^c and its trailing end point in D_r^c.

Proof: The plane T cuts D^σ into the parts D_r and D_s, where D_r lies to the left of T and D_s lies to the right of T. The construction in Lemma 6.1 ensures that $D_s^{c\prime}$ is a subset of D_s. Hence $D_s^{c\prime}$ also lies to the right of T. Moreover, X^c is a subset of the set X_s of lines of X that touch $D_s^{c\prime}$. Thus every line of X^c touches $D_s^{c\prime}$, and hence has its leading end point in $D_s^{c\prime}$, which is contained in D_s^c. Thus the trailing end points of the $L_j \in X^c$ must lie in D_r^c.

Definition 6.4: Let V_r and V_s be two vertices of a Landau diagram D^σ. Let η be a plus or minus sign. A $V_r \overset{\eta}{-} V_s$ cut set of D^σ is a set of internal lines of D^σ such that D^σ is separated by the cutting precisely once of every line of this set into two connected diagrams $D^\sigma(V_r)$ and $D^\sigma(V_s)$, where V_r lies in $D^\sigma(V_r)$ and V_s lies in $D^\sigma(V_s)$, and such that each line L_j of the cut set points from $D^\sigma(V_r)$ to $D^\sigma(V_s)$, and has a sign $\sigma_j = \eta$ or $\pm$.

Theorem 6.1: Let B be any connected bubble diagram. Let D^σ be any $D^\sigma \subset B$ that has a flow line $V_r - V_s$. Let η be either plus or minus. Suppose D^σ has no $V_r \overset{\eta}{-} V_s$ cut set. Then for every space–time representation of D^σ the vectors ω_r and ω_s to the vertices V_r and V_s satisfy

$$\omega_s - \omega_r \in V^{-\eta}, \tag{6.1}$$

where V^+ is the open forward light cone and V^- is the open backward light cone.

Proof: Suppose η is plus. And suppose that the points V_r and V_s both lie inside some single minus bubble b (i.e., they are vertices of the D_b^- that replaces b in the construction of $D^\sigma \subset B$). In this case the flow line $V_r - V_s$ must consist wholly of line segments L_j that also lie inside this minus bubble, and hence have signs $\sigma_j = -$. This is because the condition that the bubbles of B be partially ordered precludes the possibility that a flow line $V_r - V_s$ begin and end in the same b, but pass outside b. But if the lines L_j of $V_r - V_s$ all carry minus signs, then the Landau equations that define the space–time representations of D^σ entail that each of the corresponding Δ_j point into the backward light cone. Thus, by virtue of the ordering conditions on the L_j that make up a flow line, condition (6.1) will be satisfied.

Suppose, on the other hand, that V_r and V_s do not lie in the same minus bubble. Then one can construct a $V_r \overset{+}{-} V_s$ cut set of D^σ. To do this, simply draw the flow diagram D^σ by first making all the minus bubbles b of B extremely tiny, and then replacing each tiny minus bubble b by a tiny D_b^-. The D_b^+ are not made tiny. Since the minus bubbles are tiny, and V_r and V_s do not lie inside the same minus bubble, one can draw a plane T that lies perpendicular to the flow axis, that lies between V_r and V_s, and that does not touch any vertex of D^σ or any line that lies inside any of the tiny minus bubbles. The X^c of Lemma 6.2 is then a $V_r \overset{+}{-} V_s$ cut set of D^σ. Thus the assumption of the theorem is not satisfied in this case, and (6.1) need not be proved. This completes the proof for the case $\eta = +$. For the case $\eta = -1$ the proof is completely analogous.

C. Skeleton diagrams

Each flow diagram D has a unique skeleton diagram D_s, which is constructed as follows. Consider the set of maximal flow lines of D. Regard as equivalent any two of them that touch exactly the same set of external vertices. (External vertices are vertices that touch ex-

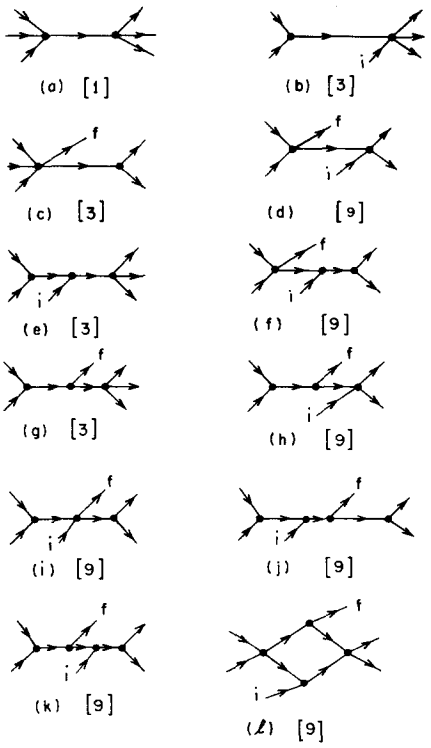


FIG. 6.1. The 76 skeleton diagrams for $3 \to 3$ processes. The indices i and f run over (1,2,3), and (4,5,6), respectively. The number in square brackets below each figure is the number of skeleton diagrams represented by that figure.

ternal lines.) Draw a diagram consisting of the external vertices of D, the external lines of D, and one internal (flow) line l for each equivalence class of maximal flow lines of D. This line l is drawn so that it touches precisely those external vertices that are touched by each member of the corresponding equivalence class. Now delete any line l that touches a set of external vertices that is a proper subset of the set of external vertices touched by any other line l'. The resulting diagram D_s called the skeleton of D.

Each flow diagram D having three incoming lines and three outgoing lines has a unique skeleton diagram D_s that is one of the 76 skeleton diagrams shown in Fig. 6.1.

In constructing these diagrams use is made of the stability requirements, which demand that each nontrivial vertex have at least two incoming lines, and at least two outgoing lines. Each maximal flow line must therefore begin at a vertex that has at least two incoming external lines, and must end at a vertex that has at least two outgoing external lines.

D. Path of continuation (off-mass-shell)

The aim of the present subsection is to construct for each of the 26 018 boundary values M^G and for each possible singularity surface $L(D^\sigma)$ of M^G a path of continuation that continues the function M^G into itself around $L(D_\sigma)$. However, the complications arising from the mass-shell constraints are ignored. When these constraints are ignored the rule of continuation can be formulated so that it depends only on G and on the skeleton of D^σ. And for a given skeleton D_s the rule depends on G only through the question of whether certain critical g's associated with D_s lie in G or $\bar{G}$. These critical g's are those that label the critical channels of D_s, which are now described.

For each skeleton D_s there is a unique set of critical channels g. The critical channels g corresponding to a given D_s are the channels g such that D_s can be separated, by cutting some of its internal lines l, into two connected parts $D_s(J_g)$ and $D_s(\bar{J}_g)$, where $D_s(J_g)$ contains all the external lines $j \in J_g$, and $D_s(\bar{J}_g)$ contains all the external lines $j \in \bar{J}_g$. Moreover, all the cut lines l run from $D_s(\bar{J}_g)$ to $D_s(J_g)$. The critical channels corresponding to several skeletons D_s are indicated in Fig. 6.2. (For each g the sets J_g and $\bar{J}_g$ must be selected so that J_g contains at least two indices f, and $\bar{J}_g$ contains at least two indices i. Otherwise the conditions given above cannot be satisfied.)

For any given G a set of signs η_g is defined as follows [see (2.25)]: The sign η_g is plus if g lies in G and is minus if g lies in $\bar{G}$. Symbolically, η_g can be defined by the set of conditions

$$g \in G^{\eta_g} \quad \text{(for every } g \in E\text{)} \tag{6.2a}$$

where

$$G^+ = G \tag{6.2b}$$

and

$$G^- = \bar{G}. \tag{6.2c}$$

The rules of continuation to be constructed here apply only to the 26 018 boundary values M^G defined by (2.25). For these functions one has, according to (5.60),

$$M^G = T^G = \bar{T}^{\bar{G}}. \tag{6.3}$$

Thus M^G can, by virtue of proposition 5.2 and Eq. (4.14), be written in two alternative forms:

$$M^G = \sum_{B \in \mathcal{B}^G_+} F^B \tag{6.4+}$$

and

$$M^G = \sum_{B \in \mathcal{B}^G_-} F^B, \tag{6.4–}$$

where the notation $\mathcal{B}^G_+ \equiv \mathcal{B}^G$ and $\mathcal{B}^G_- \equiv \bar{\mathcal{B}}^{\bar{G}}$ is used. The sets $\mathcal{B}^G_+$ and $\mathcal{B}^G_-$ have the following properties: No $D_\sigma \subset \mathcal{B}^G_+$ has a positive-α g-channel cut set for any g in G; no $D_\sigma \subset \mathcal{B}^G_-$ has a negative-α g-channel cut set for any g in $\bar{G}$.

If one introduces the notation of (6.2) and writes $+\alpha$ for positive-α and $-\alpha$ for negative-α then the properties of $\mathcal{B}^G_+$ and $\mathcal{B}^G_-$ can be combined into the following statement:

> For any G and g let $\eta = \eta_g$. Then no $D^\sigma \subset \mathcal{B}^G_\eta$ has an $\eta\alpha$ g-channel cut set.

The rule for continuing M^G past any $L(D^\sigma)$ will be derived by combining this property of $\mathcal{B}^G_\eta$ with Theorem 6.1, and then using the structure theorem described in Sec. IV.

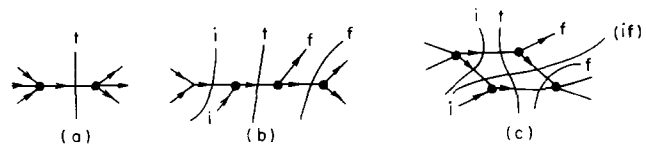


FIG. 6.2. The critical channels g of several skeleton diagrams D_s are indicated by lines g that separate D_s into the two parts $D_s(\bar{J}_g)$ and $D_s(J_g)$.

Consider any fixed G. Let D_s be the skeleton shown in Fig. 6.2(a). The rule for continuing $M^G(p)$ past all $L(D^\sigma)$ that correspond to D^σ having this skeleton D_s will now be derived.

The diagram D_s [i.e., Fig. 6.2(a)] has only one critical channel g, namely the channel $g=t$. Let η be the η_t defined in (6.2). Let M^G be represented by the formula (6.4η). Then according to the statement just given of property (4.12), no $D^\sigma \subset \mathcal{B}_\eta^G$ has an $\eta\alpha$ t-channel cut set.

This result implies that the suppositions of Theorem 6.1 are valid for every $B \in \mathcal{B}_\eta^G$ and every $D^\sigma \subset B$ such that D^σ has the skeleton D_s, provided V_r and V_s are identified as the initial and final vertices of D_s, respectively. This is because any $V_r \overset{\eta}{\not\to} V_s$ cut set of D^σ is also an $\eta\alpha$ t-channel cut set of D^σ [see (4.12).] Indeed, since D^σ has only the two external vertices V_r and V_s shown in Fig. 6.2(a), any cutting of the internal lines of D^σ that separates D^σ into two parts $D^\sigma(V_r)$ and $D^\sigma(V_s)$ with V_r in $D^\sigma(V_r)$ and V_s in $D^\sigma(V_s)$ must also separate D^σ into two parts $D^\sigma(\bar{J}_g)$ and $D^\sigma(J_g)$ such that $D^\sigma(\bar{J}_g)$ contains all the external lines $j \in \bar{J}_g$ and $D^\sigma(J_g)$ contains all the external lines $j \in J_g$. In fact, $D^\sigma(\bar{J}_g)$ would be identical to $D^\sigma(V_r)$, and $D^\sigma(J_g)$ would be identical to $D^\sigma(V_s)$.

Since the suppositions of Theorem 6.1 hold, the conclusion holds: Equation (6.1) is true for all space–time representations of all $D^\sigma \subset \mathcal{B}_\eta^G$ that have as skeleton the D_s of Fig. 6.2(a), with V_r and V_s identified in the manner described.

According to (4.6) and (6.4η), the singularities of M^G are confined to the union of $L(D^\sigma)$ over $D^\sigma \subset \mathcal{B}_\eta^G$. Suppose p lies on $L(D^\sigma)$ for $D^\sigma \subset \mathcal{B}_\eta^G$ only if D^σ has skeleton D_s. (This supposition will be removed later.) Then, by virtue of the conclusion stated in the preceeding paragraph, Eq. (6.1) holds for all ω in the set

$$\Omega(p) = \bigcup_{B \in \mathcal{B}_\eta^G} \Omega^B(p), \tag{6.5a}$$

where $\Omega^B(p)$ is defined in (4.8). Thus if $C(p)$ is defined by

$$C(p) = \bigcap_{B \in \mathcal{B}_\eta^G} C^B(p), \tag{6.5b}$$

where $C^B(p)$ is defined in (4.9), then (6.1) and (4.9) show that $C(p)$ contains all points q that satisfy, with $\epsilon_f = +1$ and $\epsilon_i = -1$,

$$\sum_{j \in J_t} \epsilon_j q_j = -\sum_{j \in \bar{J}_t} \epsilon_j q_j \in V^{-\eta}. \tag{6.6}$$

But then, according to consequence (2) of the structure theorem, which is described below (4.9), all of the (finite number of) bubble diagram functions F^B that occur in the expression (6.4η) are boundary values of functions that are analytic at all mass-shell points sufficiently near p for which q satisfies (6.6). (Strictly speaking, the cone of analyticity is not $V^{-\eta}$ itself but rather any cone that is contained with its closure, apart from the apex at $q=0$, in $V^{-\eta}$. This slight diminuation of all cones of analyticity will always be taken as understood in the discussion that follows.)

This means that M^G can be continued about all $L(D^\sigma)$ that correspond to D^σ having the skeleton D_s of Fig. 6.2(a) by means of a small detour through the region (6.6).

Next let D_s be Fig. 6.2(b). The rule for continuing any fixed M^G past all $L(D^\sigma)$ associated with this D_s is now derived.

Consider first the critical channel i of D_s. Let η be η_i, and use the representation of M^G given by (6.4η). Let the B of Theorem 6.1 be any $B \in \mathcal{B}_\eta^G$, and let D^σ be any $D^\sigma \subset B \in \mathcal{B}_\eta^G$ having skeleton D_s. Finally, let the vertices V_r and V_s of Theorem 6.1 be the vertices of Fig. 6.2(b) that stand just to the left and just to the right of the i-channel line of Fig. 6.2(b), respectively. [That is, V_r is the left-most vertex of Fig. 6.2(b), and V_s is the vertex at which the external line i terminates.]

We know that no $D^\sigma \subset \mathcal{B}_\eta^G$ has an $\eta\alpha$ i-channel cut set. This immediately implies that no $D^\sigma \subset \mathcal{B}_\eta^G$ having skeleton D_s has a $V_r \overset{\eta}{\not\to} V_s$ cut set in which the last two vertices of Fig. 6.2(b) lie in $D^\sigma(V_s)$. And the possibility of any other $V_r \overset{\eta}{\not\to} V_s$ cut set is ruled out by the existence of the flow lines represented by the two right-most lines l of Fig. 6.2(b): These lines would run the wrong way across any cut separating a $D^\sigma(V_r)$ containing V_r from $D^\sigma(V_s)$ containing V_s if either of the remaining two vertices were to lie in $D^\sigma(V_r)$. This is shown in Fig. 6.3, which is explained in the text that follows.

The possible $V_r \overset{\eta}{\not\to} V_s$ cut sets can be examined by placing V_r and V_s on the left- and right-hand sides of a vertical line, and placing the other vertices on the two sides of this line in all possible combinations. There can be a $V_r \overset{\eta}{\not\to} V_s$ with the parts $D^\sigma(V_r)$ and $D^\sigma(V_s)$ of D^σ containing the external vertices lying, respectively, on the left- and right-hand side of the vertical line only if all the lines of the skeleton diagram run from left to right across the vertical line. Only figure (a) satisfies this condition. However, this way [Fig. 6.3(a)] of achieving a $V_r \overset{\eta}{\not\to} V_s$ cut set of a $D^\sigma \subset \mathcal{B}_\eta^G$ having D_s as skeleton is ruled out by the property of $\mathcal{B}_\eta^G$.

Since the suppositions of Theorem 6.1 hold, the conclusion holds: Every space–time representation of every $D^\sigma \subset \mathcal{B}_\eta^G$ that has D_s [Fig. 6.2(b)] as its skeleton is such that the vector $\Delta = \omega_s - \omega_r$ lies in $V^{-\eta}$. Here ω_s and ω_r are the space–time vectors to the leading and trailing vertices V_s and V_r of l.

Essentially the same argument can be made for the pair of vertices lying on either side of the t line in Fig.

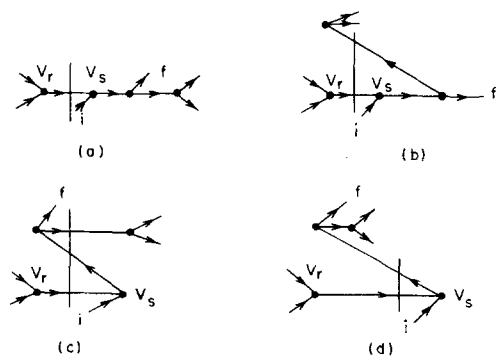


FIG. 6.3. Diagrams for examining possible $v_s \overset{\eta}{\not\to} v_s$ cut sets.

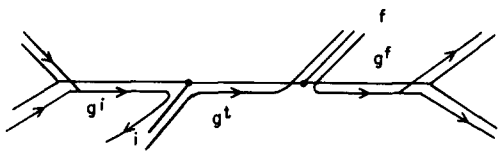


FIG. 6.4. Each q^g can be considered to be the imaginary energy carried along an open path that contains only one internal line.

6.2(b), and also for the pair of vertices lying on either side of the f line. The results of the three similar arguments can be summarized in a systematic way.

For each of the three critical channels g associated with Fig. 6.2(b) define

$$\Delta^g \equiv \omega_s^g - \omega_r^g, \tag{6.7}$$

where ω_s^g and ω_r^g are the 4-vectors to the vertices V_s^g and V_r^g that lie just to the right and just to the left of the line g in Fig. 6.2(b) corresponding to channel g. (For example, ω_s^t and ω_r^t are the vectors to the vertices lying at end points of the lines f and i, respectively.) Then the property of β_η^G for $\eta = \eta_g$, together with Theorem 6.1, shows that, for each critical channel g of D_s,

$$\Delta^g \in V^{-\eta_g} \tag{6.8}$$

for every space–time representation of every $D^\sigma \subset \beta_{\eta_g}^G$ having as skeleton the D_s of Fig. 6.2(b).

For any D^σ having as skeleton the D_s of Fig. 6.2(b) the function occurring on the right-hand side of (4.9a) can be written in the form

$$-\sum_{\substack{j\in \mathrm{Ext}\\ r\in \mathrm{Ver}}} q_j \epsilon_{jr} \omega_r = \sum_g q^g \Delta^g, \tag{6.9}$$

where g runs over the critical channels associated with Fig. 6.2(b), and q^g is the imaginary part of the momentum–energy that runs along the line l of Fig. 6.2(b) that is cut by the line g corresponding to channel g. In particular,

$$q^i = q_4 + q_5 + q_6 - q_i, \tag{6.10a}$$

$$q^t = q_4 + q_5 + q_6 = q_1 + q_2 + q_3, \tag{6.10b}$$

and

$$q^f = q_4 + q_5 + q_6 - q_f. \tag{6.10c}$$

The various q^g are indicated in Fig. 6.4.

We shall now temporarily ignore the mass-shell constraints, and suppose that our functions M^G can be extended some small finite distance off-mass-shell, and that the only singularity surfaces encountered in some sufficiently small neighborhood of any real mass-shell point p are the singularity surfaces obtained in the mass-shell theory. If continuations through these off-mass-shell regions are thus allowed, then the rules of continuation can be stated in a simple way, which will be described next. The complications associated with the restriction to the mass shell, and with the possible conspiracies among singularities corresponding to different skeleton graphs will be considered in the following subsection.

The results (6.8) and (6.9) together with the $i\epsilon$ rule stated below (4.9) show that if g labels any one of the three critical channels of D_s and if p lies on $L(D^\sigma)$ for $D^\sigma \subset \beta_{\eta_g}^G$ only if D^σ has skeleton D_s, then M^G is the boundary value of a function that is analytic near p in the q-space region satisfying

$$q^g \in V^{-\eta_g} \tag{6.11a}$$

and

$$q^{g'} = 0 \tag{6.11b}$$

where g' runs over the critical channels $g' \neq g$. This equation gives three different regions of analyticity, one for each critical channel of D_s. The boundary values are all the same (distribution) M^G. Thus the generalized edge of the wedge theorem[19] implies that the functions in these three domains are parts of one single analytic function that is analytic near p also in the q-space region restricted only by

$$q^g \in V^{-\eta_g} \quad \text{all critical } g. \tag{6.12}$$

Equation (6.12) gives a domain through which M^G can be continued (off-mass-shell) past all singularities surfaces $L(D^\sigma)$ of M^G that correspond to D^σ having as skeleton the diagram D_s shown in Fig. 6.2(b). Essentially the same argument applies for any skeleton diagram of Fig. 6.1 that does not have closed loops [i.e., the D_s of Figs. 6.2(a)–(k)]. For any given one of these *tree* diagrams D_s the q^g associated with each critical channel g of D_s is a well-defined combination of external vectors, and the domain defined as the intersection of some sufficiently small neighborhood of p with the domain of the form (6.12) is a domain of analyticity of M^G, provided p lies on only those singularity surfaces $L(D^\sigma)$ of M^G such that the skeletons of these D^σ are all the specified tree diagram D_s.

The rule obtained above is simple and explicit. M^G can be continued into itself around all the singularity suarfaces $L(D^\sigma)$ that correspond to any specified tree skeleton diagram D_s by following any infinitesimal detour that satisfies the following conditions: For each critical channel g of D_s the variable q^g is shifted into the forward light cone if M^G is evaluated above the g cut (i.e., if $g \in \bar{G}$), and into the backward light cone if M^G is evaluated below the cut g (i.e., if $g \in G$).

Now let D_s be the box diagram of Fig. 6.2(c). Let G be any fixed G that satisfies the conditions (2.25). These conditions require [with i and f fixed as in Fig. 6.2(c)] that the conditions

$$\eta_i = \eta_f = -\eta_t = -\eta_{(if)} \tag{6.13}$$

do not all hold. This requirement ensures that for some *preferred* internal line segment l of Fig. 6.2(c) the following condition holds: The two signs η_g corresponding

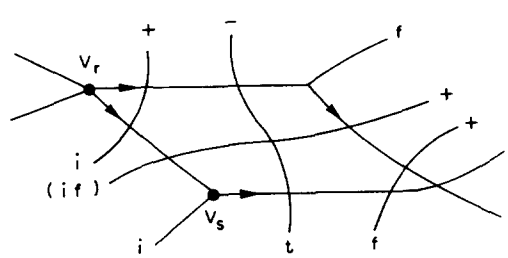


FIG. 6.5. Diagram showing the signs η_g associated with the lines g.

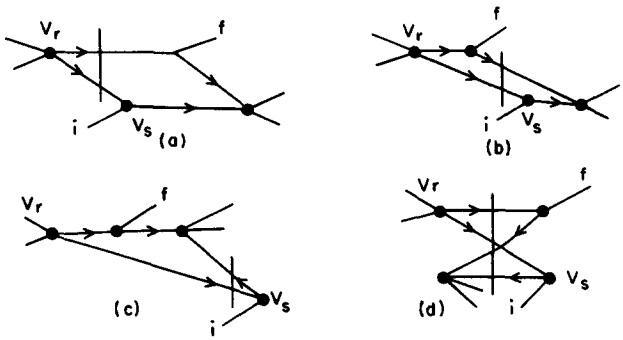


FIG. 6.6. Diagrams for examining possible $v_r \xrightarrow{\eta} v_s$ cut sets.

to the two lines g of Fig. 6.2(c) that cross the preferred line l are equal. In fact, the requirement that (6.13) be false ensures that at least two internal line segments l of Fig. 6.2(c) enjoy this property. In the special case illustrated in Fig. 6.5, in which $\eta_i = \eta_f = -\eta_t = \eta_{(if)} = +$, one of these two preferred lines l is the line on the right-hand side of the box, and the other is the line on the left-hand side of the box.

The flow line $V_r \to V_s$ of Theorem 6.1 is now taken to be one of the preferred lines l, and the vertices V_r and V_s are taken to be the initial and final vertices on this flow line, respectively. Thus if the flow line $V_r \to V_s$ of Theorem 6.1 is the preferred line l on the left-hand side of the box in Fig. 6.5 then the vertices V_r and V_s are those indicated in Fig. 6.5.

The two lines g that cut across the preferred line l correspond to two of the critical channels of D_s, which are called g_1 and g_2. The η in Theorem 6.1 is taken to be $\eta = \eta_{g_1} = \eta_{g_2}$.

With this choice of η the following property holds: No $D^\sigma \subset \beta_\eta^G$ has an $\eta\alpha$ g_1-channel cut set or an $\eta\alpha$ g_2-channel cut set. These two conditions, together with the conditions on the directions of the lines of D_s imply that no $D^\sigma \subset \beta_\eta^G$ has a $V_r \xrightarrow{\eta} V_s$ cut set. (See Fig. 6.6)

In Fig. 6.6 the flow line $V_r \to V_s$ is assumed to be the line l that lies on the left-hand side of the box, as in Fig. 6.5. Cases (c) and (d) cannot give a $V_r \xrightarrow{\eta} V_s$ cut set because a flow line points from $D(V_s)$ to $D(V_r)$. On the other hand, the cases (a) and (b) cannot give a $V_r \xrightarrow{\eta} V_s$ cut set for any $D^\sigma \subset \beta_\eta^G$, because no $D^\sigma \subset \beta_\eta^G$ can have a $\eta\alpha$ g_1-channel or g_2-channel cut set, for $\eta = \eta_{g_1} = \eta_{g_2}$. Hence the suppositions of Theorem 6.1 are satisfied.

From Theorem 6.1 one concludes that

$$\Delta = \omega_s - \omega_r \in V^{-\eta} \tag{6.14}$$

for all space–time representations of $D^\sigma \subset \beta_\eta^G$ having the box diagram skeleton diagram shown. Similar arguments show that (6.14) holds for each preferred line l of D_s, with $V_r \to V_s$ the flow line l and η now $\eta(l)$, the common signs η_g associated with l.

The function occurring on the right-hand side of (4.9a) can be written

$$-\sum_{\substack{j\in \mathrm{Ext} \\ r\in \mathrm{Ver}}} q_j \epsilon_{jr} \omega_r = \sum_l \Delta^l q^l. \tag{6.15}$$

Here q^l is the imaginary part of the momentum–energy flowing along line l of the box diagram D_s. (Momentum–energy is required to be conserved at each of the four vertices. See Fig. 6.7.)

The imaginary part of the momentum–energy flowing in and out of the diagram can be considered to flow along the internal lines $l = 1, 2, 3, 4$ of the box skeleton diagram D_s. The amount flowing along line l is q^l. This decomposition is not unique. However, the sum occurring on the right-hand side of (6.15) is not affected by this lack of uniqueness since the contribution associated with conserved flow around the loop is zero.

The $i\epsilon$ rule (4.9) ensures that M^G is analytic at points sufficiently near p that lie in the region

$$q^l \in V^{-\eta(l)} \tag{6.16a}$$

$$q^{l'} = 0 \quad (l' \neq l) \tag{6.16b}$$

where l is any preferred side of the box. This rule holds for all points p that lie on no $L(D^\sigma)$ for $D^\sigma \subset \beta_\eta^G$ except those having as skeleton the particular box diagram D_s under consideration.

This completes the derivation of the off-mass-shell $i\epsilon$ rules associated with the various individual skeleton diagrams.

E. Path of continuation (on-mass-shell)

The rule obtained above for continuing M^G past $L(D^\sigma)$ is simple because it prescribes a set of allowed detours that is independent of the particular point p on $L(D^\sigma)$. However, the rule is deficient because in many cases the set of allowed detours contains none that remain on the mass shell. Moreover, possible conflicts between the rules associated with different skeleton diagrams have not been considered: It is conceivable that a set of points lying on surfaces $L(D^\sigma)$ associated with different skeleton diagrams might conspire to block the continuation. These two deficiencies can usually be remedied by considering paths of continuation that depend on the real point p around which the continuation takes place. However, there is a set of exceptional Landau surfaces past which no mass-shell continuations are expected. This set of exceptional Landau surfaces is now described.

Two space–time representations r and r' are said to be *externally similar* if and only if every external trajectory line l_j' of r' can be brought into coincidence with the corresponding line l_j of r by means of a single (positive or negative) scale change and a single overall space–time translation of r' relative to r. The external trajectory line l_j is the space–time line that contains the external trajectory associated with the external variable $p_j \in p$. Two representations can be externally

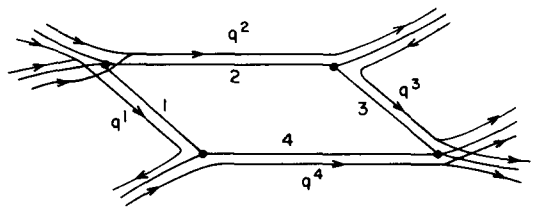


FIG. 6.7. Flow of imaginary momentum energy.

similar only if they generate the same point $p \equiv (p_1, \cdots, p_n)$, since only then are the corresponding lines l_j and l_j' parallel.

Two space–time representations r and r' are said to be *externally equivalent* if and only if (1) the two representations are externally similar, and (2) any scale change that brings every pair of lines l_j and l_j' into coincidence also brings the corresponding pairs of external vertices v_j and v_j' in coincidence.

Each point on any Landau surface is generated by many different space–time representations. In the most common case the various representations generating a given point are all externally equivalent. A simple example is provided by the normal threshold diagram having two vertices connected by two internal lines, and a second diagram in which these two internal lines are joined also at a third (internal) vertex, to give a sausage diagram. These two diagrams give coincident Landau surfaces.

The surfaces generated in this way only by externally equivalent space–time representations are unexceptional Landau surfaces. And surfaces generated only by representations no two of which are externally similar are also unexceptional. The *exceptional Landau surfaces* are the codimension-one analytic manifolds each point of which is generated by at least two externally similar representations that are not externally equivalent. An important example is the ice cream cone diagram surface in the equal mass case. In the general unequal mass case this surface has two nonpositive-α branches. These coincide when all internal and external masses become equal. The $i\epsilon$ rules associated with these two branches are opposite in cases of interest, e.g., for M^i or M^f. Thus the mass-shell continuation is blocked in the equal mass case by the coincidence of these two surfaces having opposite $i\epsilon$ rules.

In this equal mass case every point on the coincident pair of surfaces is generated by an infinite set of externally similar representations no two of which are externally equivalent. In these representations two of the three vertices coincide, and the external trajectory line attached to the third vertex passes through all three vertices. The infinite set of externally similar representations is generated by a scale change.

In the example just described all of the external trajectory lines pass through a common point. Such representations are called star-graphs. For star-graphs a scale transformation always generates an infinite set of externally similar space–time representations no two of which are externally equivalent. Thus all codimension-one (or zero) Landau surfaces generated by star-graphs are exceptional. Conversely, all exceptional Landau surfaces known to us are generated by star graphs.

The 26 018 boundary values M^G continue on-mass-shell into themselves past all but the exceptional Landau surfaces. That is, if $\mathcal{D}_G$ is the set of diagrams

$$\mathcal{D}_G \equiv \mathcal{B}_+^G \cup \mathcal{D}_-^G \tag{6.17}$$

and

$$L_G = \bigcup_{D^\sigma \in \mathcal{D}_G} L(D^\sigma). \tag{6.18}$$

Then M_G continues into itself, on-mass-shell, past the surface

$$\hat{L}_G \equiv L_G - L_G{}^E$$

where $L_G{}^E$ is the closure of the union of all exceptional Landau surfaces generated by representations of $D^\sigma \in \mathcal{D}_G$.

Remark: The existence of a space–time representation is a necessary condition for a singularity, but it is not sufficient. Representations that do not correspond to singularities of M^G can be systematically excluded from the set of all representations $D^\sigma \in \mathcal{D}_G$ without invalidating our arguments.

It is shown in appendix A that there is, for each G, a codimension-two algebraic variety W_G such that for any point p on $\hat{L}_G - W_G$ the following three properties hold:

(a) There is a neighborhood U of p such that $\hat{L}_G \cap U$ is a real codimension-one analytic submanifold of $\mathcal{M}$. This means, in particular, that the surface L_G near p is a smooth codimension-one surface with a well-defined smoothly varying normal.

(b) If $\mathcal{R}_G$ is the set of all representations of all $D^\sigma \subset \mathcal{D}_G$, then all of the $r(p) \in \mathcal{R}_G$ that generate p are externally equivalent.

(c) If $r(p) \in \mathcal{R}_G$ generates p then the external trajectory lines l_j of $r(p)$ do not all pass through a common point; i.e., there is no space–time point v that lies on every external trajectory line l_j of $r(p)$.

[Property (c) follows from (b), but is explicitly mentioned for later use.]

The set W_G is a codimension-two algebraic subvariety of $\mathcal{M}$, and hence it can not block the analytic continuation. Thus it is sufficient to prove that each M^G can be continued past $\hat{L}_G - W_G \equiv \hat{L} - W$.

Inspection of (4.9) shows that the $i\epsilon$ rule at any point p is determined by the locations of the external vertices of all of the representations $r(p)$ that generate p. It will be noted later that the condition on the domain of analyticity $C(p)$ that arises from any *individual* $r(p) \in \mathcal{R}$ that generates any $p \in \hat{L} - W$ defines a full upper-half plane of analyticity in appropriate mass-shell variables. Thus it is sufficient to show that the conditions on $C(p)$ associated with all of the different representations that generate any $p \in \hat{L} - W$ are identical.

Property (b) ensures that, apart from positive or negative scale changes and overall space–time translations, the location of each individual external vertex v_j is the same in every representation $r(p) \in \mathcal{R}$ that generates any fixed $p \in \hat{L} - W$. The positive scale changes and the overall space–time translations do not effect the $i\epsilon$ rule. Thus, in view of (6.4), it need only be shown that for each point $p \in \hat{L} - W$, and each G, there is a sign η such that if $\mathcal{R}_\eta^G$ is the restriction of $\mathcal{R}$ to representations of diagrams $D^\sigma \subset \mathcal{B}_\eta^G$, then no two $r(p) \in \mathcal{R}_\eta^G$ that generate p are related by a negative scale change, where a negative scale change is a scale change that changes the signs of all of the difference vectors $\omega_s - \omega_r$.

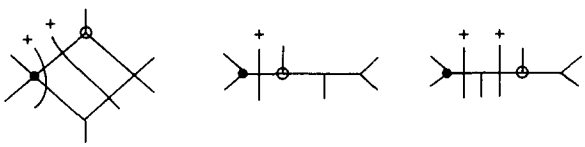


FIG. 6.8. The three skeleton diagrams for the case $n'=4$.

Let $R_\eta^G(D_s)$ be the set of all representations of all diagrams $D^\sigma \subset \mathcal{B}_\eta^G$ that have skeleton D_s. It was shown in the preceding subsection that for each G and D_s there is a pair of external vertices V_r and V_s and a sign η such that for all representations $r(p) \in R_\eta^G(D_s)$ the following condition holds:

$$\omega_s - \omega_r \in V^{-\eta}. \tag{6.19}$$

This condition is not maintained by a negative scale change. Thus for any G, D_s, and $p \in \hat{L} - W$, no two of the externally equivalent $r(p) \in R_\eta^G(D^s)$ that generate p can be related by a negative scale change. This result would complete the proof, were it not for the possibility of conspiracies among the singularities associated with different skeletons D_s.

To complete the proof it will now be shown that for each G and $p \in \hat{L} - W$ there is a pair of external vertices V_r and V_s and a sign η such that (6.19) holds for all representations $r(p) \in R_\eta^G$ that generate p. This condition precludes the possibility that any two of the externally equivalent $r(p) \in R_\eta^G$ that generate p are related by a negative scale change. Therefore, all of the representations $r(p) \in R_\eta^G$ that generate p give exactly the same constraint on the region $C(p)$ of analyticity.

To obtain this result let p be any fixed point in $\hat{L} - W$. Consider all of the representations $r(p) \in R$ that generate p. The six external vertices V_i of any such representation lie on a set of n' space–time points, where $n' = 2, 3,$ or 4. [The case $n'=1$ is ruled out by property (c).]

If $n'=4$ then each representation $r(p) \in R$ that generates the fixed p must be a representation of a $D^\sigma \in D$ that has one of three possible skeletons. Apart from trivial modifications these three skeletons are the three indicated in Fig. 6.8. The different lines in this figure are identified by their slopes. The vertical lines can be either initial or final.

Consider first the box diagram D_s of Fig. 6.8. It was shown in Sec. VI.D that for each G the skeleton diagram D_s has a preferred line $V_r - V_s$. Let this preferred line be, for example, the upper left-hand internal line of D_s. Then the vertices V_r and V_s are indicated by the heavy dot and little circle, respectively. In the figure the common signs η_g of the two lines g that cross the preferred line are shown as plus. Equation (6.14) gives (6.19), with η equal to these two common signs η_g. For the second skeleton in Fig. 6.8 this same condition (6.19) holds as a consequence of (6.8), and the condition that the sign η be equal to these two common signs η_g. For the third skeleton the same condition (6.19) holds for the same reasons, together with the fact that the sum of two vectors in $V^{-\eta}$ also lies in $V^{-\eta}$. Thus all of the $r(p) \in R_\eta^G$ that generate this arbitrary $n'=4$ point p satisfy the same condition (6.19). Hence no two can be related by a negative scale change.

Consider next the case $n'=2$. The possible skeletons are shown in Fig. 6.9.

Let η be fixed to be the sign η_g associated with the channel g defined by the first of these diagrams. In the figure η is shown as plus. The various signs η_g of the lines g in the second (i.e., box) diagram are then fixed by (2.25), together with the requirement that the line joining the two heavy dots cannot be a preferred line, nor can the line joining the two little circles be a preferred line. For with every preferred line there is a light cone condition (6.14), and such a condition would contradict the $n'=2$ condition that the two vertices represented by heavy dots lie at the same space–time point, or the analogous condition on the little circles. The remaining two lines of the box are then the two preferred lines, and the condition (6.14) associated with them gives the same condition (6.19) as was obtained from the first skeleton.

For the next three diagrams of Fig. 6.9 the condition (6.8) immediately gives the same condition as was obtained from the first two diagrams. For the final diagram the $n'=2$ condition that the two vertices represented by heavy dots lie at the same space–time point, and the similar condition for the little circles, together with (6.8), require that the signs associated with the lines g be alternating, as shown. The upper case gives the desired condition (6.19). The lower case contradicts (2.25). Thus for all $n'=2$ points there is a common condition (6.19), and a negative scale change is again precluded.

For the $n'=3$ points similar arguments work. There are essentially two cases. The first is shown in Fig. 6.10.

In this first case the two signs in the first diagram are the same, say plus. In this case the V_r and V_s are chosen to be the heavy dot and little circle, respectively. Then the remaining three diagrams shown, by arguments essentially the same as those given above, that the same condition (6.19) holds in all four cases.

The second $n'=3$ case is shown in Fig. 6.11.

