The $\lambda\phi_3^4$ Euclidean quantum field theory in a periodic box

Yong Moon Park*

Department of Mathematics, Yonsei University, Seoul, Korea
(Received 5 May 1975)

The ultraviolet cutoff (the lattice cutoff) normalized Schwinger functions converge as the ultraviolet cutoff (the lattice cutoff) is removed. The limit Schwinger functions are the moments of the normalized physical measure. As a consequence of the lattice approximation, the Lee–Yang theorem and various correlation inequalities hold for the $\lambda\phi_3^4$ field theory in a periodic box.

1. INTRODUCTION AND MAIN RESULTS

This paper studies the $\lambda\phi_3^4$ quantum field theory in a periodic box. The main progress on the $\lambda\phi_3^4$ model has been the proof of the existence [G1][1] and semiboundedness [GJ1][2] of the spatially cutoff Hamiltonian and the proof of the convergence of the ultraviolet cutoff [F1][3] and the lattice cutoff [P1][4] unnormalized Schwinger functions with the free boundary condition as the cutoffs are removed. It also has been proven that for small coupling constant (depending on the spatial cutoff) the normalized Schwinger functions exist for the free boundary condition [F1, P1].[3,4] The main purpose of this paper is to demonstrate that the ultraviolet cutoff (the lattice cutoff) normalized Schwinger functions in a periodic box converge as the ultraviolet cutoff (in respect to the lattice cutoff) is removed. To do this, it is desirable to show that the corresponding partition function does not vanish for $\lambda \in R^+$. It then follows that the Lee–Yang theorem and the various correlation inequalities hold for the given model as a consequence of the lattice approximation of the boson field theory [GRS1, L1, S1].[5–7] Then one may develop the $\lambda\phi_3^4$ field theory parallel to the $P(\phi)_2$ model. The next step in the program might involve the use of methods developed for $P(\phi)_2$ [GJS1, 2][8,9] to take the infinite volume limit of the periodic box and verify the Wightman axioms. (See the remark at the end of this section.)

We will be concerned solely with the Euclidean approach to the $\lambda\phi_3^4$ theory for $\lambda \in R^+$ in a periodic box. Let $\Lambda \subset R^3$ be a box of volume $|\Lambda| = \Pi_{i=0}^2 L^{(i)}$ centered at the origin. We define $\Delta_{\Lambda,\delta}$ to be the lattice approximation of the Laplacian with the periodic boundary condition on Λ and the lattice spacing parameter δ. Let $T_{\Lambda,\delta}$ be the torus obtained by identifying the lattices of the opposite sides of Λ and let $T_\Lambda = T_{\Lambda,\delta=0}$. Throughout this paper we fix the box Λ and supress Λ in the notation. Let dq_δ^0 be the Gaussian measure of mean zero and covariance $(-\Delta_\delta + m_0^2)^{-1}$, where m_0 is the free mass of the boson under consideration. Let $\phi_\delta(f)$, $f \in \mathcal{S}(T_\Lambda)$, be the corresponding free fields and let $\phi_{\kappa,\delta}(f)$, $f \in \mathcal{S}(T_\Lambda)$, be the double cutoff (the ultraviolet cutoff function κ and the lattice cutoff δ) free fields. Let $Z_{\kappa,\delta}$ and $S^{un}_{\kappa,\delta}(f_1, \ldots, f_n)$ be the corresponding partition function and unnormalized Schwinger functions given by

$$Z_{\kappa,\delta} = \int d\bar{q}_{\kappa,\delta},$$
$$S^{un}_{\kappa,\delta}(f_1, \ldots, f_n) = \int \phi_\delta(f_1) \cdots \phi_\delta(f_n)\, d\bar{q}_{\kappa,\delta}, \tag{1.1}$$

where

$$d\bar{q}_{\kappa,\delta} = \exp[-V(\kappa, \delta)]\, dq_\delta^0 \tag{1.2}$$

is the triple cutoff (including the space cutoff Λ) unnormalized intraction measure for the $\lambda\phi_3^4$ model. See Sec. 2 for the detailed definitions. The corresponding normalized Schwinger functions are defined by

$$S_{\kappa,\delta}(f_1, \ldots, f_n) = (Z_{\kappa,\delta})^{-1} S^{un}_{\kappa,\delta}(f_1, \ldots, f_n). \tag{1.3}$$

The ultraviolet cutoff Schwinger functions and the lattice cutoff Schwinger functions are defined by

$$\begin{aligned} S_\kappa(f_1, \ldots, f_n) &= S_{\kappa,\delta=0}(f_1, \ldots, f_n), \\ S_\delta(f_1, \ldots, f_n) &= S_{\kappa=1,\delta}(f_1, \ldots, f_n), \end{aligned} \tag{1.4}$$

respectively. The above expressions are well defined by virtue of momentum cutoffs κ and δ. We now give the main results in this paper.

Theorem 1.1: (a) Let $f_i \in \mathcal{S}(T_\Lambda)$. There exists a constant $K_1(\lambda, \Lambda, m_0^2)$ independent of κ and a Schwartz space norm $|\cdot|$ such that

$$|S_\kappa(f_1, \ldots, f_n)| \leq n! \prod_{i=1}^n |f_i| \exp(K_1).$$

(b) For all $\lambda \in R^+$, the limits

$$S(f_1, \ldots, f_n) = \lim_{\kappa \to 1} S_\kappa(f_1, \ldots, f_n)$$

exist and obey the above bounds.

(c) There exists a unique measure dq on $\mathcal{S}'(T_\Lambda)$ such that

$$S(f_1, \ldots, f_n) = \int \phi(f_1) \cdots \phi(f_n)\, dq.$$

Theorem 1.2: (a) Let $f_i \in \mathcal{S}(T_\Lambda)$. Then for all $\lambda \in R^+$

$$S_\delta(f_1, \ldots, f_n) \to S(f_1, \ldots, f_n) \quad \text{as } \delta \to 0.$$

(b) (Lee–Yang theorem for the $\lambda\phi_3^4$ field theory.) Let $f \geq 0$ and let

$$F(\mu f) = \int \exp[\mu \phi(f)]\, dq.$$

Then $F(\mu f) \neq 0$ if $\mathrm{Re}\ \mu \geq 0$.

(c) (Correlation inequalities for the $\lambda\phi_3^4$ field theory.) The GKS inequalities ([S1],[7] Theorem VIII. 15), the FKG inequalities ([S1],[7] Theorem VIII. 18), and the Lebowitz inequalities ([S1],[7] Theorem IX. 16 and [L1][6]) hold for the $\lambda\phi_3^4$ field theory with the periodic boundary conditions.

The reader is referred to [F1],[3] [GJ1],[2] [GRS1],[5] [P1],[4] and [S1][7] for further background materials, notation, and references. Throughout this paper we will adopt these notation and results.

Remark: After completing this paper I have learned

that both Magnen and Seneor[10] and Feldman and Osterwalder[11] have independently shown the existence of the infinite volume limit of the $\lambda\phi_3^4$ model for sufficiently small λ and sufficiently large bare mass m_0.

2. NOTATION, DEFINITIONS, AND BASIC ESTIMATES

In this section we introduce more detailed notation and definitions. At the end of section we collect technical lemmas which will be used in the next section. For technical reasons we assume that each side of Λ has integral length: $L^{(i)} \in Z^+$, $i = 0, 1, 2$. The same results in more general cases hold by a straightforward modification of the method in the above case. Let Δ_Λ be the Laplacian on the torus T_Λ and let dq^0 be the Gaussian measure on $\mathcal{S}'(T_\Lambda)$ of mean zero and covariance $(-\Delta_\Lambda + m_0^2)^{-1}$. We note that

$$(-\Delta_\Lambda + m_0^2)^{-1}(x-y) = \frac{1}{(2\pi)^3}\int \mu(k_\Lambda)^{-2}\exp[-ik_\Lambda\cdot(x-y)]\,dk$$
$$= \frac{1}{|\Lambda|}\sum_{k_\Lambda \in Z^3_\Lambda} \mu(k_\Lambda)^{-2}\exp[-ik_\Lambda\cdot(x-y)],$$

where k_Λ is the lattice point in Z^3 close to k,

$$Z^3_\Lambda = \left(\frac{2\pi}{L^{(0)}}\right) Z \times \left(\frac{2\pi}{L^{(1)}}\right) Z \times \left(\frac{2\pi}{L^{(2)}}\right) Z \tag{2.2}$$

and $\mu(k)^2 = (k^2 + m_0^2)$. The free theory with periodic boundary condition on Λ is given on the path space $L^2(\mathcal{S}'(T_\Lambda), dq^0)$. The Euclidean fields are the linear coordinate functions on $\mathcal{S}'_R(Z_\Lambda)$:

$$\phi_\Lambda(f)(q) = \langle q, f\rangle \quad \text{for all } q \in \mathcal{S}'_R(T_\Lambda) \text{ and } f \in \mathcal{S}(T_\Lambda).$$

For $h \in L^\infty(Z^3_\Lambda)$ we write

$$\hat{h}(x) = \int h(k_\Lambda)\exp(ik_\Lambda\cdot x)\,dk. \tag{2.3}$$

We now introduce the ultraviolet cutoff free field by

$$\phi_{\Lambda,\kappa}(x) = \frac{1}{|\Lambda|}\int_\Lambda \phi_\Lambda(y)\,\hat{\kappa}_\Lambda(x-y)\,dy. \tag{2.4}$$

We assume that the cutoff function κ is of the form as in [GJ1].[2]

From now on we suppress Λ and m_0^2 in the notation. The partition function and unnormalized Schwinger functions of the double cutoff interaction theory are just the mass and moments of the unnormalized measure $d\bar{q}_\kappa$:

$$Z_\kappa = \int d\bar{q}_\kappa,$$
$$S^{\mathrm{un}}_\kappa(f_1,\ldots,f_n) = \int \phi(f_1)\cdots\phi(f_n)\,d\bar{q}_\kappa. \tag{2.5}$$

The measure is given by [F1, GJ1][3,2]

$$d\bar{q}_\kappa = \exp[-V(\kappa)]\,dq^0,$$
$$V(\kappa) = V_I(\kappa) - \tfrac{1}{2}\lambda^2\delta m_\kappa^2 :\phi_\kappa^2: + E_2(\kappa) + E_3(\kappa),$$
$$V_I(\kappa) = \lambda :\phi_\kappa^4:, \tag{2.6}$$
$$E_2(\kappa) = \tfrac{1}{2}\int V_I(\kappa)^2\,dq^0, \quad E_3(\kappa) = -\tfrac{1}{6}\int V_I(\kappa)^3\,dq^0,$$
$$\delta m_\kappa^2 = -4^2\cdot 6\cdot(2\pi)^{-9}\int \delta(k_2+k_3+k_4)\times\prod_{i=2}^4 \mu(k_{\Lambda,i})\,\kappa(k_{\Lambda,i})\,d^3k_i.$$

Here : : means Wick ordering with respect to dq^0 and

$$:\phi_\kappa^n: = \int_\Lambda :\phi_\kappa^n:(x)\,dx. \tag{2.7}$$

We next consider the lattice approximation with the periodic boundary condition on Λ. We assume that the the lattice spacing parameter δ has the form $\delta = 2^{-n}$ for some $n \in Z^+$. From the assumption on Λ it follows that

$$L^{(i)}/\delta = 2^n m^{(i)} \text{ for some } m^{(i)} \in Z^+. \tag{2.8}$$

We consider the finite Fourier transformation from $l^2(\Lambda_\delta)$ to $l^2(Z^3_{\Lambda,\delta})$ where $\Lambda_\delta = \{n\delta \mid n\delta \in \Lambda, n \in Z^3\}$ and $Z^3_{\Lambda,\delta} = \{k_\Lambda \mid k_\Lambda \in Z^3_\Lambda, k_\Lambda^{(i)} \in (-\pi/\delta, \pi/\delta]\}$. Then the finite Fourier transformation is defined by

$$\tilde{h}_\delta(k_\Lambda) = (\delta^3/|\Lambda|)\sum_{n\delta\in\Lambda_\delta} h(n\delta)\exp(-ik_\Lambda\cdot n\delta). \tag{2.9}$$

Similarly we define the inverse transformation by

$$\hat{h}(n\delta) = \frac{1}{(2\pi)^3}\int_\delta h(k_\Lambda)\exp(ik_\Lambda\cdot n\delta)\,dk,$$

where $\int_\delta$ means that the range of integration is $k^{(i)} \in (-\pi/\delta, \pi/\delta)$. We define the Laplacian $\Delta_{\Lambda,\delta}$ on $T_{\Lambda,\delta}$, where $T_{\Lambda,\delta}$ is the discrete torus, by

$$(-\Delta_{\Lambda,\delta}f)(n\delta) = \delta^{-2}[6f(n\delta) - \sum_{|n-n'|=1} f(n'\delta)].$$

Then the image of $(-\Delta_{\Lambda,\delta} + m^2)$ on $l^2(Z^3_{\Lambda,\delta})$ is the multiplication by

$$\mu_\delta(k_\Lambda)^2 = \delta^{-2}\left(6 - 2\sum_{i=0}\cos(\delta k_\Lambda^{(i)})\right) + m_0^2. \tag{2.10}$$

The free fields $\phi_\delta(f) = \delta^3\sum_{n\delta\in\Lambda_\delta}\phi_\delta(n)f(n\delta)$ are Gaussian random variables with mean zero and covariance $(-\Delta_{\Lambda,\delta} + m_0^2)^{-1}$. Let dq^0_δ be the measure corresponding to the above covariance and let $d\bar{q}_\delta$ be the unnormalized interacting measure given by

$$d\bar{q}_\delta = \exp[-V(\delta)]\,dq^0_\delta, \tag{2.11}$$

where $V(\delta)$ is the lattice cutoff interaction defined by replacing $:\phi_\kappa^n:$ by

$$:\phi^n_\delta: = \delta^3\sum_{n\delta\in\Lambda_\delta} :\phi^n_\delta:(n) \tag{2.12}$$

and δm_κ^2 by δm_δ^2, where δm_δ^2 is defined by replacing k by k_Λ in the definition of δm_δ^2 in [P1].[4] The lattice cutoff partition function and unnormalized Schwinger functions are given by $Z_\delta = \int d\bar{q}_\delta$ and

$$S^{\mathrm{un}}_\delta(f_1,\ldots,f_n) = \int \phi_\delta(f_1)\cdots\phi_\delta(f_n)\,d\bar{q}_\delta. \tag{2.13}$$

Similarly we define the triple cutoff unnormalized interacting measure by

$$d\bar{q}_{\kappa,\delta} = \exp[-V(\kappa,\delta)]\,dq^0_\delta, \tag{2.14}$$

where $V(\kappa,\delta)$ is defined by replacing $\mu_\delta(k_\Lambda)^{-1}$ by $\mu_\delta(k_\Lambda)^{-1}\kappa(k_{\Lambda,\delta})$ in the definition of $V(\delta)$, where

$$k^{(i)}_{\Lambda,\delta} = 2\delta^{-1}\sin(\delta k^{(i)}/2), \quad i = 0, 1, 2. \tag{2.15}$$

The corresponding partition function and Schwinger functions are defined by

$$Z_{\kappa,\delta} = \int d\bar{q}_{\kappa,\delta},$$
$$S^{\mathrm{un}}_{\kappa,\delta}(f_1,\ldots,f_n) = \int \phi_\delta(f_1)\cdots\phi_\delta(f_n)\,d\bar{q}_{\kappa,\delta}. \tag{2.16}$$

In the remainder of this section we collect some technical lemmas which we will use in the next section.

Lemma 2.1: (a) For each $k \in R^3$

$$\mu_\delta(k_\Lambda) \to \mu(k_\Lambda), \kappa(k_{\Lambda,\delta}) \to \kappa(k_\Lambda) \text{ as } \delta \to 0.$$

(b) $\mu_\delta(k_\Lambda)^{-1} \leq (\pi/2)\,\mu(k_\Lambda)^{-1}$, if $|k^{(i)}| \leq \pi/\delta$, $0 \leq i \leq 2$.

(c) $[\mu(k)/\mu(k_\Lambda)]^{\pm 1} \leq O(1)$, if $L^{(i)} \geq 1$, $0 \leq i \leq 2$.

Proof: (a) and (c). These follow from an elementary observation. (b) This follows from the following inequality [GRS1,[5] (IV. 11)]:

$$1 - \cos y \geq (2/\pi)\, y^2 \quad \text{if } y \in [-\pi, \pi]. \blacksquare \tag{2.17}$$

Lemma 2.2: Let $f \in \mathcal{S}(T_\Lambda)$. Then for $|k^{(i)}| \leq \pi/\delta$

$$|\tilde{f}_\delta(k_\Lambda)| \leq O(1)\ \mu(k)^{-2}.$$

Proof: From the definitions in (2.9) and (2.10) we have that

$$|\mu_\delta(k_\Lambda)^2 \tilde{f}_\delta(k_\Lambda)| \leq O(1)(\delta/|\Lambda|) \sum_{n\delta \in \Lambda_\delta} \Big| \delta f(n) - \sum_{|n-n'|=1} f(n'\delta) \Big|.$$

We use the method used in proving [P1][4] Lemma 2.2 to bound the above by $O(1)$. The lemma follows from Lemma 2.1 (c). ■

In the next section we will divide the periodic box Λ into union of cubes. To prevent double counts of lattice points on the boundary of cubes, we consider half open cubes of the form

$$\Lambda \supset \Delta = \{x \,|\, x^{(i)} \in [a^{(i)}, b^{(i)}),\ |a^{(i)} - b^{(i)}| = d,\ 0 \leq i \leq 2\}. \tag{2.18}$$

We write

$$(\tilde{\chi}_\Delta)_\delta(k_\Lambda) = (\delta^3/|\Lambda|) \sum_{n\delta \in \Delta} \exp(-ik_\Lambda \cdot n\delta)$$

$$F_{\delta,\Delta}(k_\Lambda) = \prod_{i=0}^{2} [|\Delta|^{1/3} \mu_\delta(k_\Lambda^{(i)}) + 1]^{-1} \tag{2.19}$$

$$\mu_\delta(k_\Lambda^{(i)})^2 = \delta^{-2}[2 - 2\cos(\delta k_\Lambda^{(i)})] + m_0^2, \quad 0 \leq i \leq 2,$$

where $|\Delta|$ is the volume of Δ. We also introduce a discrete version of derivatives with respect to k_Λ variables:

$$D_\Lambda^m = \prod_{i=0}^{2} D_{\Lambda,i}^{m(i)}, \quad m = (m^{(0)}, m^{(1)}, m^{(2)})$$

$$(D_{\Lambda,0} f)(k_\Lambda) = (L^{(0)}/2\pi)[f(k^{(0)} + \pi/L^{(0)}, k^{(1)}, k^{(2)}) - f(k^{(0)} - \pi/L^{(0)}, k^{(1)}, k^{(2)})], \tag{2.20}$$

and $D_{\Lambda,1}$ and $D_{\Lambda,2}$ are defined by a similar manner. The following is the result corresponding to [P1],[4] Lemma Lemma 2.3.

Lemma 2.4: (a) For $|k^{(i)}| \leq \pi/\delta$

$$|D_\Lambda^m \mu_\delta(k_\Lambda)^{-2}| \leq O(1)\ \mu(k)^{-2-|m|}.$$

(b) We assume that $|\Delta| \leq 1$ and the center of Δ is at the origin:

$$|D_\Lambda^m (\tilde{\chi}_\Delta)_\delta(k_\Lambda)| \leq O(1)\,|\Delta|^{1+|m|/3} F_{\delta,\Delta}(k_\Lambda).$$

(c) Let κ be the momentum cutoff function defined in (2.4):

$$|D_\Lambda^m \kappa(k_{\Lambda,\delta})| \leq O(1) \min\{(\alpha^{(1)})^{-m} \,|\, i = 0, 1, 2\}\ \chi_\kappa,$$

where χ_κ is the characteristic function of the support of $\kappa(k_{\Lambda,\delta})$.

Proof: The proof follows by replacing D^m in the proof of [P1],[4] Lemma 2.3 by D_Λ^m in (2.20) and adapting a method similar to that of [P1],[4] Lemma 2.3. For the detailed proof we refer the reader to [P1].[4] ■

3. CONVERGENCE OF THE SCHWINGER FUNCTIONS

We now turn to prove Theorem 1.1 and Theorem 1.2 by employing a method similar to that developed in [GJ1, F1, P1].[2–4] The notation G may refer, depending on context, to the topological graph G, the function $G(q)$ on $\mathcal{S}'(T_\Lambda)$, or the kernel $G(k_l)$ [F1].[3] Following [F1],[3] we introduce two deferent estimates on $\int G\, d\bar{q}_{\kappa,\delta}$. Given $\gamma > 2\alpha > 0$, we define

$$\|G\|_{1,\gamma,\alpha} = \sup_{P_\alpha^e} \sup_{C} \|P_\alpha^e C M^\gamma |G|\ \|_{H\cdot S}$$

$$\|G\|_{2,\gamma,\alpha} = \sup_{P_\alpha^e} \sup_{C} \sup_{D} \|P_\alpha^e C M^\gamma |DJG|\ \|_{H\cdot S}. \tag{3.1}$$

P_α^e, C, D, J, M, and $|\cdot|$ are "operators" that modify the graph G and its kernel [F1].[3] In our case D is monomial of discrete differential operator in the variables $\{k_l\}$, that is, at most fourth order in $\{k_l^{(0)}, k_l^{(1)}, k_l^{(2)}\}$ for each fixed l. With above notations we obtain that

$$\|\phi(f)\|_{1,\gamma,\alpha} = K_1 \|(-\Delta_\Lambda + m_0^2)^{-1/2+\gamma} f\|_{L^2} \equiv |f|_\gamma,$$

$$\|\phi_\delta(f)\|_{1,\gamma,\alpha} = K_1 \|(-\Delta_{\Lambda,\delta} + m_0^2)^{-1/2+\gamma} f\|_{L^2} \equiv |f_\delta|_\gamma. \tag{3.2}$$

Notice that $|f|_\gamma$ and $|f_\delta|_\gamma$ are finite for $f \in \mathcal{S}(T_\Lambda)$. We use the method of [GJ1][2] for decomposing big graphs into little graphs to obtain

$$\|\prod_{i=1}^{n} \phi(f_i)\|_{1,\gamma,\alpha} \leq n! \prod_{i=1}^{n} |f_i|_\gamma,$$

$$\|\prod_{i=1}^{n} \phi_\delta(f_i)\|_{1,\gamma,\alpha} \leq n! \prod_{i=1}^{n} |f_{i,\delta}|_\gamma. \tag{3.3}$$

For the detailed derivation of the above estimates we refer to [F1].[3]

Theorem 1.1 and Theorem 1.2 will follow as corollaries of the following results:

Theorem 3.1. Suppose G_1 is a graph having N external legs and G_2 is a graph having $N(\Delta)$ external legs in $\Delta \subset \Lambda$. Let κ_1 and κ_2 be the ultraviolet cutoff functions of the form in (2.4) and let $a \geq 0$ be a given constant. Then, for $\delta \geq 0$ there is a constant $K_1(\lambda, \gamma_1, \gamma_2, \alpha, a)$ independent of κ_1, κ_2, and δ such that

$$\Big| \int G_1 G_2 \exp(-a : \phi^2_{\kappa_1,\delta} :)\, d\bar{q}_{\kappa_2,\delta} \Big|$$

$$\leq N^N \prod_{\Delta} N(\Delta)^{N(\Delta)} \|G_1\|_{1,\gamma_1,\alpha} \|G_2\|_{2,\gamma_2,\alpha} \exp[K_1 A(\Lambda)],$$

where $A(\Lambda)$ is the volume of the set of points within a distance one from Λ.

Corollary 3.2:

$$\Big| \int G_1 G_2 \exp[\phi_\delta(f)]\, d\bar{q}_{\kappa,\delta} \Big|$$

$$\leq N^N \prod_{\Delta} N(\Delta)^{N(\Delta)} \|G_1\|_{1,\gamma_1,\alpha} \|G_2\|_{2,\gamma_2,\alpha} \exp[K_2 A(\Lambda)].$$

Theorem 3.3: Let $\|G\|_{1,\gamma,\alpha} < \infty$ and let $|f_\delta|_\gamma < \infty$ for some $0 < \alpha < \alpha_0$ (α_0 to be chosen as in [F1][3]). Then, for $\delta \geq 0$, $\int G \exp[\mu\phi_\delta(f)]\, d\bar{q}_{\kappa,\delta}$ converges uniformly in $\delta \geq 0$ as $\kappa \to 1$ and obeys the bound of Corollary 3.2. The limit is continuous in f and analytic in μ.

Theorem 3.4: The limit

$$\lim_{\kappa_1\to 1}\lim_{\kappa_2\to 1}\int \exp(-a:\phi^2_{\kappa_1,\delta}:)\,d\bar{q}_{\kappa_2,\delta}$$
$$=\lim_{\kappa_2\to 1}\lim_{\kappa_1\to 1}\int \exp(-a:\phi^2_{\kappa_1,\delta}:)\,d\bar{q}_{\kappa_2,\delta}$$

exists.

Theorem 3.5: $Z=\lim_{\kappa\to 1} Z_\kappa > 0$.

The proofs of Theorems 3.1–3.5 are delayed to later in this paper. We now prove Theorem 1.1 and Theorem 1.2.

Proof of Theorem 1.1: (a) We set $G_1=1$ $[G_1=\Pi_{i=1}^n\phi(f_i)]$, $G_2=1$, $a=0$, and $\delta=0$ in Theorem 3.1 to obtain

$$Z_\kappa \le \exp[K_1 A(\Lambda)],$$
$$|S^{\mathrm{un}}_\kappa(f_1,\ldots,f_n)| \le n!\prod_{i=1}^n |f_i|\exp[K_1A(\Lambda)]. \tag{3.4}$$

Since

$$Z=\lim_{\kappa\to 1} Z_\kappa \quad\text{and}\quad S^{\mathrm{un}}(f_1,\ldots,f_n)=\lim_{\kappa\to 1} S^{\mathrm{un}}_\kappa(f_1,\ldots,f_n) \tag{3.5}$$

exist by Theorem 3.3 for the case in which $G=1$ [and $G_2=\Pi_{i=1}^n\phi(f_i)$], $u=0$ and $\delta=0$, it follows that there exists an ultraviolet cutoff function κ_0 such that, for $\kappa\ge\kappa_0$,

$$Z_\kappa \ge \tfrac{1}{2}Z>0 \tag{3.6}$$

by Theorem 3.5 and (3.5). The theorem follows from (3.4) and (3.6).

(b) This follows from (3.5) and (3.6).

(c) This follows from Theorem 3.5 and the argument in the proof of [F1],[3] Theorem 1(d). ■

Proof of Theorem 1.2: (a) From (3.3) and Theorem 3.3 we obtain that

$$S_{\kappa,\delta}(f_1,\ldots,f_n)\to S_\delta(f_1,\ldots,f_n) \text{ uniformly in } \delta. \tag{3.7}$$

Here we have also used Theorem 3.5. Since $S_{\kappa,\delta}\to S_\kappa$ as $\delta\to 0$ by the virtue of momentum cutoff κ, the theorem follows by the 3ϵ argument.

(b) The theorem follows from the Lee–Yang theorem for the lattice cutoff theory [GS1, S1],[12,7] Hurwitz's theorem, and the fact that

$$\int \exp[\mu\phi(f)]\,dq=\lim_{\delta\to 0} Z_\delta^{-1}\int \exp[\mu\phi_\delta(f)]\,d\bar{q}_\delta. \tag{3.8}$$

The proof of (3.8) follows from Theorem 3.3, Theorem 3.5, and a method similar to that used in the proof of Theorem 1.2(a).

(c) Since the corresponding inequalities hold for the lattice cutoff theory [GS1, S1],[12,7] the theorem follows from Theorem 1.2 (a)–(b). ■

In the rest of this section we prove Theorems 3.1–3.5. The main method of the proof rests on the so-called "inductive construction" developed in [GJ1].[2] Using the construction, we will follow the main steps used in [F1, P1][3,4] with a modification which is necessary for the given case. From now on we assume that the allowed cubes have the form

$$\Delta=2^{-j}\Delta_0+2^{-j}n\delta,\quad j\in Z,\ n\in Z^3, \tag{3.9}$$

where $\Delta_0=\{x\,|\,x^{(i)}\in[-\frac{1}{2},\frac{1}{2})\}$. From the assumption on the size of δ it follows that the length of each side of cubes in a multiple of δ. We next consider the ultraviolet cutoff function $\kappa(k_{\Lambda,\delta})$ in more detail. We define λ, u, and U—the maximum lower cutoff, the minimum upper cutoff, and the maximum upper cutoff of a group of legs

$$\lambda=\max_{i,l}\{2,\alpha_l^{(i)}\},\quad u=\min_{i,l}\{\beta_l^{(i)}\},\quad U=\max_{i,l}\{\beta_l^{(i)}\}, \tag{3.10}$$

where $\eta_{\beta,\alpha}(k^{(i)}_{l,\Lambda,\delta})=\eta(k^{(i)}_{l,\Lambda,\delta}/\beta_l^{(i)})-\eta(k^{(i)}_{l,\Lambda,\delta}/\alpha_l^{(i)})$ is the momentum cutoff function in the ith space–time direction for the leg l. We note that the above definition is independent of Λ and δ. Because of the lattice cutoff δ, we may assume that

$$\beta_l^{(i)}\le 2\pi/\delta \text{ for all } i \text{ and } l. \tag{3.11}$$

Let

$$\mathrm{supp}\eta_{\beta,\alpha}(k^{(i)}_{l,\Lambda,\delta})\cap[-\pi/\delta,\pi/\delta]$$
$$=[-\beta^{(i)}_{l,\delta},-\alpha^{(i)}_{l,\delta}]\cup[\alpha^{(i)}_{l,\delta},\beta^{(i)}_{l,\delta}].$$

We define δ-dependent maximum lower cutoff, minimum upper cutoff, and maximum upper cutoff of a group of legs by

$$\lambda_\delta=\max_{i,l}\{2,\alpha^{(i)}_{l,\delta}\},\quad u_\delta=\min_{i,l}\{\beta^{(i)}_{l,\delta}\},\quad U_\delta=\max_{i,l}\{\beta^{(i)}_{l,\delta}\}. \tag{3.12}$$

From Lemma 2.1(c) and (2.17) it is easy to check that

$$(|k^{(i)}|+1\,|\,k^{(i)}_{\Lambda,\delta}|+1)^{\pm 1}\le O(1) \text{ for } |k^{(i)}|\le\pi/\delta.$$

Hence, under the assumption (3.11) we obtain that

$$|\lambda/\lambda_\delta|^{\pm 1}\le O(1),\quad |u/u_\delta|^{\pm 1}\le O(1),\quad |U/U_\delta|\pm 1\le O(1) \tag{3.13}$$

uniformly in δ. Following the main steps in [F1, P1],[3,4] we now summarize our proofs.

Sketch of the proof of Theorem 3.1: In [P1][4] we have proved the corresponding theorem with the free boundary condition in the case in which $a=0$. Therefore, we will only point out where we must pay special attention to the periodic boundary condition and to the case of $a>0$. Since $a\ge 0$ and since $\|:\phi^2_{\kappa,\delta}:\|_2\le$ const uniformly in κ and δ, the form $\exp(-a:\phi^2_{\kappa_1,\delta}:)$ introduces no difficulties in the inductive construction. Hence we only consider the case of $a=0$ for the simplification of our discussion. We follow the same steps used in the proof [P1],[4] Theorem 3.1.

Step 1: *The inductive expansion*: By Lemma 2.1 we may employ the expansion of [GJ1][2] to obtain

$$\left|\int G_1G_2\,d\bar{q}_{\kappa,\delta}\right|\le\sum_G I(G),$$

where $I(G)$ is the elementary integration labeled by the Feynman graph G.

Step 2: *The combinatoric estimates*: The combinatoric bounds given in [F1],[3] Lemma 4.1 [with notations in (3.10)] apply equally well to our case. To show this, we note that

$$d(\Delta,\Delta'')\le 4d(\Delta,\Delta')\,d(\Delta',\Delta''), \tag{3.14}$$

where $d(\Delta,\Delta')=1+\gamma(\Delta,\Delta')$ and $\gamma(\Delta,\Delta')$ is the distance from the center of Δ to the center of Δ' in the forus T_Λ, and so a result similar to that in [GJ1],[2] p. 338, holds in our case (if necessary, a redefinition of c in [GJ1][2] does not essentially effect the combinatoric estimates). Also the results in [GJ1],[2] Sec. 3.3, and [GJ1],[2] Lemma 4.2, holds in our case. Using the combinatoric estimates, we have that

$$\sum_G I(G) \le \sup_G c(G)\,|I(G)|,$$

where $c(G)$ is the combinatoric coefficients given above.

Step 3: *Localization factors*: We note that for $|n^{(i)}\delta| \le L^{(i)}/2\pi$,

$$(n\delta)^2 \le \frac{\pi}{2}\sum_{i=0}^{2}\left(\frac{L^{(i)}}{2\pi}\right)^2\left[2-2\cos\left(\frac{2\pi}{L^{(i)}}\,n^{(i)}\delta\right)\right],$$

$$\left(\frac{L^{(i)}}{2\pi}\right)^2\left[2-2\cos\left(\frac{2\pi}{L^{(i)}}\,n^{(i)}\delta\right)\right]\exp(ik_\Lambda\cdot n\delta)$$
$$=D^2_{\Lambda,i}\exp(ik_\Lambda,n\delta),$$

where $D_{\Lambda,i}$ is the discrete differential operator in $T_{\Lambda,\delta}$. We use the above results, the periodic property of $k_{\Lambda,\delta}$ (period $2\pi/\delta$) and a method similar to that of [P1].[4] We then isolate the distance factors to obtain

$$|I(G)| \le \Big[\prod_{\text{lines }l}(\alpha_l d)^{-m}\Big]\bar{I}(G)$$

Step 4: *Estimate of* $\bar{I}(G)$: We assert that $\bar{I}(G)$ is bounded by a product of factors given by those of [F1],[3] Lemma 5.1 [with notations in (3.10)]. In [P1][4] the above assertion has followed as a consequence of [P1], Lemmas 2.1–2.3, the corresponding estimate of (3.13) and the periodic property of $F_{\delta,\Delta}(k)$ (see the Appendix of [P1][4] for the detailed discussion). Similary the above assertion follows as a consequence of Lemmas 2.1–2.3, (3.13), and the periodic property of $F_{\delta,\Delta}(k_\Lambda)$.

The theorem now follows from the assertion and the method used in proving [F1],[3] Theorem 2, from [F1],[3] Lemma 5.1. ■

Proof of Corollary 3.2. *and Theorem* 3.3: The proof follow from the methods used in the proof of [F1],[3] Corollary 3.2, and [F1],[3] Theorem 3, by using [F1],[3] Theorem 3.2, and from the method used in proving Theorem 3.1. For more detailed discussion we refer the reader to [F1].[3] ■

Proof of Theorem 3.4: Using Theorem 3.1 and using a method similar to that in the proof of [F1],[3] Theorem 3, we conclude that $\int \exp(-a:\phi^2_{\kappa_1}:)\,d\bar{q}_{\kappa_2}$ converges uniformly in κ_1 as $\kappa_2 \to 1$. Since $\lim_{\kappa_2\to 1}\lim_{\kappa_1\to 1}\int \exp(-a:\phi^2_{\kappa_1}:)\,d\bar{q}_{\kappa_2}$ exists as consequences of Theorem 3.1, the theorem follows by the 3ϵ argument. ■

To show Theorem 3.5, we need technical lemmas.

Lemma 3.6: There exists a constant $\bar{m}$ such that, for $m_0 > \bar{m}$, $Z(m_0^2) > 0$, where $Z(m_0^2)$is the partition function corresponding to the free boson mass m_0.

Proof of Lemma 3.6: We construct a sequence of momentum cutoffs $0=\kappa_0 \le \kappa_1 \le \cdots \le \kappa_M = \kappa$. We then arrive at

$$|Z_\kappa - 1| \le |Z_\kappa - Z_{\kappa_0}|$$
$$\le \sum_{i=0}^{M-1}\sum_{\sigma\in\bar{G}}\int_0^1 ds\,\Big|\int G_{\sigma,i}\,d\bar{q}_{\kappa(s)}\Big|$$
$$\le \sup_{i,\sigma}\|G_{\sigma,i}\|_{1,\gamma,\alpha}\exp[K_5 A(\Lambda)],$$

where each graph $G_{\sigma,i}$ contains one P vertex, and at most 16 C vertices, $\bar{G}$ is a finite index set, and K_5 is a constant independent of m_0^2. The last inequality follows from the method used in proving [F1],[3] Theorem 3. Since each kernel of $G_{\sigma,i}$ contains $\mu(k_\Lambda,m_0^2)^{-2} = (k_\Lambda^2+m_0^2)^{-1}$ factors, it is easy to show that

$$\|G_{\sigma,i}\|_{1,\gamma,\alpha} \to 0 \quad \text{as } m_0^2 \to \infty.$$

The theorem follows from the above result. ■

We will show Theorem 3.5 by first assuming $Z(m_0^2) = 0$ for some $m_0^2 > 0$ and then making a contradiction to Lemma 3.6. We first introduce more notation. We write

$$d\bar{q}_{\kappa,\delta}(m_1^2,m_2^2) = \exp[-V(\kappa,\delta,m_1^2)]\,dq^0_\delta(m_2^2), \tag{3.15}$$

where $dq^0(m_2^2)$ is the free measure with covariance $(-\Delta_\delta+m_2^2)^{-1}$ and

$$V(\kappa,\delta,m_1^2) = \lambda:\phi^4_{\kappa,\delta}:_{m_1^2} + E_2(\kappa,\delta,m_1^2) + E_3(\kappa,\delta,m_1^2) - \tfrac{1}{2}\lambda^2\delta m^2_{\kappa,\delta}(m_1^2):\phi^2_{\kappa,\delta}:_{m_1^2} \tag{3.16}$$

where $:\ :_{m_1^2}$ is the Wick ordering with respect to the measure $dq^0(m_1^2)$, and $E_2(\kappa,\delta,m_1^2)$, $E_3(\kappa,\delta,m_1^2)$, and $\delta m^2_{\kappa,\delta}(m_1^2)$ are obtained by replacing m_0^2 by m_1^2 in the definitions of $E_2(\kappa,\delta)$, $E_3(\kappa,\delta)$, and $\delta m^2_{\kappa,\delta}$. We also write

$$Z_{\kappa_1,\kappa_2,\delta}(a,m_1^2,m_2^2) = \int \exp(-a:\phi^2_{\kappa_1,\delta}:_{m_1^2})\,d\bar{q}_{\kappa_2,\delta}(m_1^2,m_2^2),$$
$$Z_{\kappa_1,\kappa_2}(a,m_1^2,m_2^2) = Z_{\kappa_1,\kappa_2,\delta=0}(a,m_1^2,m_2^2),$$
$$Z_\delta(a,m_1^2,m_2^2) = Z_{\kappa_1=\kappa_2=1,\delta}(a,m_1^2,m_2^2), \tag{3.17}$$
$$Z(a,m_1^2,m_2^2) = Z_{\kappa_1=\kappa_2=1}(a,m_1^2,m_2^2).$$

We first verify the following technical lemma.

Lemma 3.7: There are constants $b_1(\kappa)$ and $b_2(\kappa)$ such that

$$V(\kappa,m_1^2) = V(\kappa,m_2^2) + b_1(\kappa):\phi^2_\kappa:_{m_2^2} + b_2(\kappa)$$

and

$$|b_1(\kappa)| + |b_2(\kappa)| < \infty \text{ for all } \kappa.$$

Proof: We first consider the change in Wick polynomials. It is not hard to check that

$$:\phi^4_\kappa:_{m_1^2} = :\phi^4_\kappa:_{m_2^2} + 6\delta C_\kappa(m_2^2,m_1^2):\phi^2_\kappa:_{m_2^2} + 3[\delta C_\kappa(m_2^2,m_1^2)]^2|\Lambda|,$$
$$:\phi^2_\kappa:_{m_1^2} = :\phi^2_\kappa:_{m_2^2} + \delta C_\kappa(m_2^2,m_1^2)|\Lambda|, \tag{3.18}$$

where

$$\delta C_\kappa(m_2^2,m_1^2) = C_\kappa(m_2^2) - C_\kappa(m_1^2),$$
$$C_\kappa(m^2) = [1/(2\pi)^3]\int \kappa^2(k_\Lambda+m^2)^{-1}\,d^3k.$$

If we expand $V(\kappa,m_1)$ by using (3.18), it then easy to check that

$$V(\kappa,m_1) - V(\kappa,m_2) = b_1(\kappa):\phi^2_\kappa:_{m_2^2} + b_2'(\kappa) + [E_2(\kappa,m_1) - E_2(\kappa,m_2) - \tfrac{1}{2}\lambda^2\delta m^2_\kappa(m_1)\delta C_\kappa(m_2,m_1)]|\Lambda|, \tag{3.19}$$

where $|b_1(\kappa)| + |b_2'(\kappa)| <$ const uniformly in κ. We now assert that the second term of (2.19) is bounded uniformly in κ. The proof follows by the method used in the mass renormalization cancellation (and its straightforward modification). We leave the detailed proof to the reader. Combining (3.19) and the assertion, we have proved the lemma. ■

Proof of Theorem 1.6: (i) In the case of $b_1(\kappa) \geq 0$ in Lemma 3.7: We assume that $Z(m_0^2) = 0$ for some m_0. Since $\exp(-a:\phi_\kappa^2:_{m_2^2}) \leq \text{const}(\kappa)$, we conclude that

$$\lim_{\kappa_2 \to 1} Z_{\kappa_1,\kappa_2}(a, m_0^2, m_0^2) \to 0 \tag{3.20}$$

by the assumption. Hence by Theorem 3.4 it follows that that

$$\lim_{\kappa_1 \to 1} \lim_{\kappa_2 \to 1} Z_{\kappa_1,\kappa_2}(a, m_0^2, m_0^2) = Z(a, m_0^2, m_0^2) = 0. \tag{3.21}$$

Since $Z(a, m_0^2, m_0^2) = \lim_{\delta \to 0} Z_\delta(a, m_0^2, m_0^2)$ by a method similar to that used in proving Theorem 1.2(a), and since [Sec. 3 of SP1],[13]

$$C_\delta^{-1} Z_\delta(a, m_0^2, m_0^2) = Z_\delta(0, m_0^2, m_0^2 + 2a),$$

by a consequence of the lattice approximation, where $C_\delta = \int \exp(-a:\phi_\delta^2:) dq_\delta^0 > 1$ uniformly in δ, it follows that

$$C^{-1} Z(a, m_0^2, m_0^2) = Z(0, m_0^2, m_0^2 + 2a) = 0, \tag{3.22}$$

where $C = \lim_{\delta \to 0} C_\delta$. We use Theorem 3.4 and Lemma 3.7 to conclude that

$$Z(0, m_0^2 + 2a, m_0^2 + 2a) = dZ(b_1, m_0^2, m_0^2 + 2a), \tag{3.23}$$

where $d = \exp[-b_2(1)]$ and $b_1 = b_1(1)$. We again use (3.22) and a method similar to that in proving (3.21) to obtain $Z(b_1, m_0^2, m_0^2 + 2a) = 0$ and so $Z(0, m_0^2 + 2a, m_0^2 + 2a) = 0$ for any $a \geq 0$ by (3.23). This contradicts Lemma 3.6.

(ii) In the case of $b_1(\kappa) < 0$ in Lemma 3.7: Replacing a by $a - b_1$ in (3.21) and following the method used above, we obtain

$$0 = C^{-1} Z(a - b_1, m_0^2, m_0^2) = Z(-b, m_0^2, m_0^2 + 2a)$$
$$= d_1 Z(0, m_0^2 + 2a, m_0^2 + 2a).$$

Hence it follows that $Z(0, m_0^2 + 2a, m_0^2 + 2a) = 0$. This contradicts Lemma 3.6. ■

ACKNOWLEDGMENTS

The author wishes to thank Professor J. Glimm and Dr. T. Spencer for many enlightening discussions and helpful suggestions. I also would like to express my appreciation for their hospitality at the Courant Institute of Mathematical Science, N. Y. U., during the summer of 1974.

*Supported by Korean Traders Scholarship Foundation.

[1] [G1] J. Glimm, Commun. Math. Phys. **10**, 1 (1968).
[2] [GJ1] J. Glimm and A. Jaffe, Fortschr. Physik **21**, 327 (1973).
[3] [F1] J. Feldman, Comm. Math. Phys. **37**, 94 (1974).
[4] [P1] Y. Park, J. Math. Phys. **16**, 1065 (1975).
[5] [GRS1] F. Guerra, L. Rosen, and B. Simon, "The $P(\phi)_2$ Euclidean quantum field theory as classical statistical mechanics," to appear in Ann. Math.
[6] [L1] J. Lebowitz, Comm. Math. Phys. **35**, 87 (1974).
[7] [S1] B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, N. J., 1974).
[8] [GJS1] J. Glimm, A. Jaffe and T. Spencer, Ann. Math. **100**, No. 3 (1974).
[9] [GJS2] J. Glimm, A. Jaffe, and T. Spencer, "The particle structure of the weakly coupled $P(\phi)_2$ model and other applications of high temperature expansion. I and II," in *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973).
[10] J. Magnen and R. Seneor, "Infinite volume limit of the ϕ_3^4 model," preprint.
[11] J. Feldman and K. Osterwalder, "The Wightman axioms and the mass gap for the weakly coupled $(\phi^4)_3$ quantum field theory," Harvard preprint.
[12] [GS1] R. Griffiths and B. Simon, Commun. Math. Phys. **33**, 145 (1973).
[13] [SP1] T. Spencer, Comm. Math. Phys. **35**, 74 (1974).

New Jacobian θ functions and the evaluation of lattice sums

I. J. Zucker

Department of Physics, University of Surrey, Guildford, Surrey, England
(Received 19 May 1975)

The properties of some infinite series are discussed. They are used to evaluate some two- and higher-dimensional lattice sums.

I. INTRODUCTION

In recent work initiated by Glasser[1] on the exact evaluation of lattice sums much use has been made[2,3] of the Jacobian θ functions of zero argument. They are given below in their infinite series representations as presented by Whittaker and Watson[4]:

$$\theta_2(0,q)=\theta_2=\sum_{-\infty}^{\infty}{}^{(n-1/2)^2}=2q^{1/4}(1+q^2+q^6+q^{12}\cdots), \tag{1a}$$

$$\theta_3(0,q)=\theta_3=\sum_{-\infty}^{\infty}q^{n^2}=1+2q+2q^4+2q^9\cdots, \tag{1b}$$

$$\theta_4(0,q)=\theta_4=\sum_{-\infty}^{\infty}(-1)^n q^2=1-2q+2q^4-2q^9\cdots, \tag{1c}$$

$$\theta_1'(0,q)=\theta_1'=2\sum_{0}^{\infty}(-1)^n(2n+1)q^{(n+1/2)^2}$$
$$=2q^{1/4}(1-3q^2+5q^6\cdots). \tag{1d}$$

To evaluate lattice sums, the latter have been expressed as Mellin transforms of products and powers of the above series. Then using identities established by Jacobi,[5] these products and powers are expressed as a single series. This decomposes a multiple lattice sum into a product of simple sums. The method is illustrated below. Consider $S_1=\sum_{(m,n\neq 0,0)}(m^2+n^2)^{-s}$.

Let

$$\Gamma(s)M_s[f]=\int_0^{\infty}t^{s-1}f\,dt; \tag{2}$$

then

$$S_1=M_s[\theta_3^2(q)-1] \text{ with } q=e^{-t}. \tag{3}$$

By using the identity[5]

$$\theta_3^2-1=4\sum_{n=0}^{\infty}\frac{q^{n+1}}{(1+q^{2n+2})}=4\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}(-1)^m q^{(1+2m)(1+n)}, \tag{4}$$

it is elementary to show that

$$S_1=4L_1(s)L_4(s). \tag{5}$$

L_1 and L_4 are Dirichlet L functions,[6] better known in their notation and series representation now given, namely

$$L_1(s)=\zeta(s)=\sum_{0}^{\infty}(n+1)^{-s},\ s>1,$$
$$L_4(s)=\beta(s)=\sum_{0}^{\infty}(-1)^n(2n+1)^{-s},\ s>0. \tag{6}$$

In the evaluation of lattice sums much use is made of many beautiful identities which exist among (1a—1d).

These relations can be divided roughly into two kinds: (a) additive relations, e.g.,

$$\theta_3+\theta_4=2\theta_3(q^4), \tag{7a}$$

$$\theta_3-\theta_4=2\theta_2(q^4), \tag{7b}$$

$$\theta_3=\theta_3(q^4)+\theta_2(q^4); \tag{7c}$$

(b) multiplicative relations, e.g.,

$$\theta_3\theta_4=\theta_4^2(q^2), \tag{8a}$$

$$2\theta_2\theta_3=\theta_2^2(q^{1/2}). \tag{8b}$$

The additive relations are easily deduced from the series (1a—1d). Multiplicative identities are more readily established from another remarkable way of representing the θ functions, namely by infinite products. With the notation

$$Q_0=\prod_1^{\infty}(1-q^{2n}),\quad Q_1=\prod_1^{\infty}(1+q^{2n}),$$
$$Q_2=\prod_1^{\infty}(1+q^{2n-1}),\quad Q_3=\prod_1^{\infty}(1-q^{2n-1}), \tag{9}$$

the following relations are given by Whittaker and Watson[7]:

$$\theta_2=2q^{1/4}Q_0Q_2^2, \tag{10a}$$

$$\theta_3=Q_0Q_2^2, \tag{10b}$$

$$\theta_4=Q_0Q_3^2, \tag{10c}$$

$$\theta_1'=2q^{1/4}Q_0^3. \tag{10d}$$

It is elementary to establish that

$$Q_0Q_3=Q_0(q^{1/2}),\quad Q_1Q_2=Q_1(q^{1/2}),$$
$$Q_2Q_3=Q_3(q^2),\quad Q_0Q_1=Q_0(q^2), \tag{11}$$

whence

$$Q_1Q_2Q_3=1. \tag{12}$$

It is simple to find (8a) and (8b) from these results and the famous identity $\theta_1'=\theta_2\theta_3\theta_4$ follows immediately.

II. A NEW q SERIES

From (8a) and (8b) simple q series for $\theta_2\theta_3$ and $\theta_3\theta_4$ may be found since q series for θ_2^2 and θ_4^2 are known. A series for $\theta_2\theta_4$ was required by Glasser[1] to evaluate a certain lattice sum and this did not appear to be known. But Glasser[8] was able to evaluate his sum by means of number theoretic techniques, and the writer,[9] working backwards from Glasser's result, was able to find the q series for $\theta_2\theta_4$. It was also found by a direct approach

involving manipulation of q series.[9] Subsequently, it was pointed out by Joyce[10] that the q series for $\theta_2\theta_4$ could be found in Tannery and Molk,[11] who indicate that the series can be established by forming the Fourier series for $\theta_1'(z,q)/\theta_3(z,q)$ and putting $z=0$. The series is

$$\theta_2\,\theta_4 = 2q^{1/4}\sum_{n=0}^{\infty}(-1)^n\left(\frac{q^n}{1+q^{4n+1}} - \frac{q^{3n+2}}{1+q^{4n+3}}\right). \tag{13}$$

Having a q series for $\theta_2\theta_4$ and looking at relation (8a) and (8b) it seems natural to ask the following question. Is there a q series of which $\theta_2\theta_4$ is the square? The answer is in the affirmative, but the series cannot be expressed in terms of (1a—1d), and it would seem legitimate to name this series θ_5. It may be found as follows. From (10a)—(10d)

$$\theta_2\theta_4 = 2q^{1/4}Q_0^2Q_1^2Q_3^2 = 2q^{1/4}Q_0^2/Q_2^2. \tag{14}$$

Now there is a famous identity of Gauss[12] namely

$$Q_0/Q_3 = 1 + q + q^3 + q^6 + \cdots + q^{n(n+1)/2}\cdots. \tag{15}$$

Put $-q^2$ for q in this expression. Now, since $Q_0(-q^2) = Q_0(q^2)$ and $Q_3(-q^2) = Q_2(q^2)$, the following holds:

$$Q_0(q^2)/Q_2(q^2) = 1 - q^2 - q^6 + q^{12} + q^{20}\cdots. \tag{16}$$

Therefore, from (14) and (16) we may write

$$\begin{aligned} 2\theta_2(q^2)\theta_4(q^2) &= 4q^{1/2}Q_0^2(q^2)/Q_2^2(q^2) \\ &= [2q^{1/4}Q_0(q^2)/Q_2(q^2)]^2 = \theta_5^2, \end{aligned} \tag{17}$$

where

$$\theta_5 = 2q^{1/4}(1 - q^2 - q^6 + q^{12} + q^{20}\cdots) \tag{18}$$

In (17) the indices of q are $n(n+1)$ and the sign of the terms alternate in *pairs*. The series θ_5 does not appear to have been considered by Jacobi, and it would seem that it should be added to (1a)—(1d) as another basic q series. No additive relations involving θ_5 have yet been found, but the following multiplicative results have been deduced

$$\theta_2\theta_5 = 2\theta_2(q^2)\theta_4(q^4), \tag{19a}$$

$$\theta_2\theta_5\theta_4(q^4) = 2\theta_1'(q^2). \tag{19b}$$

Knowing that even powers of θ_5 can be expressed in powers of $\theta_2\theta_4$ enables the following sums to be evaluated, Namely:

$$\sum_{-\infty}^{\infty}\cdots\sum_{-\infty}^{\infty}(-1)^{m+n\cdots}[(2m-\tfrac{1}{2})^2 + (2n-\tfrac{1}{2})^2\cdots]^{-s}, \tag{20}$$

provided there are an even number of terms up to 8 only in the sum. For example, we evaluate the two-dimensional sum

$$\sum_{-\infty}^{\infty}\sum(-1)^{m+n}[(2m-\tfrac{1}{2})^2 + (2n-\tfrac{1}{2})^2]^{-s} = M_s[\theta_5^2/4]. \tag{21}$$

From (17) and (13)

$$\begin{aligned} M_s[\theta_5^2/4] &= M_s\left[\sum_{n=0}^{\infty}\frac{(-1)^n q^{2n+1/2}}{1+q^{8n+2}} - \frac{(-1)^n q^{6n+1/2}}{1+q^{8n+1}}\right] \\ &= 2^s[L_{8a}(s) + L_{8b}(s)], \end{aligned} \tag{22}$$

where

$$L_{8a}(s) = 1 + 3^{-s} - 5^{-s} - 7^s\cdots, \quad L_{8b}(s) = 1 - 3^{-s} - 5^{-s} + 7^{-s}\cdots \tag{23}$$

are Dirichlet L series of period 8.[6] When $s=1$, this gives the curious result

$$\begin{aligned} &\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}(-1)^{m+n}[(2m-\tfrac{1}{2})^2 + (2n-\tfrac{1}{2})^2]^{-1} \\ &\quad = [\pi + 2\ln(1+\sqrt{2})]/\sqrt{2}. \end{aligned} \tag{24}$$

The four-, six-, and eight-dimensional forms of (20) may be similarly evaluated. (19b) also enables us to find a three-dimensional result, namely,

$$\begin{aligned} &\sum\sum\sum(-1)^{m+n}[4m^2 + 4n^2 + 2(p-\tfrac{1}{2})^2]^{-s} \\ &\quad = M_s[\theta_4^2(q^4)\theta_2(q^2)] \\ &\quad = M_s[\theta_1'(q^2)] = 2^{s+1}\beta(2s-1). \end{aligned} \tag{25}$$

For $s=\frac{1}{2}$ this equals $\sqrt{2}$ while for $s=1$ it is equal to π.

III. OTHER Q SERIES

It will be observed that, among all the relations involving the θ functions discussed above, the powers of q involved are of the form 2^n only. This prompted us to explore q series in which this was not the case. It seemed fruitful to consider

$$\theta_6 = 2\sum_{-\infty}^{\infty}q^{(n-1/3)^2} = 2\sum_{-\infty}^{\infty}q^{(n-2/3)^2}, \tag{26}$$

$$\theta_7 = 2\sum_{-\infty}^{\infty}(-1)^n q^{(n-1/3)^2} = -2\sum_{-\infty}^{\infty}(-1)^n q^{(n-2/3)^2}. \tag{27}$$

Additive relations have been found among these series. They are

$$\theta_6(q^9) = \theta_3 - \theta_3(q^9), \tag{28a}$$

$$\theta_7(q^9) = \theta_4(q^9) - \theta_4, \tag{28b}$$

$$\theta_6 - \theta_7 = 2\theta_5(q^4), \tag{28c}$$

but as yet no multiplicative results have been found. This limits the number of lattice sums which can be evaluated. Thus only two-dimensional sums have been found in the following cases:

$$\begin{aligned} \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}[(3m-1)^2 + (3n-1)^2]^{-s} &= \tfrac{1}{4}M_s[\theta_6^2(q^9)] \\ &= \tfrac{1}{4}M_s[\theta_3^2 - 2\theta_3\theta_3(q^9) + \theta_3^2(q^9)]. \end{aligned}$$

The evaluation of $M_s[\theta_3\theta_3(q^9) - 1]$ has been given elsewhere[6] by the writer. The final result for the above sum may be written

$$[(1 - 3^{-2s})L_1L_4 - L_3L_{12}]/2, \tag{29}$$

where L_3 and L_{12} are further Dirichlet L series of period 3 and 12 given by

$$L_3 = 1 - 2^{-s} + 4^{-s} - 5^{-s}\cdots, \quad L_{12} = 1 - 5^{-s} - 7^{-s} + 11^{-s}\cdots. \tag{30}$$

Similarly it can be shown

$$\begin{aligned} &\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}(-1)^{m+n}[(3m-1)^2 + (3n-1)^2]^{-s} \\ &\quad = -[(1 - 3^{-2s})L_2L_4 - (1 + 2^{1-s})L_3L_{12}]/2, \end{aligned} \tag{31}$$

where $L_2 = 1 - 2^{-s} + 3^{-s} - 4^{-s}\cdots$. For $s=1$ the value is $-\pi\ln(4+2\sqrt{3})/9$.

It would seem likely from the foregoing that other q series might be of use in evaluating lattice sums.

ACKNOWLEDGMENT

The author expresses his thanks to Dr. M. L. Glasser for much encouraging and stimulating correspondance.

[1]M. L. Glasser, J. Math. Phys. **14**, 409 (1973).
[2]I. J. Zucker, J. Math. Phys. **15**, 187 (1974).
[3]I. J. Zucker, J. Phys. A **7**, 1568 (1974).
[4]E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U. P., Cambridge, 1958), 4th ed., p. 464. Jacobi, *Fundamental Nova Theoriae Functionum Ellipticarum* (Konigsberg, 1829), Sec. 40.
[6]I. J. Zucker and M. M. Robertson, J. Phys. A **8**, 874 (1975).
[7]Ref. 4, p. 472.
[8]M. L. Glasser, J. Math. Phys. **14**, 701 (1973).
[9]I. J. Zucker, J. Lond. Math. Soc. 11(2) (1975), to be published.
[10]G. S. Joyce, private communication (1974).
[11]J. Tannery and J. Molk, *Elements de la Théorie des fonctions elliptiques* (Paris, 1893) **4**, I, p. 108.
[12]Ref. 4, p. 185.

Self-adjoint operators, derivations and automorphisms on C^*-algebras

Heide Narnhofer*

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 8 May 1975)

We discuss how the concept of C^*-algebras with a strongly continuous one parameter group of automorphisms can be realized, if the automorphism is implemented by an unbounded operator.

1. INTRODUCTION

A system of infinitely many particles is often described by a suitable chosen C^*-algebra $\mathcal{A}$ and the states over it. Dynamics is supposed to be defined by a one parameter automorphism group τ_t.[1/2] For many results one even strengthens this assumption by the claim that this group is strongly continuous, i.e.,

$$\lim_{t\to 0}\|\tau_t A - A\| = 0, \quad \forall\ A\in\mathcal{A}. \tag{1.1}$$

For instance this continuity can be used to construct operators analytic under time automorphism. Usually this form of continuity is not essential but can be weakened by the assumption that for a given state and its corresponding representation

$$s - \lim_{t\to 0}(\pi(\tau_t A) - \pi(A)) = 0, \quad \forall\ A\in\mathcal{A}. \tag{1.2}$$

Nevertheless this restriction is representation dependent and it is an open question whether it is reasonable to assume that it holds in all representations. At least one cannot prove it referring to unitary operators implementing the automorphism, because their existence would also imply weak operator continuity in A for fixed t. But for noninvariant states time automorphism will usually not be continuous in this way, which can be easily demonstrated by a counterexample.

Assume ω is space translation invariant, but not time translation invariant. Assume further that it is a factor state. Let σ_x be the automorphism group of space translation. Then

$$w - \lim_{x\to\infty}\pi(\sigma_x A) = \omega(A)\, I_\pi. \tag{1.3}$$

Assume τ_t can be enlarged as automorphism on $\pi(\mathcal{A})''$. We know $\tau_t\sigma_x A = \sigma_x\tau_t A$. Therefore,

$$w\lim_{x\to\infty}\pi(\sigma_x\tau_t A) = \tau_t w\lim\pi(\sigma_x A) = \omega(A)\, I_\pi \tag{1.4a}$$

$$= \omega(\tau_t A)\, I_\pi \tag{1.4b}$$

which leads to a contradiction. Therefore, $\tau_t(\cdot)$ cannot be continuous in the weak operator topology.

Another argument why it should be nice if we could assume strong continuity is the following: Similar to the relation between self-adjoint operators and groups of strongly continuous unitary operators we can find a connection between automorphism groups and derivations,[3] and as perturbation theory is easier to handle for self-adjoint operators than for unitary operators it might be easier to consider the effect of additional interactions on the level of derivation theory rather than on the level of the theory on automorphism groups.

We ask how the assumption of strong continuity is realized for known physical models: It is satisfied for spin systems with short range interaction,[4] it is satisfied for continuous free Fermi systems, where $\mathcal{A}$ is built up by creation and annihilation operators. It is not satisfied for continuous free Bose systems with $\mathcal{A}$ constructed by the Weyl operators.[5] For interacting continuous systems it is not even proven whether time automorphisms exist (and in fact it is our hope that derivations might be helpful in this respect).

We turn now to the case of one particle moving in an external field. If we construct our C^*-algebra by $\exp(iq\alpha)$, $\exp(ip\beta)$, the free time automorphism is well defined but not strongly continuous. If we add an external field, then usually time evolution will not be representable as time automorphism but will lead to a new C^*-algebra.[6]

It is the aim of this paper to find models of C^*-algebras, subalgebras of $\mathcal{B}(H)$, where H is some Hilbert space, dense in strong operator topology in $\mathcal{B}(H)$, such that time automorphism is continuous in the topology introduced by the norm. So we should get some feeling how restrictive these assumptions are. We will not consider the problem how to construct a quasilocal algebra such that at least the automorphisms implied by the local Hamiltonians are strongly continuous (though to find such an algebra was the motivation of this paper). In fact it will turn out that a simple imbedding $\mathcal{A}(\Lambda)\subset\mathcal{A}(\Lambda')$ for $\Lambda\subset\Lambda'$ does not hold and therefore the usual construction is not possible. So it looks quite probable that, though the concept of strongly continuous automorphism groups works for finite systems, it may fail in general for the quasilocal algebra of statistical mechanics.

2. TIME AUTOMORPHISMS AND DERIVATIONS

We want to summarize here some results about the relation between automorphism groups and derivations:

Definition: A strongly continuous automorphism group is the representation of R in the set of automorphisms over a C^*-algebra $\mathcal{A}$ satisfying

$$\tau_s\cdot\tau_t(A) = \tau_{s+t}(A), \quad \|\tau_t A\| = \|A\|,$$
$$\lim_{t\to 0}\|\tau_t(A) - A\| = 0. \tag{2.1}$$

Definition: A derivation δ is a map of a dense subset $D(\delta)\subset\mathcal{A}$ into $\mathcal{A}$ satisfying

$$\delta(AB) = \delta(A)\cdot B + A\,\delta(B). \tag{2.2}$$

It is symmetric if $A\in D(\delta)$ implies $A^\dagger\in D(\delta)$ and $\delta(A^\dagger) = \delta(A)^\dagger$. It is closable if $A_n\to 0$, $A_n\in D(\delta)$ and

$\|\delta(A_n - A_m)\| < \epsilon$ for $n, m > n_0$ implies $\delta(A_n) \to 0$. It is maximal, if $R(\delta+1) = R(\delta-1) = \mathcal{A}$.

Lemma[15]: If $R(\delta \pm 1) = \mathcal{A}$ and if $\|\delta A + A\| \geq \|A\|$, then $R(\delta \pm c) = \mathcal{A}$ for $c > 0$. This lemma expresses the stability of deficiency which is proven in Ref. 7 [Kato, IV 5.17] for general Banach spaces (e.g., this is true for bounded and therefore inner derivations and therefore also for derivations that can be approximated by inner derivations).

Lemma: If $R(\delta+1) = R(\delta-1) = \mathcal{A}$, then

$$\tau_t = \lim_{n\to\infty}(1 - t\delta/n)^{-n} \tag{2.3}$$

exists for $t \in R$ and defines a strongly continuous automorphism group. (If this expression makes sense, $\tau_t = \exp t\delta$.) The proof corresponds to the proof for self-adjoint operators (Kato, IX 1.2).

Lemma: If τ_t is a strongly continuous automorphism group, then a maximal derivation is defined by

$$\frac{1}{1-\delta} A = \int_0^\infty \exp(-t)\tau_t A dt. \tag{2.4}$$

It is equal to

$$\delta(A) = n \lim_{t\to 0} \frac{\tau_t A - A}{t} \tag{2.5}$$

if this limit exists, and the automorphism group defined by δ coincides with τ. Again the proofs can be adopted from Ref. 7 (IX Secs. 1.2, 3).

3. THE C^*-ALGEBRA

We start with a Hilbert space H and a self-adjoint operator H on H. We want to find an algebra $\mathcal{A}$, closed in norm topology, dense in strong operator topology, such that $\mathcal{A}$ is stable under time evolution τ_t defined by H and such that τ_t acts strongly continuous on $\mathcal{A}$.

Evidently $\mathcal{B}(H)$ satisfies our requirements only if H is bounded. The biggest algebra $\hat{\mathcal{A}}$ is defined in the following way: Take $\hat{a}$ to be the set of all bounded operators A such that δA is bounded. $A \in \hat{a}$, so is $A^\dagger$ and AB and $\tau_t A$. Therefore, $\hat{\mathcal{A}} = \overline{\hat{a}}$ satisfies all our claims. Nevertheless, $\hat{\mathcal{A}}$ is defined in a slightly abstract way so that its properties cannot be discussed so nicely.

We will restrict our interest to another algebra. Let $\{E_\lambda\}$ be the spectral family of H. Define

$$\mathcal{A} = \overline{\cup_\lambda E_\lambda \mathcal{B}(H) E_\lambda \cup I}. \tag{3.1}$$

This algebra is well defined being obtained as an inductive limit. It is invariant under time automorphism and this automorphism is strongly continuous.

Remark: Other possibilities would be

$$\mathcal{A}_1 = \overline{\cup_\lambda E_\lambda \mathcal{B}(H) E_\lambda \cup \{H\}''},$$
$$\mathcal{A}_2 = \overline{\cup_\lambda E_\lambda \mathcal{B}(H) E_\lambda \cup \{H\}'}. \tag{3.2}$$

Evidently $\mathcal{A} \subset \mathcal{A}_1 \subset \mathcal{A}_2 \subset \hat{\mathcal{A}}$ where it is an open problem whether $\mathcal{A}_2 = \hat{\mathcal{A}}$.

Our definition of $\mathcal{A}$ depends very much on H and therefore we want to give an equivalent definition where it is easier to see what happens if H is changed. Define the two norms

$$\|A\|_{H,1} = \sup_{\substack{\psi \in D(H) \\ \|\psi\|=1}} \|A(H+i)\psi\|,$$
$$\|A\|_{H,2} = \infty \quad \text{if } \exists \psi \text{ with } A\psi \notin D(H)$$
$$= \sup_{\substack{\psi \in H \\ \|\psi\|=1}} \|(H-i)A\psi\| \quad \text{otherwise} \tag{3.3}$$

and define $\underline{a} = \{A, \|A\| < \infty,\ \|A\|_{H,1} < \infty,\ \|A\|_{H,2} < \infty\}$. Since $\|A\|_{H,1} = \|A^\dagger\|_{H,2}$, $\underline{a}$ is stable under conjugation. Since

$$\|AB\|_{H,1} \leq \|A\|\,\|B\|_{H,1},$$
$$\|AB\|_{H,2} \leq \|B\|\,\|A^\dagger\|_{H,1}, \tag{3.4}$$

it is also stable under taking the product. It is dense in the strong operator topology, because $\underline{a} \supset \cup_\lambda E_\lambda \mathcal{B} E_\lambda$. But it is also stable under time automorphism and

$$\delta A = n \lim \frac{\tau_t A - A}{t} = i[H, A] \quad \forall\ A \in \underline{a}.$$

Therefore the algebra $\tilde{\mathcal{A}} = \overline{\underline{a} \cup 1}$ satisfies our claims. We have to prove that both definitions lead to the same algebra: Since $\underline{a} \supset \cup_\lambda E_\lambda \mathcal{B} E_\lambda \Rightarrow \tilde{\mathcal{A}} \supset \mathcal{A}$. We want to show that $\mathcal{A} \supset \tilde{\mathcal{A}}$. Take $A \in \underline{a}$. Then we know that $\exists c$ with ($\|\psi\| = 1$)

$$c \geq \|HE_\lambda^\perp A\psi\| \geq \lambda \|E_\lambda^\perp A\psi\|. \tag{3.5}$$

This implies

$$\|E_\lambda^\perp A\psi\| \leq c/\lambda \quad \forall\ \psi, \quad \|\psi\| = 1. \tag{3.6}$$

With the same argument for $A^\dagger$ we obtain

$$\|AE_\lambda^\perp\| \leq c/\lambda. \tag{3.7}$$

Therefore,

$$A \subset \cup_\lambda (E_\lambda \mathcal{B} \cap \mathcal{B} E_\lambda) = \cup_\lambda E_\lambda \mathcal{B} E_\lambda.$$

The advantage of the second definition is the following: In the first definition we had to refer to the spectral family which depends greatly on the self-adjoint operator. But now we used H only to show that A and $A^\dagger$ are bounded operators from H on $D(H)$, $D(H)$ equipped with the norm which makes it complete. But if therefore another operator $\hat{H}$ has $D(H)$ as domain, where $\hat{H}$ is self-adjoint too, then not only τ_t but also $\hat{\tau}_t$ will be strongly continuous on $\mathcal{A}$.

There is another candidate for an algebra satisfying our requirements. That is $C = C_0 \cup 1$, where C_0 is the algebra of compact operators. Evidently it is invariant for any (unitary implemented) automorphism. The problem is whether the automorphism acts as strongly continuous representation of R.

Assume first that our self-adjoint operator H has a discrete spectrum, i.e., the spectrum consists only of eigenvalues of finite multiplicity. Then $E_\lambda B E_\lambda$ is a compact operator for all B, so $\overline{\cup_\lambda E_\lambda \mathcal{B} E_\lambda} \subset C_0$. On the other hand, we know that C is the smallest subalgebra of $\mathcal{B}(H)$, closed in norm, such that the identical representation is irreducible (Ref. 8, IV 1). Therefore $\mathcal{A} = C$.

Evidently C satisfies our claims also for every bounded operator H, though in this case it is really a subalgebra of $\mathcal{A} = \mathcal{B}(H)$. The problem remains what

happens if H is unbounded and has an essential spectrum. We can restrict ourselves on those C which can be written

$$C=\sum f_{nm}|F_n\rangle\langle F_m| \tag{3.8}$$

where the F_n are one-dimensional projection operators and the sum runs over a finite number of indices, since these operators are already norm dense in $\mathcal{C}$. For every ϵ we can find projection operators G_n with $\|G_n-F_n\|\leq\epsilon f$ such that $\|C-C'\|<\epsilon$ with $C'=\sum f_{nm}|G_n\rangle\langle G_m|$ and with the property that $\exists\lambda$ with $G_n<E_\lambda\ \forall n$.

Evidently C' belongs to α and therefore C belongs to $\mathcal{A}$. But again $\mathcal{C}$ will now be a subalgebra of $\mathcal{A}$ because $\exists\lambda_0$ such that $E_\lambda\beta E_\lambda$ is the algebra of bounded operators over an infinite-dimensional Hilbert space $\forall\lambda>\lambda_0$. One should mention another useful property of $\mathcal{C}$. It is not only stable under any time automorphism and this automorphism acts strongly continuously but there are also only two irreducible representations, namely the trivial one $[\pi(C+\lambda 1)=\lambda]$ and the identical one.[14]

4. PERTURBATION OF THE DERIVATION

Similarly for self-adjoint operators, we want to add to a maximal derivation a symmetric one and ask for a theorem that guarantees that the new derivation is again maximal (or its closure is maximal). We are led by the theorem of Ref. 7 (V, 4) which tells us: Let K be a self-adjoint operator and V be a symmetric operator satisfying

$$\|V\psi\|\leq a\|\psi\|+b\|K\psi\|\ \forall\ \psi\in D(K),\quad b<1, \tag{4.1}$$

then $K\pm V$ is self-adjoint on $D(K)$. If $b=1$, then $K\pm V$ is essentially self-adjoint on $D(K)$.

There are two ways to translate this theorem for derivations: Either adapt the proof so that the result looks formally the same, or ask what this condition implies on the corresponding derivations so that the result is the same in its contents. The first way is trivial or does not work.

Theorem: Let δ_1 be a maximal derivation and δ_2 be a symmetric derivation satisfying

$$\|\delta_2 A\|\leq a\|A\|+b\|\delta_1 A\|\ \forall\ A\in D(\delta_1),\quad b<1, \tag{4.2}$$

and $\|\alpha(\delta_1+\delta_2)A+A\|\geq\|A\|$. Then $\delta_1\pm\delta_2$ is a maximal derivation on $D(\delta_1)$.

The proof is exactly the same as for operators, using the Banach space structure. For the case $b=1$ the Hilbert space structure became important for the proof so that it fails in our case and the problem remains open. The theorem covers especially all bounded derivations δ_2 (corresponding to bounded V).

This theorem can be generalized in some sense.

Theorem. Let δ_1 again be a maximal derivation and δ_2 be a symmetric derivation with $D(\delta_1)\subset D(\delta_2)$. Then there exists an a_0 such that for all a, $|a|\leq a_0$, $R(\delta_1+a\delta_2\pm 1)$ is dense in $\mathcal{A}$ and therefore $\delta_1+a\delta_2$ defines an automorphism group.

The proof is given for general Banach spaces in Ref. 9. We try now to follow the other way and to find a general characterization of what properties derivations have corresponding to operators that are Kato-bounded. First we make the following observation:

Lemma: Kato boundedness of the operators does not imply Kato boundedness of the derivations. One can easily construct a counter example. Take as Hilbert space $\mathcal{H}$, the Fock space of fermions (Sec. 5), and let K be the kinetic energy, V some nice two-body interaction, and N the number operator. Then the derivations with respect to K and to $K+aN$ are defined on all creation and annihilation operators $a(f)$ and $a^\dagger(f)$ with f sufficiently smooth (though they do not belong to our constructed algebra $\mathcal{A}$). But this does not hold for the derivation with respect to V though we can assume V to be bounded with respect to $K+aN$. (This in fact demonstrates that we really had to choose our algebra $\mathcal{A}$ in this way.) We can only state the following:

Lemma: Let $\|V\psi\|\leq a\|\psi\|+b\|K\psi\|$. Then define $\delta_1+\alpha\delta_2$ on $\mathcal{A}$. The closure $\overline{\delta_1+\alpha\delta_2}$ is a maximal derivation for $|\alpha|<b^{-1}$ and strongly continuous in α for all $A\in\mathcal{A}$. So is $(\overline{\delta_1+\alpha\delta_2}+\gamma)^{-1}$ for all $A\in\mathcal{A}$. $\delta_1+\alpha\delta_2$ corresponds to $K+\alpha V$, which defines an automorphism with $\mathcal{A}$ being invariant.

For operators we know that Kato boundedness is not a necessary condition that the perturbed Hamiltonian exist as a self-adjoint operator. In fact, from the physical point of view one need not insist that $K+aV$ is essentially self-adjoint on $D(K)\cap D(V)$ to make good sense. We can define our new Hamiltonian, e.g., by the Friedrichs extension or by the Trotter formula. Especially the last method is also applicable for automorphism groups.

Theorem (Trotter): If $D(\delta_1)\subset D(\delta_2)$ and if further $R(\delta_1\pm 1)$, $R(\delta_2\pm 1)$ and $R(\delta_1+\delta_2\pm 1)$ are dense in $\mathcal{A}$, then the automorphism defined by $\delta_1+\delta_2$ can be obtained as norm limit

$$\tau_t(A)=n\lim_{n\to\infty}(\tau_{1t/n}\,\tau_{2t/n})^n A. \tag{4.3}$$

For self-adjoint operators this limit can exist, even when the assumptions of the theorem are not satisfied (see, e.g., Refs. 10, 11). Similarly, we have the result:

Theorem. If τ_1 and τ_2 are strongly continuous automorphisms and if

$$\tau_t(A)=n\lim_{n\to\infty}(\tau_{1t/n}\,\tau_{2t/n})^n(A)$$

exists for a dense set of t in a compact neighborhood of 0 for all A, then τ_t can be enlarged on all t and is an automorphism group.

Contrary to the operator case, here the assumptions can be rather restrictive. Throughout the previous theorems we did not assume that δ_2 itself defines an automorphism. For all usual potentials in physics, the operators we add are not only symmetric but essentially self-adjoint on the restricted domain. But this does not imply that this domain is invariant under $\exp(iVt)$, an assumption which was essential in our definition of $\mathcal{A}$.

On the other hand, if we restrict our interest on $\mathcal{C}$, then we know that $\mathcal{C}$ is stable under τ_1 and τ_2 and that they act strongly continuously, so that the input is the same as for operators.