In this case the two signs in the first diagram are opposite, say plus and minus as shown. The signs in the box diagram are then forced to be those shown by Eq. (2.25) and the $n'=3$ condition that the lower right-hand line

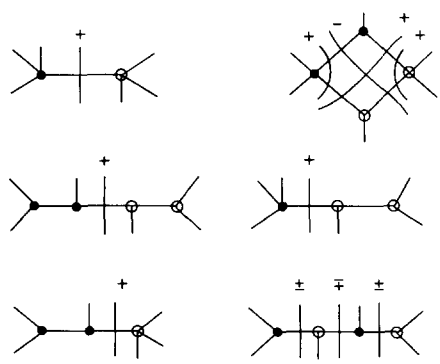


FIG. 6.9. The skeletons for the case $n'=2$.

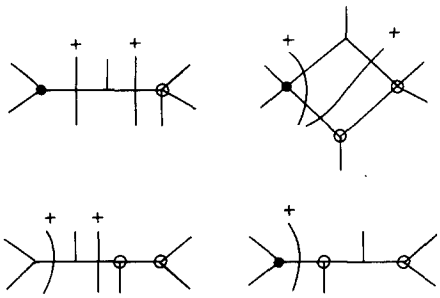


FIG. 6.10. The skeletons and signs for the first case $n'=3$.

not be a preferred line. The signs in the last diagram are forced to be those shown by Eq. (2.25) and the $n'=3$ condition that the last two signs be opposite. These diagrams show that the same condition (6.20) holds in all four cases.

The arguments just given show that for any fixed G and fixed $p \in \hat{L} - W$ there is a sign η such that all of the $r(p) \in \mathcal{R}_\eta^G$ that generate p give precisely the same condition on the cone of analyticity $C(p)$. It remains to show that the corresponding domain of analyticity contains a path that remains in the mass shell. It was shown in Ref. 15 that each individual representation $r(p)$ restricts $C(p)$ to a full half plane in appropriate mass-shell variables, provided the external trajectory lines l_j of $r(p)$ do not all intersect at a common point. This condition is satisfied at each point $p \in \hat{L} - W$, by virtue of property (c).

APPENDIX A: PROPERTIES OF LANDAU SURFACES

Several properties of Landau surfaces are derived in this appendix. A principal result is that each Landau surface $L(D^\sigma)'$ lies in a codimension-one algebraic subvariety of $\mathcal{M}$, where $\mathcal{M}$ is the $(3n-4\equiv N)$-dimensional mass shell restricted by momentum—energy conservation, and where the prime on $L(D^\sigma)'$ means that contributions to $L(D^\sigma)$ from star-graph representations are excluded.

An earlier proof that Landau surfaces lie in algebraic varieties has been given by Chandler[20] and Kershaw.[21] However, in those works the Landau surfaces were defined by the α-form of the Landau equations.[22] These α-form equations are not equivalent to the original Landau equations, for they have extra solutions, which arise from points in α space where a certain discriminant $C(\alpha)$ vanishes. These extra solutions are called second-type singularities.[23] There are cases, for example the Landau surface associated with the ice cream cone diagram, where these so-called second-type singularities cover the entire mass shell $\mathcal{M}$ (i.e., the resolvent is identically zero[24]). These second-type singularities spoil the earlier proofs, but are not pertinent in the present work.

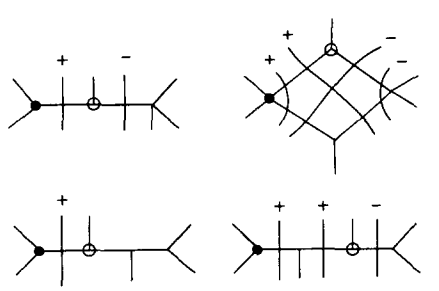


FIG. 6.11. The skeletons and signs for the second case $n'=3$.

The singularities of the functions M^G lie, by virtue of the structure theorem, on a union Landau surfaces $L(D^\sigma)$. Each Landau surface $L(D^\sigma)$ is defined by a corresponding set of equations (4.2a—h). These equations are essentially the original Landau equations. Hence the functions M^G do not have second-type singularities in their original real domain of definition.

Each real Landau surface $L(D^\sigma)$ lies in a corresponding complex Landau surface $L(D)$ defined by Eqs. (4.2a—d). It is convenient to eliminate the conservation law equation (4.2b) by introducing loop momenta. Let q_l be the momentum—energy associated with closed loop l and let k_e be the momentum—energy associated with the open loop e. Then the momentum—energy p_j of line j is

$$p_j = p_j(k, q) \equiv \sum_l \eta_{jl} q_l + \sum_e \eta_{je} k_e, \tag{A1}$$

where the η_{jl} are as in (4.2c), and the η_{je} are the analogous quantities for the open loops e. The $(n-1)$ independent open loops e enter and leave the diagram D via the n external lines of D. For definiteness the open loop e is assumed to leave D via external line e and to enter D via external line n. Then k_e equals $\pm p_e$ for $e=1, \cdots, n-1$. (The lower sign is for initial e.)

With the aid of (A1) the Landau equations (4.2a—d) can be reduced to the equations

$$\sum_j \alpha_j p_j(k, q) \eta_{jl} = 0 \quad (\text{all } l), \tag{A2}$$

$$p_j^2(k, q) - m_j^2 = 0 \quad (\text{all } j), \tag{A3}$$

and

$$\sum \alpha_j^2 - 1 = 0. \tag{A4}$$

The coefficient η_{jl} is zero if j labels an external line. Thus the α_j needed in (A2) include only those corresponding to internal lines. However, parameters α_j corresponding to external lines can be introduced, and the external-loop analog of (A2) considered:

$$\sum \alpha_j p_j(k, q) \eta_{je} = 0 \quad (\text{all } e). \tag{A5}$$

Equations (A2)—(A4) define the Landau surface $L(D)$. Equations (A2)—(A5) define an associated subsurface $L^x(D) \subset L(D)$. Equation (A5) is the requirement that all of the external trajectory lines pass through some common point. That is, each external line is assigned a parameter α_j, and hence also a space—time interval $\Delta_j = \alpha_j p_j$, and all of the external loops are required to be closed loops. This condition is property (c) of Sec. VI. E.

Let $z \equiv (z_1, \cdots, z_s) \equiv (k, q, \alpha)$ denote the set of components of the vectors k_e and q_l together with the α_j for $j \in$ Int. Then the Landau equations (A2)—(A4) are a set of algebraic (i.e., polynomial) equations

$$g_j(z) = 0. \tag{A6}$$

The set of points $z \in \mathbb{C}^s$ that satisfy all of Eqs. (A6) is the algebraic subvariety $V(g) \equiv V$ of $\mathbb{C}^s$. The set V^x defined by

$$V^x \equiv \{z \in V : \text{(A5) is soluble}\} \tag{A7}$$

is the set of point $z \in \mathbb{C}^s$ that satisfy both (A6) and (A5) for some set of α_j with $j \in$ Ext.

Let k be the set of open loop momenta $(k_1, \cdots, k_{n-1})$, and let $g_j^{ex}(k)$ be the functions

$$g_j^{ex}(k) = p_j^2(k) - m_j^2, \quad j \in \text{Ext}. \tag{A8}$$

By the mass shell $\mathcal{H}$ we mean now the algebraic subvariety of k space

$$\mathcal{H} \equiv V(g_j^{ex}) \equiv \{k : g_j^{ex}(k) = 0 \text{ all } \ j \in \text{Ext}\}. \tag{A9}$$

The *restricted mass shell* $\mathcal{H}'$ consists of all points $k \in \mathcal{H}$ except those rare points for which each external momentum–energy vector $p_j(k)$ is parallel to every other one. The restricted mass shell $\mathcal{H}'$ is an N-dimensional analytic submanifold of k space.[15] Thus for each point $k \in \mathcal{H}'$ there is a mass-shell neighborhood $U \subset \mathcal{H}'$ of k, and a set $(Z_1, \cdots, Z_N)$ consisting of N of the components of the loop momenta k_e such that the equation

$$\text{rank}\left(\frac{\partial(g_j^{ex}, Z_m)}{\partial k_e^{\ \mu}}(k)\right) = 4(n-1) \tag{A10}$$

holds for all points k in U. These variables $Z_1, \cdots, Z_N$ are local analytic coordinates of the mass shell in U. That is, U is isomorphic[25] to its image $Z(U) \subset \mathbb{C}^N$. The mapping $Z : \mathbb{C}^{4(n-1)} \to \mathbb{C}^N$ is the projection of k space onto Z space.

Let $\phi : \mathbb{C}^s \to \mathbb{C}^{4(n-1)}$ denote the projection of z space onto k space. Then

$$\phi(V) = L(D), \tag{A11}$$

where $V \equiv V[D]$ is the subvariety of $\mathbb{C}^s$ defined by the Landau equations (4.2a–d) associated with the Landau diagram D.

Some properties of Landau surfaces will now be developed in a series of lemmas. The quantities ϕ, V, Z_m, g_j, etc., occurring in these lemmas are the quantities defined above. And in particular U is an open subset of the restricted mass shell $\mathcal{H}'$ that is isomorphic to its image $Z(U)$ in Z space. Equation (A10) holds for all $k \in U$, and $\phi^{-1}U$ is the set of points z such that $\phi(z) = k$ lies in U. The function $Z(z) \equiv Z(k, q, \alpha)$ is defined by $Z(k, q, \alpha) = Z(k)$.

Lemma A1: For any point $z \in [V \cap \phi^{-1}U]$ there is a set of coefficients $\alpha_j'(z)$ and a set of coefficients $\beta_m(z)$ such that for all $h \in (1, \cdots, s)$

$$\sum_{j \in \text{Int} \cup \text{Ext}} \alpha_j'(z) \frac{\partial g_j}{\partial z_h}(z) = \sum_{m=1}^{N} \beta_m(z) \frac{\partial Z_m}{\partial z_h}(z). \tag{A12}$$

The sum on the left-hand side of (A12) is restricted to the indices $j \in \text{Int} \cup \text{Ext}$, and for any such j

$$g_j(z) = p_j^2(k, q) - m_j^2. \tag{A13}$$

Furthermore, the $\alpha'_j(z)$ for $j \in$ Int are given by

$$\alpha_j'(z) = \alpha_j \quad j \in \text{Int}. \tag{A14}$$

With these restrictions imposed the remaining n coefficients $\alpha_j'(z)$, $j \in$ Ext, and N coefficients $\beta_m(z)$, $m = 1, \cdots, N$, are uniquely defined functions of $z \in [V \cap \phi^{-1}U]$.

Proof: The $g_j(z)$ in (A13) do not depend on the parameters α_j. Neither do the $Z_m(z) \equiv Z_m(k)$. Thus if z_h is any one of the parameters α_j, then (A12) holds trivially: Both sides vanish. If z_h is any component of one of the q_l then the right-hand side of (A12) vanishes because the $Z_m(z)$ depend only on the k_e, not the q_l, and the left-hand side vanishes by virtue of (A1), (A2), and (A13). Thus (A12) holds also for these z_h. Finally, consider those cases in which the z_h are components of the k_e. In these cases (A12) can be written in the form

$$\sum_{j \in \text{Ext}} \alpha_j'(z) \frac{\partial g_j^{ex}(k)}{\partial k_e^{\ \mu}} - \sum_{m=1}^{N} \beta_m(z) \frac{\partial Z_m(k)}{\partial k_e^{\ \mu}} = -\sum_{j \in \text{Int}} \alpha_j \frac{\partial g_j(z)}{\partial k_e^{\ \mu}}, \tag{A15}$$

where the z_k of (A12) are now the $4(n-1)$ components of the $n-1$ vectors k_e. For fixed $z = (k, q, \alpha)$ there are $4(n-1)$ unknowns, namely the $N = 3n-4$ coefficients $\beta_m(z)$ together with the n coefficients $\alpha_j'(z)$ for $j \in$ Ext. Thus, for fixed z, (A15) is a set of $4(n-1)$ linear equations for $4(n-1)$ unknowns. These equations have a unique solution provided

$$\text{rank}\left(\frac{\partial(g_j^{ex}, Z_M)}{\partial k_e^{\ \mu}}(k)\right) = 4(n-1). \tag{A16}$$

This rank condition is just (A10), which holds for all $z \in \phi^{-1}U$. Thus (A12)–(A14) has a unique solution for each $z \in [V \cap \phi^{-1}U]$. QED

Lemma A2: For any fixed $z \in [V \cap \phi^{-1}U]$ the unique set of $\beta_m(z)$ defined in Lemma A1 satisfies the N equations

$$\beta_m(z) = 0, \quad m = 1, \cdots, N, \tag{A17}$$

if and only if $z \in V^x$.

Proof: If z lies in $V \cap \phi^{-1}U$ then the unique set of $\beta_m(z)$ defined in Lemma A1 is defined by (A15), and the points $z \in V^x$ are defined by (A7). But, given (A15), condition (A7) is equivalent to (A17).

Lemma A3: Let $\sigma \subset V$ be any analytic submanifold of $\mathbb{C}^s$ that lies in V. Let $Z \mid \sigma : \sigma \to \mathbb{C}^N$ be the restriction of the mapping $Z(z)$ to the submanifold σ. Let z be any point on $\sigma \cap (V - V^x) \cap \phi^{-1}U$. Then the rank of the mapping $Z \mid \sigma$ at z is less than N:

$$\text{rank}_z Z \mid \sigma < N. \tag{A18}$$

Proof: Let z be any point on $\sigma \cap (V - V^x) \cap \phi^{-1}U$. Equation (A12) holds for all $k \in (1, \cdots, s)$. Thus it holds also if the z_h are replaced by a coordinate set[25] $(z_1', \cdots, z_s')$ such that the coordinates $(z_1', \cdots, z_t')$ are local analytic coordinates of the (t-dimensional) analytic submanifold σ at z. Every function $g_j(z')$ vanishes identically on σ, near $z' = z'(z)$, since σ lies in $V \equiv V(g)$. Thus, for every j,

$$\frac{\partial g_j}{\partial z_h'}(z') = 0 \quad \text{for } k = 1, \cdots, t \tag{A19}$$

at $z' = z'(z)$. Hence the left-hand side of (A12), with the z_h replaced by the z_h', vanishes for $h = 1, \cdots, t$ at $z' = z'(z)$. Thus the right-hand side also vanishes there:

$$\sum_{m=1}^{N} \beta_m(z') \frac{\partial Z_m}{\partial z_h'}(z') = 0 \quad \text{for } h = 1, \cdots, t \tag{A20}$$

at $z'=z'(z)$. Lemma A2 and the requirement of this lemma $z\in(V-V^x)\cap\phi^{-1}U$ ensure that the $\beta_m(z')$ are not all zero at $z'=z'(z)$. But (A20) with some $\beta_m(z')\neq 0$ at $z'=z'(z)$ is equivalent to the condition (A18): Each says that the rank of the matrix $(\partial Z_m/\partial z_i')$ at $z'=z'(z)$, where the z_h' are local analytic coordinates of σ at $z'=z'(z)$, is less than N. QED

Any point on any analytic subvariety is either a manifold point or a nonmanifold point. A *manifold point* of an analytic subvariety is a point having a neighborhood such that the restriction of the subvariety to the neighborhood is an analytic submanifold. The remaining points of the variety are the nonmanifold points.

Almost every point of any analytic subvariety is a manifold point. In fact, the nonmanifold points of any analytic subvariety are confined to an analytic subvariety of lower dimension.[26] This latter subvariety can be similarly decomposed into manifold and nonmanifold points. By this process any analytic subvariety of $\mathbb{C}^s$ can be decomposed into a finite set of disjoint analytic submanifolds σ.

For algebraic subvarieties a similar but more detailed result holds:

Lemma A4: Let t be any integer. Any algebraic subvariety of $\mathbb{C}^t$ can be decomposed into a finite set of disjoint analytic submanifolds σ each of the form $\sigma_{ij}=V_i-V_j$, where V_i is an irreducible algebraic subvariety of $\mathbb{C}^t$, V_j is a proper algebraic subvariety of V_i, and the dimensions of V_i and V_j satisfy

$$\dim V_j<\dim V_i=\dim\sigma_{ij}.$$

Proof: Any algebraic subvariety of $\mathbb{C}^t$ can be decomposed into a finite union of irreducible algebraic subvarieties V_i.[27,28] Each irreducible algebraic subvariety V_i of $\mathbb{C}^t$ has a well-defined dimension $\dim V_i$, which is a nonnegative integer $d\leq t$. The dimension of any algebraic subvariety is the maximum of the dimensions of its irreducible components, and the following property holds[27,28]: If V_i is irreducible and V_j is a proper subvariety of V_i (i.e., $V_j\subset V_i$ but $V_j\neq V_i$), then $\dim V_j<\dim V_i$.

Let V_i be any irreducible component of any algebraic subvariety of $\mathbb{C}^t$. The set V_i is defined as the set of common zeros of some set of polynomials $F\equiv(f_1,\cdots,f_m)$. Let rank_zF be the rank of $(f_1,\cdots,f_r)$ at z. Let r_i be the maximum of rank_zF over $z\in V_i$. Let

$$V_j\equiv\{z\in V_i:\mathrm{rank}_zF<r_i\}.$$

This set V_j is a proper subset of V_i and is defined by the zeros of certain determinants, which are polynomials. Thus V_j is a proper algebraic subvariety of V_j, and hence $\dim V_j<\dim V_i$. Evidently $\mathrm{rank}_zF=r_i$ on σ_{ij}, and hence σ_{ij} is a $(t-r_i)$-dimensional analytic submanifold of $\mathbb{C}^t$.[25,29] Examination of the definitions shows that $\dim\sigma_{ij}=\dim V_i$.

The algebraic subvariety V_j can be decomposed into its irreducible components and the same procedure applied. The dimension of V_j is the maximum of the dimensions of its irreducible components. Thus the dimension decreases at each step, and the process terminates after some finite number of analytic submanifolds σ of the required form are obtained. QED

Definition: Let the *measure* of a subset of $\mathbb{C}^N$ be the real $2N$-dimensional measure of this set considered as a subset of R^{2N}.

Lemma A5:

$$\mathrm{meas}\,Z[(V-V^x)\cap\phi^{-1}U]=0. \tag{A21}$$

Proof: The set V is an algebraic subvariety of $\mathbb{C}^s$. Thus, according to Lemma A4, it can be decomposed into a finite set of analytic submanifolds σ.

Consider any one of these submanifolds $\sigma\subset V$. Sard's theorem,[30] generalized to complex mappings,[31] asserts that

$$\mathrm{meas}\,Z[\{z\in\sigma:\mathrm{rank}_zZ\,|\,\sigma<N]=0. \tag{A22}$$

Lemma A3 asserts that

$$\mathrm{rank}_zZ\,|\,\sigma<N\ \text{ at all } z\in\sigma\cap(V-V^x)\cap\phi^{-1}U.$$

Thus $\mathrm{meas}\,Z[\sigma\cap(V-V^x)\cap\phi^{-1}U]=0$. But V is a finite union of sets σ. Thus $\mathrm{meas}\,Z[(V-V^x)\cap\phi^{-1}U]=0$. QED

Lemma A6: The set $\phi(V-V^x)$ lies in a codimension-one algebraic subvariety V^k of $\mathcal{H}$.

Proof: Chevalley's theorem[32] ensures that the image under the algebraic mapping ϕ of a constructible set in $\mathbb{C}^s$ is a constructible set in $\mathbb{C}^{4(n-1)}$. For any t a constructible set in $\mathbb{C}^t$ is a subset of $\mathbb{C}^t$ that can be constructed as a finite union of sets $\sigma_{ij}\equiv V_i-V_j$, where V_i and V_j are algebraic subvarieties of $\mathbb{C}^t$, V_i is irreducible, and V_j is a proper subvariety of V_i. The set $V-V^x$ is constructible in $\mathbb{C}^s$. Thus $\phi(V-V^x)$ is constructible in $\mathbb{C}^{4(n-1)}$. Let $\sigma_{ij}=V_i-V_j$ be any one of the finite number of sets from which $\phi(V-V^x)$ is constructed. By repeated use of Lemma A4 one can arrange that σ_{ij} is also an analytic submanifold of $\mathbb{C}^{4(n-1)}$ of dimension $\dim\sigma_{ij}=\dim V_i$. Since σ_{ij} lies in $\mathcal{H}$, $\dim\sigma_{ij}\leq\dim\mathcal{H}=N$. If $\dim\sigma_{ij}=N$ then σ_{ij} contains a set that is isomorphic to an open set in $\mathbb{C}^N$.[33] This set must intersect $\mathcal{H}'$, since $\mathcal{H}-\mathcal{H}'$ is a set of dimension less than N. Thus if $\dim\sigma_{ij}=N$ then $\sigma_{ij}\cap\mathcal{H}'$ must contain a set σ' that is isomorphic to an open set in $\mathbb{C}^N$. Let k be a point in σ', and consider the local analytic coordinates $Z_1,\cdots,Z_N$ of $\mathcal{H}'$ in a neighborhood $U\subset\sigma'$ of k. The set $U\cap\sigma'$ is isomorphic to an open set in $\mathbb{C}^N$, and hence must be mapped onto an open set in $\mathbb{C}^N$ by the mapping Z that maps $U\subset\mathcal{H}'$ onto $Z(U)\subset\mathbb{C}^N$. But this contradicts the conclusion of Lemma A5 that $\mathrm{meas}\,Z[(V-V^x)\cap\phi^{-1}U]$ is zero. Thus $\dim\sigma_{ij}=\dim V_i<N$. Hence the constructible set $\phi(V-V^x)$ is a finite union of sets $\sigma_{ij}=V_i-V_j$ of dimension $\dim\sigma_{ij}=\dim V_i<N$. Each set σ_{ij} lies in the mass shell $\mathcal{H}$. Hence the closure of each σ_{ij} lies in the closed set $\mathcal{H}$. But σ_{ij} is dense in V_i and its closure is V_i. Thus $\phi(V-V^x)$ lies in a finite union of irreducible algegraic varieties V_i each of which lies in $\mathcal{H}$ and has dimension $\dim V_i<N$. Thus $\phi(V-V^x)$ lies in a single algebraic subvariety of $\mathcal{H}$ of dimension less than N. Any algebraic variety of dimension less than N is contained in an algebraic variety of dimension $N-1$.[27]

Lemma A7: Let V_i be any irreducible component of V. Let $V_i^x\equiv V_i\cap V^x$. Then $\phi(V_i)$ lies in the V^k of Lemma A6 unless $V_i\equiv V_i^x$.

Proof: Suppose $V_i \neq V_i^x$. Then V_i^x is a proper subvariety of V_i, and $\dim V_i^x < \dim V_i$. But then $V_i - V_i^x$ is dense in V_i, and V_i lies in the closure of $V_i - V_i^x$. But then $\phi(V_i)$ lies in the closure of $\phi(V_i - V_i^x)$, since the mapping ϕ is continuous. Thus $\phi(V_i)$ lies in the closed set V^k that contains $\phi(V_i - V_i^x)$. QED

Lemma A8: $\phi(V)$ lies in a codimension-one algebraic subvariety of $\mathcal{M}$ unless for some i the set V coincides with V_i^x in some neighborhood U of some point $\bar{z} \in V_i^x$, and, moreover, $\phi(V_i^x)$ contains almost all points of $\mathcal{M}$.

Proof: The image $\phi(V_i^x)$ of any V_i^x is constructible and hence either lies in an algebraic subvariety of dimension less than N, or contains almost all points of the mass shell. The former sets $\phi(V_i^x)$ can be combined with V^k to give an algebraic variety $V^{k'}$ of dimension less than N. If any V_i remains then this V_i must coincide with V_i^x, by Lemma A7. But there are points on V_i that lie on no other irreducible component of V. Any such point $\bar{z} \in V_i \equiv V_i^x$ has a neighborhood U such that $V \cap U = V_i \cap U = V_i^x \cap U$. QED

The above lemma says that $\phi(V) = L(D)$ lies in a codimension-one algebraic subvariety of $\mathcal{M}$ unless the conditions (A5) that the external trajectory lines l_j pass through a common point are redundant near some point of V; i.e., the conditions (A5) are implied by the condition (A6).

The quantities $\beta_m(z)$ that occur in (A12) have a simple but important interpretation, which is now described. Let the $p_1, \cdots, p_n$ be the external p_j. The mass-shell constraints ensure that at least one component of each p_j is nonzero. Near real points the energy components p_j^0 are nonzero. In the following discussion it is assumed that these energy components do not vanish, but minor changes would allow the other cases to be encompassed.

For each point $k \in \mathcal{M}'$ the external $p_j(k)$ are not all parallel. Thus for some $j < n$ the vector p_j is not parallel to p_n. Let the labelling be such that p_{n-1} is not parallel to p_n. Then for some space component, which is taken to be the $\mu = 3$ component,

$$p_{n-1}^3/p_{n-1}^0 \neq p_n^3/p_n^0. \tag{A23}$$

If the labels are arranged in this way, then the set $(Z_1, \cdots, Z_N)$ can be taken to be the set consisting of all of the space components of all of the $n-1$ vectors k_e except for k_{n-1}^3. For with this choice the quantity

$$\det\left(\frac{\partial(g_j^e, Z_m)}{\partial k_e^\mu}\right) = 2^n(p_n^0 p_{n-1}^3 - p_{n-1}^0 p_n^3)\prod_{j=1}^{n-2} p_j^0 \tag{A24}$$

is nonzero, by virtue of condition (A23), and our convention about p_j^0. This same condition also ensures that the projections onto (x^0, x^3) space of the external trajectory lines l_{n-1} and l_n of the space–time representation of the point z of V must intersect at some point $(\bar{x}^0, \bar{x}^3)$. Let the origin of time be chosen so that $\bar{t} \equiv \bar{x}^0 = 0$. Let the origin of space be the point at which the trajectory line of particle n intersects the plane $t \equiv x^0 = 0$. Then for each index $j \in \mathrm{Ext}$ let x_j be the 3-vector from the origin of space to the point at which the trajectory line l_j of external particle j intersects the plane $t = 0$.

In general, each of the three components x_j^μ ($\mu = 1, 2, 3$) of each vector x_j is a complex number. However, the components x_n^1, x_n^2, x_n^3, and x_{n-1}^3 all vanish, by virtue of the choice of origin. The remaining $3n-4$ coordinates x_e^μ ($\mu = 1, 2, 3$) are, apart from a factor of two, the coefficients $\beta_m(z)$. In particular, if $Z_m = k_{e(m)}^{\mu\,(m)} = p_{e(m)}^{\mu\,(m)}$, then

$$\beta_m(z) = 2x_{e(m)}^{\mu\,(m)}(z), \tag{A25}$$

where the argument z of $x_e^\mu(z)$ signifies that x_e^μ depends on the point $z \in V$, since, of course, the space–time representation of D depends on $z \in V$. For each $j \in \mathrm{Ext}$ the parameter $\alpha_j'(z)$ is the complex number such that $\Delta_j = \alpha_j' p_j$ is the space–time vector to the point $(0, x_j)$ from the point v_j on the external trajectory j where this trajectory joins D. These determinations of the $\alpha_j'(z)$ and of the $\beta_m(z)$ ensure that (A15), and hence (A12), holds.

The identification (A25) means that the N-vector $\beta(z)$ with components $\beta_m(z)$, is determined in a simple way by the locations of the external trajectories l_j of the space–time representation corresponding to $z \in V$. This vector $\beta(z)$ is zero, as noted in Lemma A2, if and only if the external trajectory lines all pass through a common point.

The N-vector $\beta(z)$ discussed above is also essentially the normal to the Landau surface at the image of z:

Lemma A9: Suppose σ is an analytic submanifold of $\mathbb{C}^s$ that lies in V. Suppose z is a point of $\sigma \cap (V - V^x) \cap \phi^{-1}U$ such that the restriction $Z \mid \sigma$ of the mapping $Z(z)$ to σ has rank $N-1$ at z. [Rank $> N-1$ is excluded by (A18)]. Then there is a σ-neighborhood of z, $U_z \subset [\sigma \cap (V - V^x) \cap \phi^{-1}U]$, such that $Z(U_z)$ is an analytic submanifold of $\mathbb{C}^N$ of dimension $N-1$. This $N-1$ dimensional analytic submanifold $Z(U_z)$ lies in $L(D)$. The normal to $Z(U_z)$ at $Z(z)$ is well defined and nonzero, and it is equal to $\beta(z)$, apart from a nonzero scale factor.

Proof: The rank of $Z \mid \sigma$ on $\sigma \cap (V - V^x) \cap \phi^{-1}U$ is, by virtue of Lemma A3, at most $N-1$. Since the rank can decrease only on the zeros of certain determinants the rank must be $N-1$ in some σ-neighborhood of z. Thus there is a σ-neighborhood of z, $U_z \subset [\sigma \cap (V - V^x) \cap \phi^{-1}U]$, such that $Z(U_z)$ is an $(N-1)$-dimensional analytic submanifold of $\mathbb{C}^N$.[34] The normal to any codimension-one analytic submanifold of $\mathbb{C}^N$ is well defined and nonzero. Let $(z', \cdots, z_s')$ be a set of local analytic coordinates[25] such that $z_1', \cdots, z_t'$ are local coordinates of σ near the point z. Then the set of t N-vectors T^h, with $h = 1, \cdots, t$, whose components are $T_m^h = \partial Z_m/\partial z_h(z)$ span an $(N-1)$-dimensional subspace of $\mathbb{C}^N$, since the rank of $Z \mid \sigma$ at z is $N-1$. This $(N-1)$-dimensional subspace of $\mathbb{C}^N$ is the $(N-1)$-dimensional tangent space to $Z(U_z)$ at z. But then (A20) is the condition that the N-vector $\beta(z)$ be a nonzero multiple of the normal to $Z(U_z)$ at z. QED

This result that the positions of the external trajectory lines l_j determine the normal to the Landau surface was derived earlier[15] by another method, for positive-α surfaces. That earlier method involves nonalgebraic functions that develop singularities when any of the internal α_j vanish. The present purely algebraic method

extends the earlier results to the points z where one or more $\alpha_j = 0$.

Definitions: R is a real $R \subset V$; $\phi_R \equiv \phi/R$; and $\hat{L}(D^\sigma) \equiv \phi_R(V) \equiv \phi(R)$.

Lemma A10: Let D be any Landau diagram, and let V be the corresponding variety in $\mathbb{C}^s$. Suppose $\phi_R(V^x)$ is confined to a codimension-two algebraic subvariety of $\mathcal{M}$. Then there is a codimension-two algebraic subvariety $W(D)$ of $\mathcal{M}$ with the following four properties:

(1) $\mathcal{M} - \mathcal{M}'$ lies in $W(D)$.

(2) If z is a real point of R that is mapped by ϕ into $\hat{L}(D^\sigma) - W(D)$, and if $r(z)$ is the space–time representation of D that corresponds to z, then the external lines l_j of $r(z)$ do not all intersect at a common point; i.e., $\beta(z) \neq 0$.

(3) The set $\hat{L}(D^\sigma) - W(D)$ is either empty or a codimension-one analytic submanifold of $\mathcal{M}'$ that is confined to a codimension-one algebraic subvariety of $\mathcal{M}$.

(4) If z and z' are two points of R that correspond to the same point $k \in \hat{L}(D^\sigma) - W(D)$, then $r \equiv r(z)$ and $r' \equiv r(z')$ are externally similar, i.e., there is a positive or negative scale change and an overall space–time translation that brings the external trajectory lines l_j' of r' into coincidence with the corresponding lines l_j of r.

Proof: The required $W(D)$ is

$$W(D) = W_1(D) \cup W_2(D) \cup W_3(D) \cup W_4(D), \tag{A26}$$

where $W_1(D)$ is $\mathcal{M} - \mathcal{M}'$; $W_2(D)$ is the codimension-two algebraic subvariety of $\mathcal{M}$ that by hypothesis contains $\phi_R(V^x)$; $W_3(D)$ is the union of the algebraic subvarieties V_i and V_j of dimension less than $N-1$ that arise in the decompotition, via the Chevalley theorem and Lemma A4, of the constructible set $\phi(V - V^x)$ into analytic submanifolds $\sigma_{ij} = V_i - V_j$; and $W_4(D)$ is the closure of the image in $\mathcal{M}$ of the set of points $z \in R$ such that $\text{rank}_z \phi \mid \sigma_z < N-1$. Here σ_z is the analytic submanifold $\sigma_{ij} \subset V_i \subset V$ that arises in the decomposition of V by means of Lemma A4, and that contains z, and $\phi \mid \sigma_z$ is the restriction of ϕ to σ_z.

The set $W_1(D)$ is an algebraic subvariety of $\mathcal{M}$ of dimension less than $N-2$. Its presence in $W(D)$ ensures property (1).

The set $W_2(D)$ is a codimension-two algebraic subvariety of $\mathcal{M}$ by hypothesis. Its presence in W ensures property (2), because the conditions that z be real point of V and that the external trajectories l_j of $r(z)$ have a common point is precisely the condition that z belong also to V^x.

The set $W_3(D)$ is a finite union of algebraic subvarieties of $\mathcal{M}$ of dimension less than $N-1$, and is thus an algebraic subvariety of $\mathcal{M}$ of dimension less than $N-1$. The remaining sets V_i and σ_{ij} in the decomposition of $\phi(V - V^x)$ have dimension $N-1$, by virtue of Lemma A6, and they include all points of $L(D^\sigma) - W(D)$, since $\phi_R(V^x)$ lies in $W(D)$.

The set $W_4(D)$ is

$$W_4(D) = \text{closure}\, \phi\{z \in R : \text{rank}_z \phi \mid \sigma_z < N-1\}. \tag{A27}$$

The set $\{z \in R : \text{rank}_z \phi \mid \sigma_z < N-1\}$ is a constructible set, and hence so, by virtue of the Chevalley theorem,[32] is its image $\phi\{z \in R : \text{rank}_z \phi \mid \sigma_z < N-1\}$ in $\mathcal{M}$. But arguments analogous to those of Lemma A5 show that the image under ϕ of the set $\{z : \text{rank}_z \phi \mid \sigma_z < N-1\}$ cannot contain any set $\sigma_{ij} = V_i - V_j$ of dimension $N-1$. Hence the closure of $\phi\{z \in R : \text{rank}_z \phi \mid \sigma_z < N-1\}$ is a finite union of sets V_i of $\dim V_i < N-1$. Thus $W_4(D)$ is an algebraic subvariety of $\mathcal{M}$ of dimension less than $N-1$.

Each point $k \in \hat{L}(D^\sigma) - W(D)$ is a point near which $\hat{L}(D^\sigma)$ is a codimension-one analytic submanifold of $\mathcal{M}$, by property (3). Hence the normal to $\hat{L}(D^\sigma)$ at any point of $\hat{L}(D^\sigma) - W(D)$ is well defined. On the other hand, every point $z = R$ that maps to any point $k \in \hat{L}(D^\sigma) - W(D)$ is a point $z \in R$ where $\text{rank}_z \phi \mid \sigma_z = N-1$. Thus, by virtue of Lemma A9, the vector $\beta(z)$ associated with any $z \in R$ that maps to any $k \in \hat{L}(D^\sigma) - W(D)$ is a multiple of the well-defined normal to $\hat{L}(D^\sigma)$ at k. Therefore, all of the vectors $\beta(z)$ that are associated in this way with any given $k \in \hat{L}(D^\sigma) - W(D)$ are nonzero multiples of each other. Since all of the representations $r(z)$ that generate any fixed point $k \in L(D^\sigma)$ have their corresponding external trajectory lines parallel, this equality (up to scale change) of the N-vectors $\beta(z)$ guarantees that these representations $r(z)$ are all externally similar. Thus property (4) holds. QED

The results obtained above refer to a single Landau surface $L(D)$. Let B be any bubble diagram and let $F^B(k)$ be the corresponding bubble-diagram function [see Eq. (3.4)]. Let $L(B)$ be

$$L(B) = \bigcup_{D^\sigma \subset B} L(D^\sigma). \tag{A28}$$

The structure theorem says that the singularities of $F^B(k)$ (at real k) are confined to $L(B)$ [see (4.6)].

For any given B there is an infinite set of $D^\sigma \subset B$. One trivial way in which this set becomes infinite is illustrated in Fig. A1. As n runs from 2 to infinity the diagram in this figure generates an infinite set of diagrams. If the multiple loops in all these diagrams are formed from a single fixed pair of particles, and if the signs σ_j are all plus, or are all minus, then each of these diagrams gives the same surface $L(D^\sigma)$. These D^σ are examples of nonbasic diagrams, which can, as we shall see, be ignored.

A *nonbasic diagram* D^σ is a diagram that has a reducible part. A *reducible* part is a part R such that: (1) All the lines of R have the same sign + or −; (2) in every space–time representation of D^σ all the lines of R lie on one common space–time line; and (3) some vertex of D^σ is connected only to lines of R.

Properties (1) and (2) ensure that all the vertices of R that satisfy (3) can be shifted along the common

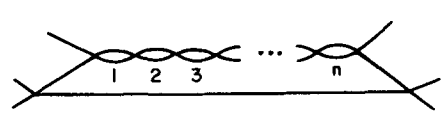


FIG. A7. A typical nonbasic diagram.

space–time line, and brought up against vertices that do not satisfy (3). By merging these coincident vertices one can construct for each nonbasic D^σ a basic $D_\beta{}^\sigma$ with the property that each representation r of D^σ is externally equivalent to some representation r' of $D_\beta{}^\sigma$.

A contractible part of a $D^\sigma \subset B$ must be a part of the diagram D_b^+ or D_b^- that replaces some plus or a minus bubble b of B. Otherwise some sign would be a "sign" $\sigma_j = \pm$, hence not a well-defined sign plus or minus. Thus the restriction of $D^\sigma \subset B$ to basic diagrams is equivalent to the restriction of the constituent diagrams D_b^+ and D_b^- to basic diagrams $D_{b\beta}^+$ and $D_{b\beta}^-$.

Since every representation r of a nonbasic diagram D^σ is externally equivalent to a representation r' of a basic diagram $D_\beta{}^\sigma$ one may write

$$L(B) = \bigcup_{D_\beta{}^\sigma \subset B} L(D^\sigma), \tag{A29}$$

where $D_\beta{}^\sigma$ is a basic diagram.

Lemma A11: Let $\mathcal{R}$ be any bounded region in k space. Then only a finite set of basic D_β^+ gives surfaces $L(D_\beta^+)$ that intersect $\mathcal{R}$. And only a finite set of basic D_β^- gives surfaces $L(D_\beta^-)$ that intersect $\mathcal{R}$.

Proof: This was proved in Ref. 35.

Remark: Because of the mass-shell and positive-energy ($p_j{}^0 > 0$) constraints a region $\mathcal{R}$ in k space is bounded if and only if the total energy of the process $\frac{1}{2}\sum p_j{}^0$ is bounded.

Lemma A12: Let $\mathcal{R}$ be any bounded region in k space. Let B be any fixed bubble diagram. Then only a finite number of basic $D_\beta{}^\sigma \subset B$ give surfaces $L(D_\beta{}^\sigma)$ that intersect $\mathcal{R}$.

Proof: The flow-diagram ordering condition of Sec. VI. A, together with the conservation law condition (4. 2b), and the positivity energy condition (4. 2g), ensures that the total energy entering any bubble b of B is no greater than the total energy entering B. That is, a bound in k space implies a bound on the energy entering each individual bubble. But the Landau equations corresponding to $D_\beta{}^\sigma$ cannot be satisfied unless the Landau equations for each of the constituent parts $D_{b\beta}^+$ and $D_{b\beta}^-$ can be satisfied. Thus the bound, mentioned above, on the energy entering each bubble b, together with Lemma A11, implies that for each b the corresponding $D_{b\beta}^+$ or $D_{b\beta}^-$ must be one of a certain finite set of diagrams. However, only a finite set of diagrams $D_\beta{}^\sigma$ can be constructed from the finite sets of $D_{b\beta}^\pm$'s inserted in all possible ways into the finite set of bubbles b of B.