5. PHYSICAL EXAMPLES

Our considerations have been quite general, starting with a Hilbert space and a C^*-algebra with an irreducible representation in $\mathcal{B}(\mathcal{H})$. We want now to ask how $\mathcal{A}$ really looks like in physical problems.

A. One particle

If one treats the one particle problem in the framework of C^*-algebras, one usually starts with the C^*-algebra $\mathcal{A}_0$ generated by $\{\exp(ip\alpha), \exp(iq\beta)\}$.[12] For this algebra free time evolution exists as automorphism group, but it is not strongly continuous. The next shortcoming, which is worse, is the fact that for the particle interacting with an external field (with only few and rather unphysical exceptions of the field) the algebra is not stable under time evolution.

Our algebra $\mathcal{A}$ (3.1) is defined with respect to $K=-\Delta$, the free time evolution. Then we have a strongly continuous time automorphism group for all $H=K+V$ with $V(x)\in L^2\cup L^\infty$. Comparing $\mathcal{A}$ with $\mathcal{A}_0$ we observe that neither $\exp(ip\alpha)$ nor $\exp(iq\beta)$ belong to $\mathcal{A}$. On the other hand, $\theta(P^2-p^2)\in\mathcal{A}$ but $\notin\mathcal{A}_0$, so that there is no simple relation between $\mathcal{A}$ and $\mathcal{A}_0$.

If $\mathcal{H}=L^2(\Lambda)$ so that we stay in a finite region, then K has a purely discrete spectrum no matter which boundary conditions we choose (K of course self-adjoint), so that $\mathcal{A}=C$ and in fact does not depend on the boundary condition.

B. Fock space

Take as Hilbert space $\mathcal{H}=\oplus\mathcal{H}_n$, $\mathcal{H}_n$ being the space of n particles, satisfying some symmetry properties. For our purpose it does not make any difference if the particles have to stay in some finite region Λ with specified boundary conditions or can be spread over all of R^k. Since all physical Hamiltonians considered in statistical mechanics are number preserving, we can write $H=\sum_n H_n$ with H_n the Hamiltonian of n particles acting in $\mathcal{H}_n$. Then we can construct our algebra $\mathcal{A}_n$ with the corresponding time automorphism. Again it is stable under adding a potential as long as $H_n=K_n+V_n$ is self-adjoint on $D(K_n)$. We might define $\mathcal{A}_1=\overline{\cup_n\mathcal{A}_n}$. But $\mathcal{A}_1$ is too small because it does not permit a transition from an n-particle subspace to an m-particle subspace. Especially, it does not contain creation and annihilation operators, not even in its weak closure.

H being the sum of commuting self-adjoint operators is self-adjoint, too. Therefore it seems wiser to start already with H and construct $\mathcal{A}_2$ by using (3.1). But now we are led to another shortcoming.

For a two-body interaction, the best statement one can hope for is

$$\|V_n\psi\|\le a_n\|\psi\|+b_n\|K_n\psi\| \tag{5.1}$$

where $a_n\to\infty$ for $n\to\infty$. Therefore, $D(H)\neq D(K)$ and the algebra will depend on the interaction which is not satisfying. We have to look for a smaller algebra. Choose

$$\mathcal{A}=\overline{\cup_N\cup_\lambda E_{\lambda N}\mathcal{B}E_{\lambda N}\cup 1} \tag{5.2}$$

where $E_{\lambda N}$ belongs to the spectral family of $H_N=\sum_{n=1}^N H_n$. But for this algebra our previous considerations apply. Take $A\in\cup_\lambda E_{\lambda N}\mathcal{B}E_{\lambda N}\cup 1$. Then $\delta A=\delta_N A$ and $\tau_t A=\tau_{tN}A$. So we have a strongly continuous time automorphism group for the free particles as well as for the interacting particles. We should remark that it is inessential how a_n depends on n, so that the theory covers Coulomb interaction for fermions as well as for bosons.

As for the one particle case there is no simple relation to the algebra one has usually in mind, namely in the fermion case the one built up by creation and annihilation operators (here free time automorphism acts strongly continuously, but for the automorphism corresponding to an interaction this need not be, at least the derivations δ_1 and δ_2 do not have a common domain) and in the Bose case the one constructed by $\exp[ia(f)]$ and $\exp[ia^\dagger(f)]$. Here in fact even the free time evolution is represented as automorphism, which is not strongly continuous. So $\mathcal{A}$ cannot be larger than these algebras. On the other hand, $P_1(\sum a(f)a^\dagger(g))P_1$ does not construct the whole algebra $\mathcal{B}_1$ on the one particle space, so that we do not have an inclusion in the other way either.

With stronger assumptions on our potential—but still realistic ones—we can give another description of our algebra $\mathcal{A}$.

Consider the automorphism group ν_t implemented by the number operator

$$\nu_t(A)=\exp(iNt)A\exp(-iNt). \tag{5.3}$$

Evidently this automorphism group is strongly continuous for $A\in\cup\overline{\mathcal{B}(\mathcal{H}_N)}$. Furthermore, it commutes with K and H. Therefore we can consider $\bar\tau_t$ implemented by $K+aN$ and

$$\tau_t=\nu_{-at}\bar\tau_t. \tag{5.4}$$

But now we can restrict our interest to potentials satisfying

$$\|V\psi\|<b\|\psi\|+c\|(K+aN)\psi\|, \tag{5.5}$$

so that we can take

$$\mathcal{A}=\overline{\cup_\lambda F_\lambda\mathcal{B}F_\lambda\cup 1} \tag{5.6}$$

where F_λ belongs to the spectral family of $K+aN$.

One of the most important examples in physics is the Coulomb system. Here we know from the results of Dyson and Lenard,[13] that (5) holds if we have fermions and an equal amount of positive and negative charge. But within this restriction on the Hilbert space we see that our algebra has a well-defined time automorphism. If, on the other hand, we take the full Fock space $\mathcal{H}=\mathcal{H}_1\otimes\mathcal{H}_2$, $\mathcal{H}_i$ being the Fock space belonging to the different kinds of particles, so that the system need not be neutral or if we have gravitation, i.e., only attraction, then

$$V\le K+aN^{7/3} \tag{5.7}$$

or for a neutral Bose system

$$V\le K+aN^{5/3}. \tag{5.8}$$

Still $N^{7/3}(N^{5/3})$ defines an automorphism group commuting with τ. Now F_λ has to belong to the spectral family of $K+aN^{7/3}(K+aN^{5/3})$.

We ask once more whether the considerations become

simpler for finite regions. In fact, K has a pure point spectrum so that $\mathcal{A}=\mathcal{C}$ and

$$\overline{\bigcup_\lambda E_\lambda \beta E_\lambda \bigcup 1} = \overline{\bigcup_N \bigcup_\lambda E_{\lambda N} \beta E_{\lambda N} \bigcup 1} = \overline{\bigcup_\lambda F_\lambda \beta F_\lambda \bigcup 1}. \quad (5.9)$$

Though $K+V$ need not have a pure discrete spectrum, it defines a strongly continuous automorphism group on $\mathcal{A}$. In fact, $K+V+aN$ still has a pure discrete spectrum, if (5.5) holds.

To summarize our observation, we have seen that the concept of a C^*-algebra with a strongly continuous one parameter automorphism group is quite satisfying for a great class of physical problems. Of course we have to be careful in our choice of the algebra $\mathcal{A}$. It need not be unique and it is possible to choose it in a way such that it depends quite heavily on the automorphism we have in mind. On the other hand, we are also able to find one which fits for every automorphism.

Nevertheless, we should mention that we have lost one nice property we had for other algebras: We cannot define it purely algebraically, i.e., as an algebra built up by some operators with some norm satisfying some commutation or anticommutation relations. But this should not be an essential shortcoming, especially since the weak closures of the algebras in question coincide for all physical relevant representations. The real problem that remains is how strong continuity can be saved if we turn to the algebra of statistical mechanics.

ACKNOWLEDGMENTS

I want to thank Professor J.R. Klauder and Dr. K. Baumann for critical and stimulating discussions.

*On leave of absence from the Institute for Theoretical Physics, University of Vienna, Austria.

[1]R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964). D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[2]For example, R. Haag, N.M. Hugenholtz, and M. Winnink, Commun. Math. Phys. 5, 215 (1967); D. Kastler, J.C.T. Pool, and E.T. Poulsen, Commun. Math. Phys. 12, 175 (1969); R. Haag, D. Kastler, and E.B. Trych Pohlmeier, Commun. Math. Phys. 38, 173 (1974); H. Narnhofer and D.W. Robinson, "Dynamical stability," Commun. Math. Phys. (to be published).
[3]J. Dixmier, "Algebres d'operateurs," in Varenna Lecture Notes, *Local Quantum Theory* (Academic, New York, 1969); R.V. Kadison and J.R. Ringrose, Commun. Math. Phys. 4, 32 (1967).
[4]D.W. Robinson, Commun. Math. Phys. 7, 337 (1968).
[5]M. Guenin, Commun. Math. Phys. 3, 120 (1966); R.F. Streater, Commun. Math. Phys. 7, 93 (1968).
[6]M. Fannes and A. Verbeure, Commun. Math. Phys. 35, 257 (1974).
[7]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[8]J. Dixmier, *Les C*-algebres* (Gauthier-Villars, Paris, 1969).
[9]H. Trotter, Proc. Am. Math. Soc. 10, 545 (1959).
[10]E. Nelson, J. Math. Phys. 5, 332 (1964).
[11]W. Faris, Pacific J. Math. 22, 47 (1967); P.R. Chernoff, Bull. Am. Math. Soc. 76, 395 (1970).
[12]J. Slawny, Commun. Math. Phys. 24, 151 (1972); J. Ganuceau, M. Sirugue, D. Testand, and A. Verbeure, Commun. Math. Phys. 32, 231 (1973).
[13]F.J. Dyson and A. Lenard, J. Math. Phys. 8, 423 (1967); A. Lenard and F.J. Dyson, J. Math. Phys. 9, 698 (1968).
[14]R.V. Kadison, Lectures on Operator Algebras, Cargese Lectures, *Applications of Mathematics to Problems in Theoretical Physics* (Gordon and Breach, New York, 1967).
[15]O. Poratelli and D.W. Robinson, Comm. Math. Phys. 42, 253 (1975).

Calculation of special functional integrals

W. Kollmann

Department of Mechanics and Materials Science, The Johns Hopkins University, Baltimore, Maryland 21218
(Received 1 April 1975)

For the integration with respect to a Gaussian weak distribution of a special class of functionals φ defined on $L^2_{[0,1]}$ explicit formulas are derived. The φ are defined as nonnegative integer powers of a continuous linear functional B and the integration is performed under the condition that another continuous linear functional A assumes a given value.

INTRODUCTION

The application of the projection operator method by Zwanzig[1] in cases where the phase space Ω is a function space (Hilbert space) makes the evaluation of certain functional integrals necessary. The definition of this projection operator implies the definition of an integration process over all possible states Ω of the system under consideration. We confine ourselves to integration with respect to a Gaussian density and to $\Omega = L^2_{[0,1]}$. Skorohod[2] has shown that this density, defined on all finite-dimensional subspaces of the separable Hilbert space Ω, is compatible and yields a weak distribution on Ω. We consider two linear functionals A, B and derive the explicit form of the integral of B^α, where $\alpha \geq 0$ integer, over Ω under the condition that A assumes a given value a.

DEFINITIONS

Consider the real Hilbert space $\Omega = L^2_R$ where $R = [0,1]$ with the scalar product

$$(\varphi, \psi) = \int_R d\xi \varphi(\xi)\psi(\xi) \quad \text{for } \varphi, \psi \in \Omega.$$

The linear functionals have the form

$$A[\varphi(\cdot)] = \int_R d\xi a(\xi)\varphi(\xi) \tag{1}$$

and

$$B[\varphi(\cdot)] = \int_R d\xi b(\xi)\varphi(\xi), \tag{2}$$

where $a(\xi), b(\xi) \in C_R$ (continuous) and let

$$\int_R d\xi a^2(\xi) > 0 \quad \text{and} \quad \int_R d\xi b^2(\xi) > 0.$$

A and B are obviously defined for $\varphi \in \Omega$. The integral over Ω is defined as (see Ref. 2, Chap. 1, Ref. 3, I)

$$\int_\Omega \mu(d\varphi) f[\varphi(\cdot)] = \lim_{n\to\infty} \int_{\Omega_n} \mu_n(d\varphi) f[\varphi_n(\cdot)],$$

$$\varphi_n = \varphi(P_n\varphi),$$

where $\{P_n\}$ is a sequence of projections of Ω onto the finite-dimensional spaces Ω_n. P_n is defined by $P_n\varphi = (\varphi_i)_1^n$ and

$$\varphi_i = \frac{1}{l_i}\int_{\Delta_i} d\xi\, \varphi(\xi), \quad \begin{cases} i = 1 \cdots n, & \varphi \in \Omega, \\ l_i = \xi_i - \xi_{i-1}, \end{cases}$$

and Δ_i is the ith interval of the division ξ_n of R. The sequence $\{\xi_n\}$ must satisfy

$$\max_{\Delta_i \in \xi_n} l_i \to 0 \quad \text{as} \quad n \to \infty.$$

The symbol $(\varphi_i)_1^n$ denotes the step function which has in interval Δ_i the value φ_i. Therefore, $(\varphi_i)_1^n \in L^2_R = \Omega$ and $\Omega_n \subset \Omega$. The differential $\mu_n(d\varphi)$ is chosen as Gaussian and defined as

$$\mu_n(d\varphi) \equiv \exp[-\tfrac{1}{2}(P_n\varphi, P_n\varphi)] \prod_{i=1}^n \sqrt{l_i/2\pi}\, d\varphi_i.$$

We want to calculate the integral[3] for a special class of functionals that have the form

$$f[\varphi(\cdot)] = B^\alpha[\varphi(\cdot)]\delta(A[\varphi(\cdot)] - a).$$

Some remarks are necessary concerning the integration w.r.t. a weak distribution that does not define a measure on (Ω, β) (β Borel σ-algebra of subsets of Ω). From the form of $\mu_n(d\varphi)$ given above we see that for each finite-dimensional subspace Ω_n, μ_n defines a Gaussian measure with the identity I_n as correlation matrix. In the limit $n \to \infty$ we get therefore the identity I as correlation operator of the weak distribution, and since I is not nuclear, the sequence $\{\mu_n\}$ does not generate a Gaussian measure. But the integral of $f[\varphi(\cdot)]$ w.r.t. this weak distribution can be defined (see Ref. 2, Sec. 2) and can be useful in application, if we keep in mind that this integral is not σ-additive.

The main result is stated in Theorem 2. For the proof of Theorem 2 we need the formulas given in Theorem 1 below. We introduce the notations

$$X \equiv \int_R d\xi\, a^2(\xi)$$

$$Y \equiv \int_R d\xi\, a(\xi)b(\xi) \quad \text{and} \quad \Gamma \equiv \frac{Y^2}{XZ},$$

$$Z \equiv \int_R d\xi\, b^2(\xi)$$

where $a(\xi), b(\xi)$ are defined by (1) and (2).

THEOREM 1

Let $x \geq 0$, $l \geq 0$ be integers, the linear functionals be given by (1) and (2); then

$$\int_\Omega \mu(d\varphi) A^l[\varphi(\cdot)] B^\alpha[\varphi(\cdot)] = \begin{cases} 0 & \text{for } \alpha + l = 2k+1,\ k = 0,1,2,\cdots, \\ \dfrac{\alpha!\, l!}{2^{\gamma+m}} \cdot X^m \cdot Z^\gamma \cdot \displaystyle\sum_{K=0}^{T/2} \frac{\Gamma^k}{(2k)!(m-k)!(\gamma-k)!} & \text{for } \alpha = 2\gamma, l = 2m, \gamma, m = 0,1,2,\cdots,\ T \equiv \min(\alpha, l) \\ \dfrac{\alpha!\, l!}{2^{\gamma+m+1}} \cdot Y \cdot X^m \cdot Z^\gamma \cdot \displaystyle\sum_{K=0}^{T/2} \frac{\Gamma^k}{(2k+1)!(m-k)!(\gamma-k)!} & \text{for } \alpha = 2\gamma+1, l = 2m+1, \gamma, m = 0,1,2,\cdots, \\ & T \equiv \min(\alpha - 1, l - 1). \end{cases}$$

Proof: To any projection P_n of Ω into the n-dimensional subspace Ω_n there corresponds the n-fold integral

$$I^n_{l\alpha} = \int_{\Omega_n} \mu_n(d\varphi) A^l[P_n\varphi]\, B^\alpha[P_n\varphi]$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}\sum_{i=1}^{n} l_i\varphi_i^2\right) \cdot \left(\sum_{i=1}^{n} l_i a_i \varphi_i\right)^l \cdot \left(\sum_{i=1}^{n} l_i b_i \varphi_i\right)^\alpha$$
$$\times \prod_{i=1}^{n} {}^{\alpha} \sqrt{l_i/2\pi}\, d\varphi_i;$$

or

$$I^n_{lx} = \sum_{\nu_1=1}^{n} \cdots \sum_{\nu_l=1}^{n} \sum_{\mu_1=1}^{n} \cdots \sum_{\mu_\alpha=1}^{n} l_{\nu_1} \cdots l_{\nu_l} l_{\mu_1} \cdots l_{\mu_\alpha}$$
$$\times a_{\nu_1} \cdots a_{\nu_l} b_{\mu_1} \cdots b_{\mu_\alpha} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_{\nu_1} \cdots \varphi_{\nu_l} \varphi_{\mu_1} \cdots \varphi_{\mu_\alpha}$$
$$\times \exp\left(-\tfrac{1}{2}\sum_{i=1}^{n} l_i\varphi_i^2\right) \prod_{i=1}^{n} \sqrt{l_i/2\pi}\, d\varphi_i$$

where P_n, l_i, φ_i are defined as above. Since the exponent of the Gaussian density has diagonal form, we can apply the relation

$$\int_{-\infty}^{\infty} x^m e^{-x^2/2\sigma^2} \frac{dx}{\sigma\sqrt{2\pi}} = \begin{cases} \sigma^{2K} \cdot (2k)!/2^K k!, & m = 2k \\ 0 & m = 2k+1. \end{cases} \tag{3}$$

For $l+\alpha = 2k+1$, $k = 0,1,2,\cdots$,

$$I^n_{l\alpha} = 0 \Rightarrow \int_\Omega \mu(d\varphi) B^\alpha[\varphi(\cdot)] A^l[\varphi(\cdot)] = 0.$$

For $l+\alpha = 2k$ and $\alpha = 2\gamma$, $l = 2m$ we see that the integrals in $I^n_{l\alpha}$ are nonzero iff even numbers of indices coincide. Therefore, the sum can be reduced to

$$I^n_{l\alpha} = \sum_{\kappa_1=1}^{n} \cdots \sum_{\kappa_k=1}^{n} l^2_{\kappa_1} \cdots l^2_{\kappa_k} \sum_{A_i} a_{\nu_1} \cdots a_{\nu_l} b_{\mu_1} \cdots b_{\mu_\alpha}$$
$$\times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_{\nu_1} \cdots \varphi_{\mu_x} \varphi_{\mu_\alpha} \mu_n(d\varphi),$$

where $\sum_{A_i}$ is extended over all possible mappings A_i of the $2k$-element index set $\{\nu_1 \cdots \nu_l \mu_1 \cdots \mu_\alpha\}$ onto $\{\kappa_1 \cdots \kappa_k\}$ and where $A_i^{-1}(\kappa_j)$ has exactly two elements and $A_i^{-1}(\kappa_m) \cap A_i^{-1}(\kappa_j) = \phi$ for $m \neq j$ and for all i. Integration yields then

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi_{\nu_1} \cdots \varphi_{\mu_\alpha} \mu_n(d\varphi)$$
$$= \begin{cases} \prod_{i=1}^{k} l_{\kappa_i}^{-1} & \text{for all } \kappa_i \text{ different,} \\ \prod_{i=1}^{k} l_{\kappa_i}^{-1} + C \prod_{i=1}^{k} l_{\kappa_i}^{-1} & \text{at least one } \kappa_j = \kappa_m, \end{cases}$$

where C is determined by (3) and depends on the number of coincidences $\kappa_m = \kappa_j$, but not on n. Then we get

$$I^n_{l\alpha} = \sum_{s=s_0(2)}^{T} D_s \left(\sum_{i=1}^{n} l_i a_i^2\right)^{n_1} \cdot \left(\sum_{i=1}^{n} l_i a_i b_i\right)^{s} \cdot \left(\sum_{i=1}^{n} l_i b_i^2\right)^{n_2}$$
$$+ O(l_i),$$

where

$$n_i = \tfrac{1}{2}(l-s), \quad n_2 = \tfrac{1}{2}(\alpha - s),$$
$$s_0 = 0 \qquad T = \min(\alpha, l).$$

The coefficients D_s are determined inductively by counting all configurations that lead to a given exponent s in $I^n_{l\alpha}$:

$$D_s = 2^{s-k} \frac{\alpha!\, l!}{[(l-s)/2]!\, s!\, [(\alpha - s)/2]!}$$

The terms $O(l_i)$ come from diagonal terms in the sums over κ_i and are of order l_i. The desired integral over Ω can be evaluated by

$$\int_\Omega \mu(d\varphi) A^l[\varphi(\cdot)] B^\alpha[\varphi(\cdot)] = \lim_{n\to\infty} I^n_{lx},$$

where

$$\left.\begin{aligned} &\sum_{i=1}^{n} l_i a_i^2 \to X \\ &\sum_{i=1}^{n} l_i a_i b_i \to Y \\ &\sum_{i=1}^{n} l_i b_i^2 \to Z \end{aligned}\right\} \text{ as } n \to \infty$$

because $a, b \in C_R$. For $l+\alpha = 2k$ and $x = 2\gamma+1$, $l = 2m+1$ a slight variation of the same arguments applies and therefore they shall not be repeated.

THEOREM 2

Let the linear functionals $A[\varphi(\cdot)]$, $B[\varphi(\cdot)]$ be (1) and (2); then (a) for $\alpha = 2\gamma \geq 0$, $\gamma = 0,1,2,\cdots$ even

$$\int_\Omega \mu(d\varphi) \delta(A[\varphi(\cdot)] - a) B^\alpha[\varphi(\cdot)]$$
$$= \frac{\alpha!}{\sqrt{2\pi X}} \left(\frac{Z}{2}\right)^\gamma \cdot \exp\left(-\frac{a^2}{2X}\right)$$
$$\cdot \sum_{K=0}^{\gamma} \frac{(-1)^k}{(2k)!\,(\gamma-k)!} \left(\frac{\Gamma}{4}\right)^k \cdot H_{2k}\left(\frac{a}{\sqrt{2X}}\right);$$

(b) for $\alpha = 2\gamma+1 > 0$, $\gamma = 0,1,2,\cdots$ odd

$$\int_\Omega \mu(d\varphi) \delta(A[\varphi(\cdot)] - a) B^\alpha[\varphi(\cdot)]$$
$$= \frac{\alpha!}{4\sqrt{\pi}} \cdot \frac{Y}{X} \left(\frac{Z}{2}\right)^\gamma \cdot \exp\left(-\frac{a^2}{2X}\right)$$
$$\cdot \sum_{K=0}^{\gamma} \frac{(-1)^k}{(2k+1)!\,(\gamma-k)!} \left(\frac{\Gamma}{4}\right)^k \cdot H_{2k+1}\left(\frac{a}{\sqrt{2X}}\right)$$

where X, Y, Z, Γ are defined above and $H_n(y)$ denotes the Hermite polynomial of order n.

Proof of (a): We use the well-known fundamental sequence $\{\delta_n\}$ defining the delta distribution

$$\delta_k(x-t) = (k/\sqrt{\pi}) \exp[-k^2(x-t)^2],$$

where

$$\delta(x-t) = [\delta_n(x-t)].$$

Then we can write the integral (3) as

$$\int_\Omega \mu(d\varphi) f[\varphi(\cdot)]$$
$$= \lim_{k\to\infty} \lim_{n\to\infty} \int_{\Omega_n} \mu_n(d\varphi) \delta_k(A[P_n\varphi] - a)\, B^\alpha[P_n\varphi].$$

Consider $I_{nk} \equiv \int_{\Omega_n} \mu_n(d\varphi) \delta_k(A[P_n\varphi] - a) B^\alpha[P_n\varphi]$ for $n,k > 0$ fixed. I_{nk} is bounded as $n \to \infty$:

$$|I_{nk}| = \frac{k}{\sqrt{\pi}} \left| \int_{\Omega_n} \mu_n(d\varphi) \exp\{-k^2(A[P_n\varphi] - a)^2\} B^\alpha[P_n\varphi] \right|$$
$$\leq \frac{k}{\sqrt{\pi}} \int_{\Omega_n} \mu_n(d\varphi) \left| B^\alpha[P_n\varphi] \right|$$

for $\alpha = 2\gamma \geq 0$; it follows from Theorem 1 that

$$|I_{nk}| \leq \frac{k}{\sqrt{\pi}} \int_{\Omega_n} \mu_n(d\varphi) B^{\alpha}[P_n\varphi]$$

$$= \frac{k}{\sqrt{\pi}} \frac{\alpha!}{2^{\gamma}\gamma!} \{[\int_R d\xi b^2(\xi)]^{\gamma} + O(l_i)\} < \infty$$

for $\alpha = 2\gamma + 1 > 0$:

$$|I_{nk}| \leq \frac{k}{\sqrt{\pi}} \int_{\Omega_n} \mu_n(d\varphi) B^{2\gamma}[P_n\varphi] |B[P_n\varphi]|$$

$$\leq \frac{k}{\sqrt{\pi}} \int_{\Omega_n} \mu_n(d\varphi)[B^{2\gamma}[P_n\varphi] + B^{2\gamma+2}[P_n\varphi]] < \infty$$

using $|B[P_n\varphi]| \leq 1 + B^2[P_n\varphi]$ and Theorem 1. Then the series

$$\frac{k}{\sqrt{\pi}} \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} k^{2m} \int_{\Omega_n} \mu_n(d\varphi)(A[P_n\varphi] - a)^{2m} B^{\alpha}[P_n\varphi]$$

is convergent to I_{nk} for all finite a and for $n \to \infty$ because for $\epsilon > 0$ exists $N_\epsilon(a)$:

$$\left| \delta_k(A[P_n\varphi] - a) - \frac{k}{\sqrt{\pi}} \sum_{m=0}^{n} \frac{(-1)^m}{m!} \cdot k^{2m} \cdot (A[P_n\varphi] - a)^{2m} \right| < \epsilon$$

for $N > N_\epsilon$.

It follows that

$$\left| \int_{\Omega_n} \mu_n(d\varphi) \delta_k(A[P_n\varphi] - a) B^{\alpha}[P_n\varphi] \right.$$

$$\left. - \frac{k}{\sqrt{\pi}} \sum_{m=0}^{N} \frac{(-1)^m}{m!} k^{2m} \int_{\Omega_n} \mu_n(d\varphi)(A[P_n\varphi] - a)^{2m} B^{\alpha}[P_n\varphi] \right|$$

$$< \epsilon \int_{\Omega_n} \mu_n(d\varphi) |B^{\alpha}[P_n\varphi]| \text{ for } N > N_\epsilon.$$

This holds for $n \to \infty$ because

$$\lim_{n \to \infty} \int_{\Omega_n} \mu_n(d\varphi) |B^{\alpha}[P_n\varphi]| < \infty$$

and

$$I_k = \lim_{n \to \infty} I_{nk} < \infty,$$

as has been shown above. Therefore, the right-hand side of the inequality can be made as small as desired independent of n and the series converges to I_{nk} as $n \to \infty$. Since

$$\int_{\Omega_n} \mu_n(d\varphi)(A[P_n\varphi] - a)^{2m} B^{\alpha}[P_n\varphi]$$

is analytic in a, the series may be rearranged in powers of a. After carrying out the integration using Theorem 1, we get

$$I_k = \frac{k}{\sqrt{\pi}} \alpha! \left(\frac{Z}{2}\right)^{\gamma} \cdot \sum_{\delta=0}^{\infty} C_\delta W^\delta,$$

where

$$W \equiv a^2/X, \quad V \equiv k^2 X$$

and

$$C_\delta = \sum_{r=\delta}^{\infty} (-1)^r \cdot \frac{(2r)!}{r!} \cdot \frac{V^r}{2^{r-\delta}(2\delta)!}$$

$$\cdot \sum_{\kappa=0}^{\min(r-\delta,\gamma)} \frac{\Gamma^\kappa}{(2\kappa)!(r-\delta-\kappa)!(\gamma-\kappa)!};$$

then

$$C_\delta = \frac{(-1)^\delta}{(2\delta)!} V^\delta \cdot \sum_{\kappa=0}^{\gamma} (-1)^\kappa \cdot \left(\frac{V}{2}\right)^\kappa \cdot \frac{\Gamma^\kappa}{(2\kappa)!(\gamma-\kappa)!}$$

$$\times \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \frac{(2l+2\delta+2\kappa)!}{(l+\delta+\kappa)!} \left(\frac{V}{2}\right)^l,$$

which represents C_δ only for $V < \frac{1}{2}$. To get a representation for all $V \geq 0$, we consider the series $S_\xi(x)$ in powers of x

$$S_\xi(x) = \sum_{l=0}^{\infty} h_l x^l \quad \text{with } h_l = \frac{(-1)^l}{l!} \frac{(2l+2\xi)!}{(l+\xi)!},$$

then the coefficients h_l satisfy

$$h_0 = (2\xi)!/\xi!, \quad h_{l+1} = -4h_l \cdot \frac{1 + (2\xi+1)/2l}{1 + (1/l)}, \quad l > 0.$$

From Leibniz's criterion for alternating series follows the fact that $S_\xi(x)$ converges for $0 \leq x < \frac{1}{4}$ and $S_\xi(x)$ satisfies the differential equation

$$S_\xi'(x)(1+4x) + 2S_\xi(1+2\xi) = 0$$

with initial condition $S_\xi(0) = h_0$. Then

$$S_\xi(x) = \frac{(2\xi)!}{\xi!} \cdot \frac{1}{(1+4x)^{\xi+1/2}}$$

as easily can be verified. For C_δ we get a representation valid for all $V \geq 0$,

$$C_\delta = \frac{1}{\sqrt{1+2V}} \frac{(-1)^\delta}{(2\delta)!} \cdot \left(\frac{V}{1+2V}\right)^\delta$$

$$\cdot \sum_{\kappa=0}^{\gamma} \frac{(-1)^\kappa}{(2\kappa)!(\gamma-\kappa)!} \frac{(2\delta+2\kappa)!}{(\delta+\kappa)!} \cdot \left(\frac{\Gamma}{2}\right)^\kappa \cdot \left(\frac{V}{1+2V}\right)^\kappa;$$

then

$$\sum_{\delta=0}^{\infty} C_\delta W^\delta = \frac{1}{\sqrt{1+2V}} \cdot \sum_{\kappa=0}^{\gamma} \frac{(-1)^\kappa}{(2\kappa)!(\gamma-\kappa)!} \left(\frac{V\Gamma}{2+4V}\right)^\kappa$$

$$\cdot \sum_{\delta=0}^{\infty} \frac{(-1)^\delta}{(2\delta)!} \cdot \frac{(2\delta+2\kappa)!}{(\delta+\kappa)!} \left(\frac{VW}{1+2V}\right)^\delta.$$

Rodriguez' formula for Hermite polynomials $H_n(x)$,

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}(e^{-x^2}),$$

yields ($V \geq 0$, $W \geq 0$, see definition)

$$\sum_{\delta=0}^{\infty} \frac{(-1)^\delta}{(2\delta)!} \cdot \frac{(2\delta+2\kappa)!}{(\delta+\kappa)!} \left(\left[\frac{VW}{1+2V}\right]^{1/2}\right)^{2\delta}$$

$$= H_{2\kappa}\left[\left(\frac{VW}{1+2V}\right)^{1/2}\right] \cdot \exp\left(-\frac{VW}{1+2V}\right)$$

and

$$I_k = \frac{\alpha!}{\sqrt{\pi X}} \left(\frac{Z}{2}\right)^{\gamma} \cdot \left(\frac{V}{1+2V}\right)^{1/2} \cdot \exp\left(-\frac{VW}{1+2V}\right)$$

$$\cdot \sum_{\kappa=0}^{\gamma} \frac{(-1)^\kappa}{(2\kappa)!(\gamma-\kappa)!} \left(\frac{V\Gamma}{2+4V}\right)^\kappa \cdot H_{2\kappa}\left(\left(\frac{VW}{1+2V}\right)^{1/2}\right),$$

and finally for $k \to \infty$ we arrive at

$$\int_\Omega \mu(d\varphi)\delta(A[\varphi(\cdot)]-a)B^\alpha[\varphi(\cdot)]$$
$$= \frac{\alpha!}{\sqrt{2\pi X}}\left(\frac{Z}{2}\right)^\gamma \cdot \exp\left(-\frac{a^2}{2X}\right)\cdot \sum_{K=0}^{\gamma} \frac{(-1)^k}{(2k)!(\gamma-k)!}$$
$$\cdot \left(\frac{\Gamma}{4}\right)^k \cdot H_{2k}\left(\frac{a}{\sqrt{2X}}\right).$$

The proof of (b) goes along the same lines and shall not be repeated here.

ACKNOWLEDGMENT

This work was supported by the Deutsche Forschungsgemeinschaft Fellowship no. Ko 547/1/2.

[1]R. Zwanzig, Physica **30**, 1109 (1964); Phys. Rev. **124**, 983 (1961).

[2]A. V. Skorohod, *Integration in Hilbert Space* (Springer, New York, 1974).

[3]K. O. Friedrichs *et al.*, "Integration of functionals," Seminar, New York University, 1957.

The Weyl correspondence and path integrals*

Maurice M. Mizrahi

Center for Relativity, University of Texas, Austin, Texas 78712
(Received 2 May 1975)

The method of Weyl transforms is used to rigorously derive path integral forms for position and momentum transition amplitudes from the time-dependent Schrödinger equation for arbitrary Hermitian Hamiltonians. It is found that all paths in phase space contribute equally in magnitude, but that each path has a different phase, equal to $1/\hbar$ times an "effective action" taken along it. The latter is the time integral of $p\cdot\dot{q}-h(p,q)$, $h(p,q)$ being the Weyl transform of the Hamiltonian operator $\mathbf{H}$, which differs from the classical Hamiltonian function by terms of order $\hbar^2$, vanishing in the classical limit. These terms, which can be explicitly computed, are zero for relatively simple Hamiltonians, such as $(1/2M)[\mathbf{P}-eA(\mathbf{Q})]^2+V(\mathbf{Q})$, but appear when the coupling of the position and momentum operators is stronger, such as for a relativistic spinless particle in an electromagnetic field, or when configuration space is curved. They are always zero if one opts for Weyl's rule for forming the quantum operator corresponding to a given classical Hamiltonian. The transition amplitude between two position states is found to be expressible as a path integral in configuration space alone only in very special cases, such as when the Hamiltonian is quadratic in the momenta.

I. INTRODUCTION

The Weyl correspondence between quantum-mechanical operators and ordinary functions in phase space has not received enough attention in the textbooks. This neglect stems from the fact that it was introduced by Weyl[1] (who derived it on the basis of group-theoretic arguments) primarily as a means of forming unique quantum-mechanical operators from classical functions. As such, it was only one of many such rules, with its advantages and drawbacks.[2]

However, it turns out that the Weyl correspondence, regarded simply as a way of transforming an operator into a function and vice versa, can be used as an analytical tool for quantum-mechanical calculations. It is our purpose in this paper to use it to derive path integral forms for various transition amplitudes from the time-dependent Schrödinger equation for arbitrary Hamiltonians.

II. DEFINITIONS AND NOTATION

Let $\mathbf{Q}^i(t)$, $\mathbf{P}_i(t)$ be the (Hermitian) position and momentum operators at time t, with eigenstates $|q,t\rangle$, $|p,t\rangle$ and eigenvalues q^i, p_i:

$$\mathbf{Q}^i(t)\,|q,t\rangle = q^i\,|q,t\rangle,\quad \mathbf{P}_i(t)\,|p,t\rangle = p_i\,|p,t\rangle \tag{1}$$

The q^i's are rectangular coordinates and the p_i's their conjugate mates. We work in n dimensions and all integrals will be over $\mathbb{R}^n$. Boldface capital letters denote operators, lower case letters denote ordinary numbers.

The eigenstates are required to satisfy the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\langle q,t| = H(\mathbf{P},\mathbf{Q},t)\langle q,t|,$$

$$i\hbar\frac{\partial}{\partial t}\langle p,t| = H(\mathbf{P},\mathbf{Q},t)\langle p,t| \tag{2}$$

with the boundary conditions

$$\langle q,t|q',t\rangle = \delta(q-q'),\quad \langle p,t|p',t\rangle = \delta(p-p'). \tag{3}$$

If the (Hermitian) Hamiltonian operator $\mathbf{H}=H(\mathbf{P},\mathbf{Q},t)$ is time-independent, (2) can be solved to yield

$$|q,t\rangle = \exp(it\mathbf{H}/\hbar)\,|q\rangle,\quad |p,t\rangle = \exp(it\mathbf{H}/\hbar)\,|p\rangle, \tag{4}$$

where $|q\rangle$ and $|p\rangle$ are time-independent (stationary) states. (1) and (4) can be used to show that the operators

$$\mathbf{Q}^i \equiv \exp(-it\mathbf{H}/\hbar)\,\mathbf{Q}^i(t)\exp(it\mathbf{H}/\hbar),$$

$$\mathbf{P}_i \equiv \exp(-it\mathbf{H}/\hbar)\,\mathbf{P}_i(t)\exp(it\mathbf{H}/\hbar) \tag{5}$$

are time-independent, with eigenvalues q^i, p_i and eigenstates $|q\rangle$, $|p\rangle$.

The momentum eigenvector $|p\rangle$ is chosen to have the coordinate representation (plane wave):

$$\langle q,t|p,t\rangle = \langle q|p\rangle = h^{-n/2}\exp(ip\cdot q/\hbar), \tag{6}$$

where $h\equiv 2\pi\hbar$, $p\cdot q\equiv p_1q^1+\cdots+p_nq^n$. We assume that the eigenstates form complete sets:

$$\int |p,t\rangle\langle p,t|\,dp = \int |q,t\rangle\langle q,t|\,dq = \mathbf{I}, \tag{7}$$

where $dp\equiv dp_1\ldots dp_n$, $dq\equiv dq^1\ldots dq^n$. On the basis of (6) and (7), one can deduce the representation of $\mathbf{P}_i$ in coordinate space and the commutation relation:

$$\langle p|\,\mathbf{P}_i = -i\hbar\frac{\partial}{\partial q^i}\langle p|,\quad [\mathbf{Q}^i,\mathbf{P}_j] = i\hbar\delta^i{}_j. \tag{7'}$$

III. THE WEYL CORRESPONDENCE

Here we follow the clear presentation of DeGroot and Suttorp.[3] By expanding an arbitrary operator $\mathbf{A}=A(\mathbf{P},\mathbf{Q})$ in position and momentum eigenstates,

$$A(\mathbf{P},\mathbf{Q}) = \int dp'\,dp''\,dq'\,dq''\,|p'\rangle\langle p'|q'\rangle\langle q'|A(\mathbf{P},\mathbf{Q})\times|q''\rangle\langle q''|p''\rangle\langle p''|, \tag{8}$$

making the change of variable (of Jacobian unity),

$$p' = p-u/2,\quad p'' = p+u/2,$$

$$q' = q-v/2,\quad q'' = q+v/2, \tag{9}$$

and using (6), we get

$$A(\mathbf{P},\mathbf{Q}) = h^{-n}\int dp\,dq\,du\,dv\,|p-u/2\rangle\langle q-v/2|A(\mathbf{P},\mathbf{Q})\times|q+v/2\rangle\langle p+u/2|\exp[i(q\cdot u+p\cdot v)/\hbar] \tag{10}$$

A. Definition

The *Weyl transform* of an operator $A(\mathbf{P},\mathbf{Q})$ is

$$a(p,q)\equiv\int dv\exp(ip\cdot v/\hbar)\,\langle q-v/2|A(\mathbf{P},\mathbf{Q})|q+v/2\rangle. \tag{11}$$

The *inverse transform* is then given by (10):

$$A(\mathbf{P},\mathbf{Q})=h^{-n}\int dp\,dq\,a(p,q)\Delta(p,q), \tag{12}$$

where $\Delta(p,q)$ is the Hermitian operator

$$\Delta(p,q)\equiv\int du\exp(iq\cdot u/\hbar)\,|p-u/2\rangle\langle p+u/2|. \tag{13}$$

One can show that the following alternate expressions hold:

$$a(p,q)=\int du\exp(iq\cdot u/\hbar)\langle p+u/2|A(\mathbf{P},\mathbf{Q})|p-u/2\rangle, \tag{14}$$

$$\Delta(p,q)=\int dv\exp(ip\cdot v/\hbar)\,|q+v/2\rangle\langle q-v/2|$$
$$=h^{-n}\int du\,dv\exp\{(i/\hbar)[(q-\mathbf{Q})\cdot u+(p-\mathbf{P})\cdot v]\}. \tag{15}$$

B. Properties

a) The correspondence is one-to-one. (We shall denote it by $\longleftrightarrow$.)

b) If $A(\mathbf{P},\mathbf{Q})\longleftrightarrow a(p,q)$, then $A^{\dagger}(\mathbf{P},\mathbf{Q})\longleftrightarrow a^{*}(p,q)$. In particular, if A is Hermitian, a is real.

c) The inverse transform can be written in the compact form[4]

$$A(\mathbf{P},\mathbf{Q})=\left[\left\{\exp\left(\frac{\hbar}{2i}\,\frac{\partial}{\partial p}\cdot\frac{\partial}{\partial q}\right)\right\}a(p,q)\right]_{\substack{p=\mathbf{P},\,q=\mathbf{Q}\\ \mathbf{Q}\ before\ \mathbf{P}}}; \tag{16}$$

i.e., after evaluating the function inside the bracket, replace p and q by $\mathbf{P}$ and $\mathbf{Q}$, with all the $\mathbf{Q}$ operators placed before the P operators. This formula, which enables one to write down the operator A to arbitrary order in $\hbar$, gives the important result that in the classical limit the functions a and A are the same:

$$a(p,q)=A(p,q)+i\hbar a_1(p,q)+(i\hbar)^2a_2(p,q)+\cdots,$$

$$\lim_{\hbar\to 0}a(p,q)=A(p,q). \tag{17}$$

Hence, for Hermitian operators, $a(p,q)$ is a series in $\hbar^2$.

C. Examples

$$f(\mathbf{P})\longleftrightarrow f(p), \tag{18}$$

$$g(\mathbf{Q})\longleftrightarrow g(q), \tag{19}$$

$$\Delta(p',q')\longleftrightarrow h^n\delta(p-p')\,\delta(q-q'), \tag{20}$$

$$F(\mathbf{Q})\mathbf{P}_i\,G(\mathbf{Q})\longleftrightarrow p_iFG+(i\hbar/2)(F_{,i}G-G_{,i}F) \tag{21}$$

$$F(\mathbf{Q})\mathbf{P}_i\,V(\mathbf{Q})\,\mathbf{P}_j\,G(\mathbf{Q})$$
$$\longleftrightarrow p_ip_jFGV+(i\hbar/2)[p_iFGV_{,j}-p_jFGV_{,i}$$
$$+p_iF_{,j}GV+p_jF_{,i}GV-p_iFG_{,j}V-p_jFG_{,i}V]$$
$$+(\hbar^2/4)|F_{,j}GV_{,i}-F_{,i}GV_{,j}+FG_{,i}V_{,j}-FG_{,j}V_{,i}$$
$$+FGV_{,ij}-F_{,ij}GV-FG_{,ij}V+F_{,i}G_{,j}V+F_{,j}G_{,i}V]. \tag{22}$$

The proof of the first three is straightforward. In the last two, which we prove below, F, G, V can have any number of indices and are evaluated at q.

Proof of (21)

The transform of $F(\mathbf{Q})\,\mathbf{P}_i\,G(\mathbf{Q})$ is

$$a(p,q)=\int dv\exp(ip\cdot v/\hbar)\langle q-v/2|F(\mathbf{Q})\,\mathbf{P}_i|p'\rangle\langle p'|$$
$$\times G(\mathbf{Q})|q+v/2\rangle\,dp'$$
$$=\int dv\,F(q-v/2)\,G(q+v/2)$$
$$\times\exp(ip\cdot v/\hbar)[i\hbar\delta_{,i}(v)] \tag{23}$$

since the integral over p' yields the derivative of the delta function. The result readily follows.

Proof of (22)

The transform of $F(\mathbf{Q})\,\mathbf{P}_i\,V(\mathbf{Q})\,\mathbf{P}_j\,G(\mathbf{Q})$ is

$$a(p,q)=\int dv\exp(ip\cdot v/\hbar)\langle q-v/2|F(\mathbf{Q})|p'\rangle\,dp'$$
$$\times\langle p'|\mathbf{P}_i|q'\rangle\,dq'\,\langle q'|V(\mathbf{Q})\,\mathbf{P}_j|p''\rangle$$
$$\times dp''\langle p''|G(\mathbf{Q})|q+v/2\rangle$$
$$=h^{-2n}\int dv\exp(ip\cdot v/\hbar)\,F(q-v/2)\,G(q+v/2)$$
$$\times\int dq'\,V(q')\int p_i'\exp[ip'\cdot(q-q'-v/2)/\hbar]\,dp'$$
$$\times\int p_j''\exp[-ip''\cdot(q-q'+v/2)/\hbar]\,dp''$$
$$=-\hbar^2\{(\partial/\partial v^i)[\exp(ip\cdot v/\hbar)\,F(q-v/2)$$
$$\times G(q+v/2)\,V_{,i}(q-v/2)]$$
$$+(\partial^2/\partial v^i\partial v^j)[\exp(ip\cdot v/\hbar)\,F(q-v/2)\,G(q+v/2)$$
$$\times V(q-v/2)\}_{v=0}, \tag{24}$$

and the result follows. Special cases of (22) are

$$\mathbf{P}_i\,\mathbf{P}_j\,G(\mathbf{Q})\longleftrightarrow\left(p_i-\frac{i\hbar}{2}\,\frac{\partial}{\partial q^i}\right)\left(p_j-\frac{i\hbar}{2}\,\frac{\partial}{\partial q^j}\right)G(q), \tag{25}$$

$$\mathbf{P}_i\,V(\mathbf{Q})\,\mathbf{P}_j\longleftrightarrow\left(p_i-\frac{i\hbar}{2}\,\frac{\partial}{\partial q^i}\right)\left(p_j+\frac{i\hbar}{2}\,\frac{\partial}{\partial q^j}\right)V(q), \tag{26}$$

$$(\mathbf{P}\cdot\mathbf{P})^m\,V(\mathbf{Q})\longleftrightarrow\left\{\sum_{i=1}^{n}\left(p_i-\frac{i\hbar}{2}\,\frac{\partial}{\partial q^i}\right)\left(p_i-\frac{i\hbar}{2}\,\frac{\partial}{\partial q^i}\right)\right\}^m V(q). \tag{27}$$

If $V^{ij}(\mathbf{Q})$ is a symmetric matrix, $V^{ij}=V^{ji}$, then

$$\mathbf{P}_i\,V^{ij}(\mathbf{Q})\,\mathbf{P}_j\longleftrightarrow p_ip_j\,V^{ij}(q)+\tfrac{1}{4}\hbar^2\,V^{ij}{}_{,ij}(q) \tag{28}$$

$$F(\mathbf{Q})\,\mathbf{P}_i\,V^{ij}(\mathbf{Q})\,\mathbf{P}_j\,F(\mathbf{Q})\longleftrightarrow p_ip_j\,F^2V^{ij}+\tfrac{1}{2}\hbar^2(\tfrac{1}{2}F^2V^{ij}{}_{,ij}$$
$$+F_{,i}F_{,j}\,V^{ij}-F_{,ij}\,FV^{ij}). \tag{29}$$

The plausibility of the rule is strengthened by the following correspondences: if $\mathbf{A}\longleftrightarrow a$ and $\mathbf{B}\longleftrightarrow b$, then (Ref. 3a, p. 348)

$$\tfrac{1}{2}(\mathbf{AB}+\mathbf{BA})\longleftrightarrow ab+O(\hbar^2), \tag{29'}$$

$$-\frac{i}{\hbar}(\mathbf{AB}-\mathbf{BA})\longleftrightarrow\frac{\partial a}{\partial q}\cdot\frac{\partial b}{\partial p}-\frac{\partial b}{\partial q}\cdot\frac{\partial a}{\partial p}+O(\hbar). \tag{29''}$$

Weyl's rule is consistent with Dirac's rule that the classical Poisson brackets become commutators (within factors of $\hbar$).

We will need the matrix elements of the Δ operator, which can be easily worked out:

$$\langle q'|\Delta(p,q)\;q''\rangle=\exp[ip\cdot(q'-q'')/\hbar]\,\delta(q-(q'+q'')/2), \tag{30}$$

$$\langle p' | \Delta(p,q)\ p'' \rangle = \exp[-iq\cdot(p'-p''/\hbar]\,\delta(p-(p'+p'')/2), \tag{31}$$

$$\langle p' | \Delta(p,q) | q' \rangle = 2^n h^{-n/2} \exp\{i[2q\cdot(p-p') - q'\cdot(2p-p')]/\hbar\} \tag{32}$$

IV. THE PATH INTEGRAL

Feynman's original presentation of the path integral[5] as an alternate formulation of quantum mechanics was based on postulates which were then shown to be consistent with the Schrödinger equation in the special case of a particle in flat space in a potential. It would be interesting to see whether, given an arbitrary Hamiltonian operator, one can write a formal path integral (over configuration, momentum, or phase space) for a given transition amplitude satisfying the Schrödinger equation. This is the question that we propose to answer here.

A. Transition between position states

The transition amplitude (or transformation function, or propagator) for a particle at q_a at time t_a to arrive at q_b at time t_b is $\langle q_b, t_b | q_a, t_a \rangle$. It satisfies the Schrödinger equation at (q_b, t_b) and its complex conjugate at (q_a, t_a). By splitting up the time interval $[t_a, t_b]$ into $m+1$ subintervals,

$$t_a \equiv t_0 < t_1 < t_2 < \cdots < t_m < t_{m+1} \equiv t_b, \tag{33}$$

we can use (4) and the completeness relation (7) to write

$$\begin{aligned}\langle q_b, t_b | q_a, t_a \rangle &= \int dq_1 \cdots dq_m \prod_{j=0}^{m} \langle q_{j+1}, t_{j+1} | q_j, t_j \rangle \\ &= \int dq_1 \cdots dq_m \\ &\quad \times \prod_{j=0}^{m} \langle q_{j+1} | \exp[-i/\hbar (t_{j+1}-t_j) H(\mathbf{P},\mathbf{Q})] | q_j \rangle,\end{aligned} \tag{34}$$

where $q_{m+1} \equiv q_b$ and $q_0 \equiv q_a$. (34) is an illustration of the theorem that amplitudes for events occurring in succession in time multiply—this follows from the requirement of completeness of position states.

Now let $h(p,q)$ be the Weyl transform of $H(\mathbf{P},\mathbf{Q})$. To *first order* in the time interval $t_{j+1}-t_j$, we have

$$\exp[(-i/\hbar)(t_{j+1}-t_j)H(\mathbf{P},\mathbf{Q})] \longrightarrow \exp[(-i/\hbar)(t_{j+1}-t_j)\,h(p,q)]. \tag{35}$$

Indeed, the Weyl transform of $\mathbf{I} - (i/\hbar)(t_{j+1}-t_j)\,\mathbf{H}$ is $1 - (i/\hbar)(t_{j+1}-t_j)h$, which can again be written as an exponential. Now, the crucial point is to rewrite the above time-evolution operator as the inverse transform of its transform, using (12); then the matrix element of the right-hand side of (34) reduces to a definite integral:

$$\begin{aligned}&\langle q_{j+1} | \exp[-i(t_{j+1}-t_j)\mathbf{H}/\hbar] | q_j \rangle \\ &\quad = \langle q_{j+1} | (2\pi\hbar)^{-n} \int \exp[-i(t_{j+1}-t_j)\,h(p,q)/\hbar] \\ &\qquad \times \Delta(p,q)\,dp\,dq\,|\,q_j\rangle \\ &\quad = (2\pi\hbar)^{-n} \int dp \exp\left\{\frac{i}{\hbar}\left[p\cdot\left(\frac{q_{j+1}-q_j}{t_{j+1}-t_j}\right)\right.\right. \\ &\qquad \left.\left. - h\left(p, \frac{q_{j+1}+q_j}{2}\right)\right](t_{j+1}-t_j)\right\},\end{aligned} \tag{36}$$

where we have made use of (30). (34) now becomes

$$\begin{aligned}&\langle q_b, t_b | q_a, t_a \rangle \\ &\quad = \int dq_1 \cdots dq_m \left(\prod_{j=0}^{m} \int \frac{dp}{(2\pi\hbar)^n} \right. \\ &\qquad \left. \times \exp\left\{\frac{i}{\hbar}\left[p\cdot\frac{(q_{j+1}-q_j)}{t_{j+1}-t_j} - h\left(p, \frac{q_{j+1}+q_j}{2}\right)\right](t_{j+1}-t_j)\right\}\right)\end{aligned} \tag{37}$$

$$\begin{aligned}&= \int \frac{dq_1 \cdots dq_m\, dp_0\, dp_1 \cdots dp_m}{(2\pi\hbar)^{(m+1)n}} \\ &\quad \times \exp\left\{\frac{i}{\hbar}\sum_{j=0}^{m}\left[p_j\cdot\left(\frac{q_{j+1}-q_j}{t_{j+1}-t_j}\right)\right.\right. \\ &\qquad \left.\left. - h\left(p_j, \frac{q_{j+1}+q_j}{2}\right)\right](t_{j+1}-t_j)\right\}.\end{aligned} \tag{38}$$

Relations (37) and (38) have been rigorously derived from the Schrödinger equation and are valid to first order in $\max(t_{j+1}-t_j)$. In the limit $m \to \infty$ when $\max(t_{j+1}-t_j) \to 0$ together with $\sum_{j=0}^{m}(t_{j+1}-t_j) = t_b - t_a$, the integrand in (38) tends to a well-defined functional in phase space, and we can write the transition amplitude as a path integral in *phase space* (the dot over the equal sign indicates that this is a symbolic notation):

$$\begin{aligned}\langle q_b, t_b | q_a, t_a \rangle \doteq \int \left[\frac{dp\,dq}{(2\pi\hbar)^n}\right] \\ \times \exp\left(\frac{i}{\hbar}\int_{t_a}^{t_b} [p(t)\cdot\dot{q}(t) - h(p(t),q(t))]\,dt\right).\end{aligned} \tag{39}$$

The passage to the limit raises serious and difficult mathematical problems about which there is an abundant literature. We shall not be concerned with them here, as we only wish to make a qualitative analysis. A number of general statements can be made on the basis of (38) and (39), direct consequences of the Schrödinger equation.

a) "Principle of democracy": Each path in phase space contributes equally in magnitude. Note even the absence of a mesh-dependent "normalization factor." Note also that this is true only if $h(p,q)$ is real, i.e., if $H(\mathbf{P},\mathbf{Q})$ is Hermitian. This property may not be true for non-Hermitian Hamiltonians.

b) Each path in phase space contributes with a phase equal to $1/\hbar$ times an "effective action" taken along it. Since, by virtue of (17) and the Hermiticity of $\mathbf{H}$, we have

$$h(p,q) = H_c(p,q) + O(\hbar^2) \tag{40}$$

where $H_c(p,q)$ is the classical Hamiltonian, we see that in the classical limit the phase is proportional to the classical action. As one can easily verify, many Hamiltonians of physical interest, such as $\mathbf{P}^2/2m + V(\mathbf{Q})$ or $(1/2m)[\mathbf{P} - (e/c)A(\mathbf{Q})]^2 + e\phi(\mathbf{Q})$ are such that[6] $h(p,q) = H_c(p,q)$, a coincidence which conceals this subtle point: It is h and not H_c which rightfully belongs in the exponent of (39). $h(p,q)$ is always equal to $H_c(p,q)$ *only* if the operator $\mathbf{H} = H(\mathbf{P},\mathbf{Q})$ is *defined* to be the inverse Weyl transform of the classical Hamiltonian $H_c(p,q)$, i.e., if Weyl's rule for ordering the factors is chosen.

Therefore, accepting Feynman's postulate that each path comes with a phase equal to $1/\hbar$ times the action

functional is entirely equivalent to opting for Weyl's rule in constructing the quantum Hamiltonian.

Note that this should in no way be construed as an argument in favor of Weyl's rule. For example, if one associates the operator

$$\mathbf{H} = e\phi(\mathbf{Q}) + c(\Pi^2 + M^2c^2\mathbf{I})^{1/2} = e\phi(\mathbf{Q}) + Mc^2[\mathbf{I} + \tfrac{1}{2}\epsilon\Pi^2 - \tfrac{1}{8}\epsilon^2\Pi^4 + O(\epsilon^3)] \quad (40')$$

(where $\Pi \equiv \mathbf{P} - (e/c)A(\mathbf{Q})$ and $\epsilon \equiv (Mc)^{-2}$) with the classical Hamiltonian $H_c = e\phi(q) + c\{[p - (e/c)A(q)]^2 + M^2c^2\}^{1/2}$ for a relativistic spinless particle in three dimensions in an electromagnetic field, it is easily checked that the Weyl transform of $\mathbf{H}$ is not H_c but a nonterminating series in $\hbar^2$, where H_c is the leading term.

Expression (39) with $H_c(p, q)$ in the exponent was hinted at by Feynman, and later developed by Davies and Garrod.[7]

c) We see from (38) that the time-slicing method consistently yields the correct propagator [to first order in $\max(t_{j+1} - t_j)$] if the position-dependent terms are evaluated at the *midpoint* $(q_{j+1} + q_j)/2$ and *not* anywhere else, such as q_j or $(q_{j-1} + q_j + q_{j+1})/3$. This point is often neglected, and its neglect leads to incorrect results (sometimes used to cast doubt on the validity of the path integral formalism).[8] Here we see that the "midpoint rule" is a direct mathematical consequence of the Schrödinger equation.

d) Similar expressions are obtained if the Hamiltonian is time-dependent, provided $[\mathbf{H}(t), \mathbf{H}(s)] = 0$ for all $t, s \in [t_a, t_b]$. The time-evolution operator is then

$$\exp[(-i/\hbar)\int_{t_j}^{t_{j+1}} H(\mathbf{P}, \mathbf{Q}, s)\, ds]$$

which is

$$\exp[(-i/\hbar)H(\mathbf{P}, \mathbf{Q}, t_j)(t_{j+1} - t_j)]$$

to first order in $t_{j+1} - t_j$. The argument proceeds as before.

It is worth noting for the sake of symmetry that by inserting $m + 2$ complete sets of p states in the propagator [rather than m sets of q states as in (34)] one still obtains (38), but with the q terms evaluated at q_j and the p terms evaluated at the midpoint $(p_{j+1} + p_j)/2$.

The uncertainty principle is automatically satisfied by (39) in that we have no information whatsoever on the initial and final momenta (the particle being sharply localized), and hence the path integral is really an integral over all continuous hypersurfaces in phase space-time spanning the lines $(q = q_a,\ t = t_a)$ and $(q = q_b,\ t = t_b)$ and including them, with no motion backwards in time permitted.

Path integrals in configuration space

We now ask the question: When does the path integral in (39) reduce to a path integral over configuration space alone? One might be tempted to answer that, after performing the p integral in (37) (an ordinary integral), we would be left, in the limit $m \to \infty$, with a path integral over configuration space. However, there is no guarantee that in the limit $m \to \infty$ the integrand will tend to a well-defined functional in configuration space. This point is of crucial importance. We give an example and two counterexamples.

Counterexamples

Consider the Hamiltonian

$$H(\mathbf{P}, \mathbf{Q}) = c(\mathbf{P}^2 + M^2c^2\,\mathbf{I})^{1/2} \quad (41)$$

corresponding to a relativistic spin zero particle ($n = 3$) of mass M. Its Weyl transform is $h(p, q) = c(p^2 + M^2c^2)^{1/2}$. By use of the integral representation of the Hankel function,

$$\int_{\mathbb{R}} \exp[ixy + ib(a^2 + x^2)^{1/2}]\, dx = [-\pi ab/(b^2 - y^2)]H_1^{(1)}(a(b^2 - y^2)^{1/2}), \quad (42)$$

and the formula

$$\frac{d}{dx}\left(\frac{H_1^{(1)}(x)}{x}\right) = \frac{1}{x}H_0^{(1)}(x) - \frac{2}{x^2}H_1^{(1)}(x), \quad (43)$$

we can perform the p integral in (37) to get[9]

$$\langle \vec{q}_{j+1}, t_{j+1} | \vec{q}_j, t_j \rangle = \frac{ck^2T}{4\pi\xi^2}\left(H_0^{(1)}(k\xi) - \frac{2}{k\xi}H_1^{(1)}(k\xi)\right), \quad (44)$$

where $\xi \equiv (c^2T^2 - R^2)^{1/2}$, $\vec{R} \equiv \vec{q}_{j+1} - \vec{q}_j$, $T \equiv t_{j+1} - t_j$, $k \equiv Mc/\hbar$. This propagator satisfies the Klein–Gordon equation

$$\left(\nabla_q^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{M^2c^2}{\hbar^2}\right)\psi(q, t) = 0 \quad (45)$$

at either the initial or the final point. A somewhat less physical counterexample is

$$H(\mathbf{P}, \mathbf{Q}) = \tfrac{1}{8}[\mathbf{P}^3 f(\mathbf{Q}) + 3\mathbf{P}^2 f(\mathbf{Q})\mathbf{P} + 3\mathbf{P} f(\mathbf{Q})\mathbf{P}^2 + f(\mathbf{Q})\mathbf{P}^3] \quad (46)$$

in one dimension. Its Weyl transform is simply $h(p, q) = p^3 f(q)$. By use of the integral representation of the Airy function,

$$\int_{\mathbb{R}} \exp(-iax^3/3 + iacx)\, dx = (2\pi/a^{1/3})\,\mathrm{Ai}(-ca^{2/3}), \quad (47)$$

we can perform the p integral in (37) to get

$$\langle q_{j+1}, t_{j+1} | q_j, t_j \rangle = \hbar^{-2/3}[3(t_{j+1} - t_j)f(s_j)]^{-1/3} \times \mathrm{Ai}\{-(q_{j+1} - q_j)\hbar^{-2/3}[3(t_{j+1} - t_j)f(s_j)]^{1/3}\}, \quad (48)$$

where $s_j \equiv (q_{j+1} + q_j)/2$.

In neither of these two cases do we get a well-defined functional of the path in configuration space as $m \to \infty$. Further, (48) is real, and has no phase one could equate to the action.[10]

Example: Hamiltonian quadratic in the momenta

Consider the classical Hamiltonian function:

$$H_c(p, q) = \frac{1}{2M}\, g^{ij}(q)[p_i - A_i(q)][p_j - A_j(q)] + V(q). \quad (49)$$

The corresponding quantum-mechanical operator is not clearly defined due to the factor ordering ambiguity. Let us try the manifestly Hermitian operator:

$$H(\mathbf{P}, \mathbf{Q}) = (1/2M)\, g^{\alpha}(\mathbf{Q})[\mathbf{P}_i - A_i(\mathbf{Q})]g^{-2\alpha}(\mathbf{Q}) \times g^{ij}(\mathbf{Q})[\mathbf{P}_j - A_j(\mathbf{Q})]g^{\alpha}(\mathbf{Q}) + V(\mathbf{Q}), \quad (50)$$

where

$$g \equiv -\det(g_{ij}) \neq 0 \quad \text{and} \quad g^{ik}g_{kj} = \delta^i{}_j$$

and α is some real number. The Weyl transform of $H(\mathbf{P}, \mathbf{Q})$ is

$$h(p, q) = H_c(p, q) + (\hbar^2/2M)[\tfrac{1}{4}g^{2\alpha}(g^{-2\alpha}g^{ij})_{,ij} + \tfrac{1}{2}(g^{\alpha})_{,i}(g^{\alpha})_{,j}g^{-2\alpha}g^{ij} - \tfrac{1}{2}(g^{\alpha})_{,ij}g^{-\alpha}g^{ij}], \quad (51)$$

where we have made use of (29), and the fact that, according to (21),

$$\tfrac{1}{2}[F(\mathbf{Q})\mathbf{P}_i G(\mathbf{Q}) + G(\mathbf{Q})\mathbf{P}_i F(\mathbf{Q})] \longrightarrow p_i F(q) G(q). \tag{52}$$

We now define the Christoffel symbol

$$\Gamma^i{}_{jk} \equiv \tfrac{1}{2} g^{ia}(g_{ja,k} + g_{ka,j} - g_{jk,a}), \tag{53}$$

which we use to express the derivatives of the metric coefficients and metric determinant:

$$(g^\alpha)_{,i} = 2\alpha g^\alpha \Gamma^l{}_{li} \tag{54}$$

$$g^{ij}{}_{,k} + \Gamma^i{}_{lk} g^{lj} + \Gamma^j{}_{lk} g^{il} = 0 \ (\equiv g^{ij}{}_{;k}). \tag{55}$$

After working out the higher order derivatives and substituting in (51), we finally get

$$\begin{aligned} h(p,q) = H_c(p,q) + (\hbar^2/2M) g^{ij}[&(2\alpha + \tfrac{1}{4})\Gamma^m{}_{ml}\Gamma^l{}_{ij} \\ &- \tfrac{1}{4}\Gamma^l{}_{ij,l} - (2\alpha + \tfrac{1}{4})\Gamma^s{}_{si,j} + \tfrac{1}{2}\Gamma^l{}_{mi}\Gamma^m{}_{lj} \\ &+ (2\alpha + \tfrac{1}{2})^2 \Gamma^m{}_{ml}\Gamma^l{}_{ij}]. \end{aligned} \tag{56}$$

The only covariant combination we can form with (56) would have to involve the Ricci scalar curvature tensor:

$$R \equiv g^{ij}(\Gamma^l{}_{lj,i} - \Gamma^l{}_{ij,l} + \Gamma^l{}_{mj}\Gamma^m{}_{il} - \Gamma^l{}_{ij}\Gamma^m{}_{ml}). \tag{57}$$

For the choice

$$\alpha = -\tfrac{1}{4} \tag{58}$$

we have

$$h(p,q) = H_c(p,q) + (\hbar^2/8M)[R(q) + \Gamma^m{}_{li}(q)\Gamma^l{}_{mj}(q) g^{ij}(q)]. \tag{59}$$

Since $\mathbf{P}_i$ is represented in coordinate space by $-i\hbar\partial/\partial q^i$, we easily infer that $H(\mathbf{P},\mathbf{Q})$ in (50) with $\alpha = -\frac{1}{4}$ is covariant when operating on a scalar density of weight $\frac{1}{2}$, which is what $|q,t\rangle$ is if $|\psi(q,t)|^2 dq \equiv |\langle\psi|q,t\rangle|^2 dq$ is to be an invariant.

The Weyl transform formalism, and in particular the plane wave (6), not being covariant, we should not expect $h(p,q)$ to be covariant even when $H(\mathbf{P},\mathbf{Q})$ is.

One might think of (49) as the Hamiltonian describing a particle of mass M moving in a curved n-dimensional space with metric $g_{ij} = g_{ji}$ in the presence of a vector potential A_i and a scalar potential V. Since our formalism is valid only when the q's are rectangular coordinates, we can assume that the curved space is embedded in $\mathbb{R}^N$ ($N \geq n$), so that we can use rectangular coordinates to write the line element in our curved space.

Let us now perform the p integral in (37) with $h(p,q)$ given by (59). We need the well-known result

$$\begin{aligned} &\int_{\mathbb{R}^n} dp \exp[ib\cdot p - (B^{-1})^{kl} p_k p_l/2] \\ &\quad = (2\pi)^{n/2}\sqrt{\det B}\ \exp(-B_{ij}b^i b^j/2). \end{aligned} \tag{60}$$

for any invertible matrix B. With the change of variable $p_i + A_i(q) = p_i'$ our integral can be brought to this form, and the result is:

$$\begin{aligned} \langle q_b, t_b | q_a, t_a\rangle = \int dq_1 \cdots dq_m \Bigg[\prod_{j=0}^{m}\left(\frac{M}{2\pi i\hbar(t_{j+1}-t_j)}\right)^{n/2} \\ \times g^{1/2}\left(\frac{q_{j+1}+q_j}{2}\right)\Bigg] \exp\left(\frac{i}{\hbar}\sum_{j=0}^{m} l_j(t_{j+1}-t_j)\right), \end{aligned} \tag{61}$$

where

$$\begin{aligned} l_j \equiv \frac{M}{2}&\left(\frac{q_{j+1}-q_j}{t_{j+1}-t_j}\right)^s \left(\frac{q_{j+1}-q_j}{t_{j+1}-t_j}\right)^t g_{st}\left(\frac{q_{j+1}+q_j}{2}\right) \\ &+ A_s\left(\frac{q_{j+1}+q_j}{2}\right)\left(\frac{q_{j+1}-q_j}{t_{j+1}-t_j}\right)^s - V\left(\frac{q_{j+1}+q_j}{2}\right) - \frac{\hbar^2}{8M} \\ &\times\left[R\left(\frac{q_{j+1}+q_j}{2}\right) + \Gamma^s{}_{ta}\left(\frac{q_{j+1}+q_j}{2}\right)\right. \\ &\left.\times\ \Gamma^t{}_{sb}\left(\frac{q_{j+1}+q_j}{2}\right) g^{ab}\left(\frac{q_{j+1}+q_j}{2}\right)\right]. \end{aligned} \tag{62}$$

In the limit $m \to \infty$, the integrand indeed tends to a well-defined functional, and we can write the propagator as a path integral in configuration space alone:

$$\langle q_b, t_b | q_a, t_a\rangle \doteq \int [dq/N] \exp\{(i/\hbar) S_{\text{eff}}[q]\}, \tag{63}$$

where the "effective action" functional is

$$\begin{aligned} S_{\text{eff}}[q] \equiv \int_{t_a}^{t_b} dt\{&\tfrac{1}{2}M g_{ij}(q)\dot{q}^i\dot{q}^j + A_i(q)\dot{q}^i \\ &- V(q) - (\hbar^2/8M)[R(q) + \Gamma^m{}_{sj}(q)\Gamma^s{}_{im}(q) g^{ij}(q)]\}. \end{aligned} \tag{64}$$

This is seen to be, in the classical limit $\hbar \to 0$, the time integral of the Lagrangian associated with (49). Note the automatic appearance of the correct "normalization factor" N in (61). Note also that since the factor $g^{1/2}$ is evaluated at the midpoint, it cannot be used to make the volume element $dq_1 \cdots dq_m$ an invariant.

B. Transition between momentum states

The foregoing results are completely symmetric with respect to momentum and position. We can write the transition amplitude for a particle with momentum p_a at time t_a to have momentum p_b at time t_b in a manner similar to (34),

$$\begin{aligned} \langle p_b, t_b | p_a, t_a\rangle = \int dp_1 \cdots dp_m \\ \times \prod_{j=0}^{m} \langle p_{j+1}| \exp[(-i/\hbar)(t_{j+1}-t_j)H(\mathbf{P},\mathbf{Q})]|p_j\rangle, \end{aligned} \tag{65}$$

by inserting m complete sets of momentum states and using (4). Again, we write the time-evolution operator as the inverse transform of its transform, to first order in $\max(t_{j+1}-t_j)$ and use (31), with the result

$$\begin{aligned} \langle p_b, t_b | p_a, t_a\rangle = \int dp_1 \cdots dp_m \prod_{j=0}^{m} \int \frac{dq}{(2\pi\hbar)^n} \\ \times \exp\left\{\frac{i}{\hbar}\left[-q\cdot\left(\frac{p_{j+1}-p_j}{t_{j+1}-t_j}\right) - h\left(\frac{p_{j+1}+p_j}{2}, q_j\right)\right]\right. \\ \left.(t_{j+1}-t_j)\right\} \end{aligned} \tag{66}$$

$$\begin{aligned} = \int \frac{dp_1 \cdots dp_m dq_0 dq_1 \cdots dq_m}{(2\pi\hbar)^{(m+1)n}} \\ \times \exp\left\{-\frac{i}{\hbar}\sum_{j=0}^{m}\left[q_j\cdot\left(\frac{p_{j+1}-p_j}{t_{j+1}-t_j}\right) + h\left(\frac{p_{j+1}+p_j}{2}, q_j\right)\right]\right. \\ \left.(t_{j+1}-t_j)\right\} \end{aligned} \tag{67}$$

$$\doteq \int\left[\frac{dp\,dq}{(2\pi\hbar)^n}\right] \exp\left(-\frac{i}{\hbar}\int_{t_a}^{t_b}[q\cdot p - h(p,q)]\,dt\right) \tag{68}$$

$$\begin{aligned} \doteq \int\left[\frac{dp\,dq}{(2\pi\hbar)^n}\right] \exp\left(\frac{i}{\hbar}[p_a\cdot q(t_a) - p_b\cdot q(t_b)]\right) \\ \times\exp\left(\frac{i}{\hbar}\int_{t_a}^{t_b}[p\cdot\dot{q} - h(p,q)]\,dt\right) \end{aligned} \tag{69}$$

since $q\cdot\dot{p} = (q\cdot p)^{\cdot} - \dot{q}\cdot p$. Note that this formula differs

from (39) by insertion of an extra functional in the integrand [$q(t_a)$ and $q(t_b)$ are completely undetermined]. From (66) we see that it is the p terms this time which must be evaluated at the midpoint $(p_{j+1}+p_j)/2$ and that the path integral can be reduced to a path integral in momentum space alone only if h is quadratic *in the q's*, i. e., not beyond the harmonic oscillator (this is why path integrals in momentum space have generated little interest).

V. CONCLUSION

Weyl's correspondence has provided us with a "royal road" from the Schrödinger equation to the Feynman path integral. It is clear now that in order to generalize this path integral approach to arbitrary Hamiltonians, one must work in phase space with the Weyl transform of the Hamiltonian operator, rather than in configuration space with the classical action.

ACKNOWLEDGMENTS

I gratefully acknowledge discussions with Professor Bryce S. DeWitt and with Dr. Peter A. Hogan.

*This research was supported by the National Science Foundation, Grant No. MPS 74-03563 A01.

[1]H. Weyl, Z. Physik **46**, 1 (1927); *Theory of Groups and Quantum Mechanics* (Dover, 1950), p. 275.

[2]For a discussion of these various rules, see J.R. Shewell, Am. J. Phys. **27**, 16 (1959). For example, Weyl's rule does not satisfy Von Neumann's requirement that if $O(A)$ is the operator corresponding to the classical function A, than $O(f(A)) = f(O(A))$.

[3] (a) S.R. DeGroot and L.G. Suttorp. *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972), Chap. VI and Appendix. (b) See also B. Leaf, J. Math. Phys. **9**, 65 (1968) and H.J. Groenewold, Physica **12**, 405 (1946).

[4]N.H. McCoy, Proc. Natl. Acad. Sci. **18**, 674 (1932).

[5]R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[6]These cases were suggested as a problem in DeGroot and Suttorp, Ref. 3a, p. 366, where the procedure is outlined.

[7]R. P. Feynman, Phys. Rev. **84**, 108 (1951), Appendix B; H. Davies, Proc. Camb. Phil. Soc. **59**, 147 (1963), Claude Garrod, Rev. Mod. Phys. **38**, 483 (1966). See also unpublished notes by J. Plebański, "A Phase Space Approach to Quantum Mechanics."

[8]L.S. Schulman ["Caustics and Multivaluedness: Two Results of Adding Path Amplitudes," in *Proceedings of the International Conference on Functional Integration and Its Applications* (London, 1974) (to be published)] has given an interesting physical explanation of this rule: It follows from the requirement that a gauge transformation induce only a change in phase in the propagator (for the simple case of a particle in an electromagnetic field). Here we see that the rule can be extended to arbitrary Hamiltonians.

[9]This result is *exact* even if $t_{j+1}-t_j$ is not small. Indeed, for Hamiltonians $f(\mathbf{P})$ depending only on momentum, the propagator can be reduced to a definite integral:

$$\langle q_b, t_b \mid q_a, t_a\rangle = \int \langle q_b \mid p\rangle\langle p \mid \exp[(-i/\hbar)(t_b-t_a)f(\mathbf{P})] \mid q_a\rangle\, dp$$
$$= h^{-n}\int dp \exp\{(i/\hbar)[p\cdot(q_b-q_a) - f(p)(t_b-t_a)]\}.$$

See Kursunoglu, *Modern Quantum Theory* (Freeman, San Francisco, 1962), p. 162, for our specific example.

[10]*Discussion*: It must be pointed out, however, that when one performs the p integral in (37), there are often many different ways one can rewrite the resulting expression, since we are working only to first order in $\epsilon \equiv t_{j+1}-t_j$. One of these forms *may* result, in the limit, in an expression such as (63) for an appropriate normalization factor. We know of no systematic way to find this form, if it exists. In our counterexamples, it does not seem to exist, although we come close: a) In the case of the relativistic spinless particle (41) ($c=1$) in one dimension, one has, using (42),

$$\langle q_b, t_b \mid q_a, t_a\rangle = \int dq_1 \cdots dq_m \prod_{j=0}^{m} \frac{\pm M\epsilon}{2(\epsilon^2-\Delta^2)} H_1^{(1)}\left(\pm \frac{M}{\hbar}\sqrt{\epsilon^2-\Delta^2}\right),$$

where $\Delta \equiv q_{j+1}-q_j$. This tends to no functional in the limit $\epsilon \to 0$, but if we argue that $\hbar$ is small, and use the asymptotic formula for the Hankel function [Abramowitz and Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), p. 364],

$$H_1^{(1)}(z) \underset{|z|\to\infty}{\to} \left(\frac{2}{\pi z}\right)^{1/2} \exp(-3\pi i/4)\exp(iz),$$

we get

$$\langle q_b, t_b \mid q_a, t_a\rangle = \int dq_1 \cdots dq_m \left\{\prod_{j=0}^{m}\left(\pm\frac{Mi\hbar}{2\pi}\right)^{1/2}(\epsilon^2-\Delta^2)^{-5/4}\right\}$$
$$\exp\left[\pm\frac{i}{\hbar}\sum_{j=0}^{m} M\left(1-\frac{\Delta^2}{\epsilon^2}\right)^{1/2}\epsilon\right].$$

Since the associated Lagrangian is $-M(1-\dot{q}^2)^{1/2}$, it is obvious that in the limit we get the "Lagrangian" path integral (63). This argument, however, is strikingly artifical, and even erroneous, since, in the limit, the argument of the Hankel function behaves as $\pm M(\epsilon/\hbar)(1-\dot{q}^2)^{1/2}$ which is not so large after all, $\hbar$ being fixed. For small z, $H_1^{(1)}(z) \sim -2i/\pi z$, and we do not get anywhere near (63).

b) The propagator (48) for the Hamiltonian cubic in p is real, and hence cannot be expressed as (63). To see this better, let us use the asymptotic expansion of the Airy function for large arguments (the ϵ dependence of the argument is $\epsilon^{-1/3}$) as given by Abramowitz and Stegun, p. 448 [assume $f(q) > 0$]. This gives a result of the form

$$\langle q_b, t_b \mid q_a t_a\rangle = \int dq_1 \cdots dq_m \prod_{j=0}^{m}\{N_1(\epsilon,\Delta,s)\exp[-(1/\hbar)\epsilon L_j]Y(-\Delta)$$
$$+ Y(\Delta)N_2(\epsilon,\Delta,s)\sin[(1/\hbar)\epsilon L_j + \pi/4] + Y(\Delta)N_3(\epsilon,\Delta,s)$$
$$\times\cos[(1/\hbar)\epsilon L_j + \pi/4]\},$$

where the N's are real functions, Y is the Heaviside step function, $s=(q_{j+1}+q_j)/2$ and L_j is the discrete version of the corresponding Lagrangian, $L = 2\dot{q}^{3/2}f^{-1/2}(q)/3\sqrt{3}$. Therefore, even though the latter does appear, in the limit we do not get the "Lagrangian" path integral (63).

Invariant imbedding and the Fredholm integral equation

Alan P. Wang

Department of Mathematics, Arizona State University, Tempe, Arizona 85282
(Received 20 November 1974; revised manuscript received 18 July 1975)

The standard method of solving the Fredholm integral equation is either to use Neumann series or to convert to an equivalent algebraic equation via the eigenvalues. This paper uses an idea similar to that given by Ueno and others. The integral equation is transformed into a Cauchy problem which can be approximated effectively by modern high speed digital computer.

I. INTRODUCTION

The idea of the principle of invariance or invariant imbedding was applied to solve the Fredholm integral equation by Ueno[1,2] and others,[3–9] in a series of papers. In these papers, two-point boundary value problems are transformed into initial value Cauchy problems, with the use of the Bellman–Krein formula. The invariant imbedding approach is distinctly different from that used in the standard procedure. The standard procedure in solving integral equations is either to make use of the associated Newmann series or to transform them into a system of algebraic equations via the eigenvalues.[10]

In neutron transport, radiative transfer, rarefied gas dynamics and wave propagation, we deal frequently with mathematical models of the form of Fredholm integral equations with boundary values. The importance of the new approach is that when these integral equations are transformed into the Cauchy equation, problems can be more effectively treated by modern computers.

In this paper, we approach the problem using a different method. However, the basic idea is the same as that used by Ueno *et al.*, i.e., the idea of invariant imbedding. Our starting point is the linear-operator equations. Our results reduce to those obtained by Ueno *et al.*, and Bellman-Krein's formula is also obtained as a by-product.

II. LINEAR OPERATOR EQUATIONS

Let L be a linear operator valued function defined on a real line R. It maps a space D into itself. It is assumed that L is differentiable. We shall discuss the following equation:

$$u = g + Lu, \tag{1}$$

where $g \in D$ and g is independent of R. Let K be the resolvent of the operator L; then a unique solution u of the above equation can be expressed as

$$u = g + Kg. \tag{2}$$

It is noted that the linear operator K is differentiable with respect to $x \in R$, because L is differentiable and

$$\frac{d}{dx}[(1 - L)(1 + K)] = 0.$$

Then it follows immediately that u is also differentiable and satisfies the equation

$$u_x = K_x g, \tag{3}$$

where the subscript denotes differentiation. Let Φ be a linear operator having the same domain and range as L and satisfying the operator equation

$$\Phi = L_x + L\Phi, \tag{4}$$

i.e., for each $u \in D$,

$$\Phi u = L_x u + L\Phi u. \tag{5}$$

Upon differentiating Eq. (1), we obtain

$$u_x = L_x u + L u_x. \tag{6}$$

By the uniqueness of the solution, it follows that

$$u_x = \Phi u = \Phi(g + Kg). \tag{7}$$

Using Eqs. (3) and (7) and the fact K is differentiable, we have the operator equation

$$K_x = \Phi(1 + K). \tag{8}$$

By taking advantage of the resolvent, Φ in (5) can be expressed as

$$\Phi = (1 + K)L_x. \tag{9}$$

By combining the above two Eqs. (8) and (9), the following Riccati type operator equation is obtained:

$$K_x = (1 + K)L_x(1 + K). \tag{10}$$

Thus we have transformed an operator equation (1) to a Cauchy system (10). To show that the above analyses apply to the Fredholm integral equation, we shall consider the integral equation with an arbitrary kernel, and we shall show that some assumptions used in this section can be proved for the case of integral equations.

III. THE FREDHOLM INTEGRAL EQUATION

Let us consider the well-known Fredholm equation

$$u(t, x) = g(t) + \int_0^x l(t, y)u(y, x)\,dy, \quad 0 \le t \le x, \tag{11}$$

where $g(t)$ and $l(t, y)$ are continuous real valued functions. (Note that differentiation is not assumed.) The function g is defined on $[a, b]$ and l is defined on $[a, b] \times [a, b]$. For a fixed pair (t, x), we may view

$$(Lu)(t, x) = \int_0^x l(t, y)u(y, x)\,dy,$$

as a linear operator L on $u(\cdot, x)$. It is well known that the resolvent kernel exists, and it can be written as

$$(Kg)(t, x) = \int_0^x k(t, y\,; x)g(y)\,dy.$$

The resolvent kernel is related to $l(t, x)$ by

$$k(t, x; x) = l(t, y) + \int_0^x k(t, z; x) l(z, y)\, dz = l(t, y) + \int_0^x l(t, z) k(z, y; x)\, dz. \tag{12}$$

In the previous section we assumed that L is differentiable, and from that it follows that K is differentiable. In this case we do not make that assumption. However, to show that L is differentiable, it is sufficient to show that u is differentiable with respect to x.

Let

$$u(t, x+h) - u(t, x) = \delta(t, x; h).$$

Then from Eq. (11) we have

$$\delta(t, x; h) = \alpha(t, x; h) + \int_0^x l(t, y)\delta(t, y; h)\, dy, \tag{13}$$

where

$$\alpha(t, x; h) = \int_x^{x+h} l(t, y) u(y, x+h)\, dy.$$

By using the resolvent kernel, we have

$$\delta(t, x; h) = \alpha(t, x; h) + \int_0^x k(t, z; x)\alpha(t, z; h)\, dz.$$

The right-hand side of the above equation has a limit, as $h \to 0$, since every point x is a Lebesgue point. From this it follows that $u(t, x)$ is differentiable with respect to x almost everywhere in $[a, b]$. In fact we have that

$$\frac{\partial}{\partial x} u(t, x) = l(t, x)u(x, x) + \int_0^x k(t, z; x) l(t, z) u(z, z)\, dz \tag{14}$$

holds almost everywhere in $[a, b]$. In what follows, all the equations involving a partial derivative with respect to x hold almost everywhere in $[a, b]$.

Now, corresponding to Eq. (10), we have

$$\begin{aligned}\frac{\partial}{\partial x}\int_0^x k(t, y; x) g(y)\, dy &= l(t, x)g(x) + \int_0^x l(t, x) k(t, z, ; x) g(z)\, dz \\ &\quad + \int_0^x k(t, z; x) l(t, z) g(z)\, dz \\ &\quad + \int_0^x \int_0^x k(t, y; x) l(y, z) k(t, z; x) g(z)\, dz\, dy. \end{aligned} \tag{15}$$

This is equivalent to the equation

$$\frac{\partial}{\partial x} K = l + K + Kl + KlK$$

with the understanding that for scalar multiplication

$$(lg)(t, x) = l(t, x)g(x).$$

Upon substituting the integral representation

$$(\Phi u)(t, x) = \int_0^x \varphi(t, y) u(y, x)\, dy,$$

into Eq. (8), some algebraic manipulation yields the well-known Bellman-Krein formula

$$k_x(t, y; x) = \varphi(t, x) k(x, y; x) = k(t, x; x) k(x, y; x). \tag{16}$$

IV. THE DEGENERATE KERNEL AND ITS LIMIT

Let us consider the degenerate kernel

$$l(t, x) = \sum_{i=1}^n \alpha_i(t)\beta^i(x),$$

where $\{\alpha_i\}$ and $\{\beta^i\}$ are two orthonormal sets of continuous functions. Let P_i and P^j be two projections defined by

$$P_i : f(t, x) \to \int_0^\infty f(t, y)\beta^i(y)\, dy,$$

$$P^j : f(t, x) \to \int_0^\infty \beta^j(y) f(y, x)\, dy.$$

For convenience we shall write

$$P_i f = f_i, \quad P^j f = f^j, \quad \text{and } P^j P_i f = f_i^j.$$

Under the projection P_i, Eq. (12) reduces to

$$k_i(t, x) = \alpha_i(t) + \int_0^x l(t, z) k_i(z, x)\, dz. \tag{17}$$

This equation can be expressed in terms of the resolvent

$$k_i(t, x) = \alpha_i(t) + \int_0^x k(t, z; x)\, \alpha_i(z)\, dz. \tag{18}$$

On differentiation of Eq. (18) with respect to x, we have

$$\frac{\partial}{\partial x} k_i(t, x) = k(t, x; x)\alpha_i(x) + \int_0^x \frac{\partial}{\partial x} k(t, z; x)\alpha_i(z)\, dz. \tag{19}$$

By taking advantage of the Bellman—Krein formula and the expansion

$$k(t, x; x) = \sum_{j=1}^n \beta^j(x) k^j(t, x), \tag{20}$$

Eq. (19) reduces to the form

$$\frac{\partial}{\partial x} k_i(t, x) = \alpha_i(t, x) \sum_{j=1}^n \beta^j(x) k^j(t, x). \tag{21}$$

Note that Eq. (17) can be expressed as

$$k_i(t, x) = k_i(t) + \sum_{j=1}^n \alpha_j k_i^j(x). \tag{22}$$

On the other hand, according to the definition for $k_i^j(x)$,

$$\frac{\partial}{\partial x} k_i^j(x) = \frac{\partial}{\partial x}\int_0^x \beta^j(y) k_i(y, x)\, dy = \beta^j(x) k_i(x, x) + \int_0^x \beta^j(y) \frac{\partial}{\partial x} k_i(y, x)\, dy. \tag{23}$$

Upon a substitution of Eqs. (21) and (22) into Eq. (23), we find that a Cauchy system holds almost everywhere,

$$\frac{\partial}{\partial x} k_i^j(x) = \left(\alpha_i(x) + \sum_{j=1}^n \alpha_j(x) k_i^j(x)\right) \times \left(\beta^j(x) + \sum_{i=1}^n \beta^j k_i^j(x)\right) \tag{24}$$

with initial values

$$k_i^j(x) = 0 \quad \text{for all } i, j = 1, 2, \ldots, n.$$

By Eqs. (18) and (20) the kernel is given by

$$k(t, y; x) = \sum_{i=1}^n \beta^j(x) k_i^j(t, x) = \sum_{i=1}^n \beta^j(x)\alpha_i(x) + \sum_{i=1}^n \alpha_i(x) k_j^i(x)$$

$$= \sum_{i=1}^{n} \alpha_i(x)\beta^i(x) + \sum_{j=1}^{n}\sum_{j=1}^{n} \alpha_j(x)k_j^i(x)\beta^i(x). \tag{25}$$

Equation (25) also can be expressed in matrix form

$$k(t, y; x) = \boldsymbol{\alpha}\boldsymbol{\beta}^t + \boldsymbol{\alpha}\mathbf{K}\boldsymbol{\beta}^t, \tag{26}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)$, and K is an $n \times n$ matrix with ith row and jth column element k_j^i.

Since an arbitrary kernel can be approximated uniformly by a degenerate kernel, it follows that the resolvent kernel expressed in Eq. (26) can also be used to approximate uniformly a resolvent of an arbitrary kernel. Let L be an operator associated with an arbitrary kernel $l(t, x)$, and let L_n be the corresponding operators associated with a sequence of degenerate kernels $l_n(t, x)$, with $l_n(t, x)$ converging uniformly to $l(t, x)$. Then under the uniform norm,

$$\|K - K_n\| \leq \|L - L_n\| + \|L_n\|\|K_n - K\| + \|L - L_n\|\|K\|,$$

where K and K_n are resolvent operators of L and L_n respectively, we have

$$\|K - K_n\| \leq \frac{1 - \|K\|}{1 - \|L\|}\|L - L_n\|,$$

where $\|L\| \neq 1$. Therefore, $L_n <= L$ (converges uniformly) implies $K_n <= K$ also. That is, the result in (26) can be used to approximate an arbitrary kernel uniformly.

ACKNOWLEDGMENT

The author is pleased to acknowledge his indebtedness to Professor S. Ueno for the privilege of seeing Ref. 1 in manuscript form.

[1]S. Ueno, J. Math. Phys. **15**, 1939 (1974).
[2]S. Ueno, Dept. Elect. Eng., Univ. S. Cal., TR. No. RB 73-30, 1973, to appear in J. Math. Anal. Appl.
[3]R. Bellman and S. Ueno, J. Math. Phys. **14**, 1489 (1973).
[4]R. Bellman and S. Ueno, Space Sci. Rev. **16**, 241 (1972).
[5]R. Bellman and S. Ueno, Univ. S. Cal., TR. No. 72-22, May 1972.
[6]H.H. Kagiwada and R. Kalaba, Assoc. Comput. Mach. **17**, 412 (1970).
[7]R.E. Kalaba and E. Zagustin, J. Franklin Inst. **293**, 277 (1971).
[8]T.K. Leong and K.K. Sen, Publ. Astron. Soc. Jpn. **23**, 99 (1971).
[9]A.P. Wang, J. Math. Phys. **14**, 855 (1973).
[10]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953).

Ergodic properties of a particle moving elastically inside a polygon

Arthur Hobson

Department of Physics, University of Arkansas, Fayetteville, Arkansas 72701
(Received 16 May 1975)

The flow of a classical particle bouncing elastically inside an arbitrary polygon is investigated. If every interior angle is a rational multiple of π, there exists precisely one isolating integral in addition to the energy; this integral is described in detail; any possible third integral is nonisolating. If one or more interior angles is an irrational multiple of π, the second integral becomes everywhere nonisolating and non-Lebesgue-measurable, i.e., the second integral disappears. The flow of two hard points bouncing elastically in a finite one-dimensional box is equivalent to the flow of a point particle moving elastically inside a right triangle having interior angle $\tan^{-1}(m_2/m_1)^{1/2}$, so the preceding remarks apply to this model. Nonrigorous arguments are given in support of the notion that the polygon model is ergodic and mixing, but is not a C-system.

I. INTRODUCTION

There has been a recent resurgence of interest in the rigorous analysis of simple classical Hamiltonian systems, mainly in the hope of illuminating some basic problems of statistical mechanics.[1-5] For instance, Sinai's proof[1] that the N-body, hard sphere gas in a finite box has the strongly "random" K-system property (and is therefore also mixing and ergodic) makes it appear likely that realistic classical many-body systems have such "randomness" built into their mechanics.

Lebowitz has suggested[3] that it would be interesting to study the ergodic properties of the unequal-mass one-dimensional hard point gas, i.e., N hard points moving elastically inside a finite one-dimensional box. Even for the case $N=2$, this problem turns out to be surprisingly complicated. Casati and Ford[6] have recently performed computer experiments for the case $N=2$, and have concluded that the system appears to be ergodic and mixing when the parameter θ defined by

$$\tan\theta=(m_2/m_1)^{1/2} \tag{1}$$

is an irrational multiple of π, and that the system is not a C-system (and hence not a K-system). Of course, computer experiments cannot *prove* such results.

As is proven in the Appendix, the two-body hard point gas may be canonically transformed into a single point moving elastically (i.e., at constant speed, with equal-angle bounces from the walls) in two dimensions inside a right triangle with interior angle θ given by (1). Hence the ergodic properties of this "right-triangle model" are identical to those of the two-body hard point gas. But a brief study of the right-triangle model reveals that no extra complication is introduced by generalizing to the "polygon model," i.e., a single point of mass m moving elastically in two dimensions inside an arbitrary (but nonintersecting) polygon with interior angles $\theta_1, \theta_2, \cdots, \theta_n$ (Fig. 1).

This paper is mainly devoted to the statement and proof of some of the ergodic properties of the polygon model (see Sec. II, Theorems 1, 2, and 3). The primary question, i.e., ergodicity or nonergodicity of the model, remains open, although ergodicity is strongly indicated (see Sec. III). Hopefully, publication of these results will stimulate others to prove (or disprove!) ergodicity of the polygon model.

II. A FEW ERGODIC PROPERTIES OF THE POLYGON MODEL

The model is defined in Sec. I. It is a conservative Hamiltonian system with $H(p_1,p_2)=(p_1^2+p_2^2)/2m$, where $p_i=m\,dq_i/dt$ $(i=1,2)$. Definitions: "Configuration space" Q is the set of all points (q_1,q_2) inside or on the boundary of the polygon. The set $\Gamma=Q\times R^2$ is "phase space," with points denoted (q_1,q_2,p_1,p_2). The "energy surface" $\Gamma(E)$ means all $(q_1,q_2,p_1,p_2)\in\Gamma$ which satisfy $H(p_1,p_2)=E$. The "momentum phase angle" ϕ means the angle which the line from the origin to the point (p_1,p_2) makes with the p_1 axis in momentum space, so that

$$\phi(p_1,p_2)=\tan^{-1}(p_2/p_1)\in[0,2\pi). \tag{2}$$

"Reduced phase space" means the cylinder $X=Q\times[0,2\pi)$ (base Q, height 2π), with points labeled $x=(q_1,q_2,\phi)$. Any moving phase point is confined to a single $\Gamma(E)$; we will henceforth restrict our attention to the motion on $\Gamma(E)$. This motion can be described in terms of the motion of the "reduced phase point" $x=(q_1,q_2,\phi)\in X$, since the mapping from $\Gamma(E)$ to X is one-to-one. The flow in X will be represented by the one-parameter group of transformations T^t, i.e., T^tx means the reduced phase point at time t corresponding to the initial (at $t=0$) point x; t may be positive or negative; if the atom hits a vertex at time $t_0>0$ ($t_0<0$), T^tx is undefined for $t>t_0$ ($t<t_0$). The complete path in X passing through x at $t=0$ is denoted $P(x)$, i.e., $P(x)$

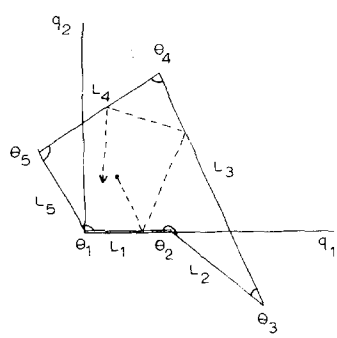

FIG. 1. The polygon model, for the case $n=5$. The dashed line is a portion of a path.

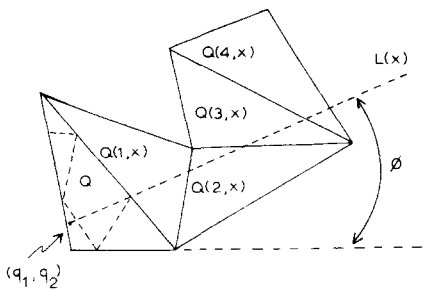

FIG. 2. A portion of the polygon sequence $A(x)$ generated from the point $x=(q_1, q_2, \phi)$ by the polygon Q. The dashed line inside Q is a portion of the actual path $P(x)$ started from x. The five path-segments shown are congruent to the five segments of $L(x)$ cut off by Q, $Q(1,x)$, ..., $Q(4,x)$.

$=\{T^t x : t \in I(x)\}$, where $I(x)$ is the time interval [generally $(-\infty, \infty)$—see Theorem 1] during which $T^t x$ is defined.

General description of the motion in X: Starting from x_0, $T^t x_0$ moves at speed $(2E/m)^{1/2}$ across the plane $\phi = \phi_0$ along a line whose angle with the q_1 axis is $\tan^{-1}(\dot{q}_2/\dot{q}_1) = \tan^{-1}(p_2/p_1) = \phi_0$; when $T^t x_0$ runs into a side of the cylinder X, it instantaneously jumps to a new plane $\phi = \phi_1$, where ϕ_1 is determined from ϕ_0 by the law of reflection; $T^t x_0$ then moves across the plane $\phi = \phi_1$ in the direction ϕ_1.

The construction in Fig. 2 (shown for the special case $n=3$) is helpful in analyzing the flow in X: Draw Q; for any initial point $x=(q_1, q_2, \phi)$, draw the ray $L(x)$ from (q_1, q_2) in the direction ϕ; reflect Q across the side through which $L(x)$ passes, and call the reflection $Q(1, x)$; repeat this operation on $Q(1, x)$, obtaining $Q(2, x)$, etc.; thus we obtain a sequence $A(x)=\{Q(i, x)\}_{i\in Z_+}$ (where Z_+ means the positive integers) of oriented polygons along the ray $L(x)$. The union $A=\cup_{x\in X}A(x)$ of all these sequences is an (overlapping!) array of oriented polygons generated by Q along all possible rays $L(x)$ $(x \in X)$. This "array A generated by Q" is contained in the "*complete* array $\bar{A}$ generated by Q," defined as the set of oriented polygons obtained by reflecting Q across *each* of its sides, then reflecting each of these reflections about each side, etc. That is, an oriented polygon Q' is an element of $\bar{A}$ if and only if Q' can be obtained from Q by some sequence of reflections about successive sides.

Study of Fig. 2 reveals that the portion of $P(x)$ lying between bounces i and $i+1$ is congruent to that portion of $L(x)$ cut off by $Q(i, x) \in A(x)$. Thus we can study the polygon models by studying the properties of rays ("free particle paths") passing through the various sequences $A(x) \subset A \subset \bar{A}$ generated by Q. Note that A is countable, because $\bar{A}$ is.

Define the sets $\Phi_v(q_1, q_2) \subset [0, 2\pi)$, and $\Phi_p(q_1, q_2) \subset [0, 2\pi)$, by

$$\Phi_v(q_1, q_2) = \{\phi : P(q_1, q_2, \phi) \text{ hits a vertex}\},$$

$$\Phi_p(q_1, q_2) = \{\phi : P(q_1, q_2, \phi) \text{ is periodic}\};$$

these sets consist of the angles leading, respectively, to vertex collisions and to periodic paths, starting from (q_1, q_2). Similarly, we define $X_v \subset X$ and $X_p \subset X$ by

$$X_v = \{x : P(x) \text{ hits a vertex}\},$$

$$X_p = \{x : P(x) \text{ is periodic}\}.$$

Theorem 1: (a) $\Phi_v(q_1, q_2)$ and $\Phi_p(q_1, q_2)$ are countable, $\forall (q_1, q_2) \in Q$; (b) X_v and X_p are Lebesgue measurable in X, and both have measure zero.

Proof: Let $\Phi_v(q_1, q_2, D)$ be those elements of $\Phi_v(q_1, q_2)$ leading to paths which hit a vertex in a distance (measured along the path) $\leq D$; similarly, define $X_v(D) \subset X_v$. Let $\Phi_p(q_1, q_2, D)$ be those elements of $\Phi_p(q_1, q_2)$ leading to paths having "periodicity length" (distance, along the path, after which the motion repeats) $\leq D$; similarly, define $X_p(D) \subset X_p$. If we can show that $\Phi_v(q_1, q_2, D)$ and $\Phi_p(q_1, q_2, D)$ are countable, we will have proved (a), since

$$\Phi_v(q_1, q_2) = \bigcup_{D>0} \Phi_v(q_1, q_2, D) = \bigcup_{D\in Z_+} \Phi_v(q_1, q_2, D)$$

is then a countable union of countable sets, and similarly for $\Phi_p(q_1, q_2)$. Similarly, if we can prove that $X_v(D)$ and $X_p(D)$ are measurable and have measure zero, we will have proved (b). Define

$$A(q_1, q_2) = \bigcup_{\phi\in[0,2\pi)} A(q_1, q_2, \phi),$$

and let $A(q_1, q_2, D)$ be those polygons in $A(q_1, q_2)$ which are inside the circle of radius D centered on (q_1, q_2). Now $P(x)$ hits a vertex in a distance $\leq D$ if and only if $L(x)$ (see Fig. 2) hits one of the vertices in $A(q_1, q_2, D)$. But $A(q_1, q_2, D)$ is countable and hence possesses only countably many vertices, and so $\Phi_v(q_1, q_2, D)$ can contain only countably many elements. Similarly, $P(x)$ is periodic, with periodicity length $\leq D$, if and only if $L(x)$ passes through a point $(q_1', q_2') \in Q(i, x) \in A(q_1, q_2, D)$, where $Q(i, x)$ has the same orientation as Q and where (q_1', q_2') $\in Q(i, x)$ is congruent to $(q_1, q_2) \in Q$. But $A(q_1, q_2, D)$ possesses only countably many such congruent points so $\Phi_p(q_1, q_2, D)$ is countable. This proves (a). Define

$$A(D) = \bigcup_{(q_1,q_2)\in Q} A(q_1, q_2, D),$$

i.e., $A(D)$ consists of those elements of A lying within distance D of Q. Then $A(D)$ is countable. Thus, the vertices in $A(D)$ are countable; label them (q_{1j}, q_{2j}) $(j = 1, 2, \cdots)$. In order that $L(x)$ pass through (q_{1j}, q_{2j}) [i.e., in order that x be in $X_v(D)$], it is necessary and sufficient that $x=(q_1, q_2, \phi)$ satisfy the two conditions

$$(q_{2j} - q_2)/(q_{1j} - q_1) = \tan\phi\,, \quad (q_{1j} - q_1)^2 + (q_{2j} - q_2)^2 \leq D^2.$$

For each j, the points x obeying these conditions form a two-dimensional measurable set, of measure zero, in X. Thus $X_v(D)$ is the countable union of sets of measure zero, and so $X_v(D)$ is measurable and has measure zero. Similarly, in order that $x=(q_1, q_2, \phi)$ be in $X_p(D)$, it is necessary and sufficient that $L(x)$ pass through a point (q_1', q_2') congruent to (q_1, q_2), where (q_1', q_2') is in a polygon $Q(i, x) \in A(D)$ which is oriented in the same way as Q. Study of Fig. 2 reveals that, for each correctly oriented $Q(i, x)$, the set of such x has fixed ϕ and covers a two-dimensional measurable region in Q. But any such set is measurable and has measure zero in X. Thus, $X_p(D)$ consists of finitely many measurable sets, each of measure zero, and so $X_p(D)$ is measurable and has measure zero. This completes the proof.

Definitions: An "integral of the motion" (or simply

"integral") in X is any phase function $f(x)$ (domain X) which (except possibly for a set of measure zero in X) is constant along every path $P(x)$ in X. A "measurable integral" means an integral $f(x)$ which is Lebesgue measurable. An "integral surface" means a surface $X' \subset X$ such that $P(x) \subset X'$ whenever $x \in X'$. An "isolating integral" means an integral $f(x)$ such that, for every $x_0 \in X$, the integral surface $\{x : f(x) = f(x_0)\}$ has at most a finite number of sheets in the neighborhood of every $x \in X$. The "number" of (unrestricted, or measurable, or isolating) integrals means the number of *functionally independent* integrals. Extending these definitions to the full phase space Γ, we see that $H(q_1, q_2, p_1, p_2)$ is an isolating integral in Γ with integral surfaces $\Gamma(E)$. We would like to know if there are others, i.e., we would like to know if there are isolating integrals in X. A "rational polygon" means any polygon having every interior angle a rational multiple of π; an "irrational polygon" is one which is not rational, i.e., having at least one interior angle an irrational multiple of π.

Theorem 2: If Q is a rational polygon, then $\exists$ precisely one isolating integral in X. The corresponding isolating integral surfaces are finite sets of cross sections of X, each cross section lying parallel to Q. More precisely, each integral surface $X(\phi) \subset X$ is of the form $X(\phi) = Q \times \Phi(\phi)$, where (for every ϕ) $\Phi(\phi) \subset [0, 2\pi)$ is a *finite* set of phase angles.

Proof: If Q is rational, then the complete polygon array $\bar{A}$ ranges over only a *finite* number of different orientations. Thus (cf. Fig. 2) all paths started in a given direction ϕ [regardless of the initial configuration (q_1, q_2)] must have their phase angles $\phi(t)$ restricted to the finite set $\Phi(\phi)$, defined as the set of phase angles obtained by drawing lines at orientation ϕ through every element of $\bar{A}$, and then reorienting that element to bring it into congruence with Q so that our line now has direction $\phi' \in \Phi(\phi)$. Any polygon $Q' \in \bar{A}$ may be considered to be the "generator" of $\bar{A}$, i.e., we can obtain $\bar{A}$ by starting with any $Q' \in \bar{A}$ and performing reflections. Thus, any path $P(q_1, q_2, \phi')$ having $\phi' \in \Phi(\phi)$ must have *all* of its phase angles in $\Phi(\phi)$. That is, $X(\phi) = Q \times \Phi(\phi)$ is an integral surface. Note that we are not asserting that every $\phi' \in \Phi(\phi)$ is actually realized on a single orbit (this property is indicated, but not proved, by the computer calculations of Casati and Ford[6]; I have been unable to prove this property). The surface $X(\phi)$ is isolating since it has only a finite number of sheets in X. Thus, there exists at least one isolating integral in X. Now suppose there existed a second isolating integral $g(x)$. Then every path $P(x)$ would lie on the intersection of one of the surfaces $X(\phi)$ with one of the isolating surfaces $g(x) = \text{const}$, and hence $P(x)$ would consist of a *finite* number of *finite* line segments and therefore be periodic [unless $P(x)$ ran into a vertex]. But this contradicts Theorem 1, so that there cannot exist a second isolating integral. This completes the proof.

We can describe the surfaces $X(\phi)$ in considerably more detail. Divide the array $\bar{A}$ into a subarray $\bar{A}_+$ consisting of polygons obtained from Q by an *even* number of reflections (these polygons differ from Q by only a rotation and a translation) and the subarray $\bar{A}_-$ consisting of polygons obtained from Q by an *odd* number of reflections (these differ from Q by a rotation, a translation, and a reflection). Correspondingly, the set $\Phi(\phi)$ breaks up into two sets,

$$\Phi(\phi) = \Phi_+(\phi) \cup \Phi_-(\phi), \tag{3}$$

which are either identical or disjoint (depending on the value of ϕ). Figure 2 then reveals that

$$\Phi_+(\phi) = \Phi_+(0) + \phi, \quad \Phi_-(\phi) = \Phi_-(0) - \phi, \tag{4}$$

i.e., any set $\Phi_\pm(\phi)$ is just a translate (through $\pm\phi$) of $\Phi_\pm(0)$. The set $\bar{A}_+$ is generated (via even numbers of reflections) from any one of its members $Q' \in \bar{A}_+$, and similarly for $\bar{A}_-$. Referring to Fig. 2, we then see that any pair of "positive" sets $\Phi_+(\phi)$ and $\Phi_+(\phi')$ is either disjoint or identical, and similarly for the "negative" sets:

$$\begin{aligned} &\Phi_+(\phi) \cap \Phi_+(\phi') = 0 \ \text{ unless } \Phi_+(\phi) = \Phi_+(\phi'), \\ &\Phi_-(\phi) \cap \Phi_-(\phi') = 0 \ \text{ unless } \Phi_-(\phi) = \Phi_-(\phi'). \end{aligned} \tag{5}$$

It should also be noted that

$$\phi \in \Phi_+(\phi) \ \text{ and } \ -\phi \in \Phi_-(\phi). \tag{6}$$

For the rational models, $\Phi_+(\phi)$ and $\Phi_-(\phi)$ each contain the same (finite) number of elements; call this number N. Properties (4) and (5) then imply that $\Phi_+(\phi)$ and $\Phi_-(\phi)$ are each equally spaced, i.e., adjacent members [of $\Phi_+(\phi)$, say] are separated by the fixed distance $d = 2\pi/N$. This fact, coupled with (6), implies that

$$\begin{aligned} &\Phi_+(\phi) = \{\phi, d + \phi, 2d + \phi, \ldots, (N-1)d + \phi\}, \\ &\Phi_-(\phi) = \{-\phi, d - \phi, 2d - \phi, \ldots, (N-1)d - \phi\}, \end{aligned} \tag{7}$$

where each element is taken mod 2π. Now that we know something about the structure of the set $\Phi(\phi)$, it is not difficult to write down the corresponding integral of the motion. In fact, there are any number of simple functions $f_2(x) = f_2(\phi)$ (function of ϕ alone) which are constant on a given surface $Q \times \Phi(\phi)$ and different on different surfaces. One such function is

$$f_2(\phi) = \cos^4(N\phi/4) + \sin^4(N\phi/4). \tag{8}$$

The function $f_2(\phi(p_1, p_2))$, with $\phi(p_1, p_2)$ given by (2) and with some appropriate integer N, is an isolating integral of any rational polygon model.

Most of the preceding properties of $X(\phi)$ also go through for *irrational* models, except that $X(\phi)$ is now everywhere dense and hence nonisolating. For the irrational case, $\bar{A}$ is countably infinite, and the orientations of the various polygons in $\bar{A}$ are dense in $[0, 2\pi)$. Thus, along any path $P(x)$, the phase angles are restricted to a countably infinite set $\Phi(\phi)$ which is dense in $[0, 2\pi)$. Furthermore, since any $Q' \in \bar{A}$ can be considered to be the generator of $\bar{A}$, any path $P(q_1, q_2, \phi')$ with $\phi' \in \Phi(\phi)$ must have all of its phase angles in $\Phi(\phi)$. That is, the everywhere-dense "surface" $X(\phi) = Q \times \Phi(\phi)$ is an integral surface in X. The array $\bar{A}$ again breaks up into $\bar{A}_+$ and $\bar{A}_-$, and the corresponding (dense) sets $\Phi_+(\phi)$ and $\Phi_-(\phi)$ again have the properties (3)—(6). We now show that there exists no measurable integral describing the surfaces $X(\phi)$.

Theorem 3: For any irrational model, the (nonisolating, everywhere-dense) integral surfaces $X(\phi) = Q \times \Phi(\phi)$ correspond to a nonmeasurable integral. More

precisely, any $f(x)$ which is constant on each $X(\phi)$ is either nonmeasurable, or else $f(x)$ is constant a.e. in X.

Proof: Assume f is measurable and constant on each $X(\phi)$. Since $X(\phi) = Q \times \Phi(\phi)$, f depends on ϕ alone. By (3), $f(\phi)$ is constant on $\Phi_+(\phi) \cup \Phi_-(\phi)$, and hence is (in particular) constant on each $\Phi_+(\phi)$. Properties (4) and (6) imply the "periodicity property" $f(\phi + \Delta) = f(\phi)$ $\forall\, \Delta \in \Phi_+(0)$ and $\forall\, \phi \in [0, 2\pi)$. Define

$$F(\phi) = \int_0^\phi f(\phi')\, d\phi' \tag{9}$$

and partition $[0, 2\pi)$ by means of a net having the constant interval width Δ, where $\Delta \in \Phi_+(0)$. Letting K be the number of complete intervals in $[0, \phi)$, we have

$$F(\phi) = \sum_1^K \int_{(k-1)\Delta}^{k\Delta} f(\phi')\, d\phi' + \int_{K\Delta}^{\phi} f(\phi')\, d\phi'$$

$$= K \int_0^\Delta f(\phi')\, d\phi' + \int_{K\Delta}^{\phi} f(\phi')\, d\phi'. \tag{10}$$

Since $\Phi_+(0)$ is everywhere dense in $[0, 2\pi)$, we can let $\Delta \to 0$ while maintaining the condition $\Delta \in \Phi_+(0)$. By taking this limit, (10) implies that the ratio $F(\phi)/F(2\pi)$ is $\phi/2\pi$. Thus $F(\phi) = \phi F(2\pi)/2\pi$, and so $dF/c\phi = \text{const}$. But, by (9), $dF/d\phi = f(\phi)$ a.e. in $[0, 2\pi)$. Thus $f(\phi) = \text{const}$ a.e in X. This completes the proof.

Thus the polygon models provide an example of an integral which is isolating for some models and which becomes not only nonisolating but also nonmeasurable for the remaining models. Our *second* integral f is analogous to the *third* integral δ for the anisotropic two-dimensional harmonic oscillator[7]; the oscillator always possesses two isolating integrals E_1 and E_2 (the energies associated with q_1 and q_2), along with a third integral δ which is isolating only if the frequency ratio ω_1/ω_2 is rational; for irrational ω_1/ω_2 δ becomes nonisolating.

III. DISCUSSION

To proceed with the analysis, we need to study the flow on the integral surfaces $X(\phi)$. Considering for the moment only the "simple" case of rational interior angles, the Poincaré recurrence theorem plus the finite separation of the different sheets of $X(\phi)$ implies that any trajectory passing through a given sheet of $X(\phi)$ will keep returning to the same sheet. Study of Fig. 2 then shows that every sheet (the kth, for instance) is the union of a countable (and perhaps finite) set of "zones" such that each trajectory passing through the ith zone of the kth sheet gets shifted by a characteristic amount ω_i upon returning to the kth sheet. Using this fact, the question of the ergodicity of the flow on $X(\phi)$ may be reduced to the question of the ergodicity of a discrete "interleaving" transformation T: $[0, b) \to [0, b)$ (on the boundary of the kth sheet) having the form $Ty = y + \omega_i$ $(y \in Z_i)$, where the zones $Z_i = [a_{i-1}, a_i)$ are intervals forming a partition of $[0, b)$. The boundaries a_i of the Z_i, as well as the shifts ω_i, are dependent on the model and on ϕ. Katok and Stepin[8] have analyzed the interleaving transformation for the particular case of three zones, and found it to be ergodic for almost every choice of the two zone boundaries a_1, a_2. Thus it seems likely that the transformation T [and hence the flow on $X(\phi)$] is ergodic for almost every ϕ, provided the number of zones is *finite*. For the rational models the flow is obviously not ergodic on the full energy surface X, since each phase path is restricted to one isolating surface $X(\phi)$.

Turning now to the irrational case (i.e., the "general" case, since the rational polygons form a "set of measure zero" among the set of all polygons), things are even more complicated since each $X(\phi)$ has an infinite number of sheets. The "vanishing" of the second integral $f(x)$ in the irrational case (Theorem 3) makes it seem likely that the flow is ergodic on the full energy surface X (cf. Lewis's theorem[4]).

A primary reason for studying simple classical mechanical models is to shed light on the general problems of statistical mechanics. In this regard, the question of whether or not a system is *mixing* is more important than the question of ergodicity, since it is mixing systems, rather than merely ergodic systems (mixing implies ergodicity but not vice versa[2,5]), that have the "irreversibility" properties which one desires for statistical-mechanical models of realistic systems.[2]

It appears that the irrational polygon models are mixing (the rational models cannot be mixing on the full energy surface since they are not even ergodic). The intuitive argument is as follows: The essence of mixing is the splitting or stretching of an initial volume element on the energy surface, in such a way that the element gets distributed around the entire surface. Let us follow the development, in the reduced phase space X, of a small volume element $V \subset X$. For an irrational polygon model, the moving element $T^t V \subset X$ will eventually be split into two parts, when some point in $T^t V$ runs into a vertex. Looking at Fig. 2, it is difficult to see how these two parts could ever reunite for irrational models, since (after splitting) the two sequences $\{Q(i, x)\}_{i \in Z_+}$ and $\{Q(i, x')\}_{i \in Z_+}$ for the two parts appear to be disjoint. Thus, the splitting process will go on indefinitely, until $T^t V$ consists of small fragments distributed (hopefully in a uniform fashion) throughout X.

On the other hand, the rational models may be nonmixing even on the two-dimensional surface $X(\phi)$, even though these models appear to be ergodic on $X(\phi)$. This is certainly true for the two-body hard point gas with equal masses (i.e., the isosceles right triangle model) and for any other polygon model for which $\bar{A}$ is nonoverlapping: study of Fig. 2 shows that if $\bar{A}$ is nonoverlapping, a two-dimensional area element started on a single sheet of $X(\phi)$ will split up (onto two or more sheets) only briefly upon hitting a vertex, and then immediately reunite on a single sheet. It is possible that similar behavior occurs for all rational models, the only difference being that several wall-collision times are required before the area element reunites on a single sheet of $X(\phi)$.

The irrational polygon models do not appear to be C-systems,[2] i.e., the mixing process seems to occur with less-than-exponential rapidity. The intuitive argument is as follows: We again need to follow the progress $T^t V$ of a small element $V \subset X$. The moving element remains essentially undistorted (except for a slight twisting and stretching due to the fact that points at different ϕ move in different directions) until it hits a vertex. The time τ at which $T^t V$ hits a vertex should be inversely proportional to the mean diameter of the projection of $T^t V$ onto

Q, since if this mean diameter is (say) halved, then $T^t V$ must go through roughly twice as many wall collisions before finally hitting a vertex. Thus, τ is inversely proportional to the square root of the area of the projection of $T^t V$ onto Q. When $T^t V$ hits a vertex, it splits into two pieces $T^t V_1$ and $T^t V_2$ which (cf. Fig. 2) achieve their maximum separation (in X) from each other within one wall-collision time after the vertex collision. Thus, mixing appears to be due to vertex collisions alone, and is not due to stretching of single volume elements or increasing separation between pairs of volume elements between vertex collisions. If τ is the time for $T^t V$ to break into two pieces, and if (as we have argued) τ is inversely proportional to the square root of the projected area, then the time for these two pieces to break into four pieces is roughly $\sqrt{2}\tau$. Continuing this argument, we see that the time for $T^t V$ to make the transition from 2^n pieces to 2^{n+1} pieces is $2^{n/2}\tau$. For "exponential mixing" due to vertex collisions alone, this transition would occur in time τ; for "linear mixing" due to vertex collisions alone, this transition would occur in time $2^{n+1}\tau - 2^n\tau = 2^n\tau$. Thus, the mixing rate for polygon models appears to be much slower than exponential, but slightly faster than linear. This slow mixing behavior has been observed "experimentally" in the computer calculations of Casati and Ford,[6] who note that a surprisingly large number of collisions is required before a set of initially close phase trajectories becomes evenly mixed.

Aizenman[9] has proven that if a polygon system is ergodic, then there must exist a finite generating partition having zero entropy,[5] and hence the model has zero entropy[5] and cannot be a K-system; on the other hand, *non*-ergodic models cannot be K-systems, and so polygon models are certainly not K-systems. Aizenman's result shows that any ergodic polygon system is "deterministic" in the following sense: partition the energy surface into any *finite* set of "phase cells" labeled $j = 1, 2, \ldots, N$, and specify the cells $j(t)$ which the system occupied at the past times $t = -1, -2, \cdots$; if the system has zero entropy, its present and future cells $j(t)$ ($t = 0, 1, 2, \cdots$) are then completely determined. That is, the "coarse-grained past" of any zero-entropy system completely determines its "coarse-grained future." Since only the coarse-grained past and future are accessible to experimental observation, systems (e.g., any ergodic polygon model) having zero entropy are "experimentally deterministic." Aizenman's result also indicates that polygon systems are not C-systems, since C-systems are "generally" also K-systems.[5]

ACKNOWLEDGMENTS

In the course of this work I have benefitted from correspondence and conversation with many people. In particular, I would like to mention Professor Joseph Ford, Dr. Michael Aizenman, and Professor Joel Lebowitz; also Professor Itrel Monroe (who provided most of the proof of Theorem 3) of the University of Arkansas Mathematics Department, and Professor Michael Lieber of the University of Arkansas Physics Department.

APPENDIX

Configuration space for the two-body hard point gas is the isosceles right triangle $0 \leq q_1 \leq q_2 \leq L$. In these variables, the model is not a polygon model since the configuration point does not obey the equal-angles law upon bouncing from the hypotenuse. Define new canonical variables y_i, u_i ($i = 1, 2$) by $y_i = q_i\sqrt{m_i}$, $u_i = p_i/\sqrt{m_i}$. After a two-body collision, the initial and final momenta are related by

$$\begin{pmatrix} u_1' \\ u_2' \end{pmatrix} = \begin{pmatrix} \beta & (1-\beta^2)^{1/2} \\ (1-\beta^2)^{1/2} & \beta \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

where $\beta = (m_1 - m_2)/(m_1 + m_2)$. Examination of the eigenvectors of the collision matrix reveals that this transformation is a reflection in momentum space across the line which makes an angle $\tan^{-1}[(1-\beta)/(1+\beta)]^{1/2} = \theta$ with the u_1 axis, where θ is defined by (1). In the new variables, the velocity vector is parallel to the momentum vector. Thus, not only the momentum vector, but also the velocity vector, reflects across the direction ϕ upon collision, so the motion obeys the equal-angles law upon collision. It is easy to see that the model also obeys the equal-angles law when either particle bounces from a wall. Between bounces, the configuration point moves in a straight line with constant speed $(\dot{y}_1^2 + \dot{y}_2^2)^{1/2} = (u_1^2 + u_2^2)^{1/2} = (2E)^{1/2}$, where E is the total energy. Thus, we have proven that the two-body hard point gas is canonically transformable to a polygon system with configuration space being the right triangle $0 \leq y_1/\sqrt{m_1} \leq y_2/\sqrt{m_2} \leq L$ [i.e., the right triangle with interior angle ϕ given by (1)].

[1] Ja. G. Sinai, in *Statistical Mechanics* (Proc. I.U.P.A.P. Conf., Copenhagen, 1966), edited by T. A. Bak (Benjamin, New York, 1967); Russ. Math. Sur. **25**, 137 (1970).
[2] I. E. Farquhar, in *Irreversibility in the Many-Body Problem* (Sitges International School of Physics, May, 1972), edited by J. Biel and J. Rae (Plenum, New York, 1972).
[3] J. L. Lebowitz, in *Statistical Mechanics* (Proc. I.U.P.A.P. Conf., Chicago, 1971), edited by S. A. Rice, K. F. Freed, and J. C. Light (University of Chicago Press, Chicago, 1972).
[4] I. E. Farquhar, *Ergodic Theory in Statistical Mechanics* (Wiley, London, 1964).
[5] V. I. Arnold and A. Avez, *Ergodic Problems of Statistical Mechanics* (Benjamin, New York, 1968).
[6] G. Casati and J. Ford, J. Comp. Phys. (to appear).
[7] P. A. Mello, Ann. Phys. (N.Y.) **71**, 1 (1972).
[8] A. B. Katok and A. M. Stepin, Sov. Math. Dokl. **7**, 1638 (1966).
[9] M. Aizenman (private communication).

The 2:1 anisotropic oscillator, separation of variables and symmetry group in Bargmann space

Charles P. Boyer and Kurt Bernardo Wolf

Centro de Investigación en Matemáticas Aplicadas y en Sistemas (CIMAS), Universidad Nacional Autónoma de México, México D. F., Mexico
(Received 3 March 1975)

We present a detailed analysis of the separation of variables for the time-dependent Schrödinger equation for the anisotropic oscillator with a 2:1 frequency ratio. This reduces essentially to the time-independent one, where the known separability in Cartesian and parabolic coordinates applies. The eigenvalue problem in parabolic coordinates is a multiparameter one which is solved in a simple manner by transforming the system to Bargmann's Hilbert space. There, the degeneracy space appears as a subspace of homogeneous polynomials which admit unique representations of a solvable symmetry algebra s_3 in terms of first order operators. These representations, as well as their conjugate representations, are then integrated to indecomposable finite-dimensional nonunitary representations of the corresponding group S_3. It is then shown that the two separable coordinate systems correspond to precisely the two orbits of the factor algebra $s_3/u(1)$ [$u(1)$ generated by the Hamiltonian] under the adjoint action of the group. We derive some special function identitites for the new polynomials which occur in parabolic coordinates. The action of S_3 induces a nonlinear canocical transformation in phase space which leaves the Hamiltonian invariant. We discuss the differences with previous works which present $su(2)$ as the algebra responsible for the degeneracy of the two-dimensional anisotropic oscillator.

1. INTRODUCTION

In this paper we will examine the quantum two-dimensional anisotropic harmonic oscillator with a 2:1 frequency ratio. This system, though particular, is interesting in two respects: First, the time-independent problem is known to separate in two coordinate systems, Cartesian and parabolic and, second, its energy levels exhibit a degeneracy pattern which has been attributed to a symmetry algebra. Both features will be shown to be related through the treatment of the problem in Bargmann space.

Winternitz *et al.* have shown[1] that there is a one-to-one correspondence between second-order differential operators which commute with quantum Hamiltonians H of the standard type [i. e., $-\Delta + V(x_1, x_2)$], and separable coordinate systems for the Schrödinger equation, that is, there exist functions $v_1(x_1, x_2)$, $v_2(x_1, x_2)$ such that the time-independent Schrödinger equation separates into two differential equations, one in v_1 and one in v_2. Reduced to a canonical form, these v's can be made to correspond to one of the four orthogonal coordinate systems in two-space: Cartesian, polar, parabolic, or elliptic. The 2:1 anisotropic oscillator, in particular, was shown to separate in Cartesian and parabolic coordinates with the corresponding "separation" operators S_C and S_P commuting with H. Section 2 recapitulates these developments in the light of the general procedure of separation of variables[2,3] involving the time, and shows that the time-dependent problem can be reduced to a study of the time-independent one. The wavefunctions of the system in parabolic coordinates are not known special functions.

In Sec. 3 we show that the introduction of Bargmann space[4] provides a very convenient tool for finding the eigenfunctions and spectra of the pair S_P and H. The parabolic basis eigenfunctions are seen to be given in terms of new orthogonal polynomials whose coefficients are given by a three-term recursion relation. The polynomials and their eigenvalues are tabulated up to the $n = 15$ level. These coefficients also give the expansion of the parabolic basis eigenfunctions in terms of the Cartesian ones and vice versa as well as other special function identities.

In Sec. 4 we relate the polynomials and degeneracy pattern to the existence of a solvable symmetry group in Bargmann space. This group, or more accurately, its infinitesimal generators are found by looking for all first-order densely defined differential operators in Bargmann space which commute with the Hamiltonian. The representations we find are indecomposable nonunitary finite-dimensional representations of the symmetry group. We also discuss the relevance of the conjugate representations. Moreover, the orbit structure of the Lie algebra is analyzed and it is shown that the orbits relate to the two separable coordinate systems in the usual configuration space. We point out that this connection breaks down for any other rational frequency ratio.

The solvable symmetry group is a group of nonlinear canonical transformations of the coordinate-momentum space which are *geometrical* symmetries in Bargmann space. This is shown in Sec. 5. Finally, in Sec. 6 some conclusions are presented about the relations and differences between our and former work.[5-7] These question the necessity of unitary representations and of $su(2)$ in describing accidental degeneracy.

2. SEPARATION OF VARIABLES

Our first aim is to find all separable coordinate systems for the equation

$$U_{x_1x_1} + U_{x_2x_2} + iU_t - (4x_1^2 + x_2^2)U = 0 \tag{2.1}$$

where $U_z \equiv \partial U/\partial z$. The procedure that follows is quite analogous to Ref. 3 with only slight modifications due to the potential term. We will thus only give a rough sketch of the method used in deriving the result (2. 5).

We look for all coordinate systems described by the change of variables

$$x_1 = X_1(v_1, v_2, v_3), \quad x_2 = X_2(v_1, v_2, v_3), \quad t = T(v_1, v_2, v_3), \tag{2.2}$$

such that a solution of (2.1) is of the form

$$U(v_1, v_2, v_3) = \exp[iS(v_1, v_2, v_3)]V_1(v_1)V_2(v_2)V_3(v_3), \tag{2.3}$$

and (2.1) reduces to three ordinary differential equations. The function $S(v_1, v_2, v_3)$ is called a multiplier and can be determined by the analysis. Moreover, the separation process[3] always allows us to choose $t = T(v_1, v_2, v_3) = v_3$ in (2.2).

For the purpose of finding the separable coordinate systems of (2.1) it is useful to consider its symmetry group. The Lie algebra of this group was determined in Ref. 8, where the integrated group H_2 (the two-dimensional harmonic oscillator group) has the structure $H_2 \approx R_1 \circledS W_2$, where R_1 is the group of additive reals, W_2 is the five-dimensional Weyl group in two-space and $\circledS$ denotes the semidirect product. The group action given in terms of the one-parameter subgroups (R_1 is generated by R and W_2 by B_i, P_i, $i = 1, 2$, and E with $[B_j, P_k] = i\delta_{jk}E\omega_j$), $\omega_1 = 2$, $\omega_2 = 1$, is

$$\exp(i\tau R)f(x_1, x_2, t) = f(x_1, x_2, t + \tau), \tag{2.4a}$$

$$\begin{aligned}\exp(i\boldsymbol{\beta}\cdot\mathbf{P})f(x_1, x_2, t) = \exp[&-i(4\beta_1 x_1 \sin 4t + 2\beta_2 x_2 \sin 2t) \\ &+ \beta_1^2 \sin 8t + \tfrac{1}{2}\beta_2^2 \sin 4t] \\ &\times f(x_1 - \beta_1 \cos 4t, x_2 - \beta_2 \cos 2t, t),\end{aligned} \tag{2.4b}$$

$$\begin{aligned}&\exp(i\boldsymbol{\alpha}\cdot\mathbf{B})f(x_1, x_2, t) \\ &\quad = \exp[i(4\alpha_1 x_1 \sin 4t + 2\alpha_2 x_2 \sin 2t) + \alpha_1^2 \sin 8t \\ &\qquad + \tfrac{1}{2}\alpha_2^2 \sin 4t]f(x_1 - \alpha_1 \sin 4t, x_2 - \alpha_2 \sin 2t, t),\end{aligned} \tag{2.4c}$$

$$\exp(\gamma E)f(x_1, x_2, t) = \exp(\gamma)f(x_1, x_2, t), \tag{2.4d}$$

where $f \in C^\infty$, $\tau, \alpha_1, \ldots, \gamma \in \mathbb{R}$. Now, by a straightforward calculation following the procedure of Ref. 3, it can be shown that the only separable coordinates with a nontrivial multiplier S (i.e., *not* a sum of functions of the individual variables) are those given precisely by the change of variables induced by the transformations of the symmetry group (2.4). Indeed, such transformations give rise to separable solutions $V_1(v_1)$, $V_2(v_2)$ which satisfy the same ordinary differential equations with the usual separation in t, i.e., $T(v_3) = c \exp iEt$. Therefore, two separable coordinate systems which differ by a transformation of the type (2.4) are said to be equivalent. Hence, our problem reduces to the separation of the time-independent Schrödinger equation[1] and we find only *two* inequivalent separable coordinates:

(i) Cartesian

$$x_1 = v_1, \quad x_2 = v_2, \quad t = v_3, \quad x_1, x_2 \in \mathbb{R}; \tag{2.5a}$$

(ii) parabolic

$$x_1 = \tfrac{1}{2}(v_1^2 - v_2^2), \quad x_2 = v_1 v_2, \quad t = v_3. \tag{2.5b}$$

$v_1 \in \mathbb{R}$, $v_2 \in \mathbb{R}^+$; thus in what follows we consider the time-independent Schrödinger equation, viz.

$$H\psi = -\psi_{x_1x_1} - \psi_{x_2x_2} + (4x_1^2 + x_2^2)\psi = E\psi, \tag{2.6a}$$

obtained from (2.1) through

$$U(x_1, x_2, t) = \exp(iEt)\psi(x_1, x_2). \tag{2.6b}$$

Winternitz and collaborators[1] have characterized the two separable systems (2.5) of (2.6) by the two second-order symmetry operators

$$S_C = -\partial_{x_1x_1} + 4x_1^2, \tag{2.7a}$$

$$S_P = x_1\partial_{x_2x_2} - x_2\partial_{x_1x_2} - \tfrac{1}{2}\partial_{x_1} + x_1x_2^2, \tag{2.7b}$$

corresponding to (i) and (ii) above, respectively. Indeed, it can be shown that (2.7) are the only second-order operators which commute with the Hamiltonian H. Now, the solutions of (2.1) in the Cartesian coordinate system (2.5a) are characterized by the equations

$$H\psi^C = E\psi^C, \tag{2.8a}$$

$$S_C\psi^C = \mu\psi^C, \tag{2.8b}$$

which give rise to the well-known eigenvalue problem for the one-dimensional harmonic oscillator with eigenfunctions normalized in the usual Hilbert space norm $L^2(\mathbb{R}^2)$ given by

$$\begin{aligned}&\psi^C_{n_1n_2}(x_1, x_2) \\ &\quad = [2^{n_1+n_2-1/2}\pi n_1!\, n_2!\,]^{-1/2}\exp[-x^2 - \tfrac{1}{2}y^2]H_{n_1}(\sqrt{2}x_1)H_{n_2}(x_2),\end{aligned} \tag{2.9a}$$

with eigenvalues

$$E = 4n_1 + 2n_2 + 3 \equiv 2n + 3, \quad n_1, n_2, n = 0, 1, 2, \cdots, \quad \mu = 4n_1 + 2. \tag{2.9b}$$

Notice that the energy level labeled by n has degeneracy $[n/2] + 1$, where $[r]$ is the integer part of r.

The solution of (2.1) in parabolic coordinates (2.5b) are

$$H\psi^P = E\psi^P, \tag{2.10a}$$

$$S_P\psi^P = \lambda\psi^P. \tag{2.10b}$$

These equations give rise to $L^2(\mathbb{R}^2)$ solutions $\psi^P_{nl}(x_1, x_2)$ which are products of the form

$$\psi^P_{nl}(x_1, x_2) = \phi_{nl}(v_1)\phi_{nl}(iv_2), \tag{2.11}$$

where $\phi(v)$ is a solution of the equation

$$\phi_{vv} + (2\lambda + Ev^2 - v^6)\phi = 0. \tag{2.12}$$

We note that since the measure in parabolic coordinates is

$$d^2\mathbf{x} = dx_1\, dx_2 = (v_1^2 + v_2^2)\, dv_1\, dv_2, \tag{2.13}$$

and (2.12) depends on both λ and E, the eigenvalue problem is a (coupled) multiparameter one. However, we know the value of E from the Cartesian separation and we can use (2.10b) to derive a recursion relation for the overlap functions between the two bases. Then, since the degeneracy for each n is $[n/2] + 1$, we look for the recursion relation to be cut off. Rather than implement this procedure here, in the next section we will analyze the system in Bargmann's Hilbert space where our problem reduces to a single Sturm–Liouville problem and the degeneracy of states appears simply as a

subspace of homogeneous polynomials. It is further noticed that (2.12) is the equation for an anharmonic oscillator with a $-\omega^2 v^2 + v^6$ potential, and can be related to the confluent Heun equation. [9]

Equations (2.19) also exhibit an interesting discrete symmetry: It is easily seen from (2.6a) and (2.7b) that under the parity transformation $x_1 \to -x_1$, $H \to H$ while $S_P \to -S_P$; hence if $\psi^P_{nl}(x_1, x_2)$ is a solution of the eigenvalue problem (2.10) with eigenvalues E and λ then $\psi^P_{nl}(-x_1, x_2) = \phi_{nl}(v_2)\phi_{nl}(iv_1)$ is also a solution with eigenvalues E and $-\lambda$. In addition, Eqs. (2.10) are invariant under $x_2 \to -x_2$ ($v_1 \to -v_1$) and the parity properties of the Cartesian basis (2.9a) are well known.

Note: Our Eq. (2.12) has been recently treated in an interesting work by Truong through the use of harmonic analysis on the Weyl group. [See T. T. Truong, a Weyl quantization of anharmonic oscillators, J. Math, Phys. **16**, 1034 (1975).]

3. SOLUTION IN BARGMANN SPACE

In this section we shall show that the treatment of the anisotropic oscillator in Bargmann's Hilbert space of analytic functions[4] allows a simple interpretation of the degeneracy pattern as well as a reduction for the parabolic coordinates to a simple Sturm—Liouville problem. [10] For an oscillator of frequency ω we define[11] out of the canonically conjugate operators $\hat{x}$ and $\hat{p}$ (with $[\hat{x}, \hat{p}] = i\mathbb{1}$),

$$\hat{x} = (2\omega)^{-1/2}(\hat{\eta} + i\hat{\xi}) \tag{3.1a}$$

$$\hat{p} = (2/\omega)^{-1/2}(i\hat{\eta} + \hat{\xi}), \tag{3.1b}$$

so that $\hat{\eta}$ and $\hat{\xi}$ also constitute a canonical pair ($[\hat{\eta}, \hat{\xi}] = i\mathbb{1}$). Under (3.1), the Hamiltonian becomes

$$H_\omega = \hat{p}^2 + \omega^2\hat{x}^2 = 2\omega(i\hat{\eta}\hat{\xi} + \tfrac{1}{2}). \tag{3.1c}$$

Upon introducing a scalar product over the complex plane $\mathbb{C}$

$$(f, g)_\omega = \omega^{1/2}\pi^{-1}\int_{\mathbb{C}} d^2\mu_\omega(\eta) f(\eta)^* g(\eta), \tag{3.2a}$$

where

$$d^2\mu_\omega(\eta) = \exp[-\omega|\eta|^2]d^2\eta, \quad d^2\eta \equiv d\,\mathrm{Re}\eta\, d\,\mathrm{Im}\eta, \tag{3.2b}$$

f and g are analytic functions in η over $\mathbb{C}$ of growth $(2, \omega/2)$, and completing with respect to the norm induced by (3.2) we obtain the Bargmann space $\mathcal{J}_\omega$. Bargmann has shown that the operators given by (3.1a, b, c) are self-adjoint in $\mathcal{J}_\omega$ defined with the domains

$$\mathcal{D}(\mathcal{O}) = \{f \in \mathcal{J}_\omega : \mathcal{O}f \in \mathcal{J}_\omega\}, \tag{3.3a}$$

where $\mathcal{O}$ is one of the operators (3.1). Thus in $\mathcal{J}_\omega$ we have the representation

$$\hat{\eta} f(\eta) = \eta f(\eta), \quad \hat{\xi} f(\eta) = -i\partial f(\eta)/\partial\eta, \tag{3.3b}$$

with the Hilbert space adjoint

$$\hat{\eta}^\dagger = i\hat{\xi}/\omega, \quad \hat{\xi}^\dagger = i\omega\hat{\eta}. \tag{3.3c}$$

The unitary mapping between $\mathcal{L}^2(\mathbb{R})$ and $\mathcal{J}_\omega$ is given by

$$\tilde{f}(\eta) = (\mathbf{A}_\omega f)(\eta) = \int_{\mathbb{R}} dx A(\eta, x) f(x) \tag{3.4a}$$

with the inverse

$$f(x) = (\mathbf{A}_\omega^{-1}\tilde{f})(x) = \int_{\mathbb{C}} d^2\mu(\eta) A(\eta, x)^* \tilde{f}(\eta) \tag{3.4b}$$

where

$$A_\omega(\eta, x) = \omega^{1/2}\pi^{-1/4}\exp[\omega(-\tfrac{1}{2}x^2 + \sqrt{2}x\eta - \tfrac{1}{2}\eta^2)], \tag{3.4c}$$

and $\tilde{f} \in \mathcal{J}_\omega$, $f \in \mathcal{L}^2(\mathbb{R})$, and the integrals are understood to be in the sense of limit in the mean.

We can now build the space $\mathcal{J} \equiv \mathcal{J}(\mathbb{C}^2)$ with the measure $d\mu_1^2 d\mu_2^2$ and the two-dimensional Hamiltonian which is the image under the unitary mapping $\mathbf{A} \equiv \mathbf{A}_2 \otimes \mathbf{A}_1$ of the Hamiltonian (2.6a) and the sum of two Hamiltonians (3.1c) with $\omega_1 = 2$, $\omega_2 = 1$. Hence in $\mathcal{J}$ our Hamiltonian is

$$\tilde{H} = 4\eta_1\partial_{\eta_1} + 2\eta_2\partial_{\eta_2} + 3. \tag{3.5}$$

Now the simple form of (3.5) allows us to immediately solve the two-dimensional eigenvalue problem

$$\tilde{H}\tilde{\psi}_n = E_n\tilde{\psi}_n, \quad \tilde{\psi}_n = \tilde{\psi}_n(\eta_1, \eta_2). \tag{3.6}$$

We find by the method of characteristics and the fact that $\tilde{H}\tilde{\psi}_n \in \mathcal{J}$ the general solution with (2.9b),

$$\tilde{\psi}_n(\eta_1, \eta_2) = \eta_2^n P_n(\eta_1/\eta_2^2), \tag{3.7}$$

where P_n is a polynomial of degree $[n/2]$. Hence the degeneracy of states in $\mathcal{J}$ makes its appearance by the simple fact that the solutions of (3.6) are homogeneous polynomials P_n of degree $[n/2]$. This polynomial subspace $\mathcal{P}_n$ maps under the Bargmann transform $\mathbf{A}$ given by (3.4) onto all $\mathcal{L}^2(\mathbb{R}^2)$ solutions of the Schrödinger equation (2.6a) with fixed energy E_n. In the next section we will find a group of transformations [*not* $SU(2)$] which maps the polynomial subspace $\mathcal{P}_n \subset \mathcal{J}$ defined by (3.7) into itself. We also emphasize here that the above analysis is quite general and applies to any anisotropic oscillator whose ratio of frequencies is rational, although there will be no connection with separable coordinate systems.

In the Cartesian basis described by the self-adjoint[4] operator

$$\tilde{S}_C = \mathbf{A}S_C\mathbf{A}^{-1} = 4\eta_1\partial_{\eta_1} + 2, \tag{3.8a}$$

along with (3.6) we obtain the orthonormal eigenfunctions

$$\tilde{\psi}^C_{n_1 n_2}(\eta_1, \eta_2) = (n_1!\,n_2!)^{-1/2}2^{n_1/2}\eta_1^{n_1}\eta_2^{n_2}. \tag{3.8b}$$

with E given by (2.9b). Note since $n = 2n_1 + n_2$, it follows that $\tilde{\psi}^C_{n_1 n_2} \in \mathcal{P}_n$. Moreover, under (3.4b) the eigenfunctions (3.8b) map onto the usual harmonic oscillator eigenfunctions (2.9a).

For the parabolic coordinates we find the operator

$$\tilde{S}_P = \mathbf{A}S_P\mathbf{A}^{-1} = \sqrt{2}\eta_1\partial_{\eta_2\eta_2} + \eta_2^2\partial_{\eta_1}/\sqrt{2}. \tag{3.9}$$

which is self-adjoint on the domain $\mathcal{D}(\hat{\eta}_1 \times \hat{\eta}_2^2)$ with the $\mathcal{D}$'s given by (3.3). From the operator S_P in (2.6b) one expects in general upon inserting (3.1) that $\tilde{S}_P$ be a third order operator in $\mathcal{J}$. It is a pleasant feature of the mapping that the third order terms cancel. The eigenvalue problem for (3.9) on $\mathcal{P}_n$, upon introducing in (3.7) the change of variables

$$z = \eta_2, \quad u = \eta_1/\eta_2^2, \tag{3.10a}$$

yields the differential equation

$$4\sqrt{2}u^3 P''_{nl}(u) + [-2\sqrt{2}(2n-3)u^2 + 1/\sqrt{2}\,]P'_{nl}(u)$$

$$+[\sqrt{2}n(n-1)u-\lambda_l]P_{nl}(u)=0, \tag{3.10b}$$

where we have labelled the eigenvalue λ of $\tilde{S}_P$ through the index l in a fashion which will be described below. Expanding the polynomials $P_{nl}(u)$ as

$$P_{nl}(u)=\sum_{m=0}^{[n/2]} p_m^{nl}u^m, \tag{3.11}$$

we find three-term recursion relation for the coefficients

$$\frac{1}{\sqrt{2}}(m+1)p_{m+1}^{nl}-\lambda_l p_m^{nl}+\sqrt{2}(2m-n-1)(2m-n-2)p_{m-1}^{nl}=0. \tag{3.12}$$

We remark that the coefficients p_m^{nl} have been chosen to be real and such that $p_0^{nl}>0$. Equation (3.12) allows us to solve the eigenvalue problem for λ_l when we require that $p_{[n/2]+1}^{nl}=0$. This problem is equivalent to diagonalizing a square tridiagonal matrix of dimension $[n/2]+1$. The resulting eigenvalues λ_l can be labelled by the index l running from $-\frac{1}{2}([n/2]+1)$ to $\frac{1}{2}([n/2]+1)$ in integer steps and such that $\lambda_{l_1}<\lambda_{l_2}$ iff $l_1<l_2$. The motivation for such a labeling stems from the parity properties discussed at the end of the last section. Clearly the inversion $x_1\to -x_1$ implies $\eta_1\to-\eta_1$, or equivalently $u\to -u$; and again if under $\tilde{S}_P$, λ_l is the eigenvalue of $\tilde{\psi}_{nl}^P(\eta_1,\eta_2)$, then $-\lambda_l$ is the eigenvalue of $\tilde{\psi}_{nl}^P(-\eta_1,\eta_2)$; and if $P_{nl}(u)$ satisfied (3.10b) with λ_l, $P_{nl}(-u)$ will satisfy the same equation with $-\lambda_l$. Our labeling convention for λ_l then implies that $-\lambda_l=\lambda_{-l}$ and $P_{n,-l}(u)=P_{nl}(-u)$. The eigenvalues λ_l appear thus in symmetrical pairs. When $[n/2]+1$ is even, the l's are half-intergers, while when $[n/2]+1$ is odd, the l's are integers and $\lambda_0=0$ is among the eigenvalues. We point out that although the label l resembles a "magnetic" quantum number suggesting an $su(2)$ symmetry algebra for the system, no such construction has been made.

The eigenvalues λ_l and the properly normalized coefficients p_m^{nl} for (3.11), (3.12) for the first 15 values of n have been computed and collected in Table I. We will refer to $P_{nl}(u)$ as *parabolic* polynomials. The entries of this table also give us the needed information about the expansion of the parabolic coordinate solutions in terms of the Cartesian basis, since from (3.7), (3.8b), and (3.11)

$$\tilde{\psi}_{nl}^P(\eta_1,\eta_2)=\sum_{m=0}^{[n/2]} p_m^{nl}\eta_1^m\eta_2^{n-2m}=\sum_{m=0}^{[n/2]}[2^{-m}m!(n-2m)!]^{1/2}p_m^{nl}\tilde{\psi}_{m,n-2m}^C(\eta_1,\eta_2). \tag{3.13}$$

Choosing $\tilde{\psi}_{nl}^P(\eta_1,\eta_2)$ to be normalized in $\mathcal{J}$, we find

$$\sum_{m=0}^{[n/2]}[2^{-m}m!(n-2m)!]p_m^{nl}p_m^{nl'}=\delta_{l,l'}. \tag{3.14}$$

The expansion inverse to (3.13) is easily obtained and reads

$$\tilde{\psi}_{m,n-2m}^C(\eta_1,\eta_2)=[2^{-m}m!(n-2m)!]^{1/2}\sum_{l=-[n/2]-1}^{[n/2]+1} p_m^{nl}\tilde{\psi}_{nl}^P(\eta_1,\eta_2) \tag{3.15}$$

Again from the orthonormality properties we obtain

$$2^{-m}m!(n-2m)!\sum_{l=-[n/2]-1}^{[n/2]+1} p_{m'}^{nl}p_m^{nl}=\delta_{m,m'}. \tag{3.16}$$

The solutions (3.10) of (3.11) are instrumental for the solutions (2.11), $\phi_{nl}(v)$, of (2.12) in the following manner: Transforming (3.10) to its standard form, we find the latter to be identical with (2.12) so that its solutions are

$$\phi_{nl}(v)\sim v^n\exp(-v^4/4)P_{nl}([2\sqrt{2}v^2]^{-1}), \tag{3.17}$$

upon demanding the correct asymptotic properties for $\tilde{\psi}_{nl}^P$. It is emphasized that we have constructed polynomial solutions of the differential equation (2.12). The advantage of the Bargmann space treatment is now manifest: Through the unitary transform (3.4) we have reduced the coupled multiparameter eigenvalue problem (2.10) to the single Sturm—Liouville problem (3.10) whence upon transforming back and using (3.17), we have

$$\psi_{nl}^P(v_1,v_2)=c_{nl}(v_1v_2)^n\exp[-\tfrac{1}{4}(v_1^4+v_2^4)]P_{nl}([2\sqrt{2}v_1^2]^{-1})\times P_{nl}(-[2\sqrt{2}v_2^2]^{-1}), \tag{3.18}$$

where c_{nl} is the normalization coefficient with respect to the measure (2.14) given by

$$c_{nl}=(p_{[n/2]}^{nl})^{-2}\pi^{-1/2}2^{n+\sigma(n)/4}\sum_{k=0}^{[n/4]} p_{2k}^{nl}(-1)^k\frac{(2k)!(n-4k)!}{k!([n/2]-2k)!} \tag{3.19}$$

where $\sigma(n)\equiv(-1)^n$ and can be calculated with the use of the Table I.

Writing the transform (3.4a) explicitly, we find the integral identity

$$\int_{-\infty}^{\infty}dv_1\int_0^{\infty}dv_2(v_1^2+v_2^2)(v_1v_2)^n\exp[-\tfrac{1}{2}(v_1^4+v_2^4)+\sqrt{2}(v_1^2-v_2^2)\eta_1+\sqrt{2}v_1v_2\eta_2]P_{nl}([2\sqrt{2}v_1^2]^{-1})P_{nl}(-[2\sqrt{2}v_2^2]^{-1})=[c_{nl}\sqrt{2/\pi}\,]^{-1}\exp[\eta_1^2+\tfrac{1}{2}\eta_2^2]\eta_2^nP_{nl}(\eta_1/\eta_2^2). \tag{3.20a}$$

Equation (3.4b) yields

$$\int\int_{\mathbb{C}}d^2\eta_1\,d^2\eta_2\,\eta_2^nP_{nl}(\eta_1/\eta_2^2)\exp[-2|\eta_1|^2-|\eta_2|^2-\eta_1^{*2}-\tfrac{1}{2}\eta_2^{*2}+\sqrt{2}(v_1^2-v_2^2)\eta_1^*+\sqrt{2}\,v_1v_2\eta_2^*]=\tfrac{1}{2}\pi^{3/2}c_{nl}(v_1v_2)^nP_{nl}([2\sqrt{2}v_1^2]^{-1})P_{nl}(-[2\sqrt{2}v_2^2]^{-1}). \tag{3.20b}$$

Moreover, applying the unitary transform (3.4) to the expansions (3.13) and (3.15), we find the expansion formulae

$$\sum_{m=0}^{[n/2]} p_m^{nl}H_m(2^{-1/2}[v_1^2-v_2^2])H_{n-2m}(v_1v_2)=\pi^{1/2}2^{n/2-1/4}c_{nl}(v_1v_2)^nP_{nl}([2\sqrt{2}v_1^2]^{-1})P_{nl}(-[2\sqrt{2}v_2^2]^{-1}) \tag{3.21}$$

and

$$(v_1v_2)^n\sum_{l=-[n/2]-1}^{[n/2]+1} p_m^{nl}c_{nl}P_{nl}([2\sqrt{2}v_1^2]^{-1})P_{nl}(-[2\sqrt{2}v_2^2]^{-1})=[\pi^{1/2}2^{n/2-m-1/4}m!(n-2m)!]^{-1}H_m(2^{-1/2}[v_1^2-v_2^2])H_{n-2m}(v_1v_2). \tag{3.22}$$

These formulas allowed us to calculate c_{nl} in (3.19) by evaluating (3.21) for even n at $\mathbf{x}=0$ and for odd n,

TABLE I. Eigenvalues and eigenvectors for parabolic polynomials. The eigenvectors $p_m^{n,\pm l}$ for fixed n and l are listed from top to bottom as m runs from 0 to $[n/2+1]$, respectively.