Definitions:

$$D_\eta{}^G \equiv \{D^\sigma : D^\sigma \subset B \in \mathcal{B}_\eta{}^G\}, \tag{A30}$$

$$\mathcal{D} \equiv \bigcup_{G,\eta} D_\eta{}^G. \tag{A31}$$

Let $\mathcal{R}$ be a region in k space and let $N_\beta(\mathcal{R})$ be the number of basic $D_\beta{}^\sigma \in \mathcal{D}$ such that $L(D_\beta{}^\sigma) \cap \mathcal{R} \neq \phi$.

Lemma A13: If $\mathcal{R}$ is bounded then $N_\beta(\mathcal{R})$ is finite.

Proof: The number of pairs (G, η) is finite. For each (G, η) the set $\mathcal{B}_\eta{}^G$ is constructed so that in any bounded region $\mathcal{R}$ only a finite set of $B \in \mathcal{B}_\eta{}^G$ satisfy the conservation law, positive energy, and mass-shell conditions associated with the explicit lines of B. These conditions are included among the Landau equations associated with the diagrams $D_\beta{}^\sigma \subset B$. Thus only a finite set of $B \in \mathcal{B}_\eta{}^G$ can have $D_\beta{}^\sigma \subset B$ that give surfaces $L(D_\beta{}^\sigma)$ that intersect any bounded $\mathcal{R}$. But then the finiteness of $N_\beta(\mathcal{R})$ follows from Lemma A12. QED

Definition: $\mathcal{D}_G \equiv D_+{}^G \cup D_-{}^G$.

Lemma A14: Suppose for each $D_\beta{}^\sigma \in \mathcal{D}_G$ that there is a real region $R_\beta{}^\sigma \subset V_\beta{}^\sigma$ such that $\phi(R_\beta{}^\sigma \cap V^x)$ is confined to a corresponding codimension-two algebraic subvariety of $\mathcal{H}$. Then there is a codimension-two algebraic subvariety W of $\mathcal{H}$ such that

(1) $\mathcal{H} - \mathcal{H}'$ lies in W;

(2) if z lies in $\cup R_\beta{}^\sigma$ and $r \equiv r(z)$

corresponds to a point $k \notin W$ then the external trajectory lines l_j of r do not all pass through a common point;

(3) the set $\cup \phi(R_\beta{}^\sigma) - W$ is either empty or a codimension-one analytic submanifold of $\mathcal{H}'$ that is confined to a codimension-one algebraic subvariety of $\mathcal{H}$.

(4) if z and z' lie in $\cup R_\beta{}^\sigma$ and if both z and z' correspond to the same point $k \notin W$, then $r \equiv r(z)$ and $z' \equiv r(z')$ are externally similar.

Proof: This result follows immediately from Lemma A10 and A13, for the set

$$W \equiv \bigcup_{D_G{}^\sigma \in \mathcal{D}_G} W(D_\beta{}^\sigma)$$

restricted to any bounded region is a finite union of algebraic varieties. Hence W is an algebraic variety.

Lemma A15: Let $L_G{}^S$ be the set of points k generated by at least two externally similar but not externally equivalent representations of diagrams $D_\beta{}^\sigma \subset \mathcal{D}_G$. [The pertinent definitions lie between Eqs. (6. 16) and (6. 17).] Let $L_G{}^E$ be the closure of the union of the codimension-one analytic submanifolds contained in $L_G{}^S$. Then $L_G{}^S - L_G{}^E$ lies in a codimension-two algebraic subvariety.

Proof: It is sufficient to prove that $L_G{}^S$ is constructible, for in this case $L_G{}^S$ is a finite union analytic submanifold of the form $\sigma_{ij} = V_i - V_j$ where V_i and V_j are algebraic varieties and V_j is a proper subvariety of V_i. The set $L_G{}^E$ is then the closure of the union of the σ_{ij} of codimension-zero or codimension-one, and $L_G{}^S - L_G{}^E$ is contained in the finite union of the codimension-two-or-more algebraic varieties V_i into which $L_G{}^S$ is decomposed.

To show that $L_G{}^S$ is constructible it is sufficient, by virtue of the Chevalley theorem, to show that it is the image under an algebraic mapping of a constructible set. It is enough to deal with k space restricted to an arbitrarily large sphere centered at the origin, because

the result to be established is that $L_G{}^S - L_G{}^E$ restricted to every such sphere lies in a codimension-two algebraic subvariety. But then, by virtue of Lemma A13, it is sufficient to consider only a finite set of $D_\beta{}^\sigma \subset \mathcal{D}_G$. Let $\Gamma \equiv \pi \Gamma_\beta{}^\sigma$ be the tensor product of the spaces $\Gamma_\beta{}^\sigma$ of variables k, q, α, and v, associated with these $D_\beta{}^\sigma$. Here k is the set of external 4-momenta, q is the set of internal 4-momenta, α is the set of α parameters, and v is the set of 4-vectors that specify the positions of the vertices of the space–time representations of $D_\beta{}^\sigma$. The pre-image of $L_G{}^S$ in Γ is a constructible set. One sees this by considering in turn each pair $(D_\beta{}^\sigma, D_\beta{}^{\sigma'})$ of $D_\beta{}^\sigma$'s in the finite set, and noting that the condition for the coincidence of the two external trajectory lines l_e and l_e' connected to external line e in $D_\beta{}^\sigma$ and $D_\beta{}^{\sigma'}$, respectively, is the condition that the difference $\Delta_e \equiv v_{j(e)} - v'_{j(e)}$ of the positions of the external vertices connected to e in $D_\beta{}^\sigma$ and $D_\beta{}^{\sigma'}$ be parallel to p_e. This condition can be expressed algebraically by the vanishing of the Gram determinant of the two vectors Δ_e and p_e. Imposing this requirement for each e, and then deleting the set where all $\Delta_e = 0$ gives a constructible set. The union of these constructible sets over the finite set of pairs $(D_\beta{}^\sigma, D_\beta{}^{\sigma'})$ gives a constructible set whose projection onto k space is $L_G{}^S$. Then the Chevalley theorem implies that $L_G{}^S$ is constructible. QED

The conclusions (a), (b), and (c) listed below Eq. (6.19) in the main text follow from Lemmas A14 and A15. The sets $R_\beta{}^\sigma \subset V_\beta{}^\sigma$ of Lemma A14 are taken to be the sets of real points of $V_\beta{}^\sigma$ that do not generate points k of the exceptional set $L_G{}^E$. Then, by virtue of Lemma A15, points $k \in \cup\, \phi(R_\beta{}^\sigma)$ generated by pairs of externally similar representations that are not externally equivalent are confined to some codimension-two algebraic subvariety W'. This subvariety W' contains $\cup\, \phi(R_\beta{}^\sigma \cap V_\beta{}^{\sigma x})$, which ensures the condition of Lemma A14. The choice $W_G = W \cup W'$ ensures that conclusion (b) follows from consequence (4) of Lemma A14, to the extent that the diagrams D^σ in condition (b) are restricted to basic diagrams $D_\beta{}^\sigma$. Moreover, conditions (a) and (c) follow from consequences (3) and (2), in this case. However, if the properties (a), (b), and (c) hold with the D^σ restricted to basic diagrams $D_\beta{}^\sigma$ then these three properties also hold without this restriction. For each representation of a nonbasic D^σ is associated with a unique externally equivalent representation of a basic diagram $D_\beta{}^\sigma$ that corresponds to the same point k. The fact that external equivalence is an equivalence relation then extends the properties (a), (b), and (c) restricted to the set of basic diagrams $D_\beta{}^\sigma \in \mathcal{D}_G$ to the set of all $D^\sigma \in \mathcal{D}_G$.

APPENDIX B: COMBINATORIC IDENTITIES

Let H be any finite set. Suppose that

$$A_H = \sum_{H' \subset H} (-1)^{n(H')} B^{H'} \tag{B1}$$

and

$$B^{H'} = \sum_{H'' \subset H'} (-1)^{n(H'')} C_{H''}, \tag{B2}$$

where the sums run over all distinct subsets H' of H and H'' of H', respectively, and $n(F)$ is the number of elements of F. Substitution of (B2) into (B1) gives

$$A_H = \sum_{H'' \subset H} X_{HH''} C_{H''}, \tag{B3}$$

where

$$X_{HH''} = \sum_{H'' \subset H' \subset H} (-1)^{n(H') - n(H'')}. \tag{B4}$$

Clearly $X_{HH} = 1$. Comparison of the expression for the remaining quantities $X_{HH''}$ with the binomial expansion of $(1-1)^{n(H)-n(H'')}$ shows that X is the unit matrix:

$$X_{HH''} = \delta_{HH''} \equiv \begin{cases} 1 & \text{for } H = H'', \\ 0 & \text{for } H \neq H''. \end{cases} \tag{B5}$$

This result implies (2.5). It also implies that the quantities $T_H{}^G$ defined for every G and H by

$$T_H{}^G = \sum_{G' \subset G} (-1)^{n(G')} T_{HG'} \tag{B6}$$

satisfy

$$T_{HH'} = \sum_{H'' \subset H'} (-1)^{n(H'')} T_H{}^{H''}. \tag{B7}$$

The substitution of (B7), with $H = \phi$, into (B6) gives

$$\begin{aligned} T_H{}^G &= \sum_{G' \subset G} (-1)^{n(G')} \sum_{H'' \subset H G'} (-1)^{n(H'')} T^{H''} \\ &= \sum_{G' \subset G} (-1)^{n(G')} \sum_{G'' \subset G'} (-1)^{n(G'')} \sum_{H' \subset H} (-1)^{n(H')} T^{G'' H'} \\ &= \sum_{G'' G'' \subset G' \subset G} (-1)^{n(G') - n(G'')} \sum_{H' \subset H} (-1)^{n(H')} T^{G'' H'} \\ &= \sum_{H' \subset H} (-1)^{n(H')} T^{G H'}, \end{aligned} \tag{B8}$$

where (B5) is used to get the last line. This result is (4.17b). The result (4.17c) follows from repeated application of the identity

$$T_H{}^G = T_H{}^{Gk} + T_{Hk}{}^G, \tag{B9}$$

which follows from (B6), which is (4.17a), via the equation

$$\begin{aligned} T_H{}^{Gk} &= \sum_{G' \subset G} (-1)^{n(G)} T_{HG'} + \sum_{G' \subset G} (-1)^{n(G)+1} T_{HkG'} \\ &= T_H{}^G - T_{Hk}{}^G. \end{aligned} \tag{B10}$$

*Work supported in part by the Western Illinois University.
†Work supported by the U.S. Atomic Energy Commission.

[1] A.H. Mueller, Phys. Rev. **D2**, 2963 (1970).
[2] C.I. Tan, Phys. Rev. D **5**, 1476 (1972).
[3] H.P. Stapp, Phys. Rev. D **3**, 3177 (1971).
[4] K.E. Cahill and H.P. Stapp, "Optical Theorems and Steinmann Relations," to be published in Ann. Phys. (N.Y.).
[5] C.E. DeTar and J.H. Weis, Phys. Rev. D **4**, 3141 (1971); I.G. Halliday, Nucl. Phys. **B33**, 285 (1971); J.H. Weis, Phys. Rev. **D5**, 1043 (1972); A.R. White, Nucl. Phys. **B39**, 477 (1972); A.R. White, Nucl. Phys. **B50**, 96 (1972); J.W. Weis, Phys. Rev. **D6**, 2823 (1972).
[6] J.H. Weis, Phys. Lett. **43B**, 487 (1973).
[7] J. Coster and H.P. Stapp, J. Math. Phys. **10**, 371 (1969).
[8] Ref. 7, Appendix A.
[9] J. Coster and H.P. Stapp, J. Math. Phys. **11**, 1441 (1970).
[10] Ref. 9, Appendix A.
[11] J. Coster and H.P. Stapp, J. Math. Phys. **11**, 2743 (1970), Sec. 3.

[12]Ref. 11, Sec. 2F.
[13]D. Iagolnitzer and H.P. Stapp, Commun. Math. Phys. **14**, 15 (1969).
[14]S. Coleman and R.F. Norton, Nuovo Cimento **38**, 438 (1965).
[15]C. Chandler and H.P. Stapp, J. Math. Phys. **9**, 1548 (1969).
[16]H.P. Stapp, J. Math. Phys. **9**, 1548 (1969), Eq. (3.5).
[17]Ref. 16, Sec. 3.
[18]F. Pham, Ann. Inst. Henri Poincaré **6A**, 89 (1967), or Ref. 7, Eq. (3.14).
[19]H. Epstein, J. Math. Phys. **1**, 524 (1960), Appendix; J. Bros, "Axiomatic Field Theory," in *High Energy Physics and Elementary Particles* (International Atomic Energy Agency, Vienna, 1965), p. 90.
[20]J. Wright and H.P. Stapp, Nuovo Cimento **57A**, 475 (1968), Appendix D.
[21]D.S. Kershaw, Phys. Rev. D **5**, 1976 (1972).
[22]R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne, *The Analytic S Matrix* (Cambridge, U.P., Cambridge, 1966), Eqs. (2.2.11), (2.2.5), and (1.5.30).
[23]Ref. 22, p. 116.
[24]C. Chandler, private communication.
[25]R.C. Gunning and H. Rossi, *Analytic Functions of Several of Several Complex Variables* (Prentice-Hall, Englewood Cliffs, N.J., 1965), p. 17.
[26]Ref. 25, pp. 111, 113, 115, 116, 141.
[27]David Mumford, "Introduction to Algebraic Geometry, Preliminary Version of First Three Chapters," Mathematics Department of Harvard University, 1966, p. 15. 75.
[28]F.S. Macaulay, *The Algebraic Theory of Modular Systems*, Cambridge Tracts in Mathematics and Mathematical Physics 19 (Cambridge U.P., Cambridge, 1916), p. 28.
[29]Ref. 25, p. 160.
[30]S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N.J., 1965), p. 47.
[31]Ref. 20, Appendix B. Lemma B.
[32]Ref. 27, p. 97; Ref. 20, Appendix B.
[33]Ref. 25, p. 17, 111, 116.
[34]Ref. 25, p. 160.
[35]H.P. Stapp, J. Math. Phys. **8**, 1606 (1967).

Self-interacting, boson, quantum, field theory and the thermodynamic limit in *d* dimensions*

George A. Baker, Jr.

Applied Mathematics Department, Brookhaven National Laboratory, Upton, New York 11973 †
and Theoretical Division, Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87544
(Received 2 December 1974)

By use of a finite volume, lattice approximation, we set up an approximation to the analytic continuation of a polynomial, self-interacting boson quantum field theory from Minkowski space to Euclidean space. The infinite volume limit for various boundary conditions is shown to exist and to be asymptotic to the perturbation expansion in the coupling constant g at $g = 0$. For $g{:}\phi^4{:}_d$ theory we prove mass renormalizability and show how, by use of Nelson's reconstruction theorem, the corresponding Minkowski space quantum field theory can be obtained. We discuss, at least for $d \geq 4$, how statistical mechanical techniques, used to analyze the Ising model in the critical region just above the critical temperature, can be used to compute the properties of quantum field theory.

1. INTRODUCTION AND SUMMARY

Since the introduction of quantum field theory, there has remained a question as to whether, and under what circumstances, it was a real theory. Did it make precise predictions for physical phenomena, and if so how could one calculate them. In the case of quantum electrodynamics, the electric charge, or coupling constant, is sufficiently small so that a perturbation expansion in the coupling constant gives sufficient accuracy for experimental purposes using only a few terms. The theoretical question of whether the theory defines a unique, precise result however has not been resolved. The problem is more severe and of greater practical importance in the case of the strong interactions where the coupling constant is larger and accuracy adequate for experimental purposes cannot be obtained from the first few terms of the invariant perturbation series. Indeed, as the series is likely divergent, simple series summation is of limited value in obtaining accurate answers.

A substantial amount of work has been done on this problem in the case of two-dimensional systems (one space and one time dimension). The reader is referred to Glimm and Jaffe.[1] Dimock[2] has established that for self-interacting Boson theory in two dimensions the perturbation series in the coupling constant is asymptotic to an infinitely differentiable theory, at least over some small interval in the coupling constant. However, the complete resolution of the theoretical problem has not yet been given even in two dimensions, much less the resolution of the practical question.

A fundamental advance was made by Nelson.[3] He was able to show that a quantum field theory could be constructed from a simpler object. Quantum field theories are related to the properties of Minkowski space where the distance between two points is given by $ds^2 = -dx^2 - dy^2 - dz^2 + dt^2$ instead of the usual Euclidean formula with all plus signs. (The velocity of light is taken to be unity.) Schwinger,[4] generalizing a technique used previously in perturbation theory, studied the analytic continuation of the vacuum expectation values to imaginary times. When the continuation is carried out, the characteristic symmetry group becomes the orthogonal group in four-dimensional, Euclidean space rather than the Lorentz group in Minkowski space. Symanzik[5] then made the important discovery, using multiple Wiener integrals of functionals, that, at least for certain model systems described by Lagrangians, Euclidean, Boson, quantum field theory was mathematically very similar to a classical statistical mechanical system. Nelson[3] isolated the important Markov property (roughly, if one has complete information about a system on the boundary of a region no additional information about the interior is gained by a further knowledge of the exterior) and showed how to reconstruct a quantum field theory from a Euclidean field theory with this property. It is not however particularly easy to give examples of Markov fields. Nelson[6] has, however, shown that the free fields of mass m in d dimensions have this property and hence lead by way of his reconstruction theorem to free field, Boson, quantum field theory.

If a Euclidean field theory satisfies the properties required by Nelson[3] or more generally by Osterwalder and Schrader[7] then a quantum field theory may be constructed from it. Thus their work has reduced the problem of existence and perhaps also computation to the study of the Euclidean problem. The fundamental starting point is the Feynman-Kac formula[8] for the vacuum expectation value

$$\langle \Omega_0, \exp(-tH)\Omega_0 \rangle = E(\exp(-gV)),$$

where H is $d-1$ dimensional, t is a pseudo-time, and V is the d-dimensional interaction. The left-hand side (Fock-space expression) has no time-ordering operator because that operator acts only on the real part of the times which are all zero. The right-hand side is the expectation value with respect to a Markovian, Gaussian random process. The fundamental Feynman—Kac formula has been extended by Osterwalder and Schrader[9] to boson—fermion models and, though technically more complex, should serve as the basis for an extension of the present work to such systems.

A logical next step in an effort to establish a computable quantum field theory is to introduce a partition of

space into cells and to consider only one field variable per cell. One then has the analogy to a crystal lattice. Unknown to us at the beginning of our investigation, a similar approximation was extensively studied by Guerra, Rosen, and Simon[10] for the case of two space-time dimensions. Once the lattice approximation has been introduced over a finite volume there are two limits to be taken. These limits are (i) the volume goes to infinity and (ii) the lattice spacing goes to zero. The latter limit is equivalent to the usual ultraviolet cutoff going to infinity. Our approach is first to take the thermodynamic limit, that is, the volume goes to infinity, and then to let the lattice spacing go to zero. We will work in an arbitrary ($d \geq 2$) number of dimensions. The approach of Guerra, Rosen, and Simon[10] differs from ours fundamentally in that they contemplate first taking the lattice spacing to zero and then the volume to infinity.

Our work in this paper is concerned with self-interacting, Boson field theory. Aside from $g:\phi^2:_d$ field theory, which is equivalent to changing the mass in free field theory and has no scattering, our most complete results are for $g:\phi^4:_d$ field theory. We discuss these results in Section 7. Since mass renormalization simply means adjusting a parameter of the underlying theory, the bare mass, so that the mass predicted by the theory will equal the experimental mass, a theory will be mass renormalizable if that condition can be met. We show (with a qualification) for $g:\phi^4:_d$ theory that the bare mass can be adjusted so that the mass predicted ranges over $0 < m < \infty$. Furthermore, for the lattice spacing $\Delta > 0$ the mass is a monotonic, uniformly continuous (in the volume) function of the bare mass for all values of the coupling constant. Consequently, the mass renormalization can be carried out. We further show that the cluster property, i.e., exponential decay of the field correlations at long distances, holds. Furthermore, the Schwinger functions and the free energy (the fundamental objects of the theory) are continuous functions of the coupling constant for all g, and are defined in a nonzero region by the asymptotic, perturbation series at $g=0$, at least if the lattice spacing $\Delta > 0$. The basic Wightman functions, which are the fundamental objects of axiomatic field theory, are constructed by analytic continuation of the Schwinger functions by means of Nelson's reconstruction theorem[3] or Osterwalder and Schrader's[7] results.

An outline of our procedures is as follows: In the second section we set up on a finite volume with a finite lattice spacing a set of Gaussian random fields which are coupled to each other by the d-dimensional difference approximation to the ordinary free field Lagrangian

In the third section we define the normal ordered product on a discrete lattice.

In the fourth section we show the existence, by the use of a number of inequalities known in statistical mechanics, of the infinite volume limit at fixed lattice spacing. These limits depend explicitly on the general, semibounded polynomial interaction assumed, and on the type of boundary conditions imposed. We show generally that certain inequalities hold between results for different types of boundary conditions, and for some cases we can establish equality.

In the fifth section we show the relation of the limiting functions to the perturbation series in the coupling constant. By means of generalized Padé approximants and the method of Villani's limit theorem we can construct from the volume-dependent perturbation series alone the infinite volume limit of the free energy and the Schwinger functions. These limits have the properties of being continuous for all real, nonnegative g, and strongly asymptotic to the coupling constant expansion in an angular wedge of angle at least π at $g=0$, for the lattice spacing greater than zero.

In the sixth section we discuss mass renormalization. In this section we bring out the relation of field theory to critical phenomena. The mass in field theory is directly analogous to the inverse range of correlation in a statistical mechanical lattice system. Since we seek a theory where the lattice spacing goes to zero, and physical phenomena happen at finite separation, then $\xi/\Delta \to \infty$, where ξ is the correlation length. In other words, we must be approaching the critical point as the lattice spacing tends to zero. The mass renormalization condition adjusts the bare mass so that the system behaves as if, in statistical mechanical language, it were at a temperature just above the critical temperature, and tending to the critical temperature as the lattice spacing goes to zero. We show that the renormalized mass has certain monotonicity properties as a function of the bare mass, but have not proved, for the general models, continuity which is needed to complete the proof of mass renormalizability. As we remarked above, this result is proven for the special case $g:\phi^4:_d$ in the seventh section.

In the eignth section we discuss the computability particularly for the $g:\phi^4:_d$ theory. We find that what is needed is to compute the asymptotic behavior in the critical region. The method of high-temperature expansions, such as employed by Fisher and Burford[11] in their study of the spin-spin correlation functions in the critical region, would seem adequate, at least for $d \geq 4$, where the limit of the mass and ($d > 4$) coupling-constant renormalized, $g:\phi^4:_d$ theory is a continuous-spin Ising model.

Finally, in the appendixes we give some additional technical material. In Appendix A, we review, and extend, the relevant inequalities of statistical mechanics. In Appendix B, we give a brief review and slightly extend the known results on generalized Padé approximants as they are particularly relevant to the Padé-Borel summation method. In Appendix C we give the proof of the continuity of the renormalized mass in $g:\phi^4:_d$ theory. In Appendix D we discuss Nelson's reconstruction theorem and how it relates to our results for $g:\phi^4:_d$ theory.

2. LATTICE FIELD THEORY

Nelson[3] has shown under a mild assumption that if one has a Markov, random field defined over a d-dimensional, Euclidean space, one may construct from it a

field theory which satisfies all the Wightman axioms for a relativistic field theory in Minkowski space. The problems of noncommuting operators are replaced by the problems of correlated random variables. The expectation values with respect to the random field are related to the expectation values in Fock space by the Feynman-Kac formula

$$E(\exp(-gV)) = \langle \Omega_0, \exp(-tH)\Omega_0 \rangle \tag{2.1}$$

where H is $d-1$ dimensional, t is a pseudo-time, Ω_0 is the vacuum state, and V is the d-dimensional interaction. It is this formula which is fundamental to our approach. We seek to establish the existence of such a Markov, random field. Nelson[6] has shown this can be done for the free field. As motivation for our approach, we start on this study in such a way that many of the techniques of rigorous statistical mechanics can be used, and we keep in mind the type of Euclidean-space properties that Osterwalder and Schrader[7] found were necessary and sufficient to insure the existence of a relativistic field theory.

Let us now consider a box in d-dimensional, Euclidean space with each edge of length L, and divide it into N intervals in each direction. We have N^d cells. Let us assign a field variable $\phi_{\mathbf{r}}$ to each cell. The subscript $\mathbf{r}$ is a d-dimensional vector which takes on the possible values $0 \cdot L/N$, $1 \cdot L/N$, ..., $(N-1) \cdot L/N$ for each of its d components. It will be convenient to use the notation $\Delta = L/N$ for the lattice spacing between the cells' centers of our box. It is also convenient to introduce the momentum transformed variables

$$q_{\mathbf{k}} = (\Delta)^d \sum_{\mathbf{r}} \exp(2\pi i \mathbf{k} \cdot \mathbf{r}) \phi_{\mathbf{r}} \tag{2.2}$$

for k on a lattice of spacing L^{-1} in each direction, and centered at the origin. The reverse transformation is

$$\phi_{\mathbf{r}} = (L)^{-d} \sum_{\mathbf{k}} \exp(-2\pi i \mathbf{k} \cdot \mathbf{r}) q_{\mathbf{k}} \tag{2.3}$$

by the standard Fourier transform theory. Next we introduce an action function for a free field of mass m_0. It is

$$A_+ = \tfrac{1}{2}\Delta^d \sum_{\mathbf{r}} \left(\frac{\sum_{\{\delta\}} (\phi_{\mathbf{r}+\delta} - \phi_{\mathbf{r}})^2}{\Delta^2} + m_0^2 \phi_{\mathbf{r}}^2 \right). \tag{2.4}$$

The summation over the set $\{\delta\}$ is a sum over half of the nearest-neighbor cells of the cell r. The set $\{\delta\} = \{(\Delta, 0, \ldots, 0), (0, \Delta, \ldots, 0), \ldots, (0, 0, \ldots, \Delta)\}$, except for $\mathbf{r}$ on the boundaries of the box where the term of the summation corresponding to $\delta = \eta$, the vector normal to the hypersurface of the box is missing as $\phi_{\mathbf{r}+\eta}$ lies outside the box. Formula (2.4) is the discrete analog of the usual field theory formula

$$A_+ = \int L \, dt = \tfrac{1}{2} \int dt \, d^{d-1}x \, [(\nabla\varphi)^2 + m_0^2 \varphi^2]. \tag{2.5}$$

In terms of $q_{\mathbf{k}}$ of (2.2) we may rewrite (2.4) as

$$A_+ = \tfrac{1}{2} L^{-d} \sum_{\mathbf{k}} \{ \sum_{\{\delta\}} [4\Delta^{-2} \sin^2(\pi \mathbf{k} \cdot \delta)] + m_0^2 \} q_{\mathbf{k}} q_{-\mathbf{k}} - \tfrac{1}{2}\Delta^{d-2} \sum_{S} (\phi_{\mathbf{r}+(1-N)\eta} - \phi_{\mathbf{r}})^2, \tag{2.6}$$

where the last summation is over the "upper" surfaces. that is those with coordinates $(L-\Delta, r_2, \ldots, r_d)$, $(r_1, L-\Delta, \ldots, r_d)$, ..., $(r_1, r_2, \ldots, L-\Delta)$, where $0 \le r_i \le L - \Delta$ are arbitrary. Edge and corner points will appear more than once. This action function A_+ corresponds to free boundary conditions.

In addition to Eq. (2.4), we will find it desirable to define a corresponding A for periodic boundary conditions. The formula is identical to (2.4) except that now the difference terms on the boundary, instead of being omitted when they extend to a cell outside the box, are defined by the periodic condition

$$\phi_{\mathbf{r}+\eta} \equiv \phi_{\mathbf{r}+(1-N)\eta}$$

which *is* in the box. In terms of the $q_{\mathbf{k}}$ we have the action function A for periodic boundary condictions diagonalized as

$$A = \tfrac{1}{2} L^{-d} \sum_{\mathbf{k}} \{ \sum_{\{\delta\}} [4\Delta^{-2} \sin^2(\pi \mathbf{k} \cdot \delta)] + m_0^2 \} q_{\mathbf{k}} q_{-\mathbf{k}} \tag{2.7}$$

as the surface term in (2.6) vanishes by definition. To make contact with the usual field theory form we note that for Δ very small and $|\mathbf{k}| = 0(1)$ Eq. (2.7) goes to

$$A \approx \tfrac{1}{2} \int d^d \mathbf{k} (4\pi^2 |\mathbf{k}|^2 + m_0^2) q_{\mathbf{k}} q_{-\mathbf{k}}, \tag{2.8}$$

which is the familiar type.

While it would be quite curious if the final theory were to depend on which boundary conditions we impose, we need to consider several types in order to accomplish our demonstrations. One further type of boundary conditions we have found useful are the Dirichlet boundary conditions. These boundary conditions are obtained by adding a complete additional layer of cells over the entire surface of the box which interact as in (2.4); however, the field variables in these cells are held fixed at $\phi = 0$. The corresponding action function for Dirichlet boundary conditions may be put in a symmetrical form by expanding the squared difference term in (2.4) as extended, and rearranging the sum. We obtain for the Dirichlet boundary conditions an action function

$$A_- = \tfrac{1}{2}\Delta^d \sum_{\mathbf{r}} [-2\Delta^{-2} \sum_{\{\delta\}} (\phi_{\mathbf{r}} \phi_{\mathbf{r}+\delta}) + (m_0^2 + 2d\Delta^{-2}) \phi_{\mathbf{r}}^2], \tag{2.9}$$

where any $\phi_{\mathbf{r}+\delta}$ outside the box is set to zero. The various action functions are related by

$$\begin{aligned} A_- &= A + \Delta^{d-2} \sum_{S} \phi_{\mathbf{r}} \phi_{\mathbf{r}+(1-N)\eta}, \\ A &= A_+ + \tfrac{1}{2}\Delta^{d-2} \sum_{S} (\phi_{\mathbf{r}} - \phi_{\mathbf{r}+(1-N)\eta})^2, \\ A_- &= A_+ + \tfrac{1}{2}\Delta^{d-2} \sum_{S} (\phi_{\mathbf{r}}^2 + \phi_{\mathbf{r}+(1-N)\eta}^2). \end{aligned} \tag{2.10}$$

We are now in a position to introduce our discrete random fields in terms of the following *unnormalized* distribution functions

$$\begin{aligned} d\mu_+ &= \exp(-A_+) \prod_{\mathbf{r}} d\phi_{\mathbf{r}}, \\ d\mu &= \exp(-A) \prod_{\mathbf{r}} d\phi_{\mathbf{r}}, \\ d\mu_- &= \exp(-A_-) \prod_{\mathbf{r}} d\phi_{\mathbf{r}}, \end{aligned} \tag{2.11}$$

which correspond to free, periodic, and Dirichlet

boundary conditions, respectively. Corresponding to these measures we introduce the unnormalized expectation values

$$E_+(f)=\int f d\mu_+, \quad E(f)=\int f d\mu, \quad E_-(f)=\int f d\mu_-, \tag{2.12}$$

and weakly normalized expectation values

$$\hat{E}_+(f)=E_+(f)\,F^{L^d}, \quad \hat{E}(f)=E(f)\,F^{L^d}, \quad \hat{E}_-(f)=E_-(f)\,F^{L^d}, \tag{2.13}$$

where from (2.7) we define the normalizing factor per unit hypervolume as

$$F=\left(\frac{\Delta^d}{2\pi}\right)^{\Delta^{-d}/2} \exp\left\{\tfrac{1}{2}\int_{-1/2\Delta}^{1/2\Delta}\cdots\int d\mathbf{k}\,\ln[m_0^2+4\Delta^{-2}\times\sum_{(\delta)}\sin^2(\pi\mathbf{k}\cdot\delta)]\right\}. \tag{2.14}$$

These expectation values have the property that

$$\lim_{L\to\infty}[\hat{E}_+(1)]^{L^{-d}}=\lim_{L\to\infty}[\hat{E}(1)]^{L^{-d}}=1. \tag{2.15}$$

The formula for F was derived by computing the expected value of unity using (2.7), (2.11), and (2.12), and by showing that the surface contributions for the other boundary conditions do not affect the result (2.15) provided $\Delta>0$.

3. NORMAL ORDERED PRODUCT ON A DISCRETE LATTICE

One of the things which makes field theory technically difficult to treat is the occurrence of the normal-ordered products. The normal-ordered product differs from an ordinary product in that all the creation operators are placed to the left of all the annihilation operators. These products can, by means of their commutation relations be re-expressed in terms of ordinary products. Our analysis follows that of Nelson,[12] Glimm,[13] and Glimm and Jaffe.[14] We will consider periodic boundary conditions. The action defined by Eq. (2.7) is plainly by (2.10) a positive-definite ($m_0^2>0$) quadratic form in the variables $\phi_{\mathbf{r}}$ which can be taken as symmetric. For large values of L and small Δ, as we see in (2.7) the N^d eigenvectors are combinations of $q_{\mathbf{k}}$ and $q_{-\mathbf{k}}$ and the N^d eigenvalues are the coefficients of $q_{\mathbf{k}}q_{-\mathbf{k}}$.

In order to associate Fock space operators with this discrete momentum theory[15] we introduce a creation operator, $a^*(\mathbf{k})$, and an annihilation operator, $a(\mathbf{k})$, for each k of (2.2). These operators satisfy the usual Boson commutation relations

$$[a(\mathbf{k}),\ a^*(\mathbf{k}')]=L^d\delta_{\mathbf{k},\mathbf{k}'},$$
$$[a(\mathbf{k}),\ a(\mathbf{k}')]=0,$$
$$[a^*(\mathbf{k}),\ a^*(\mathbf{k}')]=0, \tag{3.1}$$

where $\delta_{\alpha,\beta}$ is the Kronecker delta which is 1 if $\alpha=\beta$ and zero otherwise. Now using the notation

$$\omega_{\mathbf{k}}=\tfrac{1}{2}[m_0^2+4\Delta^{-2}\sum_{(\delta)}\sin^2(\pi\mathbf{k}\cdot\delta)], \tag{3.2}$$

the usual second quantization formula

$$q_{\mathbf{k}}=(2\omega_{\mathbf{k}})^{-1/2}[a(\mathbf{k})+a^*(-\mathbf{k})], \tag{3.3}$$

and Eq. (2.3), we obtain the corresponding Fock space operator

$$\Phi(\mathbf{r})=\Phi^-(\mathbf{r})+\Phi^+(\mathbf{r}), \tag{3.4}$$

where

$$\Phi^-(\mathbf{r})=L^{-d}\sum_{\mathbf{k}}[\exp(2\pi i\mathbf{k}\cdot\mathbf{r})\,a(\mathbf{k})/(2\omega_{\mathbf{k}})^{1/2}], \tag{3.5}$$
$$\Phi^+(\mathbf{r})=L^{-d}\sum_{\mathbf{k}}[\exp(2\pi i\mathbf{k}\cdot\mathbf{r})\,a^*(-\mathbf{k})/(2\omega_{\mathbf{k}})^{1/2}].$$

Now, by definition, the normal ordered product is

$$:\Phi^p(\mathbf{r}):=\sum_{j=0}^{p}\binom{p}{j}[\Phi^+(\mathbf{r})]^j\,[\Phi^-(\mathbf{r})]^{p-j}, \tag{3.6}$$

where $\binom{p}{j}$ is the binomial coefficient. As an example of the reduction by means of the commutation relations, we consider $p=2$.

$$\Phi^2(\mathbf{r})=[\Phi^+(\mathbf{r})]^2+\Phi^+(\mathbf{r})\,\Phi^-(\mathbf{r})+\Phi^-(\mathbf{r})\,\Phi^+(\mathbf{r})+\Phi^-(\mathbf{r})^2,$$
$$=:\Phi^2(\mathbf{r}):+\Phi^-(\mathbf{r})\,\Phi^+(\mathbf{r})-\Phi^+(\mathbf{r})\,\Phi^-(\mathbf{r}). \tag{3.7}$$

Using (3.5) and (3.1) we may evaluate the commutator in (3.7) as

$$C_L=\frac{1}{2L^{2d}}\sum_{\mathbf{k}}[\exp(i(\mathbf{k}-\mathbf{k})\cdot\mathbf{r})\,L^d/\omega_{\mathbf{k}}], \tag{3.8}$$
$$=\frac{1}{L^d}\sum_{\mathbf{k}}[m_0^2+4\Delta^{-2}\sum_{(\delta)}\sin^2(\pi\mathbf{k}\cdot\delta)]^{-1},$$

independent of $\mathbf{r}$. If we reduce out all the commutators in this way, we obtain by induction

$$\Phi^p(\mathbf{r})=\sum_{j=0}^{[p/2]}\frac{p!}{(p-2j)!\,j!}\,2^{-j}C_L^j:\Phi^{p-2j}(\mathbf{r}): \tag{3.9}$$

or inverting this equation

$$:\Phi^p(\mathbf{r}):=\sum_{j=0}^{[p/2]}(-1)^j\frac{p!}{(p-2j)!\,j!}\,2^{-j}\,C_L^j\,\Phi^{p-2j}(\mathbf{r}). \tag{3.10}$$

Thus the normal ordered operators for periodic boundary conditions are polynomials in the ordinary Φ's with coefficients depending the commutator C_L and independent of $\mathbf{r}$.

The first limit we will be concerned with in our construction of a Markov, random field from which to construct a self-interacting Boson field theory will be the "thermodynamic" limit, $L\to\infty$, with the lattice spacing Δ fixed. In this limit, (3.8) goes over directly to

$$C=\lim_{L\to\infty}C_L=\int_{-1/2\Delta}^{1/2\Delta}\cdots\int\frac{d\mathbf{k}}{m_0^2+4\Delta^{-2}\sum_{(\delta)}\sin^2(\pi\mathbf{k}\cdot\delta)} \tag{3.11}$$

as in this limit the spacing between the discrete k's goes to zero and they fill the hypercube indicated by (3.11). Since for a fixed value of Δ and $|\Phi(\mathbf{r})|$, we can by choosing L large enough make $:\Phi^p(\mathbf{r}):$ arbitrarily close to that obtained by substituting C for C_L in (3.10), and since it will be convenient later on to have an interaction which is independent of the box size, we define the random field equivalent normal ordered product as

$$:\Phi_{\mathbf{r}}^{p}:=\sum_{j=0}^{[p/2]}(-1)^{j}\frac{p!}{(p-2j)!j!}\,2^{-j}C^{j}\,\phi_{\mathbf{r}}^{p-2j}. \tag{3.12}$$

For future reference, we note that for Δ very small,

$$\begin{aligned} &C\propto\Delta^{2-d}, && d>2,\\ &C\propto-\ln\Delta, && d=2,\\ &C \text{ finite}, && d<2, \end{aligned} \tag{3.13}$$

gives the asymptotic behavior of the constant C as the ultraviolet cutoff ($\Delta\to 0$) is removed. This potential infinity must be borne in mind during subsequent analysis.