Level		Eigenvalue $\lambda_{\pm l}$	Eigenvectors $p_m^{n,\pm l}$
$n=2$	$l=1/2$	± 1.414214	5.000000×10^{-1}
			± 1.000000
$n=3$	$l=1/2$	± 2.449490	2.886751×10^{-1}
			± 1.000000
$n=4$	$l=1$	± 4.000000	1.250000×10^{-1}
			$\pm 7.071068\times10^{-1}$
			5.000000×10^{-1}
	$l=0$	0.000000	1.020621×10^{-1}
			0.000000×10^{-4}
			-1.224745
$n=5$	$l=1$	± 5.656854	5.103104×10^{-2}
			$\pm 4.082483\times10^{-1}$
			6.123724×10^{-1}
	$l=0$	0.000000	5.590170×10^{-1}
			0.000000×10^{-4}
			-1.118034
$n=6$	$l=3/2$	± 7.538754	1.863390×10^{-2}
			$\pm 1.986636\times10^{-1}$
			5.000000×10^{-1}
			$\pm 1.875925\times10^{-1}$
	$l=1/2$	± 1.779658	1.863390×10^{-2}
			$\pm 4.689812\times10^{-2}$
			-5.000000×10^{-1}
			$\pm 7.946545\times10^{-1}$
$n=7$	$l=3/2$	± 9.579208	6.332529×10^{-3}
			$\pm 8.578707\times10^{-2}$
			3.151144×10^{-1}
			$\pm 2.791290\times10^{-1}$
	$l=1/2$	± 2.870326	7.688005×10^{-3}
			$\pm 3.120757\times10^{-2}$
			-2.595565×10^{-1}
			$\pm 7.673029\times10^{-1}$
$n=8$	$l=2$	± 11.78082	2.008449×10^{-3}
			$\pm 3.346196\times10^{-2}$
			1.662750×10^{-1}
			$\pm 2.541743\times10^{-1}$
			6.102406×10^{-2}
	$l=1$	± 4.605675	2.518940×10^{-3}
			$\pm 1.640688\times10^{-2}$
			-8.762825×10^{-2}
			$\pm 5.183905\times10^{-1}$
			-3.183528×10^{-1}
	$l=0$	0.000000	2.010905×10^{-3}
			0.000000×10^{-4}
			-1.126107×10^{-1}
			0.000000×10^{-4}
			6.756639×10^{-1}
$n=9$	$l=2$	± 14.12795	6.007569×10^{-4}
			$\pm 1.200309\times10^{-2}$
			7.665598×10^{-2}
			$\pm 1.744409\times10^{-1}$
			-3.791902×10^{-1}
	$l=1$	± 6.356178	8.200552×10^{-4}
			$\pm 7.371471\times10^{-3}$
			-2.591292×10^{-2}
			$\pm 2.840449\times10^{-1}$
			-3.791902×10^{-1}
	$l=0$	0.000000	8.300199×10^{-4}
			0.000000×10^{-3}
			-5.976143×10^{-2}
			0.000000×10^{-4}
			5.976143×10^{-1}
$n=10$	$l=5/2$	± 16.61347	1.704169×10^{-4}
			$\pm 4.003944\times10^{-3}$
			3.169879×10^{-2}
			$\pm 9.877367\times10^{-2}$
			1.046898×10^{-1}
			$\pm 1.782333\times10^{-2}$
	$l=3/2$	± 8.353982	2.447240×10^{-4}
			$\pm 2.891216\times10^{-3}$
			-4.946480×10^{-3}
			$\pm 1.274183\times10^{-1}$
			-3.021387×10^{-1}
			$\pm 1.022969\times10^{-1}$
	$l=1/2$	± 2.050659	2.210312×10^{-4}
			$\pm 6.410059\times10^{-4}$
			-1.896333×10^{-2}
			$\pm 4.226254\times10^{-2}$
			2.538089×10^{-1}
			$\pm 3.500729\times10^{-1}$
$n=11$	$l=5/2$	± 19.22960	4.608940×10^{-5}
			$\pm 1.253390\times10^{-3}$
			1.197299×10^{-2}
			$\pm 4.837152\times10^{-2}$
			7.743018×10^{-2}
			$\pm 3.416695\times10^{-2}$
	$l=3/2$	± 10.48809	6.979429×10^{-5}
			$\pm 1.035217\times10^{-3}$
			-0.000000×10^{-5}
			$\pm 4.969040\times10^{-2}$
			-1.842569×10^{-1}
			$\pm 1.490712\times10^{-1}$
	$l=1/2$	± 3.197242	7.436786×10^{-5}
			$\pm 3.362605\times10^{-4}$
			-7.420250×10^{-3}
			$\pm 2.732426\times10^{-2}$
			1.249380×10^{-1}
			$\pm 3.315776\times10^{-1}$
$n=12$	$l=3$	± 21.97026	1.193308×10^{-5}
			$\pm 3.707685\times10^{-4}$
			4.184841×10^{-3}
			$\pm 2.109578\times10^{-2}$
			4.668927×10^{-2}
			$\pm 3.698379\times10^{-2}$
			4.761253×10^{-2}
	$l=2$	± 12.78129	1.890156×10^{-5}
			$\pm 3.416546\times10^{-4}$
			5.927794×10^{-4}
			$\pm 1.692769\times10^{-2}$
			-9.309182×10^{-2}
			$\pm 1.334035\times10^{-1}$
			-2.952144×10^{-2}
	$l=1$	± 5.093752	2.029600×10^{-5}
			$\pm 1.462053\times10^{-4}$
			-2.152465×10^{-3}
			$\pm 1.394086\times10^{-2}$
			3.416276×10^{-2}
			$\pm 2.179504\times10^{-1}$
			1.210221×10^{-1}
	$l=0$	0.000000	1.626302×10^{-5}
			0.000000×10^{-5}
			-2.146719×10^{-3}
			0.000000×10^{-5}
			-6.010814×10^{-2}
			0.000000×10^{-5}
			-2.404325×10^{-1}

(continued)

TABLE I. (*continued*)

Level		Eigenvalue	Eigenvectors $p_m^{n,\pm l}$
$n=13$	$l=3$	± 24.82997	2.968235×10^{-6}
			$\pm 1.042292\times10^{-4}$
			1.366953×10^{-3}
			$\pm 8.356661\times10^{-3}$
			2.415046×10^{-2}
			$\pm 2.921621\times10^{-2}$
			9.984215×10^{-3}
	$l=2$	± 15.21229	4.902887×10^{-6}
			$\pm 1.054779\times10^{-4}$
			3.697447×10^{-4}
			$\pm 5.083554\times10^{-3}$
			-4.065197×10^{-2}
			$\pm 8.950889\times10^{-2}$
			-4.992728×10^{-2}
	$l=1$	± 6.932454	5.644911×10^{-6}
			$\pm 5.534254\times10^{-5}$
			-6.093177×10^{-4}
			$\pm 6.049697\times10^{-3}$
			7.107675×10^{-3}
			$\pm 1.155716\times10^{-1}$
			1.414589×10^{-1}
	$l=0$	0.000000	5.582380×10^{-6}
			0.000000×10^{-6}
			-8.708512×10^{-4}
			0.000000×10^{-5}
			3.135064×10^{-2}
			0.000000×10^{-4}
			-2.090043×10^{-1}
$n=14$	$l=7/2$	± 27.80398	7.114271×10^{-7}
			$\pm 2.797386\times10^{-5}$
			4.204971×10^{-4}
			$\pm 3.049722\times10^{-3}$
			1.105699×10^{-2}
			$\pm 1.864007\times10^{-2}$
			1.158708×10^{-2}
			$\pm 1.178723\times10^{-3}$
	$l=5/2$	± 17.77729	1.220946×10^{-6}
			$\pm 3.069567\times10^{-5}$
			1.636458×10^{-4}
			$\pm 1.329819\times10^{-3}$
			-1.572227×10^{-2}
			$\pm 4.926638\times10^{-2}$
			-4.921069×10^{-2}
			$\pm 7.829590\times10^{-3}$
	$l=3/2$	± 9.040977	1.455154×10^{-6}
			$\pm 1.860541\times10^{-5}$
			-1.458948×10^{-4}
			$\pm 2.259073\times10^{-3}$
			-6.557908×10^{-4}
			$\pm 4.892627\times10^{-2}$
			1.108188×10^{-1}
			$\pm 3.466912\times10^{-2}$
	$l=1/2$	± 2.273209	1.273209×10^{-6}
			$\pm 4.093117\times10^{-6}$
			-2.251448×10^{-4}
			$\pm 6.014597\times10^{-4}$
			9.648122×10^{-3}
			$\pm 1.967606\times10^{-2}$
			-8.593878×10^{-2}
			$\pm 1.069288\times10^{-4}$
$n=15$	$l=7/2$	± 30.88805	1.647271×10^{-7}
			$\pm 7.195660\times10^{-6}$
			1.225688×10^{-4}
			$\pm 1.036347\times10^{-3}$
			4.576218×10^{-3}
			$\pm 1.013316\times10^{-2}$
			9.706253×10^{-3}
			$\pm 2.666413\times10^{-3}$
	$l=5/2$	± 20.46843	2.929299×10^{-7}
			$\pm 8.479361\times10^{-6}$
			6.120959×10^{-5}
			$\pm 2.912478\times10^{-4}$
			-5.474195×10^{-3}
			$\pm 2.330407\times10^{-2}$
			-3.579068×10^{-2}
			$\pm 1.483719\times10^{-2}$
	$l=3/2$	± 11.26606	3.638824×10^{-7}
			$\pm 5.797598\times10^{-6}$
			-3.022984×10^{-5}
			$\pm 7.634970\times10^{-4}$
			-1.378486×10^{-3}
			$\pm 1.759614\times10^{-2}$
			6.602422×10^{-2}
			$\pm 4.972759\times10^{-2}$
	$l=1/2$	± 3.470994	3.701417×10^{-7}
			$\pm 1.816924\times10^{-6}$
			-7.327036×10^{-5}
			$\pm 3.088482\times10^{-4}$
			3.650857×10^{-3}
			$\pm 1.247904\times10^{-2}$
			-4.090263×10^{-2}
			$\pm 9.999164\times10^{-2}$

$\partial^2/\partial v_1 \partial v_2$ of both sides at $\mathbf{x}=0$. We add that from (3.21) and (3.22) many $L^2(\mathbb{R}^2)$ expansions can be derived for the parabolic polynomials P_{nl}. To conclude this section we give explicitly the parabolic polynomials for the first few n values:

$$n=0: \quad P_{00}(u)=1, \tag{3.23a}$$

$$n=1: \quad P_{10}(u)=1, \tag{3.23b}$$

$$n=2: \quad P_{2,\pm1/2}(u)=\pm u+\tfrac{1}{2}, \tag{3.23c}$$

$$n=3: \quad P_{3,\pm1/2}(u)=\pm u+1/2\sqrt{3}, \tag{3.23d}$$

$$n=4: \quad P_{4,\pm1}(u)=\frac{1}{\sqrt{106}}(10u^2\pm4\sqrt{2}u+1),$$

$$P_{4,0}=\frac{\sqrt{6}}{2}(-u^2+\tfrac{1}{12}), \tag{3.23e}$$

$$n=5: \quad P_{5,\pm1}(u)=\frac{1}{\sqrt{394}}(12u^2\pm8u+1),$$

$$P_{5,0}=\frac{\sqrt{5}}{2}(-u^2+\tfrac{1}{20}). \tag{3.23f}$$

When written in terms of the variables z,u as (3.10a) we can perform the z integral and obtain orthogonality for n as well as a weight function in u which depends on n. This weight function appears in terms of parabolic cylinder functions.

4. A SYMMETRY GROUP IN BARGMANN SPACE

Here we will show how the information of the previous section can be obtained by studying the group of geometrical symmetry transformations in Bargmann space. We look for all *first order* differential operators of the form

$$A=\sum_{i=1}^{2} a_i(\eta_1,\eta_2)\partial_{\eta_i}+b(\eta_1,\eta_2) \tag{4.1}$$

which commute with the Hamiltonian (3.5), i.e., $[A,\tilde{H}]=0$. If we further demand that all our symmetry operators A be densely defined in $\mathcal{J}$ then the functions a_i and b must be analytic functions of η_1 and η_2, we find

$$A_1=\eta_1\partial_{\eta_1},\quad A_2=\eta_2\partial_{\eta_2},\quad A_3=\eta_2^2\partial_{\eta_1},\quad A_0=\mathbf{1}, \tag{4.2}$$

with domains $\mathcal{D}(A_1)$ and $\mathcal{D}(A_2)$ given by (3.3) and $\mathcal{D}(A_3)=\mathcal{D}(\eta_1\times\eta_2^2)$. When we disregard the central element A_0 we find the three-dimensional solvable algebra[12] s_3 with Lie brackets

$$[A_1,A_2]=0,\quad [A_1,A_3]=-A_3,\quad [A_2,A_3]=2A_3. \tag{4.3}$$

It is easily seen that

$$\tilde{H}=4A_1+2A_2+3A_0, \tag{4.4}$$

and so we have the structure $s_3\approx u(1)\oplus s_2$, where s_2 is the two-dimensional algebra spanned by, say, A_1 and A_3, and $u(1)$ is spanned by $\tilde{H}$. Now on the space $\mathcal{P}_n$ of homogeneous polynomials the representation (4.2) of s_3 acts on the normalized Cartesian basis in Bargmann space, calling $\tilde{\phi}_m^n\equiv\tilde{\psi}^c_{m,n-2m}$, as

$$A_1\tilde{\phi}_m^n=m\tilde{\phi}_m^n, \tag{4.5a}$$

$$A_2\tilde{\phi}_m^n=(n-2m)\tilde{\phi}_m^n, \tag{4.5b}$$

$$A_3\tilde{\phi}_m^n=[2m(n+2-2m)(n+1-2m)]^{1/2}\tilde{\phi}_{m+1}^n. \tag{4.5c}$$

This action can be integrated to a representation of the solvable Lie group S_3 as

$$\begin{aligned}T(g(\alpha\beta\gamma))f(\eta_1,\eta_2)&\equiv\exp(\alpha A_1+\beta A_2+\gamma A_3)f(\eta_1,\eta_2)\\&=\exp(\alpha A_1)\exp(\beta A_2)\exp(\delta A_3)f(\eta_1,\eta_2)\\&=f(e^{\alpha}\eta_1+e^{2\beta}\delta\eta_2^2,e^{\beta}\eta_2)\\&=e^{n\beta}\eta_2^nP_n(e^{\alpha-2\beta}\eta_1/\eta_2^2+\delta)\end{aligned} \tag{4.6a}$$

where $f\in\mathcal{P}_n$, $\alpha,\beta,\gamma\in\mathbb{C}$, and

$$\delta=\gamma(e^{\alpha-2\beta}-1)/(\alpha-2\beta). \tag{4.6b}$$

The transformations (4.6) form the group of geometrical symmetry transformations in Bargmann space. The group composition law is $g(\alpha_1,\beta_1,\gamma_1)g(\alpha_2,\beta_2,\gamma_2)=g(\alpha_1+\alpha_2,\beta_1+\beta_2,\gamma_3)$ where γ_3 is related to δ_3 through (4.6b) and $\delta_3=\delta_2+e^{\alpha_2-2\beta_2}\delta_1$. This yields the representation matrices

$$\begin{aligned}D^n_{m'm}(\alpha\beta\gamma)&\equiv(\tilde{\phi}^n_{m'},T(g(\alpha\beta\gamma))\tilde{\phi}^n_m)\\&=e^{m'\alpha}e^{(n-2m)\beta}(\sqrt{2}\delta)^{m-m'}\frac{1}{(m-m')!}\left(\frac{m!(n-2m')!}{m'!(n-2m)!}\right)^{1/2},\end{aligned} \tag{4.7}$$

where $0\le m'$, $m\le[n/2]$ and the matrix is upper-triangular, δ being given by (4.6b). Now since A_1 and A_2 are self-adjoint on $\mathcal{D}$ given by (3.3), by choosing α,β pure imaginary, the representation of the Abelian subgroup generated by them defines a unitary representation on $\mathcal{J}$. Of course A_3 is not Hermitean (symmetric) in $\mathcal{J}$, so its integrated group representation is not unitary. Moreover, $\exp(\delta A_3)$ is an unbounded operator in $\mathcal{J}$, since functions of growth $(2,1)$ in η_1 and $(2,\frac{1}{2})$ in η_2 are mapped onto functions of growth $(2,1)$ in η_1 and $(4,\delta)$ in η_2. However, it can be seen easily from (4.6a) that $T(g)$ maps $\mathcal{P}_n$ into itself and thus is densely defined on $\mathcal{J}$. Using the binomial theorem and (4.6a), it follows that all polynomials $\mathcal{P}_j$ of degree $j\le[n/2]$ form an invariant subspace under s_3. The complement $\bar{\mathcal{P}}_j$ of $\mathcal{P}_j$ in $\mathcal{P}_n$ is *not* invariant under (4.6a). Thus the representation (4.6a) of s_3 on $\mathcal{P}_n$ is indecomposable and nonunitary with dimension $[n/2]+1$. This is consistent with a a theorem of Lie which states[12] that all finite-dimensional representations of a solvable Lie group over $\mathbb{C}$ are indecomposable. From the point of view of the Lie algebra (4.3) this means that we have only a lowering operator given by A_3. It can be seen how another representation of the same algebra s_3 contains a raising operator: Indeed, consider the operators defined by

$$A_1'\equiv -A_1,\quad A_2'\equiv -A_2,\quad A_3'\equiv A_3^{\dagger}=2\eta_1\partial_{\eta_2\eta_2}, \tag{4.8}$$

where $\mathcal{D}(A_3^{\dagger})=\mathcal{D}(A_3)$. The primed operators (4.8) form a representation of s_3 conjugate to that of (4.2). In fact, we easily find

$$A_3'\tilde{\phi}_m^n=[2(m+1)(n-2m)(n-2m-1)]^{1/2}\tilde{\phi}^n_{m+1}. \tag{4.9}$$

Since A_3' is a second-order operator, its exponentiation will be represented through an integral kernel in Bargmann space. There is a striking analogy between this exponentiation and the development in time of the solutions of the heat equation.[13] Using this analogy, the general element of the conjugate representation of S_3 can thus be found as

$$\begin{aligned}T(g'(\alpha\beta\gamma))f(\eta_1,\eta_2)&\equiv\exp(\alpha A_1'+\beta A_2'+\gamma A_3')f(\eta_1,\eta_2)\\&=\exp(\delta' A_3')\exp(-\beta A_2)\exp(\alpha A_1)f(\eta_1,\eta_2)\\&=\int\int_{\mathbb{C}}d^2\mu_1(\eta_1')\,d^2\mu_2(\eta_2')\\&\quad\times K_{g'(\alpha\beta\gamma)}(\eta_1,\eta_2;\eta_1',\eta_2')f(\eta_1',\eta_2'),\end{aligned} \tag{4.10a}$$

where $\delta'=\gamma(e^{2\beta-\alpha}-1)/(2\beta-\alpha)$ and the integral kernel is

$$K_{g'(\alpha\beta\gamma)}=\exp(2e^{-\alpha}\eta_1\eta_1'^{*}+e^{-\beta}\eta_2\eta_2'^{*}+2\delta' e^{-2\beta}\eta_1\eta_2'^{*2}). \tag{4.10b}$$

Finally, it is straightforward to see that the matrix elements of the representation (4.10) are the adjoints of the matrix elements (4.7) of the group S_3. However, it must be noted that if we try to embed the two representations (4.2) and (4.8) of the algebra s_3 into a higher-dimensional Lie algebra, we are led to an algebra of infinite dimension.

Now from the relations (4.5c) and (4.9) one can derive the recursion relation (3.12). Forming the inner product of $(\tilde{S}_p-\lambda)$ from (4.9b) between the Cartesian and parabolic bases, this yields

$$\begin{aligned}\lambda(\tilde{\psi}^P_{nl},\tilde{\phi}^n_m)&=[m(n+2-2m)(n+1-2m)]^{1/2}(\tilde{\psi}^P_{nl},\tilde{\phi}^n_m)\\&\quad+[(m+1)(n-2m)(n-2m-1)]^{1/2}(\tilde{\psi}^P_{nl},\tilde{\phi}^n_{m+1}).\end{aligned} \tag{4.11a}$$

Then upon defining

$$(\tilde{\psi}^P_{nl},\tilde{\phi}^n_m)=[(n-2m)!(n-2m+1)m!]^{1/2}2^{-m/2}p^{nl}_m, \tag{4.11b}$$

we regain (3.12). We thus could have made these calculations using the harmonic oscillator raising and lowering operator formalism in ordinary configuration

space; however, the analysis of the differential equations that was made previously requires specific Lie algebra models.

Another important consequence of the symmetry algebra is the correspondence between the separable coordinate systems (2.5) and the orbits of the factor algebra $s_2 \approx s_3/u(1)$ under the adjoint action of S_3. An easy calculation shows that we have essentially two orbits:

$$\text{(i) } \alpha A_1 \text{ and (ii) } A_3. \tag{4.12}$$

As discussed previously, A_1 is a self-adjoint operator on $\mathcal{J}$; in fact, from (3.8a) we have $A_1 = \frac{1}{4}\mathcal{S}_C - \frac{1}{2}$. Thus orbit (i) describes the Cartesian basis. In constrast, the operator A_3 is not Hermitian in $\mathcal{J}$. However, by considering the Hermitian part of A_3, i.e., $\frac{1}{2}(A_3 + A_3^*)$, we find the self-adjoint operator $\mathcal{S}_P = 2^{-1/2}(A_3 + A_3^*)$. Thus orbit (ii) describes the parabolic basis and we have found the correspondence between the two orbits (4.12) of the symmetry algebra and the two separable coordinate systems (2.5). It can also be remarked that the preceeding description of the symmetry algebra also carries over to the case of any anisotropic two-dimensional oscillator whose frequency ratio is $k:1$ (k integer). What does *not* carry over is the connection with separation of variables, and the reason clearly is that for any other ratio of frequencies the operator $A_3^\dagger$ is higher than second order, giving rise to a higher than second order operator for not only the analog of $\mathcal{S}_P$ but also of S_P.

5. CANONICAL TRANSFORMATIONS INDUCED BY THE SYMMETRY GROUP ACTION

In this section we want to show explicitly that the S_3 group action induces a canonical transformation in the Bargmann phase space and point out some of its characteristics. Consider the action of $\exp(\alpha A_1)$ and $\exp(\beta A_2)$; these produce dilatations of the canonical operators $\hat{\eta}_i, \hat{\xi}_i$, i.e., $\hat{\eta}_1 \to e^{\alpha}\hat{\eta}_1$, $\hat{\xi}_1 \to e^{-\alpha}\hat{\xi}_1$ under the first one and $\hat{\eta}_2 \to e^{\beta}\hat{\eta}_2$, $\hat{\xi}_2 \to e^{-\beta}\hat{\eta}_2$ under the second one. It is clear that they preserve the canonical commutation relations $[\hat{\eta}_j, \hat{\xi}_k] = i\delta_{jk}\mathbf{1}$ and the form of the Hamiltonian (3.5). The adjoint group action of $\exp(\gamma A_3)$ gives

$$\hat{\eta}_1 \to \hat{\eta}_1' = \exp(\gamma A_3)\hat{\eta}_1 \exp(-\gamma A_3) = \hat{\eta}_1 + \gamma\hat{\eta}_2^2, \tag{5.1a}$$

$$\hat{\eta}_2 \to \hat{\eta}_2' = \hat{\eta}_2, \quad \hat{\xi}_1 \to \hat{\xi}_1' = \hat{\xi}_1, \tag{5.1b}$$

$$\hat{\xi}_2 \to \hat{\xi}_2' = \hat{\xi}_2 - 2\gamma\hat{\eta}_2\hat{\xi}_1, \tag{5.1c}$$

which can be similarly checked to preserve the canonical commutation relations and the form of the Hamiltonian. Thus S_3 can be said to induce a *canonical transformation*[11,14,15] in the η-space which, moreover, is a nonlinear *point* transformation, as $\hat{\eta}_i'$ is only a function of the $\hat{\eta}_i$ and the group element. The translation to ordinary description of phase space can be made through (3.1) and seen to mix the configuration and momentum components. The transformation (5.1) is not in general a unitary transformation since as was seen in (4.6), $\exp(\gamma A_3)$ is not unitary.

The action of the conjugate representation of S_3 can be obtained through adjunction from (5.1) and similar considerations apply as a canonical transformation. It is not a point transformation, however.

By looking at the transformations (5.1) it is seen that two new operators appear, namely

$$A_4 = \eta_2^2, \quad A_5 = \eta_2\partial_{\eta_1}. \tag{5.2}$$

The generators of s_3 together with (5.2) and the five-dimensional Heisenberg–Weyl algebra w_2 form a solvable dynamical noninvariance algebra of dimension ten, with the structure $s_{10} \approx s_5 \dot{+} w_2$ where s_5 is a five-dimensional solvable algebra with basis $A_1, \ldots, A_5$ and w_2 is an ideal in s_{10}. Similarly, one can construct the conjugate representation from the Hilbert space adjoint operators. The algebra s_{10} can be integrated to the corresponding group on a dense invariant subspace of $\mathcal{J}$. This group is a Lie subgroup of the pseudogroup of all canonical transformations.[14]

6. DISCUSSION ON SYMMETRY GROUPS AND ACCIDENTAL DEGENERACY

The degeneracy pattern for the anisotropic oscillator has usually been attributed[5–7,15] to the group $SU(2)$. We feel, however, that the role of this $SU(2)$ is still not well understood since in contradistinction to the isotropic oscillator case, the formal Lie algebra $su(2)$ for the anisotropic oscillator cannot be written in terms of finite-order differential operators in Hilbert space. The generators of $su(2)$ are written in terms of shift operators which are well defined over the finite-dimensional subspaces; however, their extension to a dense subspace of Hilbert space seems to have been overlooked. Moreover, in order to obtain a unitary irreducible representation[7,15] of the group $SU(2)$ on one of the finite-dimensional subspaces, a new norm must be introduced. This is the meaning of the factors containing the number operator and modulo numbers: One has to rescale the basis functions so that they form a properly normalized $SU(2)$ basis, for they do not do so in the ordinary norm. As a consequence, the representations are nonunitary in the usual Hilbert space norm.

Second, when we follow the procedure of Refs. 5 and 7 for n-dimensional anisotropic oscillators ($n > 2$), the group $SU(n)$ does not in general give a full account of the degeneracy of the system, that is, representations are in general reducible, in fact, completely reducible. This occurs already in the $n = 3$ case and constitutes the major failure for $SU(n)$ as the symmetry group explaining the accidental degeneracy.

Thirdly, the choice of the group $SU(2)$ [$U(2)$ including the action of the Hamiltonian] is not unique. In Ref. 7 this choice was dictated in order to find the quantum counterpart of a classical canonical transformation which maps the general anisotropic oscillator onto the isotropic oscillator whose geometrical symmetry group in Bargmann space *is* $U(2)$. It is of interest to study the former system on its own, since the two quantum problems are not unitarily equivalent.

The generators of the solvable group S_3 on the other hand, are all the first order symmetry operators in Bargmann's description of phase space. They are thus the generators of all the *geometrical* symmetry transformations in Bargmann space, and in this sense they are unique. While the representations are reducible,

they cannot be decomposed into irreducible parts, i.e., they are indecomposable. We can find no fundamental reason why, when explaining accidental degeneracy through a symmetry group, one should exclude non-unitary indecomposable representations. Clearly, completely reducible representations should be excluded. It is thus of interest to consider the n-dimensional generalization of the geometrical symmetry group discussed above.

To sum up, the connection between accidental degeneracy and symmetry groups seems to be still an open question. In this context one should understand the role played by the infinite-dimensional Lie algebras of symmetry transformations and its corresponding Lie pseudogroup. Perhaps more immediate is the possibility of finding, for all systems with discrete spectra exhibiting accidental degeneracy, a Hilbert space *á la* Bargmann such that its group of geometrical symmetry transformations explains the accidental degeneracy. Work in this direction is currently in progress.

ACKNOWLEDGMENTS

The results of Table I were obtained in collaboration with Dr. Victor Guerra (CIMAS). We would like to thank Professor Marcos Moshinsky for his interest in our work.

[1]P. Winternitz, Ya.A. Somorodinskii, M. Uhlíř, and I. Friš, Yad. Fiz. **4**, 625 (1966) [Sov. J. Nucl. Phys. **4**, 444 (1967)].
[2]E.G. Kalnins and W. Miller Jr., J. Math. Phys. **15**, 1728 (1974).
[3]C.P. Boyer, E.G. Kalnins and W. Miller Jr., J. Math. Phys. **16**, 499 (1975).
[4]V. Barmann, Commun. Pure Appl. Math. **14**, 187 (1961); **20**, 1 (1967).
[5]Yu. N. Demkov, Zh. Exp. Theor. Fiz. **44**, 2007 (1963) [Sov. Phys. JETP 17, 1349 (1963)]; L.A. Il'kaeva, Vestnik Leningrad Univ. **22**, 56 (1963); H.V. McIntosh, J. Math. Phys. **11**, 870 (1970); H.V. McIntosh, "Symmetry and Degeneracy" in *Group Theory and Its Applications* (Academic, New York, 1971), Vol. 4.
[6]I. Vendramin, Nuovo Cimento A **54**, 190 (1968); G. Maielia and G. Vilasi, Lett. Nuovo Cimento **1**, 57 (1969).
[7]J.D. Louck, M. Moshinsky, and K.B. Wolf, J. Math. Phys. **14**, 692 (1973).
[8]C.P. Boyer, Helv. Phys. Acta **42**, 589 (1974).
[9]A. Lemieux and A.K. Bose, Ann. Inst. Henri Poincaré, **10A**, 259 (1969).
[10]C.P. Boyer, E.G. Kalnins, and W. Miller Jr., J. Math. Phys. **16**, 512 (1975).
[11]K.B. Wolf, J. Math. Phys. **15**, 1295 (1974).
[12]N. Jacobson, *Lie Algebras* (Wiley Interscience, New York, 1962).
[13]D.V. Widder, Duke Math. J. **29**, 447 (1962).
[14]K.B. Wolf, "The Heisenberg-Weyl Ring in Quantum Mechanics" in *Group Theory and Its Applications*, edited by E. Loebl (Academic, New York, 1975), Vol. III.
[15]P. Kramer, M. Moshinsky, and T.H. Seligman, "Complex Extentions of Canonical Transformations in Quantum Mechanics" in *Group Theory and Its Applications*, edited by E. Loebl (Academic, New York, 1975), Vol. III.

Space averaging techniques of determinantal measures*

I. Kelson and G. Shadmon

Department of Physics and Astronomy, Tel-Aviv University, Tel-Aviv, Israel
(Received 27 June 1974; final revision 18 February 1975)

Techniques for evaluating approximate space-averages of determinantal measures are discussed. The concept of combinatorial structure diagrams is introduced and investigated.

I. INTRODUCTION

The use of the independent particle picture is very widespread in many-body theory. It provides the basis for many approaches in atomic, nuclear, molecular and solid state physics. There are two basic reasons for such wide and diversified use. The first—a physical one—is that under a variety of circumstances, the motion of the particles is indeed semi-independent, either because of the weakness of correlations and interactions, or due to statistical considerations. The second reason—a more pragmatic one—is that technically, mathematically and computationally, the nomenclature and techniques of independent particle motion are almost the only manageable tools at our disposal. It is because of the prevalence of this approach, that we consider it worthwhile to investigate in greater depth some of its quantitative and qualitative aspects. Particularly because modifications and generalizations of the theory may thus be better understood. We have already begun[1] a line of research, which we follow up here in some specialized directions. Since the subject matter of this work deals with new concepts and entities, we wish to begin by pointing out its relevance to the physics involved.

We should emphasize from the outset that we deal only with systems of fermions; primarily because the properties of an interacting assembly of bosons involve completely different concepts and techniques. Moreover, we work explicitly with finite dimensional vector spaces, with prescribed numbers of particles and of orbitals. The limitation on the number of particles simply means that we only deal with finite systems (such as nuclei, atoms and molecules) rather than infinite systems (such as solids and infinite nuclear matter). The limitation on the number of orbits is equivalent to a truncation of the complete, infinite-dimensional vector space to a finite-dimensional model space, and entails an approximation of substance. In fact, the validity of such a truncation is in itself a central problem in the theory of any finite system, along with its relation to the model Hamiltonian.

Having once limited ourselves to a finite, well-defined vector space of completely antisymmetrized many-particle wavefunctions, we turn our attention to approximate descriptions of the ground state (and other low excited states) of the system. One must remember a very important point here. The validity (though not necessarily the value) of any given approximation is determined by its ability to reproduce the exact states of the *model* problem, not those of the actual physical system. Thus, when it is possible to solve the truncated problem exactly, one has a direct way of testing various approximations, and it is irrelevant whether the exact solution does or does not provide an acceptable description of the real system under study. This is why, in our general study, the precise nature of the orbits is of no importance. All multiparticle systems with a given number of particles k, and a given number of orbitals N are thus, for our purposes, identical. In Sec. II we summarize briefly the elements of previous work concerning this space $S(N, k)$ and some concepts related to it. In particular, we discuss a general subspace $S_D^{(n)}(N, k)$ which contains all states that can be expressed as a sum of n, but not of less, independent particle determinants. The importance of these subspaces is readily apparent. As is well known, the ground state of a given Hamiltonian satisfies the variational principle, requiring that it minimizes the expectation value of that Hamiltonian. The Hartree–Fock approximation is obtained by limiting the variational principle to trial wavefunctions belonging to the subspace which we defined as $S_D^{(1)}(N, k)$. The straightforward generalizations of the Hartree-Fock theory, which we investigate, involve the application of the variational search in $S_D^{(n)}(N, k)$ with arbitrary n.

Various relevant questions now arise: How well can a given state be approximated by a state belonging to one of these subspaces, $S_D^{(n)}(N, k)$? How does the answer depend on the characteristics of that given state, particularly its energy? What is the quality of the spanning by $S_D^{(n)}(N, k)$ of the complete space $S(N, k)$? Obviously, the answers to these and similar questions, are essential in determining the usefulness of the generalized Hartree–Fock approximation. Only in the simplest of cases, and for the simplest of problems, can quantitative results be obtained directly—by statistical methods, as indicated at the end of Sec. II.

To perform such studies in general, new investigative tools are necessary, and these call for the introduction of new concepts and approximations. Thus, in Sec. III we introduce "combinatorial structure diagrams," along with a preliminary display of some of their simplest applications. In very broad terms, a combinatorial structure diagram describes the interrelationships between the various basis states appearing in the expansion of a given state (or family of states), where the multiparticle basis is composed of antisymmetrized independent particle wavefunctions in some standard one-body representation (of dimension N). The basic relationship involved between two such multiparticle basis states is the number of *single*-particle states they have in common (which clearly cannot exceed $k-1$). How this bears on our previous

discussion can be intuitively grasped by the following example. Consider two k-particle basis states which differ by precisely one single-particle orbital. Any linear combination of these two states [which would appear to belong to $S_D^{(2)}(N,k)$] can trivially be written as a single determinant in a different single-particle representation, and thus belongs—by definition—to $S_D^{(1)}(N,k)$. The combinatorial relationship between the two basis states (in this trivial example) determines the properties of the subspace which they span; namely, the $S_D^{(n)}(N,k)$ to which it belongs. Generally, of course, such a determination is neither unique nor can it be made exactly, but it nevertheless contributes directly and quantitatively to the problems which we have set out to solve.

We should also add that the purpose of this work is to introduce these concepts and to point out possible directions for their investigation, rather than to provide an exclusive and comprehensive study of their properties.

II. THE MANY FERMION SPACE

A. The independent particle description

We consider a system of k identical particles, fermions, which are restricted to a finite, N-dimensional single-particle complex vector space.[2] This space $S(N,1)$ is spanned by an orthonormal set of single-particle states

$$\phi_i \equiv |i\rangle \equiv a_i^+|_0\rangle,\ i=1,\ldots,N, \tag{1}$$

where nothing is specified about the detailed properties of these basis states. The multiparticle space $S(N,k)$ is defined as the antisymmetrized direct product of k single-particle spaces $(S(N,1)\times\cdots\times S(N,1))_a$. As a basis for $S(N,k)$ we may choose the $B(N,k)$ functions[3]

$$\Phi_{i_k\cdots i_k} \equiv \mathcal{A}\{\phi_{i_1}(1)\cdots\phi_{i_k}(k)\},\quad i\le i<i_2\cdots<i_k\le N. \tag{2}$$

The antisymmetrizing operator is defined as

$$\mathcal{A}\equiv\frac{1}{\sqrt{k!}}\sum(-1)^P P$$

where the summation runs over all the permutations of the arguments $1\cdots k$. Clearly, this is an orthonormal basis of $S(N,k)$, and any vector ψ in it can be expanded,

$$\psi=\sum_{1\le i_1<\cdots<i_k\le N} d_{i_1 i_2\cdots i_k}\phi_{i_1\cdots i_k} \tag{3}$$

It is worthwhile, for clarity of notation and ease of formal manipulation, to work with unrestricted sums in Eq. (3); namely to substitute the explicit expansion of the determinantal wavefunctions. Thus,

$$\psi=\sum_{i_1,\ldots,i_k=1}^{N} C_{i_1\cdots i_k}\phi_{i_1}(1)\cdots\phi_{i_k}(k) \tag{4}$$

where C is a totally antisymmetric tensor, which—up to a phase—uniquely specifies ψ. The scalar product of two states $\psi^{(1)}$ and $\psi^{(2)}$ is simply given by contracting over all the indices of their representative tensors

$$\langle\psi^{(1)}|\psi^{(2)}\rangle=\sum_{i_1,\ldots,i_k=1}^{N} C^{(1)*}_{i_1\cdots i_k}C^{(2)}_{i_1\cdots i_k} \tag{5}$$

Similarly, the n-body density matrix corresponding to a given state ψ, is obtained by contracting over k-n indices,

$$\rho_{\alpha_1\cdots\alpha_n,\beta_1\cdots\beta_n}=\sum_{i_{n+1},\ldots,i_k=1}^{N} C_{\alpha_1\cdots\alpha_n i_{n+1}\cdots i_k}C^{*}_{\beta_1\cdots\beta_n i_{n+1}\cdots i_k}. \tag{6}$$

Upon a change of the single-particle basis (1) the many-particle basis is transformed, along with the tensor $C_{i_1\cdots i_k}$. Let the single particle transformation be effected by the orthogonal matrix O

$$\phi_i=\sum_{j=1}^{N}O_{ij}\phi'_j,\quad \phi'_j=\sum_{i=1}^{N}O_{ij}\phi_i, \tag{7}$$

then

$$C'_{i_1\cdots i_k}=\sum_{j_1\cdots j_k} C_{j_1\cdots j_k}O_{j_1 i_k}\cdots. \tag{8}$$

The antisymmetry is obviously conserved in the transformation; however, the apparent structural simplicity of a state (say, its being a pure determinant) may be destroyed by such a transformation.

B. Determinantal measures

Let $S_D^{(1)}(N,k)$ be a subspace of $S(N,k)$, containing all states which are pure Slater determinants. Namely, all those states for which a single particle orthonormal basis exists, such that the expansion (3) contains only one element. More generally, let $S_D^{(n)}(N,k)$ be a subspace of $S(N,k)$ containing all states which can be written as a linear combination of n, but not of less, states belonging to $S_D^{(1)}(N,k)$. It must be remembered that the n Slater determinants appearing in this linear combination are in general not mutually orthogonal.

We define as the n-determinantal measure of a given state ψ

$$D^{(n)}(\psi)=\frac{1}{\langle\psi|\psi\rangle^{1/2}}\max_{\Phi\in S_{D^{(n)}(N,k)}}\frac{|\langle\Phi|\psi\rangle|}{\langle\Phi|\Phi\rangle^{1/2}}. \tag{9}$$

Since $|\langle\Phi|\psi\rangle|/\langle\Phi|\Phi\rangle^{1/2}$ is bounded from above, and since the variation is carried over a closed domain (which $S_D^{(n)}$ trivially is), the maximum $D^{(n)}(\psi)$ is always attained. It may, however, be attained at multiple points of $S_D^{(n)}$.

In Ref. 1 we have shown an explicit procedure for solving for $D^{(1)}(\psi)$ and the corresponding determinantal wavefunction Φ. Let

$$\Phi=\mathcal{A}(\hat\phi_1(1)\cdots\hat\phi_k(k))$$

where the single particle states $\{\hat\phi_i\}$ are related to the basis states $\{\phi_i\}$ through an orthogonal transformation matrix O [see Eq. (7)]. The variational parameters are thus the elements $\{O_{\alpha i_\alpha},\ \alpha=1,\ldots k;\ i_\alpha=1,\ldots,N\}$, and they have to satisfy the equations

$$\sum_{i_1\cdots i_{\alpha-1}i_{\alpha+1}\cdots i_k}^{N} C_{i_1\cdots i_k}O_{1i_1}\cdots O_{\alpha-1\,i_{\alpha-1}}O_{\alpha+1\,i_{\alpha+1}}\cdots O_{ki_k} = D^{(1)}(\psi)O_{\alpha i_\alpha}, \tag{10}$$

$$\alpha=1,\ldots,k,\quad i_\alpha=1,\ldots,N$$

where $C_{i_1\cdots i_k}$ is the totally antisymmetric tensor of Eq. (4).

Furthermore, we described a general, numerical

(Monte-Carlo) procedure for evaluating the average of $D^{(1)}(\psi)$ over the entire space $S(N,k)$. For this procedure to be well defined, one must have a weight attached to each infinitesimal segment dS of the space. One falls back on the Cartesian picture of $S(N,k)$, which utilizes the $B(N,k)$ basis states as primary axes. A point in S is thus characterized by the Cartesian components $\{X_i\}$, which are assumed to be normalized to unity. Then,

$$\langle D^{(1)}\rangle = \frac{\int_{S(N,k)} D^{(1)}(\psi)\, dS}{\int_{S(N,k)} dS} \tag{11}$$

where the weight is simply the surface area of dS.

The immense dimensionalities involved, even for primitive spaces, inhibit the direct performance of the integration. Rather, one resorts to Monte-Carlo methods, that is, to the utilization of selection procedures, which reflect, through their probability, the weighting agreed upon. Essentially, one has to select a point (a state) within dS, with a probability proportional to the area of this infinitesimal element. This is achieved by successive choices of the $B(N,k)$ components of ψ. The probability of finding the ith component with absolute magnitude between x and $x+dx$ is

$$P(x \leq x_i \leq x+dx) \sim \frac{d}{dx}\, V_{B(N,k)-i}[(1-R_{i-1}^2-x^2)^{1/2}]dx \tag{12}$$

where

$$R_{i-1}^2 = \sum_{j=1}^{i-1} X_j^2 \tag{13}$$

and $V_n(R)$ is the volume of an n-dimensional sphere of radius R

$$V_n(R) = C_n R^n = \frac{\pi/2}{(n/2)!} R^n. \tag{14}$$

Thus, the magnitudes of the components are selected, then their phases, followed by solving Eq. (10). Some numerical results of this entire procedure are quoted in the next section, where approximate methods to circumvent this cumbersome apparatus are discussed.

III. APPROXIMATE SPACE AVERAGING TECHNIQUES

A. Combinatorial structure diagrams (C.S.D.)

Let there be N orbits, k particles (fermions) distributed amongst them, and M component states. Disregarding the coefficients of these states, we assume they all are in the same single particle representation; we are therefore dealing with an assembly of M k-sets chosen from among N objects. We are not interested in the specific identity of the $M\times k$ objects, but rather in their interrelations, which will be characterized by the combinatorial structure diagram (C.S.D.).

Given m k-sets, we can specify the total number Γ of different objects which appear in their union (or "circumference"). For example, if $N=8$, $k=3$, $m=3$, and the three 3-sets are (1, 2, 3), (1, 4, 5), and (3, 4, 6), then $\Gamma=6$.

This number, Γ, is clearly invariant under all permutations of the symmetric group P_N. This is why we only have to specify Γ's associated with any given assembly and not the actual objects appearing in it. All assemblies which have the same Γ's are said to have the same combinatorial structure diagram.

There are altogether $B(N,k)$ ways of forming k-sets. We arbitrarily agree, for the sake of the following, that no sign change is associated with obtaining equivalent sets by permuting the k objects of the set among themselves. There are $B(B(N,k),M)$ ways of forming different assemblies. We shall associated a sign change with all odd permutations of the M k-sets among themselves. As we indicated, each C.S.D. is characterized by the set of all circumferences (Γ) of all possible subassemblies of m k sets, with $m=0,\ldots,M$. The circumferences Γ appearing are clearly not all independent. To name a few restrictions (given N and k):

for $m=0$, trivially $\Gamma\equiv 0$,

for $m=1$, trivially $\Gamma\equiv k$,

for $m=2$, $k+1\leq\Gamma\leq 2k$.

This is a particular example of a general triangular inequality holding between Γ's. To make this point clearer, we may talk about the combinatorial "distance" between two sets, which is equal to $\Gamma-k$, and which gives the number of different objects in each set. If the distance is larger than 1, we shall refer to these as "alien" combinations. We note here, that two component states whose combinations are alien—cannot be connected by a one-body operator. Similarly, any n-body operator can only have nonvanishing matrix elements between component states no more than n apart. In Appendix A, we discuss the relation between the set of $\{\Gamma\}$ and the monomial representations of the permutation group P_N.

In view of the importance assigned to independent particle description, which in turn emphasized the difference between alien and nonalien (allied) combinations, we shall group together all C.S.D.'s which have the same number L, and topological structure, of pairs of combinations (component states) which are allied, that is differ from each other by one object (single particle state) only. The range of L is $0\leq L\leq B(M,2)$, but both limits are sometimes not realized, depending on N and k. We shall utilize tree, or link, diagrams, in which M points represent the component states, and lines connect those which are allied. Clearly, the diagrams are characterized only by M and L (aside from the link-topology), while N and k only enter into the evaluation of the weights of the diagrams, as we shall see below.

In Fig. 1, we give for example all different link-diagrams of C.S.D.'s with $2\leq M\leq 5$.

B. Enumeration of C.S.D.'s

We shall begin by some general observations. Having restricted ourselves to the $(N,k)M$, L classification of C.S.D.'s, we may state that the smaller L is, the larger is its relative weight (or its total number of combinations). This is so because each link means another imposition, or restriction, which limits the

FIG. 1. All the topologically distinct combinatorial structure diagrams, with the number of component states (points) M, less than or equal to 5. A straight line linking two points means that they are allied, namely, they differ from each other by precisely one single particle state. The number L is the total number of such links in a diagram.

freedom of choice of the particular structure diagram. This, however, is only true when there is little correlation, or self-restriction, amongst the k-sets. To make this point clearer, let $\mathcal{N}(N, k)$ be the largest number of k-sets, all mutually alien, that can be constructed out of N objects (an interesting combinatorial problem, which is closely related to questions of state representation). Then, in particular, when $M > \mathcal{N}(N, k)$, we have $C(L=0)=0$, where $C(\ )$ stands for the absolute combinatorial count. In this case, we would say, the mere smallness of the single particle space imposes descriptive restrictions on multiple-component many-particle states. However, when $M \ll \mathcal{N}(N, k)$, $C(L)$ is a monotonically decreasing function of L. We can estimate the ratio of $C(L+1)$ to $C(L)$, by assuming all links do correspond to independent restrictions, in which case

$$\frac{C(L+1)}{C(L)} \approx \frac{\text{measure of } (\bullet\!-\!\circ)}{\text{measure of } (\circ\ \bullet)} = \frac{B(N, k+1)B(k+1, 2)}{B(B(N, k), 2) - B(N, k+1)B(k+1, 2)}, \tag{15}$$

$$L \ll M \ll \mathcal{N}(N, k).$$

For $N=10$, $k=5$, for example, this ratio is about 0.025. For fixed k and large N, the ratio behaves asymptotically as $N^{-(k-1)}$. The key point is that the $L=0$ C.S.D. is dominant, or at most the lowest few L-values. When the conditions stated above do not hold, $C(L)$ will have a maximum at some value of L different from zero.

We give below some specific examples of interest.

$M=2$: This can be trivially evaluated, since the $L=0$ and $L=1$ are the only possibilities. We have

$$C(L=1) = B(B(N, k), 2) \times \frac{k(N-k)}{k(N-k) + B(k, 2)B(N-k, 2) + \cdots + B(k, k)B(N-k, k)} \tag{16}$$

$$C(L=0) = B(B(N, k), 2) - C(L=1). \tag{17}$$

$L=B(M, 2)$, $M>2$: This corresponds to a C.S.D. in which all combinations are allied to all others. We note that there are two distinct types of M; $L=B(M, 2)$ C.S.D.'s which have a different total Γ, one with $\Gamma = k+1$ and the other with $\Gamma = k+M-1$, which may be viewed as particle-hole conjugates. By adding a different element (state) to each of the k-sets in the $\Gamma = k+1$ C.S.D. one arrives at a common $(k+1)-$ set. On the other hand, by subtracting a different element from each of the k-sets in the $\Gamma = k+M-1$ C.S.D., one arrives at a common $(k-1)$-set. Obviously, the two kinds become identical for $M=2$.

The corresponding count is

$$C(L=B(M, 2), \Gamma = k+1) = B(N, k+1)B(k+1, M) \tag{18}$$

$$C(L=B(M, 2), \Gamma = k+M-1) = B(N, k-1)B(N-k+1, M), \tag{19}$$

and

$$C(M>2, L=B(M, 2)) = B(N, k+1)B(k+1, M) + B(N, k-1)B(N-k+1, M). \tag{20}$$

Of the two components, the $\Gamma = k+M-1$ dominates for large N.

TABLE I. Combinatorial count of diagrams for $M=3$, $k=3$, $N=6$, 8, 10. In parenthesis are given the normalized weights of the various diagrams.

	$M=3\ N=6\ k=3$	$M=3\ N=8\ k=3$	$M=3\ N=10\ k=3$
$L=0$	120 (0.105)	10 080 (0.364)	154 000 (0.548)
$L=1$	540 (0.474)	13 440 (0.485)	108 360 (0.386)
$L=2$	360 (0.316)	3360 (0.121)	15120 (0.054)
$L=3$	120 (0.105)	840 (0.030)	3360 (0.012)

$M=3$: The enumeration involves a slightly more complicated combinatorial procedure. In Table I, we give the results for some cases, where—in parentheses—we give the percentages as well.

Note that with increasing N (and hence approaching the fulfillment of the above-mentioned conditions) the $L=0$ C.S.D. becomes dominant. To demonstrate the behavior in small spaces, we also give the weights for the $M=4$, $N=6$, $k=3$ case in Table II.

In all cases the numbers sum up to $B(B(N,k),M)$, as they should.

$L=0$. As we have pointed out above, this is the predominant case, and, from the combinatorial point of view, it is the most intriguing. We shall employ here an approximation for its calculation, which may be generalized to other low L values, and which is quite amusing.

Clearly, if no two k-sets are to be allied, then no $(k-1)$-set should appear more than once, in the ensemble of M k-sets. We divide our counting into two stages. First, we consider all the $B(N,k-1)$ $(k-1)$-sets as independent objects, and only take into account their gross combinatorial coupling. Second, we consider the correlations that must exist between them so that they actually form M k-sets, but neglecting the effect of interference between them. We note that each k-set "uses" or ("uses up") k $(k-1)$-sets. Thus the total count will be approximated by giving the number of ways of selecting M equivalent groups of k $(k-1)$-sets each, and multiplying it by the probability that each of these groups, independently, forms a k-set. Thus,

$$C(N,k,M,L=0)\approx\frac{L(N,k-1)!}{M!(k!)^M(B(N,k-1)-Mk)!}\times\left(\frac{B(N,k)}{B(B(N,k-1),k)}\right)^M. \tag{21}$$

To demonstrate the power of this approximation, and its region of validity (lack of interference) we give some of the approximate results (rounded to the closest integer), along with the exact ones, in Table III.

$L=1$: Much the same logic may be applied to the $L=1$ case, with the added complication of the linked pair, which removes the approximate equivalence of all k-sets. The result is

$$C(N,k,M,L=1)\approx\frac{B(N,k-1)!}{(M-2)!(k!)^{M-2}(B(N,k-1)-Mk+1)!}\times\left(\frac{B(N,k)}{B(B(N,k-1),k)}\right)^{M-2}\frac{B(N,k+1)B(k+1,2)}{B(B(N,k-1),2k-1)}. \tag{22}$$

Table IV gives a comparison with the exact results for a few cases, where the agreement is again seen to be very good.

$L=2$: To demonstrate the power (and complexity) of this approximation we give the results for $L=2$, where one has to differentiate between the topologies of the links—joint vs. disjoint. Thus

$$C(N,k,M,L=2,\text{joint})\approx\frac{B(N,k-1)!}{(M-3)!(k!)^{M-3}(3k-2)!(B(N,k-1)-Mk+2)!}\times\left(\frac{B(N,k)}{B(B(N,k-1),k)}\right)^{M-3}\frac{2B(N,k+2)B(k+2,2)B(k,2)}{B(B(N,k-1),3k-2)} \tag{23}$$

$$C(N,k,M,L=2,\text{disjoint})\approx\frac{B(N,k-1)!}{(M-4)!(k!)^{M-4}2!(2k-1)!^2(B(N,k-1)-Mk+2)!}\times\left(\frac{B(N,k)}{B(B(N,k-1),k)}\right)^{M-4}\left(\frac{B(N,k+1)B(k+1,2)}{B(B(N,k-1),2k-1)}\right)^2 \tag{24}$$

It is easy to check that for large N, larger spaces, the approximations become increasingly valid, and that indeed the asymptotic ratio holds.

C. Reduction of C.S.D.'s

A "reduced" C.S.D. (namely one without links, or with $L=0$) is of particular interest, not only because of its dominance, but also because it represents a wavefunction whose determinantal measure is susceptible to the above-mentioned simple approximation. We therefore would like to devise procedures which will allow us to reduce, or transform, complicated C.S.D.'s to reduced ones, by removing the links in them. The procedure is not unique, and furthermore, it necessitates assigning different weights to different points in the diagram. This last added complexity may be simply ignored, as an approximate method.

The basis for the reduction procedure is as follows. Given a subset of m points, all of which are mutually linked, then for any arbitrary amplitudes assigned to them, they can be written as one single point (namely, one independent particle determinant) in some representation. These m points can therefore be replaced by one point, with m times the elementary unit weight, and its relations, or connection, with the other points in the diagram will be determined by the statistical behavior of its components. In general, one should start by reducing the largest aggregates. When more than one way of reduction is possible, leading to dif-

TABLE II. Combinatorial count for $M=4$, $N=6$, $k=3$.

	$M=4\ N=6\ k=3$
$L=0$	105 (0.022)
$L=1$	360 (0.027)
$L=2$	1440 (0.297)
$L=3$	1740 (0.359)
$L=4$	810 (0.167)
$L=5$	360 (0.074)
$L=6$	30 (0.006)

TABLE III. Approximate combinatorial count for $L=0$. The exact values, obtained by direct solution, are shown alongside the approximate one (truncated to the closest integer).

N, k, M	Exact	Approximate
6 3 3	120	119
8 3 3	10 080	9660
10 3 3	154 000	150 060
8 4 2	1855	1806
6 3 4	105	27
7 3 6	210	84
7 3 7	30	1

ferent final results, one should either devise further rules for selecting the most appropriate way, or else decide on a recipe for averaging over the diverse possibilities.

The mechanics of the reduction are best understood through a direct demonstration (Fig. 2). We keep track of the weights associated with each point, where it is understood that no weight quoted means unity. Also, in contrast to previous notation, we indicate links by wavy lines, and alien pairs by straight lines.

The complete C.S.D.: Of special interest is the complete C.S.D., that is, where M has its maximal value $M=B(N,k)$. This diagram represents the structure of a state in the entire k-particle space, and hence is unique in structure, and its reduction is important. A pictorial representation of the complete diagram, except for the most trivial cases, is unfeasible, and there is a multitude of ways for reducing it. Figure 3 gives, as a very simple example, the case $N=4$, $k=2$, $M=6$. Near each point we indicate the state represented.

D. Integrals represented by reduced C.S.D.'s

Following the definition, enumeration and reduction of C.S.D.'s, we now get to the evaluation of the (averaging) integrals represented by the reduced C.S.D.'s. A reduced C.S.D., in general, is characterized by the weights $m_1,\ldots,m_\nu$ associated with the points, which are subject to the condition

$$\sum_{i=1}^{\nu} m_i = M. \tag{25}$$

The integral represented (which is the large-space approximation to the single-determinantal measure) is

$$K_{[m_1,\cdots,m_\nu]} \equiv \frac{\int_0^1\cdots\int_0^1 dx_1\cdots dx_M\,\delta(x_1^2+\cdots+x_M^2-1)\max(r_1,\ldots,r_\nu)}{\int_0^1\cdots\int_0^1 dx_1\cdots dx_M\,\delta(x_1^2+\cdots+x_M^2-1)} \tag{26}$$

TABLE IV. Approximate combinatorial count for $L=1$. The exact values, obtained by direct solution, are shown alongside the approximate ones (Truncated to the closest integer).

N, k, M	Exact	Approximate
6 3 3	540	475
8 3 3	13 440	12 715
10 3 3	108 360	105 100

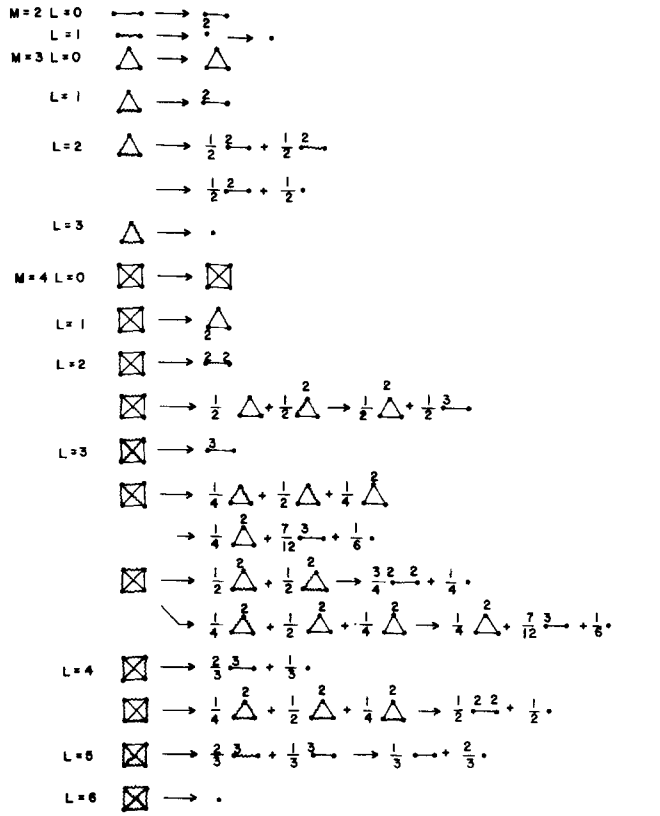

FIG. 2. Pictorial representation of the procedure of reducing combinatorial structure diagrams, to ones without links. Here, a straight line between two points means they are alien, while a wavy line means they are allied. A subset of m points, all of which are linked with wavy lines, can be collapsed into a single point with weight m. The nature of the linkage of the collapsed point to other points, is statistically determined by the initial m connections. This factor is carried in front of the diagrams, while the weights at each point are given, in case they are not unity. A repetition of this procedure, not necessarily in unique order, results in the final reduction.

where

$$r_i = \left(\sum_{j=1}^{m_i} x_j^{(i)2}\right)^{1/2}, \tag{27}$$

namely the "radius" of the ith member of the partition (m). One may, in fact, characterize the partition in much the same way as one does Young tableaux.

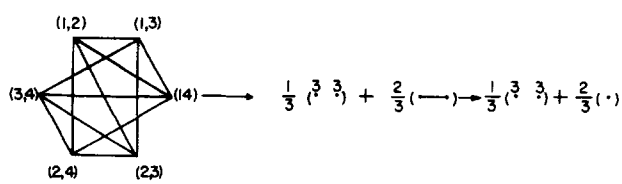

FIG. 3. The complete combinatorial structure diagram for 2 particles in 4 states. The numbers next to the vertices stand for the single particle states making up the 6 two-particle states. The reduction is effected by collapsing simultaneously two complementary triangles.

The evaluation of integrals of this form is in general not a straightforward matter. The denominator is just the surface area of the M-dimensional spherical "quadrant" (namely, 2^{-M} of the sphere). We shall designate it

$$\sigma_M \equiv \int_0^1 \cdots \int_0^1 dx_1 \circ \cdots dx_M \delta(x_1^2 + \circ\circ\circ + x_M^2 - 1) = \frac{M\pi^{[M/2]}}{2^M (M/2)!} \tag{28}$$

with the low σ_M being

$$\sigma_1 = 1, \quad \sigma_2 = \pi/2, \quad \sigma_3 = \pi/2, \quad \sigma_4 = \frac{\pi^2}{8},$$
$$\sigma_5 = \pi^2/12, \quad \sigma_6 = \pi^3/64, \quad \sigma_7 = \frac{\pi^3}{120}, \quad \sigma_8 = \pi^4/768. \tag{29}$$

Rather than try to evaluate the general integral, we develop a recurrence relation based on a simple probabilistic idea. Expressed in words it is as follows:

[The probability density that for a given partition $(m_1, \ldots, m_\nu)$ the maximum of $r_1, \ldots, r_\nu$ lies within the range $(x, x + dx)$] = [The probability that r_i lies within the range $(x, x + dx)$, times the probability that the maximum of the rest of the partition is less than x—summed over i].

Symbolically,

$$G_{\{m\}}(x) = \sum_{i=1}^{\nu} f_{m_i}^{(M)}(x) \int_0^x \hat{G}_{\{m\}-m_i}(\tau; \sqrt{1-x^2})\, d\tau \tag{30}$$

where $G_{\{m\}}(x)$ is the probability we are after.

$\hat{G}_{\{m\}}(x, R)$ is the corresponding generalized probability density function, where the variables are restricted to a spherical surface of radius R (rather than unity) and is susceptible to trivial scaling.

Hence, *assuming all functions f, G to be normalized*, we have

$$G_{\{m\}}(x) = \sum_{i=1}^{\nu} f_{m_i}^{(M)}(x) \int_0^{\min[1, x/(1-x^2)^{1/2}]} G_{\{m\}-m_i}(t)\, dt, \tag{31}$$

$$K_{\{m\}} = \frac{\int_0^1 x G_{\{m\}}(x)\, dx}{\int_0^1 G_{\{m\}}(x)\, dx}. \tag{32}$$

The probability $f_{m_i}^{(M)}(x)$ is of a simple form. It can easily be shown to be proportional to

$$f_{m_i}^{(M)}(x) \sim (\sqrt{1-x^2})^{M-m_i-2}. \tag{33}$$

The normalization is

$$\frac{1}{\xi_{M-m_i}} \equiv \int_0^1 (1-x^2)^{(M-m_i-2)/2}\, dx = \frac{(M-m_i-2)!!}{(M-m_i-1)!!} \Xi_{M-m_i}, \tag{34}$$

where

$$\Xi_\alpha = 1, \quad \alpha \text{ even}$$
$$= \frac{\pi}{2}, \quad \alpha \text{ odd}.$$

Thus

$$f_{m_i}^{(M)}(x) = \xi_{M-m_i}(1-x^2)^{(M-m_i-2)/2} \tag{35}$$

and

$$G_{\{m\}}(x) = \sum_{i=1}^{\nu} \left(\xi_{M-m_i}(1-x^2)^{(M-m_i-2)/2} \times \int_0^{\min[1, x/(1-x^2)^{1/2}]} G_{\{m\}-m_i}(t)\, dt \right). \tag{36}$$

Having calculated and normalized all previous probability density functions, this, then, provides a recurrence relation for $G_{(m)}(x)$, and it can be used numerically.

Numerical solution for string partitions: A string partition is one for which $\nu = M$ and all the weights $m_i = 1$. One can forego the necessity to normalize $G(x)$, and dispense with the summation, since all terms are equivalent. Writing $G_M(x)$ for these functions we have

$$G_M(x) \sim (1-x^2)^{(M-3)/2} \int_0^{\min[1, x/(1-x^2)^{1/2}]} G_{M-1}(t)\, dt,$$
$$G_1(x) = \delta(x-1). \tag{37}$$

Regarding the technical aspects of the integration we shall just mention that it is convenient to make a change of variables from x to $y = 1/x^2$. The first 20 values of K_M are given in Table V.

The special significance of the string partitions, is that they may serve as a working approximation for all partitions, simply by equating $K_{\{M'\}}$ where $M' > M$, to K_M where $M = \nu$, which is the number of weights different from 0 in (M').

Direct integration for simple partitions: The case $\nu = 1$ (no partition at all) gives trivially $G \equiv \delta(x-1)$ and $K = 1$. Let us consider the case $\nu = 2$, where M is partitioned into $M = \alpha + \beta$. We have

$$K_{\{\alpha,\beta\}} = \frac{\int_0^1 \cdots \int_0^1 dx_1 \cdots dx_M \delta(x_1^2 + \cdots + x_M^2 - M) \max(\sqrt{x_1^2 + \cdots + x_\alpha^2}, \sqrt{x_{\alpha+1}^2 \cdots + x_M^2})}{\int_0^1 \cdots \int_0^1 dx_1 \cdots dx_M \delta(x_1^2 + \cdots + x_M^2 - 1)}. \tag{38}$$

The denominator is $\frac{1}{2}\sigma_{\alpha+\beta}$ [the factor $\frac{1}{2}$ comes because of the Jacobian in transforming the δ-function to $\delta(\sqrt{x^2 + \cdots + x_M^2} - 1)$].

We transform the coordinates as

$$dx_1 \cdots dx_\alpha \to r_1^{\alpha-1}\, dr_1\, d\Omega_1,$$
$$dx_{\alpha+1} \cdots dx_{\alpha+\beta} \to r_2^{\beta-1}\, dr_2\, d\Omega_2 \tag{39}$$

and we divide the region of integration according to whether r_1 or r_2 is the maximum. Therefore,

$$\tfrac{1}{2} K_{\{\alpha,\beta\}} \sigma_{\alpha+\beta} = \int_{1/\sqrt{2}}^1 dr_1 r_1^{\alpha-1} \int d\Omega_1 \times \int_0^{r_1} dr_2 r_2^{\beta-1} \delta(r_2^2 - (1-r_1^2)) \int d\Omega_2 + \int_{1/\sqrt{2}}^1 dr_2 r_2^{\beta-1} r_2 \int d\Omega_2 \int_0^{r_2} dr_1 r_1^{\alpha-1} \delta(r_1^2 - (1-r_2^2))\, d\Omega_1.$$

Therefore,

$$\frac{1}{2}\frac{\sigma_{\alpha+\beta}}{\sigma_\alpha \sigma_\beta} K_{\{\alpha,\beta\}} = \frac{1}{2} \int_{1/\sqrt{2}}^1 dr [r^\alpha (1-r^2)^{(\beta-2)/2} + r^\beta (1-r^2)^{(\alpha-2)/2}] \tag{41}$$

or, substituting $r = \cos x$, we obtain (40)

TABLE V. Values of integrals K_M for string partitions.

M	K_M	M	K_M
1	1.	11	.5926
2	.9003	12	.5772
3	.8312	13	.5631
4	.7784	14	.5503
5	.7376	15	.5384
6	.7038	16	.5275
7	.6751	17	.5173
8	.6504	18	.5078
9	.6289	19	.4989
10	.6098	20	.4905

TABLE VII. Average values of determinantal measures. Along with the numerical values obtained by the procedure described in Sec. II B, the approximate values are shown.

N	k	$\mathcal{N}(N, k)$	$\langle D\rangle_{\text{numerical}}$	$\langle D\rangle_{\text{approx.}}$
4	2	2	.96	.950
6	2	3	.88	.865
8	2	4	.79	.805
10	2	5	.73	.758
6	3	4	.83	.805
8	3	7	.65	.689
10	3	13	.56	.570
8	4	14	.59	.557

$$K_{\{\alpha,\beta\}}=\frac{\sigma_\alpha\sigma_\beta}{\sigma_{\alpha+\beta}}\int_0^{\pi/4}(\cos^\alpha x\sin^{\beta-1}x+\cos^\beta x\sin^{\alpha-1}x)\,dx. \tag{42}$$

The integral, incidentally, is simply related to the incomplete β-function, as can be seen through the transformation $r^2=y$, giving

$$K_{\{\alpha,\beta\}}=\frac{\sigma_\alpha\sigma_\beta}{2\sigma_{\alpha+\beta}}\int_{1/2}^1 dy[y^{(\alpha-1)/2}(1-y)^{(\beta-2)/2}+y^{(\beta-1)/2}\times(1-y)^{(\alpha-2)/2}]. \tag{43}$$

Table VI gives some numerical examples, for all partitions such that $\alpha+\beta\leq 8$.

E. Some combined applications

(i) We begin by calculating the space average (given N and k) of the single-determinantal measure. One must remember that this is a highly nonlinear problem, with many complicated features. It can be evaluated numerically—and this was done—for simple cases, where statistical sampling is feasible. Within the framework described above, we employ the following procedure. We note that the reduction procedure of the complete $M=B(N,k)$ diagram, will end up with ν no larger than $\mathcal{N}(N,k)$. We also know that but for the smallest spaces, this is dominant. We therefore use the following approximation, which is quite *ad hoc*:

$$\langle D^{(1)}\rangle_{N,k}\sim\tfrac{1}{2}(K_{\mathcal{N}(N,k)}+K_{\mathcal{N}(N,k)-1}). \tag{44}$$

In view of the uncertainties (due to statistical sampling) in the numerical procedure, and the rough nature of this approximation, the results as shown in Table VII are rather pleasing.

(ii) An alternative approach is to look at a series of subspaces of $S(N,k)$, namely at $S_D^{(\nu)}(N,k)$—the space containing all states which are expressible as a sum of ν independent determinantal states. Although we allow, in general, both determinants (in $S_D^{(2)}$, for example) to belong to different representations, we nevertheless try to span the space, or integrate over it, by considering only the same representation, and going through the entire reduction and enumeration procedure as before.

Thus, we first find the number $\bar{\nu}$, which is necessary for $S_D^{(\bar{\nu})}$ to span the entire space. This is

$$\bar{\nu}(N,k)=\left(\frac{B(N,k)}{k(N-k)+1}\right)_+ \text{ (the containing integer)}. \tag{45}$$

We therefore view $S(N,k)$ as $S_D^{(\bar{\nu})}(N,k)$ and reduce only the $\bar{\nu}$-C.S.D. rather than the complete C.S.D. It must be remembered, however, that the relative weights associated with each point are larger than unity.

For example, in the case $N=4$, $k=2$ (admittedly trivial), we have $\bar{\nu}=2$. The relative strengths of the (• •) and (•—•) diagrams is 0.2 and 0.8, respectively. Thus

$$\langle D\rangle_{4,2}\approx 0.8(\bullet\!-\!\bullet)+0.2(\bullet^3\quad\bullet^3)\approx 0.968\,. \tag{46}$$

Slightly different results would be obtained, had we performed a binomial averaging over the relative weights. This compares with a value of 0.96 obtained statistically.

Similarly, for $N=6$, $k=2$, we also have $\bar{\nu}=2$, and

$$\langle D\rangle_{6,2}\sim\tfrac{4}{7}(\cdot\;)+\tfrac{3}{7}\begin{pmatrix}\text{“}7\frac{1}{2}\text{”} & \text{“}7\frac{1}{2}\text{”}\\ \bullet & \bullet\end{pmatrix}\approx 0.91 \tag{47}$$

TABLE VI. Values of integrals $K_{\alpha\beta}$ for $\alpha+\beta\leq 8$.

α / β	1	2	3	4	5	6	7
1	$\frac{2\sqrt{2}}{\pi}$	$\frac{1}{2}+\frac{\pi}{8}$	$\frac{2\sqrt{2}}{\pi}$	$\frac{15}{32}+\frac{9\pi}{64}$	$\frac{92\sqrt{2}}{45\pi}$	$\frac{15}{32}+\frac{75\pi}{512}$	$\frac{52\sqrt{2}}{25\pi}$
2		$\frac{4-\sqrt{2}}{3}$	$\frac{9}{16}+\frac{3\pi}{32}$	$\frac{4-\sqrt{2}}{3}$	$\frac{5}{8}+\frac{5\pi}{64}$	$\frac{46}{35}-\frac{43\sqrt{2}}{140}$	
3			$\frac{28\sqrt{2}}{15\pi}$	$\frac{15}{32}+\frac{15\pi}{128}$	$\frac{28\sqrt{2}}{15\pi}$		
4				$\frac{96-27\sqrt{2}}{70}$			

TABLE VIII. Monomial Representations of P_3.

Rep. label	States	I	P_{12}	P_{13}	P_{23}	P_{123}	P_{132}
$N=3\ k=1\ M=1$	[(1)]	$\begin{pmatrix}1&&\\&1&\\&&1\end{pmatrix}$	$\begin{pmatrix}&1&\\1&&\\&&1\end{pmatrix}$	$\begin{pmatrix}&&1\\&1&\\1&&\end{pmatrix}$	$\begin{pmatrix}1&&\\&&1\\&1&\end{pmatrix}$	$\begin{pmatrix}&1&\\&&1\\1&&\end{pmatrix}$	$\begin{pmatrix}&&1\\1&&\\&1&\end{pmatrix}$
	[(2)]						
	[(3)]						
$N=3\ k=1\ M=2$	$\begin{bmatrix}(1)\\(2)\end{bmatrix}$	$\begin{pmatrix}1&&\\&1&\\&&1\end{pmatrix}$	$\begin{pmatrix}-1&&\\&&1\\&1&\end{pmatrix}$	$\begin{pmatrix}&&-1\\&-1&\\-1&&\end{pmatrix}$	$\begin{pmatrix}1&&\\&&-1\\&-1&\end{pmatrix}$	$\begin{pmatrix}&&1\\-1&&\\&-1&\end{pmatrix}$	$\begin{pmatrix}&&-1\\&-1&\\1&&\end{pmatrix}$
	$\begin{bmatrix}(1)\\(3)\end{bmatrix}$						
	$\begin{bmatrix}(2)\\(3)\end{bmatrix}$						
$N=3\ k=1\ M=3$	$\begin{bmatrix}(1)\\(2)\\(3)\end{bmatrix}$	(1)	(−1)	(−1)	(−1)	(1)	(1)
$N=3\ k=2\ M=1$	[(12)]	$\begin{pmatrix}1&&\\&1&\\&&1\end{pmatrix}$	$\begin{pmatrix}1&&\\&&1\\&1&\end{pmatrix}$	$\begin{pmatrix}&&1\\&1&\\1&&\end{pmatrix}$	$\begin{pmatrix}&1&\\1&&\\&&1\end{pmatrix}$	$\begin{pmatrix}&1&\\1&&\\&&1\end{pmatrix}$	$\begin{pmatrix}&&1\\1&&\\&1&\end{pmatrix}$
	[(13)]						
	[(23)]						
$N=3\ k=2\ M=2$	$\begin{bmatrix}(12)\\(13)\end{bmatrix}$	$\begin{pmatrix}1&&\\&1&\\&&1\end{pmatrix}$	$\begin{pmatrix}1&&\\&&-1\\&-1&\end{pmatrix}$	$\begin{pmatrix}&&-1\\&-1&\\-1&&\end{pmatrix}$	$\begin{pmatrix}-1&&\\&&1\\&1&\end{pmatrix}$	$\begin{pmatrix}&&-1\\&-1&\\1&&\end{pmatrix}$	$\begin{pmatrix}&&1\\-1&&\\&-1&\end{pmatrix}$
	$\begin{bmatrix}(12)\\(23)\end{bmatrix}$						
	$\begin{bmatrix}(13)\\(23)\end{bmatrix}$						
$n=3\ k=2\ M=3$	$\begin{bmatrix}(12)\\(13)\\(23)\end{bmatrix}$	(1)	(−1)	(−1)	(−1)	(1)	(1)
$N=3\ k=3\ M=1$	[(123)]	(1)	(1)	(1)	(1)	(1)	(1)

compared to a calculated value of 0.88. This procedure can be easily generalized, and it may provide valuable primary information regarding the representation of states. Note, that we have used effective, noninteger weights, by way of indicating the possibility of assigning continuous (average) values to this parameter.

APPENDIX: (Γ) OF C.S.D.'s AS LABEL OF MONOMIAL REPRESENTATIONS OF P_N

A monomial representation is one in which only the numbers 0, +1, −1 appear in the representing matrices and only one number may differ from 0 in each row or column. The "states," or entities which form the basis of the representation, are only capable of being simply transposed among themselves under the operations of the group. Clearly, when each such representation is viewed upon the real (or complex) numbers, it becomes reducible; but this does not concern us here. The fact that each C.S.D. (along with its Γ's) may serve as a label for such a representation is a consequence of the fact that it is invariant under all operations in the group P_N. By the same token, if one considers $N' > N$, one also has a basis for a representation of $P_{N'}$.

As an example, the monomial representations of P_3, obtained by considering $N=3$, $k=1, 2, 3$, and $M=1, 2, 3$ are shown in Table VIII. Under each representation label (N, k, M), we write down the basis, that is, all the possible choices of M k-sets. The dimension of the representation is clearly $B(B(N, k), M)$.

Since the enumeration of all monomial representations is a formidable task by itself (with significant logical applications), it is interesting to ask whether this procedure exhausts all of them. The answer is unfortunately negative, and is simply demonstrated by dimensional considerations. For example, the basic representation of order $N!$ cannot be generated, since all dimensions appearing must be nontrivial divisors of $N!$ Another example is the one in which all even permutations are represented by $\binom{1}{1}$ and all odd ones by $\binom{1}{1}$.

*Work supported by the European Research Office, U.S. Army through Contract No. DAJA-37-72-C-4532.

[1] I. Kelson and G. Shadmon, Ann. Phys. 63, 497 (1971).

[2] N is finite, as it is for practical applications of physical models. Various complications are introduced when N tends to infinity.

[3] We use the notation $B(N,k)$ for the binomial coefficient.

G-Hilbert bundles

Maria C. Abbati, Renzo Cirelli, and Franco Gallone

Istituto di Scienze Fisiche dell'Università, 20133 Milano, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Milano
(Received 2 August 1974)

A notion of Hilbert bundle is proposed which leads to the construction of a "big" Hilbert space H starting from a family of Hilbert spaces. For this, such a family is equipped with a suitable structure, called Borel field structure. A meaningful relationship is established between the Borel structures which can be defined on the union of the Hilbert spaces of the family and the Borel field structures with which the family can be equipped. For a topological group G, the structure of G-Hilbert bundle is defined linking in a suitable way a Hilbert bundle with actions of G. In the framework of a G-Hilbert bundle, a continuous unitary representation of G in H can be constructed. The transitive G-Hilbert bundles which are often used in the theory of induced representations of groups are shown to be a subclass of the class of the G-Hilbert bundles which are proposed in this paper.

INTRODUCTION

In this paper we deal with some topics in the theory of vector bundles.[1] In fact, we are concerned with some vector bundles in which the fibers are Hilbert spaces and actions of a group G are given. We shall call them G-Hilbert bundles.

The interest of such a study is twofold. Indeed, the mathematical structure we examine has an intrinsic relevance as it is not trivial and leads to meaningful mathematical results. Moreover, even if vector bundles do not seem very popular among physicists, this structure is perhaps one of the best suited for the discussion of some fundamental physical problems. In fact, it encompasses and intertwines in a very intuitive way geometric ideas with measure-theoretic, topological, and differential concepts, thus providing a unified set of mathematical tools for the discussion of covariance, invariance, gauge transformations, and so on.[2] Besides, the interest of vector bundles for physics is shown by the role this structure plays in some rather recent approaches to classical mechanics.[3] Finally, we notice that a possible physical interest of G-Hilbert bundles, or perhaps of some generalization of them, is to provide a framework in which to study the coupling of external and internal symmetries.

Our main concern, in the present paper, is to get a unitary representation of a group G from a G-Hilbert bundle. This procedure is very well known from the theory of induced representations when the group acts transitively on the base space.[4] However, we extend the theory to the nontransitive case since many a group of physical interest do not act transitively on the base space. Moreover, we find some conditions for the construction of G-Hilbert bundles which are easier to handle than the very definition of G-Hilbert bundle and which can be met in some definite situations. For an example of such a construction of a nontransitive G-Hilbert bundle starting from phenomenological considerations, see Ref. 5.

In Sec. 1 a notion of Hilbert bundle is proposed which leads to the construction of a "big" Hilbert space H, starting from a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces. For this, such a family is equipped with a suitable structure, which is called Borel field structure. Another structure is also found which provides a way to construct a Borel field structure and which can be more amenable to tackle than the very definition of Borel field structure. Furthermore, a relationship is examined between the Borel structures which can be defined on the union of the $H(\xi)$'s and the Borel field structures with which $\{H(\xi)\}$ $(\xi \in Z)$ can be equipped. When such a relationship holds, the elements of the Borel field structure are characterized by the Borel structure.

In Sec. 2 the structure of G-Hilbert bundle is defined for a locally compact topological group G satisfying the second axiom of countability. In this structure, a Hilbert bundle and actions of G are linked in such a way that a continuous unitary representation of G in H can be constructed. This is perhaps the main item of the present work.

In Sec. 3 another structure is examined, which turns out to be equivalent to that of G-Hilbert bundle but possibly easier to handle. In fact, it can be viewed as a means of constructing G-Hilbert bundles. Finally, the transitive G-Hilbert bundles which are often used in the theory of induced representations of groups, are shown to be particular cases of the general G-Hilbert bundles which are discussed in the present paper.

1. HILBERT BUNDLES

One could be tempted to accept as a meaningful notion of Hilbert bundle that of a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces $H(\xi)$ with an index set Z, and reject further conditions as unessential sophistications.[6] However, the best we can do in general with such a family is to construct the linear space $\Pi_{\xi \in Z} H(\xi)$ of the maps x from Z into the set $S = \cup_{\xi \in Z} H(\xi)$ that have the property $x(\xi) \in H(\xi)$ for each $\xi \in Z$, in which the addition and the scalar multiplication are defined pointwise, i.e., by $(x+y)(\xi) = x(\xi) + y(\xi)$ and $(\alpha x)(\xi) = \alpha x(\xi)$. Besides, this is not a really satisfactory outcome, as what we get is just a linear space and not a Hilbert space. In fact the Hilbert space structure of the $H(\xi)$'s could as well be absent, since only the linear space structure of the $H(\xi)$'s enters the above construction. Then, we are led to search a notion of Hilbert bundle wherein the Hilbert

space structure plays a central role. In such a notion, some conditions should indeed appear in an essential way in order to glue the $H(\xi)$'s together. Our main reference is a book of Dixmier,[7] which will be hereafter denoted simply by D. The paragraphs 1 and 2 of D, II are assumed as prerequisites to this section.

We begin with some conventions, which will be held throughout the paper. Whenever a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces is given, we always assume that Z is a standard Borel (B in the following) space. With the abridgement $d(\xi)$ for $\dim H(\xi)$, we define the set $Z_p = \{\xi \in Z;\ d(\xi) = p\}$ for any $p = 1, 2, \ldots, \infty$. For any p, we denote by $H^{(p)}$ a definite p-dimensional Hilbert space (it is not important what it is, but it is important to maintain it fixed through the discussion).

Now, we recall the definition of Borel field (BF in the following) structure, reproducing Def. 1 of D, II, 1 with "Borel" in the place of "measurable." This substitution is made consistently through the paper (see the remarks at pp. 143 and 146 of D).

Definition 1.1: A BF structure on a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces is a linear subspace σ of $\Pi_{\xi \in Z} H(\xi)$ with the following properties:

(i) $Z \ni \xi \to \|x(\xi)\| \in \mathbb{R}$ (the real line) is a B function, $\forall\, x \in \sigma$;

(ii) if $y \in \Pi_{\xi \in Z} H(\xi)$ is such that $Z \ni \xi \to (x(\xi) | y(\xi)) \in \mathbb{C}$ (the complex plane) is a B function for any $x \in \sigma$, then $y \in \sigma$;

(iii) a sequence $\{x_n\}$ of elements of σ exists such that, for each $\xi \in Z$, the closed subspace of $H(\xi)$ spanned by $\{x_n(\xi)\}$ coincides with $H(\xi)$.

Whenever a BF structure σ is given on a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces and a positive measure μ is defined on Z, then we can construct a "big" Hilbert space which is called direct integral of the $H(\xi)$'s and is denoted by ${}^{\sigma}\int_Z^{\oplus} H(\xi)\, d\mu(\xi)$. Take in fact the subset σ_μ of $\Pi_{\xi \in Z} H(\xi)$ containing those elements $y \in \Pi_{\xi \in Z} H(\xi)$ such that $Z \ni \xi \to (x(\xi) | y(\xi)) \in \mathbb{C}$ is a μ-measurable function for any $x \in \sigma$. It can be shown that σ_μ defines on $\{H(\xi)\}$ $(\xi \in Z)$ the structure of μ-measurable field of Hilbert spaces (Def. 1 of D, II, 1). Then we define ${}^{\sigma}\int_Z^{\oplus} H(\xi) \times d\mu(\xi) = {}^{\sigma_\mu}\int_Z^{\oplus} H(\xi)\, d\mu(\xi)$ (see Sec. 5 of D, II, 1). The elements of this Hilbert space are equivalence classes of elements of $\Pi_{\xi \in Z} H(\xi)$. A useful remark is that each such equivalence class contains an element of σ. We shall denote with the same symbol a vector of ${}^{\sigma}\int_Z^{\oplus} H(\xi) \times d\mu(\xi)$ and a representative element thereof and we shall always choose representatives which lie in σ.

From the above considerations it follows that a reasonable notion of Hilbert bundle is provided by a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces together with a BF structure on it. However, it may happen that a BF structure can hardly be found, also in the framework of a mathematical or physical problem in which, on the contrary, a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces naturally arises. For this reason, we shall now present a constructive approach to BF structures, hence to Hilbert bundles.

Definition 1.2: We say that a basic unitarity is given on a family $\{H(\xi)\}$ $(\xi \in Z)$ of Hilbert spaces if $d(\xi)$ is a B function (which amounts to the condition that Z_p is a B set for any p) and a family $\{U(\xi)\}$ $(\xi \in Z)$ is given in which $U(\xi)$ is a unitary isomorphism from $H(\xi)$ onto $H^{d(\xi)}$, $\forall\ \xi \in Z$.

We shall now state a theorem which connects Hilbert bundles with basic unitarities. Before, we notice in the present paper that when we consider a Hilbert space as a B space, we take on it the natural B structure defined by the inner product.

Proposition 1.1: (a) If a basic unitarity is given on $\{H(\xi)\}$ $(\xi \in Z)$, then the subset σ of $\Pi_{\xi \in Z} H(\xi)$ containing those elements x for which $Z_p \ni \xi \to U(\xi) x(\xi) \in H^{(p)}$ is a B map for any p, is a BF structure on $\{H(\xi)\}$ $(\xi \in Z)$, which is said generated by the basic unitarity.

(b) If a BF structure σ is given on $\{H(\xi)\}$ $(\xi \in Z)$, then a basic unitarity exists which generates σ.

Proof: (a) We notice that $x \in \sigma$ iff x is an element of $\Pi_{\xi \in Z} H(\xi)$ such that $Z_p \ni \xi \to (u | U(\xi) x(\xi)) \in \mathbb{C}$ is a B function, for each $u \in H^{(p)}$ and any p.[8] Then, we can easily show that σ fulfills the conditions of Def. 1.1. First we notice that σ is obviously a linear subspace of $\Pi_{\xi \in Z} H(\xi)$. Next, for any $x \in \sigma$, $Z_p \ni \xi \to \|x(\xi)\| \in \mathbb{R}$ is a B function for any p, since $\|x(\xi)\|^2 = \sum_n |(u_n^{(p)} | U(\xi) x(\xi))|^2$ for an orthonormal basis $\{u_n^{(p)}\}$ in $H^{(p)}$. Hence $\|x(\xi)\|$ is a B function on Z, since the Z_p's are B sets whose (countable) union is Z. Let, moreover, $y \in \Pi_{\xi \in Z} H(\xi)$ be such that $Z \ni \xi \to (x(\xi) | y(\xi)) \in \mathbb{C}$ is a B function, $\forall x \in \sigma$. For any p take an element $x^{(p)}$ of $H^{(p)}$ and construct the map $x : Z \to \cup_{\xi \in Z} H(\xi)$, $x(\xi) = U(\xi)^{-1} x^{(d(\xi))}$, which is obviously an element of σ. Then $(x^{(d(\xi))} | U(\xi) y(\xi))$ is a B function on Z, whence $(x^{(p)} | U(\xi) y(\xi))$ is a B function on Z_p for each $x^{(p)} \in H^{(p)}$ and for any p, which amounts to $y \in \sigma$. Finally, for any p let $\{x_n^{(p)}\}$ be a sequence of elements of $H^{(p)}$ which span $H^{(p)}$. Then, for any n the map $x_n : Z \to \cup_{\xi \in Z} H(\xi)$, $x_n(\xi) = U(\xi)^{-1} x_n^{(d(\xi))}$ is obviously an element of σ and $\{x_n(\xi)\}$ span $H(\xi)$ for each $\xi \in Z$, as $U(\xi)$ is a unitary isomorphism.

(b) The proof can be easily performed taking into account Prop. 1 and the proof of Prop. 3 in D, II, 1.

QED

This proposition shows that it is equivalent to have, on $\{H(\xi)\}$ $(\xi \in Z)$, a BF structure or a basic unitarity. In particular, part (a) provides a way to construct a BF structure on $\{H(\xi)\}$ $(\xi \in Z)$, i.e., a Hilbert bundle, whenever a basic unitarity is given. We point out that, in definite problems, a hint for the construction of basic unitarities may be furnished directly by the possible presence of "natural" isomorphisms among those $H(\xi)$'s which have the same dimension. This can be a way to solve the problem of the construction of BF structures.

It is obvious that a basic unitarity which generates a BF structure does not need to be unique. Anyway, it is quite natural to state the following definition, since the objects in which we are primarily interested are the BF structures.

Definition 1.3: Two basic unitarities are said equivalent if they generate the same BF structure.

The equivalence of two basic unitarities can be

characterized in an intrinsic way, namely without referring to the BF structures they generate. Previously, for a basic unitarity defined by the family $\{U(\xi)\}$ $(\xi \in Z)$ of unitary isomorphisms we construct the map

$$\psi_{(U)} : \cup_p (Z_p \times H^{(p)}) \to S, \quad \psi_{(U)}(\xi, u) = U(\xi)^{-1} u,$$

which is obviously a bijection of $\cup_p (Z_p \times H^{(p)})$ onto S. In what follows, on $Z_p \times H^{(p)}$ we assume the product B structure and on $\cup_p (Z_p \times H^{(p)})$ the sum B structure. When we consider the set $U(H)$ of the unitary operators on a Hilbert space H as a B space, we take on it the structure generated by the strong operator topology. Therefore a map $f: X \to U(H)$ from a B space X into $U(H)$ is a B map iff $f_u : X \to H$, $f_u(x) = f(x)u$ is a B map for each $u \in H$. Now comes the announced theorem on the equivalence of two basic unitarities.

Proposition 1.2: If two basic unitarities are defined on $\{H(\xi)\}$ $(\xi \in Z)$ by the families $\{U(\xi)\}$ $(\xi \in Z)$ and $\{U'_{(\xi)}\}$ $(\xi \in Z)$ of unitary isomorphisms, the following statements are equivalent:

(a) the two basic unitarities are equivalent;

(b) $Z_p \ni \xi \to U'(\xi) U(\xi)^{-1} \in U(H^{(p)})$ is a B map for any p;

(c) $\psi_{(U')}{}^{-1} \circ \psi_{(U)}$ is a B automorphism of $\cup_p (Z_p \times H^{(p)})$.

Proof: (a) => (b). Take an orthonormal basis $\{u_n{}^{(p)}\}$ in $H^{(p)}$ for any p. Define, for any index n, the elements u_n and u'_n of $\Pi_{\xi \in Z} H(\xi)$ setting

$$u_n(\xi) = \begin{cases} U(\xi)^{-1} u_n{}^{(d(\xi))} & \text{if } n \leq d(\xi), \\ 0 & \text{if } n > d(\xi), \end{cases}$$

and analogously for u'_n with $U'(\xi)$ in the place of $U(\xi)$. For any n and m, $(u'_n(\xi) | u_m(\xi))$ is a B function on Z as u'_n and u_m belong to the same BF structure. Therefore, for any p and for any v and u in $H^{(p)}$ the function

$$(v | U'(\xi) U(\xi)^{-1} u) = \sum_{n,m} (v | u_n{}^{(p)}) \, (u'_n(\xi) | u_m(\xi)) \, (u_m{}^{(p)} | u)$$

is a B function on Z_p, whence $Z_p \ni \xi \to U'(\xi) U(\xi)^{-1} u \in H^{(p)}$ is a B function for each $u \in H^{(p)}$.[8] Then, (b) follows by the remark preceeding the proposition.

(b) $\Rightarrow$ (c). We have that $Z_p \times H^{(p)} \ni (\xi, u) \to (U'(\xi) U(\xi)^{-1}, u) \in U(H^{(p)}) \times H^{(p)}$ is a B map with respect to the product B structure on $U(H^{(p)}) \times H^{(p)}$. Therefore $Z_p \times H^{(p)} \ni (\xi, u) \to U'(\xi) U(\xi)^{-1} u \in H^{(p)}$ is a B map, whence the same holds for $Z_p \times H^{(p)} \ni (\xi, u) \to (\xi, U'(\xi) U(\xi)^{-1} u) \in Z_p \times H^{(p)}$. This last map is nothing else than $\psi_{(U')}{}^{-1} \circ \psi_{(U)}$ restricted to $Z_p \times H^{(p)}$. Moreover, in (b) $U'(\xi)$ and $U(\xi)$ may be interchanged as $U(H^{(p)})$ is a topological group with respect to the strong operator topology.[9] Then (c) follows.

(c) => (a). Projecting onto $H^{(p)}$ we have that, for any p, $Z_p \times H^{(p)} \ni (\xi, u) \to U'(\xi) U(\xi)^{-1} u \in H^{(p)}$ is a B map. If x is an element of the BF structure generated by the basic unitarity defined by $\{U(\xi)\}$ $(\xi \in Z)$, then, for any p, $Z_p \ni \xi \to U(\xi) x(\xi) \in H^{(p)}$ is a B map, whence the same holds for $Z_p \ni \xi \to (\xi, U(\xi) x(\xi)) \in Z_p \times H^{(p)}$ and for

$$Z_p \ni \xi \to U'(\xi) U(\xi)^{-1} \, (U(\xi) x(\xi)) = U'(\xi) x(\xi) \in H^{(p)}.$$

Therefore, x results into an element of the BF structure generated by the basic unitarity defined by $\{U'(\xi)\}$ $(\xi \in Z)$. As in (c) we can interchange U and U', (a) is proved. QED

In some definite case, it may happen to have a B structure on $S = \cup_{\xi \in Z} H(\xi)$ and a BF structure on $\{H(\xi)\}$ $(\xi \in Z)$. Then, a somewhat natural question is whether a relation can be defined between the two classes of BF and B structures such that, whenever it holds, the elements of the BF structure on $\{H(\xi)\}$ $(\xi \in Z)$ can be characterized by means of the B structure on $S = \cup_{\xi \in Z} H(\xi)$. We propose here an answer to this problem.

Definition 1.4: A BF structure σ on $\{H(\xi)\}$ $(\xi \in Z)$ is said to be compatible with a B structure on $\cup_{\xi \in Z} H(\xi)$ if, for a basic unitarity which generates σ, the map $\psi_{(U)}$ introduced before Prop. 1.2 is a B isomorphism.

We remark that the above definition makes sense because of the equivalence of the statements (a) and (c) of Prop. 1.2. The usefulness of the concept of compatibility now introduced is brought into evidence by the following theorem.

Proposition 1.3: Let a BF structure σ be given on $\{H(\xi)\}$ $(\xi \in Z)$ and let it be compatible with a B structure which is given on $S = \cup_{\xi \in Z} H(\xi)$. Then, for an element x of $\Pi_{\xi \in Z} H(\xi)$, $x \in \sigma$ iff x is a B map from Z into S.

Proof: Let $\{U(\xi)\}$ $(\xi \in Z)$ be the family of unitary isomorphisms of a basic unitarity which generates σ. Then $\psi_{(U)}$ is a B isomorphism and, for an element x of $\Pi_{\xi \in Z} H(\xi)$, x is a B map from Z into S iff $\psi_{(U)}^{-1} \circ x$ is a B map from Z into $\cup_p (Z_p \times H^{(p)})$, and this happens iff $Z_p \ni \xi \to U(\xi) x(\xi) \in H^{(p)}$ is a B map for any p, namely iff $x \in \sigma$. QED

We point out that whenever a BF structure σ is given on $\{H(\xi)\}$ $(\xi \in Z)$, then one and only one B structure exists on $S = \cup_{\xi \in Z} H(\xi)$ with which σ is compatible. It is in fact the B structure which can be transferred from $\cup_p (Z_p \times H^{(p)})$ onto S by means of the map $\psi_{(U)}$ defined by a basic unitarity which generates σ. If conversely a B structure is given on S, then at most one BF structure σ exists on $\{H(\xi)\}$ $(\xi \in Z)$ for which the condition of compatibility holds. If, in fact, σ' is another such B structure, then σ' is forced to coincide with σ by the (c) => (a) part of Prop. 1.2.

Finally, it can be easily shown that if a BF structure on $\{H(\xi)\}$ $(\xi \in Z)$ is compatible with a B structure on S, then for each $\xi \in Z$ the B structure induced on $H(\xi)$ by S results into the natural B structure defined on $H(\xi)$ by the inner product. The proof is left to the reader.

2. G-HILBERT BUNDLES AND REPRESENTATIONS OF GROUPS

The subject of this section is a structure which includes a Hilbert bundle together with actions of a group in such a way that a unitary representation of the group can be constructed. By G we shall denote a locally compact topological group satisfying the second axiom of countability.

Definition 2.1: A G-Hilbert bundle (G-Hb in the following) is a family $\{H(\xi)\}$ $(\xi \in Z)$ equipped in the fol-

lowing way: (α) Z is a B G-space, namely a homomorphism t of G into the group of the permutations of Z is given such that $G\times Z \ni (g,\zeta)\to t(g)\zeta \in Z$ is a B map. (β) On $S=\cup_{\zeta\in Z} H(\zeta)$ a B structure is given, and it is a G-space, namely a homomorphism T of G into the group of the permutations of S exists such that $\tau: G\times S\to S$, $\tau(g,w)=T(g)w$, is a B map.

(γ) The map $\pi: S\to Z$, $\pi(w)=\zeta$ if $w\in H(\zeta)$, intertwines T and t, namely $t(g)\circ\pi=\pi\circ T(g)$ for each $g\in G$, and $T(g)$ restricted to $H(\zeta)$ is a unitary isomorphism of $H(\zeta)$ onto $H(t(g)\zeta)$ for each $g\in G$ and $\zeta\in Z$.

(δ) On $\{H(\zeta)\}$ ($\zeta\in Z$) a BF structure σ is given which is compatible with the B structure which is given on S.

We notice that condition (δ) may be interpreted as a triviality condition, as in its essentials it states that the B space S is B isomorphic to $\cup_p (Z_p\times H^{(p)})$ through a family of unitary operators. We notice also that, as a consequence of condition (δ), the map π results in a B map. In fact, $\psi_{(U)}$ is a B isomorphism for a basic unitarity which generates σ and $\pi\circ\psi_{(U)}$ is a B map, since it is the projection $\cup_p (Z_p\times H^{(p)})\ni(\zeta,u)\to\zeta\in Z$. Hence, π results in the product $(\pi\circ\psi_{(U)})\circ\psi_{(U)}^{-1}$ of two B maps.

The next theorem shows that a continuous unitary representation (c.u.r. in the following) of G can be constructed in the framework of a G-Hb. In fact, a Hilbert space H will be constructed as well as a c.u.r. of G in H, namely a continuous homomorphism of G into the group $U(H)$ of the unitary operators on H endowed with the strong operator topology.

Proposition 2.1: (a) In a G-Hb, let a positive measure μ be given on Z and let μ be quasi-invariant with respect to the action t of G on Z. Denote by H the Hilbert space $\int_Z^{\oplus} H(\zeta)\,d\mu(\zeta)$ introduced after Def. 1.1, by μ_g the measure defined on Z for each $g\in G$ as $\mu_g(\Delta)=\mu(t(g^{-1})(\Delta))$, for each B set Δ of Z, and by λ_g a representative of the Radon–Nikodym derivative $d\mu_g/d\mu$. Then there exists exactly one c.u.r. V of G in H such that

$$(V(g)x)(\zeta)=(\lambda_g(\zeta))^{1/2}T(g)\,x(t(g^{-1})\zeta),\ \forall\ x\in H,\ \zeta\in Z,$$

for ν-almost all $g\in G$, where ν stands for a Haar measure on G.

(b) If μ is invariant, then the map V_g of H into itself defined by $(V_g x)(\zeta)=T(g)x(t(g^{-1})\zeta)$ for each $x\in H$ and $\zeta\in Z$ is a unitary operator such that $V: G\to U(H)$, $V(g)=V_g$ is a c.u.r. of G in H.

Proof: (a) Take an element x of σ. Then, taking into account the conditions (α) and (δ) in Def. 2.1 and Prop. 1.3, $G\times Z\ni(g,\zeta)\to x(t(g^{-1})\zeta)\in S$ is a B map, whence the same holds for $G\times Z\ni(g,\zeta)\to(g,x(t(g^{-1})\zeta))\in G\times S$. Therefore, $G\times Z\ni(g,\zeta)\to T(g)x(t(g^{-1})\zeta)\in S$ is a B map as a consequence of condition (β) in Def. 2.1. Next, let $G\times Z\ni(g,\zeta)\to\lambda_g(\zeta)\in\mathbb{C}$ be a nonnegative B function for which a B set G' exists in G such that $\nu(G-G')=0$ and $Z\ni\zeta\to\lambda_g(\zeta)\in\mathbb{C}$ is a representative of $d\mu_g/d\mu$, $\forall\ g\in G'$. The existence of such a function is asserted by Theorem 8.10 of Ref. 9. Then $G\times Z\ni(g,\zeta)\to(\lambda_g(\zeta))^{1/2}\,T(g)\times x(t(g^{-1})\zeta)\in S$ is a B map, since it results from the composition of

$$G\times Z\ni(g,\zeta)\to(\lambda_g(\zeta))^{1/2},\ T(g)x(t(g^{-1})\zeta)\in\mathbb{C}\times S$$

with $\mathbb{C}\times S\ni(\alpha,w)\to\alpha w\in S$, which is obviously a B map since it is translated into $\mathbb{C}\times\cup_p(Z_p\times H^{(p)})\ni(\alpha,\zeta,u)\to(\zeta,\alpha u)\in\cup_p(Z_p\times H^{(p)})$ by the B isomorphism $\psi(U)$ defined by the family $\{U(\zeta)\}$ ($\zeta\in Z$) of unitary isomorphisms of a basic unitarity which generates σ.

Take now an element x of H and an element g of G'. The map $Z\ni\zeta\to(\lambda_g(\zeta))^{1/2}T(g)x(t(g^{-1})\zeta)\in S$ is an element of $\Pi_{\zeta\in Z}H(\zeta)$ as a consequence of condition (γ) in Def. 2.1. Indeed, it is an element of σ by the observations of the preceding paragraph. Moreover, it defines an element of H since square integrability follows from

$$\int_Z \|(\lambda_g(\zeta))^{1/2}T(g)\,x(t(g^{-1})\zeta)\|^2\,d\mu(\zeta)$$
$$=\int_Z \|x(t(g^{-1})\zeta)\|^2\,\lambda_g(\zeta)\,d\mu(\zeta)$$
$$=\int_Z \|x(\zeta)\|^2\,d\mu(\zeta)=\|x\|^2,$$

which is a consequence of condition (γ) in Def. 2.1. Then, for each $g\in G'$ we can define $L_g: H\to H$, $(L_g x)(\zeta)=(\lambda_g(\zeta))^{1/2}T(g)x(t(g^{-1})\zeta)$, which is trivially shown to be a linear isometric operator on H. Now, we want to show that it is in fact a unitary operator. Previously, we notice that if ν is, for instance, a left Haar measure, then $\nu((G-G')^{-1})=\nu(G-(G')^{-1})$ is a right Haar measure of $G-G'$, whence $\nu(G-(G')^{-1})=0$ and G' may be replaced by $G'\cap(G')^{-1}$ in the above discussion. Therefore, we can assume G' invariant with respect to inversion and arrive at the conclusion that, for each $g\in G'$ and $x\in H$, the map $Z\ni\zeta\to(\lambda_{g^{-1}}(\zeta))^{1/2}T(g^{-1})x(t(g)\zeta)\in S$ defines an element of H. Applying L_g to this element, we get

$$[\lambda_g(\zeta)\,\lambda_{g^{-1}}(t(g^{-1})\zeta)]^{1/2}\ T(g)T(g^{-1})x(\zeta),$$

which equals $x(\zeta)$ μ-a.e. since $\lambda_g(\zeta)\lambda_{g^{-1}}(t(g^{-1})\zeta)$ is a representative of the Radon–Nikodym derivative of μ with respect to itself by Lemma 8.9 of Ref. 9. This argument establishes that L_g is onto, and hence that it is a unitary operator.

For $g\notin G'$, we can define L_g to be, for instance, the unit operator on H. We want now to show that, for two elements x and y of H, the function

$$\varphi^{(x,y)}: G\times Z\to\mathbb{C},\quad \varphi^{(x,y)}(g,\zeta)=((L_g x)(\zeta)\,|\,y(\zeta))$$

is $\nu\times\mu$-measurable. First we notice that

$$G\times Z\ni(g,\zeta)\to(\lambda_g(\zeta))^{1/2}T(g)x(t(g^{-1})\zeta)\in S$$

is a B map, as it has been previously shown, and that $G\times Z\ni(g,\zeta)\to y(\zeta)\in S$

turns out easily to be a B map as a consequence of Prop. 1.3. The same obviously holds for the restrictions of these maps to $G'\times Z$ with the induced B structure. Next, composing these maps with $\psi_{(U)}^{-1}$, restricting to $G'\times Z_p$ and projecting from $Z_p\times H^{(p)}$ onto $H^{(p)}$ we get that

$$G'\times Z_p\ni(g,\zeta)\to U(\zeta)\left((\lambda_g(\zeta))^{1/2}T(g)\,x(t(g^{-1})\zeta)\right)\in H^{(p)}$$

and

$$G'\times Z_p\ni(g,\zeta)\to U(\zeta)\,y(\zeta)\in H^{(p)}$$

are B maps for any p with respect to the induced B structure on $G'\times Z_p$. Then, taking an orthonormal basis $\{u_n^{(p)}\}$ in $H^{(p)}$ and by Ref. 8, we get that

$$G' \times Z_p \ni (g,\zeta) \to \Big((\lambda_g(\zeta))^{1/2} T(g)\, x(t(g^{-1})\zeta) \,|\, y(\zeta) \Big)$$
$$= \sum_n \left(\mathcal{U}(\zeta)((\lambda_g(\zeta))^{1/2} T(g)\, x(t(g^{-1})\zeta)) \,|\, u_n^{(p)} \right)$$
$$\times (u_n^{(p)} \,|\, \mathcal{U}(\zeta)\, y(\zeta))$$

is a B map for any p. Since this map is nothing else than $\varphi^{(x,y)}$ restricted to $G' \times Z_p$, since $G' \times Z_p$ is a B subset of $G' \times Z$ and since $G' \times Z$ is the union of the countable many $G' \times Z_p$'s, we have that the restriction of $\varphi^{(x,y)}$ to $G' \times Z$ is a B map. It is then enough to notice that $\nu \times \mu(G \times Z - G' \times Z) = \nu(G - G')\,\mu(Z) = 0$ to obtain the required $\nu \times \mu$-measurability of $\varphi^{(x,y)}$.

Now, let $\tilde{\nu}$ be a finite measure on G equivalent to ν (such a finite measure exists since in a locally compact topological group satisfying the second axiom of countability a Haar measure is σ-finite) and x, y two elements of $\mathcal{H}$. Then, by the Tonelli theorem (see, for instance, Theorem 22 in Chap. 12 in Ref. 10), the integral $I = \int_G \left(\int_Z |\varphi^{(x,y)}(g,\zeta)|\, d\mu(\zeta) \right) d\tilde{\nu}(g)$ exists. Moreover, by Schwarz and Hölder inequalities,

$$\int_Z |\varphi^{(x,y)}(g,\zeta)|\, d\mu(\zeta)$$
$$\leq \left(\int_Z \|(L_g x)(\zeta)\|^2\, d\mu(\zeta) \right)^{1/2} \cdot \left(\int_Z \|y(\zeta)\|^2\, d\mu(\zeta) \right)^{1/2}$$
$$= \|x\| \cdot \|y\|,$$

whence $I \leq \tilde{\nu}(G)\|x\| \cdot \|y\|$. Then, again by the Tonelli theorem, $\varphi^{(x,y)}$ is $\tilde{\nu} \times \mu$-integrable on $G \times Z$. Therefore, by the Fubini theorem (see, for instance, Theorem 21 in Chap. 12 of Ref. 10), the map

$$G \ni g \to \int_Z \varphi^{(x,y)}(g,\zeta)\, d\mu(\zeta) = (L_g x \,|\, y) \in \mathbb{C}$$

is $\tilde{\nu}$-measurable, whence ν-measurable, for all x, $y \in \mathcal{H}$.

Finally, from Theorem 8.10 of Ref. 9 we get that $\lambda_{g_1 g_2}(\zeta) = \lambda_{g_2}(t(g_1^{-1})\zeta)\lambda_{g_1}(\zeta)$ for $\nu \times \nu \times \mu$ almost all $(g_1, g_2, \zeta) \in G \times G \times Z$. For such a triple in which moreover g_1 and g_2 are such that $g_1, g_2, g_1 g_2 \in G'$, and for $x \in \mathcal{H}$ we have

$$(L_{g_1} L_{g_2} x)(\zeta) = (\lambda_{g_1}(\zeta))^{1/2} T(g_1)\,(L_{g_2} x)(t(g_1^{-1})\zeta)$$
$$= (\lambda_{g_1}(\zeta)\lambda_{g_2}(t(g_1^{-1})\zeta))^{1/2}\, T(g_1) T(g_2)$$
$$x(t((g_1 g_2)^{-1})\zeta)$$
$$= (\lambda_{g_1 g_2}(\zeta))^{1/2}\, T(g_1 g_2)\, x\,(t((g_1 g_2)^{-1})\zeta)$$
$$= (L_{g_1 g_2} x)(\zeta).$$

Therefore, $(L_{g_1} L_{g_2} x)(\zeta) = (L_{g_1 g_2} x)(\zeta)$ μ-a.e. for $\nu \times \nu$-almost all $(g_1, g_2) \in G \times G$ and each $x \in \mathcal{H}$, whence $L_{g_1} L_{g_2} = L_{g_1 g_2}$ for $\nu \times \nu$-almost all $(g_1, g_2) \in G \times G$. The result then follows from Lemma 9.6 of Ref. 9 since $\mathcal{H}$ is separable as a consequence of the corollary in the paragraph 6 of D, II, 1.

(b) As in the part (a), we get that, for each $x \in \mathcal{H}$, $G \times Z \ni (g,\zeta) \to T(g)\, x(t(g^{-1})\zeta) \in S$ is a Borel map. From this fact two consequences follow. The first one is that for each $g \in G$ we can show [taking into account condition (γ) in Def. 2.1 and the invariance of the measure] that, for any x of $\mathcal{H}$, $Z \ni \zeta \to T(g)\, x(t(g^{-1})\zeta) \in S$ defines an element of $\mathcal{H}$, and therefore we can define in a consistent way the operator V_g, which is easily shown to be unitary. The second one is that, following very closely the pattern of part (a), we can show that $G \ni g \to (V_g x \,|\, y) \in \mathbb{C}$ is measurable for all $x, y \in \mathcal{H}$ with respect to a Haar measure on G. Moreover $V_{g_1 g_2} = V_{g_1} V_{g_2}$, $\forall$ $g_1, g_2 \in G$, can be stated by an easy computation. Then V results into a c.u.r. of G in $\mathcal{H}$.[11] QED

3. A WAY TO CONSTRUCT G-HILBERT BUNDLES

Proposition 2.1 shows that G-Hb's can be useful. However, the notion of G-Hb is somewhat cumbersome and in definite cases it cannot be so easy to construct G-Hb's. We shall now propose a way to bypass this difficulty. We begin with a definition.

Definition 3.1: A pre-G-Hb is a family $\{\mathcal{H}(\zeta)\}$ $(\zeta \in Z)$ equipped in the following way:

(a) Like (α) in Def. 2.1.

(b) $S = \cup_{\zeta \in Z} \mathcal{H}(\zeta)$ is a G-space, namely a homomorphism T of G into the group of the permutations of S exists.

(c) Like (γ) in Def. 2.1.

(d) On $\{\mathcal{H}(\zeta)\}$ $(\zeta \in Z)$ a basic unitarity is given such that if $\{\mathcal{U}(\zeta)\}$ $(\zeta \in Z)$ is the family of unitary isomorphisms associated with it, then

$$G \times Z_p \ni (g,\zeta) \to \mathcal{U}(t(g)\zeta)\, T(g)\, \mathcal{U}(\zeta)^{-1} \in \mathcal{U}(H^{(p)})$$

is a B map for any p.

The structure of pre-G-Hb may happen to be easier to handle than that of G-Hb, as condition (b) in Def. 3.1 is weaker than condition (β) in Def. 2.1 and sometimes condition (d) can be fulfilled without too much trouble. This was the case, for instance, in a previous paper,[5] and this is the case, also, when a basic unitarity can be found such that, for any p,

$$\mathcal{U}(t(g)\zeta)\, T(g)\, \mathcal{U}(\zeta)^{-1} = R^{(p)}(g), \quad \forall \quad \zeta \in Z_p, \ \forall\ g \in G,$$

where $R^{(p)}$ is a c.u.r. of G in $H^{(p)}$.

We notice also that condition (d) of Def. 3.1 turns out to be equivalent to the following one:

(d') on $\{\mathcal{H}(\zeta)\}$ $(\zeta \in Z)$ a BF structure σ is given such that

$$G \times Z \ni (g,\zeta) \to \Big((y(\zeta) \,|\, T(g) x(t(g^{-1})\zeta)) \Big) \in \mathbb{C}$$

is a B map for all $x, y \in \sigma$.

Suppose in fact that (d) holds. Then, take σ to be the BF structure generated by the basic unitarity of (d). An element x in σ is an element of $\Pi_{\zeta \in Z} \mathcal{H}(\zeta)$ such that, for any p, $Z_p \ni \zeta \to \mathcal{U}(\zeta) x(\zeta) \in H^{(p)}$ is a B map, whence, by using condition (a) in Def. 3.1,

$$G \times Z_p \ni (g,\zeta) \to \mathcal{U}(t(g^{-1})\zeta)\, x(t(g^{-1})\zeta) \in H^{(p)}$$

is a B map, and this implies that

$$G \times Z_p \ni (g,\zeta)$$
$$\to \Big(\mathcal{U}(\zeta) T(g)\, \mathcal{U}(t(g^{-1})\zeta)^{-1},\ \mathcal{U}(t(g^{-1})\zeta)\, x(t(g^{-1})\zeta) \Big)$$
$$\in \mathcal{U}(H^{(p)}) \times H^{(p)}$$

is a B map, if condition (d) is taken into account. There-

fore, since $\mathcal{U}(\mathcal{H})\times\mathcal{H}\ni(V,u)\to Vu\in\mathcal{H}$ is a B map for any Hilbert space $\mathcal{H}$, for any p

$$G\times Z_p\ni(g,\zeta)\to\mathcal{U}(\zeta)\,T(g)\,x(t(g^{-1})\zeta)\in H^{(p)}$$

results in a B map for each $x\in\sigma$, whence the same holds for

$$G\times Z_p\ni(g,\zeta)\to(\mathcal{U}(\zeta)\,y(\zeta),\ \mathcal{U}(\zeta)T(g)\,x(t(g^{-1})\zeta))\in H^{(p)}\times H^{(p)}$$

and therefore also for

$$G\times Z_p\ni(g,\zeta)\to\Big(y(\zeta)\,\Big|\,T(g)\,x(t(g^{-1})\zeta)\Big)\in\mathbb{C}$$

for all $x,y\in\sigma$. Then for σ condition (d′) holds.

Suppose conversely that (d′) holds. Then, take a basic unitarity which generates the BF structure σ of (d′) and let $\{\mathcal{U}(\zeta)\}$ $(\zeta\in Z)$ be the family of unitary isomorphisms associated with it. Particular elements of σ are the elements of $\Pi_{\zeta\in Z}\,\mathcal{H}(\zeta)$ defined by $x(\zeta)=\mathcal{U}(\zeta)^{-1}x^{(q(\zeta))}$, where $x^{(p)}$ is any element of $H^{(p)}$ for any p. Then from (d′) it follows in particular that

$$G\times Z_p\ni(g,\zeta)\to\ (v\,|\,\mathcal{U}(\zeta)T(g)\,\mathcal{U}(t(g^{-1})\zeta)^{-1}u)\ \in\mathbb{C}$$

is a B map for any p and all $v,\ u\in H^{(p)}$. By the still used standard arguments condition (d) then follows.

The following theorem shows that, also if pre-G-Hb's may be more amenable then G-Hb's, these two structures are in fact equivalent.

Proposition 3.1: (a) If a pre-G-Hb is given, then a unique B structure can be constructed on S such that the conditions of Def. 2.1 hold with respect to the BF structure generated on $\{\mathcal{H}(\zeta)\}$ $(\zeta\in Z)$ by the basic unitarity of condition (d) in Def. 3.1.

(b) If a G-Hb is given, then the conditions of Def. 3.1 hold with respect to a basic unitarity which generates the BF structure of condition (δ) in Def. 2.1.

Proof: (a) By means of the basic unitarity of condition (d) in Def. 3.1 we construct the map

$$\psi_{(\mathcal{U})}:\bigcup_p\,(Z_p\times H^{(p)})\to S,\quad\psi_{(\mathcal{U})}(\zeta,u)=\mathcal{U}(\zeta)^{-1}u,$$

which is a bijection of $\cup_p\,(Z_p\times H^{(p)})$ onto S. Then, we can transport the B structure of $\cup_p\,(Z_p\times H^{(p)})$ onto S. Take on S this B structure and on $\{\mathcal{H}(\zeta)\}$ $(\zeta\in Z)$ the BF structure generated by the basic unitarity. Then condition (δ) of Def. 2.1 obviously holds. To perform the proof of this part, it is then enough to show that the map $\tau: G\times S\to S$, $\tau(g,w)=T(g)w$, is a B map. First, we notice that τ can be decomposed into the product $\tau=\psi_{(\mathcal{U})}\circ\rho\circ\eta$, where

$$\eta: G\times S\to G\times\Big(\bigcup_p\,(Z_p\times H^{(p)})\Big),$$

$$\eta(g,w)=(g,\psi_{(\mathcal{U})}^{-1}(w))=(g,\pi(w),\ \mathcal{U}(\pi(w))w),$$

and

$$\rho: G\times\Big(\bigcup_p\,(Z_p\times H^{(p)})\Big)\to\bigcup_p\,(Z_p\times H^{(p)}),$$

$$\rho(g,\zeta,u)=\Big(t(g)\zeta,\ \mathcal{U}(t(g)\zeta)\,T(g)\,\mathcal{U}(\zeta)^{-1}u\Big),$$

as can be easily shown by a direct computation. Next, we notice that $\psi_{(\mathcal{U})}$ and η are B maps as a direct consequence of the B structure defined on S and that ρ results in a B map as a consequence of conditions (a) and (d) in Def. 3.1 together with the fact that

$$\mathcal{U}(\mathcal{H})\times\mathcal{H}\ni(V,u)\to Vu\in\mathcal{H}$$

is a B map for any Hilbert space $\mathcal{H}$. This proves that τ is a B map.

(b) The only thing to be proved is that condition (d) of Def. 3.1 holds for the family $\{\mathcal{U}(\zeta)\}$ $(\zeta\in Z)$ of unitary isomorphisms of a basic unitarity which generates the BF structure of condition (δ) in Def. 2.1. The map

$$\eta^{-1}: G\times\Big(\bigcup_p\,(Z_p\times H^{(p)})\Big)\to G\times S,$$

$$\eta^{-1}(g,\zeta,u)=(g,\psi_{(\mathcal{U})}(\zeta,u))=(g,\ \mathcal{U}(\zeta)^{-1}u),$$

is a B map as a consequence of condition (δ) in Def. 2.1, whence $\psi_{(\mathcal{U})}^{-1}\circ\tau\circ\eta^{-1}$ is a B map from $G\times(\cup_p(Z_p\times H^{(p)}))$ onto $\cup_p\,(Z_p\times H^{(p)})$ if condition (β) in Def. 2.1 is taken into account. Then, restricting $\psi_{(\mathcal{U})}^{-1}\circ\tau\circ\eta^{-1}$ to $G\times Z_p\times H^{(p)}$, projecting in the range space from $Z_p\times H^{(p)}$ onto $H^{(p)}$ and fixing an element u of $H^{(p)}$ in $G\times Z_p\times H^{(p)}$, we have that $G\times Z_p\ni(g,\zeta)\to\mathcal{U}(t(g)\zeta)\,T(g)\,\mathcal{U}(\zeta)^{-1}u\in H^{(p)}$ is a B map for any p and each $u\in H^{(p)}$. This concludes the proof, by the remark preceding Prop. 1.2. QED

Thus, the class of G-Hb's can be identified with the class of pre-G-Hb's in the definite sense that a bijection can be constructed between these two classes along the lines of Prop. 3.1. We stress that, in many circumstances, it is likely that pre-G-Hb's are easier to be found than G-Hb's. Under this respect, the relation between pre-G-Hb's and G-Hb's is in fact quite similar to that between basic unitarities and BF structures examined in Sec. 1.

A structure which is often found in the literature (see, for instance, p. 86 of Ref. 9) can be obtained making some alterations in Def. 2.1.

Definition 3.2: We call a transitive G-Hilbert bundle (TG-Hb in the following) what we get from Def. 2.1 suppressing condition (δ), maintaining conditions (α), (β), (γ), letting G act transitively on Z [namely for any two elements ζ_1 and ζ_2 of Z an element $g\subset G$ exists such that $t(g)\zeta_1=\zeta_2$], and adding the conditions that the B space S is standard, that the B structure induced on $\mathcal{H}(\zeta)$ by S coincides with the natural B structure of $\mathcal{H}(\zeta)$ for each $\zeta\in Z$, and that π is a B map.

We shall now show that the class of TG-Hb's is a subclass of the class of G-Hb's. In fact, the following theorem proves that TG-Hb's are exactly those G-Hb's in which the action of G on Z is transitive.

Proposition 3.2: (a) A G-Hb in which the G-action of condition (α) in Def. 2.1 is transitive, results into a TG-Hb.

(b) If a TG-Hb is given, then a unique BF structure on $\{\mathcal{H}(\zeta)\}$ $(\zeta\in Z)$ exists for which condition (δ) of Def. 2.1 holds, namely with which we get a G-Hb.

Proof: (a) By condition (δ) in Def. 2.1, S changes into a standard B space, as it is isomorphic to the sum B space of the countable standard B spaces $\cup_p\,(Z_p\times H^{(p)})$. The assertion about the B structure induced on $\mathcal{H}(\zeta)$

follows from the last remark of Sec. 1 and the assertion about π from the remark preceding Prop. 2.1.

(b) First, we construct a basic unitarity on $\{\mathcal{H}(\zeta)\}$ $(\zeta \in Z)$. Owing to transitivity of the action of G on Z and to condition (γ) of Def. 3.2, $d(\zeta)$ is a constant function on Z. Then, choose a point ζ_0 in Z and take $H = \mathcal{H}(\zeta_0)$ as the definite space $H^{(p)}$ of the fixed dimension p common to the Hilbert spaces of the family $\{\mathcal{H}(\zeta)\}$ $(\zeta \in Z)$. Let s be a B map from Z into G such that $s(\zeta_0)$) is the unit element of G and $t(s(\zeta))\zeta_0 = \zeta$ for each $\zeta \in Z$; such a B map exists after Theorem 8.11 of Ref. 9. Then $T(s(\zeta)^{-1})_r$, where r means restriction to $\mathcal{H}(\zeta)$, is a unitary isomorphism of $\mathcal{H}(\zeta)$ onto H for each $\zeta \in Z$, as a consequence of condition (γ) in Def. 3.2. By setting $\mathcal{U}(\zeta) = T(s(\zeta)^{-1})_r$, $\{\mathcal{U}(\zeta)\}$ $(\zeta \in Z)$ is a family of unitary isomorphisms which defines a basic unitarity on $\{\mathcal{H}(\zeta)\}$ $(\zeta \in Z)$.

Next, we want to show that for this basic unitarity the condition (d) of Def. 3.1 holds. Denoting by G_0 the stability subgroup of G at ζ_0 equipped with the induced B structure of G, we have that

$$G_0 \times H \ni (g, u) \to T(g)u \in H$$

is a B map with respect to the natural B structure on H, since it is the restriction to $G_0 \times H$ of τ which is a B map from $G \times S$ into S and since the B sets of H are B sets in S as well because $H = \pi^{-1}(\{\zeta_0\})$ is a B subset of S (as π is a B map and Z is standard) and the natural B structure of H coincides with the B structure induced on it by S. Therefore, $G_0 \ni g \to T(g)u \in H$ is a B map for each $u \in H$ and from this the same follows for $G_0 \ni g \to T(g)_r \in \mathcal{U}(H)$, where r means restriction to H. Moreover, $G \times Z \ni (g, \zeta) \to s(t(g^{-1})\zeta) \in G$ is a B map, whence the same holds for

$$G \times Z \ni (g, \zeta) \to \Big(g, s(t(g^{-1})\zeta)\Big) \in G \times G$$

and for $G \times Z \ni (g, \zeta) \to g s(t(g^{-1})\zeta) \in G$. Then

$$G \times Z \ni (g, \zeta) \to \Big(s(\zeta)^{-1},\ g s(t(g^{-1})\zeta)\Big) \in G \times G$$

is a B map, whence the same holds for

$$G \times Z \ni (g, \zeta) \to s(\zeta)^{-1} g s(t(g^{-1})\zeta) \in G.$$

As $s(\zeta)^{-1} g s(t(g^{-1})\zeta) \in G_0$ for each $\zeta \in Z$ and $g \in G$, and G_0 is a B subset of G, we get that

$$G \times Z \ni (g, \zeta) \to s(\zeta)^{-1} g\, s(t(g^{-1})\zeta) \in G_0$$

is a B map. Therefore,

$$\begin{aligned} G \times Z \ni (g, \zeta) &\to T\Big(s(\zeta)^{-1} g\, s(t(g^{-1})\zeta)\Big)_r \\ &= T(s(\zeta)^{-1})_r\, T(g)\, T\Big(s(t(g^{-1})\zeta)\Big)_r \\ &= \mathcal{U}(\zeta)\, T(g)\, \mathcal{U}(t(g^{-1})\zeta) \in \mathcal{U}(H) \end{aligned}$$

is a B map, whence condition (d) of Def. 3.1 easily follows.

Then, from Prop. 3.1 (a) and its proof we know that we have a G-Hb if we define on $\{\mathcal{H}(\zeta)\}$ $(\zeta \in Z)$ the BF structure generated by the basic unitarity now constructed, and on S the B structure which makes of the map $\psi(\mathcal{U})$ a B isomorphism. Now, we want to prove that this B structure coincides with the one existing on S after condition (β) in Def. 3.2, namely that with respect to this last B structure $\psi(\mathcal{U})$ is a B isomorphism. We have in fact that $\psi(\mathcal{U})$ results from the product of the map $Z \times H \ni (\zeta, u) \to (s(\zeta), u) \in G \times H$, which is a B map by construction, with the map $G \times H \ni (g, u) \to T(g)u \in S$, which is a B map with respect to the B structure induced by $G \times S$ on $G \times H$, as a consequence of the fact that τ is a B map after condition (β) in Def. 3.2. Moreover, the B structure induced by $G \times S$ on $G \times H$ coincides with the natural one, by the arguments used above. Then $\psi(\mathcal{U})$ is a B map. As for $\psi(\mathcal{U})^{-1}$, it is the map

$$S \ni w \to \Big(\pi(w),\ T[s(\pi(w))^{-1} w]\Big) \in Z \times H,$$

which is a B map since π is a B map by hypothesis, and $S \ni w \to T[s(\pi(w))^{-1}]w \in H$ will now be proved to be a B map. In fact, by using the fact that τ is a B map, this map would easily be shown to be a B map if S were the range space. Anyway, the same holds with H as range space since H is a B subset of S and its natural B structure coincides with the one induced on it by S, as has been proved above.

Thus, the BF structure generated by the basic unitarity defined by $\{\mathcal{U}(\zeta)\}$ $(\zeta \in Z)$ fulfills condition (δ) of Def. 2.1. Its uniticity is a direct consequence of a remark following Prop. 1.3. This concludes the proof. QED

It is worthwhile to notice that, while for a general G-Hb the c.u.r. of G which is constructed in Def. 2.1 can be given in a definite form for each $g \in G$ only if the measure μ is invariant [see the part (b) of Prop. 2.1], for a TG-Hb this is possible also if the measure is not invariant but only quasi-invariant. We have in fact that if the action of G on Z is transitive, a B function $G \times Z \ni (g, \zeta) \to \lambda_g(\zeta) \in \mathbb{C}$ exists such that $Z \ni \zeta \to \lambda_g(\zeta) \in \mathbb{C}$ is a representative of $d\mu_g/d\mu$ for each $g \in G$ and $\lambda_{g_1 g_2}(\zeta) = \lambda_{g_2}(t(g_1^{-1})\zeta)\, \lambda_{g_1}(\zeta)$ for all $g_1, g_2 \in G$ and $\zeta \in Z$.[12] Therefore, proceeding as in the part (b) of Prop. 2.1, we can define in a consistent way

$$V_g : \mathcal{H} \to \mathcal{H},$$
$$(V_g x)(\zeta) = (\lambda_g(\zeta))^{1/2}\, T(g)\, x(t(g^{-1})\zeta)$$

for each $g \in G$ and show that it is a unitary operator on $\mathcal{H}$ such that $G \ni g \to V_g \in \mathcal{U}(\mathcal{H})$ is a c.u.r. of G in $\mathcal{H}$. Obviously, this is the c.u.r. whose existence is asserted in the part (a) of Prop. 2.1.

Finally, we point out that TG-Hb's are widely spread in the literature mainly because they are closely related to the theory of induced representations of groups.[4]

[1]See, for instance, D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).
[2]See, for instance: A. Trautman, Commun. Math. Phys. **6**, 248 (1967); Rep. Math. Phys. **1**, 29 (1970).
[3]See, for instance, L.H. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Mass., 1968).
[4]D.J. Simms, *Lie Groups and Quantum Mechanics* (Springer-Verlag, Berlin, 1968); G. Warner, *Harmonic Analysis on Semi-simple Lie Groups. I* (Springer-Verlag, Berlin, 1972).
[5]M.C. Abbati, R. Cirelli, F. Gallone, J. Math. Phys. **16**, 929 (1975).
[6]By "Hilbert space" we mean here and in the sequel a complex and separable Hilbert space.
[7]J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1969).
[8]M. Reed and B. Simon, *Functional Analysis, Vol. I* (Academic, New York, 1972), Theorem IV. 22.
[9]V.S. Varadarajan, *Geometry of Quantum Theory*, Vol. *II* (Van Nostrand Reinhold, New York, 1970), Lemma 8.34.
[10]H.L. Royden, *Real Analysis* (Macmillan, New York, 1963).
[11]See p. 55 of Ref. 9 or G.W. Mackey, "The Theory of Group Representations" (Department of Mathematics, University of Chicago, 1955), p. 4.
[12]G.W. Mackey, Ann. Math. **55**, 101 (1952).

Unequal mass spinor-spinor Bethe–Salpeter equation

B. J. Brennan

Department of Physics, University of Auckland, Auckland, New Zealand
(Received 1 April 1975)

Coupled radial equations are derived for the ladder approximation Bethe–Salpeter equation describing a system of two spin-(1/2) particles of unequal masses interacting to form a bound state of total mass zero. The numerical behavior of the coupling parameter λ as a function of the mass ratio is examined for known analytical equal-mass solutions. In addition a perturbation method is employed to investigate the behavior of λ for small values of the exchange mass.

I. INTRODUCTION

In this section we briefly recapitulate certain features of the Bethe-Salpeter (BS) equation for two bound spin-$\frac{1}{2}$ particles. We examine equal mass systems in Sec. II and perturbation of the mass of the exchange boson in Sec. III. In several contexts we need to refer to a paper by Brennan and Keam[1] and a series of papers by Keam.[2] The notations and conventions of the present paper are the same as for these references.

In configuration space, the ladder approximation, Wick-rotated BS equation describing the interaction of a spin-$\frac{1}{2}$ fermion of mass m_a and a spin-$\frac{1}{2}$ antifermion of mass m_b to form a bound state of total 4-momentum $P=(\mathbf{P}, iE)$, may be written

$$[\gamma\cdot(\partial+i\mu_a P)+m_a]f(x)[\gamma\cdot(\overleftarrow{\partial}-i\mu_b P)+m_b]=-\hat{V}(R)f(x), \quad (1)$$

where x is the (Euclidean) relative coordinate, $R(=\sqrt{x^2})$ is the four-dimensional radius, $\hat{V}$ describes the interaction of the two particles, and $\mu_a+\mu_b=1$. In the particular case where the interaction is due to the exchange of a boson of mass μ,[3] the potential $\hat{V}$ takes the form [we use the label n.s.(j) to show that j is not summed up in the preceding expression]:

$$\hat{V}^j=-\lambda^j\Gamma^j(4\mu/R)K_1(\mu R), \quad \text{n.s.}(j), \quad (2)$$

where λ^j, the coupling parameter, is given by[4]

$$\lambda^j=g_a^j g_b^j/(4\pi)^2, \quad \text{n.s.}(j). \quad (3)$$

Here j assumes the values 1, 2, 4, or 5 when the exchange boson is of scalar, vector, axial vector, or pseudoscalar type, respectively. In the notation of K1, Γ^j is of form $\epsilon_j\hat{\Gamma}_j$ [n.s.(j)], where ϵ_j is +1 for $j=1,4$ and −1 for $j=2,5$. g_i^j ($i=a,b$) is the coupling constant for the interaction of particle i with the exchange boson.

When the particle of mass m_b is a fermion rather than an antifermion, the modified BS amplitude $f^c(p)=f(p)C^{-1}$, where C is the charge conjugation matrix, satisfies an equation of the same form as Eq. (1). The potential due to boson exchange is the same, with the exception that $\Gamma^2=+\hat{\Gamma}_2$.[5]

In relative momentum space, the transform of Eq. (1) may be written

$$[\gamma\cdot(p+\mu_a P)-im_a]f(p)[\gamma\cdot(p-\mu_b P)-im_b]$$
$$=(2\pi)^{-4}\int d^4k\hat{W}(|p-k|)f(k), \quad (4)$$

where

$$\hat{W}(K)=\int d^4x e^{ikx}\hat{V}(R), \quad (5)$$

and

$$K^2=k^2.$$

For the exchange of a boson of "type" j,

$$\hat{W}(K)=-\lambda^j(4\pi)^2(K^2+\mu^2)^{-1}\Gamma^j, \quad \text{n.s.}(j). \quad (6)$$

Throughout this paper we consider the BS equation for the centre of mass system with E vanishing; viz $P_\mu=0$.

II. UNEQUAL MASS SYSTEMS

A. Introduction

Unequal mass systems have not previously been considered in any detail. Brennan and Keam[B1] have shown that the coupling parameter is an even function of the mass difference, m_b-m_a. Keam [K4] developed a perturbation theory and applied this to an SV sector solution, giving the approximate behavior of λ for small values of the mass difference, while in Ref. K6 the same author examined mass symmetries of the equation and λ.

In Sec. IIB we derive the coupled radial equations for unequal mass systems, and in Sec. IIC and the appendix we examine the numerical behavior of λ as a function of m_b/m_a for the known analytic equal mass solutions of Keam[K4] and Kummer.[6] The latter task has been performed using numerical methods and, for Kummer's solutions, by the use of perturbation theory.

B. Reduction of the equation

We consider the configuration form of the BS equation [Eq. (1)] with $P_\mu=0$.[7] In this case the set

$$\{\beta_0^{-1}\hat{V}, \alpha, \beta^2, \mathbf{J}^2, J_z, \overline{P}, \overline{C}_1\} \quad (7)$$

is a commuting set of operators. Here β_0 is the BS operator of the left member of Eq. (1); $\alpha, \beta^2, \mathbf{J}^2, J_z$ are the $O(4)$ operators constructed from the angular momentum operators $M_{\mu\nu}$[K1]; $\overline{P}$ is the parity operator[K2,8]; and $\overline{C}_1$ is the generalized charge parity operator.[K6] Thus we may express the BS amplitude f as a simultaneous eigenfunction of these operators. For convenience we shall consider eigenfunctions of $\overline{C}_1$ later and for the moment consider eigenfunctions of the remaining operators, viz, $f^{\overline{P}}{}_{\alpha\beta}{}^{2}{}_{Jm}$. Four distinct classes exist, as follow. We use the notation of K4 for kets, that is, we write angular kets as $|\Gamma_i(l,s^+)j^+,(l,s^-)j^-;Jm\rangle$, though for brevity we omit the quantum numbers J and m. The

angular momentum operators $\mathbf{S}^{\pm}, \mathbf{L}^{\pm}, \mathbf{J}^{\pm}$ are defined in K1. Parity eigenvalues in the following refer to fermion–antifermion systems.[8,9]

Class 1: $\alpha = 4j(j+1),\ \beta^2 = 0,\ \bar{P} = (-1)^J$

$$\begin{aligned} f_1 = {} & f_S \left| \Gamma_1(j,0)j,\ (j,0)j \right\rangle \\ & + f_{V_1} \left| \Gamma_2(j+\tfrac{1}{2},\tfrac{1}{2})j,\ (j+\tfrac{1}{2},\tfrac{1}{2})j \right\rangle \\ & + f_{V_2} \left| \Gamma_2(j-\tfrac{1}{2},\tfrac{1}{2})j,\ (j-\tfrac{1}{2},\tfrac{1}{2})j \right\rangle \\ & + f_{T^-}(1/\sqrt{2})\left(\left| \Gamma_3(j,1)j,\ (j,0)j \right\rangle - \left| \Gamma_3(j,0)j,\ (j,1)j \right\rangle \right). \end{aligned} \tag{8}$$

Class 2: $\alpha = 4j(j+1),\ \beta^2 = 0,\ \bar{P} = (-1)^{J+1}$

$$\begin{aligned} f_2 = {} & f_{T^+}(1/\sqrt{2})\left(\left| \Gamma_3(j,1)j,\ (j,0)j \right\rangle + \left| \Gamma_3(j,0)j,\ (j,1)j \right\rangle \right) \\ & + f_{A_1} \left| \Gamma_4(j+\tfrac{1}{2},\tfrac{1}{2})j,\ (j+\tfrac{1}{2},\tfrac{1}{2})j \right\rangle \\ & + f_{A_2} \left| \Gamma_4(j-\tfrac{1}{2},\tfrac{1}{2})j,\ (j-\tfrac{1}{2},\tfrac{1}{2})j \right\rangle \\ & + f_P \left| \Gamma_5(j,0)j,\ (j,0)j \right\rangle. \end{aligned} \tag{9}$$

Class 3: $\alpha = \beta^2 = 4(j+1)^2,\ \bar{P} = (-1)^J$

$$f_3 = (1/\sqrt{2})(f_{jj+1Jm} + f_{j+1jJm}), \tag{10}$$

where

$$\begin{aligned} f_{jj+1Jm} = {} & f_V \left| \Gamma_2(j+\tfrac{1}{2},\tfrac{1}{2})j,\ (j+\tfrac{1}{2},\tfrac{1}{2})j+1 \right\rangle \\ & + f_{T_1} \left| \Gamma_3(j+1,1)j,\ (j+1,0)j+1 \right\rangle \\ & + f_{T_2} \left| \Gamma_3(j,0)j,\ (j,1)j+1 \right\rangle \\ & + f_A \left| \Gamma_4(j+\tfrac{1}{2},\tfrac{1}{2})j,\ (j+\tfrac{1}{2},\tfrac{1}{2})j+1 \right\rangle \end{aligned} \tag{11}$$

and

$$\begin{aligned} f_{j+1jJm} = {} & -f_V \left| \Gamma_2(j+\tfrac{1}{2},\tfrac{1}{2})j+1,\ (j+\tfrac{1}{2},\tfrac{1}{2})j \right\rangle \\ & + f_{T_1} \left| \Gamma_3(j+1,0)j+1,\ (j+1,1)j \right\rangle \\ & + f_{T_2} \left| \Gamma_3(j,1)j+1,\ (j,0)j \right\rangle \\ & + f_A \left| \Gamma_4(j+\tfrac{1}{2},\tfrac{1}{2})j+1,\ (j+\tfrac{1}{2},\tfrac{1}{2})j \right\rangle. \end{aligned} \tag{12}$$

Class 4: $\alpha = \beta^2 = 4(j+1)^2,\ \bar{P} = (-1)^{J+1}$

$$f_4 = (1/\sqrt{2})(f_{jj+1Jm} - f_{j+1jJm}). \tag{13}$$

We consider potentials of the type

$$\hat{V} = \sum_j \hat{V}^j(R)\hat{\Gamma}_j \quad (j = 1,2,4,5) \tag{14}$$

and define

$$\bar{V}_j = \sum_i c_{ij} \hat{V}^i(R), \tag{15}$$

where the c_{ij} are defined by

$$\hat{\Gamma}_i \Gamma_j = c_{ij} \Gamma_j, \quad \text{n.s.}(j) \tag{16}$$

[cf. K1, Eq. (16) and Table I].

We also define

$$D^{\pm}_{\alpha} = \frac{\partial}{\partial R} \pm \frac{2(j+\alpha)}{R}, \tag{17}$$

as in K1, whilst

$$m = \tfrac{1}{2}(m_a + m_b), \quad \Delta = \tfrac{1}{2}(m_a - m_b). \tag{18}$$

The coupled radial equations obtained on substitution in Eq. (1) are then as follows:

Class 1:

$$D^+_{3/2} D^-_0 s + 2m[D^+_{3/2} v_1 + D^-_{-1/2} v_2] = -[m^2 - \Delta^2 + \bar{V}_1]s, \tag{19a}$$

$$\begin{aligned} & D^-_0[D^+_{3/2} v_1 + 2(j+1)D^-_{-1/2} v_2 + 2(j+1)ms + 2i\Delta\bar{w}] \\ & \quad = -(2j+1)[m^2 - \Delta^2 + \bar{V}_2]v_1, \end{aligned} \tag{19b}$$

$$\begin{aligned} & D^+_1[-D^-_{-1/2} v_2 + 2jD^+_{3/2} v_1 + 2jms - 2i\Delta\bar{w} \\ & \quad = -(2j+1)[m^2 - \Delta^2 + \bar{V}_2]v_2, \end{aligned} \tag{19c}$$

$$-D^+_{3/2} D^-_0 \bar{w} + 2i\Delta[-jD^+_{3/2} v_1 + (j+1)D^-_{-1/2} v_2] = -[m^2 - \Delta^2 + \bar{V}_3]\bar{w}, \tag{19d}$$

where

$$s = (2j+1)^{1/2} f_s, \quad v_1 = -(j+1)^{1/2} f_{V_1}, \quad v_2 = j^{1/2} f_{V_2}, \quad \bar{w} = [j(j+1)(2j+1)]^{1/2} f_{T^-}. \tag{20}$$

Class 2:

$$D^+_{3/2} D^-_0 w + 2m[-jD^+_{3/2} a_1 + (j+1)D^-_{-1/2} a_2] = -[m^2 - \Delta^2 + \bar{V}_3]w, \tag{21a}$$

$$\begin{aligned} & D^-_0[-D^+_{3/2} a_1 - 2(j+1)D^-_{-1/2} a_2 - 2mw - 2i(j+1)\Delta p] \\ & \quad = -(2j+1)[m^2 - \Delta^2 + \bar{V}_4]a_1, \end{aligned} \tag{21b}$$

$$\begin{aligned} & D^+_1[D^-_{-1/2} a_2 - 2jD^+_{3/2} a_1 + 2mw - 2ij\Delta p] \\ & \quad = -(2j+1)[m^2 - \Delta^2 + \bar{V}_4]a_2, \end{aligned} \tag{21c}$$

$$-D^+_{3/2} D^-_0 p + 2i\Delta[D^+_{3/2} a_1 + D^-_{-1/2} a_2] = -[m^2 - \Delta^2 + \bar{V}_5]p, \tag{21d}$$

where

$$w = [j(j+1)(2j+1)]^{1/2} f_{T^+}, \quad a_1 = -(j+1)^{1/2} f_{A_1}, \quad a_2 = j^{1/2} f_{A_2}, \quad p = (2j+1)^{1/2} f p. \tag{22}$$

Class 3 & 4:

$$D^+_2 D^-_{1/2} v + i\Delta[D^+_2 t_1 - D^-_0 t_2] = -[m^2 - \Delta^2 + \bar{V}_2]v, \tag{23a}$$

$$D^-_{-1/2}[D^-_0 t_2 + 2ma + 2i\Delta v] = -[m^2 - \Delta^2 + \bar{V}_3]t_1, \tag{23b}$$

$$D^+_{3/2}[D^+_2 t_1 + 2ma - 2i\Delta v] = -[m^2 - \Delta^2 + \bar{V}_3]t_2, \tag{23c}$$

$$D^+_2 D^-_{1/2} a + m[D^+_2 t_1 + D^-_0 t_2] = -[m^2 - \Delta^2 + \bar{V}_4]a, \tag{23d}$$

where

$$v = f_V, \quad t_1 = \sqrt{2} f_{T_1}, \quad t_2 = \sqrt{2} f_{T_2}, \quad a = f_A. \tag{24}$$

For each of the above classes, eigenfunctions of $\bar{C}_1$ may be extracted from f by expressing each radial function as

$$g(R) = g^+(R, \Delta^2) + \Delta g^-(R, \Delta^2). \tag{25}$$

Separating odd and even functions of Δ in the equation sets (19), (21), and (23) yields six sets of equations involving the radial functions tabulated in Table I.

For the particular case where the potential is due to the exchange of one type of particle only, certain symmetries in the equation sets (19), (21), and (23) are apparent.

Some of these arise from the fact that the operator T^b_5, where

$$T^b_5 f(p; m_a, m_b) = f(p; m_a, -m_b)\gamma_5 = f^b(p)\gamma_5, \tag{26}$$

TABLE I. Grouping of radial functions according to modified charge parity eigenvalues.

$\bar{C}_1$	Class 1	Class 2	Classes 3 and 4
$(-1)^{2j}$	$s^+, v_1^+, v_2^+, \bar{w}^-$	w^-, a_1^-, a_2^-, p^+	v^+, t_1^-, t_2^-, a^-
$(-1)^{2j+1}$	s^-, v_1^-, v_2^-, w^+	w^+, a_1^+, a_2^+, p^-	v^-, t_1^+, t_2^+, a^+

commutes with $\beta_0^{-1}\hat{V}$. Using the relation[K6]

$$\lambda^j(m_a, -m_b) = \epsilon_j' \lambda^j(m_a, m_b), \tag{27}$$

where ϵ_j' is $+1$ for $j'=2$ or 4, and -1 for $j=1$ or 5, we obtain the result that if $f(p)$ is a solution of the BS equation, then $T_5^b f(p)$ is also a solution, for the same values of λ, m_a, and m_b. We note that in the proof of Eq. (27), it is assumed that f and λ are analytic functions of m_b, and that certain integrals converge and are nonzero.

Thus $T_5^b f_1$ is a class 2 solution, with w, a_1, a_2, and p, respectively, replaced by $-\bar{w}^b$, iv_1^b, iv_2^b, and s^b (the superscript ^{b} denotes that in functions parametrically dependent on m_b, m_b has been replaced by $-m_b$). Similarly, $T_5^b f_2$ is a class 1 solution, with the converse replacement. $T_5^b f_3$ is a class 4 solution with v, t_1, t_2, and a, respectively, replaced by $-ia^b$, $-t_1^b$ $-t_2^b$, and iv^b, while $T_5^b f_4$ is a class 3 solution with the converse replacement.

However, the new solutions predicted by the action of T_5^b on known solutions in practice vanish, and this is examined briefly in Sec. II C3.

The equation set (23) is of interest for $j=2$ or 4 (V or A type exchange), where a simplification occurs. In these cases $\bar{V}_3=0$, and the terms involving v and $D_2^+ t_1 - D_0^- t_2$, are decoupled from those involving a and $D_2^+ t_1 + D_0^- t_2$. Similarly, in the momentum space transforms of Eqs. (23) [cf. Eqs. (33)], the terms involving $v(P)$ and $t_1(P) - t_2(P)$ are decoupled from those involving $a(P)$ and $t_1(P) + t_2(P)$.[10]

This decoupling is related to the properties of the operator S, defined in configuration space by

$$S_x f(x) = \frac{1}{2\pi^2}\gamma\cdot\partial\left(\int d^4y \frac{f(y)}{(x-y)^2}\right)\gamma\cdot\overleftarrow{\partial} = \frac{1}{2\pi^2}\int d^4y \frac{\gamma\cdot\partial f(y)\gamma\cdot\overleftarrow{\partial}}{(x-y)^2} \tag{28}$$

or in momentum space by its transform

$$S_P f(p) = \frac{\gamma\cdot p}{P} f(p) \frac{\gamma\cdot p}{P}, \tag{29}$$

where $P^2=p^2$. We note that, in both representations,

$$S^2 = 1 \tag{30}$$

and

$$[S_x, \beta_0] = [S_p, \mathcal{A}_0] = 0, \tag{31}$$

where $-\mathcal{A}_0$ is the momentum space BS operator of the left member of Eq. (4). In both representations, the operators $\mathcal{M}_{\mu\nu}$, $\bar{P}$, $\bar{C}_1$, and R commute with S. S_p and U (or S_x and $\hat{V}$) do not commute, though in the case of a class 3 or class 4 solution with V or or A type exchange,

$$[S_x, \hat{V}] f(x) = [S_p, U] f(p) = 0, \tag{32}$$

and hence the decoupling noted above occurs. In particular, in momentum space, the terms in the expression for $f(p)$ associated with a and $t_1 + t_2$ form an eigenfunction of S_p with eigenvalue $+1$, while those associated with v and $t_1 - t_2$ form an eigenfunction with eigenvalue -1.

C. Behavior of λ when $ma \pm mb$ for known solutions

In this section we consider the unequal mass generalizations of the equal-mass solutions of Keam and Kummer. In Sec. II C 1 we present a perturbation treatment of Kummer's solutions, and in Sec. II C 2 we present a numerical approach to the solution of the unequal mass equations. As in Sec. II B we consider the BS equation with $P_\mu = 0$, and also set $\mu = 0$.

1. Perturbation theory for Kummer's solutions

These are class 3 class 4 solutions for a fermion–fermion system with V type exchange, and $\bar{C}_1 = (-1)^{2j+1}$. If we allow negative values[4] of λ, the solutions are appropriate to both fermion–fermion and fermion–antifermion systems, for both V and A type exchange. The eigenvalues of λ are described by the two parameters j and q,[11] where $q \geq 2j+2$. The momentum space equations in this case are the transforms of Eq. (21), viz.:

$$d_2^+ d_{1/2}^-[(1-\delta^2-\sigma^2)a + i\sigma(t_1+t_2)] = -8\lambda a, \tag{33a}$$

$$\sigma^2 t_2 - (1-\delta^2)t_1 = 2i\sigma a - 2\delta\sigma v, \tag{33b}$$

$$\sigma^2 t_1 - (1-\delta^2)t_2 = 2i\sigma a + 2\delta\sigma v, \tag{33c}$$

$$d_2^+ d_{1/2}^-[(1-\delta^2+\sigma^2)v + \delta\sigma(t_1-t_2)] = 8\lambda v, \tag{33d}$$

where

$$\delta = \Delta/m, \quad \sigma = P/m, \quad \text{and} \quad d_\alpha^\pm = \frac{\partial}{\partial\sigma} \pm \frac{2(j+\alpha)}{\sigma}. \tag{34}$$

Eliminating v, t_1, and t_2 yields

$$d_2^+ d_{1/2}^- \frac{[\sigma^2+(1+\delta)^2][\sigma^2+(1-\delta)^2]}{\sigma^2-(1-\delta^2)} a = 8\lambda a. \tag{35}$$

We assume that the operand of the left member of Eq. (35) and λ may both be expanded as a convergent power series in δ^2. Equation (35) may then be written as

$$[D_0 + D_1\delta^2 + \cdots][\tau_0 + \tau_1\delta^2 + \cdots] = 0, \tag{36}$$

where

$$\tau = \tau_0 + \tau_1\delta^2 + \cdots = \frac{[\sigma^2+(1+\delta)^2][\sigma^2+(1-\delta)^2]}{\sigma^2-(1-\delta^2)} a \tag{37}$$

and the differential operators $D_0, D_1, \cdots$ depend on the terms $\lambda_0, \lambda_1, \cdots$ in the expansion

$$\lambda = \lambda_0 + \lambda_1\delta^2 + \cdots \tag{38}$$

Using as the independent variable

$$Z = (1+\sigma^2)^{-1}, \tag{39}$$

we obtain for D_0 and D_1 the expressions

$$D_0 = Z(1-Z)^2 \frac{d^2}{dZ^2} - 2Z^2(1-Z)\frac{d}{dZ} - (j+\tfrac{1}{2})(j+\tfrac{3}{2}) + 2\lambda_0(1-Z)(1-2Z) \tag{40a}$$

and

$$D_1 = 2Z(1-2Z)D_0 + 2\lambda_1(1-Z)(1-2Z) - 2\lambda_0 Z(1-Z)(1-8Z+8Z^2) \tag{40b}$$

From Eq. (37),

$$D_0\tau_0=0 \tag{41a}$$

and

$$D_0\tau_1+D_1\tau_0=0 \tag{41b}$$

We act on Eq. (41b) with the operator $L_1\{\tau_0,\cdot\}$, where

$$L_1\{A,B\}=\int_0^1 dZZ^{-2}A\cdot B. \tag{42}$$

Assuming the integrals converge, it may be shown that

$$L_1\{\tau_0,D_0\tau_1\}=L_1\{\tau_1,D_0\tau_0\}=0, \tag{43}$$

using (41a). Thus

$$L_1\{\tau_0D_1\tau_0\}=0, \tag{44}$$

and for given values of j and q, the known expressions for λ_0 and τ_0 may be used to yield a value for λ_1. In the simplest case ($q=2j+2$),

$$\lambda_1=\frac{\partial\lambda}{\partial\delta^2}\bigg|_{\delta^2=0}=\left(\frac{3j+4}{8j+13}\right)\lambda_0, \tag{45a}$$

where

$$\lambda_0=\tfrac{1}{2}(2j+3)(4j+5) \tag{45b}$$

2. Numerical calculations

The analytic solutions of the relevant equations [viz. Eq. (35) for Kummer's solutions, and the momentum space analogs of Eqs. (19) for Keam's solution (with $j=0$)] is rather difficult, and consequently a numerical approach has been adopted based on that used by Keam in finding his solution.

(i) The method

We require a solution $f(p)$ of the BS equation to satisfy the boundary conditions[12]

$$a>-2, \quad b<-3, \tag{46a}$$

where the behaviour of any radial term $g(P)$ in the expression for $f(p)$ is given by

$$g(P)\sim P^a \quad \text{as } P\to 0,$$
$$g(P)\sim P^b \quad \text{as } P\to\infty. \tag{46b}$$

Consider a system of n coupled second order differential equations in n radial functions $\mathbf{f}=(f_1,\ldots,f_n)$. This is equivalent to a system of $2n$ coupled first order differential equations in the $2n$ functions $\mathbf{f}'=(df_1,\ldots,df_n,f_1,\ldots,f_n)$. Here the operator d denotes differentiation with respect to y, where

$$y=\sigma^2. \tag{47}$$

With y as the independent variable, we may evaluate the Frobenius series for which Eq. (46a) is satisfied as P (and y)$\to 0$. This yields n_1 vectors $\mathbf{f}_i'$ $(i=1,\ldots,n_1)$. Similarly, with the independent variable

$$u=(1+\delta^2+y)^{-1}, \tag{48}$$

we may evaluate the n_2 vectors $\mathbf{g}_j'(j=1,\ldots,n_2)$ for which Eq. (46a) is satisfied as $P\to\infty$.

The variable u is chosen so that, for the cases considered, there is a region of the complex y plane in which both sets of vectors are convergent series.

Hence if $\mathbf{h}'$ represents the radial functions (and their derivatives) of a solution of the BS equation, then in this region

$$\mathbf{h}'\equiv a_i\mathbf{f}_i'\equiv b_j\mathbf{g}_j' \tag{49}$$

where i and j are summed over their respective ranges, and not all of a_i or b_j are zero.

If, as in both cases considered in this section,

$$n_1+n_2=2n, \tag{50}$$

the condition that a nontrivial solution of Eq. (49) exists is that

$$d_\lambda(y)=\det\begin{bmatrix}\mathbf{f}_1'\\ \vdots\\ \mathbf{f}_{n_1}\\ \mathbf{g}_1'\\ \mathbf{g}_{n_2}'\end{bmatrix}\equiv 0 \tag{51}$$

for each eigenvalue λ.