4. THE INFINITE VOLUME LIMIT

We have introduced in the previous two sections the free field action and the quantum field theory normal ordered product in the context of discrete random fields. Now we will introduce the fundamental objects of the Euclidean quantum field theory. First, for our interaction, we choose an even polynomial in the normal ordered products

$$V=\Delta^{d}\sum_{\mathbf{r}}\sum_{j=0}^{p}a_{j}:\phi_{\mathbf{r}}^{2j}:, \tag{4.1}$$

where the a_j are independent of $\mathbf{r}$, $a_p=1$, and the $:\phi_{\mathbf{r}}^{2j}:$ are given in terms of the $\phi_{\mathbf{r}}$ by Eq. (3.12). This interaction is the linear sum of terms which act only within a single cell.

We need the following objects. First the partition function (periodic boundary conditions)

$$Z=E(\exp(-gV)), \tag{4.2}$$

where the unnormalized expectation value defined by (2.13) is used. A similar formula defines the corresponding partition function for free and Dirichlet boundary conditions. The connection with the Fock space operators is given by

$$\langle\Omega_0,\ \exp(-\Sigma dtH)\Omega_0\rangle=[E(\exp(-gV))]/[E(1)]. \tag{4.3}$$

We further need the Schwinger functions or correlation functions in the language of statistical mechanics. The Schwinger functions are the analytic continuation of the Wightman functions to the Schwinger, imaginary time, i.e., Euclidean space, points. They are defined as

$$S(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_n)=E(\phi_{\mathbf{r}_1}\phi_{\mathbf{r}_2}\cdots\phi_{\mathbf{r}_n}\exp(-gV))/Z, \tag{4.4}$$

and similarly for the other boundary conditions.

The main tool that we will use in discussion of the limit of (4.2) and (4.4) as $L\to\infty$ will be various inequalities. In Appendix A we discuss the requirements for validity and list for reference some of those which are useful here and occur in the statistical mechanics literature. Some of them have been extended to cover our cases. First we will relate the quantities defined by the various boundary conditions. From Eq. (2.10), since $\exp(-\phi^2)\leq 1$ for all real ϕ, it follows at once that

$$Z_{+}\geq Z,\quad Z_{+}\geq Z_{-}. \tag{4.5}$$

Let us now consider the relation between Z and Z_-. We can go from Z_- to Z continuously by increasing from zero to unity, one at a time, the coefficients $\lambda_{\mathbf{r}}$ of each term in the $\sum_S$ given in (2.10). If we compute,

$$\frac{\partial}{\partial\lambda_{\mathbf{r}}}Z_{-}\{\lambda\}=E_{-}\{\lambda\}(\Delta^{d-2}\phi_{\mathbf{r}}\phi_{\mathbf{r}+\mathbf{1}(1-N)m}\exp(-gV))\geq 0 \tag{4.6}$$

by the Griffiths inequality (A3). Thus by the integration of (4.6) we can conclude

$$Z\geq Z_{-},\quad Z_{+}\geq Z\geq Z_{-} \tag{4.7}$$

For the Schwinger functions, if we do the same thing, then

$$\begin{aligned}&\frac{\partial}{\partial\lambda_{\mathbf{r}}}S_{-}\{\lambda\}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_n)\\ &=[E_{-}\{\lambda\}(\phi_{\mathbf{r}}\phi_{\mathbf{r}+\mathbf{1}(1-N)m}\,\phi_{\mathbf{r}_1}\cdots\phi_{\mathbf{r}_n}\exp(-gV))\\ &\times E_{-}\{\lambda\}(\exp(-gV))-E_{-}\{\lambda\}(\phi_{\mathbf{r}}\phi_{\mathbf{r}+\mathbf{1}(1-N)m}\exp(-gV))\\ &\times E_{-}\{\lambda\}(\phi_{\mathbf{r}_1}\cdots\phi_{\mathbf{r}_n}\exp(-gV))]\,[E_{-}\{\lambda\}(\exp(-gV))]^{-2}\geq 0\end{aligned} \tag{4.8}$$

by the Griffiths-Kelly-Sherman inequalities (A5). Thus integrating (4.8), by a similar argument on the relation between A_- and A_+, we have

$$\begin{aligned}&S(\mathbf{r}_1,\ldots,\mathbf{r}_n)\geq S_{-}(\mathbf{r}_1,\ldots,\mathbf{r}_n)\geq 0,\\ &S_{+}(\mathbf{r}_1,\ldots,\mathbf{r}_n)\geq S_{-}(\mathbf{r}_1,\ldots,\mathbf{r}_n)\geq 0,\end{aligned} \tag{4.9}$$

where the positivity comes from Griffiths inequality (A3).

Let us now derive an upper bound for S. First we introduce the notation

$$\phi_{\mathbf{r}}=x_{\mathbf{r}}\sigma_{\mathbf{r}},\quad 0\leq x_{\mathbf{r}}\leq\infty,\ \sigma_{\mathbf{r}}=\pm 1. \tag{4.10}$$

Now, by the inequality between the geometric and arithmetic means,[16] we have

$$\frac{1}{n}\sum_{i=1}^{n}x_i^{n}\geq\prod_{i=1}^{n}x_i\geq\prod_{i=1}^{n}\phi_i. \tag{4.11}$$

Since, for periodic boundary conditions every cell is equivalent, for the purpose of taking the expected value we need only consider, say, x_1^n instead of the left-hand side of (4.11). Thus we may write

$$S(\mathbf{r}_1,\ldots,\mathbf{r}_n)\leq\frac{\int_0^\infty\cdots\int x_1^{n}\sum_{\{\sigma_{\mathbf{r}}=\pm 1\}}\exp(-A-gV)\prod_{\mathbf{r}}dx_{\mathbf{r}}}{\int_0^\infty\cdots\int\sum_{\{\sigma_{\mathbf{r}}=\pm 1\}}\exp(-A-gV)\prod_{\mathbf{r}}dx_{\mathbf{r}}}. \tag{4.12}$$

Next let us consider the effect of introducing a factor of

$$\exp[\tfrac{1}{2}\lambda\Delta^{d-2}\sum_{\{\delta\}}(2x_1^2-x_1x_{1+\delta}\sigma_1\sigma_{1+\delta}-x_1x_{1-\delta}\sigma_1\sigma_{1-\delta})] \tag{4.13}$$

in the numerator of (4.12), where the reader is referred to Section 2 for the notation. Taking the derivative with respect to λ gives

$$\frac{\partial S\{\lambda\}}{\partial\lambda}=\tfrac{1}{2}\Delta^{d-2}\sum_{\{\delta\}}E(2x_1^{n+2}-x_1^{n+1}x_{1+\delta}\sigma_1\sigma_{1+\delta}-x_1^{n+1}x_{1-\delta}\sigma_1\sigma_{1-\delta})$$

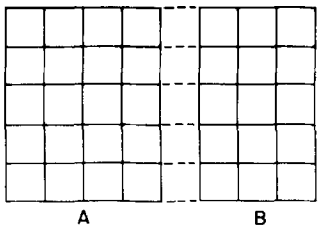


FIG. 1. The adjoining of an additional chunk B to the original lattice A. The additional coupling terms not in A or B considered separately are shown by the broken lines.

$$\geq \tfrac{1}{2}\Delta^{d-2} \sum_{\{\delta\}} E(2x_1^{n+2} - x_1^{n+1} x_{1+\delta} - x_1^{n+1} x_{1-\delta}). \tag{4.14}$$

Now a special case of the geometric mean-arithmetic mean inequality gives

$$x^{n+1}y \leq \frac{1}{n+2}\,[(n+1)x^{n+2} + y^{n+2}], \tag{4.15}$$

which applied to (4.14) gives

$$\frac{\partial S\{\lambda\}}{\partial \lambda} \geq \frac{\Delta^{d-2}}{2(n+2)} \sum_{\{\delta\}} [2E(x_1^{n+2}) - E(x_{1+\delta}^{n+2}) - E(x_{1-\delta}^{n+2})] \tag{4.16}$$
$$= 0,$$

for $\lambda = 0$, as all cells are equivalent for periodic boundary conditions. If we now look at the structure of $S\{\lambda\}$ as a function of λ, it must be

$$S\{\lambda\} = \int_{-\infty}^{+\infty} \exp(-\lambda\xi)\, d\rho(\xi), \tag{4.17}$$

where $d\rho \geq 0$. In this case the $B_{0,-1}(\lambda)$ generalized Padé approximant based on $\exp(-xs)$ (see Appendix B) forms a lower bound to $S\{\lambda\}$. As we have shown near $\lambda = 0$ that

$$S\{\lambda\} = A + B\lambda, \quad A,\ B > 0, \tag{4.18}$$

we conclude that

$$S\{\lambda\} \geq A \exp(B\lambda/A), \tag{4.19}$$

and so

$$S\{\lambda = 1\} \geq S\{\lambda = 0\}. \tag{4.20}$$

But, by the definition of A, Eq. (2.7), the integration over x_1 in the modified numerator of (4.12) is decoupled from the rest of the integrations. Thus, if we multiply numerator and denominator by a convenient factor, inequalities (4.12) and (4.20) combine to imply

$$S(\mathbf{r}_1, \ldots, \mathbf{r}_n)$$
$$\leq \frac{\int_0^\infty x_1^n \exp[-\frac{1}{2}m_0^2\Delta^d x_1^2 - g\sum_{j=0}^{p} a_j : x_1^{2j} :]\, dx_1}{\int_0^\infty \exp[-(\frac{1}{2}m_0^2\Delta^d + d\Delta^{d-2})x_1^2 - g\sum_{j=0}^{p} a_j : x_1^{2j} :]\, dx_1}$$
$$\times \frac{\int_0^\infty \cdots \int \sum_{\{\sigma_r = \pm 1\}} \exp(-A^* - gV) \prod_{\mathbf{r}} dx_{\mathbf{r}}}{\int_0^\infty \cdots \int \sum_{\{\sigma_r = \pm 1\}} \exp(-A - gV) \prod_{\mathbf{r}} dx_{\mathbf{r}}}. \tag{4.21}$$

The action function A^* differs from A by not having the terms

$$\tfrac{1}{2}\Delta^{d-2} \sum_{\{\delta\}} (x_1 x_{1+\delta}\sigma_1\sigma_{1+\delta} + x_1 x_{1-\delta}\sigma_1\sigma_{1-\delta}). \tag{4.22}$$

If we add them, by an argument identical to that which allowed us to conclude $Z \geq Z_-$, (4.7), we conclude that this addition will increase the numerator of the second factor of (4.21). As the addition of (4.22) makes the second factor increase to unity we conclude

$$S(\mathbf{r}_1, \ldots, \mathbf{r}_n)$$
$$\leq \frac{\int_0^\infty x^n \exp[-\frac{1}{2}m_0^2\Delta^d x^2 - g\sum_{j=0}^{p} a_j : x^{2j} :]\, dx}{\int_0^\infty \exp[(\frac{1}{2}m_0^2\Delta^d + d\Delta^{d-2})x^2 - g\sum_{j=0}^{p} a_j : x^{2j} :]\, dx} \tag{4.23}$$

independent of L.

The same upper bound will also hold for S_+, by the same proof, provided that when we wrote (4.12) we picked x_1 such that both $E_+(x_1^n)$ and $E_+(x_1^{n+2})$ are simultaneously maximum. I conjecture that such a choice is substantially possible, but have not pursued a proof.

We next turn our attention to the behavior of the various objects of the theory with the size of the system. First, let us consider the case of Dirichlet boundary conditions. Referring to Eq. (2.9), we see that if we adjoin an additional chunk (Fig. 1) to our lattice by increasing, one at a time, the coefficients $\lambda_{\mathbf{r}}$ of those $\phi_{\mathbf{r}}\phi_{\mathbf{r}+\delta}$ which link the original part of the lattice with the new chunk, then as

$$\frac{\partial Z_-\{\lambda\}_{A+B}}{\partial \lambda_{\mathbf{r}}} = E_-\{\lambda\}(\phi_{\mathbf{r}}\phi_{\mathbf{r}+\delta}) \geq 0 \tag{4.24}$$

by Griffith's inequality (A3), we conclude

$$Z_{-,A+B} \geq Z_{-,A}\, Z_{-,B}, \tag{4.25}$$

where A and B are the two pieces of the lattice.

Next consider doing the same thing to the Schwinger functions. First, the addition of an *uncoupled part* of the lattice does not change the Schwinger function as

$$\frac{E_{-,A+B}\{\phi_{\mathbf{r}_1} \cdots \phi_{\mathbf{r}_n} \exp[-g(V_A + V_B)]\}}{E_{-,A+B}\{\exp[-g(V_A + V_B)]\}}$$
$$= \frac{E_{-,A}\{\phi_{\mathbf{r}_1} \cdots \phi_{\mathbf{r}_n} \exp(-gV_A)\}E_{-,B}\{\exp(-gV_B)\}}{E_{-,A}\{\exp(-gV_A)\}E_{-,B}\{\exp(-gV_B)\}} \tag{4.26}$$
$$= S_{-,A}(\mathbf{r}_1, \ldots, \mathbf{r}_n)$$

Now, as we increase the coefficient of one coupling term at a time, we have

$$\frac{\partial S_{-,A+B}\{\lambda\}(\mathbf{r}_1, \ldots, \mathbf{r}_n)}{\partial \lambda_{\mathbf{r}}}$$
$$= E_{-,A+B}[\lambda]\,\{\phi_{\mathbf{r}_1} \cdots \phi_{\mathbf{r}_n}\phi_{\mathbf{r}}\phi_{\mathbf{r}+\delta} \exp[-g(V_A + V_B)]\}/$$
$$E_{-,A+B}\{\lambda\}\{\exp[-g(V_A + V_B)]\}$$
$$- E_{-,A+B}\{\lambda\}\{\phi_{\mathbf{r}_1} \cdots \phi_{\mathbf{r}_n} \exp[-g(V_A + V_B)]\}$$
$$\times E_{-,A+B}\{\lambda\}\{\phi_{\mathbf{r}}\phi_{\mathbf{r}+\delta} \exp[-g(V_A + V_B)]\}$$
$$\times [E_{-,A+B}\{\lambda\}\{\exp[-g(V_A + V_B)]\}]^{-2} \geq 0 \tag{4.27}$$

by the Griffiths–Kelly–Sherman inequality (A5). Thus we conclude also that $S_-(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ increases monotonically with the system size.

Now let us consider Z_+. Again we adjoin an additional

piece of lattice to what we had. The partition function for $A+B$ contains, in the integrand, an additional factor of

$$\exp[-\tfrac{1}{2}\Delta^{d-2}\sum_{\partial}(\phi_{\mathbf{r}}-\phi_{\mathbf{r}+\delta})^2]\leq 1, \tag{4.28}$$

where the sum is over the boundary. Thus

$$Z_{+,A+B}\leq Z_{+,A}\,Z_{+,B} \tag{4.29}$$

If $N(X)$ is Δ^d times the number of cells in lattice section X, then we may rewrite Eq. (4.29) immediately as

$$\frac{1}{N(A+B)}\ln Z_{+,A+B}\leq\frac{1}{N(A)}\ln Z_{+,A}+\frac{N(B)}{N(A+B)}\left(\frac{1}{N(B)}\ln Z_{+,B}-\frac{1}{N(A)}\ln Z_{+,A}\right), \tag{4.30}$$

which implies

$$\frac{1}{N(2A)}\ln Z_{+,2A}\leq\frac{1}{N(A)}\ln Z_{+,A}, \tag{4.31}$$

provided we choose $A=B$. Thus, at least for doubling of the size of the set we have shown that $[\ln Z_{+,A}/N(A)]$ is monotonically decreasing with $N(A)$. For any dimension d we can return to our original hypercubical shape by repeated applications of the above argument. The new hypercube will be twice as big on each edge (2^d times the volume). Applying the above argument to (4.25) we find

$$\frac{1}{N(2A)}\ln Z_{-,2A}\geq\frac{1}{N(A)}\ln Z_{-,A}, \tag{4.32}$$

so that we have $[\ln Z_{-,A}/N(A)]$ monotonically increasing with $N(A)$. Since a bounded monotonic sequence must tend to a limit we can now conclude

$$\frac{1}{L^d}\ln Z_{+}\geq\lim_{L\to\infty}\frac{1}{L^d}\ln Z_{+}\geq\lim_{L\to\infty}\frac{1}{L^d}\ln Z_{-}\geq\frac{1}{L^d}\ln Z_{-}, \tag{4.33}$$

where the passage to the limit is by way of size doublings, with g and Δ fixed. Clearly the same results hold with $\hat{Z}_{\pm}$ replacing $Z_{\pm}$, by the nature of inequalities (4.25) and (4.29).

We may remove the restriction on the sequence of L's over which the limit is taken by the following argument. Let us fix a box D of edge K, and consider another box C of edge L, which is going to infinity. If we write

$$L=mK+R, \tag{4.34}$$

where m is a positive integer and $0\leq R<K$, then we may dissect C into m^d full boxes of edge K plus no more than $d(m+1)^{d-1}$ partial boxes D_τ of edge K. By repeated applications of (4.29) we may write, after taking the logarithm and dividing by $N(C)$,

$$\frac{1}{N(C)}\ln Z_{+,C}\leq\frac{m^d}{N(C)}\ln Z_{+,D}+\frac{1}{N(C)}\sum_{\tau}\ln Z_{+,D_\tau}. \tag{4.35}$$

If we take the limit $L\to\infty$ for fixed K, then $\lim[N(C)/m^d]=N(D)$. Furthermore, for fixed K there are only a finite number of distinct types of partial boxes D_τ, so that the sum of (4.35) is less than the maximum of $\ln Z_{+,D_\tau}$ over this finite set times the number of terms. Since $N(C)$ is proportional to m^d, the summation term goes to zero as $L\to\infty$. Thus the largest possible limit is

$$\limsup_{L\to\infty}[1/N(C)]\ln Z_{+,C}\leq[1/N(D)]\ln Z_{+,D}. \tag{4.36}$$

But since D is arbitrary we may select a subsequence of all possible box sizes so that we get the smallest possible limit. Thus, by (4.36)

$$\liminf_{K\to\infty}[1/N(D)]\ln Z_{+,D}\geq\limsup_{L\to\infty}[1/N(C)]\ln Z_{+,C}. \tag{4.37}$$

But Eq. (4.37) implies that all limits coincide, so that the limit specified in (4.33) exists when taken over all L, as one would have naively supposed. The above arguments also work with the sense of the inequalities reversed for Z_- and justify the existence of the other limit in (4.33) when it too is taken over all L.

Equation (4.33) gives meaning to the free energy per unit volume

$$f_{\pm}=-\lim_{L\to\infty}(1/L^d)\ln Z_{\pm}, \tag{4.38}$$

with fixed ultraviolet cutoff $\Delta>0$, for every nonnegative, real coupling constant g, and every real unrenormalized mass m_0^2. By inequality (4.7) we can similarly give meaning to the case of periodic boundary conditions. We note from inequality (4.8) and bound (4.23) that if we integrate the differential equation that carries us from Z_- to Z_+ that

$$\ln Z-\ln Z_{-}\propto KL^{d-1}, \tag{4.39}$$

where the constant is proportional to bound (4.23) on the derivative. Consequently, we conclude that

$$f=-\lim_{L\to\infty}\frac{1}{L^d}\ln Z=f_{-} \tag{4.40}$$

as the surface-to-volume ratio goes to zero for a large box. If, as conjectured above, (4.23) bounds S_+ as well, then $f=f_+=f_-$.

It follows from the monotonicity of S_- and the upper bounds (4.9) and (4.23) that we have the existence of the limit

$$\mathcal{S}_{-}(\mathbf{r}_1,\ldots,\mathbf{r}_n)=\lim_{L\to\infty}S_{-}(\mathbf{r}_1,\ldots,\mathbf{r}_n), \tag{4.41}$$

where the limit is taken over all L.

In the same way, we can define $\mathcal{S}$ for periodic boundary conditions. (We have not proved that the whole sequence converges, but if it does not, we can, by standard arguments, find a subsequence which does.) Presumably S_+ is bounded and so we can also define $\mathcal{S}_+$ for free boundary conditions.

Consequently, we have shown in this section that the infinite volume, or thermodynamic limit, exists, and therefore the fundamental quantities, free energy per unit volume, and the Schwinger functions, are well defined for Dirichlet boundary conditions. The free energy

per unit volume has been shown to exist for free boundary conditions and for periodic boundary conditions as well.

5. CONSTRUCTION FROM THE PERTURBATION SERIES

In this section we undertake the construction of the free energy per unit volume and the Schwinger functions in the infinite volume limit in terms of their perturbation series expansions in g, the coupling constant. We will show that there exist functions which are continuous for all real positive g and whose formal power series expansions are asymptotic at $g=0$ to those of the infinite volume limits established in the previous section. It is most convenient to start with our system once again finite in size, and that is what we now do.

In order to connect the perturbation series with the limiting functions we will use generalized Padé approximants[18,19] with an exponential kernel. For the convenience of the reader, we have given their definition and have detailed some of their properties in Appendix B. In order to apply them we note the following property of our interaction. Since the polynomial implied by (4.1) and (3.12)

$$P(\phi)=\sum_{j=0}^{p} a_j :\phi^{2j}:=\sum_{j=0}^{p} b_j \phi^{2j} \tag{5.1}$$

has finite coefficients, and $b_p=1$, as long as $\Delta>0$, there must exist a constant K such that

$$P(\phi)+K\geq 0 \tag{5.2}$$

for all ϕ. Thus, if $M=N^dK$, we must have for our interaction V

$$V+M\geq 0 \tag{5.3}$$

for all $\phi_{\mathbf{r}}$, i.e., V is semibounded from below. Thus there exist distribution functions, simply related to those of Section 2, such that

$$E_-(\exp[-g(V+M)])=\int_0^\infty \exp(-g\lambda)\,d\rho_-(\lambda), \tag{5.4}$$

for Dirichlet boundary conditions, where $d\rho_-\geq 0$. A similar structure also holds for free and periodic boundary conditions. Now for functions of this form we may construct upper and lower bounding, generalized Padé approximants

$$B_{n,0}(g)\geq E_-(\exp[-g(V+M)])\geq B_{n,-1}(g), \tag{5.5}$$

and so too for free and periodic boundary conditions. We observe explicitly that the lower bounds are independent of M as

$$\begin{aligned} &E_-(\exp(-gV))\\ &\quad=\exp(gM)\,\hat{E}_-(\exp[-g(V+M)])\\ &\quad\geq \exp(gM)\sum_{j=1}^{n}\alpha_j\exp[-g(\beta_j+M)]\\ &\quad=\sum_{j=1}^{n}\alpha_j\exp(-g\beta_j). \end{aligned} \tag{5.6}$$

We now use the ideas of Villani's limit theorem[20,21] as applied by us[22] to statistical mechanics. Now, since we may consider $B_{n,-1}(g)$ as a function of the box size L, and (5.5) holds for all L, and since, by (4.32) we have monotonic increase as L doubles, we must have

$$\begin{aligned} &\max_L\{[\exp(gM)\,B_{n,-1}(g)]^{1/L^d}\}\\ &\quad\leq[E_-(\exp(-gV))]^{1/L^{d(n)}}\leq\exp(-f_-), \end{aligned} \tag{5.7}$$

where $L(n)$ is the value of L, selected from all size doublings for an initial L_0 for which the maximum approximant is obtained. Similarly, by (4.31) we may also conclude

$$\begin{aligned} &\min_L\{[\exp(gM)\,B_{n,0}(g)]^{1/L^d}\}\\ &\quad\geq[E_+(\exp(-gV))]^{1/L^{d(n)}}\geq\exp(-f_+), \end{aligned} \tag{5.8}$$

where here the approximants are defined by the $E_+(\exp(-gV))$.

For any fixed, finite, L, g, Δ, we have

$$\lim_{n\to\infty}\exp(gM)\,B_{n,-1}(g)=E_-(\exp(-gV)) \tag{5.9}$$

by the convergence properties of the generalized Padé approximants. We likewise can impose (5.9) uniformly over any range of L, $\Delta\leq L\leq L_0$. We can, by choosing L_0 large enough, make

$$[E_-(\exp(-gV))]^{1/L_0^d}\geq\exp(-f_-)-\tfrac{1}{2}\epsilon, \tag{5.10}$$

for any $\epsilon>0$, which forces $L(n)$ to infinity as n goes to infinity. Thus we conclude that

$$\begin{aligned} \exp(-f_-)&\geq\lim_{n\to\infty}[\max_L\{[\exp(gM)\,B_{n,-1}(g)]^{1/L^d}\}]\\ &\geq\exp(-f_-)-\epsilon. \end{aligned} \tag{5.11}$$

But as $\epsilon>0$ is arbitrary, the limit exists, and is equal to the function $\exp(-f_-)$ defined in the previous section. Thus we have shown [also using (4.40)]

$$f=f_-=-\lim_{n\to\infty}\left[\max_L\left(\frac{1}{L^d}\ln[\exp(gM)\,B_{n,-1}(g)]\right)\right], \tag{5.12}$$

where L ranges over doublings and the approximants are determined by the L-dependent, $E_-(\exp[-g(V+M)])$ series coefficients. Similarly we have by analogous arguments for the free boundary conditions

$$\begin{aligned} f_+&=-\lim_{n\to\infty}\left[\min_L\left(\frac{1}{L^d}\ln[\exp(gM)\,B_{n,0}(g)]\right)\right]\\ &\leq f=f_-, \end{aligned} \tag{5.13}$$

where L ranges over doublings and the approximants are determined by the L-dependent, $E_+(\exp[-g(V+M)])$ series. By the construction of the B's

$$\begin{aligned} f_{+,L}-(1/L^d)\ln[\exp(gM)\,B_{n,0}(g)]&=O(g^{2n+1}),\\ f_{-,L}-(1/L^d)\ln[\exp(gM)\,B_{n,-1}(g)]&=O(g^{2n}) \end{aligned} \tag{5.14}$$

as $B_{n,0}(0)\neq 0$. Since explicit expressions can be given for the coefficient of every power of g in the formal power series expansion of $f_{+,L}$ in terms of expectation values over the distribution functions for the free field, it remains to indicate how all the coefficients tend to finite limits as $L\to\infty$.

If we formally expand

$$Z_{-,L}(g)=\sum_{j=0}^{\infty}\eta_{-}(L,j)g^{j}, \tag{5.15}$$

then we see at once in terms of orders of magnitude for L large that

$$\xi_{-}(L,j)\equiv\eta_{-}(L,j)/\eta_{-}(L,0)\sim L^{jd} \tag{5.16}$$

as

$$\eta_{-}(L,j)=\int\left[\Delta^{d}\sum_{\mathbf{r}}\sum_{j=0}^{p}a_{j}:\phi_{\mathbf{r}}^{2j}:\right]^{j}d\mu_{-} \tag{5.17}$$

by (2.12), (4.1), (4.2), and (4.23). The calculation of the free energy

$$-L^{d}f_{-}=\ln Z_{-,L}=\ln Z_{-,L}(0)+\sum_{j=1}^{\infty}\lambda_{-}(L,j)g^{j} \tag{5.18}$$

gives formal power series coefficients which are related to those of (5.15) as are cummulants to moments. That is,

$$\begin{aligned}
&\lambda_{-}(L,1)=\xi_{-}(L,1),\\
&\lambda_{-}(L,2)=\xi_{-}(L,2)-[\xi_{-}(L,1)]^{2},\\
&\lambda_{-}(L,3)=\xi_{-}(L,3)-3\xi_{-}(L,1)\,\xi_{-}(L,2)+2[\xi_{-}(L,1)]^{3},\\
&\lambda_{-}(L,4)=\xi_{-}(L,4)-4\xi_{-}(L,3)\,\xi_{-}(L,1)\\
&\qquad\qquad-3[\xi_{-}(L,2)]^{2}+12\xi_{-}(L,2)[\xi_{-}(L,1)]^{2}\\
&\qquad\qquad-6[\xi_{-}(L,1)]^{4},
\end{aligned} \tag{5.19}$$

$$\cdots$$

Because the free field ($m_0^2>0$) satisfies a cluster property, it can be shown that the $\lambda(L,j)\propto L^{d}$ only and that the higher powers of L^{d} cancel. This result can be demonstrated in the following manner. First, let, from (5.1),

$$X_{\mathbf{r}}=P(\phi_{\mathbf{r}})+K \tag{5.20}$$

for notational convenience. Then

$$\xi_{-}(L,j)=\left\langle\sum_{\mathbf{r}_1}\sum_{\mathbf{r}_2}\cdots\sum_{\mathbf{r}_j}\prod_{k=1}^{j}X_{\mathbf{r}_k}\right\rangle_{-} \tag{5.21}$$

where $\langle\ \rangle_{-}$ is the normalized expectation value with respect to the free field distribution function with Dirichlet boundary conditions. We may then write, by (5.19),

$$\lambda_{-}(L,j)=\sum_{\mathbf{r}_1}\cdots\sum_{\mathbf{r}_j}U_{-}(L,X_{\mathbf{r}_1},\ldots,X_{\mathbf{r}_j}) \tag{5.22}$$

where the U's are the Ursell functions, i.e.,

$$\begin{aligned}
&U_{-}(L,X_{\mathbf{r}})=\langle X_{\mathbf{r}}\rangle_{-},\\
&U_{-}(L,X_{\mathbf{r}},X_{s})=\langle X_{\mathbf{r}}X_{s}\rangle_{-}-\langle X_{\mathbf{r}}\rangle_{-}\langle X_{s}\rangle_{-},\\
&U_{-}(L,X_{\mathbf{r}},X_{s},X_{t})=\langle X_{\mathbf{r}}X_{s}X_{t}\rangle_{-}-\langle X_{\mathbf{r}}\rangle_{-}\langle X_{s}X_{t}\rangle_{-}\\
&\qquad\qquad-\langle X_{s}\rangle_{-}\langle X_{\mathbf{r}}X_{t}\rangle_{-}-\langle X_{t}\rangle_{-}\langle X_{\mathbf{r}}X_{s}\rangle_{-}\\
&\qquad\qquad+2\langle X_{\mathbf{r}}\rangle_{-}\langle X_{s}\rangle_{-}\langle X_{t}\rangle_{-}.
\end{aligned} \tag{5.23}$$

The Ursell functions have the property that if the $X_{\mathbf{r}_k}$ fall into two groups $\{Y_i\}$, $\{Z_j\}$ such that for every subset of the $\{X_i\}$ and every subset of the $\{Z_i\}$ that

$$\langle Y_i\cdots Y_j Z_k\cdots Z_l\rangle=\langle Y_i\cdots Y_j\rangle\langle Z_k\cdots Z_l\rangle, \tag{5.24}$$

then $U=0$. If we define

$$\sigma_{A}=\prod_{k\in A}X_{\mathbf{r}_k} \tag{5.25}$$

where A is a subset of $k=1,\ldots,j$, then it follows[23,24] from the definition of the Ursell functions that for any partition of $1,\ldots,j$ into two disjoint sets

$$U_{-}(L,X_{\mathbf{r}_i},\ldots,X_{\mathbf{r}_j})=\sum_{\text{partitions}}(\pm1)\,(\Pi_{\alpha}\langle\sigma_{C_\alpha}\rangle_{-})\,[\langle\sigma_{A'\cup B'}\rangle_{-}-\langle\sigma_{A'}\rangle_{-}\langle\sigma_{B'}\rangle_{-}] \tag{5.26}$$

where $A'\subset A$, $B'\subset B$, $\{1,\ldots,j\}=A'\cup B'\cup C_1\cup\cdots\cup C_n$. By direct calculation we can bound the Π_α factor. For the free field the last factor decays exponentially with the distance between A' and B' by the properties of the lattice Green's function,[25] and the inequality (A23). This decay is roughly proportional to $\exp[-m_0 d(A',B')]$, where $d(A',B')$ is the distance between A' and B'. Thus we only get a significant contribution to the sum when all the $\mathbf{r}_i$ are near each other and so the $\lambda_{-}(L,j)$ are of the order L^{d} when $\Delta>0$ and $m_0^2>0$. Thus the coefficients (λ/L^{d}) of f_{-} are finite, and the explicit expressions tend as $L\to\infty$ to the usual momentum representation integrals as expected; the lattice Green's functions [as in Eq. (3.11)] play the role of propagators. It follows in the same way that the $L\to\infty$ limit for the coefficients is the same for f_{+} and f as it is for f_{-}.

Since, by (4.23), and direct calculation, $|f'_{-}|$ and $|f'|$ are bounded uniformly for all L and g, if $\Delta>0$, $m_0^2>0$, we must have that f_{-} and f are continuous functions of g for all real positive g. Since $\exp(gM)B_{n,-1}(g)$ is an entire function and bounded in absolute value for complex g by its value at $\mathrm{Re}(g)$, the uniform bound on the derivative (see Appendix E), together with $B_{n,-1}(0)\neq0$, implies that there are no zeros in an angular wedge, W, $|\arg(z)|\leq\pi/2$, $|z|\leq\omega$, for some $\omega>0$, uniformly in L. Thus, by standard theorems on the convergence of a sequence of bounded analytic functions,[26] and the properties of the generalized Padé approximants, it follows that $f_{-}=f$ is analytic in the interior of W, and continuous on its boundary.

Therefore, thinking of Dirichlet boundary conditions where our results are most complete, we conclude that our procedure (5.12) constructs a function f_{-} which is analytic in the interior of an angular wedge W, continuous on the whole positive real g axis, and asymptotic to the usual perturbation theory expansion at $g=0$ in W.

Next we consider the Schwinger functions. To this end we add to the action A_{-} given by Eq. (2.9)

$$H_n=\sum_{\mathbf{r}}h_{\mathbf{r}}\phi_{\mathbf{r}}, \tag{5.27}$$

where there are only n nonzero $h_{\mathbf{r}}$ in the sum. We require that either all $h_{\mathbf{r}}\geq0$ or all $h_{\mathbf{r}}\leq0$. In this case the Griffiths-Kelly-Sherman inequalities still hold, and so does Eq. (4.27). Thus S_{-} increases monotonically with box size L as before. We can give an upper bound on the magnitude of the S's (and S_{-}'s). By the Griffiths-Kelly-Sherman inequality, the S's increase if we add enough to $h_{\mathbf{r}}$ at every cell to make them all equal to h, the maximum $|h_{\mathbf{r}}|$ in (5.27). Since this procedure leaves a

translationally invariant lattice, the arguments leading to (4.23) apply and yield for this case

$$S(\mathbf{r}_1, \ldots, \mathbf{r}_n) \leq \frac{\int_0^\infty x^n \exp[-\frac{1}{2}m_0^2 \Delta^d x^2 - g\sum_{j=0}^{p} a_j : x^{2j} : + hx]\,dx}{\int_0^\infty \exp[-(\frac{1}{2}m_0^2 \Delta^d + d\Delta^{d-2})x^2 - g\sum_{j=0}^{p} a_j : x^{2j} : + hx]\,dx}. \tag{5.28}$$

Thus, the limit as $L \to \infty$ exists for each Schwinger function. We proceed by a variant of the previous method. Consider

$$Y_-(\mathbf{h}_\mathbf{r}, L) = \ln[Z_-(\mathbf{h}_\mathbf{r})/Z_-(0)]. \tag{5.29}$$

The partial derivatives of Y_- with respect to the $h_\mathbf{r}$ give the Ursell functions

$$\begin{aligned} U_-(\mathbf{r}) &= S_-(\mathbf{r}), \\ U_-(\mathbf{r}, \mathbf{s}) &= S_-(\mathbf{r}, \mathbf{s}) - S_-(\mathbf{r})S_-(\mathbf{s}), \\ U_-(\mathbf{r}, \mathbf{s}, \mathbf{t}) &= S_-(\mathbf{r}, \mathbf{s}, \mathbf{t}) - S_-(\mathbf{r})S_-(\mathbf{s}, \mathbf{t}) \\ &\quad - S_-(\mathbf{s})S_-(\mathbf{r}, \mathbf{t}) - S_-(\mathbf{t})S_-(\mathbf{r}, \mathbf{s}) \\ &\quad + 2S_-(\mathbf{r})S_-(\mathbf{s})S_-(\mathbf{t}), \end{aligned} \tag{5.30}$$

$$\cdots$$

from which the Schwinger functions can be directly constructed as

$$\begin{aligned} S_-(\mathbf{r}) &= U_-(\mathbf{r}), \\ S_-(\mathbf{r}, \mathbf{s}) &= U_-(\mathbf{r}, \mathbf{s}) + U_-(\mathbf{r})U_-(\mathbf{s}), \\ S_-(\mathbf{r}, \mathbf{s}, \mathbf{t}) &= U_-(\mathbf{r}, \mathbf{s}, \mathbf{t}) + U_-(\mathbf{r})U_-(\mathbf{s}, \mathbf{t}) \\ &\quad + U_-(\mathbf{s})U_-(\mathbf{r}, \mathbf{t}) + U_-(\mathbf{t})U_-(\mathbf{r}, \mathbf{s}) \\ &\quad + U_-(\mathbf{r})U_-(\mathbf{s})U_-(\mathbf{t}), \end{aligned} \tag{5.31}$$

$$\cdots$$

$$S(\mathbf{r}_1, \ldots, \mathbf{r}) = \sum_{\text{partitions}} \prod U_-.$$

We next wish to show that Y_- is monotonically increasing in L. First $Y_-(0, L) = 0$. If we differentiate Y_- with respect to a single $h_\mathbf{r}$, we obtain

$$\begin{aligned} \frac{\partial Y_-}{\partial h_\mathbf{r}} &= E_-(\phi_\mathbf{r} \exp(H_n - gV))/E_-(\exp(H_n - gV)), \\ &= S_-(\mathbf{r}). \end{aligned} \tag{5.32}$$

If we follow the line of argument given at (4.26) and (4.27), we conclude that if λ is a parameter which adds terms to the action A_- so as to increase the system size, that

$$\frac{\partial S_-(\mathbf{r})}{\partial \lambda} \geq 0, \tag{5.33}$$

as the Griffiths–Kelly–Sherman inequalities continue to hold provided all $h_\mathbf{r} \geq 0$. Thus, integrating Eq. (5.33) with respect to $h_\mathbf{r}$, we conclude

$$\frac{\partial Y_-}{\partial \lambda} \geq 0. \tag{5.34}$$

We complete the argument on each $h_\mathbf{r}$ separately until the desired value of $\mathbf{h}_\mathbf{r}$ is reached.