If $n_1+n_2>2n$, Eq. (49) may be satisfied for all appropriate values of λ, while if $n_1+n_2=2n+1-m$ $(m\geq 2)$, m distinct conditions must be satisfied in order that an eigenvalue exist. The latter possibility is considered unlikely, and the former in not encountered for the potentials considered.

The series involved in the expression for $d_\lambda(y)$ were numerically summed, term by term, for two different values of y, until the magnitude of the terms fell below a cut off value, and the determinant $d_\lambda(y)$ was evaluated. Changes in the sign of $d_\lambda(y)$ as λ was varied were used to locate eigenvalues. The use of two values of y can distinguish cases where $d_\lambda(y)$ has a zero, at one value of y, but is not identically zero, or where d_λ approaches an asymptote, and also provides a check on the effects of computer roundoff. The programmes were run on the University of Auckland Burroughs B6700 computer.

(ii) Keam's solution

This is a class 1 solution with $j=J=0$, A type exchange, and $\bar{C}_1=+1$. The momentum space radial functions for this case are

$$d_{3/2}^+d_0^-[(1-\delta^2-\sigma^2)s+2i\sigma v_1]=-16\lambda s, \tag{52a}$$

$$d_2^+d_{1/2}^-[(1-\delta^2-\sigma^2)v_1+2i\sigma s]=-8\lambda v_1. \tag{52b}$$

The system has regular singular points at $y=0$, ∞ and $r-2\pm\sqrt{1-r}$, where

$$r=1-\delta^2. \tag{53}$$

Here $n_1=n_2=2$, and the match may be tested on the interval $y\in(0,(1-\sigma)^2)$ for $\delta>0$.

The eigenvalue was determined for values of $1/r$ up to 40 (note that for large values of $1/r$ the mass ratio $m_a/m_b\cong 4/r$, assuming $\delta>0$). A relative cutoff value |term/sum of series| of 10^{-12} was used for the Frobenius series involved.

For large values of $1/r$, the matching region becomes small, and consequently a new variable $y/[2(1-\delta)^2+y]$ was used to perform the evaluation of the series that are

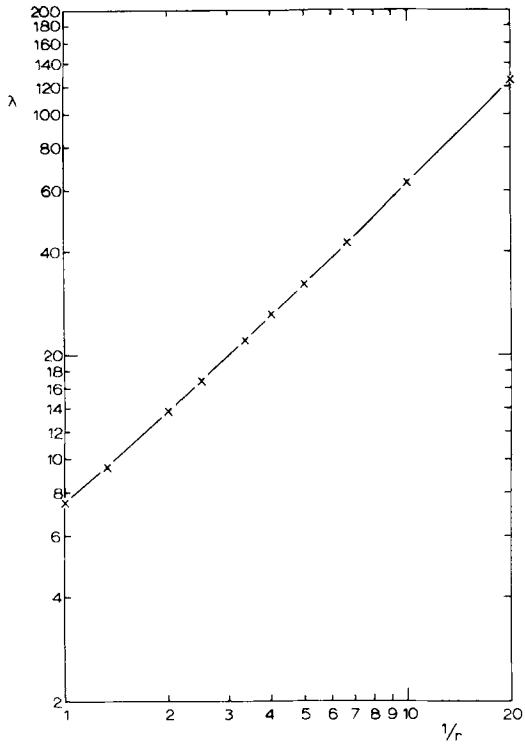

FIG. 1. The coupling parameter as a function of $1/r$ for Keam's solution.

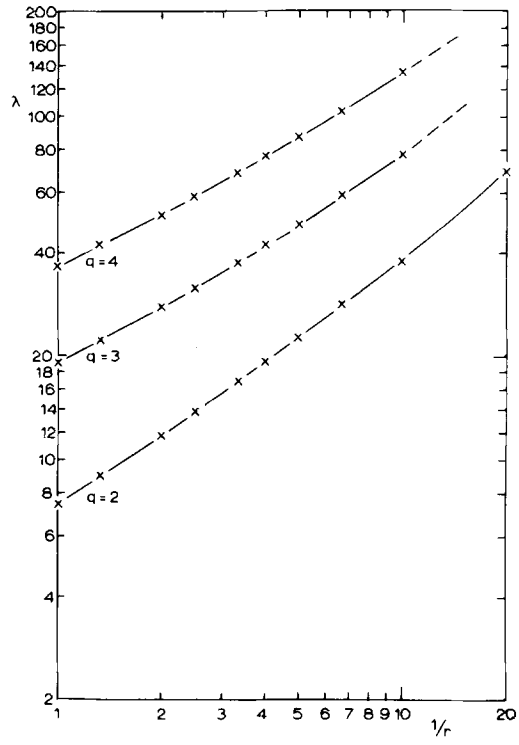

FIG. 2. The coupling parameter as a function of $1/r$ for Kummer's solutions, $j=0$.

valid as $y \to 0$. This provides more rapidly convergent series, with a matching region $y \in (0, \infty)$. When $1/r = 40$, the errors due to roundoff were apparent (the B6700 has 22 significant figures in double precision). For all other values of $1/r$ the agreement between the zeroes of $d_\lambda(y)$ for the two different values of y was excellent, and for $1/r < 10$, the two values agreed to within 1 in 10^8.

The results are summarized in Fig. 1, where λ is plotted against $1/r$ on a log-log scale. The numerical value of $\partial\lambda/\partial\delta^2|_{\delta^2=0}$ agress well with that obtained by Keam.[K4] It appears probable that $\lambda \to \infty$ as $1/r \to \infty$ (and $m_b \to 0$). This is consistent with the fact that no acceptable solutions to Eqs. (52) have been found for $\delta = 1$.[13]

The graph of λ vs $1/r$ is very nearly linear. The gradient of the regression line of $\ln(\lambda - 7.5)$ on $\ln(1/r - 1)$ is 0.997, and the regression line of λ on $1/r$ yields estimates of λ that are accurate to within 0.05 for $1/r$ in the range $[1, 20]$. Thus λ is approximately given by

$$\lambda = 1.357 + 6.194\ (1 - \Delta^2/m^2)^{-1}. \tag{54}$$

(iii) Kummer's solutions

Equation (35) has four regular singular points at the same values of y as for the system of Eqs. (52) for Keam's solution, and may be reduced to Heun's equation. We examine certain properties of Eq. (35), and present an alternative method of determining the eigenvalues, in the appendix.

We again consider the differential equation with dependent variable τ, rather than a. In this case $n_1 = n_2 = 1$, and the match may be tested for $y \in (0, (1-\delta)^2)$. The zeroes of d were numerically determined for all cases in which $\lambda < 50$ when $m_a = m_b$, and for $1/r \leq 10$. Again the agreement between the values of λ obtained for two different values of y is excellent.

The results are summarized in Figs. 2, 3, 4, and 5,

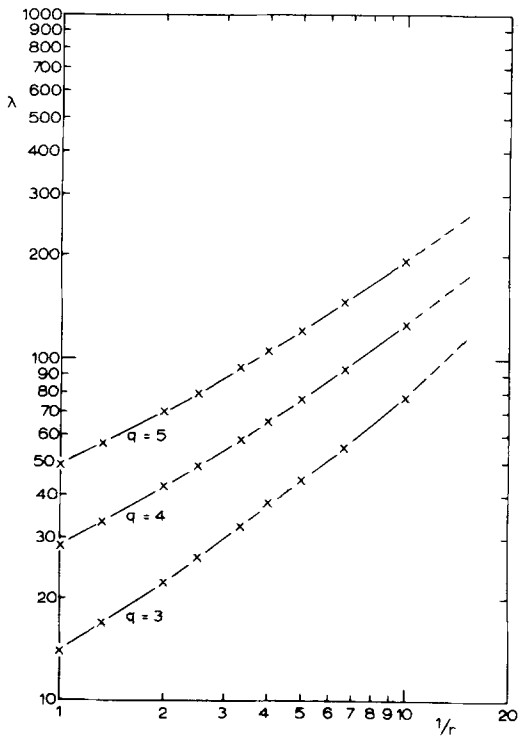

FIG. 3. The coupling parameter as a function of $1/r$ for Kummer's solutions, $j=1/2$.

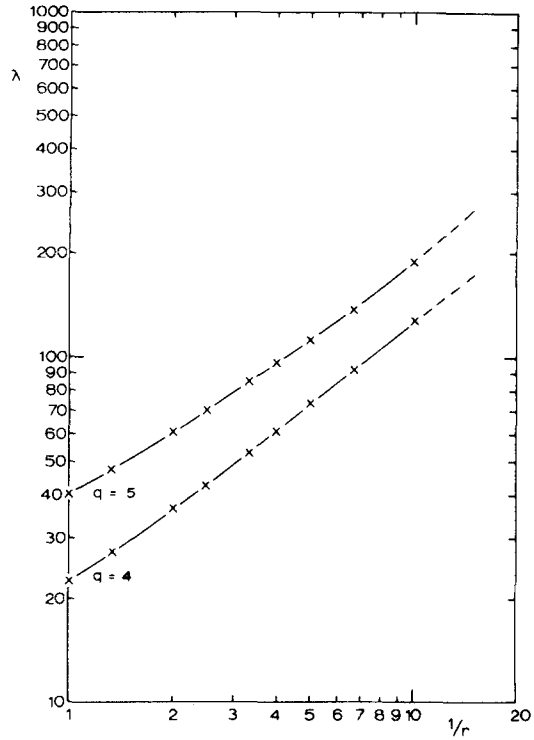

FIG. 4. The coupling parameter as a function of $1/r$ for Kummer's solutions, $j=1$.

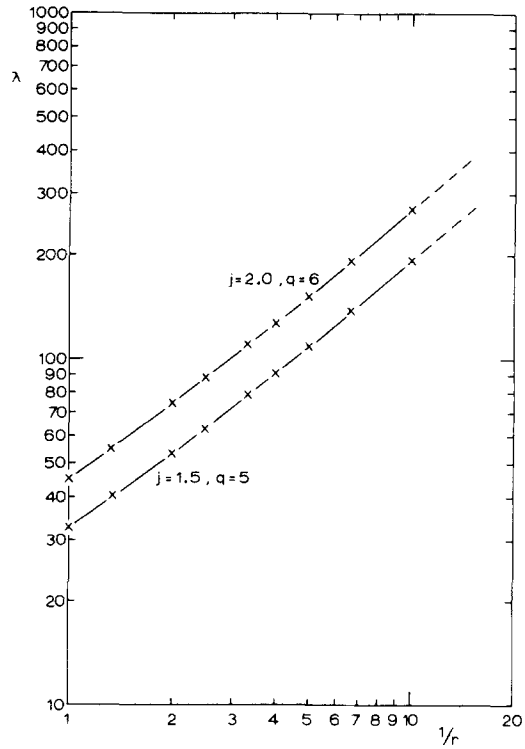

FIG. 5. The coupling parameter as a function of $1/r$ for Kummer's solutions, $j=3/2$ and 2.

again using a log—log scale. Qualitatively they are similar to the results for Keam's solution, though the departure from a linear relationship between λ and $1/r$ is more marked.

In the case $q=2j+2$, the value of $\partial\lambda/\partial\delta^2|_{\delta^2=0}$ agrees well with that given by Eq. (45). In all cases, it appears probable that $\lambda\to\infty$ as $m_b\to 0$. The author has investigated the hypergeometric equation obtained from Eq. (35) when $\delta=1$ (and $m_b=0$), and has found no cases in which acceptable solutions exist, with either positive or negative eigenvalues.

The regression lines of $\ln(\lambda-\lambda_0)$ on $\ln(1/r-1)$, where $\lambda_0=\lambda|_{\delta^2=0}$, yield estimates of λ accurate to within 2% for $1/r$ in the range $[1,10]$. These regression lines yield expressions of type

$$\lambda\approx\lambda_0+A(1/r-1)^B. \tag{55}$$

The values of λ_0, A and B are tabulated in Table II. B is always slightly less than 1.0, increasing with j for fixed q, and decreasing as q increases for fixed j. A increases with q, but is almost constant with respect to j, for fixed q.

3. Action of T_5^b on known solutions

As noted in Sec. IIB, we expect that $T_5^b f$ be a solution of the BS equation, where f itself is a solution. We note that the radial equations for the radial functions of $T_5^b f$ are equivalent to those for the radial functions of f, with m_b replaced by $-m_b$, and λ^j replaced by $\epsilon_j'\lambda^j$. Applying the methods of the previous section to Eqs. (52) and Eq. (35) with m_b negative yields no eigenvalues. Equation (A4) of the appendix has also been analyzed by splitting the operator and the operand into odd and even functions of m_b/m_a and solving the resulting pair of coupled equations numerically, as in the previous section. This yields the expected solutions and eigenvalues for m_b positive, but the solution vanishes identically for m_b negative.

Thus it appears that $T_5^b f\equiv 0$ for the solutions considered in this section, and hence the integrals involved in K6, Eqs. (23), (25), and (26) vanish, so that Eq. (27) of Sec. IIB is not valid.

III. PERTURBATION OF EXCHANGE MASS

A. Introduction

Previous studies of the spinor—spinor BS equation with nonzero exchange mass have been performed by Narayanaswamy and Pagnamenta,[14] who numerically solved the eigenvalue problem in λ using a high momentum cutoff, and Guth,[15] who performed the same task without a momentum cutoff by the addition of regulating

TABLE II. Parameters λ_0, A and B in the expressions $\lambda_{jq}=\lambda_0+A(1/r-1)^B$.

j	q	λ_0	A	B
0	2	7.500	4.212	0.9172
0	3	19.187	8.456	0.8916
0	4	36.682	14.792	0.8766
$\frac{1}{2}$	3	14.000	8.308	0.9308
$\frac{1}{2}$	4	28.651	13.694	0.9118
$\frac{1}{2}$	5	49.083	21.118	0.8976
1	4	22.500	13.816	0.9347
1	5	40.127	20.375	0.9122
$\frac{3}{2}$	5	33.000	20.715	0.9416
2	6	45.500	29.024	0.9431

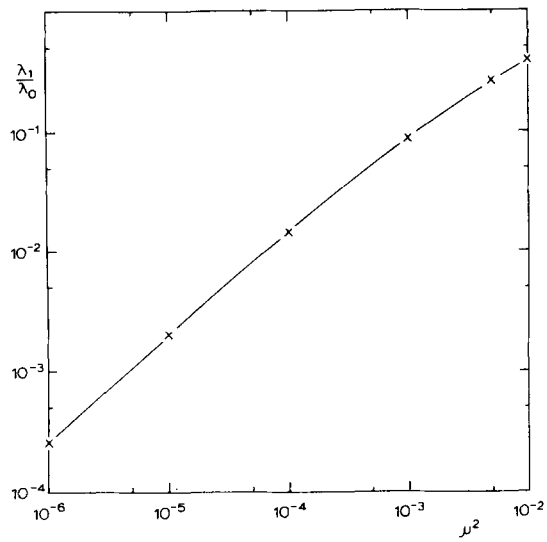

FIG. 6. The fractional change in the coupling parameter as a function of the exchange mass for Keam's solution.

terms to the one particle exchange propagator. The author's approach has been to apply perturbation theory, and use numerical methods to evaluate the integrals encountered. This analysis has been applied to Keam's solution and to the cases $j=0$, $q=2$, $J^P=1^+$ of Kummer's solutions.

B. The method

We consider the momentum space form Eq. (4) of the BS equation for $ma=mb$, $P_\mu=0$. When the exchange mass μ is nonzero, the right member involves terms of type $(\mu^2/m^2)\ln(\mu^2/m^2)$ and thus a perturbation expansion in powers of μ^2 is not valid. Rather we assume

$$f(p)=f_0(p)+f_1(\mu^2,p), \tag{56a}$$

$$\lambda=\lambda_0+\lambda_1(\mu^2), \tag{56b}$$

where $f_0(p)$ and λ_0 are appropriate to the case $\mu^2=0$, and $f_1(\mu^2,p)$ and $\lambda_1(\mu^2)\to 0$ as $\mu^2\to 0$. We assume also that the integrals encountered are convergent,[16] and set

$$\int d^4k\,\frac{\Gamma^j f_\alpha(k)}{(p-k)^2+\mu^2}=\int d^4k\,\frac{\Gamma^j f_\alpha(k)}{(p-k)^2}+\Delta I_\alpha(\mu^2,p) \qquad (\alpha=0,1) \tag{57}$$

where Γ^j is appropriate to the interaction type considered. It is assumed that for small μ $\Delta I_1(\mu^2,p)$ and $\lambda_1(\mu^2)\Delta I_0(\mu^2,p)$ may be neglected.

Equating the remaining terms involving μ^2 yields

$$(\gamma\cdot p-im)f_1(\mu^2,p)(\gamma\cdot p-im)+\frac{\lambda_0}{\pi^2}\int d^4k\,\frac{\Gamma^j f_1(\mu^2,k)}{(p-k)^2}$$
$$=-\frac{\lambda_0}{\pi^2}\cdot\Delta I_0(\mu^2,p)-\frac{\lambda_1}{\pi^2}\int d^4k\,\frac{\Gamma^j f_0(k)}{(p-k)^2}; \tag{58}$$

we multiply Eq. (58) on the left with the adjoint $\bar f(p)$ where[K4]

$$\bar f(\mathbf{p},p_4)=\gamma_4 f^*(\mathbf{p},-p_4)\gamma_4, \tag{59}$$

take the trace and integrate over momentum space. The left member yields zero [cf. K4, Eq. (52)]. We therefore obtain, with the use of Eq. (57) for $\alpha=0$,

$$\frac{\lambda}{\lambda_0}=\frac{\lambda_0+\lambda_1(\mu^2)}{\lambda_0}=2-\mathcal{J}(\mu^2)/\mathcal{J}(0), \tag{60}$$

where

$$\mathcal{J}(\mu^2)=\int d^4p\,d^4k\,Tr[f_0(p)\Gamma^j f_0(k)]/[(p-k)^2+\mu^2]. \tag{61}$$

For the solutions considered in Sec. III C the evaluation of the integrals in $\mathcal{J}(\mu^2)$ is most conveniently performed by numerical integration. In particular, we consider the integrals

$$A_r(P)=\int d^4k\{[(p-k)^2+\mu^2][1+K^2/m^2]^r\}^{-1}$$
$$=m^2\,\frac{\pi^2}{r-1}\int_0^1 dx\left(\frac{x}{z}\right)^{r-1}, \tag{62}$$

where

$$z=x+(1-x)(\mu^2/m^2)+x(1-x)(P^2/m^2), \tag{63}$$

and

$$b_r^\mu(P)=\int d^4k\,k_\mu\{[(p-k)^2+\mu^2][1+K^2/m^2]^r\}^{-1}$$
$$=p_\mu B_r(P), \tag{64}$$

where

$$B_r(P)=m^2\,\frac{\pi^2}{r-1}\int_0^1 dx(1-x)\left(\frac{x}{z}\right)^{r-1}. \tag{65}$$

The latter forms for $A_r(P)$ and $B_r(P)$ are obtained by using the Feynman method.[17]

C. Application to known solutions

(i) Keam's solution

This may be written,[K4] to within a normalization factor,

$$f_0(p)=s(P)+v_1(P)\gamma\cdot p/P, \tag{66}$$

where

$$s(P)=2u^7(1-14u+56u^2-84u^3+42u^4), \tag{67a}$$

$$v_1(P)=7i(K/m)u^8(1-2u)(1-6u+6u^2), \tag{67b}$$

and

$$u=(1+P^2/m^2)^{-1}. \tag{67c}$$

In this case

$$\bar f_0(p)=f_0(p). \tag{68}$$

After some simplification, we obtain

$$\mathcal{J}(\mu^2)=8\pi^2\int_0^\infty P^3dP\{8s(P)[A_7(P)+14A_8(P)+56A_9(p)$$
$$-84A_{10}(P)+42A_{11}(P)]+14iv_1(P)(P/m)[B_8(P)$$
$$-8B_9(P)+18B_{10}(P)-12B_{11}(P)]\}. \tag{69}$$

The double integral in the right member of Eq. (69) was evaluated by Euler–Romberg integration on the University of Auckland Burroughs B6700 computer, until successive estimates agreed to within one in 10^5. For very small μ^2/m^2 the range of integration for x was divided into two, to allow for the rapid change in x/z as $x\to 0$.

The results are summarized in Fig. 6, where λ_1/λ_0 is plotted against μ^2/m^2 on a log–log scale. The approximate expression

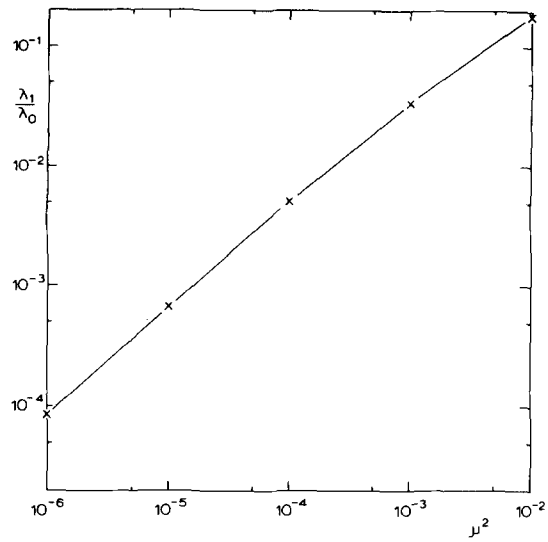

FIG. 7. The fractional change in the coupling parameter as a function of the exchange mass for Kummer's solutions $(j=0, q=2)$.

$$\lambda \cong 7.5 - \frac{\mu^2}{m^2}\left(84.27 + 24.89 \ln \frac{\mu^2}{m^2}\right), \tag{70}$$

which is fitted to the values of λ when $\mu^2/m^2 = 0$, 10^{-6} and 10^{-5}, gives good agreement in the range $\mu^2/m^2 < 0.005$ (i.e., $\mu/m < 0.07$). For larger values of μ^2, $\mathcal{J}(\mu^2)$ is substantially smaller than $\mathcal{J}(0)$, and thus the perturbation assumption is no longer appropriate.

We note that expression of Eq. (70) for λ varies more rapidly with small μ^2 than is the case for the simplest solution to the scalar–scalar BS equation.[18] This arises principally because the integrals A_r and B_r vary more rapidly with μ^2 as r increases (for the scalar–scalar solution, $r=3$).

(ii) Kummer's solution

The solutions for $j=0$, $q=2$, $J=1$, $J_3=0$, $\overline{P}=\pm 1$, may be written, to within a normalization factor (the superscript denoting the eigenvalue of $\overline{P}$, for a fermion–fermion system), as

$$f_0^+ = -2iu^7[(\gamma_1 p_1 + \gamma_2 p_2)(\gamma_4 p_3 - \gamma_3 p_4) + (p_1^2 + p_2^2)\gamma_4\gamma_3] + [2u^7 - u^6][p_2\gamma_5\gamma_1 - p_1\gamma_5\gamma_2] \tag{71}$$

and

$$f_0^- = -2iu^7[(\gamma_1 p_2 - \gamma_2 p_1)(\gamma_3 p_3 + \gamma_4 p_4) - (p_3^2 + p_4^2)\gamma_1\gamma_2] + [2u^7 - u^6][p_4\gamma_5\gamma_3 - p_3\gamma_5\gamma_4]. \tag{72}$$

Only the A sector terms contribute to $\mathcal{J}(\mu^2)$, since $\Gamma^2(=\hat{\Gamma}_2)$ yields zero when acting on a T sector matrix. For both f_0^+ and f_0^-,

$$\mathcal{J}(\mu^2) = 8\pi^2 \int_0^\infty P^3 dP P^2 (2u^7 - u^6)(A_6 - 2A_7). \tag{73}$$

Using the same methods as for Keam's solution, we obtain the results summarized in Fig. 7. In this case λ may be approximately expressed as

$$\lambda \cong 7.5 - \frac{\mu^2}{m^2}[15.14 + 7.20 \ln(\mu^2/m^2)] \tag{74}$$

for $\mu^2/m^2 < 10^{-2}$. We note that the variations of λ with μ^2 is less rapid, and the expression is a more accurate approximation, than is the case with Eq. (70).

ACKNOWLEDGMENT

I should like to thank Professor R. F. Keam for his interest and for many helpful discussions.

APPENDIX: HEUN'S EQUATION AND KUMMER'S SOLUTIONS

We transform Eq. (35), using as independent variable

$$u' = (1 + P^2/m_a^2)^{-1} = (1+\delta)^2[(1+\delta)^2 + \sigma^2]^{-1}, \tag{A1}$$

and defining F by

$$\tau = (u')^{(\mu+1)/2}(1-u')^{-(2j+3)/2}F(u'), \tag{A2}$$

where

$$\mu = [8\lambda + (2j+2)^2]^{1/2} \tag{A3}$$

Thus we obtain Heun's equation[19]

$$\frac{d^2F}{du'^2} + \left(\frac{\gamma}{u'} + \frac{\delta}{u'-1} + \frac{\epsilon}{u'-b}\right)\frac{dF}{du'} + \frac{\alpha\beta u' - q}{u'(u'-1)(u'-b)} = 0 \tag{A4}$$

where

$$b = \left[1 - \frac{m_b}{m_a}\right]^2 - 1,$$
$$\alpha = \frac{\mu}{2} - j, \quad \beta = \frac{\mu}{2} - (j+1),$$
$$\gamma = \mu + 1, \quad \delta = -(2j+1),$$
$$\epsilon = 0, \quad q = b\left[\alpha\beta - 2\lambda\frac{m_b}{m_a}\left(1 + \frac{m_b}{m_a}\right)\right]. \tag{A5}$$

We consider a solution of form [cf. Eq. (10) of Ref. (19)].

$$F(u') = (u'-1)^{1-\delta}\sum_{m=0}^{\infty} a_m (u')^m, \quad a_0 = 1. \tag{A6}$$

If the series converges for $u' \in [0,1]$ then this solution gives acceptable behaviour of $a(P)$ for $P \to 0$ and $P \to \infty$. The recurrence relation for the series is

$$A_m a_{m-1} + B_m a_m + C_{m+1} a_{m+1} = 0 \tag{A7}$$

where

$$A_m = (m-1)(m-2) + (2+\gamma-\delta)(m-1) + \alpha\beta + \gamma(1-\delta), \quad m = 1, 2, \cdots, \tag{A8a}$$

$$B_m = -\{(a+1)m(m-1) + [a(2+\gamma-\delta)+\gamma]m + q + a\gamma(1-\delta)\}, \quad m = 0, 1\cdots, \tag{A8b}$$

$$C_{m+1} = a(m+1)(m+\gamma), \quad m = 0, 1, \cdots. \tag{A8c}$$

The condition for convergence for $u' \in [0,1]$ is[19]

$$B_0 = q_1 C_1 \tag{A9a}$$

where

$$q_1 = -\cfrac{A_1}{B_1 - \cfrac{A_2 C_2}{B_2 - \cdots}}. \tag{A9b}$$

The infinite continued fraction q_1 may be evaluated

approximately by numerical means, and a search made for zeroes in $B_0 - q_1 C_1$. This was done for several values of m_b/m_a and j, and in each case the eigenvalue λ agreed with that obtained in Sec. IIB (iii) to the accuracy expected.

[1]B. J. Brennan and R. F. Keam, Prog. Theor. Phys. 49, 1679 (1973). Hereafter referred to as B1.

[2]R. F. Keam, J. Math. Phys. 9, 1962 (1968); 10, 594 (1969); 11, 394 (1970); 12, 515 (1971); Prog. Theor. Phys. 50, 957 (1973); 50, 967 (1973). Hereafter referred to as K1 to K6, respectively.

[3]The notation used for higher transcendental functions is that of *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).

[4]We note that negative values of λ are acceptable, as g_a and g_b need not have the same sign.

[5]N. Nakanishi, Prog. Theor. Phys. Suppl. 43, 61 (1969).

[6]W. Kummer, Nuovo Cimento 31, 219 (1964); Erratum 34, 1840 (1964).

[7]We note that since $E < |(m_b - m_a)|$, the more massive particle is unstable with respect to the decay $a \to a\bar{b} + b$ or $\dot{a}b + \bar{b}$. We assume that our solutions have analytic behavior in the region around $E = |(m_b - m_a)|$, and that there are no pathological consequences of the above.

[8]Note that the parity operator for fermion—fermion systems is given by $^{(j)}f^c(x) = -\gamma_4 f^c(-\mathbf{x}, x_4)\gamma_4$, with the opposite sign to that for fermion—antifermion systems.

[9]There are some differences in notation to K1; viz. (i) a_1, a_2, w_2, and $\bar{w}$ are defined in a different manner to the same functions in K1 and (ii) in class 4, t_1 and t_2 are interchanged by comparison with case B, Sec. 5 of K1.

[10]We note that the (Fourier) transform $f(p)$ of a configuration space solution $f(x)$ involves the same angular kets, with radial terms that are related to those of $f(x)$ by a Hankel transform [cf. K3, Eqs. (11a) and (11b)].

[11]See Eq. 2.34 of Ref. 3, where $n = 2j + 1$, $\lambda' = 4\lambda$.

[12]See K3, Sec. 2, and B1, Sec. 2.

[13]R. F. Keam, Private Communication.

[14]P. Narayanaswamy and A. Pagnamenta, Nuovo Cimento 53A, 635 (1968).

[15]A. H. Guth, Ann. Phys. 82, 407 (1974).

[16]We note that the boundary condition Eq. (46) is a sufficient condition for this to be so.

[17]R. P. Feynman, Phys. Rev. 76, 769 (1949).

[18]See Ref. (11), Eqs. (19) and (19).

[19]A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw Hill, New York, 1955), Vol. III, p. 57 *et seq.*

Kinks and extensions*

David Finkelstein and Gin McCollum

Belfer Graduate School of Science, Yeshiva University, New York, New York 10033
(Received 4 November 1974)

Kinks, homotopically nontrivial lightcone fields on R^4, can be black holes without curvature singularities, satisfying the weak energy condition. Kinks and all other spherically symmetric stationary spacetimes on R^4 with roots of g_{00} have incomplete geodesics which need extension. For simple roots the Kruskal extension method works and the topology around each root is that of the Kruskal manifold. For multiple roots another extension method is given, based on symmetry, and another topology.

I. INTRODUCTION FOR THE VISUAL IMAGINATION

Kinks in the gravitational field arise in the homotopy classification of spacetimes on a given topology, in the search for conserved particlelike structures in general relativity, in the study of the spin of the nonlinear gravitational field, and as possible internal structures for black holes. The present work studies the properties of some simple examples of kinks as orientation for more general considerations, just as the detailed study of the Schwarzschild solution was important in the development of the theory of black holes. A typical spherically symmetric stationary kink is illustrated in Fig. 1.

The kink[1,2] spacetime in Fig. 1 has lightcones tumbling over and up again: starting asymptotically Minkowski at infinity; gradually turning toward the center so that they are upside down, that is, future toward $-t$, at $r=0$; and turning back up symmetrically on the other side. Radii with lightcones pointing toward the center are trapped; nothing can escape. Timelike and null geodesics falling radially toward the center gradually turn to the side and then backwards, travel toward $-t$ in the middle. There are no anomalies: The lightcones retain their topology; there are no curvature singularities; background topology is R^4.

Can a physical source cause such a configuration of lightcones? In Sec. III the curvature is computed and a mass distribution given such that the stress tensor obeys the weak energy positivity conditions. Part of this problem was solved by Williams and Zia.[3]

There is a problem common to many spacetimes: incomplete geodesics. At certain radii the lightcones are turned so that one branch is parallel to the t axis, and there are null geodesics parallel to the t axis at those radii. In Fig. 1 there are two such radii, duplicated on the other side of $r=0$. In one direction nothing can cross such radii; Fig. 1 is a black kink, so that nothing can get out of the middle. Near such a one-way surface there are three sets of null geodesics: those that cross the surface, those that stay in the surface, and those that approach but do not cross. The last category are all incomplete. The others may or may not be.

Incomplete geodesics in an otherwise acceptable lightcone field on a manifold, within local field theory, constitute no reason to extend the geodesics or the manifold. But general relativity is not a local theory: Geodesic incompleteness and topology of the background manifold are both global considerations. And in general relativity geodesics have physical meaning. If they are incomplete, test particles fall off the edge of the manifold in finite proper time, or finite affine parameter. So it may be assumed that a manifold with incomplete timelike or null geodesics is never the full physical manifold, though we do not learn this from Einstein's field equations.

It is not possible, given half a geodesic, to find the other half; information about the rest of the universe is unavailable. But certain simple assumptions can be made which lead to extensions which indicate the kind of phenomena that may occur. For example, the null manifold may be assumed to be composed of two copies of the given part joined along a surface in such a manner that there are no shock waves or other violent physical indications of the juncture.

This kind of extension is performed in Sec. VI on the kink of Fig. 1. First, each one-way surface is considered separately, as in Figs. 2(a) and 2(b). Each patch in Fig. 2(a) or 2(b) is half of the patch below it in Fig. 2(c), the half with time appropriately oriented toward the future and not crossed by the U axis. Both the U and V axes are at the radius of the one-way surface. The U axis is at $t=-\infty$. Any null geodesic in a UV plane is parallel to either the U or the V axis. Null geodesics in Fig. 2(a) or 2(b) parallel to the t axis in the one-way surface transform into the V axis. Null geodesics crossing the one-way surface transform into lines crossing the V axis, parallel to the U axis. Incomplete null geodesics approaching the one-way surface but not crossing transform into lines crossing the U axis, crossing $t=-\infty$ into the new region, the added isomorph, the new sheet. t increases from $-\infty$ in each sheet away from the U axis: Time is backward in the other sheet.

In each sheet there is one region of r less than the one-way surface radius and one with r greater. r is

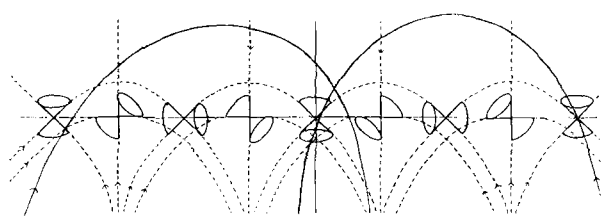

FIG. 1. Kink: —, timelike line; ---, null geodesic.

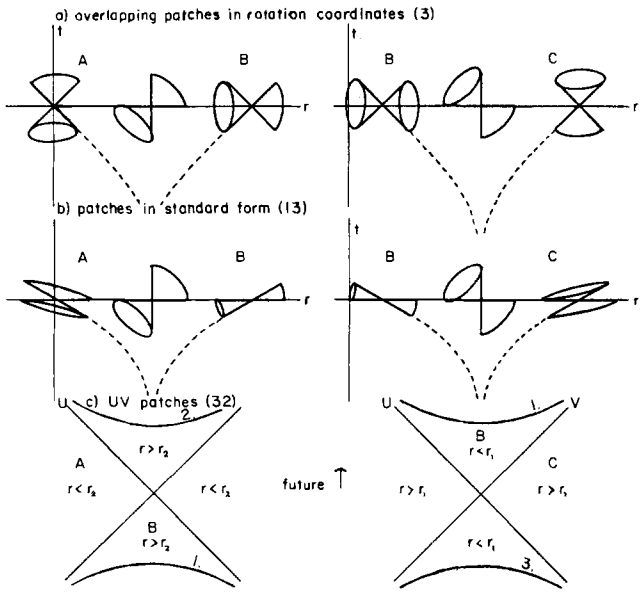

FIG. 2. Kink extended.

constant along hyperbolae as indicated. The two UV patches overlap between the two-oneway surfaces in the original sheet. Hyperbolae 1, 2, and 3 are all at the same radius, between the radii of the one-way surfaces. Hyperbolae 1 are in the original sheet; they are to be identified since they are two transformations of the same line of constant r in the original sheet. Hyperbolae 2 and 3 are in two new sheets. They may be identified or not. The manifold continues across hyperbolae 2 and 3. Identifying them provides a continuation. If they are not identified, there will be three sheets total, and perhaps more if there are additional one-way surfaces in the new sheets.

The original tr plane is stationary. Since each half of the UV plane differs from the tr plane only by a coordinate transformation, each half of the UV plane is stationary also: A consistent flow can be defined along the lines of constant r. But the full UV plane is not stationary; the flow contradicts itself at the center.

A spacetime with one point gravitating source and no other fields has only monotonically turning lightcones, turning on through the one-way surface. A quite different situation arises when the lightcones turn back up inside the one-way surface as in Fig. 3(b). This is the case in the Reissner—Nordström solution when the mass is equal to the electric charge and when a gravitating sphere has its radius equal to its Schwarzschild radius. These two examples are akin: the Reissner—Nordström solution has a distributed mass source.

The two one-way surfaces of Fig. 3(c) are extended in the UV patches of 3(d). By changing parameters the one-way surfaces can be made to coalesce: 3(c) deforms into 3(b). 3(b) cannot be extended by the same method as Figs. 1, 2, and 3(c). The one-way surface of 3(b) can be removed by changing parameters: 3(b) deforms into 3(a). 3(a) does not need to be extended. The extension of 3(b) will be of an intermediate topology to 3(a) and 3(d). Regions which deform and extend into each other are marked with the same capital letter.

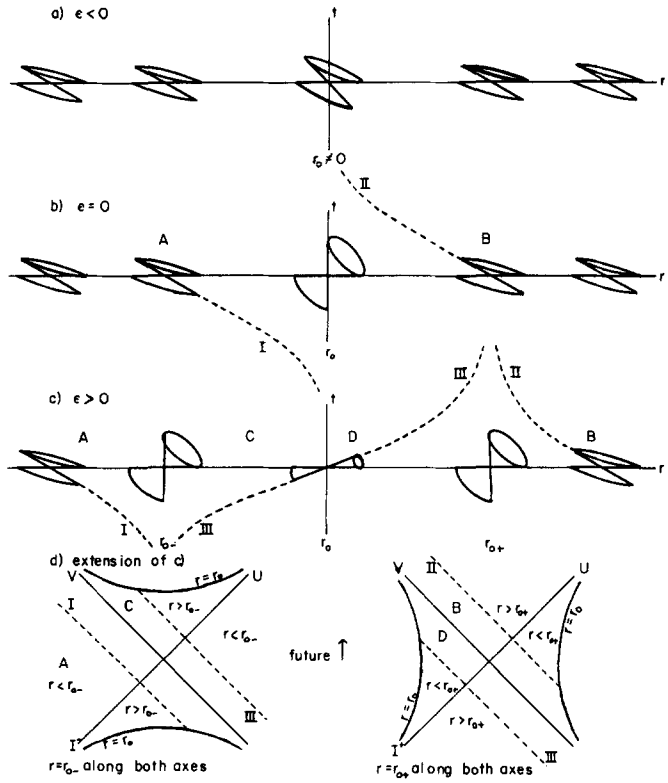

FIG. 3. Double root and its deformations.

3(b) extends in Fig. 4(a). The two sides of the one-way surface in the original spacetime are marked 1 and 2. The new pieces of the manifold are isomorphs of the old. They are labelled by r, according to the side of the one-way surface at r_0. Null geodesics crossing r are drawn as horizontal dashed lines. Geodesics formerly incomplete cross into new areas of the manifold, region

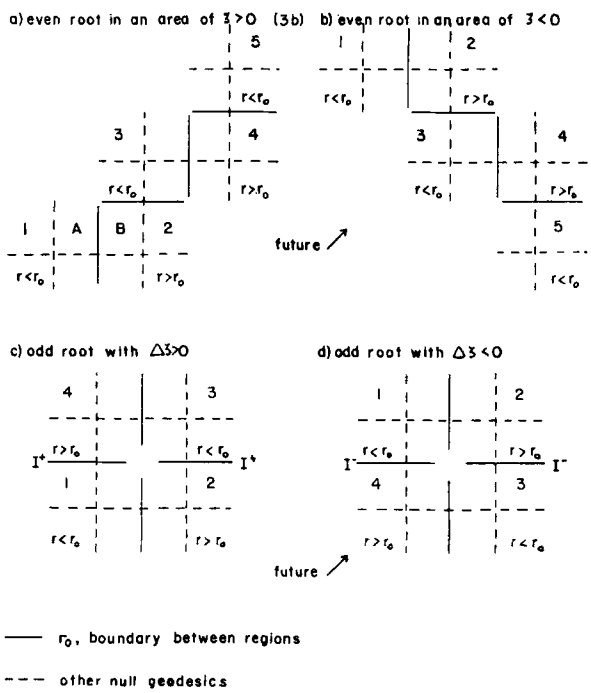

FIG. 4. Symmetry extensions for white multiple roots.

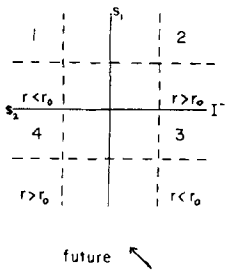

FIG. 5. Symmetry extension for a black simple root with $\zeta' > 0$.

3 or farther; null geodesics formerly incomplete are vertical dashed lines. New piece 5 can be identified with region 1. Then the extended manifold becomes a kind of cylinder. No identification need be made; new pieces can iterate, piece 6 and on.

The other extended manifolds in Figs. 4 and 5 can be understood by a similar transformation of geodesics. All null geodesics in the tr plane are parallel to those drawn as dashed lines. All null geodesics crossing the one-way surface are horizontal in Figs. 4 and 5; all null formerly incomplete are vertical, crossing lines of infinite t. For odd roots the identification of regions 5 and 1 has already been made, hence the nonappearance of region 5. Figures 2(c) and 5 are extensions of the same kind of root by this method and by analytic extension. They are the same.

It is possible that the essentials of extensions and their topologies depend on spherical symmetry or stationarity. These assumptions are, however, no more unreasonable than the assumption that half the universe is a copy of the other, and are made to simplify the problem. The answer is interesting as an indication of physical possibilities. If the original manifold is analytic, the symmetry extension is an analytic extension.

II. KINKS[1,2]

Models of our spacetime commonly have physical impossibilities such as incomplete geodesics or curvature singularities. Perhaps the only physically plausible spacetimes are those with nontrivial topologies, either of the lightcone field or of the background manifold, or both.

For example, the Schwarzschild solution,

$$ds^2 = \zeta\, dt^2 - (1/\zeta)\, dr^2 - r^2\, d\omega^2 \tag{1}$$

with $\zeta = 1 - 2m/r$ and where $d\omega^2 = d\theta^2 + \sin^2\theta\, d\varphi^2$, which is used to compute the properties of spacetime near the sun, both has incomplete geodesics approaching $r = 2m$ and, in form (1), has a singularity in g_{11} at $r = 2m$, the lightcone collapses [Fig. 6(a)]. If the Schwarzschild solution is transformed to Eddington[4] form,

$$ds^2 = \zeta\, dt^2 + 2(\zeta \pm 1)\, dt\, dr + (\zeta \pm 2)\, dr^2 - r^2\, d\omega \tag{2}$$

so that the coordinate patch extends to $r = 0$, there is still a real singularity, a curvature singularity, at $r = 0$ [Fig. 6(b)].

Inside $r = 2m$ the lightcones can turn back up or continue turning all the way over. The interior Schwarzschild solution has the former possibility; the latter involves a kink. To see the rotation, transform the Schwarzschild solution into

$$ds^2 = \cos 2\alpha\, (dt^2 - dr^2) \pm 2 \sin 2\alpha\, dt\, dr - r^2\, d\omega^2, \tag{3}$$

where $\alpha(r)$ is the angle of turning of the lightcone in the tr plane, gives $\cos 2\alpha = 1 - 2m/r$. The patch now ends at $r = m$, where $\cos 2\alpha = -1$; a cosine cannot get any smaller [Fig. 6(c)].

The coordinate system (3) unnecessarily restricts the Schwarzschild solution; the Eddington form covers more of r. But (3) is suggestive: Let α continue to increase as r decreases part $r = m$ until $\alpha = \pi$ at $r = 0$ (Fig. 1). The resulting manifold has homotopically nontrivial lightcone field, and has one kink:

Consider spacetimes which are cross sections of a fiber bundle, where the background manifold is $S^3 \times R$ and the fiber is $S_{1,3}$, the manifold of symmetric matrices of signature 1—3. The cross sections will have equivalence classes, homotopy classes, according to winding number around S_3. This winding number is called the kink number (or topological charge).

Matter enters the trapped region inside $\cos 2\alpha = 3\pi/4$ and flows into the past in the black case, for the minus sign, and in the white case, the plus sign, it flows radially outward from the future. A single kink has no timelike loops and no global Cauchy hypersurface.

It is not at all clear that such a construction will have a physically acceptable stress tensor.

III. PHYSICALLY PLAUSIBLE KINKS: CALCULATION

The Einstein tensor for the kink (3) is computed to check the physical acceptability of the stress tensor. The tetrad procedure is described by Flanders[6] and Misner.[7] The kink tetrad is

$$\omega^0 = \cos\alpha\, dt - \sin\alpha\, dr,$$

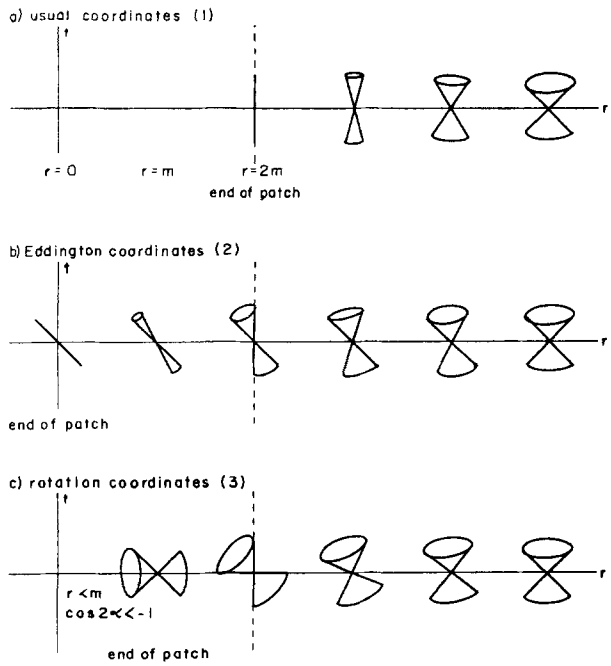

FIG. 6. Schwarzschild line element.

$$\omega^1 = \cos\alpha\, dr + \sin\alpha\, dt,$$
$$\omega^2 = r\, d\theta, \tag{4}$$
$$\omega^3 = r\sin\theta\, d\phi,$$

where

$$ds^2 = (\omega^0)^2 - (\omega^1)^2 - (\omega^2)^2 - (\omega^3)^2. \tag{5}$$

Taking $d\omega^\mu = \omega_\nu{}^\mu \wedge \omega^\nu$ gives the matrix of 1-form $\omega_\nu{}^\mu$:

$\nu = 0$	1	2	3
$\mu = 0$ 0	$-\alpha^1\cos\alpha\omega^1 - \alpha^1\sin\alpha\omega^0$	$-(1/r)\sin\alpha\omega^2$	$-(1/r)\sin\alpha\omega^3$
1 $-\alpha^1\cos\alpha\omega^1 - \alpha^1\sin\alpha\omega^0$	0	$-(1/r)\cos\alpha\omega^2$	$-(1/r)\cos\alpha\omega^3$
2 $-(1/r)\sin\alpha\omega^2$	$(1/r)\cos\alpha\omega^2$	0	$-(1/r)\cot\theta\omega^3$
3 $-(1/r)\sin\alpha\omega$	$(1/r)\cos\alpha\omega^3$	$(1/r)\cot\theta\omega^3$	0

And $d\omega_\nu{}^\mu - \omega^\mu{}_\alpha \;\; \omega_\tau{}^\alpha = \frac{1}{2}R^\mu_{\nu\rho\sigma}\omega^\rho \wedge \omega^\sigma$ gives:

$$R^0_{110} = [\cos 2\alpha]'', \tag{6}$$
$$R^0_{202} = R^1_{212} = R^0_{303} = R^1_{313} = -(1/r)[\cos 2\alpha]', \tag{7}$$
$$R^2_{323} = (2/r^2)[1 - \cos 2\alpha]. \tag{8}$$

It is convenient to introduce a function

$$\mu(r) = (r/2)(1 - \cos 2\alpha). \tag{9}$$

This is the function f of Williams and Zia.[3] Then the nonzero components of the Einstein tensor can be written simply:

$$G^0{}_0 = G^1{}_1 = \frac{4}{r^2}\frac{d\mu}{dr},$$
$$G^2{}_2 = G^3{}_3 = \frac{2}{r}\frac{d^2\mu}{dr^2}. \tag{10}$$

The indices refer to the tetrad but could as well refer to the t, r, θ, ϕ coordinate system since $G^0{}_0 = G^1{}_1$. Choose units and signs so that the gravitational equations are $G^\mu{}_\nu = T^\mu{}_\nu$, with stress tensor T. Then

$$\mu = \tfrac{1}{4}\int_0^r r^2 T^0{}_0\, dr. \tag{11}$$

The choice of $\mu(r)$ for gravitational field variable linearizes the gravitational equations. $\mu(r)$ measures the gravitating mass within the sphere of radius r. The simple relation (9) between the mass and the angle of tilt of the lightcone $\alpha(r)$ was obtained by Williams and Zia. For empty spacetime $\mu = \text{const}$.

There are several physical conditions on μ. For (3) to be physical we demand that for arbitrary timelike n^μ, $T_{\mu\nu}n^\mu n^\nu$ be nonnegative and finite everywhere. Necessary and sufficient conditions are

(a) μ must never decrease (as a function of r in $0 \le r < \infty$)[7],

(b) $r^2\mu'$ must never decrease,

(c) $\mu = O(r^3)$ at $r \to 0$,

(d) $0 \le \mu/r \le 1$,

because

(a) taking $n = \omega^0$ gives $T^0{}_0 = (4/r^2)(d\mu/dr) > 0$,

(b) taking n an arbitrary timelike linear combination of ω^0 and ω^2 gives $0 \le T^0{}_0 + T^2{}_2 = (2/r^3)(d/dr) \times r^2\, d\mu/dr)$,

(c) $T^\mu{}_\nu$ would blow up at the origin if μ were not $O(r^3)$,

(d) this follows from (9).

For the simplest kink, $\cos 2\alpha$ goes to 1 at 0 and ∞, and to -1 with zero derivative at some r_0.

One function satisfying these criteria is

$$\mu = \left(\frac{3}{2 + 1/r}\right)^3. \tag{12}$$

This defines a kink with physically reasonable stress tensor in the entire tr plane. Mass is concentrated at the center. $\mu(r)$, the mass inside radius r, becomes constant as $r \to \infty$.

The condition that leads to (a)–(d) is the weak energy condition. The stronger energy condition, that $R_{\mu\nu}n^\mu n^\nu \ge 0$ for all timelike n^μ, is violated by (12), which exhibits strong azimuthal tensions near $r = 0$. A kink of the following more general family may be still more physical.

IV. TRANSFORMATION TO STANDARD FORM

To consider trapping, incompleteness, and extension in general for spherically symmetric stationary spacetimes on R^4, it is convenient to transform to a standard form such that any two spacetimes equivalent under coordinate transformation are equal in standard form. The standard form should have no unnecessary singularities, like the ones in the Schwarzchild form. The rotation form (3) is too restricting. The Eddington form (2) is linear in ξ, no spurious singularities or unnecessary restrictions. But there is a simpler form to choose as standard:

$$ds^2 = e^{2\beta}[\xi\, dt^2 + 2\kappa_i\, dt\, dr] + g_{22}\, d\omega^2, \tag{13}$$

where β and ξ are functions of r, $\kappa_i = \pm 1$, and $g_{22} = -r^2$ except in neighborhoods of extremal r.

The most general spherically symmetric stationary spacetime can be written

$$ds^2 = g_{00}(dx^0)^2 + 2g_{01}\, dx^0\, dx^1 + g_{11}(dx^1)^2 + g_{22}\, d\omega^2, \tag{14}$$

where $g_{\mu\nu}$ is a function of x^1 alone. (14) contains four arbitrary functions, $g_{00}, g_{01}, g_{11}, g_{22}$, but admits a group of coordinate transformations

$$x^{0\prime} = \eta x^0 + a^0(x^1),$$
$$x^{1\prime} = a^1(x^1) \tag{15}$$

with two arbitrary functions of x^1 and an arbitrary constant η, and hence can be brought to standard form (13), which has only two arbitrary functions:

In standard form, g_{22} is no longer an arbitrary function of x^1. Transform (14) by $x^1 \to r$ such that $g_{22} = -r^2$ where possible, where $\sqrt{-g_{22}}$ is an admissible coordinate. Around extremal g_{22}, that is, in neighborhoods of $x_0{}^1$ such that $g_{22,1}(x_0{}^1) = 0$, and in areas of constant g_{22}, transform $x^1 \to r$ so that

$$\sqrt{-g_{22}} = (r - r_0)^p + \sqrt{-g_{22}(r_0)} \tag{16}$$

and p is minimized.

For convenience of future transformations, factor the $x^0 r$ block so that the determinant of the quotient is -1; that is,

$$ds^2 = e^{2\beta}[\zeta(dx^0)^2 + 2\kappa\, dx^0\, dr + \iota\, dr^2] + g_{22}\, d\omega^2, \tag{17}$$

where

$$\zeta\iota - \kappa^2 = 1 \tag{18}$$

and β, ζ, κ, and ι are functions of r alone. Transformation preserving this form leave ζ fixed within a sign, but can change κ and ι. If (17) is supposed linear in ζ,

$$ds^2 = e^{2\beta}[\zeta\, dt^2 + 2(\lambda\zeta + \kappa_i)\, dt\, dr + (\lambda^2\zeta + 2\lambda\kappa_i)\, dr^2] + g_{22}\, d\omega^2. \tag{19}$$

Eddington chose $\lambda = 1$, the simplest choice approaching Minkowski when $\zeta \to 1$. $\lambda = 0$ is the choice for the standard form (13), given from (17) by the transformation

$$x^0 = t - \int_l^r \frac{\kappa - \kappa_i}{\zeta}\, dr, \tag{20}$$

where κ_i is the sign of κ at a root r_i of ζ. One coordinate patch with appropriate κ_i will surround each root of ζ; each root of ζ must have a rootless neighborhood. The Schwarzschild manifold can be extended through its root of ζ with either sign for κ_i.

That standard form (13) is accessible only patch by patch is not an inelegance but the main point. Only, but not all, spacetimes of zero kink number are covered in one patch. If neighboring roots are of different κ_i, they belong to different patches. It will be convenient also to divide adjacent roots of the same κ_i by patch. Standard form (13) draws attention to root surfaces, which is appropriate because incomplete geodesics approach root surfaces. Root surfaces are trapped surfaces and boundaries of trapped surfaces.

V. COORDINATE CONDITIONS

Coordinate conditions facilitate the discussion of a general class of spacetimes, but raise the question, are the results special to the coordinate system? A spacetime symmetry or characteristic is invariant if it consists in the existence of coordinate system displaying such a symmetry. For example, a spacetime is spherically symmetric if it admits a coordinate system of the form

$$ds^2 = g_{00}(dx^0)^2 + 2g_{01}\, dx^0\, dx^1 + g_{11}(dx^1)^2 + g_{22}\, d\omega^2 \tag{21}$$

where $g_{\mu\nu}$ is a function of x^0 and x^1. If a spacetime admits such a coordinate system on only a part of the manifold, it is spherically symmetric only on that part. If it admits one globally, or in overlapping patches, it is all spherically symmetric. Further, coordinate conditions may result in a coordinate system exhibiting invariants of the spacetime. For example, in the spherically symmetric coordinate system of (21), the number of zeroes of $g_{22,1}$ is an invariant of the spacetime which may be displayed by requiring that $g_{22} = -(x^1)^2$ except in neighborhoods of x_0^1, where $g_{22,1}(x_0^1) = 0$. Then each x_0^1 is a special case, listed separately in an atlas of the manifold.

The standard form (13) is chosen for both reasons. The spacetimes considered are spherically symmetric and stationary. The form (14) is chosen to display the restriction to spherical symmetry. $g_{\mu\nu}$ is a function of x^1 alone because the spacetimes considered are invariant under time translation: x^0 is chosen to be the direction of invariance under translation.

Beyond spherical symmetry and stationarity, there are no restrictions placed on the class of spacetimes considered, except that they have background manifold R^4. The rest of the coordinate conditions are imposed to display particularly interesting characteristics of each spacetime.

The problem of incomplete geodesics involves surfaces of $g_{00} = 0$, root surfaces, stationary characteristic surfaces. For this reason, the root surfaces are singled out by the standard form, separated one to each patch. To count kink number, it is sufficient to know two signs and an indicator for each root in order: κ_i, the sign of g_{11} at a root of g_{00}; f_i, the sign of x^0 in the direction of future; and $\Delta\zeta_i$, the change in ζ at r_i. For simple roots, $\Delta\zeta_i$ is the sign of $\zeta'(r_i)$. For even roots, $\Delta\zeta_i = 0$. For odd roots, $\Delta\zeta_i$ is the sign of the change in ζ as r increases through r_i. For a qualitative picture of the extension near a root, like Figs. 2(c), 3(d), 4, and 5, it is sufficient to know κ_i, f_i, $\Delta\zeta_i$, and the order of the root.

κ_i is evident when the spacetime is written in normal form. f_i is known for one patch, usually patch 1 at I^0, and must be followed through the rest of the patches. After g_{22} is standardized and $e^{2\beta}$ is divided out, ζ is left, the only arbitrary function in the tr block. Each root must be examined for the order and $\Delta\zeta_i$.

The same κ_i, f_i, $\Delta\zeta_i$, and order of the root would be found by analysis of any spacetime in the form (14), without imposing any more coordinate conditions, as would be found by first transforming to standard form (13) and then analyzing: They are invariants under coordinate transformation for spherically symmetric stationary spacetimes.

Functions of invariants are also invariant. Kink number was defined in Sec. II without reference to coordinate system; it is a property of the manifold. Kink number can also be counted by seeing how many times the lightcones tumble in the tr plane. Pictures like Figs. 1, 2(a), 2(b), 3(b), and 3(c) can be drawn qualitatively from κ_i, f_i, and $\Delta\zeta_i$ alone; then the number of whole turns can be counted.

The order of the root is also an invariant: The

topology of the extended patch is invariant because it depends only on the order of the root. Many other properties of the extension depend further only on κ_i, f_i, and $\Delta\zeta_i$, so that they also are invariant.

VI. TRAPPED SURFACES AND INCOMPLETE GEODESICS

A shell of light emitted normal to a spherical surface of constant r grows or shrinks in surface area as $\sqrt{-g_{22}}$ increases or decreases in the direction in which the light is emitted. If g_{22} is constant, the surface area of the shell of light is constant. If $\sqrt{-g_{22}}$ has a minimum, the shell grows no matter in which direction in r it is emitted; if $\sqrt{-g_{22}}$ has a maximum, the shell can only shrink. A trapped surface is one from which a shell of normally emitted light either can never grow or can never shrink. A surface of extremal $\sqrt{-g_{22}}$ is always a trapped surface.

Trapped surfaces also occur where the future part of the lightcone points in only one direction in r. There matter can flow in only one direction in r. Since the boundary conditions are time-symmetric, static, this appears as a spontaneous break in time symmetry. [1] For the standard form (13),

$$\begin{aligned} ds^2 &= \zeta\, dt^2 + 2\kappa_i\, dt\, dr \\ &= dt(\zeta\, dt + 2\kappa_i\, dr) \\ &= 0 \end{aligned} \tag{22}$$

gives the lightcone in the tr plane. dr is always null. The other branch of the lightcone approaches dt as r approaches a root of ζ. There is a simple physical difference between the odd and even roots of ζ. At odd roots of ζ the turning branch of the lightcone rotates through the dt direction, and there is an interval in r of trapped surfaces on one side of the root, where dt is spacelike, where $\zeta < 0$. At even roots of ζ the branch only touches dt; even roots never bound regions of trapped surfaces.

The direction in r of the positive t half of the lightcone at a root of ζ is given by κ_i: $\kappa_i = -1$ when it points toward decreasing r; $\kappa_i = 1$ when it points toward increasing r. If f is the sign of t pointing toward the chosen future and $\sqrt{-g_{22}}$ is not extremal at a root of ζ, then $f\kappa_i = 1$ indicates a white hole and $f\kappa_i = -1$ a black hole.

Roots of ζ are accompanied by incomplete geodesics. For the standard form the geodesic equations restricted to the tr plane reduce to

$$\begin{aligned} &\ddot{\xi}^0 - (\beta' x_i \zeta + \tfrac{1}{2}\kappa_i \zeta')(\dot{\xi}^0)^2 = 0, \\ &\ddot{\xi}^1 + (\beta'\zeta^2 + \tfrac{1}{2}\zeta'\zeta)(\dot{\xi}^0)^2 \\ &\quad + 2(\beta'\kappa_i\zeta + \tfrac{1}{2}\kappa_i\zeta')\dot{\xi}^0\dot{\xi}^1 + 2\beta'(\dot{\xi}^1)^2 = 0, \end{aligned} \tag{23}$$

where $x^\mu = \xi^\mu(s)$ is the equation of the geodesic, $\dot{\xi}^\mu$ is the derivative with respect to an affine parameter s, and $\zeta' = d\zeta/dr$.

There are three sets of null geodesics:

1. Those parallel to the r axis,

$$\xi^0 = \text{const}, \quad \xi^1 = \int be^{-2\beta} ds, \tag{24}$$

where b is a constant.

2. Those parallel to the t axis at a root of ζ,

$$\xi^1 = r_i, \tag{25}$$

$$\zeta = \dot{\xi}^1 = 0 \tag{26}$$

with ξ^0 depending on the order of the root. For simple roots,

$$\xi^0 = (\ln as)/a, \tag{27}$$

where

$$a \equiv -\tfrac{1}{2}\kappa_i \zeta'(r_i); \tag{28}$$

half the geodesic is missing. For multiple roots

$$\xi^0 = s; \tag{29}$$

the geodesic is complete.

3. Those approaching r_i but not crossing, going to $I^\pm$ rather than $\mathcal{I}^\pm$, [8]

$$\xi^0 = -2\kappa_i b \int (e^{-2\beta}/\zeta)\, ds, \quad \xi^1 = \int be^{-2\beta} ds \tag{30}$$

The sign of ξ^0 as ξ^1 approaches r_i is given by

$$\operatorname{sgn}\xi^0 \to -\kappa_i \operatorname{sgn}(\zeta \cdot \dot{\xi}^1) \text{ as } \xi^1 \to r_i. \tag{31}$$

For geodesics 1 and 3, radial null geodesics, when $\beta = 0$, r is an affine parameter. Since any tr block can be transformed into standard form without changing the determinant of the tr block, and affine parameters are invariant under coordinate transformation, this is a general result: if the tr block of a spherically symmetric stationary $g_{\mu\nu}$ has determinant -1, then radial null geodesics have the radial coordinate as affine parameter.

VII. EXTENDED KINK

The kink of form (3) already has $g_{22} = r^2$ and tr block determinant -1, so that it can be transcribed directly into standard form (13) with $\zeta = \cos 2\alpha$. There are two roots; for indexing, let $r_1 > r_2$. Choose the black kink of Fig. 2, so $\kappa_1 = -1$, and in patch 1

$$ds^2 = \cos 2\alpha\, dt^2 - dt\, dr - r^2 d\omega^2. \tag{32}$$

The second root surface is also black, but $-t$ is future at r_2, $f = -1$. See Fig. 2(b). So $\kappa_2 = 1$. There are no further roots for positive r.

There are incomplete geodesics approaching each root. For each root all incomplete geodesics go off to I^-, as is evident from Fig. 1 and as is given in Eq. (31).

In order to complete the geodesics, follow the example of Kruskal[9] and Graves and Brill, [10] and transform the standard form (13) to the form

$$ds^2 = e^{2\beta}(2f^2\, dU\, dV) + g_{22}\, d\omega^2 \tag{33}$$

where f and r are functions of U and V. Then

$$U = e^{\sigma t} \exp\left(2\kappa_i \sigma \int_1^r \frac{dr}{\zeta}\right), \tag{34}$$

$$V = -\frac{1}{\sigma} e^{-\sigma t}, \tag{35}$$

$$2f^2 = \frac{1}{\sigma} \exp\left(\int_1^r \frac{\zeta' - 2\kappa_i\sigma}{\zeta} dr\right), \tag{36}$$

with undetermined constant σ.

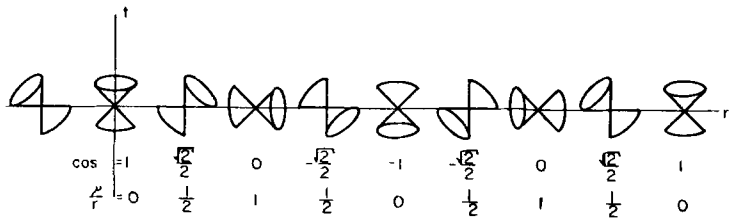

FIG. 7. Onion.

In order that $2f^2$ be regular at r_i, let $2\kappa_i\sigma = \zeta'(r_i)$. If $2f^2$ is only regular for a simple root of ζ. The original tr plane covers only half of the UV plane. Extend the tr plane by letting it also cover the other half of the UV plane with the same l and with t reversed. Then geodesics crossing the U axis will be continuous.

As an example take μ of (12). Then

$$r_1 \approx 5.1, \quad \zeta'(r_1) \approx 0.0168, \quad \sigma_1 \approx -0.0084. \tag{37}$$

The root is simple, so that $2f^2$ is regular. The null geodesics parallel to the r axis transform into straight lines crossing the V axis. The incomplete null geodesics approaching r_1 and I^- transform into straight lines crossing the U axis; $V(I^-) = 0$. [Fig. 2(c)].

There are both timelike and null geodesics crossing the U axis. Although only null geodesics were enumerated, both timelike and null have been extended.

In patch 2 where $\kappa_2 = 1$,

$$r_2 = 0.25, \quad \zeta'(r_2) \approx -4.96, \quad \sigma_3 \approx -2.48. \tag{38}$$

The orientation of the U_2V_2 plane should be taken consistent with future toward $-t$. Again $V(I^-)$ will be 0. Now $\sigma < 0$, so that $2f^2 < 0$, and so positive U and negative V are future.

Two new sheets have been attached or recognized. They may be separate and ramify to many more sheets, or they may be identified, a total of two sheets. Manifolds with higher kink number may need more extensions.

While there is no natural law of composition for $g_{\mu\nu}$'s, group elements of $GL(4, R)$ have a product. This distinguishes a combination of two one-kink manifolds to make one two-kink manifold: multiplication of group parametrizations. In general, kink numbers add when the group parametrizations multiply. For (3)'s, group composition reduces to α addition.

A continuous process creating a black kink and a white kink from an originally kinkless spacetime can be constructed: $0 = 1 - 1$. Their formation can be pictured as a pulling apart or as a nesting of spheres within spheres like an onion, as in Fig. 7. An onion (3) of monotonically increasing α is unphysical: μ would decrease, violating condition (a). So for a physical (3) of monotonically increasing α, at most two extensions are necessary.

VIII. MULTIPLE ROOTS

Neither the most general kink nor the most general spherically symmetric stationary spacetime on R^4 has been extended; multiple roots have been excluded. Since any two-dimensional manifold is conformally flat, locally any spacetime becomes (33) under some coordinate transformation. To find that transformation from the standard form (13), set

$$\begin{aligned} &\zeta\, dt^2 + \kappa_i\, dt\, dr = 2f^2\, dU\, dV, \\ &dU = U_t\, dt + U_r\, dr, \\ &dV = V_t\, dt + V_r\, dr. \end{aligned} \tag{39}$$

Then

$$\begin{aligned} &2f^2\, U_t V_t = \zeta, \\ &2f^2(U_t V_r + V_t U_r) = \kappa_i, \\ &2f^2 U_r V_r = 0. \end{aligned} \tag{40}$$

Without loss of generality choose $V_r = 0$. Then

$$U_t = \kappa_i \zeta U_r \tag{41}$$

and

$$\begin{aligned} &V = V(t), \\ &U = U(t + \textstyle\int_l^r \kappa_i\, dr/\zeta), \\ &2f^2 = \zeta/U_t V_t. \end{aligned} \tag{42}$$

The problem is to choose U so that $2f^2$ neither vanishes nor blows up.

If the root of ζ in question is n-fold, $n > 1$, choose

$$V = t, \quad U = \left(t + \int_l^r \frac{dr}{\zeta}\right)^{-p}, \tag{43}$$

where

$$p = 1/(n-1). \tag{44}$$

A transformation to these coordinates is a restriction of the manifold, not an extension, since $U \to \pm\infty$ for finite values of t and r in the original coordinates. The lower limit of integration may be set so that $U \to \pm\infty$ in both directions of V, or it may be chosen to leave U finite at the edge of the patch of original coordinates without an extension of the manifold.

In the case of a simple root of ζ, form (32) provides the extension. If n were 1 in p of (44), p would blow up, suggesting the necessary exponential of (34)–(36).

Discussion of an even root, the topology of the manifold and the configuration of the incomplete geodesics, clarifies the problem of multiple roots. Consider the tr block

$$ds^2 = (\zeta - \epsilon)\, dt^2 + dt\, dr, \tag{45}$$

where ζ has an even root at r_0 and $\zeta > 0$ around r_0 (Fig. 3). For negative ϵ there is no root. As $\epsilon \to 0$ from below, the lightcone at r_0 tilts until, when $\epsilon = 0$, one of its branches is dt and g_{00} has a root. As ϵ increases from 0, the lightcones tip farther so that there are roots at r_{0+}

and r_{0-}, both of odd order. For $\epsilon < 0$, the manifold needs only one sheet; for $\epsilon > 0$ two tunnels to a new sheet are added: $\epsilon = 0$ is the boundary between topologies.

Two planes used to extend the $\epsilon > 0$ case are shown in Fig. 3(d). As $\epsilon \to 0$, r_{0+} and r_{0-} approach r_0, and the r_0 hyperbolae approach the U and V axes. In the limit the axes of the two planes will be identified. But the two roots differ in an important respect, the sign of ξ^0 as the incomplete geodesics approach the roots. For odd roots, all incomplete geodesics go to the same infinity of t, which then becomes $U = 0$ in the UV plane. For even order roots some incomplete geodesics go to I^+ and some I^- as they approach the roots, so that no consistent sign choice can be made for the U axis.

Another extension method and, it turns out, different extended topology will be needed for multiple roots. The method works as well for simple roots, and gives again the UV plane.

IX. SYMMETRY

The Schwarzschild solution is invariant under coordinate t, $t \to -t$. Stationary spacetimes are invariant under a time translation and not necessarily under a time reflection. But the Schwarzschild solution also can be written in coordinates without x^0 relection invariance. When do stationary spacetimes have a time-reflection invariance hidden by a coordinate transformation?

Transformation to Eddington form extends the region of regularity of the Schwarzschild solution past $r = 2m$ to $r = 0$. The Eddington form lacks t-reflection invariance. What in the transformation breaks the invariance? A choice of a sign must be made in the transformation from Schwarzschild to Eddington, a choice of the sign of g_{01}, the color of the hole. This is the breaking of the invariance.

In the Schwarzschild solution, Fig. 6(a), there are two sets of incomplete null geodesics in the tr plane approaching $r = 2m$, one set approaching I^+ and the other approaching I^-. The two sets of geodesics are reflected into each other by a t-reflection. Under transformation to Eddington coordinates, Fig. 6(b), one set of geodesics remains incomplete by approaching but not crossing $r = 2m$. The other now crosses $r = 2m$. Which set crosses $r = 2m$ depends on the choice of sign in the transformation.

If the Eddington form is transformed back to Schwarzschild and then back to Eddington but with the opposite sign, a kind of reflection has been made: The original and the transformed Eddington forms differ only in the sign of t. The null geodesics crossing $r = 2m$ have been transformed into null geodesics approaching but not crossing.

Transforming back and forth between Eddington and Schwarzschild is not a proper sort of operation, because the Schwarzschild solution is not meaningful for $r < 2m$. The transformation can be carried out between the two Eddington forms with opposite signs of g_{01}, without Schwarzschild for intermediary. But the transformation still blows up at $r = 2m$: The time reflection symmetry survives outside $r = 2m$ and inside, but not at $r = 2m$.

Any spherically symmetric stationary spacetime exhibits the same symmetry, and the same symmetry breaking when $g_{00} = 0$. This is easy to show when the spacetime is written in standard form (13), or in other forms without unnecessary singularities.

X. EXTENSION BY SYMMETRY

The fact that the given manifold is almost piecewise static but not static can be used to extend the manifold when writing in the form (33) fails.

If the transformation

$$t \to -t + \int_i^r t_r dr, \quad t_r = t_r(r), \tag{46}$$

is required to take (13) into itself, t_r is determined:

$$t_r = -2\kappa_i/\zeta. \tag{47}$$

t_r belows up at r_i only: The transformation may map one side of the root into itself, but not both. Identify the original $r > r_i$ region with the new $r > r_i$ region. Then a new $r > r_i$ region has been found and the manifold has been symmetrized with respect to (46). A similar second transformation applied to the new $r < r_i$ region gives a new $r > r_i$ region, which cannot be identified with the original. A third transformation gives an $r < r_i$ region which can be identified with the original. In this way two $r > r_i$ regions and two $r < r_i$ regions are covered.

The boundaries between regions are the null geodesic parallel to the t axis at r_i (25) and (29). In the case of a multiple roots, all the null geodesics are complete in this new manifold, a cylinder cut off along $r = 0$ crossed with a sphere (Fig. 4). This cylinder is the boundary topology between the half-plane of Fig. 3(a) and the two planes identified along r_0 of Figs. 3(c) and 3(d). Regions which deform into each other as ϵ changes have the same capital letter. For a simple root, the null geodesic (25) and (27) remains incomplete. Let the first r_i boundary be the positive s_1 axis, the new boundary after one transformation be the negative s_2 axis, the new boundary after the second transformation be the negative s_1 axis, the new boundary after the third transformation be the positive s_2 axis (Fig. 5). Add the point $s_1 = s_2 = 0$. Let null geodesics crossing the s_1 (s_2) axis be lines of constant s_1 (s_2). All the null geodesics are now complete for an odd root in a manifold of the topology of the Kruskal-like UV plane.

The above method demonstrates the existence of an extension with a change in a topology of the manifold at each root of ζ isolated in r. Further, the topology of an r interval of the extended manifold depends on the multiplicity of the root, $R \times R \times S^2$ becoming $R^2 \times S^2$ for simple roots and $R \times S^1 \times S^2$ for multiple roots.

The Reissner—Nordström solution has

$$\zeta = 1 - 2m/r + e^2/r^2 \tag{48}$$

with a double root when $m = e$. The tr block determinant is already -1 and $g_{22} = -r^2$. Call the root radius r_0. κ_0 may be chosen to be either sign for the transformation to standard form (13). For $\kappa_0 = -1$, a black one-way

surface, an inward lightcone field turns back up just inside r_0; an outward field turns back up for $\kappa_0 = 1$, white. For $\kappa_0 = 1$ the root looks like Fig. 3(b) and the extension looks like Fig. 4(a). For $\kappa_0 = -1$ the root is a t-reversal of Fig. 3(b). The incomplete geodesics approaching r_0 from above approach I^- rather than I^+. Future is toward positive t, but toward decreasing r rather than increasing. The extension steps in the opposite direction from future, toward negative t and increasing r.

XI. TIMELIKE GEODESICS

Only null geodesics are enumerated in Sec. VI. Only their affine parameters are given. That is because the affine parameters of the null geodesics, specifically of the ones at r_i, determine the topology of the extended manifold. The physical nature of the extension is much clearer if timelike geodesics are extended also.

In Fig. 5 it is easy to see that there are timelike geodesics continuing from one region to the other: Timelike geodesics have been extended; in the original manifold, timelike geodesics were incomplete. It is possible not only to send light into the new sheet, but also to travel there.

In the unextended manifold, a root surface has been assumed to be a line a constant r and changing t crossed with a sphere of radius r_i.

In the extension of multiple roots the root surface retains its topology, but multiplies: There appear several of them. The case of a simple root is exceptional in this respect; the root surface changes topology to two intersecting lines crossed with a sphere.

The change in topology of the root surface allows not only null geodesics but also timelike to cross the center. This is impossible for multiple roots.

XII. SYMMETRY AND KRUSKAL-LIKE EXTENSIONS

A stationary spherically symmetric spacetime on R^4 with a simple root of ζ can be extended in two ways to a spherically symmetric spacetime not stationary and not on R^4: by the Kruskal-like extension, writing in coordinates of the form (33) and extending across the U axis; and by the symmetry extension.

The UV plane of the Kruskal-like extension [Figs. 2(c) and 3(d)] is a plane with any null geodesic in the plane parallel to one of the two axes.

The symmetry extension, Figs. 4 and 5, gives an arbitrary or infinite number of regions. The first and fifth regions are the first that can be identified; the minimum number of regions is four. In the case of a simple root, there are still incomplete geodesics after the regions are proliferated. A point must be added which continues one null geodesic at r_i into the one two regions away, as in Fig. 5.

The added point must then be covered by a coordinate patch which overlaps the other coordinate patches, the copies of the original patch. In the original patch, the geodesic along r_i is crossed by null geodesics at constant t. The other radial null geodesics in the tr plane all approach either I^+ and I^- as they approach r_i; they do not cross the geodesic along $r = r_i$, but cross the point $(I^{\pm}, r_i)$. That point extends into a line in either extension, into the U axis in the Kruskal-like extension and into the boundary between regions in the symmetry extension. The line does not intersect any of the null geodesics at constant t, and so it is parallel to them.

Let the plane coordinates of a point on the boundary between regions in the symmetry extension be $(s_i, 0)$, where s_i is the affine parameter. Let null geodesics crossing the boundary be lines of constant s_i. Then the coordinates of a point in the extended plane of Fig. 5 are (s_1, s_2), where the two null geodesics in the plane which pass through the point cross the two null geodesics at r_i at affine parameters s_1 and s_2.

The UV patch and the coordinate patch covering the center of the plane in the symmetry extension for the simple root are the same: Any null geodesic in the plane is parallel to one of the axes. The coordinates are the affine parameters of null geodesics along the axes. All null geodesics are lines of one constant coordinate.

In the one case in which the two extension methods are both applicable, the results are the same.

XIII. EXAMPLE: REISSNER–NORDSTRÖM

There is a well-known example of the phenomenon illustrated in Fig. 3, the Reissner–Nordström solution.[12] It is usually given as

$$ds^2 = \left(1 - \frac{2n}{r} + \frac{e^2}{r^2}\right)dt^2 - \frac{dr^2}{(1 - 2n/r + e^2/r^2)} - r^2\,d\omega^2, \tag{48'}$$

that is, in Schwarzschild form (1) but with $\zeta = 1 - 2m/r + e^2/r^2$. The order of roots of ζ depends on the discriminant $(m^2 - e^2)^{1/2}$. If $e > m$, there are no roots of ζ for real r. If $e = m$, there is one real double root. If $e < m$, there are two real simple roots.

In (48), g_{22} is already t^2 and the determinant of the tr block is already -1. So (48) can be immediately written in standard form (13) with $\zeta = 1 - 2m/r + e^2/r^2$. But κ_i must be chosen. In (48), g_{22} blows up at a root of g_{00}; the coordinate patch ends at the root of g_{00}. Transforming to standard form (13) extends the coordinate patch to $r = 0$, where there is a curvature singularity. Each extension past a root requires a choice. For $e > m$ there is no root, no choice, and no extension, as in Fig. 3(a).

For $e = m$, there is one root, one choice, and one extension, as in Fig. 3(b). If $f_1 = 1$ and the choice $\kappa_1 = 1$ is made then the hole is white, as in Fig. 3(b). For a while double root in an area of $\zeta > 0$, the extension looks like Fig. 4(a), with new regions added toward increasing t and increasing r. If the choice $\kappa_1 = -1$ is made, both future and new regions are in the direction of increasing t but decreasing r. Two regions with the same domain of r can be identified, forming a sort of cylinder.

For $e < m$, there are two real roots, two choices, and two extensions. Regions in the extensions can be identified or not, a matter of choice and eventually of experiment. The two roots are both simple, so the extensions will both be planes, like Fig. 5. For Fig. 5, the choice

$\kappa_i = -1$ has been made. If the choice $\kappa_i = 1$ had been made, the extension would wind the opposite way around the center, and have perpendicular direction of future, Fig. 4(c) with a point added in the middle.

In Fig. 3(c), the choice $\kappa_1 = \kappa_2 = 1$ has been made. The extension is shown in Fig. 3(d), with capital letters labelling matching regions in the two parts of the extension and in the original patch, Fig. 3(c). Also, regions in 3(b) deforming into regions of 3(c) are labelled with the same capital letter. The deformation of the extension of 3(b), Fig. 4(a), into the extension of 3(c), 3(d), can be followed by studying matching regions.

The choice $\kappa_1 = \kappa_2 = -1$ can also be made, and the resulting extensions are similar, with different directions in t of geodesics approaching r_i and with different direction of f. Can the choices $\kappa_1 \neq \kappa_2$ be made? They correspond to a time reversal between the two roots, so that they are contradictory. There would have to be either another root between the two roots, or a curvature singularity, or time reversal, essentially setting the κ_i equal. So there are really only two choices for the signs of the two simple roots: both white or both black.

And there are two possible extensions, with many possible identifications. In Fig. 3(d) the two hyperbolae not in regions C or D can be identified or not. If they are not identified, the spacetime must be continued beyond each of them. As usual, there is insufficient information for the extension. One heuristic assumption that can be made is that the extended manifold is several copies of the original part. Then the extension beyond the hyperbola of Fig. 3(d) would contain another root surface and another extension.

This work agrees in two of four dimensions with the work of Godfrey in two dimensional spacetimes. [11] It also agrees with the results Carter obtained by analytic extension[12] and by an extension similar to the present method. [13] Carter has gone on to extend the Kerr solution. [13]

XIV. RESULTS

Study of a convenient spherically symmetric stationary standard form has shown that there are incomplete geodesics approaching every root of g_{00} and that the manifold can be extended to a new sheet at every root of g_{00}. The direction in time of the incomplete geodesics approaching a root of g_{00} is given by (31). An extension is given for all cases, in the Kruskal fashion for simple roots and by symmetry for multiple roots. A simple invariant physical distinction between the odd roots and the even is found, and a topological distinction between extensions of simple roots and extensions of multiple roots.

Kinks, trapped surface manifolds without singularities, are used as an example. A specific kink with positive energy density is put into standard form, then extended: Nontrivial lightcone field homotopy is found to require a change in the topology of the background manifold.

Question: Are kinks always incomplete, or does this incompleteness stem from symmetry?

*Based on the Ph.D. thesis of G. McCollum, Yeshiva University, 1975.

[1]D. Finkelstein, Phys. Rev. **110**, 965 (1958).
[2]D. Finkelstein and C.W. Misner, Ann. Phys. (N.Y.) **6**, 230 (1959).
[3]J.G. Williams and R.K.P. Zia, J. Phys. A **6**, 1 (1973), compute the stress tensor of a kink and impose the condition $T^0{}_0 > 0$.
[4]A.S. Eddington, Nature **133**, 192 (1924).
[5]N. Steenrod, *The Topology of Fiber Bundles* (Princeton U.P., Princeton, N.J., 1951), p. 205.
[6]H. Flanders, *Differential Forms* (Academic, New York, 1963).
[7]C.W. Misner, J. Math. Phys. **4**, 924 (1963).
[8]R. Penrose, in *Relativity; Groups, and Topology, Les Houches* 1963, edited by DeWitt and DeWitt (Gordon and Breach, New York, 1964).
[9]M.D. Kruskal, Phys. Rev. **119**, 1743 (1960).
[10]J.C. Graves and D.R. Brill, Phys. Rev. **120**, 1507 (1960).
[11]B. Godfrey, J. Math. Phys. **12**, 606 (1971). Thank you, referee.
[12]B. Carter, Phys. Lett. **21**, 423 (1966).
[13]B. Carter, in *Black Holes, Les Houches*, 1972, edited by DeWitt and DeWitt (Gordon and Breach, New York, 1973).

Neutron transport problems in a spherical shell

D. C. Sahni

Theoretical Reactor Physics Section, Bhabha Atomic Research Centre, Bombay, India
(Received 23 October 1973; revised manuscript received 23 January 1975)

The density transform method has been extended to cover spherically symmetric transport problems in a spherical shell. The density transform is expanded in plane geometry normal modes and explicit singular integral equations are derived for the expansion coefficients. It is shown that the Green's function method, introduced by Case *et al.*, gives the same representation of total flux. The singular integral equations for the expansion coefficients are rederived using the analytic properties of some sectionally holomorphic functions introduced by the above authors.

1. INTRODUCTION

The normal mode expansion method of Case[1] has been applied to many transport problems. In principle the method can be applied to any transport problem in plane geometry by expanding the angular flux $\psi(x, \mu)$ in the singular eigenfunctions of homogeneous plane geometry Boltzmann equation. However, the expansion coefficients are defined as solutions of singular integral equations resulting from the application of boundary conditions. These integral equations can be solved only for a few special cases. Thus strictly speaking Case's method replaces the problem of solving the integro-differential transport equation by the solution of a singular integral equation. Nevertheless, Case's representation gives considerable insight into the mathematical structure of transport equation.

Although many generalizations of Case's work have been reported (including anisotropic scattering, time and energy dependence) most of them have been for the plane geometry problems. The straightforward extentions of Case's method to nonplanar geometries[2-5] have not been very successful. Mitsis,[3] obtained a set of normal modes for solutions of Boltzmann equation in spherical and cylindrical geometries whose completeness could not be proved. He, however, showed that the criticality problems of homogeneous spheres and cylinders could be reduced to the solution of certain singular integral equation. He also derived the same results by starting from the integral transport equation and the use of density transforms.[6] In this method one constructs integral transforms of the total neutron flux that satisfy the plane geometry transport equation in the transform variables. These transforms are then expanded in Case's singular eigenfunctions of plane geometry. The boundary conditions for these transforms can, in some cases, be inferred from their definition and can be used to obtain the singular integral equations that determine the expansion coefficients. The method gives a useful representation for the total neutron flux though the expansion coefficients have generally to be obtained by numerically solving the singular integral equations. This method is a consequence of the "replication property"[7] of the kernel of integral transport equation and has been exploited by Gibbs[8] for obtaining a general formulation of the method for an arbitrary convex body. In spherical and cylindrical geometries Gibbs analysis can be used only for transport problems in homogeneous spheres and cylinders though some variation in scattering properties of the medium can be allowed for if the macroscopic total cross section is held constant.[9] The general problem of particle transport in spherical and cylindrical shells requires an extension of density transform method to nonconvex bodies. This has been done only for the problem of flux distribution in an infinite moderator containing a black sphere. Sahni[10] transformed the integral transport equation to an integral equation whose kernel had the replication property. Smith[11] and Sheaks[12] have given the explicit density transforms for this problem using similar transformation.

In a recent paper Case *et al.*[13] have applied the Green's function approach to the spherically symmetric boundary value problems of one speed neutron transport theory. This results in a representation of angular flux in terms of eigenfunctions of spherically symmetric Boltzmann equation, the expansion coefficients being related to the angular distribution on the surface of the region. By using a reduction operator they transform the equations defining the expansion coefficients into singular integral equations resembling those encountered in plane geometry problems. Their results, however, are correct only for some of the interior problems (homogeneous solid spheres). Indeed the solution given by them for the exterior problems is incorrect. This can be seen most readily from their expression for the linear extrapolation distance for spherical Milne problem [Eq. (122) of their paper] which is independent of the radius of the black sphere. This is contrary to the well-known results of Davison.[14] This discrepancy stems from the use of a reduction operator which is not appropriate to this situation.

In this paper we generalize the density transform method for treating neutron transport problems in a spherical shell. In Sec. 2 we show that the usual problem of solving spherically symmetric Boltzmann equation with prescribed incoming angular distributions on the surface of a spherical shell is indeed equivalent to assuming a black medium in the region interior to the spherical shell. The only difference is the presence of uncollided flux terms in the integral equation of total neutron density. We cast these terms in a form which is suitable for defining the density transforms that are given in Sec. 3. The completeness property of plane geometry singular eigenfunctions gives a representa-

tion for the density (and hence the angular flux) while the boundary conditions for the density transforms give the singular integral equations for the determination of expansion coefficients. In Sec. 4 we examine the Green's function method and show that it gives the same representation of total flux as given in Sec. 3. Further we rederive the singular integral equations for the expansion coefficients from the analytic properties of the functions introduced by Case *et al.* in Ref. 13 and explain the discrepancy contained in their paper.

2. THEORY

We consider the one speed spherically symmetric Boltzmann equation with isotropic scattering and isotropic sources

$$\mu\frac{\partial\psi}{\partial r}+\frac{1-\mu^2}{r}\frac{\partial\psi}{\partial\mu}+\psi(r,\mu)=\frac{c}{2}\rho(r)+\tfrac{1}{2}Q(r). \tag{1}$$

The angular flux $\psi(r,\mu)$ depends only on the distance r measured in units of mean free path, and μ the cosine of the angle between the position vector $\mathbf{r}$ and the direction of motion of neutron Ω. c is the mean number of secondaries per collision and $Q(r)$ is the source density. The total flux (density) $\rho(r)$ is given by

$$\rho(r)=\int_{-1}^{1}\psi(r,\mu)\,d\mu. \tag{2}$$

The transport equation (1) has to be solved for the range $a\le r\le b$ and $-1<\mu<1$ with the prescribed incoming distributions $\psi(a,\mu)$, $\mu\in(0,1)$ and $\psi(b,\mu)$, $\mu\in(-1,0)$. Introducing the neutron path coordinates

$$\xi=r\sqrt{1-\mu^2},\quad \eta=\mu r,$$

the transport equation becomes

$$\frac{\partial\psi}{\partial\eta}+\psi(\xi,\eta)=\frac{c}{2}\rho(\sqrt{\xi^2+\eta^2})+\tfrac{1}{2}Q(\sqrt{\xi^2+\eta^2}) \tag{3a}$$

where

$$\psi(\xi,\eta)=\psi(r,\mu) \tag{3b}$$

while the domain $a\le r\le b$, $-1\le\mu\le1$ is transformed into the half annulus $\sqrt{\xi^2+\eta^2}\in(a,b)$, $\xi\ge0$. Equation (3a) can be easily integrated to give

$$\begin{aligned}\psi(r,\mu)=\psi(a,\sqrt{1-r^2(1-\mu^2)/a^2})\exp[-\mu r+\sqrt{a^2-r^2(1-\mu^2)}]\\ +\frac{1}{2}\int_a^r\frac{y[c\rho(y)+Q(y)]\,dy}{\sqrt{y^2-r^2(1-\mu^2)}}\\ \times\exp\{-[\mu r-\sqrt{y^2-r^2(1-\mu^2)}]\}\end{aligned} \tag{4a}$$

for $\sqrt{1-a^2/r^2}\le\mu\le1$ and

$$\begin{aligned}\psi(r,\mu)=\psi(b,-\sqrt{1-r^2(1-\mu^2)/b^2})\\ \times\exp\{-[\mu r+\sqrt{y^2-r^2(1-\mu^2)}]\}\\ +\frac{1}{2}\int_r^b\frac{y[c\rho(y)+Q(y)]\,dy}{\sqrt{y^2-r^2(1-\mu^2)}}\\ \times\exp\{-[\mu r+\sqrt{y^2-r^2(1-\mu^2)}]\}\\ +\Theta(\mu)\int_{r\sqrt{1-\mu^2}}^r\frac{y[c\rho(y)+Q(y)]}{\sqrt{y^2-r^2(1-\mu^2)}}\exp(-\mu r)\\ \times\cosh[\sqrt{y^2-r^2(1-\mu^2)}]\end{aligned} \tag{4b}$$

for $-1\le\mu\le\sqrt{1-a^2/r^2}$.

Integrating equation (4a) and (4b) over the intervals $(\sqrt{1-a^2/r^2},1)$ and $(-1,\sqrt{1-a^2/r^2})$ and adding, we find that $r\rho(r)$ is given by the equation

$$\begin{aligned}r\rho(r)=\frac{1}{2}\int_a^b y[c\rho(y)+Q(y)]\,dy\int_0^1\frac{d\nu}{\nu}\left[\exp(-|r-y|/\nu)-\exp\left(-\frac{1}{\nu}(\sqrt{r^2-a^2}+\sqrt{y^2-a^2})\right)\right]\\ +r\int_{\sqrt{1-a^2/r^2}}^1\psi(a,\sqrt{1-r^2(1-\mu^2)/a^2})\exp[-\mu r+\sqrt{a^2-r^2(1-\mu^2)}]\,d\mu\\ +\int_{-1}^{\sqrt{1-a^2/r^2}}\psi(b,-\sqrt{1-r^2(1-\mu^2)/b^2})\exp\{-[\mu r+\sqrt{b^2-r^2(1-\mu^2)}]\}\,d\mu.\end{aligned} \tag{5}$$

Equation (5) is in fact the integral form of transport equation

$$\rho(\mathbf{r})=\int_{V(\mathbf{r})}[c\rho(\mathbf{r}')+Q(\mathbf{r}')]\frac{\exp(-|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|^2}\,d\mathbf{r}'+\int_S dS\int\psi_{\text{in}}(\mathbf{r}_S,\Omega)\delta\left(\Omega-\frac{\mathbf{r}-\mathbf{r}_S}{|\mathbf{r}-\mathbf{r}_S|}\right)\frac{\Omega\cdot\mathbf{n}_1(\mathbf{r}_S)\exp(-|\mathbf{r}-\mathbf{r}_S|)}{|\mathbf{r}-\mathbf{r}_S|^2}\,d\Omega \tag{6}$$

where $V(\mathbf{r})$ is the volume of the spherical shell, $a\le r\le b$ which is not shielded by the surface $r=a$ from $\mathbf{r}$. The surface S consists of the inner surface $r=a$ and the outer surface $r=b$. The integration in the first term of (6) can be extended to the volume V of the shell (or even the sphere $r\le b$) if the region $r\le a$ is assumed to be a black medium. Thus the usual transport problem (1) with prescribed incoming distributions implies the existence of a black medium in the region interior to the surface $r=a$. Further we note that the incoming distributions are not always specified *a priori* but may be given by the boundary conditions at the interface of two media. In that case they serve to relate the unknown coefficients (introduced later) in the two adjoining media. However in what follows we shall regard these incoming distributions as known.

Using the formula[15]

$$\exp(-\sqrt{r^2-a^2}/\nu)=\int_0^{\nu}\exp(-r/\mu')\frac{d\mu'}{\mu'^2}\left(\nu^2\delta(\mu'-\nu)+\frac{a}{\nu}\frac{I_1(a\sqrt{1/\mu'^2-1/\nu^2})}{\sqrt{1/\mu'^2-1/\nu^2}}\right), \tag{7}$$

Eq. (5) can be written as

$$\begin{aligned}r\rho(r)=&\frac{1}{2}\int_0^1 d\nu\int_a^b y[c\rho(y)+Q(y)]\,dy\left[\frac{1}{\nu}\exp(-|r-y|/\nu)-\exp(-r/\nu)K(\nu,y)\right]\\&+r\int_{\sqrt{1-a^2/r^2}}^1\psi(a,\sqrt{1-r^2(1-\mu^2)/a^2})\exp[-\mu r+\sqrt{a^2-r^2(1-\mu^2)}]\,d\mu\\&+r\int_{-1}^{\sqrt{1-a^2/r^2}}\psi(b,-\sqrt{1-r^2(1-\mu^2)/b^2})\exp\{-[\mu r+\sqrt{b^2-r^2(1-\mu^2)}]\,d\mu,\end{aligned} \tag{8}$$

where

$$K(\nu,y)=\frac{1}{\nu^2}\int_{\nu}^1\frac{d\mu'}{\mu'}\left(\nu^2\delta(\nu-\mu')+\frac{a}{\mu'}\frac{I_1(a\sqrt{1/\nu'^2-1/\mu^2})}{\sqrt{1/\nu'^2-1/\mu^2}}\right)\times\exp(-\sqrt{y^2-a^2}/\mu'). \tag{9}$$

We now consider the last two terms in rhs of (8) and cast them in a suitable form so as to define the density transforms. The second term can be written as

$$r\int_{\sqrt{1-a^2/r^2}}^1\psi(a,\sqrt{1-r^2(1-\mu^2)/a^2})\times\exp\{-[\mu r-\sqrt{a^2-r^2(1-\mu^2)}]\}\,d\mu$$

$$=a^2\int_0^1\psi(a,\mu')\frac{\exp[a\mu'-\sqrt{r^2-a^2(1-\mu'^2)}]}{\sqrt{r^2-a^2(1-\mu'^2)}}\,\mu'\,d\mu'. \tag{10}$$

Again using the formula[15]

$$\frac{\exp[-\sqrt{r^2-a^2(1-\mu^2)}]}{\sqrt{r^2-a^2(1-\mu^2)}}=\int_0^1\exp(-r/\nu)I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)\frac{d\nu}{\nu^2}, \tag{11}$$

it can be written as

$$r\int_{\sqrt{1-a^2/r^2}}^1\psi(a,\sqrt{1-r^2(1-\mu^2)/a^2})\exp[-\mu r+\sqrt{a^2-r^2(1-\mu^2)}]\,d\mu=\int_0^1\exp(-r/\nu)\frac{d\nu}{\nu^2}\int_0^1 a^2\psi(a,\mu')\exp(a\mu')I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)\mu'\,d\mu'. \tag{12}$$

Now we write the last term on rhs of (9) as

$$r\int_{-1}^{\sqrt{1-a^2/r^2}}\psi(b,-\sqrt{1-r^2(1-\mu^2)/b^2})\times\exp\{-[\mu r+\sqrt{b^2-r^2(1-\mu^2)}]\}=\mathcal{J}_1-\mathcal{J}_2 \tag{13}$$

where $\mathcal{J}_1$ gives the total flux in a purely capturing, source-free solid sphere of radius b due to an incident angular flux $\psi(b,-\mu)$, $\mu\in(0,1)$. Thus

$$\mathcal{J}_1=r\int_{-1}^1\psi(b,-\sqrt{1-r^2(1-\mu^2)/b^2})\times\exp\{-[\mu r+\sqrt{b^2-r^2(1-\mu^2)}]\}\,d\mu \tag{14}$$

and

$$\mathcal{J}_2=r\int_{\sqrt{1-a^2/r^2}}^1\psi(b,-\sqrt{1-r^2(1-\mu^2)/b^2})\times\exp\{-[\mu r+\sqrt{b^2-r^2(1-\mu^2)}]\}\,d\mu. \tag{15}$$

On using the transformations used in obtaining Eq. (12), we get

$$\mathcal{J}_2=\int_0^1\exp(-r/\nu)\frac{d\nu}{\nu^2}\int_{\sqrt{1-a^2/b^2}}^1 b^2\psi(b,-\mu')\exp(-b\mu')\times I_0\left(\frac{b}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)\mu'\,d\mu'. \tag{16}$$

In order to write $\mathcal{J}_1$ in a convenient form let us assume that the angular flux $\psi(b,-\mu)$, $\mu\in(0,1)$ is due to an isotropic, spherically symmetric source density $\chi(r)$ in the region $r\geq b$. This source density can always be chosen for a given $\psi(b,-\mu)$ and if $\psi(b,-\mu)$ is a polynomial of μ of degree n then $\chi(r)$ will also be a polynomial of degree n.

Thus if

$$\psi(b,-\mu)=\int_0^{\infty}\exp(-R)\chi(\sqrt{R^2+b^2+2bR\mu})\,dR, \tag{17}$$

then we have (Appendix A)

$$\mathcal{J}_1=\int_0^1\frac{d\nu}{\nu}[\exp(r/\nu)-\exp(-r/\nu)]\int_b^{\infty}\chi(r')r'\times\exp(-r'/\nu)\,dr'. \tag{18}$$

Hence Eq. (8) gives

$$\begin{aligned}r\rho(r)=&\frac{1}{2}\int_0^1 d\nu\int_a^b y[c\rho(y)+Q(y)]\,dy\left[\frac{1}{\nu}\exp(-|r-z|/\nu)-\exp(-r/\nu)K(\nu,y)\right]\\&+a^2\int_0^1\exp(-r/\nu)\frac{d\nu}{\nu^2}\int_0^1\mu'\psi(a,\mu')\exp(a\mu')I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'\end{aligned}$$

$$+b^2\int_0^1 \exp(-r/\nu)\frac{d\nu}{\nu^2}\int_{\sqrt{1-a^2/b^2}}^1 \psi(b,\mu')\mu'\exp(-b\mu')I_0\left(\frac{b}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'$$
$$+\int_0^1 \frac{d\nu}{\nu}[\exp(r/\nu)-\exp(-r/\nu)]\int_b^\infty \chi(r')r'\exp(-r'/\nu)\,dr'. \tag{19}$$

3. DENSITY TRANSFORMS

We are now in a position to define the density transforms $\phi(r,\mu)$ that are solutions of plane geometry transport equation. We have

$$\phi(r,\mu)\exp(r/\mu)=\frac{1}{2\mu}\int_a^r y\exp(y/\mu)[c\rho(y)+Q(y)]dy-\frac{1}{\mu}\int_b^\infty \chi(r')r'\exp(-r'/\mu)\,dr'$$
$$-\frac{1}{2}\int_a^b K(\mu,y)y[c\rho(y)+Q(y)]dy+a^2\int_0^1 \psi(a,\mu')\mu'\exp(a\mu')I_0\frac{a}{\mu}\sqrt{1-\mu^2}\sqrt{1-\mu'^2}\,d\mu'$$
$$-b^2\int_{\sqrt{1-a^2/b^2}}^1 \psi(b,\mu')\mu'\exp(-b\mu')I_0\frac{b}{\mu}\sqrt{1-\mu^2}\sqrt{1-\mu'^2}\,d\mu \tag{20}$$

for $\mu\in(0,1)$ and

$$\phi(r,\mu)\exp(r/\mu)=-\frac{1}{2\mu}\int_r^b y[c\rho(y)+Q(y)]\exp(y/\mu)\,dy$$
$$-\frac{1}{\mu}\int_b^\infty \chi(r')r'\exp(r'/\mu)\,dr' \tag{21}$$

for $\mu\in(-1,0)$. Then from Eq. (19) we have

$$r\rho(r)=\int_{-1}^1 \phi(r,\mu)\,d\mu. \tag{22}$$

Further, from Eqs. (20), (21), and (22) it is clear that $\phi(r,\mu)$ satisfies the plane geometry transport equation, namely

$$\mu\frac{\partial\phi}{\partial r}+\phi(r,\mu)=\frac{c}{2}\int_{-1}^1 \phi(r,\mu)\,d\mu+\frac{rQ(r)}{2}. \tag{23}$$

The normal mode solutions of the homogeneous part of Eq. (23) are well known. It has been shown by Case[1] that the normal modes

$$\phi_\nu(r,\mu)=\exp(-r/\nu)\Phi_\nu(\mu) \tag{24}$$

where

$$\Phi_\nu(\mu)=P\frac{c\nu}{2(\nu-\mu)}+\lambda(\nu)\delta(\nu-\mu) \tag{24'}$$

for $\nu\in(-1,1)$ and $\lambda(\nu)$ given by

$$\lambda(\nu)=1-c\nu\tanh^{-1}\nu$$

and the discrete mode

$$\phi_\pm(r,\mu)=\exp(\mp r/\nu_0)\Phi_{0\pm}(\mu)=\exp(\mp r/\nu_0)\frac{c\nu_0}{2(\nu_0\mp\mu)} \tag{24''}$$

where ν_0 satisfies the equation

$$\frac{c\nu_0}{2}\log\frac{\nu_0+1}{\nu_0-1}=1$$

have partial range and full range completeness properties. They also satisfy half and full range orthogonality relations.

The general solution of Eq. (23) can be obtained by expanding the transform $\phi(r,\mu)$ in terms of $\Phi_\nu(\mu)$ over the full range. This general solution will contain a particular integral and a complementary function which is the solution of homogeneous part of Eq. (23). The particular integral can be obtained by using the Fourier transform method.[13] Thus

$$\phi_{P.I}(r,\mu)=\frac{1}{2\pi}\int_{-\infty}^\infty \frac{\bar{Q}(k)\exp(ikr)}{2(1+ik\mu)}\frac{dk}{\Lambda(k)} \tag{25}$$

where

$$\Lambda(k)=1-\frac{c}{2}\int_{-1}^1 \frac{d\mu}{1+ik\mu} \tag{25'}$$

and

$$\bar{Q}(k)=\int_{-\infty}^\infty rQ(r)\exp(-ikr)\,dr. \tag{25''}$$

We assume that $Q(r)$ vanishes for $r\notin(a,b)$. For $r\in(a,b)$ Eqs. (25) and (25') give

$$\phi_{P.I.}(r,\mu)=\frac{\exp(-r/\nu_0)\Phi_{0+}(\mu)}{2N_{0+}}\int_a^r r'Q(r')\exp(r'/\nu_0)\,dr'-\frac{\exp(r/\nu_0)\Phi_{0-}(\mu)}{2N_{0-}}\int_r^b r'Q(r')\exp(-r'/\nu_0)\,dr'$$
$$+\int_0^1 \frac{\exp(-r/\nu)\Phi_\nu(\mu)}{2N_\nu}d\nu\int_a^r r'Q(r')\exp(r'/\nu)\,dr'-\int_{-1}^0 \frac{\exp(-r/\nu)\Phi_\nu(\mu)}{2N_\nu}d\nu\int_r^b r'Q(r')\exp(r'/\nu)\,dr' \tag{26}$$

where

$$N_{0\pm}=\int_{-1}^1 \mu\Phi_{0\pm}^2(\mu)\,d\mu=\pm\frac{c\nu_0^3}{2}\left(\frac{c}{\nu_0^2-1}-\frac{1}{\nu_0^2}\right) \tag{27a}$$

and N_ν is the normalization integral for the singular eigenfunctions $\Phi_\nu(\mu)$, i.e.,

$$\int_{-1}^1 \mu\Phi_\nu(\mu)\Phi_{\nu'}(\mu)\,d\mu=N_\nu\delta(\nu-\nu'),\qquad N_\nu=\nu\,\lambda^2(\nu)+\frac{c^2\pi^2\nu^2}{4}. \tag{27b}$$

The general solution of Eq. (23) therefore can be written as

$$\phi(r,\mu)=\left[a_{0+}+\frac{1}{2N_{0+}}\left(\int_a^r r'Q(r')\exp(r'/\nu_0)\,dr'-\int_a^b r'Q(r')\exp(-r'/\nu_0)\,dr'\right)\right]\exp(-r/\nu_0)\Phi_{0+}(\mu)$$
$$+\left(a_{0-}-\frac{1}{2N_{0-}}\int_r^b r'Q(r')\exp(-r'/\nu_0)\,dr'\right)\exp(r/\nu_0)\Phi_{0-}(\mu)$$
$$+\int_0^1\exp(-r/\nu)\Phi_\nu(\mu)\,d\nu\left[A(\nu)+\frac{1}{2N_\nu}\left(\int_a^r r'Q(r')\exp(+r'/\nu)\,dr'-\int_a^b r'Q(r')\exp(-r'/\nu)\,dr'\right)\right]$$
$$+\int_{-1}^0\exp(-r/\nu)\Phi_\nu(\mu)\,d\nu\left(A(\nu)-\frac{1}{2N_\nu}\int_r^b r'Q(r')\exp(r'/\nu)\,dr'\right). \tag{28}$$

Integrating Eq. (28) we get a representation for the total neutron flux

$$r\rho(r)=a_{0+}\exp(-r/\nu_0)+a_{0-}\exp(r/\nu_0)+\int_{-1}^1 A(\nu)\exp(-r/\nu)\,d\nu+\frac{1}{2N_{0+}}\int_a^b r'Q(r')\,dr'[\exp(-|r-r'|/\nu_0-\exp(-(r+r')/\nu_0)]$$
$$+\int_0^1\frac{d\nu}{2N_\nu}\int_a^b r'Q(r')\,dr'[\exp(-|r-r'|/\nu)-\exp(-(r+r')/\nu)]. \tag{29}$$

The unknown coefficients $a_{0\pm}$; $A(\nu)$ can be determined from the following boundary conditions for $\phi(r,\mu)$ which can be inferred from the definition of density transforms:

(i) $\mu\in(-1,0)$

$$\phi(b,\mu)=-\frac{\exp(-b/\mu)}{\mu}\int_b^\infty\chi(r')r'\exp(r'/\mu)\,dr' \tag{30a}$$

(ii) $\mu\in(0,1)$

$$\phi(a,\mu)\exp(a/\mu)=-\frac{1}{\mu}\int_b^\infty\chi(r')r'\exp(-r'/\mu)\,dr'-\frac{1}{2}\int_a^b y[c\rho(y)+Q(y)]K(\mu,y)\,dy$$
$$+\frac{a^2}{\mu^2}\int_0^1\psi(a,\mu')\mu'\exp(+a\mu')I_0\left(\frac{a}{\mu}\sqrt{1-\mu^2}\sqrt{1-\mu'^2}\right)d\mu'$$
$$-\frac{b^2}{\mu^2}\int_{\sqrt{1-a^2/b^2}}^1\psi(b,-\mu')\mu'\exp(-b\mu')I_0\left(\frac{b}{\mu}\sqrt{1-\mu^2}\sqrt{1-\mu'^2}\right)d\mu'. \tag{30b}$$

These boundary conditions determine the singular integral equations for the determination of the coefficients $a_{0\pm}$ and $A(\nu)$, $\nu\in(-1,1)$. These same equations can also be derived by substituting the form (29) directly in the Eq. (19). They are

$$\lambda(\nu)A(\nu)+\frac{c}{2}\int_{-1}^1\frac{\nu'A(\nu')\exp[-a(1/\nu'-1/\nu)]}{\nu'-\nu}\,d\nu'+\frac{c}{2}\int_{-1}^1 A(\nu')\,d\nu'\int_a^b K(\nu,y)\exp(-y/\nu')\,dy$$
$$=-\left(a_{0+}-\frac{1}{2N_{0+}}\int_a^b r'Q(r')\exp(-r'/\nu_0)\,dr'\right)\exp[-a(1/\nu_0-1/\nu)]\,\Phi_{0+}(\nu)$$
$$-\left(a_{0-}+\frac{1}{2N_{0+}}\int_a^b r'Q(r')\exp(-r'/\nu_0)\,dr'\right)\exp[a(1/\nu_0+1/\nu)]\Phi_{0-}(\nu)$$
$$+\int_0^1\frac{d\nu'}{2N_{\nu'}}\int_a^b r'Q(r')\exp(-r'/\nu')\,dr'\{\exp[a(1/\nu-1/\nu')]\Phi_{\nu'}(\nu)-\exp[a(1/\nu+1/\nu')]\Phi_{-\nu'}(\nu)\}$$
$$+\frac{a^2}{\nu^2}\int_0^1\psi(a,\mu')\exp(a\mu')\mu'I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'-\frac{1}{\nu}\int_b^\infty\chi(r')r'\exp(-r'/\nu)\,dr'$$
$$-\frac{b^2}{\nu^2}\int_{\sqrt{1-a^2/b^2}}^1\psi(b,-\mu')\exp(-b\mu')\mu'I_0\left(\frac{b}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'$$
$$-\frac{1}{2}\int_a^b K(\nu,y)\,dy\left(c[a_{0+}\exp(-y/\nu_0)+a_{0-}\exp(y/\nu_0)]+\frac{c}{2N_{0+}}\int_a^b r'Q(r')\,dr'\{\exp(-|y-r'|/\nu_0)-\exp[-(y+r')/\nu_0]\}\right.$$
$$\left.+yQ(y)+\int_0^1\frac{c\,d\nu'}{2N_{\nu'}}\int_a^b r'Q(r')\,dr'\{\exp(-|y-r'|/\nu')-\exp[-(y+r')/\nu']\}\right) \tag{31}$$

for $\nu\in(0,1)$ and in the range $\nu\in(-1,0)$

$$\lambda(\nu)A(\nu)+\frac{c}{2}\int_{-1}^1\frac{\nu'A(\nu')\exp[-b(1/\nu'-1/\nu)]}{\nu'-\nu}\,d\nu'$$
$$=-\left(a_{0+}+\frac{1}{N_{0+}}\int_a^b r'Q(r')\sinh(r'/\nu_0)\,dr'\right)\exp[-b(1/\nu_0-1/\nu)]\Phi_{0+}(\nu)-a_{0-}\exp[b(1/\nu+1/\nu_0)]\Phi_{0-}(\nu)$$

$$-\frac{1}{\nu}\int_b^\infty \chi(r')r'\exp(r'/\nu)\,dr' - \int_0^1 \frac{d\nu'}{N_{\nu'}}\exp[-b(1/\nu'-1/\nu)]\Phi_{\nu'}(\nu)\int_a^b r'Q(r')\sinh(r'/\nu')\,dr'. \tag{32}$$

Solution of singular integral equations (31) and (32) gives the expansion coefficients of the density transform. We notice that the kernels of these singular integral equations contain complicated function like $K(\nu, y)$. Same function was encountered in the treatment of black sphere problem.[10-12] It is known[10] that for actual numerical solution of these equations only first two or three terms need be retained in the series expansion of the Bessel function $I_1(x)$ occurring in the expression for $K(\nu, y)$. This greatly simplifies the equations and also yields excellent numerical results. The free terms, i. e., the rhs of Eqs. (31) and (32) contain two arbitrary constants a_{0+} and a_{0-} which are the coefficients of the discrete modes. The solutions of Eqs. (31) and (32) do not exist for all values of these constants but only for one definite value for them. Thus Eqs. (31) and (32) determine all the expansion coefficients $a_{0\pm}$ and $A(\nu) \in (-1, 1)$.

In two special cases of interior (when $a = 0$) and exterior (when $b \to \infty$) problems these equations can be reduced to only one equation over the interval $(0, 1)$. Thus for interior problem we have

$$K(\nu, y) = \frac{1}{\nu}\exp(-y/\nu),$$

$$\phi(0, -\mu) = -\phi(0, \mu) \Rightarrow a_{0+} = -a_{0-}, \quad A(-\nu) = -A(\nu).$$

Both Eqs. (31) and (32) give for $\nu \in (0, 1)$

$$\lambda(\nu)A_<(\nu)\exp(b/\nu) + \frac{c}{2}\int_0^1 \nu' A_<(\nu')\exp(b/\nu')\left(\frac{1}{\nu'-\nu} - \frac{\exp(-2b/\nu')}{\nu'+\nu}\right)d\nu'$$

$$= \left(a_{0+}^< + \frac{1}{N_{0+}}\int_0^b r'Q(r')\sinh(r'/\nu_0)\,dr'\right)\exp(-b/\nu_0)\Phi_{0-}(\nu) - a_{0+}^<\exp(b/\nu_0)\Phi_{0+} - \frac{\exp(b/\nu)}{\nu}\int_b^\infty r'\chi(r')\exp(-r'/\nu)\,dr'$$

$$+ \int_0^1 \exp(-b/\nu')\Phi_{-\nu'}(\nu)\frac{d\nu'}{N_{\nu'}}\int_0^b r'Q(r')\sinh(r'/\nu')\,dr'. \tag{33}$$

For exterior problems, since the total flux $\rho(r)$ should approach the form given by diffusion theory for $r \gg a$, we have

$$a_{0-} = 0, \quad A(\nu) = 0 \ \forall\ \nu \in (-1, 0)$$

and for $\nu \in (0, 1)$ the function $A(\nu)$ is the solution of the equation

$$\lambda(\nu)A_>(\nu) + \frac{c}{2}\int_0^1 \frac{\nu' A_>(\nu')\exp[-a(1/\nu'-1/\nu)]}{\nu'-\nu}\,d\nu' + \frac{c}{2}\int_0^1 A_>(\nu')\,d\nu'\int_a^\infty K(\nu, y)\exp(-y/\nu')\,dy$$

$$= -a_{0+}^>\exp[-a(1/\nu_0 - 1/\nu)]\Phi_{0+}(\nu) + \frac{a^2}{\nu^2}\int_0^1 \mu'\psi(a, \mu')\exp(a\mu')I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'$$

$$+ \frac{\exp(a/\nu)}{2N_{0+}}\int_a^\infty r'Q(r')\exp(-r'/\nu_0)\,dr'[\exp(-a/\nu_0)\Phi_{0+}(\nu) - \exp(a/\nu_0)\Phi_{0-}(\nu)]$$

$$+ \exp(a/\nu)\int_0^1 \frac{d\nu'}{2N_{\nu'}}[\exp(-a/\nu')\Phi_{\nu'}(\nu) - \exp(a/\nu')\Phi_{-\nu'}(\nu)]\int_a^\infty r'Q(r')\exp(-r'/\nu')\,dr'$$

$$- \frac{1}{2}\int_a^\infty K(\nu, y)\,dy\left(ca_{0+}\exp(-y/\nu_0) + \frac{c}{2N_{0+}}\int_a^\infty r'Q(r')\,dr'\{\exp(-|r'-y|/\nu_0) - \exp[-(r'+y)/\nu_0]\}\right.$$

$$\left. + yQ(y) + \int_0^1 \frac{c\,d\nu'}{2N_{\nu'}}\int_a^\infty r'Q(r')\,dr'\{\exp(-|r'-y|/\nu') - \exp[-(r'+y)/\nu']\}\right). \tag{34}$$

4. GREEN'S FUNCTION METHOD

In this section we consider the Green's function approach[13] to the spherically symmetric neutron transport problems. In the notation of Ref. 13 the solution of one speed Boltzmann equation with isotropic scattering, i. e.,

$$(1 + \Omega\cdot\nabla)\psi(\mathbf{r}, \Omega) = \frac{c}{4\pi}\int \psi(\mathbf{r}, \Omega)\,d\Omega + Q(\mathbf{r}, \Omega)$$

$$= \frac{c}{4\pi}\rho(\mathbf{r}) + Q(\mathbf{r}, \Omega) \tag{35}$$

can be obtained by using the infinite medium Green function which is the solution of the equation

$$(1 + \Omega\cdot\nabla)G(\mathbf{r}, \Omega; \mathbf{r}_0, \Omega_0)$$

$$= \frac{c}{4\pi}\int G(\mathbf{r}, \Omega'; \mathbf{r}_0, \Omega_0)\,d\Omega' + \delta(\mathbf{r}-\mathbf{r}_0)\delta_2(\Omega\cdot\Omega_0). \tag{36}$$

The Green's function can be obtained by using Fourier transform; thus

$$G(\mathbf{r}, \Omega; \mathbf{r}_0, \Omega_0)$$

$$= \frac{1}{8\pi^3}\int \frac{\exp[i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0)]}{1 + i\mathbf{k}\cdot\Omega}\left(\delta(\Omega\cdot\Omega_0) + \frac{c}{4\pi}\frac{1}{1+i\mathbf{k}\cdot\Omega_0}\frac{1}{\Lambda(k)}\right) \tag{37}$$

where $\Lambda(k)$ has been defined in (25a). Thus

$$\psi(\mathbf{r}, \Omega) = \psi_q(\mathbf{r}, \Omega) + \psi_0(\mathbf{r}, \Omega) + \psi_c(\mathbf{r}, \Omega) \tag{38}$$

where

$$\psi_q(\mathbf{r}, \Omega) = \int d\Omega' \int_V d\mathbf{r}' Q(\mathbf{r}', \Omega') G(\mathbf{r}, \Omega; \mathbf{r}', \Omega') \tag{39}$$

and

$$\begin{aligned}\psi_0(\mathbf{r}, \Omega) &= \frac{1}{8\pi^3}\int \frac{\exp(i\mathbf{k}\cdot\mathbf{r})\,d\mathbf{k}}{1 + i\mathbf{k}\cdot\Omega} \int_S \Omega \cdot \mathbf{n}_1(\mathbf{r}_S)\psi(\mathbf{r}_S, \Omega) \\ &\quad \times \exp(-i\mathbf{k}\cdot\mathbf{r}_S)\,dS, \\ \psi_c(\mathbf{r}, \Omega) &= \frac{1}{8\pi^3}\int \frac{\exp(i\mathbf{k}\cdot\mathbf{r})\,d\mathbf{k}}{\Lambda(k)(1 + i\mathbf{k}\cdot\Omega)} \int\int_S \frac{c\Omega'\cdot\mathbf{n}_1(\mathbf{r}_S)\psi(\mathbf{r}_S, \Omega')}{4\pi(1 + i\mathbf{k}\cdot\Omega')} \\ &\quad \times \exp(-i\mathbf{k}\cdot\mathbf{r}_S)\,dS\,d\Omega'.\end{aligned} \tag{40}$$

$\mathbf{n}_1(\mathbf{r}_S)$ is the normal to the surface S at the point $\mathbf{r}_S$ on the surface, directed into the volume V in which the transport problem is being solved. The volume V is bounded by the surface S. For spherically symmetric problems with isotropic sources we have

$$\psi(\mathbf{r}, \Omega) = \frac{\psi(r, \mu)}{2\pi}, \quad Q(\mathbf{r}, \Omega) = \frac{Q(r)}{4\pi}, \quad \rho(\mathbf{r}) = \rho(r)$$

and $\psi(r, \mu)$ satisfies Eq. (1). Equations (38), (39), and (40) give

$$\begin{aligned}\psi_c(r, \mu) &= \frac{c}{8\pi}\int_{-\infty}^{\infty} \frac{k^2 T(k)}{\Lambda(k)} I(k, r, \mu)\,dk, \\ \psi_0(r, \mu) &= \frac{1}{4\pi}\int_{-\infty}^{\infty} k^2\,dk \int_{-1}^{1} \frac{\exp(ikrt\mu) J_0(kr\sqrt{1-\mu^2}\sqrt{1-t^2})}{1 + ikt} \\ &\quad \times H(k, t)\,dt, \\ \psi_q(r, \mu) &= \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{k^2\,dk}{\Lambda(k)} \int_a^b r'^2 Q(r') \frac{\sin kr'}{kr'} I(k, r, \mu)\,dr'\end{aligned} \tag{41}$$

where

$$\begin{aligned}I(k, r, \mu) &= \int_{-1}^{1} \frac{\exp(ikrt\mu)}{1 + ikt} J_0(kr\sqrt{1-\mu^2}\sqrt{1-t^2})\,dt, \\ H(k, t) &= \epsilon r_S^2 \int_{-1}^{1} \mu' \psi(r_S, \mu') \exp(-ikr_S t\mu') \\ &\quad \times J_0(kr_S\sqrt{1-\mu'^2}\sqrt{1-t^2})\,d\mu' \\ T(k) &= \int_{-1}^{1} \frac{H(k, t)\,dt}{1 + ikt}.\end{aligned} \tag{42}$$

The functions $T(k)$, $\Lambda(k)$, and $I(k, r, \mu)$ are all sectionally holomarphic in the complex k plane with the branch cuts $(-i\infty, i) \cup (i, i\infty)$. By taking the contour of integration along the branch cut from $(i, i\infty)$, Case *et al.* have given the following expressions for $\psi(r, \mu)$ for interior and exterior problems. ϵ equals $+1$ for exterior and -1 for interior problems.

1. Interior problems

$$\begin{aligned}\psi(r, \mu) &= \psi_q(r, \mu) + \Gamma_<^0(1/\nu_0)\phi(i/\nu_0, r, \mu) \\ &\quad + \int_0^1 \Gamma_<(i/\nu)\phi(i/\nu, r, \mu)\,d\nu/\nu\end{aligned} \tag{43}$$

where

$$\begin{aligned}\phi(i/\nu, r, \mu) &= \frac{2}{c}\int_{-1}^{1} \left(P\frac{c\nu}{2(\nu - t)} + \lambda(\nu)\delta(\nu - t)\right) \\ &\quad \times \exp[-(r/\nu)\mu t] I_0\left(\frac{r}{\nu}\sqrt{1-\mu^2}\sqrt{1-t^2}\right) dt \\ \Gamma_<(i/\nu) &= \frac{ic}{8\pi\nu^3}\left(\frac{T_1^+(i/\nu)}{\Lambda^+(i/\nu)} - \frac{T_1^-(i/\nu)}{\Lambda^-(i/\nu)}\right),\end{aligned}$$

$$\Gamma_<^0(i/\nu_0) = -\frac{icT_1(i/\nu_0)}{4\nu_0^2\Lambda'(i/\nu_0)}. \tag{44}$$

The function $T_1(k)$ is that part of $T(k)$ which vanishes as $k \to \infty$ in the upper half of complex k plane ($\mathrm{Im}k > 0$). It is related to the angular distribution $\psi(b, \mu)$ of neutrons on the surface $r = r_S = b$ of the solid sphere $r \le b$ by the relation

$$\begin{aligned}T_1(k) &= -b^2\int_{-1}^{1} \mu'\psi(b, \mu')\,d\mu' \sum_{n=0}^{\infty}(-i)^n(2n+1)P_n(\mu')\xi_n^{(1)}(kb) \\ &\quad \times \int_{-1}^{1} \frac{P_n(t)\,dt}{1 + ikt},\end{aligned} \tag{45}$$

$\xi_n^{(1)}(kb)$ being the Hankel functions of first kind. On integrating Eq. (43) we get the same representation for $\rho(r)$ as given by Eq. (29) [with $a_{0-} = -a_{0+}$; $A(-\nu) = -A(\nu)$], and

$$A_<(\nu) = -\frac{2}{c}\Gamma_<(i/\nu), \quad a_{0+}^< = -\frac{2\nu_0}{c}\Gamma_<^0(1/\nu_0). \tag{46}$$

2. Exterior problem

For exterior problems $\psi(r, \mu)$ is given by the expression

$$\begin{aligned}\psi(r, \mu) &= \psi_q(r, \mu) + \Gamma_>^0(1/\nu_0)Z(i/\nu_0, r, \mu) \\ &\quad + \int_0^1 \Gamma_>(i/\nu)Z(i/\nu, r, \mu)\,d\nu/\nu\end{aligned} \tag{47}$$

where

$$\begin{aligned}Z(i/\nu, r, \mu) &= \sum_{n=0}^{\infty} i^n(2n+1)P_n(\mu)\xi_n^{(1)}\left(\frac{ir}{\nu}\right) \\ &\quad \times \left(P\int_{-1}^{1} \frac{\nu P_n(t)\,dt}{\nu - t} + 2\frac{\lambda(\nu)}{c}P_n(\nu)\right), \\ \Gamma_>(i/\nu) &= \frac{c}{8\pi i\nu^3}\left(\frac{T^-(i/\nu)}{\Lambda^-(i/\nu)} - \frac{T^+(i/\nu)}{\Lambda^+(i/\nu)}\right), \\ \Gamma_>^0(1/\nu_0) &= -\frac{icT(i/\nu_0)}{4\nu_0^2\Lambda'(i/\nu_0)}.\end{aligned} \tag{48}$$

Equation (48) also gives the representation for $\rho(r)$ which is identical to the one obtained for exterior problems [$a_{0-} = A(-\nu) = 0$], and

$$A_>(\nu) = -\frac{4}{c}\Gamma_>(i/\nu), \quad a_{0+}^> = -\frac{4\nu_0}{c}\Gamma_>^0(1/\nu_0).$$

The expansion coefficients $a_{0+}^>$ and $A_<(\nu)$ [$a_{0+}^>$ and $A_>(\nu)$] depend upon the surface flux $\psi(b, \mu)$ [$\psi(a, \mu)$ for exterior problem]. On taking the limit as $r \to b$ ($r \to a$), we get regular integral equations for the determination of these coefficients. In order to derive the equations with singular kernels Case *et al.* introduce the reduction operator

$$\begin{aligned}\theta &\equiv \int_{-\infty}^{\infty} r'\,dr' K(r - r', \mu)\left(1 + \mu\frac{\partial}{\partial r'} + \frac{1-\mu^2}{r'}\frac{\partial}{\partial\mu}\right), \\ K(r - r', \mu) &= \frac{\exp[-(r - r')/\mu]}{\mu}\,[\Theta(r - r')\Theta(\mu) \\ &\quad - \Theta(r' - r)\Theta(-\mu)].\end{aligned}$$

Applying this operator for $\mu < 0$ ($\mu > 0$ for exterior problem) on rhs of (38) gives the expressions involving the expansion coefficients and planar eigen functions. However, it is not easy to find the result of applying this operator on lhs of (38). Case *et al.* have equated it

to the result obtained by applying the operator to $\psi(r,\mu)\Theta(b-r)$ $[\psi(r,\mu)\Theta(r-a)]$. The resulting equations are therefore incorrect. We can derive the correct singular integral equations as follows:

1. Interior problem

$$a_{0+}^{<} = \frac{iT_1(i/\nu_0)}{2\nu_0\Lambda'(i/\nu_0)} \tag{49}$$

and

$$A_<(\nu) = \frac{i}{4\pi\nu^3}\left(\frac{T_1^-(i/\nu)}{\Lambda^-(i/\nu)} - \frac{T_1^+(i/\nu)}{\Lambda^+(i/\nu)}\right). \tag{50}$$

Let us extend the definition of $A_<(\nu)$ to the interval $(-1,1)$ as given by the relation (50) and also define

$$a_{0-}^{<} = iT_1(-i/\nu_0)/2\nu_0\Lambda'(-i/\nu_0). \tag{49'}$$

Now define the functions $F_<(k)$ and $\varphi_<(k)$ by the relation

$$F_<(k) = \frac{2i}{k}\left(\int_{-1}^{1}\frac{\nu' A_<(\nu')\exp(b/\nu')}{1+ik\nu'}\,d\nu' + \frac{\nu_0 a_{0+}^{<}\exp(b/\nu_0)}{1+ik\nu_0} + \frac{\nu_0 a_{0-}^{<}\exp(-b/\nu_0)}{1-ik\nu_0}\right) \tag{51}$$

and

$$\frac{1}{k}\varphi_<(k) = F_<(k) - \frac{T_1(k)}{\Lambda(k)}\exp(-ikb). \tag{52}$$

These functions have the following properties:

(i) $T_1(k)$, $\Lambda(k)$, $F_<(k)$, and $\varphi_<(k)$ have branch cuts extending from $(-i\infty,-i)\cup(i,i\infty)$. In the cut plane $kT_1(k)$ is analytic everywhere while $\Lambda(k)$ and $kF_<(k)$ have simple poles at $\pm i/\nu_0$. It can be seen from the definition (49) and (49') that the residue of $\varphi_<(k)$ at either of these poles vanishes. Thus $\varphi_<(k)$ is analytic in the cut plane.

(ii) Along the branch cut $(-i\infty,-i)\cup(i,i\infty)$ we have from Plemelj formulas and Eq. (50)

$$\varphi_<^+(i/\nu) - \varphi^-(i/\nu) = 0.$$

Thus $\varphi_<(k)$ is analytic in the entire complex plane.

(iii) if $k\to\infty$

$$\Lambda(k)\to 1, \quad T_1(k)\to\frac{\exp(ikb)}{k^2}, \quad F_<(k)\to 1/k^2.$$

Hence $\varphi_<(k)$ vanishes at infinity. Since $\varphi_<(k)$ is analytic in the entire complex plane it vanishes everywhere and hence

$$F_<(k)\Lambda(k) = T_1(k)\exp(-ikb) \tag{53}$$

Thus using Plemelj formulae and the relation

$$F^+\Lambda^+ - F^-\Lambda^- = \tfrac{1}{2}[\Lambda^+ + \Lambda^-][F^+ - F^-] + \tfrac{1}{2}[\Lambda^+ - \Lambda^-][F^+ + F^-]. \tag{54}$$

We get

$$\lambda(\nu)A_<(\nu)\exp(b/\nu) + \frac{c}{2}\int_{-1}^{1}\frac{\nu' A_<(\nu')\exp(b/\nu')}{\nu'-\nu}\,d\nu' = \frac{H_1(i/\nu,\nu)\exp(b/\nu)}{2\nu^2} + \frac{c\nu_0}{2}\left(\frac{a_{0+}^{<}\exp(b/\nu_0)}{\nu-\nu_0} + \frac{a_{0-}^{<}\exp(-b/\nu_0)}{\nu+\nu_0}\right). \tag{55}$$

It is shown in Appendix B that for $\nu\in(0,1)$

$$A_<(-\nu) = -A_b(\nu) - \frac{1}{N_\nu}\int_0^b r'Q(r')\sinh(r'/\nu)\,dr',$$

$$a_{0-}^{<} = a_{0+}^{<} + \frac{1}{N_{0+}}\int_0^b r'Q(r')\sinh(r'/\nu_0)\,dr', \tag{56}$$

$$\frac{H_1(i/\nu,\nu)}{2\nu^2} = -\frac{1}{\nu}\int_b^\infty \chi(r')r'\exp(-r'/\nu)\,dr'.$$

Substituting (56) in (55) we get Eq. (33).

2. Exterior problems

In this case

$$A_>(\nu) = \frac{i}{2\pi\nu^3}\left(\frac{T^-(i/\nu)}{\Lambda^-(i/\nu)} - \frac{T^+(i/\nu)}{\Lambda^+(i/\nu)}\right),$$

$$a_{0+}^{>} = iT(i/\nu_0)/\nu_0\Lambda'(i/\nu_0).$$

Let us define

$$B(\nu) = \frac{1}{4\pi}\left(\frac{i}{\nu}\right)^3\left(\frac{T_1^+(i/\nu)}{\Lambda^+(i/\nu)} - \frac{T_1^-(i/\nu)}{\Lambda^-(i/\nu)}\right),$$

$$b_{0\pm} = \frac{i}{\nu_0}\frac{T_1(\pm i/\nu_0)}{\Lambda'(\pm i/\nu_0)}. \tag{57}$$

Since $\Lambda(k)$ is an even function of k, while $T(k)$ equals $[T_1(k)+T_1(-k)]/2$, we get

$$A_>(\nu) = B(\nu) + B(-\nu),$$

$$a_{0+}^{>} = b_{0+} - b_{0-}. \tag{58}$$

Again we define $\varphi_>(k)$ by the relation

$$\frac{1}{k}\varphi_>(k) = \frac{2i}{k}\left(\int_{-1}^{1}\frac{B(\nu')\nu'\exp(a/\nu')}{1+ik\nu'}\,d\nu' + \frac{\nu_0 b_{0+}\exp(a/\nu_0)}{1+ik\nu_0} + \frac{\nu_0 b_{0-}\exp(-a/\nu_0)}{1-ik\nu_0}\right) - \frac{T_1(k)}{\Lambda(k)}\exp(-ika). \tag{59}$$

Following the same treatment as before we can show that $\varphi_>(k)$ vanishes everywhere and applying the Plemelj formulas we get

$$\lambda(\nu)B(\nu)\exp(a/\nu) + \frac{c}{2}\int_{-1}^{1}\frac{\nu' B(\nu')\exp(a/\nu')}{\nu'-\nu}\,d\nu' = \frac{H_1(i/\nu,\nu)}{2\nu^2}\exp(a/\nu) + \frac{c\nu_0}{2}\left(\frac{b_{0+}\exp(a/\nu_0)}{\nu_0-\nu} - \frac{b_{0-}\exp(-a/\nu_0)}{\nu_0+\nu}\right) \tag{60}$$

or

$$\lambda(\nu)A(\nu) + \frac{c}{2}\int_{-1}^{1}\frac{\nu' B(\nu')\exp[a(1/\nu'-1/\nu)]}{\nu'-\nu} + \frac{c}{2}\int_{-1}^{1}\frac{\nu' B(-\nu')\exp[-a(1/\nu'-1/\nu)]}{\nu'-\nu} = \frac{c\nu_0}{2}\left[b_{0+}\exp(a/\nu_0)\left(\frac{\exp(-a/\nu)}{\nu-\nu_0} - \frac{\exp(+a/\nu)}{\nu+\nu_0}\right) - b_{0-}\exp(-a/\nu_0)\left(\frac{\exp(a/\nu)}{\nu-\nu_0} - \frac{\exp(-a/\nu)}{\nu+\nu_0}\right)\right] + \frac{H(i/\nu,\nu)}{\nu^2}. \tag{61}$$

It is shown in Appendix C that for $\nu\in(0,1)$

$$B(\nu)=\frac{1}{2N_\nu}\int_a^\infty r'Q(r')\exp(-r'/\nu)\,dr',$$

$$b_{0+}=\frac{1}{2N_{0+}}\int_a^\infty r'Q(r')\exp(-r'/\nu_0)\,dr', \tag{62}$$

$$H(i/\nu,\nu)=a^2\int_0^1\mu'\psi(a,\mu')\exp(a\mu')\times I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'+\frac{1}{2}\int_a^\infty y[c\rho(y)+Q(y)]\,dy\times\left[\frac{1}{\nu}\exp(-y/\nu)-K(\nu,y)\right].$$

Using the relations (62) and (58) in Eq. (60) we get Eq. (34) for $a_{0+}^{>}$ and $A_>(\nu)_j$, $\nu>0$. Thus for both the interior and exterior problems we obtain the same singular integral equation that were derived by density transform method in Sec. 3.

5. CONCLUDING REMARKS

We have treated above the spherically symmetric one speed transport equation with isotropic scattering and sources by the density transform method and the Green's function method. The expansion coefficients introduced by both these methods are identical and they are given as the solution of Cauchy type singular integral equations, much like the plane geometry problems. Although the parallelism of the two methods has been shown only for the interior and exterior problems it holds good for the general transport problem in a spherical shell. In general these singular integral equations are not amenable to analytic solutions and numerical methods are necessary as was done for the spherical Milne problem.[10] However, some analytical methods have already been indicated by Case *et al.*[13] for interior problems, though their results for Albedo problem are in error.

ACKNOWLEDGMENT

I wish to express my thanks to Dr. S. V. Lawande for many helpful discussions.

APPENDIX A

We give here the derivation of Eq. (18). Changing to the variable

$$\mu'=\sqrt{1-r^2(1-\mu^2)/b^2}$$

in Eq. (14), we get

$$\mathcal{J}_1=2b^2\int_{\sqrt{1-r^2/b^2}}^1\frac{\psi(b,-\mu')\mu'\exp(-b\mu')}{\sqrt{r^2-b^2(1-\mu'^2)}}\times\cosh[\sqrt{r^2-b^2(1-\mu'^2)}]\,d\mu'.$$

Now we put

$$b\mu'\pm\sqrt{r^2-b^2(1-\mu'^2)}=\sqrt{r^2+b^2-2br\cos\theta}$$

and get

$$\mathcal{J}_1=rb^2\int_0^\pi\psi\left(b,-\frac{b-r\cos\theta}{\sqrt{r^2+b^2-2br\cos\theta}}\right)\times\frac{\sin\theta(b-r\cos\theta)\exp(-\sqrt{r^2+b^2-2br\cos\theta})}{(\sqrt{r^2+b^2-2br\cos\theta})^3}\,d\theta \tag{A1}$$

Substituting (17) in (A1) we get

$$\mathcal{J}_1=rb^2\int_0^\infty dR\int_0^\pi d\theta\chi\left(\left\{R^2+b^2+\frac{2bR(b-r\cos\theta)}{\sqrt{r^2+b^2-2br\cos\theta}}\right\}^{1/2}\right)\times\frac{\sin\theta(b-r\cos\theta)\exp\{-[R+\sqrt{(r^2+b^2-2br\cos\theta)}]\}}{(\sqrt{r^2+b^2-2br\cos\theta})^3}. \tag{A2}$$

We now change to the new coordinates (r',α) defined by

$$\sqrt{r^2+r'^2-2rr'\cos\alpha}=R+\sqrt{r^2+b^2-2br\cos\theta},$$
$$r'\cos\alpha=b\cos\theta-\frac{R(r-b\cos\theta)}{\sqrt{r^2+b^2-2br\cos\theta}}. \tag{A3}$$

Then the argument of the function χ is simply r', while the element of integration is given by

$$dR\,d\theta=\begin{vmatrix}\partial R/\partial r' & \partial\theta/\partial r'\\ \partial R/\partial\alpha & \partial\theta/\partial\alpha\end{vmatrix}dr'\,d\alpha. \tag{A4}$$

Relation (A4) gives

$$\frac{b^2\sin\theta(b-r\cos\theta)}{(\sqrt{r^2+b^2-2br\cos\theta})^3}\,dR\,d\theta=\frac{r'^2\sin\alpha}{r^2+r'^2-2rr'\cos\alpha}\,dr'\,d\alpha \tag{A5}$$

and therefore

$$\mathcal{J}_1=\int_b^\infty\chi(r')r'\,dr'\times\int_0^\pi\frac{rr'\sin\alpha\exp(-\sqrt{r^2+r'^2-2rr'\cos\alpha})}{r^2+r'^2-2rr'\cos\alpha}\,d\alpha.$$

Carrying out the integration over α we obtain Eq. (18).

APPENDIX B

We will prove the relations (56) of text. Let us consider the third of these relations first. By definition

$$H_1(k,t)=\epsilon b^2\int_{-1}^1\mu'\psi(b,\mu')\,d\mu'\times\sum_{n=0}^\infty(-i)^n(2n+1)P_n(\mu')P_n(t)\zeta_n^{(1)}(kb). \tag{B1}$$

From the expansion of $\zeta_n^{(1)}(kb)$

$$\zeta_n^{(1)}(kb)=\sqrt{\pi/2kb}\,H_{n+1/2}^{(1)}(kb)=\frac{\exp(ikb)}{ikb}\sum_{m=0}^n\frac{(-i)^{n-m}(n+m)!}{m!(n-m)!}\left(\frac{1}{2kb}\right)^m. \tag{B2}$$

We get

$$H_1(i/\nu,\nu)=\nu b\exp(-b/\nu)\int_{-1}^1\mu'\psi(b,\mu')\,d\mu'\sum_{n=0}^\infty\frac{(\nu/2b)^n}{n!}\times\sum_{m=n}^\infty(-)^m(2m+1)\frac{(m+n)!}{(m-n)!}P_m(\mu')P_m(\nu). \tag{B3}$$

Since

$$\frac{1}{2}\sum_{m=n}^\infty(-)^m(2m+1)\frac{(m+n)!}{(m-n)!}P_m(\nu)P_m(\mu')=D(D-1.2)(D-2.3)\cdots(D-n.\overline{n+1})\delta(\mu'+\nu), \tag{B4}$$

$$D\equiv\frac{d}{d\mu'}(1-\mu'^2)\frac{d}{d\mu'}.$$

We observe that $H_1(i/\nu, \nu)$, $\nu > 0$ depends only upon the incoming angular distribution $\psi(b, \mu')$, $\mu' \in (-1, 0)$. Now the angular flux $\psi_0(r, \mu)$ in a purely capturing sphere of radius b free from all internal sources is given by

$$\psi_0(r, \mu) = \psi(b, -\sqrt{1 - r^2(1-\mu^2)/b^2}) \times \exp\{-[\mu r + \sqrt{b^2 - r^2(1-\mu^2)}]\}. \tag{B5}$$

Further, if $\psi(b, -\mu)$ is given by the Eq. (17), then

$$\psi(b, -\mu') = \int_b^\infty \chi(r') r' \frac{\exp\{+[\mu' r' - \sqrt{r'^2 - b^2(1-\mu'^2)}]\}}{\sqrt{r'^2 - b^2(1-\mu'^2)}} dr' \tag{B6}$$

and

$$\psi_0(r, \mu) = \int_0^1 \frac{d\nu}{\nu^2} \int_b^\infty \chi(r') r' \exp(-r'/\nu)\, dr' \exp(-\mu r) \times I_0\left(\frac{r}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu^2}\right). \tag{B7}$$

If we change the contour of integration (along the real line) in the definition of $\psi_0(r, \mu)$ as given by equation (41) of text, we get

$$\psi_0(r, \mu) = \frac{1}{2}\int_0^1 H_1(i/\nu, \nu) \frac{d\nu}{\nu^3} \exp(-\mu r) I_0\left(\frac{r}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu^2}\right). \tag{B8}$$

Comparing (B7) and (B8) we get (since they are true for $-1 \le \mu \le 1$, $0 \le r \le b$)

$$\frac{H_1(i/\nu, \nu)}{2\nu^2} = -\frac{1}{\nu}\int_b^\infty \chi(r') r' \exp(-r'/\nu)\, dr'. \tag{B9}$$

In order to derive the remaining equations of the set (56) we add Eq. (55) for ν and $-\nu$ to obtain

$$\lambda(\nu)[A_<(\nu) + A_<(-\nu)] + \frac{c}{2}\int_0^1 \nu' A_<(\nu') \exp(b/\nu')\, d\nu' \left(\frac{\exp(-b/\nu)}{\nu' - \nu} + \frac{\exp(b/\nu)}{\nu' + \nu}\right) + \frac{c}{2}\int_0^1 \nu' A_<(-\nu') \exp(-b/\nu)\, d\nu' \left(\frac{\exp(-b/\nu)}{\nu' + \nu} + \frac{\exp(b/\nu)}{\nu' - \nu}\right) = \frac{H(i/\nu, \nu)}{\nu^2} - \frac{c\nu_0}{2} a_{0+}^< \exp(b/\nu_0)\left[\frac{\exp(b/\nu)}{\nu_0 + \nu} + \frac{\exp(-b/\nu)}{\nu_0 - \nu}\right] + \frac{c\nu_0}{2} a_{0-}^< \exp(-b/\nu_0)\left(\frac{\exp(-b/\nu)}{\nu_0 + \nu} + \frac{\exp(b/\nu)}{\nu_0 - \nu}\right) \tag{B10}$$

where

$$H(i/\nu, \nu) = \tfrac{1}{2}[H_1(i/\nu, \nu) + H_1(-i/\nu, -\nu)] = -b^2 \int_{-1}^1 \mu' \psi(b, \mu') \exp(b\mu') I_0[(b/\nu)\sqrt{1-\nu^2}\sqrt{1-\mu'^2}]\, d\mu'. \tag{B11}$$

From Eq. (4b) of the text and Eq. (B6) we have, for interior problems,

$$\psi(b, \mu') \exp(b\mu') = \int_b^\infty r' \chi(r') \frac{\exp[-\sqrt{r'^2 - b^2(1-\mu'^2)}]}{\sqrt{r'^2 - b^2(1-\mu'^2)}} dr' + \Theta(\mu') \int_{b\sqrt{1-\mu^2}}^b y[c\rho(y) + Q(y)]\, dy \frac{\cosh[\sqrt{y^2 - b^2(1-\mu'^2)}]}{\sqrt{y^2 - b^2(1-\mu'^2)}}. \tag{B12}$$

Substituting (B12) in (B11) we find that the contribution from the first term vanishes, while the second term gives

$$H(i/\nu, \nu) = -\nu \int_0^b y[c\rho(y) + Q(y)] \sinh(y/\nu)\, dy. \tag{B13}$$

Using the expression for $y\rho(y)$ as given by Eq. (29), the integral of Eq. (B13) can be evaluated and hence we get from (B10)

$$\lambda(\nu) A_\zeta^*(\nu) \exp(-b/\nu) + \frac{c}{2}\int_0^1 \nu' A_\zeta^*(\nu') \exp(-b/\nu')\, d\nu' \times \left(\frac{1}{\nu' - \nu} + \frac{\exp(-2b/\nu)}{\nu' + \nu}\right) = a_{0+}^* \exp(-b/\nu_0)\left(\frac{1}{\nu_0 - \nu} + \frac{\exp(-2b/\nu)}{\nu_0 + \nu}\right) \tag{B14}$$

where

$$A_\zeta^*(\nu) = A_<(\nu) + A_<(-\nu) + \frac{1}{N_\nu}\int_0^b r' Q(r') \sinh(r'/\nu)\, dr' \tag{B15}$$

and the constant a_{0+}^* given by

$$a_{0+}^* = a_{0-}^< - a_{0+}^< - \frac{1}{N_{0+}}\int_0^b r' Q(r') \sinh(r'/\nu_0)\, dr'. \tag{B16}$$

The index of Eq. (B14) is unity and hence its adjoint homogeneous equation has one nontrivial solution. The solution of (B14) exists if and only if its rhs is orthogonal to this adjoint function. This requirement can be met only by letting a_{0+}^* vanish in which case $A_\zeta^*(\nu)$ also vanishes and hence the relations quoted in (56).

APPENDIX C

For the exterior problem $H_1(i/\nu, \nu)$ is given by

$$H_1(i/\nu, \nu) = \epsilon a^2 \int_{-1}^1 \mu' \psi(a, \mu')\, d\mu' \times \sum_{n=0}^\infty (-i)^n (2n+1) P_n(\mu') P_n(\nu) \zeta_n^{(1)}(ia/\nu). \tag{C1}$$

This expression is same as the one given by (B1) (a replacing b). As shown in Appendix B, $H_1(i/\nu, \nu)$ depends only upon $\psi(a, \mu)$, $\mu \in (-1, 0)$. Further, for exterior problem we have from (4b), $\mu > 0$,

$$\psi(a, -\mu) = \frac{\exp(a\mu)}{2}\int_a^\infty \frac{\exp[-\sqrt{y^2 - a^2(1-\mu^2)}]}{\sqrt{y^2 - a^2(1-\mu^2)}} \times y[c\rho(y) + Q(y)]\, dy. \tag{C2}$$

Using (C2), (B6), and (B9) we have for exterior problems

$$H_1(i/\nu, \nu) = \frac{1}{2\nu}\int_a^\infty y[c\rho(y) + Q(y)] \exp(-y/\nu)\, dy. \tag{C3}$$

Using the expansion (29) for $y\rho(y)$ for evaluating the integral in (C3) and substituting the result in Eq. (60) of the text, we get

$$\lambda(\nu)B^*(\nu)\exp(a/\nu)+\frac{c}{2}\int_0^1 B^*(\nu')\nu'\exp(a/\nu')\,d\nu'\times\left(\frac{1}{\nu'-\nu}-\frac{\exp(-2a/\nu')}{\nu'+\nu}\right)=b_0^*\frac{c\nu_0}{2}\left[\frac{\exp(-a/\nu_0)}{\nu_0+\nu}-\frac{\exp(a/\nu_0)}{\nu_0-\nu}\right] \tag{C4}$$

where $\nu\in(0,1)$

$$B^*(\nu)=B(\nu)-\frac{1}{2N_\nu}\int_a^\infty r'Q(r')\exp(-r'/\nu)\,dr' \tag{C5}$$

and

$$b_0^*=b_{0+}-\frac{1}{N_{0+}}\int_a^\infty r'Q(r')\exp(-r'/\nu_0)\,dr'. \tag{C6}$$

Again since Eq. (C4) has only trivial solution $B^*(\nu)$ and b_0^* vanish and hence the first two relations of the set (62), the function $H(i/\nu,\nu)$ is given by the relation

$$H(i/\nu,\nu)=a^2\int_{-1}^1\mu'\psi(a,\mu')\exp(a\mu')I_0[(a/\nu)\sqrt{1-\nu^2}\sqrt{1-\mu^2}]\,d\mu'. \tag{C7}$$

Since the angular flux $\psi(a,\mu)$, $\mu\in(0,1)$ is prescribed by boundary conditions while $\psi(a,-\mu)$ is given by (C2), we get

$$H(i/\nu,\nu)=a^2\int_0^1\mu'\psi(a,\mu')\exp(a\mu')I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'-\frac{a^2}{2}\int_a^\infty y[c\rho(y)+Q(y)]\,dy\times\int_0^1\mu'\frac{\exp[-\sqrt{y^2-a^2(1-\mu'^2)}]}{\sqrt{y^2-a^2(1-\mu'^2)}}\times I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu' \tag{C8}$$

On using the Eq. (11) of the text and the formula[16]

$$\int_0^{\pi/2}J_\lambda(z\sin\theta)J_{\lambda'}(\omega\cos\theta)\sin^{\lambda+1}\theta\cos^{\lambda'+1}\theta\,d\theta=z^\lambda\omega^{\lambda'}\frac{J_{\lambda+\lambda'+1}(\sqrt{z^2+\omega^2})}{(\sqrt{z^2+\omega^2})^{\lambda+\lambda'+1}}. \tag{C9}$$

The integral in the second term of (C8) can be evaluated. The result is

$$\int_0^1\mu'\frac{\exp[-\sqrt{y^2-a^2(1-\mu'^2)}]}{\sqrt{y^2-a^2(1-\mu'^2)}}I_0\left(\frac{a}{\nu}\sqrt{1-\nu^2}\sqrt{1-\mu'^2}\right)d\mu'=\int_\nu^1\exp(-\sqrt{y^2-a^2}/\mu')\frac{I_1(a\sqrt{1/\nu^2-1/\mu'^2})}{a\sqrt{1/\nu^2-1/\mu'^2}}\frac{d\mu'}{\mu'^2}+\int_0^\nu\exp(-\sqrt{y^2-a^2}/\mu')\frac{J_1(a\sqrt{1/\mu'^2-1/\nu^2})}{a\sqrt{1/\mu'^2-1/\nu^2}}\frac{d\mu'}{\mu'^2}. \tag{C10}$$

The second integral in rhs of (C10) can be evaluated in closed form

$$\int_0^\nu\exp(-\sqrt{y^2-a^2}/\mu')\frac{J_1(a\sqrt{1/\mu'^2-1/\nu^2})}{a\sqrt{1/\mu'^2-1/\nu^2}}\frac{d\mu'}{\mu'^2}=[\exp(-\sqrt{y^2-a^2}/\nu)-\exp(-y/\nu)]\frac{\nu}{a^2}. \tag{C11}$$

Using (C10) and (C11) in (C8), we get the Eq. (62) of the text.

[1]K.M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[2]W. Kofink, Report ORNL-3216, Oak Ridge National Laboratory, Oak Ridge, Tennessee (1961).
[3]G.J. Mitsis, Report ANL-6787, Argonne National Laboratory, Argonne, Illinois (1963).
[4]E.H. Bareiss and I.K. Abu-Shumays, Report ANL-7328, Argonne National Laboratory, Argonne, Illinois (1967).
[5]R.C. Erdmann and C.E. Siewert, J. Math. Phys. **9**, 81 (1968).
[6]A. Leonard and T.W. Mullikin, Proc. Natl. Acad. Sci. (U.S.) **52**, 683 (1964).
[7]K.M. Case and P.F. Zweifel, *An Introduction to Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967), Appendix I.
[8]A.G. Gibbs, J. Math. Phys. **10**, 875 (1969).
[9]O.J. Smith and C.E. Siewert, J. Nucl. Energy **22**, 579 (1968).
[10]D.C. Sahni, J. Nucl. Energy **20**, 915 (1966).
[11]O.J. Smith, *Neutron Transport Theory Conference, January* 20-24, ORO-3858-1 , Virginia Polytechnic Institute, Blacksburg (1969), p. 653.
[12]O.J. Sheaks, J. Math. Phys. **13**, 203 (1972).
[13]K.M. Case, R. Zelazny, and M. Kanal, J. Math. Phys. **11**, 223 (1970).
[14]B. Davison, Proc. Phys. Soc. Lond. **64**, 981 (1951).
[15]A. Ardelyi, W. Magnus, F. Oberhettinger, and F.G. Tricomi, *Tables of Integral Transforms*, Bateman Manuscript Project (McGraw-Hill, New York, 1954), p. 200.
[16]G.N. Watson, *Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1952), 2nd ed., p. 376.

The Gel'fand states of certain representations of $U(n)$ and the decomposition of products of representations of $U(2)$

Christopher J. Henrich

Department of Mathematics, Fordham University, Bronx, New York 10458
(Received 14 November 1974)

The representations of $U(n)$, as realized by Bargmann and Moshinsky on spaces of polynomials ("boson calculus"), are the main subject of this paper. We consider them from a global point of view, pointing out the connection with induced representations. To compute the detailed structure of the representations, we find the reproducing kernels of the function spaces and the operators that connect them according to Weyl's branching law. Using these results, we compute the boson polynomials of representations of $U(3)$, and arrange them in a generating function. We extend this generating function to the boson polynomials of representations of $U(n)$ of the form $\langle(m_1 m_2 0\cdots 0)\rangle$. By considering these polynomials from a different viewpoint, we are able to obtain an explicit decomposition of the Kronecker product of $n-1$ representations of $SU(2)$.

I. INTRODUCTION

A. Summary

The irreducible representation of the compact simple Lie groups have been studied from many viewpoints. In one sense, they might seem a "dead" subject, since they were completely classified by the work of Cartan and Weyl. But in a larger perspective, Cartan and Weyl only began the study of these representations. A complete understanding of one representation requires an explicitly described space on which the group elements act in a completely specified way. A more detailed description would include a basis of the vector space, and the matrix entries of the operators corresponding to elements of the group. Such a level of detail is of interest for applications of the theory, and is more and more perceptibly required for developments within representation theory itself.