We are now in a position to construct a set of generalized Padé approximants. Let us fix our attention on $\mathbf{h}_\mathbf{r}$ belonging to some finite dimensional closed region $\mathcal{H}$ which contains the point $\mathbf{h}_\mathbf{r} = 0$, and in which all $h_\mathbf{r} \geq 0$. We now have, as with (5.5)

$$B_{n,0}(g, \mathbf{h}_\mathbf{r}, L) \geq E_-(\exp[H_n - g(V + M)]) \geq B_{n,-1}(g, \mathbf{h}_\mathbf{r}, L). \tag{5.35}$$

Thus it follows directly that

$$b_n(g, \mathbf{h}_\mathbf{r}, L) \equiv \ln\left(\frac{B_{n,-1}(g, \mathbf{h}_\mathbf{r}, L)}{B_{n,0}(g, 0, 1)}\right) \leq Y_-(h_\mathbf{r}, L). \tag{5.36}$$

Let us define $L(\mathbf{h}_\mathbf{r}, n)$ as that L for which $b_n(g, \mathbf{h}_\mathbf{r}, L)$ is a maximum, and $L(n)$ the minimum $L(\mathbf{h}_\mathbf{r}, n)$ over all $\mathbf{h}_\mathbf{r}$ in $\mathcal{H}$. Since for some point in $\mathcal{H}$ the minimum is obtained, by arguments like those at (5.10) it follows that $L(n)$ tends to infinity with n. By integration of the uniform bounds in L for the S's and by use of Eq. (5.30) we can show that Y_- is uniformly bounded for all $\mathbf{h}_\mathbf{r}$ in $\mathcal{H}$. Thus as the Y_- are monotonic in L they tend to a limit. By arguments like those at (5.12) we conclude

$$Y_-(h_\mathbf{r}) = \lim_{n\to\infty} b_n(g, \mathbf{h}_\mathbf{r}, L(n)) = \lim_{L\to\infty} Y_-(h_\mathbf{r}, L) \tag{5.37}$$

exists. By standard theorems on the limit of a bounded sequence of analytic functions (in the $h_\mathbf{r}$) we conclude that $Y_-(h_\mathbf{r})$ is analytic in the interior of $\mathcal{H}$ and continuous on the boundaries. By the uniform bounds (5.28) we may differentiate $Y_-(h_\mathbf{r})$ and use (5.31) to construct the Schwinger functions. They can be defined at $\mathbf{h}_\mathbf{r} = 0$ by continuity. The reason for this careful construction is that in the infinite volume limit there might possibly be a phenomenon analogous to a phase transition so that $S(\mathbf{r}) \neq 0$ and $S(\mathbf{r})$ with $h = 0^+$ might equal minus that for $h = 0^-$. We have thus constructed for all real positive g the Schwinger functions for Dirichlet boundary conditions for our lattice theory.

By the method of construction, the formal series expansions of the Schwinger functions so constructed agree with those directly computed for the Schwinger functions, provided the latter are finite. If, before taking the limit $L \to \infty$, we expand the numerator and denominator in Eq. (4.4) for Dirichlet boundary conditions, we obtain, in the notation of (5.20), and letting

$$\phi_S = \prod_{\mathbf{r}_i \in S} \phi_{\mathbf{r}_i} \tag{5.38}$$

be the product of the fields involved in the Schwinger function under consideration. We again use $\langle\ \rangle_-$ to denote a normalized expectation value. Thus

$$\begin{aligned} S_-(r_1, \ldots, r_n) &= \langle\phi_S\rangle_- - g\Delta^d \sum_\mathbf{r} (\langle\phi_S X_\mathbf{r}\rangle_- \\ &\quad - \langle\phi_S\rangle_-\langle X_\mathbf{r}\rangle_-) + \frac{g^2}{2!}\Delta^{2d}\sum_\mathbf{r}\sum_\mathbf{s}[\langle\phi_S X_\mathbf{r} X_\mathbf{s}\rangle \\ &\quad - \langle\phi_S X_\mathbf{r}\rangle_-\langle X_\mathbf{s}\rangle_- - \langle\phi_S X_\mathbf{s}\rangle_-\langle X_\mathbf{r}\rangle_- - \langle\phi_S\rangle_-\langle X_\mathbf{r} X_\mathbf{s}\rangle_- \\ &\quad + 2\langle\phi_S\rangle_-\langle X_\mathbf{r}\rangle_-\langle X_\mathbf{s}\rangle_-] \\ &\quad - \frac{g^3}{3!}\Delta^{3d}\sum_\mathbf{r}\sum_\mathbf{s}\sum_\mathbf{t} U_-(L, \phi_S, X_\mathbf{r}, X_\mathbf{s}, X_\mathbf{t}) + \cdots. \end{aligned} \tag{5.39}$$

We may again apply the arguments at Eq. (5.26) which

now imply that a significant contribution to the sums comes only in the neighborhood of the set S, and hence we obtain for $m_0^2 > 0$, $\Delta > 0$, a formal power series in g with all terms finite. Again we find, by the boundedness of the derivative with respect to g, that the S_-'s are analytic in g in the interior of the same angular wedge W as was f_-, and asymptotic in W at $g = 0$. This result completes the construction of the infinite volume limit of the free energy per unit volume and the Schwinger functions for our lattice theory.

6. MASS RENORMALIZATION

In this section we will give the general framework for mass renormalization for a Euclidean lattice theory. The results will not be as complete as in the previous sections. Much fuller results will be given in the next section for $g:\phi^4:$ theory. We remind the reader that the power counting arguments applied to standard, quantum field, perturbation theory indicate that for interaction (4.1) infinite mass renormalizations are not expected when the degree $2p$ satisfies

$$p < 1 + 1/(d-2), \tag{6.1}$$

where d is the space-time dimension. Consequently this problem has not really entered into the extensive literature of $:P(\phi)_2:$ theory, although it has certainly not been ignored.

The usual practice in Minkowski space field theory is to define the mass renormalization in terms of a pole in the propagator at $k^2 = -m^2$. As this definition, at least for a finite mass, refers to the behavior at low momenta, we need to study the behavior at large distances. We would expect that

$$0 \leq \langle \phi_0 \phi_{\mathrm{j}} \rangle \propto \exp(-m\Delta|\mathrm{j}|), \tag{6.2}$$

where j is the number of lattice steps $(\mathbf{r}/\Delta)$ and Δ is the lattice spacing. In the language of statistical mechanics, the renormalized mass m is proportional to the inverse correlation length

$$\xi = (m\Delta)^{-1}. \tag{6.3}$$

Physically we view mass renormalization as the adjustment of the mass parameter m_0^2 in the Hamiltonian in such a way so as to make the renormalized mass m take on its experimental value for given values of Δ and g. When Δ is chosen smaller and smaller to correspond to a vanishing lattice spacing, the correlation length must, by (6.3), tend toward infinity. This behavior of the correlation is characteristic of statistical mechanical systems which are approaching a critical point, such as the Curie point where spontaneous magnetization becomes possible. In order to illustrate the analogy we will briefly review the Gaussian model and the lattice free field.

The nearest-neighbor Gaussian model[27] is an assembly of spins whose average magnitude is unity and which are distributed as

$$(2\pi)^{-1/2} \exp(-\nu_i^2/2)\, d\nu_i \tag{6.4}$$

and interact as

$$-J \sum_{\substack{\text{nearest} \\ \text{neighbors}}} \nu_i \nu_j. \tag{6.5}$$

The partition function is

$$Z = (2\pi)^{-N^d/2} \int_{-\infty}^{+\infty} \cdots \int \exp\left(\beta J \sum_{\substack{\text{nearest} \\ \text{neighbors}}} \nu_i \nu_j - \tfrac{1}{2} \sum_i \nu_i^2\right) \prod_i d\nu_i, \tag{6.6}$$

where $\beta = 1/k_B T$, k_B is Boltzmann's constant, and T is the absolute temperature. For a d-dimensional, hypercubic lattice the spin-spin correlation function is

$$Z^{-1}(2\pi)^{-N^d/2} \int_{-\infty}^{+\infty} \cdots \int \nu_{\mathrm{a}} \nu_{\mathrm{b}} \exp\left(\beta J \sum_{\substack{\text{nearest} \\ \text{neighbors}}} \nu_i \nu_j - \tfrac{1}{2} \sum_i \nu_i^2\right) \prod d\nu_i$$

$$= \langle \nu_{\mathrm{a}} \nu_{\mathrm{b}} \rangle = \frac{1}{N^d} \sum_{k_1=1}^{N} \cdots \sum_{k_d=1}^{N} \frac{\cos\{2\pi[k \cdot (\mathrm{a}-\mathrm{b})]/N\}}{1 - 2\beta J \sum_{\tau=1}^{d} \cos(2\pi k_\tau / N)}, \tag{6.7}$$

where the critical temperature at which all spin-spin correlations diverge for this model is given by

$$2d\beta J = 1, \quad T_c = k_B/(2dJ). \tag{6.8}$$

For $|\mathrm{a}-\mathrm{b}|$ large and $T \gtrsim T_c$ we may compute the asymptotic behavior of (6.7) by expanding in the denominator $\cos x = 1 - \frac{1}{2}x^2 + 0(x^4)$ and integrating.[28] In the limit of a large system, we obtain

$$\langle \nu_{\mathrm{a}} \nu_{\mathrm{b}} \rangle \approx \frac{\beta J (|\mathrm{a}-\mathrm{b}|/\xi)^{(d-1)/2}}{(2\pi)^{1/2d} |\mathrm{a}-\mathrm{b}|^{d-2}} K_{(d-1)/2}(|a-b|/\xi)$$

$$\approx \frac{\beta J \xi^{(3-d)/2} \exp(-|\mathrm{a}-\mathrm{b}|/\xi)}{2^{(d+1)/2} \pi^{(d-1)/2} |\mathrm{a}-\mathrm{b}|^{(d-1)/2}}, \tag{6.9}$$

where $K_l(x)$ is a modified Bessel function of the second kind and ξ is the correlation length given by

$$\xi = [\beta J/(1 - 2d\beta J)]^{1/2}. \tag{6.10}$$

The expression analogous to (6.6) for lattice field theory is the periodic boundary condition analog of (2.4). If we introduce the change of scale

$$\psi_{\mathbf{r}}^2 = (2d\Delta^{d-2} + \Delta^d m_0^2)\phi_{\mathbf{r}}^2, \tag{6.11}$$

then by direct substitution and comparison we identify

$$\beta J = 1/(2d + \Delta^2 m_0^2), \quad \xi = (\Delta m_0)^{-1}, \tag{6.12}$$

which tends to the critical point as $\Delta \to 0$. For the correlations we have

$$\langle \phi_{\mathrm{a}} \phi_{\mathrm{b}} \rangle \approx (d\Delta^{d-2} + \Delta^d m_0^2)^{-1} \langle \psi_{\mathrm{a}} \psi_{\mathrm{b}} \rangle,$$

$$\approx \frac{m_0^{(d-3)/2} \exp(-m_0 \Delta |\mathrm{a}-\mathrm{b}|)}{d^2 2^{(d+3)/2} \pi^{(d-1)/2} (\Delta |\mathrm{a}-\mathrm{b}|)^{(d-1)/2}}$$

$$\approx \frac{m_0^{(d-3)/2} \exp(-m_0 |\mathbf{r}|)}{d^2 2^{(d+3)/2} \pi^{(d-1)/2} |\mathbf{r}|^{(d-1)/2}} \tag{6.13}$$

in the case where $\xi \gg 1$. We denote the distance as the lattice spacing times the number of cells as in Sec. 2; $|\mathbf{r}| = \Delta |\mathrm{a}-\mathrm{b}|$. Thus we see that the analogy is that as the lattice spacing tends to zero (ultraviolet cutoff tends to infinity) the "temperature" tends to the critical temperature from above in such a way as to produce an

exponential decay on a fixed (i.e., not lattice dependent) scale.

We remark that although the underlying lattice does not possess rotational symmetry, the long-range (compared to the cell size) correlations do. As this property is a usual one in critical phenomena, we anticipate it here also. In the above discussion we have not been careful about the boundary conditions and the results are for short, long-range behavior, i.e., long compared to the cell size but short compared to the total system size. Clearly, $|\mathbf{r}|$ in (6.13) for periodic boundary conditions would have to refer to the shortest distance between a and all the periodic images of b.

We introduce, for the case of Dirichlet conditions the following definition of the renormalized mass for a system of size $L < \infty$

$$m(L) \equiv \min_{\mathbf{r},\mathbf{s}} \left\{ \frac{-\ln[\langle \phi_{\mathbf{r}} \phi_{\mathbf{s}} \rangle_{-}(A + m^{\alpha}|\mathbf{r}-\mathbf{s}|^{(d-1)/2}}{|\mathbf{r}-\mathbf{s}|} \right\}, \tag{6.14}$$

where the $\mathbf{r}$ and $\mathbf{s}$ are taken in the conventions of Sec. 2, and we select, independent of the system size,

$$A^{-1} = \min \left(\frac{\int_0^{\infty} x^2 \exp[-\frac{1}{2}m_0^2\Delta^d x^2 - g\sum_{j=0}^{p} a_j :x^{2j}:]\,dx}{\int_0^{\infty} \exp[-(\frac{1}{2}m_0^2\Delta^d + d\Delta^{d-2})x^2 - g\sum_{j=0}^{p} a_j :x^{2j}:]\,dx} \right), \tag{6.15}$$

where the min is over a preselected range of m_0^2, in accordance with (4.23) to insure $m(L) \geq 0$, and in (6.14) $\alpha > 0$ is arbitrary. This definition has not been proven to be the same as that of the true renormalized mass, m_T. It is true at least that $m_T \geq m$, and that they go to zero together, although the rate may possibly not be the same.

We now proceed to show that (6.14) does define a value of m in terms m_0^2, Δ, and g in the limit of infinite system size. First adjoin an additional hyperplane of cells which are uncoupled to the rest of the system. Then $\langle \phi_{\mathbf{r}} \phi_{L+\Delta} \rangle = 0$ and so $-\ln\langle \phi_{\mathbf{r}} \phi_{L+\Delta} \rangle = +\infty$. Therefore, this addition does not change the location of the minimum in (6.14). Now by (4.27) every $\langle \phi_{\mathbf{r}} \phi_{\mathbf{s}} \rangle$ increases as we couple the additional hyperplane of cells. Therefore every term in the minimum decreases, and so

$$m(L+1) \leq m(L). \tag{6.16}$$

But since $m(L) \geq 0$ for all L, we may define in the infinite volume limit the renormalized mass as

$$m = \lim_{L\to\infty} m(L). \tag{6.17}$$

An additional property[29] is that the renormalized mass is a monotonic increasing function of the explicit dependence on m_0^2. This result follows easily from the Griffiths (A3) and Griffiths-Kelly-Sherman inequalities (A5) as the derivative of every term in the minimum in (6.14) is proportional to

$$\sum_{\mathbf{t}} \frac{\langle \phi_{\mathbf{r}} \phi_{\mathbf{s}} \phi_{\mathbf{t}}^2 \rangle - \langle \phi_{\mathbf{r}} \phi_{\mathbf{s}} \rangle \langle \phi_{\mathbf{t}}^2 \rangle}{\langle \phi_{\mathbf{r}} \phi_{\mathbf{s}} \rangle} \geq 0. \tag{6.18}$$

Since the minimum is realized for some $\mathbf{r}$, $\mathbf{s}$ it must increase as long as m_0^2 is in the preselected range for (6.15). There is also an implicit dependence on m_0^2 through the constant C, Eq. (3.11), which enters into the definition of the normal ordered product. C decreases as m_0^2 increases. For the special case of a $g:\phi^4:$ interaction there is a $-6gC\phi^2$ term so that the complete coefficient of ϕ^2 increases monotonically with m_0^2, and the above argument holds for the complete dependence on m_0^2.

We will next show that by a proper choice of m_0^2 we can make m as large as we please or as small (>0) as we please. If we could also show that $m(m_0^2)$ is continuous (true for $g:\phi^4:_d$ theory), then we would be able to conclude that the theories we have been discussing are mass renormalizable. We consider now two cases. In the first case $g=0$. Then $m = |m_0|$ and $m_0^2 > 0$, so the proposition is true.

In the second case $g > 0$. Here we can choose $m_0^2 > 0$ so large, independent of L, so that to whatever accuracy we desire, the entire contribution to the integrand for

$$Z_{-} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp[-\tfrac{1}{2}\Delta^d \sum_{\mathbf{r}} [-2\Delta^{-2} \sum_{\{\delta\}} (\phi_{\mathbf{r}}\phi_{\mathbf{r}+\delta}) + (m_0^2 + 2d\Delta^{-2})\phi_{\mathbf{r}}^2 + 2gP(\phi_{\mathbf{r}})]] \prod_{\mathbf{r}} d\phi_{\mathbf{r}} \tag{6.19}$$

comes from the central peak $\phi_{\mathbf{r}} \approx 0$, and in that region $|gP(\phi_{\mathbf{r}})| < \epsilon$, for any $\epsilon > 0$. [$P(\phi)$ is defined by Eq. (5.1) and C by Eq. (3.11). The constant C decreases as m_0^2 increases.] Thus for m_0^2 sufficiently large and positive, $m \approx m_0$ and can be made as large as we like. We remark that if $p=1$ [degree of $P(\phi)$ is $2p$] then $m^2 = m_0^2 + 2g$ so that clearly the proposition is true. Otherwise, if $p > 1$, we select m_0^2 to be very small and positive. By the structure of $P(\phi_j)$ there will be two equal peaks in the integrand of (6.19) for each $\phi_{\mathbf{r}}$ near $\phi_{\mathbf{r}} = \pm\hat{\phi}$. We compute

$$P'(\hat{\phi}) = 0, \quad \hat{\phi} \propto \sqrt{C}\,, \tag{6.20}$$

where C is given by (3.11). There are two cases to distinguish: first, if $d=2$, $C \to \infty$ as $m_0^2 \to 0$, and, secondly, if $d > 2$, C remains finite as $m_0^2 \to 0$. If we introduce the scaled variables

$$\sigma_{\mathbf{r}} = \phi_{\mathbf{r}}/\hat{\phi}, \tag{6.21}$$

then, by selecting m_0^2 small enough, $d=2$, for all Δ, or by selecting Δ small enough, by (3.13), for $d > 2$, we have to any accuracy we like insofar as the computation of $\langle \phi_{\mathbf{r}} \phi_{\mathbf{s}} \rangle$ is concerned when $d > 2 + 2/(p-1)$, that the expression (6.19) is equivalent to

$$Z_{-} \propto \sum_{\{\sigma_{\mathbf{r}}=\pm 1\}} \exp(\Delta^{d-2}\hat{\phi}^2 \sum_{\mathbf{r}} \sum_{\{\delta\}} \sigma_{\mathbf{r}}\sigma_{\mathbf{r}+\delta}), \tag{6.22}$$

where

$$\Delta^{d-2}\hat{\phi}^2 \geq (0.99)\,3\Delta^{d-2}C, \tag{6.23}$$

for all $p \geq 2$ and $\Delta > 0$, sufficiently small.

For $d = 2 + 2/(p-1)$, instead of a sum over $\{\sigma_{\mathbf{r}} = \pm 1\}$ we have an integral over a bimodal spin weight distribution of finite peak height, even in the limit $\Delta \to 0$. For $0 \leq (d-2)(p-1) < 2$ the peak height decreases as $\Delta \to 0$. However, as softer spins, e.g., compare the Gaussian "soft spins" with the Ising "hard spins," allow spin states with higher correlation energy for the same mean spin, we should have a higher critical temperature or a smaller critical value of $\Delta^{d-2}\hat{\phi}^2$ than for the "hard spin" problem (6.22). Thus we will continue to work

TABLE I. Comparison with Ising critical values.

Dimension	$K_c(d)$	$3c(d)$
3	0.22171	0.63273
4	0.14988	0.37623
5	0.11403	0.28209
6	0.09236	0.23023

with (6.22) for all $d \geq 2$, even though we have not given a rigorous justification for all dimensions.

For $d=2$ we can make the right-hand side of (6.23) as large as we please by choosing Δ small enough. For $d>2$, by choosing Δ and m_0^2 small enough, we can make C from (3.11) as close as we like to

$$\Delta^{d-2}C = \frac{1}{(2\pi)^d}\int_{-\pi}^{\pi}\cdots\int \frac{d\mathbf{k}}{2d-2\sum_{\tau=1}^{d}\cos(k_\tau)}$$

$$= \tfrac{1}{2}\int_0^\infty [\exp(-y)I_0(y)]^d\,dy, \tag{6.24}$$

by the integral representation of the Bessel function. We may obtain a lower bound by observing[30] that

$$\exp(-y)I_0(y) \geq \begin{cases} 1-y, & 0 \leq y \leq 0.3, \\ (2\pi y)^{-0.5}, & y \geq 0.3, \end{cases} \tag{6.25}$$

which implies, by integrating (6.24),

$$\Delta^{d-2}C \geq c(d) = \frac{1-(0.7)^{d+1}}{2d+2} + \frac{0.3}{d-2}(0.6\pi)^{-0.5d}. \tag{6.26}$$

We seek to show that (6.22) is equivalent to a low-temperature ($\Delta^{d-2}\hat{\phi}^2$ large) Ising model. As we remarked above, for $d=2$ we can make $\Delta^{d-2}\hat{\phi}^2$ as large as we please by choosing Δ small enough, in particular, larger than $K_c = 0.440\,687$, the critical value which separates the high and low temperature regions. With accuracy sufficient for our purposes we give in Table I the critical values for various d's from Fisher and Gaunt[31] for the Ising model, and the lower bound given by (6.26) and (6.23) for $\Delta^{d-2}\hat{\phi}^2$. For dimensions higher than those listed in Table I, the same result holds, i.e., $3c(d) > K_c(d)$. Thus by adjusting $m_0^2 > 0$, and $\Delta > 0$ sufficiently small, we have a low-temperature Ising model. However, for $d \geq 2$, by the Onsager solution[32] and Griffiths inequality (A3), it is well known that

$$\liminf_{|\mathbf{r}|\to\infty} \langle \sigma_0 \sigma_{\mathbf{r}} \rangle = \mu > 0 \tag{6.27}$$

as long-range order exists. Therefore, there will be terms in (6.14) at least as small as

$$-\ln[\mu(A + m^{\alpha}L^{(d-1)/2})]/L, \tag{6.28}$$

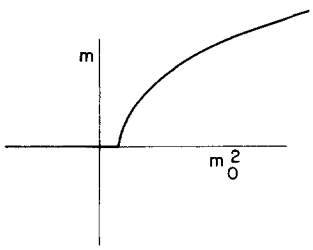


FIG. 2. Expected dependence of the renormalized mass on the bare mass.

which tends to zero as $L \to \infty$. Thus we can prechoose the range on m_0^2 for (6.15) from the required accuracy and hence we have shown that we can select m_0^2 such that m is zero or m is as large and positive as we please. As in addition, $m(m_0^2)$ is monotonic, mass renormalizability for any $\Delta > 0$, $g \geq 0$ would follow provided $m(m_0^2)$ is also continuous. The question of continuity is not wholly an idle mathematical refinement, in as much as there are examples for which it fails. We expect a curve as shown in Fig. 2, however, the result shown in Fig. 3 occurs for the two-dimensional Ising model[33] given by the Hamiltonian

$$H = -J \sum_{\substack{\text{nearest} \\ \text{neighbors}}} \sigma_i\sigma_j - b(\sum_i \sigma_i)^2/N, \tag{6.29}$$

where $\sigma_i = \pm 1$, and N is the number of spins. The second term causes every spin to interact with the "mean field" of all spins. The critical temperature is the solution of the equation

$$2b\chi(\beta_c J) = 1, \tag{6.30}$$

where χ is the standard, reduced, magnetic susceptibility[34] of the nearest-neighbor part. The transition is of the familar Bragg—Williams type, and at the temperature determined by (6.30) the inverse correlation length as defined by (6.14) drops discontinuous from the value determined by the nearest-neighbor part of the Hamiltonian above to zero.

7. $g{:}\phi^4{:}$ THEORY

In this section we give much more complete results for $g{:}\phi^4{:}_d$ in arbitrary dimension than we have obtained for more general theories. Our discussion encompasses $g(:\phi^4:_d + a:\phi^2:_d)$ but for ease of presentation we do not describe this generalization explicitly. The theory of $g:\phi^2:_d$ follows immediately by direct calculation. As was pointed out in the previous section the renormalized mass is given by $m^2 = m_0^2 + 2g$. The scattering amplitude vanishes. That is to say, $U_-(\mathbf{r}, \mathbf{s}, \mathbf{t}, \mathbf{u}) = 0$, in the notation of (5.30).

The first property we establish is the mass renormalizability of this theory. From the work of the previous section, $m^2(m_0^2)$, the infinite volume limit for fixed lattice spacing $\Delta > 0$, and coupling constant $0 \leq g < \infty$ is monotonic increasing in m_0^2 from 0 to an indefinitely large value. We show in Appendix C that it is also a continuous function of m_0^2. Thus, by Bolzano's theorem, there exists a solution of the equation

$$\mu = m(m_0^2), \quad 0 \leq \mu < \infty. \tag{7.1}$$

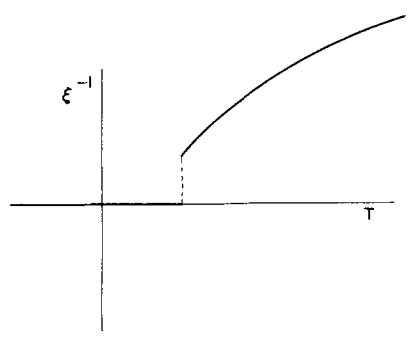


FIG. 3. Discontinuity exhibited in the inverse correlation length (analogous to the renormalized mass) for a special Ising model.

For definiteness we select the largest such solution, since when, for example, $\mu = 0$ many solutions are possible. Thus we can select

$$m_0^2 = m^{-1}(\mu, g, \Delta) \tag{7.2}$$

as that value of the unrenormalized mass which makes the renormalized mass take on the preselected (experimental) value μ.

The next property we show is the cluster property. We will assume that m_0^2 is chosen in accordance with Eq. (7.2) for $\mu > 0$. By the cluster property one means that if there are two groups of cells $\mathbf{r}_1, \ldots, \mathbf{r}_j$ and $\mathbf{s}_1, \ldots, \mathbf{s}_k$ which are separated by a distance ρ, then

$$\begin{aligned} 0 &\leq S_-(\mathbf{r}_1, \ldots, \mathbf{r}_j, \mathbf{s}_1, \ldots, \mathbf{s}_k) - S_-(\mathbf{r}_1, \ldots, \mathbf{r}_j) S_-(\mathbf{s}_1, \ldots, \mathbf{s}_k) \\ &\leq K \exp(-\mu\rho), \end{aligned} \tag{7.3}$$

where K depends only on j and k. The first inequality sign follows from the Griffiths–Kelly–Sherman inequalities (A5). The second inequality which embodies the cluster property follows from the Lebowitz inequalities via the derivation in Appendix A of (A23) and bound (4.23) on the other Schwinger functions.

The cluster property allows us to bound Eq. (4.8) and the analogous one for the relation between Dirichlet and free-boundary conditions by a factor which decreases exponentially with the system size. Since the surface area and hence the number of terms is only of the order of $2dL^{d-1}/\Delta^{d-1}$, the difference in the S's and f's due to boundary conditions disappears in the limit as $L \to \infty$. Thus, the construction procedures we gave in Sec. 5 construct only a single, free-energy density in this case and allows us to conclude that the Schwinger functions for periodic boundary conditions and free boundary conditions also converge as $L \to \infty$ and to the same limit as for Dirichlet boundary conditions.

Next the cluster property allows us to deduce for any fixed $0 < g < \infty$, $\Delta > 0$ that the S's and f are continuous functions of g. In order to see this result we observe that for L finite

$$\left.\frac{\partial S_-(\mathbf{r}_1, \ldots, \mathbf{r}_n)}{\partial g}\right|_{m_0^2} = -\Delta^d \sum_{\mathbf{r}} (\langle \phi_S X_\mathbf{r} \rangle_- - \langle \phi_S \rangle_- \langle X_\mathbf{r} \rangle_-) \tag{7.4}$$

in the notation of Sec. 5. By the application of inequality (A23) and inequality (C4) for the two-point correlation function implied by the definition of mass renormalization, we can compute a finite bound for (7.4) which is uniform in L and hence holds as $L \to \infty$. Next we need to consider

$$\frac{\partial m(L)}{\partial g} = \Delta^d \sum_{\mathbf{t}} \frac{\langle \phi_\mathbf{r}\phi_\mathbf{s} X_\mathbf{t} \rangle - \langle \phi_\mathbf{r}\phi_\mathbf{s} \rangle \langle X_\mathbf{t} \rangle}{\langle \phi_\mathbf{r}\phi_\mathbf{s} \rangle\, |\mathbf{r} - \mathbf{s}|}\,. \tag{7.5}$$

By applying inequality (A23) we can bound (7.5) by a numerical factor times (C2). From there on the proof of Appendix C implies that $|\partial m(L)/\partial g|$ is bounded uniformly in L. Now if we compute the second derivative of $m(L)$ we get from Eq. (C1) and summing first over the $\sigma_\mathbf{r}$ of (4.10) by Griffiths inequalities (A3),

$$\frac{\partial^2 m(L)}{\partial (m_0^2)^2} \propto -\Delta^{2d} \frac{\langle \phi_\mathbf{r}\phi_\mathbf{s} (\sum_\mathbf{t} \phi_\mathbf{t}^2 - \langle \phi_\mathbf{t} \rangle)^2 \rangle}{\langle \phi_\mathbf{r}\phi_\mathbf{s} \rangle\, |\mathbf{r} - \mathbf{s}|} < 0, \tag{7.6}$$

taking account of the explicit dependence and the dependence through C, Eq. (3.11). Thus, as by (6.19) *et seq.* for large enough m_0, $m \approx m_0$, we have, by integrating (7.6) from m_0 to $\bar{m}_0$

$$\frac{\partial m(L)}{\partial m_0^2} \geq \tfrac{1}{2}(\bar{m}_0)^{-1} > 0 \tag{7.7}$$

for some large $\bar{m}_0$. Thus by the rules of partial derivatives we may bound

$$\frac{dS_-}{dg} = \left.\frac{\partial S_-}{\partial g}\right|_{m_0^2} + \left.\frac{\partial S_-}{\partial m_0^2}\right|_g \left.\frac{\partial m(L)}{\partial g}\right|_{m_0^2} \Big/ \left.\frac{\partial m(L)}{\partial m_0^2}\right|_g \tag{7.8}$$

and thus conclude that S_- is a continuous function of g when we renormalize the mass. The same type of analysis also applies to f and we conclude that f is also continuous.

The general proof of mass renormalizability in Section 6 assumed that g was fixed and independent of Δ. Since $m = m_0$ for $g = 0$, and the two-point correlation function is continuous in g, we may also conclude mass renormalizability if $g \to 0$ as Δ goes to zero. Such a situation has to be considered if a coupling constant renormalization is required.

We next show that in addition to the independence of the boundary conditions demonstrated above, the $g:\phi^4:$ theory for $\Delta > 0$ is uniquely determined, in so far as it is analytic, by its formal power series in g. This result is related to the Borel summability proved by Simon[35] in two dimensions, and by Glimm and Jaffe[36] in four dimensions on the basis of a conjectured bound. It is most convenient to use the momentum transforms of the Schwinger functions

$$G(k_1, \ldots, k_n) = \Delta^{nd} \sum_{\mathbf{r}} \cdots \sum_{\mathbf{r}_n} \exp\left(2\pi i \sum_{j=1}^{n} \mathbf{k}_j \cdot \mathbf{r}_j\right) S(\mathbf{r}_1, \ldots, \mathbf{r}_n) \tag{7.9}$$

for periodic boundary conditions. Then the formal expansion (5.39) becomes in the usual way[15] an expansion in terms connected graphs with n-external lines. The standard procedure[37] is to bound the number and size of such terms. Direct enumeration[37,38] of the number of such diagrams corresponding to the coefficient of g^n in (5.39) leads to the conclusion that there are at most a $B^n(2n!)$ such terms. Each term differs from the usual form by having

$$\frac{1}{m_0^2 + 4\Delta^{-2} \sum_{(\delta)} \sin^2(\pi \mathbf{k} \cdot \delta)} \tag{7.10}$$

for a propagator and

$$\int_{-\infty}^{+\infty} d\mathbf{k} \to L^{-d} \sum_{\mathbf{k}=-1/2\Delta}^{1/2\Delta} \tag{7.11}$$

replacing the infinite momentum integrals. Since every momentum integral is less than $(1/\Delta)^d$ times the maximum integrand and every propagator is less than $1/m_0^2$, we conclude directly that the nth term in the expansion (5.39) is bounded by

$$(B')^n (2n)!/n! \approx (2B')^n (n!), \tag{7.12}$$

for some $B'(\Delta) < \infty$, when $\Delta > 0$. This result will hold, uniformly in the angular wedge W defined in Sec. 5, i.e., $|z| \leq \omega$, $|\arg(z)| \leq \pi/2$ for some $\omega > 0$, uniformly in L. We may then apply[39]:

Carleman's Theorem: If

$$|f(z)| \leq \alpha_n^n |z|^n \text{ for } 0 \leq |z| \leq \rho, \quad |\arg(z)| \leq \pi/2 \tag{7.13}$$

holds, then it is necessary and sufficient, for suitably regularized α_n, that $\sum \alpha_n^{-1}$ diverge, to conclude that $f(z) = 0$.

This theorem implies that in view of (7.12) there can be at most one function in the angular wedge W which is asymptotic to the formal power series in g. But we constructed such a function in Section 5. Thus it is uniquely determined by analytic continuation from the $\Delta > 0$ power series in the infinite volume limit without regard to the construction process which depends on the $L \to \infty$ behavior of the terms as far as that continuation will carry us. [We have not proved that the $S(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ are analytic in g for $0 < g < \infty$, so that the analytic continuation might not carry us very far. Dimock[2] has shown that distance does not vanish as $\Delta \to 0$ in two dimensions.] Since we get a series of the same structure for $m_0^2(\mu, g, \Delta)$, the mass renormalization does not interfere with the above arguments $\Delta > 0$. Also, the same conclusions as above hold for the free energy per unit volume, f.

In the limit as $\Delta \to 0$, we have by mass renormalization a bound, Eq. (C4) on the two-point function. If we set $A = 0$, this bound is uniform in Δ and finite as long as the points are not coincident. The bound is integrable in d-dimensional space. By inequalities (A23) we get an integrable bound in terms of that for the two-point function on all the high order Schwinger functions which is again finite if no two points are coincident, and uniform in Δ. Thus a limit $\Delta \to 0$ for the $g:\phi^4:$ theory can be defined. (If the limit does not exist directly, we can consider the sequence $\Delta_n = \epsilon/2^n$. Since then the lattice points for all values of n in any closed region are denumerable, and thus the different Schwinger functions defined on that region are denumerable, we can select at least a subsequence of the Δ_n for which every Schwinger function converges at every noncoincident point.)

In the cases $2 \leq d < 4$, as is well known, mass renormalization suffices to leave all the terms finite. The defining integrals converge, which allows the ultraviolet cutoff to be removed ($\Delta \to 0$). The procedure is to have formally rearranged the series to express the propagators in terms of the renormalized mass m, instead of the bare mass m_0, by summing out all the self-energy diagrams. Consequently bound (7.12) holds for the derivatives in the limit as $\Delta \to 0$ for $B'(0) < \infty$, as shown by Hurst.[37] However, we do not know, except in two dimensions[2] that the angular wedge W does not shrink to zero. If it does not, Carleman's theorem combined with the existence already shown gives us constructibility at least in W, directly from the power series. This result falls short of a proof of construction from the invariant perturbation theory, but we have established above the existence of a Euclidean theory and will return in the next section to alternate means of construction.

The final property is rotational invariance. We do not prove this property but observe that it is characteristic of phenomena near the critical point. For the two-dimensional Ising model[40] on the square lattice if has been explicitly obtained. To the extent that construction is possible directly from the perturbation series, rotational invariance follows from the term-by-term invariance. With this property assumed we can apply Nelson's[3] reconstruction theorem and construct a relativistic $g:\phi^4:$ field theory from our Euclidean one. This theory is certainly nontrivial at least for $2 \leq d < 4$, as it is asymptotic to a perturbation series which is nontrivial. We give a brief discussion of Nelson's reconstruction theorem as it relates to our case in Appendix D.

8. STATISTICAL MECHANICAL COMPUTATIONAL METHODS

Having established in the previous sections that the mass-renormalized, infinite-volume limit ($L \to \infty$) exists, and, for $g:\phi^4:_d$ theory at least, that it is independent of the boundary conditions, we now consider the actual computation of the theory ($\Delta > 0$). The results so far show us that the ultraviolet cutoff model is a continuous-spin Ising model. Mass renormalization forces this model to behave as though it were at a temperature just above the critical temperature and tending to the critical temperature as the ultraviolet cutoff is removed. One of the most successful methods to treat this type of problem has been the exact, high temperature series expansions[41] summed by the method of Padé approximants.[18] To illustrate the procedure, we will discuss a pure $g:\phi^4:_d$ theory. The partition function will be given by the expression ($\Delta > 0$)

$$Z_- = \int_{-\infty}^{+\infty} \cdots \int \prod_{\mathbf{r}} d\phi_{\mathbf{r}} \exp\Big\{\sum_{\mathbf{r}} \Big[\Delta^{d-2} \sum_{\{\delta\}} (\phi_{\mathbf{r}}\phi_{\mathbf{r}+\delta}) + h_{\mathbf{r}}\phi_{\mathbf{r}} - (d\,\Delta^{d-2} + \tfrac{1}{2}m_0^2\Delta^d)\,\phi_{\mathbf{r}}^2 - g\,\Delta^d(\phi_{\mathbf{r}}^4 - 6C\phi_{\mathbf{r}}^2 + 3C^2)\Big]\Big\}, \tag{8.1}$$

where C is defined by Eq. (3.11). The fundamental quantities that we will be concerned with are the correlation functions and their various moments

$$U_-(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \frac{\partial^n}{\partial h_1 \cdots \partial h_n} \ln Z_-\big|_{h=0},$$

$$\mu_t = \Delta^d \sum_{\mathbf{r}} |\mathbf{r}|^t U_-(0, \mathbf{r}),$$

$$G(\mathbf{k}) = \Delta^d \sum_{\mathbf{r}} \exp(2\pi i \mathbf{k}\cdot\mathbf{r})\, U_-(0, \mathbf{r}),$$

$$T(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \Delta^{3d} \sum_{\mathbf{r}} \sum_{\mathbf{s}} \sum_{\mathbf{t}} \exp(2\pi i(\mathbf{k}\cdot\mathbf{r} + \mathbf{k}_i\cdot\mathbf{s} + \mathbf{k}_3\cdot\mathbf{t}) \times U_-(0, \mathbf{r}, \mathbf{s}, \mathbf{t}). \tag{8.2}$$

The U's are the Ursell functions and directly related to the Schwinger functions by (5.30). The spherical moments, μ_t, are useful in determining the critical-point properties of the model. The moment for $t = 0$ is directly related to the magnetic susceptibility in the magnetic Ising model. The quantity $G(\mathbf{k})$ is just the propagator and

$T(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ the scattering function. In defining these quantities we use the proven translational invariance of the theory. It is convenient to introduce a change of scale

$$\psi_{\mathbf{r}} = \Delta^{(d-2)/2}\,\phi_{\mathbf{r}}, \tag{8.3}$$

and to rewrite (8.1) as

$$Z_- \propto \int_{-\infty}^{+\infty} \cdots \int \prod_{\mathbf{r}} d\psi_{\mathbf{r}} \exp\{\sum_{\mathbf{r}} [\lambda \sum_{\{\delta\}} \psi_{\mathbf{r}}\psi_{\mathbf{r}+\delta} + \Delta^{(2-d)/2} h_{\mathbf{r}}\psi_{\mathbf{r}}$$

$$- (\tfrac{1}{2}m_0^2\Delta^2 + d)\psi_{\mathbf{r}}^2 - g\Delta^{4-d}(\psi_{\mathbf{r}}^4 - 6C\Delta^{d-2}\psi_{\mathbf{r}}^2$$

$$+ 3C^2\Delta^{2d-4})]\}, \tag{8.4}$$

where $\lambda = 1$. By a change of variables we may rewrite (3.11) as

$$\Delta^{d-2}C = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} \cdots \int \frac{d\mathbf{k}}{m_0^2\Delta^2 + 2d - 2\sum_{\tau=1}^{d} \cos k_\tau}. \tag{8.5}$$

Let us first consider the case $d > 4$ without coupling constant renormalization. The coefficient of g diverges like $\Delta^{-(d-4)}$ as $\Delta \to 0$. Thus, for Δ small enough, this effect confines the contribution to the integral in (8.4) to the points

$$\psi_{\mathbf{r}} = \pm\hat{\psi},$$
$$\hat{\psi}^2 = 3C\Delta^{d-2}, \tag{8.6}$$

if we consider $m_0^2\Delta^2$ to be of order unity. We therefore have an Ising model of spin-$\frac{1}{2}$ with

$$K = J/kT = 3C\Delta^{d-2}, \tag{8.7}$$

as can be seen from the equivalence of (8.4) in this limit to

$$Z_- \propto \sum_{\{\sigma_{\mathbf{r}}=\pm 1\}} \exp\{K \sum_{\mathbf{r}} \sum_{\{\delta\}} \sigma_{\mathbf{r}}\sigma_{\mathbf{r}+\delta} + \Delta^{(2-d)/2} K^{1/2} h_{\mathbf{r}}\sigma_{\mathbf{r}}\}. \tag{8.8}$$

Our criterion for mass renormalization (Sec. 6) tells us that the correlation length is given by (6.3) in terms of the renormalized mass. Although as a practical matter it is probably not the most effective way to do the calculation, we may directly expand[42]

$$\xi \approx (1 - \exp(-1/\xi))^{-1}$$
$$= 1 + v + (2d-1)\,v^2 + (4d^2 - 6d + 3)v^3$$
$$+ (8d^3 - 20d^2 + 20d - 7)v^4$$
$$+ (16d^4 - 56d^3 + 84d^2 - 60d + 17)v^5$$
$$+ (32d^5 - 144d^4 + 296d^3 - 340d^2 + 214d - 57)v^6$$
$$+ (64d^6 - 352d^5 + 928d^4 - 1496d^3 + 1484d^2$$
$$- 782d + 155)v^7 + \cdots, \tag{8.9}$$

where

$$v = \tanh K. \tag{8.10}$$

We need to solve, then, for a v such that

$$\xi(v) = (m\Delta)^{-1}. \tag{8.11}$$

This solution is possible in this range of dimension as was shown [(6.19) *et seq.*] in Sec. 6. According to the renormalization group calculations of Wilson[43] and collaborators, we expect the singular behavior

$$\xi(v) \approx V(1 - v/v_c)^{-1/2}, \tag{8.12}$$

where v_c is the critical value and V is the amplitude. Thus as $\Delta \to 0$ we have

$$v \approx v_c(1 - V^2m^2\Delta^2), \tag{8.13}$$

and so using (8.7) and (8.5) we have the mass renormalization condition

$$\tanh^{-1}[v_c(1 - V^2m^2\Delta^2)]$$
$$= \frac{3}{(2\pi)^d} \int_{-\pi}^{\pi} \cdots \int \frac{d\mathbf{k}}{m_0^2\Delta^2 + 2d - 2\sum_{\tau=1}^{d} \cos k_\tau}, \tag{8.14}$$

which admits a finite continuous solution for $m_0^2\Delta^2$ as a function of $m^2\Delta^2$ as $\Delta \to 0$. The amplitude V and critical value v_c are directly calculable by known Padé approximant techniques[18,31,41] from existing series expansions to a reasonable accuracy. It is to be noted that the scaling properties[44] of the solution [see Eq. (6.9) as an example] are such that a finite, nonzero, two-point Schwinger function results when the scale factor introduced in (8.3) is remembered. It will be observed that the results obtained here are independent of g. If we now consider coupling-constant renormalization, it is plainly convenient to use

$$g' = g\,\Delta^{4-d} \tag{8.15}$$

as the renormalized coupling constant for $d \geq 4$. In this way, if we again think of $m_0^2\Delta^2$ of the order of unity, the representation (8.4) becomes a continuous-spin Ising model. The properties of this model are widely believed[43] to be very similar to the spin-$\frac{1}{2}$ case. We now see that what we have found for g finite, is the limiting behavior as the renormalized coupling constant diverges, and hence, for $d \geq 4$ we find that the strong coupling limit of $g:\phi^4:$ Euclidean Boson quantum field theory is exactly given by the behavior in the high-temperature part of the critical region of the spin-$\frac{1}{2}$ Ising model.

We next discuss the calculation of mass and coupling constant renormalized $g:\phi^4:$ theory in $d \geq 4$. The computational technique is to expand the quantities in (8.2) in a power series in the parameter λ we have introduced. For small, but nonzero λ, it should not be difficult to show that the series converges.[24] Now if we expand (8.4) in powers of λ as

$$Z_- \propto \int_{-\infty}^{+\infty} \cdots \int \prod_{\mathbf{r}} \{d\psi_{\mathbf{r}} \exp[-(\tfrac{1}{2}m_0^2\Delta^2 + d)\psi_{\mathbf{r}}^2$$

$$- g'(\psi_{\mathbf{r}}^4 - 6C\Delta^{d-2}\psi_{\mathbf{r}}^2 + 3C^2\Delta^{2d-4}) + \Delta^{(2-d)/2}h_{\mathbf{r}}\psi_{\mathbf{r}}]\}$$

$$\times\{1 + (\lambda/1!)\sum_{\mathbf{r},\delta} \psi_{\mathbf{r},\delta}$$

$$+ (\lambda^2/2!)\sum_{\mathbf{r},\delta}\sum_{\mathbf{s},\delta'} \psi_{\mathbf{r}}\psi_{\mathbf{r}+\delta}\psi_{\mathbf{s}}\psi_{\mathbf{s}+\delta'} + \cdots, \tag{8.16}$$

then the $\psi_{\mathbf{r}}$'s are all independently distributed. If h is zero, then any term in (8.16) which contains any $\psi_{\mathbf{r}}$ an odd number of times vanishes. The effort of analyzing

our problem is greatly reduced by means of the finite cluster method[41] introduced by Domb[45] and proven by Rushbrooke.[46] The basis for this method is the theorem that if all the spin weight distributions are identical, then

$$\ln Z_-(G)=\sum_\tau T_{G,\tau} f(\tau), \tag{8.17}$$

where the summation is over all unlabeled, connected subgraphs τ of the graph G. Only basic graphs with single lines are included. By $Z_-(G)$ we mean (8.4) with a term $\psi_{\mathbf{r}}\psi_{\mathbf{r}+\delta}$ for every line in G and an integral over $\psi_{\mathbf{r}}$ for each vertex in G. The function $f(\tau)$ is independent of G and depends only on the subgraph τ. The matrix $T_{G,\tau}$ is the number of distinct ways an unlabeled subgraph τ can be (weakly) embeded on G such that every line of τ corresponds to a line of G. This data has been extensively tabulated by Baker *et al.*[47]

If the spin weight distributions are not identical, then (8.17) generalizes to

$$\ln Z_-(G)=\sum_{\tau\subset G} f(\tau), \tag{8.18}$$

where the summation is over every unlabeled connected subgraph τ. Of course (8.17) and (8.18) are tautologies as so far stated. Their content comes in the properties of the $f(\tau)$. If we compute

$$\frac{\partial^n}{\partial h_{\mathbf{r}_1}\cdots\partial h_{\mathbf{r}_n}} \ln Z_-(G)\Big|_{\mathbf{h}=0} = \sum_{\tau\subset G}{}' \frac{\partial^n f(\tau)}{\partial h_{\mathbf{r}_1}\cdots\partial h_{\mathbf{r}_n}}\Big|_{\mathbf{h}=0}, \tag{8.19}$$

then the summation only goes over those τ which contain the points $(\mathbf{r}_1,\ldots,\mathbf{r}_n)$ as the other $f(\tau)$ are independent of those $h_{\mathbf{r}}$'s. If we truncate the expansion (8.16) at the νth term in λ, then to this order all properties of Z_- are given exactly in terms of the embeddings of up to ν line graphs. Consequently, as Rushbrooke's methods involve formal identities, (8.18) must also hold through the νth order term in λ when only graphs of ν lines and less are considered. Therefore we must have

$$f(\tau)=\sum_{n\geq\nu} f_{\tau,n}\lambda^n \tag{8.20}$$

when τ is a ν line graph. Thus we can expand any Ursell function through order λ^ν by considering only ν line graphs. If we denote the set $\{\mathbf{r}_1,\ldots,\mathbf{r}_n\}$ by A, then by (8.19)

$$U_A(G)=\sum_{A\subset\tau\subset G} N(A,\tau;G) f_A(\tau), \tag{8.21}$$

where the summation goes over all τ which contain the set A as vertexes and are embeddable on G. In (8.21) $N(A,\tau;G)$ is the number of ways of embedding on G all unlabeled subgraphs τ with exactly n odd vertexes which lie on the vertex set A of G. The functions $f_A(\tau)$ are independent of G. By repeated use of (8.21) on finite clusters we may successively compute the required $f_A(\tau)$. By considering only subgraphs of up to ν lines we may compute the series expansions thru order λ^ν.

The evaluation of these functions will depend on the parameters of the theory only through the vertex weight functions

$$w_n=\frac{\int_{-\infty}^{+\infty}\psi^n \exp[-(\frac{1}{2}m_0^2\Delta^2+d)\psi^2-g\Delta^{4-d}(\psi^4-6C\Delta^{d-2}\psi^2)]\,d\psi}{\int_{-\infty}^{+\infty}\exp[-(\frac{1}{2}m_0^2\Delta^2+d)\psi^2-g\Delta^{4-d}(\psi^4-6C\Delta^{d-2}\psi^2)]\,d\psi}, \tag{8.22}$$

where n is the number of lines at a vertex. These w_n vanish if n is odd. Thus the restriction to even order vertexes in (8.21) except where the appearance of a ψ from the set A gives an initial odd contribution.

Following Fisher and Burford[11] we introduce for convenience the effective range of correlation

$$\xi^2\approx\Lambda^2=\mu_2/(2d\mu_0), \tag{8.23}$$

where the μ's are defined by (8.2). In general we expect

$$\Lambda\propto(1-\lambda/\lambda_c)^{-\nu}, \tag{8.24}$$

where for $d\geq 4$, $\nu=\frac{1}{2}$, with possible logarithmic corrections for $d=4$. The analysis proceeds as above. We analyze the series expansion for Λ, and determine λ_c and ν as functions of $m_0^2\Delta^2$ for fixed g' and Δ. One expects ν to be independent of $m_0^2\Delta^2$. From the asymptotic expansion near $\lambda=\lambda_c$ we fix $\Lambda(\lambda=1)\approx(m\Delta)^{-1}$, and again solve for $m_0^2\Delta^2$. Having done this we can then explicitly determine the asymptotic behavior of, say $T(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$, Eq. (8.2), as $\lambda\to\lambda_c$, and by substitution obtain the limiting behavior of the scattering amplitude as $\Delta\to 0$.

For $2\leq d<4$, in principal the analysis proceeds as outlined above. The difficulty is that the coefficient of g vanishes as $\Delta\to 0$, and so the critical fluctuations must indeed magnify this vanishing effect to produce a finite result. The hasty conclusion that only a free field theory results does not coincide with the results of perturbation theory in these cases. Even though the coefficients are small if $m_0^2<O(\Delta^{2-d})$, the normal ordered product destabilizes the small momentum states and works to induce cooperative behavior.

The arguments of the renormalization group approach of Wilson[43,48] make it extremely likely that systems of this type have a "universal" behavior shared by the Ising model. Thus we expect[44]

$$(m\Delta)^{-1}\propto(m_0^2-m_{0c}^2)^{-\nu}, \tag{8.25}$$

where $\nu=1$ for $d=2$ and $\nu\approx 0.64$ for $d=3$, to impose the normalization condition for a fixed value of g. The empirical formula of Fisher and Burford[11] for the propagators is, in the limit $\Delta\to 0$,

$$G(\mathbf{p})=[1+a(\mathbf{p})^2/m^2]^{\eta/2}/[m^2+(\mathbf{p})^2], \tag{8.26}$$

where $\eta=0.25$, $d=2$, and $\eta=0.06$, $d=3$, and $a\approx 10^{-4}$ for $d=2$ and 10^{-6} for $d=3$. A direct numerical evaluation of the exact results in two dimensions has been made by Tracy and McCoy.[40] The calculation of higher order correlation functions, i.e., the scattering function T of (8.2), should have their decay properties determined by the closest pole in (8.26), their amplitudes a function of the coupling constant g, and their structure given by the appropriate long-wavelength, critical-point limit of the "universal" Ising model function of the appropriate number of dimensions.

ACKNOWLEDGMENTS

The author is happy to acknowledge many extensive

and helpful discussions with J. L. Gammel and J. Nuttall. He is also happy to acknowledge helpful discussions with R. Baumel, M. E. Fisher, J. Glimm, G. R. Golner, R. B. Griffiths, J. L. Lebowitz, B. Simon, A. S. Wightman, and K. G. Wilson.

APPENDIX A: INEQUALITIES

There are a variety of inequalities that are most useful in this work. We shall not, by and large, derive them, but will catalogue them and indicate the conditions germane to our needs required for them to hold.

We define an unnormalized probability distribution for a system of "spins" by means of the factor

$$P(\mathbf{s}) = \exp\left(\sum_{i<j}^{M} K_{ij} s_i s_j + \sum_{i=1}^{M} h_i s_i\right) \prod_{i=1}^{M} F_i(s_i), \tag{A1}$$

where we impose the restrictions that

$$\begin{aligned} &K_{ij} \geq 0, \quad h_i \geq 0, \\ &F_i(s) = F_i(-s), \\ &0 < \int_{-\infty}^{+\infty} \cdots \int P(\mathbf{s}) \prod_{i=1}^{M} ds_i < \infty. \end{aligned} \tag{A2}$$

In physical terms (A1) and (A2) imply that the energy is lower if $s_k = s_l$ than if $s_k = -s_l$. Therefore, one expects

$$E_A(s_k^{j_k} \cdots s_l^{j_l}) \geq 0, \quad \langle s_k^{j_k} \cdots s_l^{j_l} \rangle \geq 0, \tag{A3}$$

which is Griffith's[49] inequality, and was proved for this general case by Ginibre.[50] The j's are nonnegative integers and E_A is the unnormalized expectation value with respect to the distribution (A1). The brackets $\langle\ \rangle$ denote the normalized expectation value ($\langle 1 \rangle = 1$) with respect to (A1). Again in physical terms, if we increase any of the K_{ij} the correlated state is energetically more favored so the correlation should increase, or

$$\frac{\partial}{\partial K_{ij}} \langle s_k s_l \rangle \geq 0. \tag{A4}$$

This idea leads to the Griffiths,[49] Kelly, Sherman[51] inequalities

$$\langle s_k^{i_k + j_k} \cdots s_l^{i_l + j_l} \rangle \geq \langle s_k^{i_k} \cdots s_l^{i_l} \rangle \langle s_k^{j_k} \cdots s_l^{j_l} \rangle, \tag{A5}$$

where the i and j are nonnegative integers, and the proof for this general form is again due to Ginibre.[50]

There are further inequalities due to Fortuin, Kasteleyn, and Ginibre[52] which hold without the restriction on the h_i imposed in (A2). Suppose $f(s_k, \ldots, s_l)$ and $g(s_k, \ldots, s_l)$ are increasing functions of each argument separately, then

$$\langle fg \rangle \geq \langle f \rangle \langle g \rangle \tag{A6}$$

are the Fortuin, Kasteleyn, Ginibre inequalities.

Recently, Lebowitz[53] has used (A5) and (A6) to prove the following family of inequalities. We have, for this family, in addition to restriction (A1) the further restriction

$$F_i(s) = \tfrac{1}{2}[\delta(s-1) + \delta(s+1)], \tag{A7}$$

that is, $s = \pm 1$ are the only allowed values, and we have a ferromagnetic spin-$\frac{1}{2}$ Ising model. In order to state Lebowitz's inequalities it is convenient to introduce a duplicate set of variables σ with a probability distribution identical to the $\mathbf{s}$, i.e., we use the normalized measure

$$\mu(\boldsymbol{\sigma}, \mathbf{s}) = [P(\boldsymbol{\sigma}) P(\mathbf{s})]/[E_A(1)]^2. \tag{A8}$$

Then, defining

$$\begin{aligned} &q_i = \tfrac{1}{2}(\sigma_i - s_i), \quad t_i = \tfrac{1}{2}(\sigma_i + s_i), \\ &q_A = \prod_{i \in A} q_i, \quad t_A = \prod_{i \in A} t_i, \end{aligned} \tag{A9}$$

we may write the Lebowitz inequalities as

$$\langle q_C \rangle \geq 0, \tag{A10}$$

$$\langle q_C t_D \rangle \leq \langle q_C \rangle \langle t_D \rangle, \tag{A11}$$

$$\langle q_C q_D \rangle \geq \langle q_C \rangle \langle q_D \rangle. \tag{A12}$$

We remark that (A11) has the unusual property of providing an upper bound on a higher-order correlation function in terms of lower-order ones, and is a generalization of the Griffiths, Hurst, Sherman[54] inequality which shares this property.

We may weaken the restriction (A7) very considerably by combining the following simple observation with the method of Simon and Griffiths.[55] The Lebowitz inequalities are linear in every factor q_i and t_i. Thus, if we sum the inequalities over many choices for, say, the first factor, we may replace it by

$$q_i \to \tfrac{1}{2}\left(\sum_{j \in A} \sigma_j - \sum_{j \in A} s_j\right), \tag{A13}$$

and so on for every other factor. Thus (A10)—(A12) hold equally well if we replace the σ_i by sums of σ_i, or equivalently replace condition (A7) by

$$F_k(s) = \sum_{n=-N}^{+N} \left(\sum_{\sigma_i = \pm 1, \Sigma \sigma_i = n} \exp(K_{ij} \sigma_i \sigma_j)\ \delta(s-n)\right). \tag{A14}$$

Now Simon and Griffiths[55] have shown that by a proper choice of the K_{ij} in (A14) we can approximate the distribution

$$F_i(s) = \exp(-a_i s^4 - b_i s^2) \tag{A15}$$

arbitrarily well as far as expected values are concerned, if $a_i > 0$, and b_i is real. This distribution is the generalization we seek to treat $g:\phi^4:_d$ field theory. Thus the Lebowitz inequalities hold as well when only restrictions (A2) and (A15) are assumed.

Finally in this appendix we show that the generalized Lebowitz inequalities imply the cluster property. If

$$s_A = \prod_{i \in A} s_i, \tag{A16}$$

then the cluster property says that the "truncated" correlation function

$$\langle s_A\, s_B \rangle^T = \langle s_A\, s_B \rangle - \langle s_A \rangle \langle s_B \rangle, \tag{A17}$$

which is necessarily positive by (A5), decays to zero when the two-spin correlation function

$$u(A, B) = \max_{\substack{i \in A \\ j \in B}} \langle s_i s_j \rangle \tag{A18}$$

does. First of all we restrict ourselves to zero magnetic field ($h_i = 0$, all i). Otherwise we do not expect

$u(A, B) \to 0$ even though the distance between A and B gets large. In this case

$$\langle s_A s_B\rangle^T = \langle s_A s_B\rangle = 0 \tag{A19}$$

if the number of spins in A plus the number in B is odd. Thus we only need to treat the case where the sum is even. Let us therefore select $\alpha \subset A$, and $\beta \subset B$ and define the sets

$$C = \{\alpha, \beta\}, \quad D = \{d : d \in A \text{ or } d \subset B,\ d \neq \alpha,\ d \neq \beta\}. \tag{A20}$$

Then (A11) can be written out explicitly as

$$\begin{aligned}
&2^n\langle q_C t_D\rangle \\
&= 2 \sum_{\substack{\text{all partitions} \\ a_1 \cup a_2 = a \\ b_1 \cup b_2 = b}} \{\langle s_{b_2}\rangle\langle s_{a_2}\rangle [\langle \sigma_{a_1 \cup b_1 \cup \alpha \cup \beta}\rangle - \langle \sigma_{b_1 \cup \beta}\rangle\langle \sigma_{a_1 \cup \alpha}\rangle] \\
&\quad + \langle \sigma_{a_1 \cup \alpha}\rangle\langle \sigma_{b_1 \cup \beta}\rangle [\langle s_{b_2 \cup a_2}\rangle - \langle s_{b_2}\rangle\langle s_{a_2}\rangle] \\
&\quad - \langle s_{a_2}\rangle\langle s_{b_1 \cup \beta}\rangle [\langle \sigma_{a_1 \cup b_2 \cup \alpha}\rangle - \langle \sigma_{a_1 \cup \alpha}\rangle\langle \sigma_{b_2}\rangle] \\
&\quad - \langle \sigma_{a_1 \cup \alpha}\rangle\langle \sigma_{b_2}\rangle [\langle s_{a_2 \cup b_1 \cup \beta}\rangle - \langle s_{a_2}\rangle\langle s_{b_1 \cup \beta}\rangle] \\
&\quad + [\langle s_{b_2 \cup a_2}\rangle - \langle s_{b_2}\rangle\langle s_{a_2}\rangle][\langle \sigma_{a_1 \cup b_1 \cup \alpha \cup \beta}\rangle \\
&\quad - \langle \sigma_{a_1 \cup \alpha}\rangle\langle \sigma_{b_1 \cup \beta}\rangle] - [\langle s_{a_2 \cup b_1 \cup \beta}\rangle \\
&\quad - \langle s_{a_2}\rangle\langle s_{b_1 \cup \beta}\rangle][\langle \sigma_{a_1 \cup b_2 \cup \alpha}\rangle - \langle \sigma_{a_1 \cup \alpha}\rangle\langle \sigma_{b_2}\rangle]\} \\
&\quad \leq 2\langle \sigma_\alpha \sigma_\beta\rangle\langle t_D\rangle,
\end{aligned} \tag{A21}$$

where $n+2$ is the number of sites in A and B, and where use was made of the formula

$$2^n t_D = \sum_{\substack{\text{all partitions} \\ a_1 \cup a_2 = a \\ b_1 \cup b_2 = b}} \sigma_{a_1}\sigma_{b_1} s_{a_2} s_{b_2}. \tag{A22}$$

When we observe that the first term in the summation (for $a_2 = b_2 = \phi$, the empty set) is exactly $\langle \sigma_A \sigma_B\rangle^T$, then we observe that by transposing all the remaining terms to the right-hand side of the inequality in (A21) we have reduced the cluster property to the sum of terms which either are explicitly proportional to a term of the form $\langle \sigma_\alpha \sigma_\beta\rangle$, or involve the cluster property for a smaller number of sites in the two sets A and B. As we can establish the cluster property directly and obviously for the two-spin case $\langle \sigma_i \sigma_j\rangle^T = \langle \sigma_i \sigma_j\rangle$, as $\langle \sigma_i\rangle = 0$, it follows by induction for any finite size sets A and B. In fact by repeated application of the above reduction process, we can establish

$$\langle \sigma_A \sigma_B\rangle - \langle \sigma_A\rangle\langle \sigma_B\rangle \leq \sum [\prod \langle \sigma_\lambda \sigma_\eta\rangle], \tag{A23}$$

where (i) all spins occur somewhere in every term in the summation, and (ii) in every term in the summation there is at least one factor of the form $\langle \sigma_\alpha \sigma_\beta\rangle$, $\alpha \subset A$, and $\beta \subset B$. It is convenient to write out the first of the inequalities (A23)

$$\begin{aligned}
&\langle \sigma_i \sigma_j \sigma_k \sigma_l\rangle - \langle \sigma_i \sigma_j\rangle\langle \sigma_k \sigma_l\rangle \\
&\quad \leq \langle \sigma_i \sigma_k\rangle\langle \sigma_j \sigma_l\rangle + \langle \sigma_i \sigma_l\rangle\langle \sigma_j \sigma_k\rangle
\end{aligned} \tag{A24}$$

which both illustrates (A23) and is itself useful.

APPENDIX B: GENERALIZED PADE APPROXIMANTS

For the convenience of the reader, we summarize some of the properties of the generalized Padé approximants[18,19] specialized to our case. These approximants have been called Gammel-Baker approximants and, in the special case of interest, the Padé-Borel summation method.

Suppose a function can be represented as a Stieltjes integral

$$g(z) = \int_0^\infty \exp(-zs)\, d\varphi(s), \tag{B1}$$

where $d\varphi \geq 0$. We can form the approximants to $g(z)$ as

$$\begin{aligned}
B_{n,-1}(z) &= \sum_{j=1}^{n} \alpha_j \exp(-z\sigma_j), \\
B_{n\,0}(z) &= \alpha_0 + \sum_{j=1}^{n} \alpha_j \exp(-z\sigma_j),
\end{aligned} \tag{B2}$$

where the α's and σ's (different for the two approximants) are determined by the equations

$$\begin{aligned}
g(z) - B_{n,-1}(z) &= O(z^{2n}), \\
g(z) - B_{n,0}(z) &= O(z^{2n+1}).
\end{aligned} \tag{B3}$$

Then these approximants have the properties

$$B_{n,-1}(z) \leq B_{n+1,-1}(z) \leq g(z) \leq B_{n+1,0}(z) \leq B_{n,0}(z) \tag{B4}$$

for real, nonnegative z. For every such z, the $B_{n,-1}(z)$ converge to a limit function as $n \to \infty$, as is also true of the $B_{n,0}(z)$. These limits need not be the same, but as we implicitly assume finite coefficients in the formal power series for $g(z)$, it follows from Carleman's Theorem (see Sec. 7) that if

$$\sum_p |g_p|^{-1/p} < \infty \tag{B5}$$

for suitably regularized g_p, then there is a unique limit function for both the $B_{n,-1}(z)$ and the $B_{n,0}(z)$. That is to say roughly that if the g_p don't diverge more than $(p!)$ we have a unique limiting sum defined by this method of approximation.

We can relax to some extent the condition that the lower bound in (B1) be zero. Consider

$$\begin{aligned}
h(z) &= \int_{-M}^\infty \exp(-zs)\, d\varphi(s) = \exp(Mz) \int_{-M}^\infty \exp[-z(s+M)] \\
&\quad \times d\varphi(s).
\end{aligned} \tag{B6}$$

We can apply the lower bound (B4) to the right-most integral as it is of form (B1), and thus using (B2) and (B4)

$$\begin{aligned}
h(z) &\geq \exp(Mz) \sum_{j=0}^{n} \alpha_j \exp(-z\sigma_j) \\
&= \sum_{j=0}^{n} \alpha_j \exp[-z(\sigma_j - M)] = B_{n,-1}(z),
\end{aligned} \tag{B7}$$

where the last equality follows by the uniqueness of the approximant. It will be noted that the upper bound is not form invariant and does not continue if we extend the

limit of integration. We may, for the lower bound, easily take $M=\infty$, by a limiting process.

APPENDIX C: CONTINUITY OF m IN $g:\phi^4:$ THEORY

In this appendix we take advantage of the fact that the Lebowitz inequalities hold for $g:\phi^4:$ theory to prove that the infinite volume limit $m(m_0^2)$ is continuous. This property, together with the results of Sec. 6 establish the mass renormalizability of the theory. When the system size L is finite, $m(L)$ is automatically continuous, as the minimum in the defining equation (6.14) is assumed for some $\mathbf{r}$ and s, and this term is not only continuous, but even differentiable. The plan of the proof is to supply an upper bound, uniform in L, for the derivative of m with respect to m_0^2. From (6.14) we compute, if $\mathbf{r}$, s is the minimum term,

$$\left[1+\frac{\alpha m^{\alpha-1}|\mathbf{r}-\mathbf{s}|^{(d-1)/2}}{A+m^{\alpha}|\mathbf{r}-\mathbf{s}|^{(d-1)/2}}\right]\frac{\partial m(L)}{\partial m_0^2}$$

$$=\Delta^d\sum_{\mathbf{t}}\frac{\langle\phi_{\mathbf{r}}\phi_{\mathbf{s}}\phi_{\mathbf{t}}^2\rangle-\langle\phi_{\mathbf{r}}\phi_{\mathbf{s}}\rangle\langle\phi_{\mathbf{t}}^2\rangle}{\langle\phi_{\mathbf{r}}\phi_{\mathbf{s}}\rangle\,|\mathbf{r}-\mathbf{s}|}. \tag{C1}$$

Since, by (5.1), the interaction gives a field distribution function for which we may apply the inequalities (A24), we may bound Eq. (C1) by

$$0\le\frac{\partial m(L)}{\partial m_0^2}\le 2\Delta^d\sum_{\mathbf{t}}\frac{\langle\phi_{\mathbf{r}}\phi_{\mathbf{t}}\rangle\langle\phi_{\mathbf{s}}\phi_{\mathbf{t}}\rangle}{\langle\phi_{\mathbf{r}}\phi_{\mathbf{s}}\rangle\,|\mathbf{r}-\mathbf{s}|}. \tag{C2}$$

Now by definition (6.14)

$$\langle\phi_{\mathbf{r}}\phi_{\mathbf{s}}\rangle=\frac{\exp[-m(L)]\,|\mathbf{r}-\mathbf{s}|}{A+m^{\alpha}|\mathbf{r}-\mathbf{s}|^{(d-1)/2}}, \tag{C3}$$

and for all other pairs

$$\langle\phi_{\mathbf{u}}\phi_{\mathbf{v}}\rangle\le\frac{\exp[-m(L)\,|\mathbf{u}-\mathbf{v}|]}{A+m^{\alpha}|\mathbf{u}-\mathbf{v}|^{(d-1)/2}}. \tag{C4}$$

Thus by using (C2)–(C4) we obtain the bound

$$0\le\frac{\partial m(L)}{\partial m_0^2}\le 2\Delta^d\sum_{\mathbf{t}}\frac{(A+m^{\alpha}|\mathbf{r}-\mathbf{s}|^{(d-1)/2})\exp[m(L)(|\mathbf{r}-\mathbf{s}|-|\mathbf{r}-\mathbf{t}|-|\mathbf{s}-\mathbf{t}|)]}{(A+m^{\alpha}|\mathbf{r}-\mathbf{t}|^{(d-1)/2})(A+m^{\alpha}|\mathbf{s}-\mathbf{t}|^{(d-1)/2})|\mathbf{r}-\mathbf{s}|}. \tag{C5}$$

If for the minimum term $|\mathbf{r}-\mathbf{s}|$ is not large, then (C5) gives an immediate finite bound. If, as is likely, $|\mathbf{r}-\mathbf{s}|$ tends to infinity as L does, then a little calculation is required. Let us consider the hyperellipsoids on which the argument of the exponential in Eq. (C5) is constant. They have their focii at $\mathbf{r}$ and s. If we parametrize them by ζ, the length of the semiminor axis, then the semimajor axis is

$$(\zeta^2+\tfrac{1}{4}|\mathbf{r}-\mathbf{s}|^2)^{1/2}\approx\tfrac{1}{2}|\mathbf{r}-\mathbf{s}|+\zeta^2/|\mathbf{r}-\mathbf{s}|+, \tag{C6}$$

where the right-hand side is valid if $\zeta\ll|\mathbf{r}-\mathbf{s}|$. In the limit of large but finite $|\mathbf{r}-\mathbf{s}|$, if we replace the summation in (C5) by integration, and integrate first along the direction of the major axis of the hyperellipsoid, we may reduce the dominate term in (C5) to be proportioned to

$$m^{-\alpha}\int_0^{\infty}\zeta^{d-2}\exp\left[-2m(L\zeta^2/|\mathbf{r}-\mathbf{s}|)\right]\frac{d\zeta}{|\mathbf{r}-\mathbf{s}|^{(d-1)/2}}, \tag{C7}$$

where the constant of proportionality is dimension dependent, and use has been made of the decrease in volume of a hypercylindrical slice near the ends of the hyperellipsoid to offset the decrease in the denominator in (C5). The form (C7) is correct for $d\ge 2$; for $d=1$ a separate calculation leads to the desired result. For all higher values of d, (C7) is correct. We conclude, therefore, that, integrating (C7),

$$0\le\frac{\partial m(L)}{\partial m_0^2}\le\Gamma_d/[m(L)]^{\alpha+(d-1)/2}, \tag{C8}$$

where Γ_d is independent of L and finite. Thus

$$0\le\frac{\partial}{\partial m_0^2}\,[m(L)]^{\alpha+(d+1)/2}\le\Gamma_d,\ \alpha>0,$$

is uniformly bounded for all L, and hence $[m(L)]^{\alpha+(d+1)/2}$ and [therefore $m(L)$] is a continuous function of m_0^2.

APPENDIX D: NELSON'S RECONSTRUCTION THEOREMS

In this appendix we will discuss briefly a theorem of Nelson[3] on the construction of quantum field theory from Markov field theory, and show the relation of his result to our case. In particular we will confine our discussion to the $g:\phi^4:_d$ theory where we have fuller results. Nelson[6] has already treated the $g:\phi^4:_d$ theory in his discussion of the free Markov field. Osterwalder and Schrader[7] have given a set of Euclidean axioms which imply the existence of a quantum field theory in the case $d=4$; their axioms can be verified for $g:\phi^4:_4$ theory. The connection with relativistic quantum field theory in Minkowski space is made directly by analytic continuation on the time variable from real in the Euclidean-space framework to pure imaginary in the Minkowski-space framework. By a Minkowski space we mean a Euclidean space where we replace the usual vector product

$$\mathbf{x}\cdot\mathbf{y}=\sum_{n=1}^{d}x_n y_n$$

by the Lorentz invariant formula

$$(x,y)=x_d y_d-\sum_{n=1}^{d-1}x_n y_n.$$

The two products can be related to each other by letting the dth component become pure imaginary and changing the sign. We distinguish timelike vectors by $(x,x)>0$ and spacelike vectors by $(x,x)<0$. There are two separate parts to the set of timelike vectors which may be characterized as $x_d>0$ and $x_d<0$. The idea of causality tells us that unless a point lies in the forward light cone [closure of all $x_d>0$, $(x,x)>0$] of another

point, its behavior is unaffected by that of the other point. This principle is reflected by the vanishing of the correlation functions in the appropriate regions.

One must then establish the necessary properties, that is the Wightman axioms,[56] of the analytic continuation of the Schwinger functions. The two properties which are the hardest to establish for the limit of the Euclidean lattice theory are the rotational invariance and the vanishing outside the forward light cone of the Wightman functions (analytic continuation of the Schwinger functions).

In order to establish these results, Nelson makes use of the transfer matrix which adds a $(d-1)$-dimensional hyperplane to the system. For clarity, we write, in the notation of Sec. 2 and (5.1), for finite L and Δ,

$$\underset{\sim}{T}_0 = \exp\Bigg(-\tfrac{1}{2}\Delta^d \sum_{\mathbf{r}}{}' \Bigg\{\sum_{\{\delta\}} [(\phi_{\mathbf{r}+\delta} - \phi_{\mathbf{r}})^2\Delta^{-2}] + \tfrac{1}{2}m_0^2(\phi_{\mathbf{r}}^2 + \phi_{\mathbf{r}+\delta'}^2) + g[P(\phi_{\mathbf{r}}) + P(\phi_{\mathbf{r}+\delta'})]\Bigg\}\Bigg), \tag{D1}$$

where the $\sum_{\mathbf{r}}'$ is over a hyperplane, and δ' is perpendicular to that hyperplane. The normalized transfer matrix is

$$\mathbf{T} = \mathbf{T}_0/\mathrm{spr}(\mathbf{T}_0) \tag{D2}$$

where the spectral radius of an operator is defined as (independent of the norm)

$$\mathrm{spr}(\mathbf{A}) = \lim_{n\to\infty} \|\mathbf{A}^n\|^{1/n}. \tag{D3}$$

However, for free and Dirichelet boundary conditions, the limit in Eq. (D3) is substantially the limit we have shown to exist (as $\Delta > 0$) in Eq. (4.33). The difference is that in (D3) we let $L \to \infty$ in only one dimension, but the proof also works for this case. By construction $\mathbf{T}$ is Hermitian and positivity perserving, that is, if u is everywhere positive, the $\mathbf{T}u$ is also. Further, by construction, $\mathrm{spr}(\mathbf{T}) = 1$, so that every eigenvalue is real, nonnegative, and less than or equal to unity; therefore there exists a unique, Hermitian, positive operator $\mathbf{H}$ such that

$$\mathbf{T} = \exp(-\Delta\mathbf{H}). \tag{D4}$$

We may use this representation to re-express for Dirichlet boundary conditions the Schwinger functions which we had shown at (4.41) to exist. Since the order of the $\mathbf{r}_1$ is immaterial to the definition of the Schwinger functions, we can, without loss of generality, order them by their time, or dth, component. Then, if τ_i is the time component of $\mathbf{r}_i$, and ρ_i is $\mathbf{r}_i$ with the time component set to τ_1, we may write

$$\mathcal{S}_-(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \lim_{L\to\infty} \langle \phi_{\rho_1} \exp[-(\tau_2 - \tau_1)\mathbf{H}]\phi_{\rho_2} \cdots \times \exp[-(\tau_n - \tau_{n-1})\mathbf{H}]\phi_{\rho_n}\rangle_-, \tag{D5}$$

where $\langle\ \rangle_-$ is the normalized expectation over a system of size L with Dirichlet boundary conditions. The analytic continuation in form (D5) follows because when the τ_i become pure imaginary [approached from $\mathrm{Re}(\tau_i - \tau_{i-1}) > 0$], by the properties of $\mathbf{H}$

$$|\exp[-(\tau_i - \tau_{i-1})\mathbf{H}]| = 1,$$

and so as $\mathcal{S}_-(\rho_1, \ldots, \rho_n)$ has previously been proven to exist in (4.41), so also can its analytic continuation to imaginary time be established. Thus the Wightman functions $\mathbf{r}_1 \to (\rho\ ,\ i\tau_i)$ are defined for the theory.

As we have shown in Sec. 7 for $g:\phi^4$: theory, the Schwinger functions defined by periodic boundary conditions and Dirichlet boundary conditions are the same. Therefore, since it is an exact property for periodic boundary conditions, translational invariance in multiples of the lattice spacing holds. Consequently, we may write

$$\mathcal{S}_-(\mathbf{r}_1, \ldots, \mathbf{r}_n) = S_-(\mathbf{r}_1 - \mathbf{r}_2, \ldots, \mathbf{r}_{n-1} - \mathbf{r}_n). \tag{D6}$$

If we complete the Fourier transform

$$\begin{aligned}\tilde{S}_-(\mathbf{p}_1, \ldots, \mathbf{p}_n) &= \int \cdots \int d\mathbf{r}_1 \cdots d\mathbf{r}_n \times \exp\Big(-i\sum_{j=1}^n \mathbf{p}_j \cdot \mathbf{r}_j\Big)\, S_-(\mathbf{r}_1 - \mathbf{r}_2, \ldots, \mathbf{r}_{n-1} - \mathbf{r}_n) \\ &= \int \cdots \int d\mathbf{r}_1 \cdots d\mathbf{r}_n \times \exp\Big\{-i\Big[\mathbf{p}_1 \cdot \times(\mathbf{r}_1 - \mathbf{r}_2) + (\mathbf{p}_1 + \mathbf{p}_2)\cdot(\mathbf{r}_2 - \mathbf{r}_3)\cdots + \sum_{j=1}^n \mathbf{p}_j \cdot \mathbf{r}_n\Big]\Big\} S_-(\mathbf{r}_1 - \mathbf{r}_2, \ldots, \mathbf{r}_{n-1} - \mathbf{r}_n) \\ &= (2\pi)^d \delta\Big(\sum_{j=1}^n \mathbf{p}_j\Big) \times \tilde{S}_-(\mathbf{p}_1, \mathbf{p}_1 + \mathbf{p}_2, \ldots, \mathbf{p}_1 + \cdots + \mathbf{p}_{n-1}).\end{aligned} \tag{D7}$$

Now since $\mathbf{H}$ is a positive operator, the Fourier transform of $\langle v, \exp(i\tau\mathbf{H})u\rangle$ vanishes on the negative half-axis. In particular, the Fourier transform of

$$\langle \phi_{\rho_1} \cdots \phi_{\rho_j} \exp(i\tau\mathbf{H})\, \phi_{\rho_{j+1}} \cdots \phi_{\rho_n}\rangle \tag{D8}$$

vanishes on the negative half-axis so that $\tilde{S}_-(\mathbf{q}_1, \ldots, \mathbf{q}_j, \ldots, \mathbf{q}_{n-1}) = 0$ if there is an imaginary time component for the jth argument. As j is arbitrary, this result holds for all j's.

As we discussed in Sec. 7, rotational invariance is an unproven consequence of the approach to the critical point implied by mass renormalization and the limiting process $\Delta \to 0$. In so far as analytic continuation in the coupling constant g will carry us, this invariance is a rigorous result as it holds term-by-term for the limiting coefficients in the perturbation series. This invariance amounts to the result that rotations in d-dimensional Euclidean space leave the Schwinger functions invariant. Since the inner product for Minkowski space is the analytic continuation of that for Euclidean sapce, Lorentz invariance without time reversal follows by analytic continuation from rotational invariance, since we have already established the existence of the continuation.

By the Lorentz invariance and the argument at (D8) $\tilde{S}_-(\mathbf{q}_1, \ldots, \mathbf{q}_{n-1}) = 0$, unless each $\mathbf{q}_j$ is in the forward light cone. Thus we have the Wightman properties[56,57] of (a) relativistic invariance and (b) the spectral condition. If we let $W(x_1, \ldots, x_n)$, with x_1 in Minkowski space, be the Wightman functions defined above as the analytic continuation of (D5), then (c) the Hermitian property

$$W(x_1, \ldots, x_n) = W^*(x_n, \ldots, x_1) \tag{D9}$$

follows easily from the definitions and random variable nature of the ϕ_p. (The * means complex conjugate.) Also (d) the local commutativity property

$$W(x_1, \ldots, x_j, x_{j+1}, \ldots, x_n) = W(x_1, \ldots, x_{j+1}, x_j, \ldots, x_n), \tag{D10}$$

when x_j and x_{j+1} have spacelike separation, follows by using Lorentz invariance to put them in the same hyperplane, and by using the commutativity of the random variables ϕ_ρ. The same procedure gives (e) the cluster property for space-like separations directly from the Eucledian cluster property of Section 7. Finally (f), the positivity property,

$$\sum_{j=0}^{n}\sum_{k=0}^{n}\int\cdots\int dx_1\cdots dx_j dy_1\cdots dy_k f_j^*(x_1, \ldots, x_j) \times W(x_j, \ldots, x_1, y_1, \ldots, y_k) f_k(y_1, \ldots, y_k) \geq 0, \tag{D11}$$

follows by the properties of W in terms of expectation values.

APPENDIX E: COMPLEX COUPLING CONSTANT

In this appendix we use the exponential decay properties of the free-field correlation functions to bound the Schwinger functions for small complex values of the coupling constant in the closed right half-plane, $\mathrm{Re}(g) \geq 0$. Similar results have previously been obtained by Glimm *et al.*[58] and Eckmann *et al.*[35] for $:P(\phi)_2:$ theories. Although their arguments can be applied to the present case, the transcription is quite long and not completely transparent. Basically, however, their proof is dimension independent except that certain combinatorial estimates must be changed, but these changes do not affect the validity of the proof. Also, all the special properties of the interaction in two-dimensional space-time, which they used, hold in any number of dimensions, provided $\Delta > 0$.

We present a different argument. Consider the Schwinger function

$$S_B = \frac{\int_{-\infty}^{+\infty}\int (\prod_{\mathbf{r}\in B}\phi_\mathbf{r})\exp(-A_-)\prod_{\mathbf{j}=0}^{L-\Delta}\{\exp[-g(P(\phi_\mathbf{j})+K)]\,d\phi_\mathbf{j}\}}{\int_{-\infty}^{+\infty}\int \exp(-A_-)\prod_{\mathbf{j}=0}^{L-\Delta}\{\exp[-g(P(\phi_\mathbf{j})+K)]\,d\phi_\mathbf{j}\}} \tag{E1}$$

in the notation of (2.9), (5.1), and (5.2). If we write

$$\exp[-gP(\phi_\mathbf{j})] = 1 + t_\mathbf{j}\{\exp[-g(P(\phi_\mathbf{j})+K)] - 1\} = 1 + t_\mathbf{j} f_\mathbf{j} \tag{E2}$$

for $t_\mathbf{j} = 1$, then

$$0 \leq |f_\mathbf{j}| \leq 1, \quad \mathrm{Re}(g) \geq 0 \tag{E3}$$

by the definition of K. Also, for a fixed value of $\phi_\mathbf{j}$, $f_\mathbf{j} \to 0$ as $g \to 0$, so for g small, the $f_\mathbf{j}$ are in a sense small. With this notation we may rewrite Eq. (E1) as

$$S_B = \frac{\int_{-\infty}^{+\infty}\int (\prod_{\mathbf{r}\in B}\phi_\mathbf{r})\exp(-A_-)\prod_{\mathbf{j}=0}^{L-\Delta}[(1+t_\mathbf{j}f_\mathbf{j})\,d\phi_\mathbf{j}]}{\int_{-\infty}^{+\infty}\int \exp(-A_-)\prod_{\mathbf{j}=0}^{L-\Delta}[(1+t_\mathbf{j}f_\mathbf{j})\,d\phi_\mathbf{j}]}. \tag{E4}$$

Let us now consider moving from the free-field case ($t_\mathbf{j} = 0$) to the $P(\phi)$ case ($t_\mathbf{j} = 1$) by changing one $t_\mathbf{j}$ at a time from 0 to 1. If we change any one, we may write

$$S_B(t_\mathbf{k} = 1,\ t_\mathbf{j} = 0,\ \mathbf{j} \neq \mathbf{k}) = \langle \prod_{\mathbf{r}\in B}\phi_\mathbf{r}\rangle_- + \frac{\langle f_\mathbf{k}\prod_{\mathbf{r}\in B}\phi_\mathbf{r}\rangle_- - \langle f_\mathbf{k}\rangle_-\langle\prod_{\mathbf{r}\in B}\phi_\mathbf{r}\rangle_-}{1+\langle f_\mathbf{k}\rangle_-} \tag{E5}$$

where $\langle\ \rangle_-$ is the free-field expectation value with Dirichlet boundary conditions. The second, or correction term, in (E5) can be small for two reasons. First, by the exponential decoupling of the free field, if k is far from the set B, then a factor $\exp(-m_0|\mathbf{k}-B|)$ occurs where $|\mathbf{k}-B|$ means the distance between k and the closest point of B. Secondly, as we can make $|f_\mathbf{k}|$ as small as we please by selecting $|g| < \omega$ and $\mathrm{Re}(g) \geq 0$ with probability $1-\epsilon$, and since it is bounded by 1 everywhere, we can control $\langle f_\mathbf{k}\rangle_-$, etc. In order to compute the change in S_B from the free-field case, we apply (E5) repeatedly. Clearly, by the exponential decoupling the sum of all first-order corrections in $f_\mathbf{j}$ behaves like

$$G = \int k^\sigma\, dk \exp(-m_0|\mathbf{k}-B|), \tag{E6}$$

where σ is dimension dependent. The sum of the second-order corrections will have the structure[59]

$$\frac{1}{2!}\int k^\sigma\, dk \int j^\sigma\, dj \exp(-m_0|\mathbf{k}-B|) \times (\exp(-m_0|\mathbf{j}-B|) + \exp(-m_0|\mathbf{j}-\mathbf{k}|)) \approx G^2 \tag{E7}$$

and so on for the higher-order corrections. The only remaining problem is to obtain a lower bound on the denominators

$$\left|\prod_{\mathbf{j}\in D}(1+f_\mathbf{j})_-\right|. \tag{E8}$$

This may be done by considering

$$\left\langle\prod_{\mathbf{j}\in D}(1-|f_\mathbf{j}|)\right\rangle_- \tag{E9}$$

and by discarding the "gradient" term so that we may write the decoupled bound, as

$$\left|\left\langle\prod_{\mathbf{j}\in D}(1+f_\mathbf{j})\right\rangle_-\right| \geq \prod_{\mathbf{j}\in D}\left(1 - \frac{\int_{-\infty}^{+\infty}|f_\mathbf{j}|\exp(-\frac{1}{2}m_0^2\Delta^d\phi_\mathbf{j}^2)\,d\phi_\mathbf{j}}{F^{\Delta^d}}\right), \tag{E10}$$

where F is the normalizing factor per unit hypervolume given in (2.14). Therefore, repeated application of Eq. (E5) generates a series of the structure

$$\sum[G(\langle f\rangle_- M)]^n, \tag{E11}$$

which converges provided $[0 \leq |g| \leq \omega,\ \mathrm{Re}(g) \geq 0]$ ω is small enough. Note that the factors of $n!$ in the numerator from different terms as in (E7) and $n!$ in the denominator from the overcounting of permutations cancel.

We have not tried to put in the many steps to make the above discussion rigorous, but only to sketch the general pattern.

*Work performed under the auspices of the U.S. Atomic Energy Commission.
†Address to which reprint requests are to be sent.
[1]J. Glimm and A. Jaffe, in *Statistical Mechanics and Quantum Field Theory*, *Les Houches* 1970, edited by C. DeWitt and R. Stora (Gordon and Breach, New York, 1971).
[2]J. Dimock, Comm. Math. Phys. **35**, 347 (1974).
[3]E. Nelson, J. Funct. Anal. **12**, 97 (1973).
[4]J. Schwinger, Proc. Natl. Acad. Sci. U.S. **44**, 956 (1958); Phys. Rev. **115**, 721 (1959).
[5]K. Symanzik, J. Math. Phys. **7**, 510 (1966).
[6]E. Nelson, J. Funct. Anal. **12**, 211 (1973).
[7]K. Osterwalder and R. Schrader, Comm. Math. Phys. **31**, 83 (1973).
[8]M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience, New York, 1959).
[9]K. Osterwalder and R. Schrader, Phys. Rev. Lett. **29**, 1423 (1972), and Harvard University preprint.
[10]F. Guerra, L. Rosen, and B. Simon, Ann. Math. **101**,111, 191 (1975).
[11]M.E. Fisher and R.J. Burford, Phys. Rev. **156**, 583 (1967).
[12]E. Nelson, *Mathematical Theory of Elementary Particles*, edited by R. Goodman and J. Segel (M.I.T., Cambridge, Mass., 1966), pp. 69—73.
[13]J. Glimm, Comm. Math. Phys. 8, 12 (1968).
[14]Ref. 1, Lemma 2.15.
[15]N.N. Bogoliubov and D.V. Shirkov, *Introduction to the Theory of Quantized Fields*, transl. by G.M. Volkoff (Interscience, New York, 1959).
[16]E.F. Beckenbach and R. Bellman, *Inequalities* (Springer, New York, 1965).
[17]For results in $:P(\phi):_2$ theory see the work of F. Guerra, L. Rosen, and B. Simon, Bull. of Am. Math. Soc. **80**, 1205 (1974).
[18]G.A. Baker Jr., *Essential of Padé Approximants* (Academic, New York, 1975).
[19]G.A. Baker Jr., Phys. Rev. **161**, 434 (1967).
[20]M. Villani, *Cargèse Lectures in Physics in Physics*, edited by D. Bessis (Gordon and Breach, New York, 1972), Vol. 5, pp. 461—474.
[21]G.L. Folgi, M.F. Pelicoro, and M. Villani, Nuovo Cimento A **6**, 79 (1971)
[22]G.A. Baker Jr., J. Math. Phys. **13**, 1862 (1972).
[23]J.L. Lebowitz and O. Penrose, Comm. Math. Phys. **11**, 99 (1968).
[24]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[25]See the references detailed in the introduction by S. Katsura, T. Morita, S. Inawashiro, T. Horiguchi, and Y. Abe, J. Math. Phys. **12**, 892 (1971).
[26]E.C. Titchmarsh, *The Theory of Functions* (Oxford, London, 1952).
[27]T.H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).
[28]G.S. Joyce, *Phase Transitions and Critical Phenomena*, edited by C. Domb and M.S. Green (Academic, New York, 1972), Vol. 2, pp. 375—440.
[29]For results in two dimensions, see Sec. V. 4 of Ref. 10.
[30]M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graph, and Mathematical Tables*, Nat. Bur. of Standards, A.M.S. 55 (U.S. Govt. Print Office, Washington, D.C., 1964).
[31]M.E. Fisher and D.S. Gaunt, Phys. Rev. **133**, A224 (1964).
[32]L. Onsager, Phys. Rev. **65**, 117 (1944).
[33]G.A. Baker Jr., Phys. Rev. **130**, 1406 (1963).
[34]See, B.M. McCoy and T.T. Wu, *The Two-Dimensional Ising Model* (Harvard U.P., Cambridge, Mass., 1973).
[35]B. Simon, Phys. Rev. Lett. **25**, 1583 (1970); J.P. Eckmann, J. Magnen, and R. Seneor, "Decay Properties and Borel Summability for the Schwinger Functions in $P(\phi)_2$-Theories" (preprint).
[36]J. Glimm and A. Jaffe, Phys. Rev. Lett. **33**, 440 (1974).
[37]By contraction on the photon lines with $g=\lambda^2$, we have the enumeration from the work of C.A. Hurst, Proc. Roy. Soc. A **214**, 44 (1952).
[39]G.H. Hardy, *Divergent Series* (Oxford, London, 1956).
[40]C.A. Tracy and B.M. McCoy, Phys. Rev. Lett. **31**, 1500 (1973).
[41]See C. Domb and M.S. Green, eds., *Phase Transitions and Critical Phenomena* (Academic, New York, 1974), Vol. 3.
[42]G.A. Baker, Jr., Phys. Rev. B **9**, 4908 (1974).
[43]K.G. Wilson and J. Kogut, Physics Reports **12C**, 75 (1974).
[44]M.E. Fisher, Reports Prog. Phys. **30**, 615 (1967).
[45]C. Domb, Advan. Phys. **9**, Nos. 34, 35 (1960).
[46]G.S. Rushbrooke, J. Math. Phys. **8**, 1106 (1964).
[47]G.A. Baker Jr., H.E. Gilbert, J. Eve, and G.S. Rushbrooke, *A Data Compendium of Linear Graphs with Application to the Heisenberg Model*, Brookhaven National Lab. Report 50053 (*T*-460) (1967).
[48]K.G. Wilson, Phys. Rev. Lett. **28**, 548 (1972).
[49]R.B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).
[50]J. Ginibre, Comm. Math. Phys. **16**, 310 (1970).
[51]D.G. Kelly and S. Sherman, J. Math. Phys. **9**, 446 (1968).
[52]C.M. Fortuin, P.W. Kasteleyn, and J. Ginibre, Comm. Math. Phys. **22**, 89 (1971).
[53]J.L. Lebowitz, Comm. Math. Phys. **35**, 87 (1974).
[54]R.B. Griffiths, C.A. Hurst, and S. Sherman, J. Math. Phys. **11**, 790 (1970).
[55]B. Simon and R.B. Griffiths, Comm. Math. Phys. **33**, 145 (1973); Phys. Rev. Lett. **30**, 931 (1973).
[56]R.F. Streater and A.S. Wightman, *PCT, Spin, Statistics, and All That* (Benjamin, New York, 1965).
[57]A.S. Wightman, Phys. Rev. **101**, 860 (1956).
[58]J. Glimm, A. Jaffe, and T. Spencer in *Lecture Notes in Physics, Vol. 25 Constructive Quantum Field Theory* (Erice, 1973), edited by G. Velo and A. Wightman (Springer, New York, 1973), pp. 132—242.
[59]By the special properties of the free field correlation functions, the Ursell functions vanish for more than two points. Then the structure follows from (5.31). See also, M. Duneau, D. Iagolnitzer, and B. Souillard, Comm. Math. Phys. **31**, 191 (1973) and **35**, 307 (1974), for related discussions of strong cluster properties.

Minimal coupling and complex line bundles

Werner Greub

Department of Mathematics, University of Toronto, Toronto, Canada

Herbert-Rainer Petry

Institut für Theoretische Kernphysik, University of Bonn, Bonn, Germany
(Received 2 August 1974)

The concept of minimal coupling, which leads to the Schrödinger equation of a particle in an external electromagnetic field, is reformulated within the theory of complex line bundles. The possible generalizations are discussed, and the case of the magnetic monopole is investigated with the help of the new formalism.

INTRODUCTION

It seems that Dirac was the first who studied the Schrödinger equation for a particle in an electromagnetic field, which could not be described globally by suitable vector potentials. He considered the magnetic monopole and found that a quantum mechanical description could only be given, if the pole strength was quantized.

In this paper we want to show that if the usual concept of gauge invariant or minimal coupling is formulated within the theory of complex line bundles, the conditions under which a well-defined Schrödinger equation exists, can be completely classified. As an example we will use the mathematical techniques provided by bundle theory to reinvestigate the magnetic monopole.

1. MINIMAL COUPLING

Assume we are given an electromagnetic field described by the field tensor F which satisfies Maxwell's equations and hence is a closed two-form. In general, F will have certain singularities; that is, it is only defined on a subset D_F of R^4, which we assume to be open (and hence a smooth manifold). For this reason we cannot expect F to be exact, i.e., of the form $F=dA$ where A, the vector potential, is a globally defined one-form in D_F (d denotes the exterior derivative).

Next, recall how the quantum-mechanical coupling of the motion of a charged particle to the external field F is achieved. Consider the free Schrödinger equation (∂_0 denotes derivative with respect to the time, ∂_i with respect to Cartesian space coordinates)

$$i\partial_0\Psi = -\frac{1}{2M}\sum_{i=1}^{3}\partial_i^2\Psi \quad (\hbar = c = 1,\ M \text{ particle mass}).$$

(We could as well use the Dirac or Klein–Gordon equation.) Replacing ∂_μ by

$$\nabla_\mu =: \partial_\mu + ieA_\mu \quad (\mu = 0,1,2,3)\ (e \text{ particle charge}), \tag{1}$$

we obtain the equation

$$i\nabla_0\Psi = -\frac{1}{2M}\sum_{i=1}^{3}\nabla_i^2\Psi \tag{2}$$

which is gauge-invariant; i.e., it does not change under the substitutions

$$\Psi \to \Psi\exp(-ie\lambda), \quad A_\mu \to A_\mu + \partial_\mu\lambda$$

for arbitrary real functions λ. The particular way in which the vector potential appears here (and in other comparable equations like Dirac's, etc.) is known as minimal coupling.

Equation (2), which now describes the motion in the external field F, should be well defined at least in the region D_F, where F does not contain a singularity. But in this case one must obviously insist on the global existence of the vector potential A on D_F, i.e., F is exact. However, we will show that if the mathematical interpretation of Eq. (2) is slightly changed, it remains meaningful even for a larger set of two-forms F, namely those for which $eF/2\pi$ represents an integral cohomology class.

By virtue of the ideas of geometric quantization (see Refs. 1–6) the new interpretation consists in regarding Ψ as a cross section and ∇ as the covariant derivative in a complex line bundle L over D_F with curvature ieF. (∇_μ then denotes the covariant derivative in the direction of the vector field ∂_μ.) The mathematical definition of these notions is the subject of the next section.

2. COMPLEX LINE BUNDLES

A complex line bundle is a collection $L=(E,\pi,M)$, where E and M are smooth manifolds and $\pi: E\to M$ is a smooth map satisfying the following conditions:

(i) For every $x\in M$ the set $F_x=\pi^{-1}(x)$ (the *fibre* over x) is a one-dimensional complex vector space.

(ii) There exists a covering of M by open sets U_α and a family of diffeomorphisms $\Psi_\alpha: U_\alpha\times\mathbb{C}\xrightarrow{\cong}\pi^{-1}(U_\alpha)$, which restrict for every $x\in U_\alpha$ to a (complex) linear isomorphism

$$\Psi_{\alpha x}: \mathbb{C}\xrightarrow{\cong} F_x.$$

Every such family (Ψ_α, U_α) is called a *trivialization* of the complex line bundle L.

The simplest example of a complex line bundle over a manifold M is the product $M\times\mathbb{C}$ where $\pi: M\times\mathbb{C}\to M$ is the obvious projection. This is called the *trivial bundle*.

A *cross-section* in a complex line bundle is a smooth map $\sigma: M\to E$ such that $\pi\circ\sigma(x)=x$, i.e., σ carries every point $x\in M$ into a vector of its fibre. (Therefore, one can define addition of sections and multiplication of

sections with functions on M.)

A *Hermitian metric* is a map $\langle\ ,\ \rangle_L$ which assigns to every pair of cross sections σ_1, σ_2 a function $\langle \sigma_1, \sigma_2\rangle_L$ linear in σ_2, antilinear in σ_1, with $\langle\sigma_1, \sigma_2\rangle_L = \overline{\langle\sigma_2, \sigma_1\rangle}_L$ and such that

$$\langle\varphi\sigma_1, \sigma_2\rangle_L = \bar{\varphi}\langle\sigma_1, \sigma_2\rangle_L$$

for arbitrary functions φ on M. $\langle\ ,\ \rangle_L$ is called *positive definite* if $\langle\sigma, \sigma\rangle_L \geqq 0$ and $\langle\sigma, \sigma\rangle_L = 0$ implies that $\sigma = 0$.

A *linear connection* in a complex line-bundle L is an operator ∇ which assigns to every pair σ, X (σ a cross section, X a vector field on M) another cross section $\nabla_X\sigma$, called the *covariant derivative of σ with respect to X*, which is linear in X and σ and satisfies for arbitrary functions φ the conditions

$$\nabla_{\varphi X}\sigma = \varphi\,\nabla_X\sigma,$$
$$\nabla_X\,\varphi\sigma = \varphi\nabla_X\sigma + X(\varphi)\cdot\sigma.$$

[$(X(\varphi))$ denotes the derivative of φ in the direction of X.]

Let ∇ be a linear connection in L, X and Y be arbitrary vector fields on M with Lie product $[X, Y]$, and let σ be an arbitrary cross section. One can show that there is a unique two-form R, called the curvature form, on M such that

$$R(X, Y)\sigma = \nabla_X\nabla_Y\sigma - \nabla_Y\nabla_X\sigma - \nabla_{[X,Y]}\sigma. \tag{3}$$

The proof follows directly from the fact that the right-hand side of this equation is function-linear and skew-symmetric in X and Y. The *Bianchi identity* states that this two-form is closed, $dR = 0$, as can be verified from the definition of R.

Thus R represents a de Rham cohomology class $[R]$ of M. Using Čech cohomology theory one can show that the class $[R]/2\pi i$ is *integral*. Conversely, we have the following:

Existence theorem: Let M be a manifold and let Φ be a closed two-form on M such that $\Phi/2\pi i$ represents an integral class. Then there exists a line bundle L over M and a linear connection ∇ in L such that the corresponding curvature form coincides with Φ. Moreover, if M is simply connected, then L and ∇ are uniquely determined up to strong bundle isomorphisms. Finally, if $\Phi/2\pi i$ is real-valued, then L admits a positive definite hermitian metric $\langle\ ,\ \rangle_L$ such that

$$X(\langle\sigma_1, \sigma_2\rangle_L) = \langle\nabla_X\sigma_1, \sigma_2\rangle_L + \langle\sigma_1, \nabla_X\sigma_2\rangle_L \tag{4}$$

for arbitrary cross sections σ_1, σ_2 and vector fields X. This metric is determined up to a positive constant. (For the proofs see Ref. 1.)

3. THE GEOMETRIC INTERPRETATION OF MINIMAL COUPLING

We return now to the interpretation given to Eq. (2) at the end of the first section. Remember that we assumed F to be exact. Now regard Ψ more specifically as a section in the trivial bundle $D_F\times\mathbb{C}\to D_F$. In this case we can represent $\Psi: D_F\to D_F\times\mathbb{C}$ by a function denoted by the same letter:

$$\Psi(x) = (x, \Psi(x)), \quad x\in D_F.$$

The most general form of $\nabla_X\Psi$ in a trivial bundle is given by

$$\nabla_X\,\Psi(x) = (x, \varphi(x)),$$
$$\varphi = B(X)\cdot\Psi + X(\varphi),$$

where B is an arbitrary one-form on D_F. This implies that $R = dB$ [R curvature form, compare Eq. (3)]. If we put $B = ieA$, we recover the coordinate-free version of Eq. (1) and find the important relation $R = ieF$ which was anticipated in the first section.

Further, by the uniqueness part of the existence theorem of Sec. 2, the trivial bundle with the above connection is the *only* complex line bundle with curvature ieF, provided that D_F is simply connected. This will be assumed for simplicity, though even in the non-simply-connected case there is a complete classification of complex line bundles over D_F (See Ref. 1. This is also of physical interest because it explains flux quantization). The only remaining freedom, namely that the vector potentials are determined up to gradients, is as usual compensated by the gauge-invariance of Eq. (2). Thus the case when F is exact has been shown to fit completely into the line bundle description of Eq. (2). Now one observes that this description has a wider range of applicability, since it extends to the case when $eF/2\pi$ is not exact but represents an integral cohomology class (integrality condition). In fact, though the vector potentials cease to be defined globally on D_F, the operator ∇_μ, when interpreted as the covariant derivative with respect to ∂_μ, remains meaningful. It is no longer determined by an explicit formula like (1) but more indirectly by the condition that ∇ has curvature $R = ieF$.

This is sufficient to construct L and ∇ provided the integrality condition is fulfilled.

Conversely, if the relation $R = ieF$ between the electromagnetic field and the curvature form of a complex line bundle is valid, we get automatically the homogeneous set of Maxwell's equations as a consequence of the Bianchi identity for R. Moreover, the existence theorem tells us that $eF/2\pi$ defines an integral cohomology class. This is stated more explicitely by the condition

$$\int_\sigma eF/2\pi \in \mathrm{Z}$$

for an arbitrary closed two-dimensional surface in D_F. (For the relation between the integrality condition and charge quantization, see Ref. 6.) Furthermore, we can use the Hermitean metric $\langle\ ,\ \rangle_L$ in L which is determined by (4) up to a positive constant, to establish the usual probability interpretation of quantum mechanics: The set of sections obeying Eq. (2) is made into a Hilbert space with scalar product

$$(\sigma_1, \sigma_2) = \int_{t_0}\langle\sigma_1, \sigma_2\rangle_L.$$

The integral in this formula is taken over the three-dimensional space part of D_F at time t_0. This definition can be shown to be independent of the particular choice of t_0 as a consequence of (2) and (4). If the section Ψ is a solution of (2) normalized to $(\Psi, \Psi) = 1$, we can as usual interpret $\langle\Psi, \Psi\rangle_L(x)$ as the probability to find the particle at the point $x\in D_F$.

Of course, the case when F is integral but not exact will present many more difficulties than the case when F is exact. Even before one starts to solve Eq. (2) these difficulties start with the construction of the line bundle and the covariant derivative. This will now be illustrated for the magnetic monopole.

4. THE MAGNETIC MONOPOLE

As an example which will really provide us with nontrivial line bundles, we want to discuss the quantum-mechanical motion of a charged particle in a magnetic monopole field. This problem has been treated by several authors (see Refs. 7–12) but without replacing wavefunctions by sections in a complex line bundle. As a consequence one meets in the older literature vector potentials with unphysical singularities (the Dirac string) which only arise because of the improper use of coordinates. This can be avoided completely by use of the bundle formalism.

The magnetic monopole field is described by the two-form $F=\lambda\omega$, where λ is a pole strength and ω is given (in Cartesian space coordinates x_i) by

$$\omega=|x|^{-3}(x_1\,dx_2\wedge dx_3+x_3\,dx_1\wedge dx_2+x_2\,dx_3\wedge dx_1). \quad (5)$$

F is defined on $R\times\dot{R}^3$. Since it is time-independent, we are essentially left with the problem of finding line bundles of the form $R\times L\xrightarrow{l\times\pi} R\times\dot{R}^3$ ($\dot{R}^3=R^3-0$) where l denotes the identity map and $L\xrightarrow{\pi}\dot{R}^3$ is a complex line bundle over $\dot{R}^3$.

Therefore, we set $\nabla=\partial_0$ in Eq. (2). Next, we eliminate the time-dependence by the ansatz

$$\Psi(t,x)=(t,\ \exp(-iEt)\,\psi_0(x)),\quad t\in R,\quad x\in\dot{R}^3.$$

Then Eq. (2) yields the stationary equation for the cross section $\Psi_0:\dot{R}^3\to L$

$$E\Psi_0=-\frac{1}{2M}\sum_{i=1}^{3}\nabla_i^2\Psi_0, \quad (6)$$

where ∇_i denotes the covariant derivative in the direction of the vector fields $\partial/\partial x_i$ ($i=1,2,3$).

We still have to ensure the existence of line bundles with curvature ieF. In our particular example, the condition of the existence theorem in Sec. 2 reads

$$\int_{S^2} eF/2\pi\in\mathbf{Z}$$

where S^2 is a sphere in $\dot{R}^3$ centered at the origin. A simple calculation shows that this is equivalent to

$$\lambda=-m/2e,\quad m\in\mathbf{Z};$$

i.e., we get line bundles with the desired curvature if and only if the magnetic pole strength λ is an integral multiple of $1/2e$. This amazing quantization of the pole strength was first obtained by Dirac.

Applying the existence theorem once more and observing that $\dot{R}^3$ is simply connected, we obtain for each integer m a *unique* line bundle L_m over $\dot{R}^3$ with linear connection whose curvature is given by

$$R=-im\omega/2.$$

This line bundle and the corresponding covariant derivative will be constructed explicitly in the next two sections.

5. THE UNDERLYING PRINCIPAL BUNDLE

To describe the complex line bundles L_m we start with a single principal bundle over $\dot{R}^3$ which is very closely related to the Hopf fibering of S^3 over S^2. First recall the definition of the quaternion algebra. Let Q be a four-dimensional vector space with inner product $\langle\ ,\ \rangle$ and choose an orthonormal basis e, e_1, e_2, e_3. Define a multiplication in Q by setting

$$e_1e_2=e_3,\quad ee_i=e_ie=e_i,$$
$$e^2=-e_i^2=e\quad(i=1,2,3).$$

This multiplication makes Q into an associative division algebra, with unit element e, the *algebra of quaternions*. The *conjugate* of a quaternion $x=\lambda e+\sum_{i=1}^{3}\lambda_ie_i$, $\lambda,\lambda_i\in R$ is defined by

$$\bar{x}=\lambda e-\sum_{i=1}^{3}\lambda_ie_i.$$

Next, let R^3 denote the subspace of Q generated by e_1, e_2, and e_3 and let $\pi:\dot{Q}\to\dot{R}^3$ be the map given by

$$\pi(x)=xe_3\bar{x},\quad x\in\dot{Q}(=Q-0).$$

Since $\langle e,\pi(x)\rangle=0$ and $\pi(x)\neq0$ whenever $x\neq0$, π is indeed a map from $\dot{Q}$ to $\dot{R}^3$. It is easy to check that this map defines a principal fibration with fibre

$$S_1=\{\epsilon\,;\epsilon=\epsilon(t)=e\cos t+e_3\sin t,\ t\in[0,2\pi)\}$$

and that the right action of S^1 on $\dot{Q}$ is given by right multiplication.

Now we can represent every vector field on $\dot{Q}$ by a quaternion-valued function on $\dot{Q}$. The group action generates a vector field K on $\dot{Q}$ given by

$$K_x=xe_3,\quad x\in\dot{Q}.$$

A vector field Y on $\dot{Q}$ is called *horizontal* if

$$\langle Y_x,K_x\rangle=0,\quad x\in\dot{Q}.$$

To every vector field X on $\dot{R}^3$ there is a unique horizontal vector field X^* on $\dot{Q}$ satisfying

$$(d\pi)_xX_x^*=X_{\pi(x)},\quad x\in\dot{Q}.$$

It is called the *horizontal lift of* X and is explicitely given by

$$X_x^*=-X_{\pi(x)}\cdot x\cdot e_3/2|x|^2,\quad x\in\dot{Q}. \quad (7)$$

A straightforward calculation shows that if X and Y are vector fields on $\dot{R}^3$, then

$$[X^*,Y^*]_x=\tfrac{1}{2}\omega(X,Y)_{\pi(x)}\cdot K_x, \quad (8)$$

where ω is the two-form defined by (5). (The Lie-bracket should not be confused with the commutator of quaternions!)

6. THE BUNDLES L_m

Fix an integer m and let Φ_m denote the representation of S^1 in $\mathbf{C}$ given by

$$\Phi_m(\epsilon(t))=\tilde{\epsilon}(t)^m\cdot z,\quad z\in\mathbf{C},$$
$$\epsilon(t)=e\cos t+e_3\sin t,$$

$$\tilde{\epsilon}(t) = \exp(it), \quad t \in [0, 2\pi).$$

Then a right action of S^1 in $\dot{Q}\times\mathbb{C}$ is given by

$$\psi_m(\epsilon)(x, z) = (x\epsilon, \tilde{\epsilon}^{-m} z), \quad x \in \dot{Q}, \quad z \in \mathbb{C}.$$

This action defines an equivalence relation in $\dot{Q}\times\mathbb{C}$ as follows:

$$(x, z) \sim (x\epsilon, \tilde{\epsilon}^{-m} z).$$

Let $\dot{Q}\times_m\mathbb{C}$ denote the quotient manifold and let

$$q: \dot{Q}\times\mathbb{C} \to \dot{Q}\times_m\mathbb{C}$$

be the corresponding projection. Then $\dot{Q}\times_m\mathbb{C}$ becomes a line bundle over $\dot{R}^3$ and the projection ρ is determined by the commutative diagram

$$\begin{array}{ccc} \dot{Q}\times\mathbb{C} & \xrightarrow{q} & \dot{Q}\times_m\mathbb{C} \\ {\scriptstyle \pi_1}\downarrow & & \downarrow{\scriptstyle \rho} \\ \dot{Q} & \xrightarrow{\pi} & \dot{R}^3 \end{array}$$

where π_1 denotes the projection onto the first factor. The complex line bundle $L_m = (\dot{Q}\times_m\mathbb{C}, \rho, \dot{R}^3)$ so obtained is called *the complex line bundle associated to the principal bundle* $\dot{Q} \to \dot{R}^3$ *via the representation* Φ_m. (See Ref. 13.) This will turn out to be the line bundle we are looking for. We still have to define a covariant derivative and to ensure that it has the desired curvature.

A complex valued function f on $\dot{Q}$ is called *equivariant* (with respect to the representation Φ_m) if it satisfies

$$f(x\epsilon) = \tilde{\epsilon}^{-m} f(x), \quad x \in \dot{Q}, \quad \epsilon \in S^1.$$

Now we shall establish an isomorphism between the cross sections in L_m and the equivariant functions in $\dot{Q}$. To do so observe that the map q restricts to isomorphisms

$$q_x: \mathbb{C} \xrightarrow{\cong} \rho^{-1}(\pi(x)), \quad x \in \dot{Q},$$

on the fibres. Thus, if σ is a cross section in L_m, then a function $f_\sigma: \dot{Q} \to \mathbb{C}$ is defined by

$$f_\sigma(x) = q_x^{-1}(\sigma(\pi(x))).$$

This function is equivariant and, conversely, *every equivariant function on* $\dot{Q}$ *corresponds to precisely one cross section in* L_m. Thus q induces an isomorphism $q^\#$ from the cross sections in L_m to the equivariant functions in $\dot{Q}$. (See Ref. 14.)

Using this isomorphism we can define a linear connection in L_m by setting

$$\nabla_X\sigma = q^{\#-1} X^*(q^\#\sigma) \tag{9}$$

(X a vector field on $\dot{R}^3$). In fact, one easily checks that the operator ∇ satisfies the axioms of a linear connection.

To compute the corresponding curvature, let X and Y be constant vector fields on $\dot{R}^3$. Then we have by definition [compare Eq. (3)]

$$K(X, Y)\sigma = (\nabla_X\nabla_Y - \nabla_Y\nabla_X)\sigma.$$

Setting $q^\#\sigma = f$ yields in view of (9),

$$R(X, Y)\sigma = q^{\#-1}[X^*, Y^*](f).$$

Thus, using formula (8), we obtain

$$R(X, Y)\sigma = \tfrac{1}{2}\omega(X, Y)\, q^{\#-1} K(f). \tag{10}$$

Now, since f is equivariant, it satisfies

$$K(f) = -imf,$$

and so Eq. (10) yields

$$R(X, Y)\sigma = -i\tfrac{1}{2}m\,\omega(X, Y)\sigma.$$

Since this holds for every cross section, we obtain the formula

$$R = -\tfrac{1}{2}im\,\omega$$

for the curvature form of ∇. QED

Finally, we endow L_m with a Hermitian metric which has property (4). If σ_1, σ_2 are any two cross sections in L_m set

$$\varphi_i = q^{\#-1}\sigma_i, \quad i = 1, 2,$$

and define

$$\langle \sigma_1, \sigma_2 \rangle_L(y) = \overline{\varphi}_1(x)\,\varphi_2(x) \tag{11}$$

where $y \in \dot{R}^3$ and $x \in \pi^{-1}(y)$.

This definition makes sense because, in view of the equivariance of φ_1 and φ_2, the right-hand side of the equation above does not depend on the choice of $x \in \pi^{-1}(y)$. If ∇ is given by (9) it is trivial to check the validity of (4).

7. SOLUTION OF THE SCHRÖDINGER EQUATION

In this section we shall use the machinery developed above to solve Eq. (6) for the cross section Ψ_0 in L_m. Applying $q^\#$ to both sides of (6) and setting $\varphi_0 = q^\#\Psi_0$, we obtain with the help of (9) and (7)

$$E\varphi_0 = -\frac{1}{2M}\sum_{i=1}^{3} D_i^2\,\varphi_0, \tag{12}$$

where D_i denotes the ordinary derivative in the direction of the horizontal lift

$$e_i^* = -e_i x e_3/2|x|^2, \quad x \in \dot{Q}$$

[compare Eqs. (7) and (9)].

Recall that φ_0 is an equivariant function in Q,

$$\varphi_0(x(\cos t\cdot e + \sin t\cdot e_3)) = \exp(-imt)\,\varphi_0(x), \quad x \in \dot{Q}, \ t \in [0, 2\pi).$$

If $\overline{\varphi}_0(x)\varphi_0(x)$ is to be interpreted as the probability to find the particle at the point $\pi(x) \in \dot{R}^3$ [compare Sec. 3 and formula (11)], φ_0 must obviously be bounded on $\dot{Q}$.

For φ_0 we try the ansatz

$$\varphi_0(x) = f(|x|^2)\cdot g(x/|x|), \quad x \in \dot{Q},$$

with f a function on the positive reals and g a function on the group of unit quaternions, that is the group $SU(2)$. We will need the angular momentum operators J_l and K_l ($l = 1, 2, 3$):

$$(J_l g)(x_0) = \frac{i}{2}\frac{d}{dt}\, g(\exp(-te_l)x_0)\big|_{t=0},$$

$$(K_l g)(x_0) = \frac{i}{2}\frac{d}{dt}\, g(x_0\exp(te_l))\big|_{t=0},$$

$$J^2=\sum_{i=1}^{3} J_i^2, \quad x_0\in SU(2).$$

Observe that $[J_m,K_n]=0$. Insertion of the ansatz for φ_0 into (12) yields after a straightforward calculation ($r=|x|^2$, $x_0=x/|x|$)

$$(2MEf(r)+f''(r)+2f'(r)/r)g(x_0)$$
$$=[(J^2g)(x_0)-(K_3^2g)(x_0)]f(r)/r^2.$$

This shows that g must be an eigenfunction of the operator $J^2-K_3^2$. Since φ_0 is equivariant it follows that $K_3g=-mg/2$. Thus g must be an eigenfunction of J^2. From the representation theory of the group $SU(2)$ (see Ref. 15) we know that the eigenvalues of J^2 are of the form $j(j+1)$, $j=|m/2|$, $|m/2|+1$, ... A complete set of eigenfunctions is given by the Wigner coefficients $D^j_{n,-m/2}$ of the irreducible representations of $SU(2)$ (see Ref. 15), $n=-j$, $-j+1$, ..., j.

We are now left with the radial equation

$$2MEf+f''+2f'/r-[j(j+1)-m^2/4]f/r^2=0,$$

which was already obtained by Tamm (see Ref. 8). Following his arguments we have to restrict the energy E to positive values, for only under this assumption we get solutions which remain bounded when r tends to infinity.

For positive energy two linearly independent solutions are given by

$$f_\pm(r)=r^{-\frac{1}{2}}J_{\pm\lambda}(\sqrt{2ME}\,r), \quad \lambda=[(2j+1)^2-m^2]^{1/2}/2.$$

$J_{\pm\lambda}$ are Bessel-functions (See Ref. 16). f_- has to be omitted as it blows up when r tends to zero. Thus a complete set of equivariant solutions of Eq. (12) is given by the following functions $\varphi^{Ej}_{nm}:\dot{Q}\to\mathbb{C}$:

$$\varphi^{Ej}_{nm}(x)=|x|^{-1}J_\lambda(\sqrt{2ME}\ |x|^2)D^j_{n,-m/2}(x/|x|),$$

$$\lambda=[(2j+1)^2-m^2]^{1/2}/2, \quad j=|m/2|, \ |m/2|+1, \ \cdots,$$

$$n=-j, \ -j+1, \ \cdots, j, \quad E>0.$$

Obviously, we get a complete set of cross sections Ψ^{Ej}_{nm} in L_m which satisfy Eq. (6) by means of the formula (compare Sec. 6)

$$\Psi^{Ej}_{nm}=q^{\#-1}(\varphi^{Ej}_{nm}).$$

These cross sections can be described in a different way by means of local trivializations. Write ($i=1,2,3$)

$$y_i=\langle y,e_i\rangle$$

for $y\in\dot{R}^3$ and define the local cross-sections $\sigma_\pm$ in the principal bundle $\dot{Q}\xrightarrow{\pi}\dot{R}^3$ as follows:

$$\sigma_+(y)=(|y|e_3+y)[2(|y|+y_3)]^{-1/2},$$

where

$$y\in D_+=\dot{R}^3-\{x;\ |x|+x_3=0\}$$

and

$$\sigma_-(y)=\sigma_+(-y)e_2,$$

where

$$y\in D_-=\dot{R}^3-\{x;\ |x|-x_3=0\}.$$

Then $D_+\cup D_-=\dot{R}^3$. Furthermore, the following transition law is valid in $D_+\cap D_-$:

$$\sigma_-(y)=\sigma_+(y)(-y_1e+y_2e_3)\cdot(y_1^2+y_2^2)^{-1/2}.$$

The cross sections $\sigma_\pm$ induce local cross sections $\sigma^m_\pm$ in L_m (compare Sec. 6) given by

$$\sigma^m_\pm(y)=q(\sigma_\pm(y),1), \quad y\in D_\pm.$$

From this we can deduce the following identity:

$$\psi^{Ej}_{nm}(y)=\varphi^{Ej}_{nm}(\sigma_\pm(y))\cdot\sigma^m_\pm(y), \quad y\in D_\pm.$$

What has been calculated in the older literature was indeed the function $\varphi^{Ej}_{nm}\circ\sigma_\pm:D_\pm\to\mathbb{C}$ which we may call the local description of the section Ψ^{Ej}_{nm} because it is not defined on the whole of $\dot{R}^3$.

After having established the relation to former work, we close the discussion of the magnetic monopole.

CONCLUSION

We have shown that minimal coupling of an external electromagnetic field, which is not exact, to the Schrödinger equation, provides us with a nontrivial physical application of the theory of complex line bundles. Once the language and the results of modern differential geometry are adopted, the Schrödinger equation can be described in a consistent way, i.e., without introducing unphysical singular vector potentials. In the special case of the magnetic monopole the actual solutions of the Schrödinger equation are even more easily obtained and interpreted.

ACKNOWLEDGMENTS

Finally, we want to thank K. Bleuler and D. Schütte for many interesting discussions. One of us (W. G.) expresses his gratitude to the Heinrich–Hertz foundation for financial support during his stay at the Institute für Theoretische Kernphysik, Bonn.

[1]B. Kostant, "Quantization and unitary representations," in *Springer Lecture Notes in Mathematics* (Springer-Verlag, Heidelberg, 1970).
[2]B. Kostant, "Line bundles and the prequantized Schrödinger equation," preprint, 1973.
[3]E. Onofri and M. Pauri, J. Math. Phys. **13**, 538 (1972).
[4]R.J. Blattner, "Quantization and representation theory," *Proceedings of A.M.S. 1972 Summer Institute on Group Representation.*
[5]D.J. Simms, Z. Naturforsch. 28a, 538 (1973).
[6]J. Sniatycki, J. Math. Phys. **15**, 619 (1974).
[7]P.A.M. Dirac, Proc. Roy. Soc. A **133**, 60 (1931).
[8]I. Tamm. Z. Physik **71**, 141 (1931).
[9]P.A.M. Dirac, Phys. Rev. **74**, 817 (1948).
[10]P.P. Banderet, Helv. Phys. Acta **19**, 503 (1946).
[11]A.S. Goldhaber, Phys. Rev. **140**, 1407 (1965).
[12]D.M. Stevens, "The magnetic monopole, a bibliography," University of Virginia Preprint TVPI-EPP-70-6, 1970.
[13]W. Greub *et al.*, *Connections, Curvature and Cohomology* (Academic, New York, 1973), p. 198
[14]Reference 13, p. 406.
[15]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N.J., 1957).
[16]A. Erdelyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York 1953).

Higher symmetry of certain classical quadratic Hamiltonians

L. S. Wollenberg

Department of Mathematics, Queen Elizabeth College, London W.8, England
(Received 16 July 1974)

It is proved explicitly that the classical n-dimensional isotropic harmonic oscillator is invariant under $U(n)$. Two new examples of higher symmetry are found. One is the n-dimensional free particle which is shown to be invariant under a semidirect product which contains the n-dimensional Euclidean group as a proper subgroup.

1. INTRODUCTION

Certain classical systems exhibit obvious geometric symmetry. For example, a particle moving in three dimensions in a spherically symmetric potential has as obvious geometric symmetry group the three-dimensional orthogonal group $O(3)$. There are also certain classical systems for which the obvious symmetry group is only a subgroup of some larger not-so-obvious symmetry group. Such a system (and its Hamiltonian) is said to exhibit *higher* (or *accidental*) *symmetry*, and the larger group is called the *higher symmetry group* or the *dynamical group*.

Oft-quoted examples of this are (for all $n \geq 2$) the n-dimensional Kepler problem, which for negative energies is invariant under $O(n+1)$ and the n-dimensional isotropic harmonic oscillator (henceforth abbreviated to n-oscillator), which is invariant under $U(n)$. In both cases the obvious symmetry group is just $O(n)$.

But these systems are usually discussed in a quantum-mechanical context, and in fact most of the results stated here have been proved only for the corresponding *quantum-mechanical* systems (see Refs. 1, 2). Some proofs have been given in a classical mechanical context. Thus it has been shown (see Ref. 3) that the classical three-dimensional Kepler problem is invariant under $SO(4)$ [the connected component of the identity of $O(4)$], and (see Refs. 4, 5) that the classical two-oscillator is invariant under $SU(2)$ [the connected component of the identity of $U(2)$]. But to our knowledge no explicit proof has yet been given for any of the other cases mentioned that the stated higher symmetry group is actually exhibited by the corresponding *classical* system.

Such explicit proof would be of interest, for although many authors [see for example Refs. 3(a), 6—9] have worked on the problem of constructing $O(n+1)$ and $SU(n)$ Lie algebras (i.e., generators) for classical systems, even if such are constructed there remains [see Refs. 3(a), 6] the problem of finding the finite canonical transformations generated. Indeed it is pointed out in Ref. 3(a) that if construction of a given Lie algebra is possible for a given classical system, it is necessarily possible for all classical systems with the same number of degrees of freedom, due to the fact that all $2n$-dimensional symplectic manifolds are locally isomorphic.

In this paper we use simple methods to find the *linear symmetry group* (defined in Sec. 2) of classical Hamiltonians of form

$$H(q,p,\mu)=\tfrac{1}{2}\sum_{i=1}^{n}(p_i^2+\mu q_i^2) \quad (\mu = \text{a real constant}).$$

For $\mu>0$ this represents the n-oscillator, and after identifying groups which are isomorphic we obtain the group $U(n)$. Although, as noted, this result has long been "well-known," it has not previously been proved explicitly for general n. Our method is easier than those in Refs. 4, 5.

We also get some completely new results. For $\mu<0$ we obtain the group $GL(n)$ of real $(n\times n)$ nonsingular matrices, and for $\mu=0$ (representing the n-dimensional free particle) if we include also q-translations we obtain a semidirect product which contains the n-dimensional Euclidean group $E(n)$ as a proper subgroup. Since the obvious symmetry groups in these cases are just $O(n)$, $E(n)$ respectively, this proves that for all $n \geq 1$ the Hamiltonian $H(q,p,\mu)$ exhibits higher symmetry no matter what value μ has.

In Sec. 2 we give detailed definitions. In Sec. 3 we prove our results for the $\mu\neq 0$ cases, and in Sec. 4 we deal with the case $\mu=0$. Finally in Sec. 5 we make some brief remarks concerning quantum-mechanical implications.

2. LINEAR SYMMETRY GROUP OF A HAMILTONIAN

Consider a holonomic classical system with n degrees of freedom, described by canonical variables $(q,p)=(q_1,\ldots,q_n,\ p_1,\ldots,p_n)$. Let $H(q,p)$ denote its Hamiltonian (assumed not to depend explicitly on time), and let Ω denote *phase space*, the space of all points (q,p). Let $W=C^\infty(\Omega)$. Elements of W are called *dynamical variables*. For every $f,\ g\in W$ define the *Poisson bracket* $\{f,g\}$ by (see Ref. 10)

$$\{f,g\}=\sum_{i=1}^{n}\left(\frac{\partial f}{\partial q_i}\circ\frac{\partial g}{\partial p_i}-\frac{\partial f}{\partial p_i}\circ\frac{\partial g}{\partial q_i}\right).$$

A *canonical tranformation* (see Ref. 10) is a one-to-one map γ from Ω onto Ω, which we shall write as

$$\gamma:(q,p)\to(q',p')=(q_1',\ldots,q_n',\ p_1',\ldots,p_n'),$$

such that each $q_i',\ p_j'\in W$ and

$$\{q_i',q_j'\}=\{p_i',p_j'\}=0,\ \{q_i',p_j'\}=\delta_{ij}.\quad (i,j=1,2,\ldots,n). \tag{1}$$

We interpret a canonical transformation γ as a change of label of points of Ω. Thus $(q',p')=\gamma(q,p)$ are new

canonical variables labelling the point previously described by (q,p).

Canonical transformations in which the q_i', p_i' are all homogeneous linear polynomials in the variables (q,p) are known as *symplectic tranformations*. These are discussed in greater detail below.

Let Γ denote the group of all *time-independent* canonical transformations, and $Sp(n,R)$ the group (=subgroup of Γ) of all *time-independent* symplectic transformations. We say that the Hamiltonian $H(q,p)$ of our system is *invariant* under $\gamma \in \Gamma$ if the new Hamiltonian $H'(q',p')$ [i.e., corresponding to the new canonical variables $(q',p') = \gamma(q,p)$] has the same functional form as $H(q,p)$, that is, if $H'(q',p') = H(q',p')$. But since γ is time-independent, the new Hamiltonian is given (see Ref. 10) by $H'(q',p') = H(q,p)$. Hence the invariance condition can be written

$$H(q',p') = H(q,p). \tag{2}$$

Let $S(H)$ denote the group of all $\gamma \in \Gamma$ which leave $H(q,p)$ invariant. We say that the system (and its Hamiltonian) has *symmetry group* $S(H)$, or is *invariant under* $S(H)$.

In what follows we use $\approx$ to denote group isomorphism, and we will often find it convenient to identify isomorphic groups. For example, if $S(H) \approx O(3)$ we may find it convenient to say $H(q,p)$ has symmetry group $O(3)$.

As mentioned in Sec. 1 we can often spot $S(H)$ geometrically, but sometimes we obtain only a subgroup of $S(H)$ this way. Note that if $H(q,p)$ itself generates globally defined canonical transformations, then (see Refs. 11, 12) $H(q,p)$ will be invariant under these transformations (which can be considered as elements of Γ since in this situation we can regard time t merely as a real parameter), and hence $S(H)$ will be nontrivial.

In fact $S(H)$ may be quite large [cf. (23) below], and for this reason we find it convenient to introduce the group $L(H)$ of all $\gamma \in Sp(n,R)$ which leave $H(q,p)$ invariant. We call $L(H)$ the *linear symmetry group* of $H(q,p)$.

It if often assumed tacitly that $S(H) = L(H)$, but this may not be so [cf. (23) again]. In fact it should be noted that, provided we keep to time-independent changes of label, $S(H)$ is independent of the canonical variables (q,p) chosen to describe our system, but $L(H)$ is not. Fortunately this does not affect any of the arguments or conclusions which follow.

In Secs. 3 and 4 we find $L(H)$ when

$$H(q,p) = \tfrac{1}{2}\sum_{i=1}^{n} (p_i^2 + \mu q_i^2) \quad (\mu = \text{a real constant}). \tag{3}$$

This $H(q,p)$ generates globally defined canonical transformations no matter what value μ has. In fact they are symplectic in each case and thus (from previous remarks) elements of $L(H)$.

Now the symplectic condition can be expressed in a convenient manner if (as we will often do from now on) we denote points of Ω by $(2n\times 1)$ column matrices and introduce the $(2n\times 2n)$ block matrix

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

where I_n denotes the $(n\times n)$ identity matrix. We can now state (see Refs. 12, 13) that the homogeneous linear transformation

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = T\begin{pmatrix} q \\ p \end{pmatrix} \quad [T = \text{a real } (2n\times 2n) \text{ matrix}] \tag{4}$$

is symplectic if and only if

$$T'JT = J \quad (' \text{ denoting matrix transpose}) \tag{5}$$

[i.e., conditions (1) take this form when the transformation γ has form (4)]. Now in such a case T is necessarily nonsingular, for (5) implies

$$\det(T'JT) = (\det T')(\det J)(\det T) = (\det J) \neq 0,$$

i.e.,

$$(\det T) \neq 0.$$

The Hamiltonian (3) can now be regarded as a quadratic form with matrix K having the block form

$$K = \tfrac{1}{2}\begin{pmatrix} \mu I_n & 0 \\ 0 & I_n \end{pmatrix},$$

and thus a time-independent transformation of form (4) leaves $H(q,p)$ invariant [see (2)] if and only if

$$T'KT = K. \tag{6}$$

Hence, identifying the linear transformations (4) with their matrices in the usual way, we find that $L(H)$ is the multiplicative group of time-independent matrices T satisfying both the symplectic condition (5) and the invariance condition (6). In what follows we will assume that T has block form

$$T = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{7}$$

where A, B, C, D are real time-independent $(n\times n)$ matrices.

3. DERIVATION OF $L(H)$ WHEN $\mu \neq 0$

A. Introduction

In this case direct use of (5) and (6) leads to complicated matrix equations. Instead, since here K is nonsingular (we have already noted that T must be nonsingular) we proceed as follows. From (6)

$$T' = K(KT)^{-1} = KT^{-1}K^{-1},$$

and substituting this into (5) gives

$$KT^{-1}K^{-1}JT = J, \quad \text{i.e.,} \quad K^{-1}JT = TK^{-1}J. \tag{5'}$$

Conversely, if T satisfies (5') and (6) then it satisfies (5) and (6). Now the block matrix (7) satisfies (5') if and only if

$$\begin{pmatrix} \mu^{-1}I_n & 0 \\ 0 & I_n \end{pmatrix}\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} \mu^{-1}I_n & 0 \\ 0 & I_n \end{pmatrix}\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

which becomes

$$\begin{pmatrix} \mu^{-1}C & \mu^{-1}D \\ -A & -B \end{pmatrix} = \begin{pmatrix} -B & \mu^{-1}A \\ -D & \mu^{-1}C \end{pmatrix},$$

which in turn reduces to $C=-\mu B$ and $D=A$. The resulting matrix

$$T=\begin{pmatrix} A & B \\ -\mu B & A \end{pmatrix} \tag{8}$$

satisfies (6) if and only if

$$\begin{pmatrix} A & B \\ -\mu B & A \end{pmatrix}' \begin{pmatrix} \mu I_n & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} A & B \\ -\mu B & A \end{pmatrix} = \begin{pmatrix} \mu I_n & 0 \\ 0 & I_n \end{pmatrix},$$

which becomes

$$\begin{pmatrix} \mu A'A+\mu^2 B'B & \mu A'B-\mu B'A \\ \mu B'A-\mu A'B & \mu B'B+A'A \end{pmatrix} = \begin{pmatrix} \mu I_n & 0 \\ 0 & I_n \end{pmatrix},$$

which in turn reduces to

$$A'A+\mu B'B=I_n \tag{9}$$

and

$$A'B-B'A=0. \tag{10}$$

Thus $L(H)$ is the multiplicative group of time-independent matrices T having block form (8) with A, B satisfying (9) and (10). The identity element in $L(H)$ corresponds to $A=I_n$, $B=0$. We are now in a position to prove the following results, which are valid for all $n \geq 1$.

Theorem 1: (i) If $\mu<0$, then $L(H)=GL(n)$.

(ii) If $\mu>0$, then $L(H)=U(n)$.

B. Proof of Theorem 1(i)

Suppose $\mu<0$, and write $\mu=-\lambda^4$. Then condition (9) becomes

$$A'A-\lambda^4 B'B=I_n \tag{9'}$$

Lemma 1: The $(n\times n)$ time-independent matrices A, B satisfy (9') and (10) if and only if

$$A+\lambda^2 B=E \quad \text{and} \quad A-\lambda^2 B=E'^{-1} \tag{11}$$

for some time-independent $E\in GL(n)$.

Proof of Lemma 1: First note that A, B satisfy (9') and (10) if and only if

$$(A-\lambda^2 B)'(A+\lambda^2 B)=I_n. \tag{12}$$

For if (9') and (10) hold, then

$$(A-\lambda^2 B)'(A+\lambda^2 B)=(A'A-\lambda^4 B'B)+\lambda^2(A'B-B'A)=I_n.$$

Conversely, if (12) holds, then

$$(A'A-\lambda^4 B'B)+\lambda^2(A'B-B'A)=I_n,$$

and taking transposes we find also

$$(A'A-\lambda^4 B'B)+\lambda^2(B'A-A'B)=I_n.$$

Together these give (9') and (10). But (12) is satisfied if and only if $(A+\lambda^2 B)\in GL(n)$ and $(A-\lambda^2 B)=(A+\lambda^2 B)'^{-1}$, i.e., if and only if (11) holds for some time-independent $E\in GL(n)$. This proves Lemma 1.

Thus $L(H)$ is the multiplicative group of time-independent matrices T having block form

$$T=\begin{pmatrix} A & B \\ \lambda^4 B & A \end{pmatrix} \tag{8'}$$

with A, B satisfying (11) for some time-independent $E\in GL(n)$.

It can be shown that the elements of $L(H)$ generated by $H(q,p)$ itself correspond to our taking $E=(e^{\lambda^2 t})I_n$ (t = a real parameter) in (11).

Now express all elements of $L(H)$ in the form (8') and consider the map $\phi: L(H)\to GL(n)$ defined by $\phi(T)=A+\lambda^2 B$. Let T_1, $T_2\in L(H)$. Then, since

$$\begin{pmatrix} A_1 & B_1 \\ \lambda^4 B_1 & A_1 \end{pmatrix}\begin{pmatrix} A_2 & B_2 \\ \lambda^4 B_2 & A_2 \end{pmatrix} = \begin{pmatrix} A_1A_2+\lambda^4 B_1B_2 & A_1B_2+B_1A_2 \\ \lambda^4(A_1B_2+B_1A_2) & A_1A_2+\lambda^4 B_1B_2 \end{pmatrix},$$

we obtain the homomorphism property

$$\begin{aligned} \phi(T_1T_2) &= (A_1A_2+\lambda^4 B_1B_2)+\lambda^2(A_1B_2+B_1A_2) \\ &= (A_1+\lambda^2 B_1)(A_2+\lambda^2 B_2) \\ &= \phi(T_1)\phi(T_2), \quad \text{all } T_1,\ T_2\in L(H). \end{aligned}$$

Also for every $E\in GL(n)$ there exists an element $T\in L(H)$ such that $\phi(T)=E$. To construct T (in fact it is unique), just define A, B by (11). In particular, if $\phi(T)=I_n$, we must have $(A+\lambda^2 B)=(A-\lambda^2 B)=I_n$, which gives $A=I_n$ and $B=0$, i.e., T must be the identity element in $L(H)$.

We have shown that ϕ is a group isomorphism between $L(H)$ and $GL(n)$, and identifying these isomorphic groups we obtain Theorem 1(i).

Remark: Theorem 1(i) can also be proved by reducing (3), via a symplectic transformation, to the form $\lambda^2\sum p_i q_i$. The method used here is easier, and also has the advantage that the $\mu>0$ case can be dealt with (see Sec. 3C next) along roughly similar lines.

C. Proof of Theorem 1 (ii)

Suppose $\mu>0$, and write $\mu=\lambda^4$. Then condition (9) becomes

$$A'A+\lambda^4 B'B=I_n. \tag{9''}$$

Lemma 2: The real $(n\times n)$ matrices A, B satisfy (9'') and (10) if and only if $(A+i\lambda^2 B)\in U(n)$.

Proof of Lemma 2: First note that A, B satisfy (9'') and (10) if and only if

$$(A-i\lambda^2 B)'(A+i\lambda^2 B)=I_n. \tag{13}$$

For if (9'') and (10) hold, then

$$(A-i\lambda^2 B)'(A+i\lambda^2 B)=(A'A+\lambda^4 B'B)+i\lambda^2(A'B-B'A)=I_n.$$

Conversely, if (13) holds, then

$$(A'A+i\lambda^4 B'B)+i\lambda^2(A'B-B'A)=I_n,$$

and taking real and imaginary parts gives (9'') and (10). Now (13) is the condition $(A+i\lambda^2 B)\in U(n)$. Conversely, if $U\in U(n)$ is expressed in the form

$$U=A+i\lambda^2 B \quad (A,B \text{ real matrices}), \tag{14}$$

then $(A-i\lambda^2 B)'=\bar{U}'=U^{-1}$, and (13) is satisfied. Thus (13) itself is satisfied if and only if $(A+i\lambda^2 B)\in U(n)$. This proves Lemma 2.

Thus $L(H)$ is the multiplicative group of time-inde-

pendent matrices T having block form

$$T=\begin{pmatrix} A & B \\ -\lambda^4 B & A \end{pmatrix} \tag{8''}$$

with $(A+i\lambda^2 B)\in U(n)$.

It can be shown that the elements of $L(H)$ generated by $H(q,p)$ itself correspond to our taking $U=(e^{i\lambda^2 t})I_n$ (t = a real parameter) in (14).

Now express all elements of $L(H)$ in the form (8'') and consider the map $\psi: L(H)\to U(n)$ defined by $\psi(T)=A+i\lambda^2 B$. It is straightforward to verify the homomorphism property

$$\psi(T_1T_2)=\psi(T_1)\psi(T_2), \quad \text{all } T_1,\ T_2\in L(H).$$

Also for every $U\in U(n)$ there exists an element $T\in L(H)$ such that $\psi(T)=U$. To construct T (in fact it is unique), just define A, B by (14). In particular, if $\psi(T)=I_n$, we must have $A=I_n$ and $B=0$, i.e., T must be the identity element in $L(H)$.

We have shown that ψ is a group isomorphism between $L(H)$ and $U(n)$, and identifying these isomorphic groups we obtain Theorem 1(ii).

Remark: Theorem 1(ii) is essentially the statement[14] $Sp(n,R)\cap O(2n)\approx U(n)$. A similar result to this is proved in Helgason's book [see Ref. 13, p. 342, Lemma 4.1., part (c)], and when $\mu=1$ our proof reduces to a more detailed version of his.

D. Comments on Theorem 1

Note that in both cases the set N of matrices having block form (8) with $B=0$ and $A'A=I_n$ (i.e., $A\in O(n)$) forms a subgroup of $L(H)$ isomorphic to $O(n)$, the obvious symmetry group for $H(q,p)$. Thus in both cases we get higher symmetry for all $n\geq 1$.

It can be shown that in both cases $L(H)$ is generated by the n^2 conserved quantities consisting of the $\frac{1}{2}n(n-1)$ angular momenta

$$q_ip_j-q_jp_i\ (i\neq j)\quad (i,j=1,2,\ldots,n) \tag{15}$$

which generate N and the $\frac{1}{2}n(n+1)$ quantities

$$p_ip_j+\mu q_iq_j\quad (i,j=1,2,\ldots,n). \tag{16}$$

For $\mu>0$ these can be combined to give the Hamiltonian (3) itself, together with (n^2-1) quantities which generate $SU(n)$. A possible set of $SU(n)$ generators for the $\mu=1$ case was given by Jauch and Hill,[15] who were the first to connect the $SU(n)$ Lie algebra and the n-oscillator.

4. DERIVATION OF $L(H)$ WHEN $\mu = 0$

A. Introduction

Suppose $\mu=0$. In this case we make direct use of (5) and (6). The block matrix (7) satisfies (6) if and only if

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}'\begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix},$$

which becomes

$$\begin{pmatrix} C'C & C'D \\ D'C & D'D \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix},$$

which in turn reduces to $D'D=I_n$ and $C=0$. The resulting matrix

$$T=\begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$$

satisfies (5) if and only if

$$\begin{pmatrix} A & B \\ 0 & D \end{pmatrix}'\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}\begin{pmatrix} A & B \\ 0 & D \end{pmatrix}=\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

which becomes

$$\begin{pmatrix} 0 & A'D \\ -D'A & B'D-D'B \end{pmatrix}=\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

i.e.,

$$A'D=D'A=I_n \tag{17}$$

and

$$B'D-D'B=0. \tag{18}$$

Since $D'D=I_n$, condition (17) reduces to $A=D'^{-1}=D$ and

$$A'A=I_n, \tag{19}$$

and (18) reduces to

$$B'A-A'B=0. \tag{20}$$

Thus $L(H)$ is the multiplicative group of time-independent matrices T having block form

$$T=\begin{pmatrix} A & B \\ 0 & A \end{pmatrix}, \tag{21}$$

with A, B satisfying (19) and (20). [Although (19) can be obtained by putting $\mu=0$ into (9), and (20) is just (10), our derivation of (9) and (10) was valid only for $\mu\neq 0$.] Note that when $A=I_n$ condition (20) becomes $B'=B$. It can be shown that the elements of $L(H)$ generated by $H(q,p)$ itself correspond to our setting $A=I_n$, $B=tI_n$ (t = a real parameter) in (21).

Now the q-translations, i.e., the canonical transformations of form

$$(q,p)\to(q_1+\alpha_1,\ q_2+\alpha_2,\ldots,q_n+\alpha_n,\ p_1,\ldots,p_n)$$

$$(\alpha_i=\text{arbitrary})$$

clearly also leave the n-dimensional free particle Hamiltonian invariant. Combining these with $L(H)$, we obtain a subgroup $IL(H)$ of $S(H)$ consisting of all time-independent inhomogeneous linear canonical transformations

$$\begin{pmatrix} q \\ p \end{pmatrix}\to\begin{pmatrix} A & B \\ 0 & A \end{pmatrix}\begin{pmatrix} q \\ p \end{pmatrix}+\begin{pmatrix} \boldsymbol{\alpha} \\ 0 \end{pmatrix}, \tag{22}$$

where A, B satisfy (19) and (20), and

$$\begin{pmatrix} \boldsymbol{\alpha} \\ 0 \end{pmatrix}=(\alpha_1,\alpha_2,\ldots,\alpha_n,0,\ldots,0)',$$

with the α_i arbitrary.

It is worth noting that $IL(H)$ is a proper subgroup of $S(H)$, for $S(H)$ also contains all canonical transformations of form

$$(q,p)\to(q_1+f_1(p_1),\ q_2+f_2(p_2),\ldots,q_n+f_n(p_n),\ p_1,\ldots,p_n), \tag{23}$$

and not all of these are in $IL(H)$.

Now let $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_n)'$, and denote the transformation (22) by (A,B,α). Then the identity element in $IL(H)$ is the transformation $(I_n,0,0)$, and we find that products and inverses in $IL(H)$ are given by

$$(A_1,B_1,\alpha_1)(A_2,\ B_2,\ \alpha_2)=(A_1A_2,\ A_1B_2+B_1A_2,\ \alpha_1+A_1\alpha_2) \tag{24}$$

and

$$(A,B,\alpha)^{-1}=(A',\ -A'BA',\ -A'\alpha). \tag{25}$$

B. Semidirect product structure of $IL(H)$

Suppose a group G (denote its identity element by e) has a normal subgroup M and an ordinary subgroup S such that $M\cap S=\{e\}$ and such that every element $g\in G$ can be expressed in at least one way as a product $g=ms$ with $m\in M$ and $s\in S$. Then (see Ref. 16, p. 100) we say G is a *semidirect product* of M and S. We will denote this by $G=M*S$. It can be shown that because $M\cap S=\{e\}$, the expression of g as a product ms is unique.

To reconstruct G from M and S we need to know, for each $s\in S$, the automorphism $m\to sms^{-1}$ of M (for details see Ref. 16).

Theorem 2: For all $n\geq 1$,

$$IL(H)=R^{n(n+1)/2}*E(n).$$

C. Proof of Theorem 2

Lemma 3: The set $S=\{(A,0,\alpha): A'A=I_n$ and $\alpha=$ arbitrary$\}$ is a subgroup of $IL(H)$, and $S\approx E(n)$.

Proof of Lemma 3: Certainly S is a subset of $IL(H)$, and from (24) and (25) we get

$$(A_1,0,\alpha_1)(A_2,0,\alpha_2)=(A_1A_2,0,\alpha_1+A_1\alpha_2), \tag{26}$$
$$(A,0,\alpha)^{-1}=(A',0,-A'\alpha),$$

showing that S is a subgroup of $IL(H)$.

Now the n-dimensional Euclidean group $E(n)$ is just the set of transformations on R^n of form

$$(A,\alpha):\mathbf{x}\to A\mathbf{x}+\alpha,$$

where $A'A=I_n$, α is arbitrary, and the $(n\times 1)$ column matrix $\mathbf{x}$ denotes a typical point of R^n. In particular the product of elements of $E(n)$ is given by

$$(A_1,\alpha_1)(A_2,\alpha_2)=(A_1A_2,\alpha_1+A_1\alpha_2). \tag{27}$$

By using (26) and (27) it is easy to verify that the map $(A,0,\alpha)\to(A,\alpha)$ is a group isomorphism between S and $E(n)$. This proves Lemma 3.

Lemma 4: The set $M=\{(I_n,B,0): B'=B\}$ is a normal subgroup of $IL(H)$, and $M\approx R^{n(n+1)/2}$.

Proof of Lemma 4: Certainly M is a subset of $IL(H)$. But from (24) and (26) it is easy to see that the map $(A,B,\alpha)\to(A,0,\alpha)$ is a group homomorphism from $IL(H)$ onto S, and since M is just its kernel, M must be a normal subgroup of $IL(H)$. Also from (24)

$$(I_n,B_1,0)(I_n,B_2,0)=(I_n,B_1+B_2,0),$$

and it is now easy to verify that the map $(I_n,B,0)\to B$ is a group isomorphism between M and the additive group of real symmetric $(n\times n)$ matrices. Since the latter is itself isomorphic to $R^{n(n+1)/2}$, there being $n(n+1)/2$ independent entries in a symmetric $(n\times n)$ matrix, we have $M\approx R^{n(n+1)/2}$. This proves Lemma 4.

Lemma 5: Every element in $IL(H)$ can be expressed in at least one way as a product ms with $m\in M$ and $s\in S$.

Proof of Lemma 5: If the $(n\times n)$ matrices A, B satisfy (19) and (20), then also

$$A(B'A)A'=A(A'B)A'$$

i.e.

$$AB'=BA'.$$

Thus a typical element (A,B,α) in $IL(H)$ can be expressed as

$$(A,B,\alpha)=(I_n,BA',0)(A,0,\alpha),$$

which is of the form required. This proves Lemma 5.

Now every element (A,B,α) in $M\cap S$ must have $A=I_n$, $\alpha=0$, and $B=0$, i.e., must be the identity element in $IL(H)$. This, together with Lemmas 3–5, proves that $IL(H)=M*S$. Identifying M with $R^{n(n+1)/2}$ and S with $E(n)$, we obtain Theorem 2.

It is worth noting that, for $s=(A,0,\alpha)\in S$ and $m=(I_n,B,0)\in M$,

$$sms^{-1}=(A,0,\alpha)(I_n,B,0)(A',0,-A'\alpha)=(I_n,ABA',0).$$

D. Comments on Theorem 2

Since the obvious symmetry group for the n-dimensional free particle is just $E(n)$, Theorem 2 shows that it exhibits higher symmetry for all $n\geq 1$. Note that, due to our restriction to time-independent canonical transformations, the group $IL(H)$ intersects the n-dimensional Galilei group in the subgroup S only. Note also that the set N (defined in Sec. 3D) again forms a subgroup of $L(H)$ [and hence of $IL(H)$] isomorphic to $O(n)$.

It can be shown that $IL(H)$ is generated by the (n^2+n) conserved quantities consisting of the angular momenta (15) (which, as previously, generate N) and the $\frac{1}{2}n(n+3)$ quantities

$$p_ip_j,\ p_i\quad (i,j=1,2,\ldots,n). \tag{28}$$

The p_i generate the q-translations, and the remaining n^2 quantities generate $L(H)$. It can be shown that $L(H)$ is itself a semidirect product $R^{n(n+1)/2}*O(n)$.

5. QUANTUM-MECHANICAL CONSIDERATIONS

A classical mechanical symmetry will go over to quantum mechanics if the corresponding symmetry group G can be mapped isomorphically onto a group of unitary operators which commute with the quantum-mechanical Hamiltonian. In particular the generators of G must map isomorphically onto a Lie algebra of skew-adjoint operators which commute with the quantum-mechanical Hamiltonian, although this by itself

will not be sufficient as the corresponding *group* will still have to be constructed (see Ref. 17). However, even the weaker step of constructing the Lie algebra of operators cannot always be carried out (see Refs. 12, 18).

If the classical generators are all at most quadratic in the canonical variables, they *can* be mapped isomorphically onto a suitable Lie algebra of skew-adjoint operators, and the corresponding unitary group can be constructed (see Refs. 12, 18). We therefore expect [cf. (15), (16), (28)] that the groups $L(H)$, $IL(H)$ will have quantum-mechanical significance when $\mu \neq 0$, $\mu = 0$ respectively in (3). Now, when $\mu > 0$, the group $SU(n)$ [the connected component of the identity of $L(H) = U(n)$] accounts for the degeneracies (see Ref. 15), but when $\mu \leq 0$ there are no bound states, and so we cannot then invoke degeneracies as a guide.

An analysis of the quantum-mechanical significance of the free particle group $R^{n(n+1)/2} * E(n)$ would be particularly interesting, and for this reason the fact that it has a semidirect product structure seems especially convenient. We emphasize that it is an "invariance" group for the classical system.

We note that such an analysis has already been carried out thoroughly (see Refs. 19, 20) for the Galilei group.

ACKNOWLEDGMENT

The author would like to thank S. Fairthorne for several useful comments.

[1]S. P. Alliluev, Zh. Eksp. Teor. Fiz. **33**, 200 (1957) [Sov. Phys. JETP **6**, 156 (1958)].
[2]G. A. Baker, Phys. Rev. **103**, 1119 (1956).
[3](a) H. Bacry, H. Ruegg, and J-M. Souriau, Commun. Math. Phys. **3**, 323 (1966). (b) H. H. Rogers, J. Math. Phys. **14**, 1125 (1973).
[4]V. A. Dulock, H. V. McIntosh, Am. J. Phys. **33**, 109 (1965).
[5]M. Y. Han and P. Stehle, Nuovo Cimento A **48**, 180 (1966).
[6]D. M. Fradkin, Prog. Theor. Phys. (Kyoto) **37**, 798 (1967).
[7]V. A. Dulock and H. V. McIntosh, Pacific J. Math. **19**, 39 (1966).
[8]N. Mukunda, Phys. Rev. **155**, 1383 (1967).
[9]L. S. Wollenberg, Nuovo Cimento B **57**, 218 (1968).
[10]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950).
[11]L. van Hove, Acad. Roy. Belg. Classe Sci. Mem. Coll. 8° **26** (6), 1 (1951).
[12]G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[13]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[14]The fact that $Sp(n, R)$ contains a subgroup isomorphic to $U(n)$ has been exploited by M. E. Major, "Some Applications of the Theory of Group Representations in Mathematical Physics", Ph.D. thesis, Oxford (1972).
[15]J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).
[16]G. W. Mackey "Group Representations and Non-Commutative Harmonic Analysis," Berkeley (California) lecture notes (1965).
[17]E. Nelson, Ann. Math. **70**, 572 (1959).
[18]R. Arens and D. Babbitt, J. Math. Phys. **6**, 1071 (1965).
[19]J-M. Levy-Leblond, J. Math. Phys. **4**, 776 (1963).
[20]H. P. Doebner and O. Melsheimer, J. Math. Phys. **9**, 1638 (1968).

Addendum: Determination of the amplitude from the differential cross section by unitarity [J. Math. Phys. 9, 2050 (1968)]

Roger G. Newton

Physics Department, Indiana University, Bloomington, Indiana 47401
(Received 14 February 1975)

A. Martin has brought an early reference to my attention which both he and I were unaware of when we wrote our papers[1,2]: N.P. Klepikov, Zh. Eksp. Teor. Fiz. **47**, 757 (1964) [Sov. Phys. JETP **20**, 505 (1965)]. Klepikov presented a uniqueness condition similar to those of Martin and myself, and I apologize for the omission of a reference to his work.

[1]R.G. Newton, J. Math. Phys. **9**, 2050 (1968).
[2]A. Martin, Nuovo Cimento A **59**, 131 (1969).

Erratum: A minimum principle for von Neumann's equation [J. Math. Phys. 16, 158 (1975)]

Sidney Golden

Department of Chemistry, Brandeis University, Waltham, Massachusetts 02154
(Received 18 February 1975)

A typographical omission from the above paper: the last equation should be

$$\Delta_{\rho_0}(t_0) = \text{minimum}. \qquad (4.15)$$

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) *after the page number indicates an Erratum.*