For the case of $SU(n)$, some of the details have been filled in. Weyl[1], and Gel'fand and Zeitlin[2] pointed out that the "branching law"[3] provides a natural inductive method of choosing an orthonormal basis for any unitary representation. See also Ref. 4. More recently, Baird, Biedenharn, and others have made a very extensive study of the Gel'fand bases and related structures (see Refs. 5—7, and papers referred to there). A description of the representations on spaces of polynomials was introduced by Bargmann and Moshinsky[8,9]; Moshinsky, Biedenharn, and their coworkers have made some progress on the expression of the Gel'fand—Zeitlin bases in this realization, usually called boson polynomials, for $SU(3)$ and $SU(4)$ (see Refs. 4—7, 10—14). The complexity of the results for $SU(n)$ seems to grow rapidly as n increases; the difficulty of finding the results grows even faster.

Moshinsky,[15] in 1963, made an observation of the highest importance regarding the problem of outer multiplicities. This problem arises when one takes the Kronecker product of two (or more) representations of $SU(n)$ and decomposes it into irreducible subrepresentations. In general, some of the subrepresentations occur with multiplicities greater than 1; therefore, if one wishes to express the decomposition explicitly, one must somehow select a sufficient number of inequivalent irreducible subrepresentations. The search for efficient means of selection has challenged the ingenuity of researchers from Racah on down. Moshinsky observed that the Kronecker product of k representations of $SU(n)$ could be analyzed in terms of certain representations of $SU(N)$, where $N=k(n-1)$. For another approach to this problem, giving detailed results in $SU(3)$ and a partial extension to $SU(n)$, see Refs. 16 and 17.

In the present paper, the author hopes to clarify some of the theory of the representations of $SU(n)$. The reproducing kernels, of certain vector spaces and of operators between them, provide a kind of explicit description of the spaces and operators. We compute these kernels explicitly; on the one hand, we find that they are related to the maximal and semimaximal states as defined in Ref. 5, and on the other hand they can be used to compute other states. We have successfully carried out this computation for the states of certain representations, including the representations $(m_1 m_2 0\cdots 0)$ of $U(n)$. This result is enough to provide a decomposition of products of representations of $SU(2)$.

Certain steps in the argument are not new, indeed are familiar in various forms in the literature; I have tried to give correct acknowledgments. (I am indebted to the referee for several references.) Some material is repeated in order to fix notation and make this paper relatively self-contained. As to the normalization constants for the maximal and semimaximal states, it seems (to the author, at least) that the proofs given here are more conceptual than the lengthy computations which are sometimes described very briefly in the literature. This paper largely avoids the use of "infinitesimal elements," raising and lowering operators, and the like, in favor of more "global" and integral methods in the spirit of Bargmann's elegant paper.[18]

The organization of this paper is as follows. Sec. IB describes the Fock spaces of analytic functions, in which we shall work. The formalism of these spaces is known by many names; notably, the "boson calculus." Section IC is concerned with some notational conventions we shall adopt for working in Fock spaces over spaces

of matrices, and with the device which Weyl called the "unitarian trick."

Section II introduces the representation spaces, following Bargmann and Moshinsky. In Sec. IIA we give the definition. Section IIB points out how this construction is related to recent developments in pure mathematics, known as the Borel—Weil theorem. This relationship has been very little noticed, and came as a surprise to the author. Section IIC is largely devoted to a result, very similar to one proved by Moshinsky, on the structure of any element of one of the representation spaces. In Sec. IID we find the reproducing kernels of these spaces.

Proceeding in the same style, we consider the branching operators in Sec. III. In Sec. IIIA we describe the requirements which the kernel of such an operator must satisfy. In Sec. IIIB we compute the kernel.

Section IV is concerned with the main problem of this paper: the explicit expression of the boson polynomials. In Sec. IVA, we analyze the branching kernel to find the states for $U(3)$; of course, we get the same answer as everyone else. The same kind of analysis does not easily extend to $U(n)$ where $n \geq 4$. However, we notice the existence of a simple generating function for the $U(3)$ states, as explained in Sec. IVB. It cries aloud to be generalized, and in Sec. IVC we do generalize it to certain representations of $U(n)$.

In Sec. V we apply the preceding work to the study of products of representations of $SU(2)$. Section VA restates in somewhat greater detail what Moshinsky pointed out in 1963.[15] The product of $n-1$ representations of $SU(2)$ can be decomposed by finding invariant vectors in products of n representations. These vectors can be regarded as Gel'fand states of representations of $U(n)$. Only those representations are needed which correspond to Young diagrams with two rows of equal length; for these, the result of Sec. IVC is adequate. The results, which are worked out in Sec. VC, include a generalization of the generating function for the 3-j coefficients, found in Ref. 18.

In conclusion, Sec. VI describes some problems which await solution, and some of the author's plans and hopes regarding them.

B. Fock spaces

Let E be a finite-dimensional complex Hilbert space. If n is the dimension of E and $(z_1, \ldots, z_n)$ is an orthonormal coordinate system, let γ be the Gaussian measure on E defined by

$$d\gamma = \pi^{-n} \exp[-(z,z)] \prod_{k=1}^{n} d\mathrm{Re}z_k \, d\mathrm{Im}z_k. \tag{1.1}$$

Let $\mathcal{F}$ be the space of analytic functions f on E such that $\int |f|^2 d\gamma < \infty$; then $\mathcal{F}$ is a Hilbert space with the inner product

$$(f,g) = \int \bar{f}(z) g(z) \, d\gamma(z).$$

The Hilbert space $\mathcal{F}$, which we shall call Fock space over E, has been studied in detail by Bargmann[19]; a useful source is Ref. 18. For proofs of the results in this section, see either of these references.

All polynomials on E belong to $\mathcal{F}$; if we let $\mathbf{z}^\alpha$ stand for $z_1^{\alpha_1} \cdots z_n^{\alpha_n}$ and $\alpha!$ stand for $\alpha_1! \cdots \alpha_n!$, then the monomials $(\alpha!)^{-1/2}\mathbf{z}^\alpha$ make up an orthonormal basis of $\mathcal{F}$.

If $w \in E$, let us define the function $\mathbf{e}_w$ by

$$\mathbf{e}_w(z) = \exp[(w,z)].$$

We are using the convention that inner products are complex-linear in the right-hand factor. Then $\mathbf{e}_w \in \mathcal{F}$, and for any f in $\mathcal{F}$,

$$f(w) = (\mathbf{e}_w, f).$$

This relation is sometimes expressed by saying that the function $\mathbf{K}$ defined by

$$\mathbf{K}(z,w) = e_w(z)$$

is the reproducing kernel of $\mathcal{F}$. Any closed linear subspace $\mathcal{G}$ of $\mathcal{F}$ has a reproducing kernel, i.e., a function $\mathbf{K}_\mathcal{G}$ such that, if we set

$$\mathbf{K}_{\mathcal{G},w}(z) = \mathbf{K}_\mathcal{G}(z,w)$$

then

$$\mathbf{K}_{\mathcal{G},w} \in \mathcal{G}, \quad f(w) = (\mathbf{K}_{\mathcal{G},w}, f) \quad \text{if } f \in \mathcal{G}. \tag{1.2}$$

The reproducing kernel $\mathbf{K}_\mathcal{G}$ is completely characterized by (1.2).

Proposition 1.1: Let $\{f_j\}$ be an orthornormal basis of the closed linear subspace $\mathcal{G}$. Then (a bar denotes complex conjugation)

$$\mathbf{K}_\mathcal{G}(z,w) = \sum_j f_j(z) \bar{f}_j(w), \quad z, w \in M(E). \tag{1.3}$$

If M is a bounded linear operator on $\mathcal{F}$, then we can define the reproducing kernel of M as follows:

$$\mathbf{M}(z,w) = M(\mathbf{e}_w)(z) = (\mathbf{e}_z, M\mathbf{e}_w).$$

Let us note the following facts:

Proposition 1.2: If $f \in \mathcal{F}$, then

$$Mf(z) = \int \mathbf{M}(z,w) f(w) \, d\gamma(w).$$

Proposition 1.3: The reproducing kernel of M^* is

$$\mathbf{M}^*(z,w) = \mathbf{M}(w,z)^-.$$

Proposition 1.4: If $\{f_j\}$ is an orthornormal basis of $\mathcal{F}$, then

$$\mathbf{M}(z,w) = \sum (Mf_j)(z) \bar{f}_j(w);$$

the elements of f_j may be restricted to a basis of the orthogonal complement of the null space of M.

Proposition 1.5: If P is the orthogonal projection of $\mathcal{F}$ onto a closed subspace $\mathcal{G}$, then

$$\mathbf{K}_\mathcal{G}(z,w) = \mathbf{P}(z,w).$$

Proposition 1.6: If S and T are bounded operators on $\mathcal{F}$, and $V = ST$, then

$$\mathbf{V}(z,w) = \int \mathbf{S}(z,v) \mathbf{T}(v,w) \, d\gamma(v).$$

C. Fock space over a space of matrices

We shall apply the formalism of Sec. IB to the case that E is the vector space $M(n)$ of $n \times n$ square matrices, with the inner product

$$(z,w) = \mathrm{Tr}(z^* w).$$

If $z \in M(n)$, we shall denote the rows of z by $z_{1\cdot}, \ldots, z_{n\cdot}$; on occasion we shall consider the columns of z, and we shall denote them by $z_{\cdot 1}, \ldots, z_{\cdot n}$.

We shall need to refer to the determinants of submatrices of z. For the most general case, we set

$$\Delta^{i_1 \cdots i_d}_{j_1 \cdots j_d}(z) = \det(z_{i_a j_b})_{a,b=1,\ldots,d}. \tag{1.4}$$

Here we can assume $i_1 < i_2 < \cdots < i_d$ and $j_1 < j_2 < \cdots < j_d$. For the most part, we shall need (1.4) only when $i_1 = 1, \ldots, i_d = d$; we set

$$\Delta_{j_1 \cdots j_d}(z) = \Delta^{1 \cdots d}_{j_1 \cdots j_d}(z). \tag{1.5}$$

These polynomials are boson polynomials; every boson polynomial can be represented in terms of them. Finally we set

$$\Delta(d; z) = \Delta^{1 \cdots d}_{1 \cdots d}(z). \tag{1.6}$$

The vector space $M(n)$ contains $GL(n, \mathbb{C})$ as a dense open subset, and $U(n)$ as a compact subset of $GL(n, \mathbb{C})$. It is important for us that certain equations can be extended from $U(n)$ to $M(n)$. Weyl (Ref. 20, p. 265) uses such an extension to show that a complete list of finite-dimensional representations of $SL(n, \mathbb{C})$ can be derived from a complete list of representations of $SU(n)$. We shall use the following proposition several times.

Proposition 1.7: Let f and g be analytic functions on $GL(n, \mathbb{C})$. If $f(u) = g(u)$ for all u in $U(n)$, then $f(z) = g(z)$ for all z in $GL(n, \mathbb{C})$. If furthermore, f and g are analytic on $M(n)$, then $f(z) = g(z)$ for all z in $M(n)$.

Proof: Consider the analytic function $\phi(s) = f(\exp(s)) - g(\exp(s))$, where $s \in M(n)$. By hypothesis if s is skew-Hermitian, so that $\exp(s) \in U(n)$, then $\phi(s) = 0$. Thus the function ϕ vanishes on a set of matrices which is characterized by the condition that certain linear combinations of the entries have real values. But analytic continuation is possible, passing from those real values to arbitrary complex values; that is, $\phi(\exp(s)) = 0$ for all s in $M(n)$. Therefore, $f(z) = g(z)$ for all z belonging to some neighborhood of the identity, and the conclusion follows immediately.

II. THE REPRESENTATION SPACES OF BARGMANN AND MOSHINSKY

A. Definitions

The Fock space $\mathcal{F}$ lying over $M(n)$ is invariant under transformations of the base space which preserve inner products, e.g., multiplication on the right by a member of $U(n)$. If $u \in U(n)$ and we define $L_u : \mathcal{F} \to \mathcal{F}$ by

$$(L_u f)(z) = f(zu), \tag{2.1}$$

then L is a unitary representation of $U(n)$. Bargmann and Moshinsky[8,9] have shown how to define subrepresentations of L which make up a complete list of irreducible representations of $SU(n)$. The method can be traced back to Jordan[21]; see also Schwinger's notes in Ref. 22. Their method is as follows. If $1 \leq j, k \leq n$, let $\Gamma_{j,k}$ be the operator defined by

$$\Gamma_{j,k} f(z) = \sum_{h=1}^{n} z_{jh} \frac{\partial f}{\partial z_{kh}}$$

For nonnegative integers $m_1, \ldots, m_n$, let $\mathcal{B}^{m_1 \cdots m_n}$ (or, for short, $\mathcal{B}^m$) be the space of functions f in $\mathcal{F}$ satisfying

$$\Gamma_{j,j} f = m_j f, \quad 1 \leq j \leq n, \tag{2.2}$$

$$\Gamma_{j,k} f = 0, \quad 1 \leq j < k \leq n. \tag{2.3}$$

Equation (2.2) implies that f is a polynomial, homogeneous of degree m_j in the elements of the row $z_{j\cdot}$. Moshinsky[9] shows that $\mathcal{B}^m$ is zero unless $m_1 \geq m_2 \geq \cdots \geq m_n$, and if these inequalities are satisfied then $U(n)$ acts irreducibly on $\mathcal{B}^m$. We shall let L^m denote the restriction of L to $\mathcal{B}^m$:

$$L^m f(z) = f(zu), \quad f \in \mathcal{B}^m.$$

It is also shown in Ref. 9 that L^m is isomorphic to the representation denoted $\langle P(m_1, \ldots, m_n) \rangle$ by Weyl (Ref. 20, p. 130ff.)

We get a complete list (up to equivalence) of unitary representations of $SU(n)$ by taking those L^m with $m_n = 0$. As representations of $U(n)$, $L^{m_1 \cdots m_n}$ and $L^{m_1+h, \ldots, m_n+h}$ differ only in a power of the determinant of the group element. Thus by taking all n-tuples satisfying $m_1 \geq \cdots \geq m_n \geq 0$, we get a redundant list of representations of $SU(n)$, or a partial list of representations of $U(n)$; the others differ by powers of the determinant.

Let us consider the definition of $\mathcal{B}^m$ in a slightly different way. Let $\mathcal{N}_n$ be the subgroup of $GL(n, \mathbb{C})$ consisting of lower-triangular matrices with 1's on the main diagonal:

$$\mathcal{N}_n = \{ g \in GL(n, \mathbb{C}) : g_{jj} = 1, \ g_{jk} = 0 \text{ if } 1 \leq j < k \leq n \}. \tag{2.4}$$

Then (2.3) is formally equivalent to the condition

$$f(gz) = f(z), \quad g \in \mathcal{N}_n. \tag{2.5}$$

(As we are dealing with polynomials, we may put out of our minds all scruples connected with the unboundedness of $\Gamma_{j,k}$.) Similarly, if we let $\mathcal{D}_n$ be the group of diagonal matrices

$$\delta = \operatorname{diag}(\delta_1, \ldots, \delta_n) \in GL(n, \mathbb{C})$$

and write δ^m for $\delta_1^{m_1} \cdots \delta_n^{m_n}$, then (2.2) can be expressed as

$$f(\delta z) = \delta^m f(z). \tag{2.6}$$

Therefore $\mathcal{B}^m$ is the space of holomorphic functions on $M(n)$ which transform in the following way under the left action of the group $\mathcal{D}_n \mathcal{N}_n$ of lower triangular matrices:

$$f(\delta g z) = \delta^m f(z), \quad z \in M(n), \quad \delta \in \mathcal{D}_n, \quad g \in \mathcal{N}_n. \tag{2.7}$$

This description of the representation spaces is also found in Refs. 23 and 24.

B. A digression on induced representations

The reader who is acquainted with the representation theory of the noncompact simple Lie groups (e.g., Ref. 23 or Sec. 9 of Ref. 25) may recognize that the definition of L^m is very strongly analogous to the construction of the principal series of representations of $GL(n, \mathbb{C})$ as induced representations. Let us recall that construction. First, one considers a homomorphism ρ of the Abelian group $\mathcal{D}_n$ into $\mathbb{C}^*$; ρ defines a homomorphism $\tilde{\rho}$ of the solvable group $\mathcal{P} = \mathcal{D}_n \mathcal{N}_n$ into $\mathbb{C}^*$, by

setting $\tilde{\rho}(\delta g)=\rho(\delta)$ if $\delta\in(D)_n$ and $g\in(N)_n$. One then considers functions f defined on $GL(n,\mathbb{C})$ such that

$$f(pz)=\tilde{\rho}(p)f(z),\quad z\in GL(n,\mathbb{C}),\quad p\in(P). \tag{2.8}$$

Then $GL(n,\mathbb{C})$ acts on the space of such functions by right translation. If ρ satisfies $|\rho(\delta)|=1$, then an inner product can be defined on the set of f satisfying (2.8) so that the functions of finite norm constitute a Hilbert space, and the representation of $GL(n,\mathbb{C})$ on this Hilbert space is unitary. It is the representation induced by $\tilde{\rho}$.

The most important differences between this construction and that of Moshinsky[9,15] are (a) the homomorphism ρ is not unitary but is analytic on the complex Lie group $(D)_n$, and (b) the functions which make up the representation space not only satisfy (2.8) but also are holomorphic on $GL(n,\mathbb{C})$. (It may clarify matters to note that Moshinsky does not mention ρ explicitly, but the equations (4.3a-b) of Ref. 9 are equivalent to our (2.2)–(2.3).) The representation on $(B)^m$ can be defined for $GL(n,\mathbb{C})$, but only on the compact subgroup $U(n)$ can it be unitary. The construction of L^m is sometimes known as holomorphic induction.

The theorem that every irreducible representation of a compact Lie group can be realized by holomorphic induction from a one-dimensional representation of a group such as (P) is a particular case of what is known as the Borel–Weil theorem. It was first suggested (so far as I know) by Godement[26] in 1952, and was developed further for the classical groups by Zhelobenko[24] in 1962. The formulation of Borel and Weil[27], and its generalization by Bott[28], use highly sophisticated mathematical concepts. A general algebraic form of the theorem is found in Ref. 29; the book,[30] especially chapter 6, contains a detailed exposition. A book by the present author to be published by Marcel Dekker, contains an introductory exposition. It seems curious that on the one hand, Bargmann and Moshinsky make no explicit reference to induced representations, while on the other hand it is not apparent from the statement of the Borel–Weil theorem that the spaces which it describes can be made explicit in a way suitable for computation.

When this paper was essentially finished, the author saw Ref. 31, which also remarks the identity of Zhelobenko's construction with that of Bargmann and Moshinsky.

C. The structure of a function in $(B)^m$

Let us consider the space (B) of all elements of (F) satisfying (2.5). This is the orthogonal direct sum of the spaces $(B)^m$. The polynomial members of (B) form a dense subspace $(B)_0$ of (B); in fact $(B)_0$ is the set of finite linear combinations of members of the spaces $(B)^m$.

Theorem 2.1 Every member f of $(B)_0$ can be expressed as a polynomial function of the determinants $\Delta_{j_1\cdots j_d}(z)$ as defined in (1.5).

Proof: Let us regard the matrix z as organized into its n columns $z_{\cdot 1},\ldots,z_{\cdot n}$. An element of $(B)_0$ can be considered as a polynomial in these n vector arguments, invariant under the action upon them of the group $(N)_n$. The theorem claims that the determinants $\Delta_{j_1\cdots j_d}(z)$ are (in the terminology of Ref. 20) a basic set of invariants, i.e., they generate the ring of polynomial invariants $(B)_0$. Now $\Delta_{j_1,\ldots,j_d}(z)$ can be obtained from $\Delta(d;z)=\Delta_{1\ldots d}(z)$ by simply substituting $z_{\cdot j_a}$ for $z_{\cdot a}$ in the latter determinant; thus (Ref. 20, p. 44) we want to prove that $\Delta(1;z),\ldots,\Delta(n;z)$ make up a complete list of "typical basic invariants." We shall do this by an induction argument. To reduce the case for n-rowed vectors to that for $(n-1)$-rowed vectors, we use the following lemma.

Lemma 2.1: Let $f\in(B)$ and let $f(z)$ be independent of $z_{\cdot n}$. Then $f(z)$ is independent of $z_{n\cdot}$.

Proof: We may consider f restricted to the subspace on which $z_{\cdot n}=0$. Suppose first that $\Delta(n-1;z)\neq 0$, which implies that the rows $z_{1\cdot},\ldots,z_{n-1\cdot}$ are linearly independent. Then they span the space of all $1\times n$ row vectors with the last entry equal to 0, so we can write

$$z_{n\cdot}=\sum_{j=1}^{n-1}c_j z_{j\cdot}.$$

If we set

$$g=\begin{pmatrix}1&0&\cdots&0&0\\0&1&\cdots&0&0\\\vdots&\vdots&\cdots&\vdots&\vdots\\0&0&\cdots&1&0\\c_1&c_2&&c_{n-1}&1\end{pmatrix},$$

then $g\in(N)_n$, and by (2.5)

$$f(z)=f\begin{pmatrix}z_{1\cdot}\\\vdots\\z_{n-1\cdot}\\z_{n\cdot}\end{pmatrix}=f\left(g\begin{pmatrix}z_{1\cdot}\\\vdots\\z_{n-1\cdot}\\0\end{pmatrix}\right)=f\begin{pmatrix}z_{1\cdot}\\\vdots\\z_{n-1\cdot}\\0\end{pmatrix}.$$

Thus, if z belongs to an open dense subset of the matrices with vanishing nth column, then $f(z)$ is independent of the nth row of z. By continuity the same thing is true for all such matrices. This proves the lemma.

We now prove by induction that $\Delta(1,z),\ldots,\Delta(n;z)$ are a complete list of typical basic invariants for $(N)_n$. If $n=1$, the statement is vacuous. Suppose that it is true for $n-1$, i.e., that $\Delta(1,z),\ldots,\Delta(n-1;z)$ are a complete list of typical basic invariant functions of column vectors of dimension $n-1$, under the group $(N)_{n-1}$.

Consider a polynomial f, depending on $n-1$ column vectors $z_{\cdot 1},\ldots,z_{\cdot n-1}$ of dimension n, and invariant under $(N)_n$. Such an f can be regarded as an invariant function of $z_{\cdot 1},\ldots,z_{\cdot n}$ that just happens to be independent of $z_{\cdot n}$. By the lemma, f is also independent of $z_{n\cdot}$, that is f does not depend on the last element of each column $z_{\cdot 1},\ldots,z_{\cdot n-1}$. Thus f can be regarded as a function defined on $M(n-1)$, invariant under the group N_{n-1}. By the induction hypothesis, f can be expressed in terms of $\Delta(1;z),\ldots,\Delta(n-1;z)$ and the results of column substitutions.

We have just shown that $\Delta(1;z),\ldots,\Delta(n-1;z)$ are a complete list of typical basic invariants for $n-1$

column vectors. By theorem (2.5A) on page 44 of Ref. 20, such a list, augmented by the inclusion of $\Delta(n;z)$, is a complete list for any number of column vectors. This completes the induction step, and finishes the proof of the theorem.

For another proof, see (Ref. 32, p. 219). See also Refs. 11 and 33. Moshinsky Ref. 9 proves a theorem which is similar to Theorem 2.1, starting from (2.3). His conclusion is that f may be expressed as a rational function of certain determinants of the form (1.5). Moshinsky's result is more economical in that the invariant is expressed in terms of a smaller number of arguments; but it has the disadvantage that there is some requirement on the rational function of the Δ's which must be solved if the invariant, as a function of the z_{jk}, is to be a polynomial.

For later reference we note this result. (For convenience we set $m_{n+1}=0$.)

Proposition 2.2: $\mathcal{B}^{m_1\cdots m_n}$ is nonzero if and only if $m_1\geq m_2\geq\cdots\geq m_n\geq 0$. If that is the case, then $f\in\mathcal{B}^{m_1\cdots m_n}$ if and only if

$$f(z)=P(\Delta_1(z),\ldots,\Delta_{j_1\cdots j_d}(z),\ldots,\Delta_{1\cdots n}(z))$$

where P is a polynomial which, for each $d=1,2,\ldots,n$, is homogeneous of degree m_d-m_{d+1} in the arguments

$$\Delta_{j_1\cdots j_d}(z).$$

Proof: We know that f, being a member of $\mathcal{B}_0$, is expressible as a polynomial in the $\Delta_{j_1\cdots j_d}(z)$. We have, because $f\in\mathcal{B}^m$, if $\delta=\mathrm{diag}(\delta_1,\ldots,\delta_n)$

$$\begin{aligned}f(\delta z)&=\delta_1^{m_1}\cdots\delta_n^{m_n}f(z)\\&=\delta_1^{m_1-m_2}(\delta_1\delta_2)^{m_2-m_3}\cdots(\delta_1\cdots\delta_n)^{m_n-m_{n+1}}f(z).\end{aligned}\tag{2.9}$$

But

$$\Delta_{j_1\cdots j_d}(\delta z)=(\delta_1\cdots\delta_d)\Delta_{j_1\cdots j_d}(z).$$

So if f is to have the homogeneity indicated by (2.9), P must be homogeneous of degree m_1-m_2 in the arguments $\Delta_j(z)$, and so on.

D. The reproducing kernel of $\mathcal{B}^m$

In this section, we shall compute the function $\mathbf{K}^m(z,w)$ which is the reproducing kernel of the space $\mathcal{B}^m$. We know that $\mathbf{K}^m(z,w)$ is an analytic function of z, and because $\mathbf{K}^m(z,w)=\mathbf{K}^m(w,z)^-$ (see Proposition 1.3), $\mathbf{K}^{m-}$ is analytic as a function of w; or, $\mathbf{K}^m(z,w)$ is analytic as a function of w^*.

Proposition 2.3: $\mathbf{K}^m(z,w)$ can be expressed as a function of zw^*; in fact,

$$\mathbf{K}^m(z,w)=\mathbf{K}^m(zw^*,I).$$

Proof: $\mathbf{K}^m$ is the kernel of the projection operator P^m whose image is $\mathcal{B}^m$ (see Proposition 1.5), so

$$\mathbf{K}^m(z,w)=(\mathbf{e}_z,P^m\mathbf{e}_w).$$

Now P^m commutes with L_u if $u\in U(n)$, so

$$(\mathbf{e}_z,P^mL_u\mathbf{e}_w)=(\mathbf{e}_z,L_uP^m\mathbf{e}_w)=(L_{u^*}\mathbf{e}_z,P^m\mathbf{e}_w).\tag{2.10}$$

We have $\mathbf{e}_w(x)=\exp(w,x)=\exp[\mathrm{Tr}(w^*x)]$; therefore,

$$\begin{aligned}L_u\mathbf{e}_w(x)&=e_w(xu)=\exp[\mathrm{Tr}(w^*xu)]=\exp[\mathrm{Tr}(uw^*x)]\\&=\mathbf{e}_{wu^*}(x).\end{aligned}$$

Thus (2.10) says

$$(\mathbf{e}_z,\ P^m\mathbf{e}_{wu^*})=(\mathbf{e}_{zu},P^m\mathbf{e}_w),$$

or

$$\mathbf{K}^m(z,wu^*)=\mathbf{K}^m(zu,w).\tag{2.11}$$

Now both sides of this equation depend analytically on the matrix elements of u, and may be expressed as polynomials. Therefore by Proposition 1.7, the equation is valid for all u in $M(n)$. If we set $w=I$, then we have $\mathbf{K}^m(z,u^*)=\mathbf{K}^m(zu,I)$, which is what we wanted to prove with u^* in place of w.

We can now concentrate our attention on the function $\mathbf{K}^m(z,I)$. It is completely determined by the fact that it belongs to $\mathcal{B}^m$, and the inner product relation

$$(\mathbf{K}^m(z,I),\ f(z))=f(I),\quad f\in\mathcal{B}^m.\tag{2.12}$$

By Theorem 2.1 and Proposition 2.2 we can represent $\mathbf{K}^m(z,I)$ as a linear combination of monomials in the determinants $\Delta_{j_1\cdots j_d}(z)$, with $j_1<\cdots<j_d$; the monomials are to be homogeneous of degree e_d in the Δ's with exactly d subscripts, where

$$e_d=m_d-m_{d+1},\quad d\leq n-1,\ e_n=m_n.\tag{2.13}$$

Among these monomials we distinguish one in particular: for brevity, we set

$$\Delta^e(z)=\Delta(1;z)^{e_1}\,\Delta(2;z)^{e_2}\cdots\Delta(n;z)^{e_n}.$$

The set of all other monomials with the given homogeneity will be denoted by S.

Proposition 2.4: If $M\in S$, then $M(I)=0$. On the other hand, $\Delta^e(I)=1$.

Proof: We verify that $\Delta(d;I)=1$, whereas if $(j_1,\ldots,j_d)\neq(1,\ldots,d)$, then $\Delta_{j_1\cdots j_d}(I)=0$. Therefore, a product which includes any such factors $\Delta_{j_1\cdots j_d}(z)$ must vanish when $z=I$.

Proposition 2.5: If $M\in S$ then M is orthogonal to Δ^e.

Proof: Every monomial in the Δ's is an eigenvector of the action of L^m restricted to the diagonal unitary matrices. This follows because, if $\delta=\mathrm{diag}(\delta_1,\ldots,\delta_n)$,

$$(L_\delta\Delta_{j_1\cdots j_d})(z)=\Delta_{j_1\cdots j_d}(z\delta)=\delta_{j_1}\cdots\delta_{j_d}\Delta_{j_1\cdots j_d}(z).$$

So, for each monomial M, there is a multi-index $\mu(M)=(\mu_1(M),\ldots,\mu_n(M))$ such that [with the notation of 2.6]

$$L_\delta M=\delta^{\mu(M)}M,\quad \delta\in\mathcal{D}.$$

In particular, $\mu(\Delta^e)=(m_1,m_2,\ldots,m_n)$. We place a partial order on such multi-indices, as follows: let $\mu'=(\mu'_1,\ldots,\mu'_n)$ and $\mu''=(\mu''_1,\ldots,\mu''_n)$; then $\mu'>\mu''$ if $\mu'_1=\mu''_1,\ldots,\mu'_j=\mu''_j$, $\mu'_{j+1}>\mu''_{j+1}$ for some j. Then if $M\in S$, we have $\mu(\Delta^e)>\mu(M)$; for $\mu(\Delta_{1\cdots d})>\mu(\Delta_{j_1\cdots j_d})$ unless $(1,\ldots,d)=(j_1,\ldots,j_d)$, and both the multi-indices and their ordering are additive. But $\mu(M)\neq\mu(\Delta^e)$ implies that M and Δ^e are orthogonal. This completes the proof.

Of course the analysis of a representation space into eigenspaces of the subgroup $\mathcal{D}$ is the classical theory of weight spaces of representations of semisimple Lie

groups. In the terminology of that theory, we have shown that Δ^e is a generator of the maximal weight space (cf. Refs. 9 and 34).

Proposition 2.6: With the notation of (2.13), we have

$$K^m(z, I) = \|\Delta^e\|^{-2}\, \Delta^e(z). \tag{2.14}$$

Proof: The right-hand side of this equation belongs to $(B)_m$; we must show that it makes (2.12) true. Let f be any element of $(B)^m$; by Proposition 2.2 we can express f as

$$f = \alpha \Delta^e + f_2$$

where f_2 is a combination of monomials in the Δ's other than Δ^e, but belonging to $(B)^m$. By Proposition 2.4, $f(I) = \alpha$. By Proposition 2.5,

$$(\Delta^e, f) = \|\Delta^e\|^2 \alpha + (\Delta^e, f_2) = \|\Delta^e\|^2 \alpha.$$

Therefore we have

$$f(I) = \alpha = \|\Delta^e\|^{-2} (\Delta^e, f). \qquad \text{QED}$$

We must now turn our attention to calculating $\|\Delta^e\|^2$. We shall express it in terms of an integral over $U(n)$. Such integrals are notoriously difficult to evaluate; but we shall be able to sidestep the task, by reducing the integrand to another one of simpler form.

The key to the reduction to $U(n)$ is the fact that when we are finding the norm of a function f in (B), it is possible to reduce the integration over $M(n)$ to integration over a submanifold, namely the set of all z such that the row vectors $z_{1\cdot}, \ldots, z_{n\cdot}$ are pairwise orthogonal. We do this reduction by integrating first with respect to $z_{n\cdot}$, then with respect to $z_{n-1\cdot}$, and so on back to $z_{1\cdot}$. We may assume that $z_{1\cdot}, \ldots, z_{n\cdot}$ are linearly independent; by doing so we are merely excluding a set of measure zero.

Consider the stage in the process at which we have integrated with respect to $z_{j+1\cdot}, \ldots, z_{n\cdot}$. Because of the $(N)_n$-invariance of the function f with which we started, our integrand, a function of $z_{1\cdot}, \ldots, z_{j\cdot}$, is unchanged by the addition to $z_{j\cdot}$ of any linear combination of $z_{1\cdot}, \ldots, z_{j-1\cdot}$. So as a function of $z_{j\cdot}$ it is constant along the subspaces parallel to the linear subspace generated by $z_{1\cdot}, \ldots, z_{j-1\cdot}$. Now the measure with respect to which we are integrating is the Gaussian measure; it can be regarded as the product of the Gaussian measures on that $(j-1)$-dimensional subspace and its orthogonal complement. When we integrate over a translate of the former subspace, because the integrand is constant and the total measure is 1, we simply get the value of the integrand at one point. So we might as well not bother, and just integrate over the latter subspace, that is, the vectors orthogonal to $z_{1\cdot}, \ldots, z_{j-1\cdot}$.

When we have, in this manner, reduced the integration over $z_{j\cdot}$, for each j running from n back to 1, we have a result which can be expressed as follows. Let F_1 be the space of all n-dimensional row vectors; given $z_{1\cdot}$ in F_1, let F_2 be the space of all vectors in F_1 orthogonal to $z_{1\cdot}$; given $z_{2\cdot}$ in F_2, let F_3 be the set of all vectors in F_2 orthogonal to $z_{2\cdot}$; and so on. Then if $f \in (B)$,

$$\int_{M(n)} |f(z)|^2\, d\gamma(z_{n\cdot}) \cdots d\gamma(z_{1\cdot})$$

$$= \int_{F_1} \int_{F_2} \cdots \int_{F_n} |f(z)|^2\, d\gamma_{F_n}(z_{n\cdot}) \cdots d\gamma_{F_2}(z_{2\cdot})\, d\gamma_{F_1}(z_{1\cdot}). \tag{2.15}$$

When we apply (2.15) to a function f in $(B)^m$, we can make a further reduction by carrying out the integration over F_j in two steps. To do this, we set $z_{j\cdot} = \rho_j u_j$, where $\rho_j \geq 0$ and u_j is a vector of unit length, and let $d\sigma_j$ denote a uniform measure on the unit sphere S_j in the $(n-j+1)$-dimensional space F_j; we normalize $d\sigma_j$ so that the total measure of the sphere is $1/(n-j)!$. Then

$$d\gamma_{F_j}(z_{j\cdot}) = 2\rho_j^{2n-2j+1} \exp(-\rho_j)^2\, d\sigma_j(u_{j\cdot})\, d\rho_j. \tag{2.16}$$

Now the vectors $u_{1\cdot}, \ldots, u_{n\cdot}$ are pairwise orthogonal and of length 1; therefore they are the rows of a unitary matrix u. The measure $d\sigma$ defined by

$$\int \psi(u) d\sigma(u) = \int_{S_1} \int_{S_2} \cdots \int_{S_n} \psi(u)\, d\sigma_n(u_{n\cdot}) \cdots d\sigma_2(u_{2\cdot})\, d\sigma_1(u_{1\cdot})$$

is a Haar measure on $U(n)$ (Ref. 34). We may arrange (2.15), with the help of (2.16), in this way:

$$\int_{M(n)} |f(z)|^2\, d\gamma(z)$$

$$= \int_0^\infty \cdots \int_0^\infty \int_{U(n)} |f(\mathrm{diag}(\rho_1, \ldots, \rho_n)u)|^2\, d\sigma(u)$$

$$\times \prod_{j=1}^{n} (2\rho_j^{2n-2j+1} \exp(-\rho_j^2)\, d\rho_j). \tag{2.17}$$

Now if $f \in (B)^m$, then $f(\mathrm{diag}(\rho_1, \ldots, \rho_n)u) = \rho_1^{m_1} \cdots \rho_n^{m_n} f(u)$, and in (2.17) we can carry out the integration over $\rho_1, \ldots, \rho_n$, to find

$$\int_{M(n)} |f(z)|^2\, d\gamma(z) = \int_{U(n)} |f(u)|^2\, d\sigma(u) \prod_{j=1}^{n} (m_j + n - j)! \tag{2.18}$$

Let us apply (2.18) if $f = \Delta^e$, where $e = (e_1, \ldots, e_n)$ is related to $(m_1, \ldots, m_n)$ by (2.13). We set

$$A(m_1, \ldots, m_n) = \int_{M(n)} |\Delta^e(z)|^2\, d\gamma(z),$$

$$B(e_1, \ldots, e_n) = \int_{U(n)} |\Delta^e(u)|^2\, d\sigma(u).$$

Equation (2.18) tells us

$$A(m_1, \ldots, m_n) = B(e_1, \ldots, e_n) \prod_{j=1}^{n} (m_j + n - j)! \tag{2.19}$$

We can carry out an induction on n by making two observations. First, if $m_n = 0$ then $\Delta^e(z)$ is independent of $z_{n\cdot}$, so that

$$A(m_1, \ldots, m_{n-1}, 0) = A(m_1, \ldots, m_{n-1}). \tag{2.20}$$

Second, if $u \in U(n)$ then $|\Delta(n; u)| = 1$; therefore

$$B(e_1, \ldots, e_n) = B(e_1, \ldots, e_{n-1}, 0). \tag{2.21}$$

But we can apply (2.19) with e_n changed to 0:

$$A(m_1 - m_n, \ldots, m_{n-1} - m_n, 0)$$

$$= B(e_1, \ldots, e_{n-1}, 0) \prod_{j=1}^{n-1} (m_j - m_n + n - j)! \tag{2.22}$$

From (2.19) and (2.20), we find

$$A(m_1 - m_n, \ldots, m_{n-1} - m_n, 0)$$

$$= A(m_1 - m_n, \ldots, m_{n-1} - m_n)$$

$$= B(e_1, \ldots, e_{n-1}) \prod_{j=1}^{n-1} (m_j - m_n + n - 1 - j)! \tag{2.23}$$

We can combine (2.21), (2.22), and (2.23) to find

$$B(e_1,\ldots,e_n)=B(e_1,\ldots,e_{n-1})\prod_{j=1}^{n-1}(m_j-m_n+n-j)^{-1}. \quad (2.24)$$

Now $B(e_1)=1$, so from (2.24) we find

$$B(e_1,\ldots,e_n)=\prod_{1\le j<k\le n}(m_j-m_k+k-j)^{-1}. \quad (2.25)$$

Using (2.19) we now conclude

$$A(m_1,\ldots,m_n)=\left(\prod_{j=1}^{n}(m_j+n-j)!\right)\times\left(\prod_{j=1}^{n-1}\prod_{k=j+1}^{n}(m_j-m_k+k-j)\right)^{-1}. \quad (2.26)$$

We summarize the work of this section in the following theorem.

Theorem 2.2: Let $m=(m_1,\ldots,m_n)$; let $e_j=m_j-m_{j+1}$ if $1\le j\le n-1$, and $e_n=m_n$. Then $\mathcal{B}^m=\{0\}$ unless $m_1\ge m_2\ge\cdots\ge m_n\ge 0$. If these inequalities are satisfied, then the reproducing kernel of $\mathcal{B}^m$ is

$$\mathbf{K}^m(z,w)=A(m)^{-1}\Delta^e(zw^*), \quad (2.27)$$

where Δ^e is defined before Proposition (2.4) and $A(m)$ is given by (2.26).

The identification of $\Delta^e(z)$ as a maximal weight vector is found in Moshinsky's papers[9,15]. The normalization factor $A(m)^{-1/2}$ is given in Ref. 5; the proof indicated there apparently is of a combinatorial nature. The importance of the maximal and semimaximal vectors to the representation spaces has been recognized since the publication of Refs. 5, 9, 15; their use as reproducing kernels can be seen in Refs. 11 and 35.

III. THE BRANCHING OPERATORS

A. Global description

The branching operators arise from the embedding of $U(n-1)$ in $U(n)$:

$$u=\begin{pmatrix} u_{11} & \cdots & u_{1,n-1} \\ \vdots & & \\ u_{n-1,1} & \cdots & u_{n-1,n-1} \end{pmatrix}\rightarrow\begin{pmatrix} u_{11} & \cdots & u_{1,n-1} & 0 \\ \vdots & & \vdots & \vdots \\ u_{n-1,1} & \cdots & u_{n-1,n-1} & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \quad (3.1)$$

We shall also denote the larger matrix by u. If the representation $L^{(m_1,\ldots,m_n)}$ of $U(n)$ is restricted to $U(n-1)$, then according to Weyl's branching law (Ref. 3, p. 391) it can be decomposed into a direct sum of distinct irreducible representations: one equivalent to each $L^{m'}$ where $m'=(m'_1,\ldots,m'_{n-1})$ is subject to

$$m_1\ge m'_1\ge m_2\ge\cdots\ge m'_{n-1}\ge m_n. \quad (3.2)$$

We define the branching operator $R^m_{m'}$ to be a map from $\mathcal{B}^{m'}$ to $\mathcal{B}^m$ which realizes this equivalence; that is, it intertwines the given representations of $U(n-1)$ on these two spaces, or if $u\in U(n-1)$

$$L^m_u R^m_{m'}=R^m_{m'}L^{m'}_u. \quad (3.3)$$

$R^m_{m'}$ is then determined up to a multiplicative constant, because the multiplicities in the branching law are all 0 or 1. We shall choose the constant so that $R^m_{m'}$ is a partial isometry.

We shall compute with $R^m_{m'}$ by means of its reproducing kernel $\mathbf{R}^m_{m'}(z,w)$. This is a function of z in $M(n)$ and w in $M(n-1)$. As in (3.1), we will find it convenient to identify a matrix w in $M(n-1)$ with its image in $M(n)$, under the injection which agrees with (3.1):

$$\begin{pmatrix} w_{11} & \cdots & w_{1,n-1} \\ \vdots & & \vdots \\ w_{n-1,1} & \cdots & w_{n-1,n-1} \end{pmatrix}\rightarrow\begin{pmatrix} w_{11} & \cdots & w_{1,n-1} & 0 \\ \vdots & & \vdots & \vdots \\ w_{n-1,1} & \cdots & w_{n-1,n-1} & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

The properties of $R^m_{m'}$ correspond to properties of $\mathbf{R}^m_{m'}$, which we can express as follows. First of all, the image of $R^m_{m'}$ is contained in $\mathcal{B}^m$; therefore (see Proposition 1.4) if w is fixed then $\mathbf{R}^m_{m'}(z,w)$ as a function of z is contained in $\mathcal{B}^m$. The characterization of $\mathcal{B}^m$ by (2.7) tells us, if $\delta\in\mathcal{D}_n$, $g\in\mathcal{N}_n$

$$\mathbf{R}^m_{m'}(\delta gz,w)=\delta^m\mathbf{R}^m_{m'}(z,w). \quad (3.4)$$

Second, the null space of $R^m_{m'}$ is the orthogonal complement of $\mathcal{B}^{m'}$, or equivalently the adjoint $(R^m_{m'})^*$ has $\mathcal{B}^{m'}$ as its image. By an application of Propositions 1.3 and 1.4, we find that for fixed z, $\mathbf{R}^m_{m'}(z,w)^-$ as a function of w belongs to $\mathcal{B}^{m'}$; again (2.7) tells us, if $\delta'\in\mathcal{D}_{n-1}$ and $g'\in\mathcal{N}_{n-1}$

$$\mathbf{R}^m_{m'}(z,\delta'g'w)=(\bar{\delta}')^{m'}\mathbf{R}^m_{m'}(z,w). \quad (3.5)$$

Finally, the requirement that $R^m_{m'}$ satisfy (3.1) implies, as in (2.11), that if $u\in U(n-1)$

$$\mathbf{R}^m_{m'}(z,wu^*)=\mathbf{R}^m_{m'}(zu,w). \quad (3.6)$$

The three properties of $R^m_{m'}$ that we have used characterize the operator completely, except for a constant multiplier. Therefore, Eqs. (3.4), (3.5), and (3.6) characterize $\mathbf{R}^m_{m'}$ up to a constant. The next section is devoted to finding a solution of (3.4)—(3.6), and computing a normalization constant.

B. The reproducing kernel of $R^m_{m'}$

We begin by proving an analog of Proposition 2.3. The argument which led from (2.11) to the equation $\mathbf{K}^m(z,w)=K^m(zw^*,I)$ can be applied here, and we find

$$\mathbf{R}^m_{m'}(z,w)=\mathbf{R}^m_{m'}(zw^*,I). \quad (3.7)$$

Then we can replace (3.3) and (3.4) thus:

$$\mathbf{R}^m_{m'}(gzg',I)=\mathbf{R}^m_{m'}(z,I),\quad g\in\mathcal{N}_n,\ g'\in(\mathcal{N}_{n-1})^*, \quad (3.8)$$

and

$$\mathbf{R}^m_{m'}(\delta z\delta',I)=\mathbf{R}^m_{m'}(z,I)\,\delta^m(\delta')^{m'},$$
$$\delta\in\mathcal{D}_n,\ \delta'\in\mathcal{D}_{n-1}. \quad (3.9)$$

Let us look for a polynomial function that satisfies (3.8) and (3.9). From Sec. IIC, we know that such a function must be formed of minor determinants formed from the top rows of the matrix z; this is a consequence of the invariance under N_n expressed in (3.8). The invariance under $(N_{n-1})^*$ has a similar consequence, but for the columns of z instead of the rows; and because $g'\in GL(n-1,\mathbb{C})$, only the first $n-1$ rows are affected. So, of the columns involved in a minor determinant, we should satisfy the N_{n-1} invariance by taking the first several columns of $(1,2,\ldots,n-1)$, and possibly column n. In summary, the Δ's which satisfy (3.8) are these:

$$\Delta(d;z)=\Delta_{1\dots d}(z),\quad d=1,\dots,n-1,$$

$$\Delta(d;z)=\Delta_{1\dots d-1,n}(z),\quad d=1,\dots,n.$$

We note that $\tilde{\Delta}(n;z)=\Delta(n;z)$; the notation $\tilde{\Delta}(n;z)$ will be more convenient for us. Let us form a plausible-seeming combination of these Δ's. If $f=(f_1,\dots,f_n)$ and $g=(g_1,\dots,g_{n-1})$ are ordered sets of nonnegative integers, consider

$$\tilde{\Delta}^f\,\Delta^g(z)=\tilde{\Delta}(1;z)^{f_1}\,\Delta(1;z)^{g_1}\,\tilde{\Delta}(2;z)^{f_2}\cdots \cdots\,\Delta(n-1;z)^{g_{n-1}}\,\tilde{\Delta}(n;z)^{f_n}. \tag{3.10}$$

The polynomial we have just written down satisfies the requirements of invariance with respect to $(N)_n$ and $((N)_{n-1})^*$. Let us consider the transformation properties with respect to $(D)_n$ and $(D)_{n-1}$ which we require for $\mathbf{R}^m_{m'}$. We have, if $\delta\in(D)_n$,

$$\tilde{\Delta}^f\,\Delta^g(\delta z)=\delta_1^{\tilde{m}_1}\cdots\delta_n^{\tilde{m}_n}\,\tilde{\Delta}^f\,\Delta^g(z),$$

$$\tilde{m}_j=f_j+g_j+\cdots+g_{n-1}+f_n.$$

If $\delta'\in(D)_{n-1}$, then

$$\tilde{\Delta}^f\,\Delta^g(z\delta')=(\delta'_1)^{\tilde{m}'_1}\cdots(\delta'_{n-1})^{\tilde{m}'_{n-1}}\tilde{\Delta}^f\,\Delta^g(z)$$

where

$$\tilde{m}'_j=g_j+f_{j+1}+\cdots+g_{n-1}+f_n.$$

To solve (3.9) with $\tilde{\Delta}^f\Delta^g(z)$ as $\mathbf{R}^m_{m'}(z,I)$, we must have $m_j=\tilde{m}_j$ and $\tilde{m}'_j=m'_j$. These equations can be solved as follows:

$$\begin{aligned} &f_j=m_j-m_j \quad \text{if}\quad 1\le j\le n-1,\\ &f_n=m_n,\\ &g_j=m'_j-m_{j+1}\quad \text{if}\quad 1\le j\le n-1. \end{aligned} \tag{3.11}$$

It is interesting that the "betweenness" relations (3.2) are equivalent to $f_j\ge 0$, $g_j\ge 0$. Now $\tilde{\Delta}^f\,\Delta^g(z)$ satisfies all the requirements which determine $\mathbf{R}^m_{m'}(z,I)$ up to a factor; therefore we may infer

$$\mathbf{R}^m_{m'}(z,w)=K\,\tilde{\Delta}^f\,\Delta^g(zw^*). \tag{3.12}$$

Our task is now to choose a good value of K. We shall choose K to be real and positive, and such that $R^m_{m'}$ is an isometry on $(B)^{m'}$. Then $(R^m_{m'})^*\,(R^m_{m'})$ is the projection onto $(B)^{m'}$, or in the light of Propositions 1.3–1.6

$$\int \mathbf{R}^m_{m'}(z,w)^-\,\mathbf{R}^m_{m'}(z,v)\,d\gamma(z)=K^{m'}(w,v).$$

It is sufficient to apply this formula with $v=w=I$; using Theorem 2.2 we find

$$\int\left|\mathbf{R}^m_{m'}(z,I)\right|^2 d\gamma(z)=A(m')^{-1}. \tag{3.13}$$

From (3.12) and (3.13) we see that, if we define

$$A\binom{m}{m'}=\|\tilde{\Delta}^f\,\Delta^g\|^2, \tag{3.14}$$

then

$$K=A(m')^{-1/2}A\binom{m}{m'}^{-1/2}. \tag{3.15}$$

We have now to evaluate $A\binom{m}{m'}$, by a process similar to that used in Sec. IID. Let us set $f'=(f_1,\dots,f_n,0)$ and $g'=(g_1,\dots,g_{n-2},0)$; our strategy will be to relate $\|\tilde{\Delta}^f\,\Delta^g\|$ to $\|\tilde{\Delta}^{f'}\,\Delta^{g'}\|$. First, we can relate $\|\tilde{\Delta}^f\Delta^g\|$ to an integral over the group $U(n)$; if we set

$$B(f;g)=\int_{U(n)}\left|\tilde{\Delta}^f\Delta^g(u)\right|^2 d\sigma(u),$$

then by (2.18)

$$A\binom{m}{m'}=B(f;g)\prod_{j=1}^{n}(m_j+n-j)!. \tag{3.16}$$

Now $|\tilde{\Delta}(n,u)|=1$, so $B(f;g)$ is independent of f_n, and

$$B_n(f;g)=B_n(f';g)$$

or, using (3.16),

$$\|\tilde{\Delta}^f\,\Delta^g\|^2=\|\tilde{\Delta}^{f'}\,\Delta^g\|^2\prod_{j=1}^{n}\frac{(m_j+n-j)!}{(m_j-m_n+n-j)!}. \tag{3.17}$$

We have made a reduction from f to f'; we have now to reduce from g to g'. In the function $\tilde{\Delta}^{f'}\Delta^g$, the arguments of the nth row do not appear at all, so we may regard $\|\Delta^{f'}\,\Delta^g\|^2$ as an integral over the space $M(n-1,n)$ of rectangular $(n-1)\times n$ matrices. We notice that the entries of column $n-1$ appear only in the factor $\Delta(n-1;z)^{g_{n-1}}$. We will treat that factor as we did $\tilde{\Delta}(n;z)^{f_n}$, by exploiting the fact that $\tilde{\Delta}^{f'}\,\Delta^g$ is invariant under the right action of $(N)_{n-1}{}^*$. Our reasoning is like that which led up to (2.16), but this time we orthogonalize not the rows, but rather the first $n-1$ columns. Let v stand for a unitary matrix with $n-1$ rows and columns; we find

$$\begin{aligned} \|\tilde{\Delta}^{f'}\,\Delta^g\|^2&=\int_{M(n-1,n)}\left|\tilde{\Delta}^{f'}\,\Delta^g(z)\right|^2 d\gamma(z)\\ &=\int_0^\infty\cdots\int_0^\infty\int_{\mathfrak{C}^n}\int_{U(n-1)}\left|\tilde{\Delta}^{f'}\,\Delta^g(v;z_{\cdot n})\right|^2\\ &\quad\times d\sigma'(v)\,d\gamma(z_{\cdot n})\prod_{j=1}^{n-1}(2\rho_j^{\,2m'_j-2m_n+2n-2j-1}\,e^{-\rho_j^2}\,d\rho_j)\\ &=\tilde{B}(f',g)\prod_{j=1}^{n-1}(m_j'-m_n+n-j-1)! \end{aligned} \tag{3.18}$$

where

$$\tilde{B}(f',g)=\int_{\mathfrak{C}^n}\int_{U(n-1)}\left|\tilde{\Delta}^{f'}\,\Delta^g(v;z_{\cdot n})\right|^2 d\sigma'(v)\,d\gamma(z_{\cdot n}).$$

Now $|\Delta(n-1;v,z_{\cdot n})|=1$, so $\tilde{B}(f',g)$ is independent of g_{n-1}, and

$$\tilde{B}(f,g)=\tilde{B}(f',g').$$

Using (3.18) we convert this equation to

$$\|\tilde{\Delta}^{f'}\,\Delta^g\|^2=\|\tilde{\Delta}^{f'}\,\Delta^{g'}\|^2\prod_{j=1}^{n-1}\frac{(m'_j-m_n+n-j-1)!}{(m'_j-m'_{n-1}+n-j-1)!}. \tag{3.19}$$

Combining (3.17) and (3.19), we find

$$\begin{aligned} \|\tilde{\Delta}^f\,\Delta^g\|^2&=\|\tilde{\Delta}^{f'}\,\Delta^{g'}\|^2\prod_{j=1}^{n}(m_j+n-j)!\\ &\quad\times\prod_{j=1}^{n-1}(m_j'-m_n+n-j-1)!\\ &\quad\times\prod_{j=1}^{n-1}[(m_j-m_n+n-j)!\,(m_j'-m'_n+n-j-1)!]^{-1}. \end{aligned} \tag{3.20}$$

This equation is a recursion from n to $n-1$, because on the one hand $\|\tilde{\Delta}^f\,\Delta^g\|^2=A\binom{m}{m'}$, while on the other hand if

$$\bar{m}=(m_1-m'_{n-1},\dots,m_{n-1}-m'_{n-1}),$$

$$\bar{m}'=(m_1'-m'_{n-1},\dots,m'_{n-2}-m'_{n-1}),$$

then $\|\tilde{\Delta}^{f'}\,\Delta^{g'}\|^2=A\binom{\bar{m}}{\bar{m}'}$. To express the numerical value of $A^m_{m'}$ let us use the following notation:

$$\mu_j = m_j + n - j, \quad \mu'_j = m'_j + n - j - 1,$$

$$P(\mu) = \prod_{j<k} (\mu_j - \mu_k)!$$

$$P(\mu') = \prod_{j<k} (\mu'_j - \mu'_k)!,$$

$$Q(\mu, \mu') = \prod_{j<k} (\mu'_j - \mu_k)!,$$

$$\tilde{Q}(\mu', \mu) = \prod_{j\leq k} (\mu_j - \mu'_k + 1)!.$$

Then

$$\|\tilde{\Delta}^f \Delta^g\|^2 = A\binom{m}{m'} = \frac{\prod_{j=1}^n \mu_j!\, Q(\mu;\mu')\tilde{Q}(\mu;\mu')}{P(\mu)P(\mu')}. \tag{3.21}$$

We may summarize the work of this section in a theorem

Theorem 3.1: Let $(m_1, \ldots, m_n)$ and $(m'_1, \ldots, m'_{n-1})$ satisfy the betweenness relations $m_j \geq m'_j \geq m_{j+1}$. Then there is a branching operator $R^m_{m'}$ from $(B)^{m'}$ to $(B)^m$, which is an isometry on $(B)^{m'}$. If f and g are defined by (3.11), and $A(m)$ and $A\binom{m}{m'}$ by (2.26) and (3.21), then with the notation of (3.10) the reproducing kernel of $R^m_{m'}$ is

$$\mathbf{R}^m_{m'}(z,w) = A(m')^{-1/2}\, A\binom{m}{m'}^{-1/2} \tilde{\Delta}^f \Delta^g(zw^*). \tag{3.22}$$

At this point it is clear what a strategic role is played in the internal organization of the representation spaces by the maximal and semimaximal states. The maximal state of $(B)^m$, normalized, is $A(m)^{-1/2}\Delta^e(z)$. The semimaximal states are the images in $(B)^m$ of the maximal states of the $(B)^{m'}$ for which m and m' satisfy the betweenness relations; and they are just

$$A\binom{m}{m'}^{-1/2} \Delta^f \Delta^g(z).$$

This formula appears in the literature starting in 1963 (Ref. 5, (50) and (52); Ref. 15). To prove it, we observe that the maximal state in $(B)^{m'}$ is

$$A(m')^{1/2}\, \mathbf{K}^m(w, I).$$

If we act on this by $R^m_{m'}$, then according to Proposition 1.2, the result is

$$A(m')^{1/2} \int \mathbf{R}^m_{m'}(z,w)\mathbf{K}^{m'}(w,I)\, d\gamma(w)$$
$$= A(m')\, \mathbf{R}^m_{m'}(z, I)$$
$$= A\binom{m}{m'}^{-1/2} \tilde{\Delta}^f \Delta^g(z).$$

IV. GEL'FAND BASIS VECTORS

A. Introduction: The case of $SU(3)$

Having defined the branching operators, we may define an orthonormal basis of each space $(B)^m$ in the following way. The procedure is inductive on n and assumes that orthonormal bases have been defined for the spaces $(B)^{m'}$ such that $R^m_{m'} \neq 0$. Then each of these basis elements may be mapped into $(B)^m$ by the branching operator $R^m_{m'}$; the resulting vectors are an orthonormal basis of $(B)^m$. Each basis element is thus of the form

$$\Gamma = R^m_{m'} R^{m'}_{m''} \cdots R^{m_{12}m_{22}}_{m_{11}} R^{m_{11}}_{*}(1), \tag{4.1}$$

where m, m', m'', ..., $(m_{12}m_{22})$, m_{11} are ordered sets of $n, n-1, \ldots, 1$ indices, all satisfying the betweenness relations. We may (following Refs. 1, 2, 4) use a double subscript notation

$$m = (m_{1n}, \ldots, m_{nn}),$$
$$m' = (m_{1,n-1}, \ldots, m_{n-1,n-1}), \text{ etc.}, \tag{4.2}$$

and write (4.1) as

$$\Gamma\begin{pmatrix} m_{1n} & & m_{2n} & \cdots & m_{nn} \\ & m_{1,n-1} & \cdots & m_{n-1,n-1} & \\ & & m_{11} & & \end{pmatrix} = \Gamma(M).$$

In principle, one could say that (4.1) defines $\Gamma(M)$, and indeed every $\Gamma(M)$ could be found from (4.1) by a finite number of steps. This procedure would be far from satisfactory as a practical technique. To obtain a more enlightening view of the Gel'fand bases, let us look at some low-dimensional cases. First, for future reference, we prove a proposition about the action of $(D)_n$ on $\Gamma(M)$. (This was already known, of course, to Gel'fand and his co-workers.[2,4])

Proposition 4.1: The Gel'fand vector $\Gamma(M)$ is an eigenvector of the action of $(D)_n$; if $w_j = m_{1j} + \cdots + m_{jj}$, and $\delta = \mathrm{diag}(\delta_1, \ldots, \delta_n)$, then

$$L_\delta \Gamma(M) = \delta_1^{w_1} \delta_2^{w_2 - w_1} \cdots \delta_n^{w_n - w_{n-1}} \Gamma(M). \tag{4.3}$$

Proof: Let $\delta' = \mathrm{diag}(\delta_1 \cdots \delta_{n-1}, 1) \in (D)_{n-1}$ and $\delta'' = \mathrm{diag}(1, \ldots, 1, \delta_n)$. Then from (3.3) we have

$$L^m_\delta R^m_{m'} = L^m_{\delta''} L^m_{\delta'} R^m_{m'} = L^m{}_{\delta''} R^m_{m'} L^{m'}_{\delta'}.$$

Now

$$L^m_{\delta''} R^m_{m'} = \delta_n^{w_n - w_{n-1}} R^m_{m'}. \tag{4.4}$$

Indeed

$$(L^m_{\delta''} R^m_{m'} \mathbf{e}_w)(z) = R^m_{m'}(\mathbf{e}_w)(z\delta'')$$
$$= \mathbf{R}^m_{m'}(z\delta'', w)$$
$$= \mathbf{R}^m_{m'}(z\delta'' w^*, I)$$
$$= \mathbf{R}^m_{m'}(zw^*\delta'', I)$$

where, at the last step, we have observed that the matrix δ'' commutes with anything of the form (3.1); the right action by δ'' multiplies every determinant $\Delta(d;z)$ by δ_n, and so we see from (3.12) that it multiplies $R^m_{m'}(zw^*, I)$ by this power of δ_n:

$$f_1 + \cdots + f_n = m_1 - m_1' + \cdots + m_{n-1} - m'_{n-1} + m_n$$
$$= w_n - w_{n-1}.$$

The proposition follows from (4.4) by an easy inductive argument.

Let us consider the cases of $U(1)$ and $U(2)$. The irreducible representations of $U(1)$ are one-dimensional, because the group is commutative; the Gel'fand basis vector in $(B)^{m_{11}}$ is

$$\Gamma(m_{11}; z) = (m_{11}!)^{-1/2} \Delta_1(z). \tag{4.5}$$

The Gel'fand basis vectors in a representation of $U(2)$ are the semimaximal states, computed in the previous section. We have

$$\Gamma\begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & \end{pmatrix}$$
$$= A\begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & \end{pmatrix}^{-1/2} \tilde{\Delta}(1)^{f_1} \tilde{\Delta}(2)^{f_2} \Delta(1)^{g_1}$$
$$= A\begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & \end{pmatrix}^{-1/2} \Delta_1^{m_{11} - m_{12}} \Delta_2^{m_{12} - m_{11}} \Delta_{12}^{m_{22}} \tag{4.6}$$

or

$$\Gamma \begin{matrix} m_{12} & & m_{22} \\ & m_{11} & \end{matrix}$$

$$= [(m_{12} - m_{22} + 1)!\,]^{1/2} \times [(m_{12} + 1)!\, m_{22}!\, (m_{11} - m_{22})!\, (m_{12} - m_{11})!\,]^{-1/2} \times \Delta_1^{\,m_{11} - m_{22}} \Delta_2^{\,m_{12} - m_{11}} \Delta_{12}^{\,m_{22}}. \tag{4.7}$$

For the Gel'fand vectors of $U(3)$ we cannot expect such simple expressions. The reproducing kernel $R^m_{m'}$ can be made to yield expressions for them, if we use the fact (Proposition 1.6) that

$$R^{\begin{matrix} m_{13} & m_{23} \\ m_{12} & m_{22} \end{matrix} m_{33}}(z, w) = \sum_{m_{11}} \Gamma \begin{pmatrix} m_{13} & m_{23} & m_{33} \\ & m_{12} & m_{22} \\ & m_{11} & \end{pmatrix} ; z \times \Gamma \begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & \end{pmatrix} ; w \Big)^{-}. \tag{4.8}$$

We can manipulate (3.22) into a form like (4.8), by expanding the determinants. We make use of a few simple cases of a classical formula in determinant theory; for future use we state the proposition in this way:

Proposition 4.2: Let $z \in M(n)$ and $w \in M(n-1)$; inject w into $M(n)$ as in (3.1). If $S_d(n)$ denotes the set of all ordered d-tuples $(j_1, j_2, \ldots, j_d)$ such that $j_1 < j_2 < \cdots < j_d \leq n$, then

$$\Delta(d; zw^*) = \sum_{J \in S_d(n-1)} \Delta_J(z)\, \bar{\Delta}_J(w) \tag{4.9}$$

$$\tilde{\Delta}(d; zw^*) = \sum_{J \in S_{d-1}(n-1)} \Delta_{Jn}(z)\, \bar{\Delta}_J(w). \tag{4.10}$$

Proof: By a well-known formula (cf. Ref. 36), if $J,\ J' \in S_d(n)$ and $z,\ v \in M(n)$

$$\Delta^{J'}_{J}(zv) = \sum_{J'' \in S_d(n)} \Delta^{J'}_{J''}(z)\, \Delta^{J''}_{J}(v).$$

We apply this formula with $J' = (1, \ldots, d)$ and $v = w^*$, and J either $(1, \ldots, d)$ or $(1, \ldots, d-1, n)$. We then have $\Delta^{J'}_{J''}(z) = \Delta_{J''}(z)$. If $J = (1, \ldots, d)$, then

$$\Delta^{J''}_{J}(v) = \bar{\Delta}^{J}_{J''}(w) = \bar{\Delta}_{J''}(w).$$

We arrive at the formula (4.9); Note that because w lies in $M(n-1)$, $\Delta_{J''}(w) = 0$ if J'' contains the index n. In the second case, we find

$$\tilde{\Delta}(d; zw^*) = \sum_{J'' \in S_d(n)} \Delta_{J''}(z)\, \bar{\Delta}^{1 \cdots d-1, n}_{J''}(w).$$

In this sum, the special form of w implies that $\bar{\Delta}^{1 \cdots d-1, n}_{J''}(w)$ vanishes if J'' does not contain n, while

$$\Delta^{1 \cdots d-1\ n}_{J\quad\ n}(w) = \Delta^{1 \cdots d-1}_{J}(w).$$

So we arrive at (4.10).

Now let us use (4.9) and (4.10) to expand the expression for the branching kernel, as given in (3.22). We find

$$\mathbf{R}^m_{m'}(z, w) = A(m')^{-1/2} A \binom{m}{m'}^{-1/2} \Delta_1(zw^*)^{g_1} \Delta_3(zw^*)^{f_1} \times \Delta_{12}(zw^*)^{g_2} \Delta_{13}(zw^*)^{f_2} \Delta_{123}(zw^*)^{f_3}$$

$$= A(m')^{-1/2} A \binom{m}{m'}^{-1/2} \times (\Delta_1(z)\bar{\Delta}_1(w) + \Delta_2(z)\,\bar{\Delta}_2(w))^{g_1} \Delta_3(z)^{f_1} \times \Delta_{12}(z)^{g_2}\, \bar{\Delta}_{12}(w)^{g_2+f_3} (\Delta_{13}(z)\bar{\Delta}_1(w) + \Delta_{23}(z)\bar{\Delta}_2(w))^{f_2} \Delta_{123}(z)^{f_3} \tag{4.11}$$

or [after using (3.11) and the binomial theorem]

$$\mathbf{R}^m_{m'}(z, w) = A(m')^{-1/2} A \binom{m}{m'}^{-1/2} \sum_{\alpha, \beta} \binom{m_{12} - m_{23}}{\alpha} \binom{m_{23} - m_{22}}{\beta} \times \Delta_1(z)^{\alpha} \Delta_2(z)^{m_{12} - m_{23} - \alpha} \Delta_3(z)^{m_{13} - m_{12}} \Delta_{13}(z)^{\beta} \times \Delta_{23}(z)^{m_{23} - m_{22} - \beta} \Delta_{123}(z)^{m_{33}} \Delta_{12}(z)^{m_{22} - m_{23}} \times [\bar{\Delta}_1(w)^{\alpha+\beta}\, \bar{\Delta}_2(w)^{m_{12} - m_{22} - \alpha - \beta} \Delta_{12}(w)^{m_{22}}]. \tag{4.12}$$

Now compare the factors involving w in this expression with (4.7). To extract from (4.12) an expression which multiplies

$$\Gamma \begin{pmatrix} m_{12} & & m_{22} \\ & m_{11} & \end{pmatrix} ; w \Big)^{-},$$

we must collect the terms of the sum for which $\alpha + \beta = m_{11} - m_{22}$, and for normalization we must multiply them by $A(^{m'}_{m_{11}})^{1/2}$. Thus, (4.8) implies

$$\Gamma \begin{pmatrix} m_{13} & m_{23} & m_{33} \\ & m_{12} & m_{22} \\ & m_{11} & \end{pmatrix} = [(m_{13} - m_{23} + 1)!\,(m_{13} - m_{33} + 2)! \times (m_{23} - m_{33} + 1)!\,(m_{11} - m_{22})!\,(m_{12} - m_{11})! \times (m_{12} - m_{22} + 1)\,]^{1/2} \times [(m_{13} + 2)!\,(m_{23} + 1)!\, m_{33}!\,(m_{12} - m_{23})!\,(m_{22} - m_{33})! \times (m_{13} - m_{12})!\,(m_{12} - m_{33} + 1)! \times (m_{23} - m_{22})!\,(m_{13} - m_{22} + 1)!\,]^{-1/2} \times \sum_{\alpha + \beta = m_{11} - m_{22}} \binom{m_{12} - m_{23}}{\alpha} \binom{m_{23} - m_{22}}{\beta} \times \Delta_1^{\alpha} \Delta_2^{\,m_{12} - m_{23} - \alpha} \Delta_3^{\,m_{13} - m_{12}} \times \Delta_{12}^{\,m_{22} - m_{33}} \Delta_{13}^{\,\beta} \Delta_{23}^{\,m_{23} - m_{22} - \beta} \Delta_{123}^{\,m_{33}}. \tag{4.13}$$

The requirements that $0 \leq \alpha \leq m_{12} - m_{23}$ and $0 \leq \beta \leq m_{23} - m_{22}$ are implicit in the binomial coefficients.

Now another expression for the boson polynomials of $U(3)$ is known (Ref. 5, Eqs. (43), (44)):

$$\Gamma \begin{pmatrix} m_{13} & m_{23} & 0 \\ & m_{12} & m_{22} \\ & m_{11} & \end{pmatrix} = [(m_{11} - m_{22})!\,(m_{12} - m_{23})! \times (m_{12} - m_{22} + 1)\,(m_{13} - m_{23} + 1)!\,]^{1/2}\, [(m_{11} - m_{23})!\,]^{-1} \times [(m_{12} + 1)!\, m_{22}!\,(m_{12} - m_{11})!\,(m_{23} - m_{22})! \times (m_{13} - m_{22} + 1)!\,(m_{13} - m_{12})!\,]^{-1/2} \times \Delta_1^{\,m_{11} - m_{23}} \Delta_2^{\,m_{12} - m_{11}} \Delta_3^{\,m_{13} - m_{12}} \Delta_{12}^{\,m_{22}} \Delta_{13}^{\,m_{23} - m_{22}} \times {}_2F_1(m_{22} - m_{23},\ m_{11} - m_{12};\ m_{11} - m_{23} + 1;\ \Delta_1\Delta_{23}/\Delta_2\Delta_{13}). \tag{4.14}$$

It might appear that this formula is well-defined only if $m_{11} - m_{23} \geq 0$; if $m_{11} - m_{23} < 0$ then the factorial of a

negative number appears in the normalization coefficient, and zero denominators in the hypergeometric series. Working formally, one can make these two troubles neutralize each other, by recourse to this relation (Ref. 37, Eq. 15.1.2):

$$\lim_{c\to -m}\frac{1}{\Gamma(c)}F(a,b;c;z)$$
$$=(a)_{m+1}(b)_{m+1}((m+1)!)^{-1/2}$$
$$\times z^{m+1}F(a+m+1,\ b+m+1;m+2;z).$$

We apply this formula with $-m=m_{11}-m_{23}+1$, and find a counterpart to (4.14), valid if $m_{11}\leq m_{23}$:

$$\Gamma\begin{pmatrix} m_{13} & m_{23} & 0\\ & m_{12} & m_{22}\\ & & m_{11}\end{pmatrix}=[(m_{23}-m_{22})!(m_{12}-m_{11})!$$
$$\times(m_{12}-m_{22}+1)(m_{13}-m_{23}+1)!]^{1/2}[(m_{23}-m_{11})!]^{-1}$$
$$\times[(m_{12}+1)!m_{22}!(m_{12}-m_{23})!(m_{11}-m_{22})!$$
$$\times(m_{13}-m_{22}+1)!(m_{13}-m_{12})!]^{-1/2}$$
$$\times\Delta_2^{m_{12}-m_{23}}\Delta_3^{m_{13}-m_{12}}\Delta_{12}^{m_{22}}\Delta_{13}^{m_{11}-m_{22}}\Delta_{23}^{m_{23}-m_{11}}$$
$$\times{}_2F_1(m_{22}-m_{11},m_{23}-m_{12};\ m_{23}-m_{11}+1;\ \Delta_1\Delta_{23}/\Delta_2\Delta_{13}). \tag{4.15}$$

Let us verify that (4.14) and (4.15) agree with (4.13). If $m_{11}\geq m_{23}$, then the conditions on α, implicit in (4.13), can be stated as follows: If we set

$$j=\alpha-m_{11}+m_{23},$$

then

$$0\leq j\leq\min(m_{12}-m_{11},\ m_{23}-m_{22}).$$

The same range of the summation index j is to be discerned in the hypergeometric series that occurs in (4.14). The rest of the verification is routine. In the case $m_{11}\leq m_{23}$, the condition on α is

$$0\leq\alpha\leq\min(m_{12}-m_{23},\ m_{11}-m_{22})$$

and α can be identified with the summation index in the hypergeometric series that occurs in (4.15). Again, the verification offers no difficulty except tedium.

We are now in the comfortable position of having found out something we already knew. To carry our evaluation of Gel'fand states beyond the case of $U(3)$, by the means so far employed, would be extremely difficult. One could, indeed, expand $R^{m}_{m'}(z,w)$ with $m=(m_{14}m_{24}m_{34}m_{44})$ and $m'=(m_{13}m_{23}m_{33})$ as in (4.11), but in the resulting sea of determinants it would be very difficult to recognize the Gel'fand states in w, or to collect terms properly. The trouble lies in the fact that there are identities among the determinants $\Delta_{j_1\cdots j_d}$ when $n\geq 3$, so that the expansion analogous to (4.11) is full of redundancy. To get the algebra under control, we must consider our equation (4.13) from a different viewpoint, which is explained in the next section.

B. Generating functions: for $U(1)$, $U(2)$, and $U(3)$

If one contemplates the sum in (4.13), one may be struck by the fact that every monomial of the form

$$\Delta_1^{\alpha_1}\Delta_2^{\alpha_2}\Delta_3^{\alpha_3}\Delta_{12}^{\alpha_{12}}\Delta_{13}^{\alpha_{13}}\Delta_{23}^{\alpha_{23}}\Delta_{123}^{\alpha_{123}} \tag{4.16}$$

occurs in the expression of precisely one of the Gel'fand polynomials. The exponents are related to the indices in the following way:

$$\alpha_1+\alpha_2=m_{12}-m_{23},\quad \alpha_1+\alpha_{13}=m_{11}-m_{22},$$
$$\alpha_{13}+\alpha_{23}=m_{23}-m_{22},\quad [\alpha_2+\alpha_{23}=m_{12}-m_{11}], \tag{4.17}$$
$$\alpha_3=m_{13}-m_{12},\quad \alpha_{12}=m_{22}-m_{33},\quad \alpha_{123}=m_{33}.$$

The equation in square brackets is redundant, but is included for the sake of symmetry. The coefficient of (4.17) includes

$$[\alpha_1!\,\alpha_2!\,\alpha_{13}!\,\alpha_{23}!]^{-1},$$

and other factors which are determined by the indices m_{ij}. These circumstances suggest that it may be useful to combine the boson polynomials, multiplied by suitable coefficients, in a generating function in which all the monomials (4.17) occur, with coefficients that cause the entire sum to have a simple structure.

To explain how we shall make this combination, we introduce some convenient notation. As we have done before, we use the letter M to stand for the array of indices (m_{ij}). A boldface $\boldsymbol{\alpha}$ will denote the set of seven integers $\alpha_1,\ldots,\alpha_{123}$ in (4.17), and that monomial itself will be denoted $\boldsymbol{\Delta}^{\boldsymbol{\alpha}}$. The product $\alpha_1!\cdots\alpha_{123}!$ will be denoted $\boldsymbol{\alpha}!$ Our crucial observation is that if $\Gamma(M)$, as defined by (4.13), is multiplied by an appropriate factor $N(M)$, then every term in the product is of the form $\boldsymbol{\Delta}^{\boldsymbol{\alpha}}/\boldsymbol{\alpha}!$, and the terms that occur are precisely those for which M and $\boldsymbol{\alpha}$ are related by (4.17). If we use the notation

$$S(\boldsymbol{\alpha},M) \tag{4.18}$$

to stand for the relations (4.17), then

$$N(M)\,\Gamma(M)=\sum_{\boldsymbol{\alpha}:S(\boldsymbol{\alpha},M)}\boldsymbol{\Delta}^{\boldsymbol{\alpha}}/\boldsymbol{\alpha}! \tag{4.19}$$

Now we must add together all Eqs. (4.19), multiplied by suitably chosen powers of indeterminates, to form a generating function for the set of all the boson polynomials. These being determined by six indices, the summation must be sixfold. The most convenient set of indices is not M but rather

$$F=\begin{pmatrix} f_{13} & f_{23} & f_{33}\\ & f_{12} & f_{22}\\ & & f_{11}\end{pmatrix}$$

where [cf. (3.11)]

$$f_{ij}=m_{ij}-m_{i,j-1},\quad i\leq j-1,$$
$$f_{ii}=m_{ii}. \tag{4.20}$$

We introduce indeterminates

$$\Lambda=\begin{pmatrix} \lambda_{13} & \lambda_{23} & \lambda_{33}\\ & \lambda_{12} & \lambda_{22}\\ & & \lambda_{11}\end{pmatrix}$$

and set

$$\Lambda^F=\prod_{i\leq j}\lambda_{ij}^{f_{ij}}.$$

So, from (4.19), we find

$$\sum_M\Lambda^F N(M)\Gamma(M)=\sum_M\sum_{\boldsymbol{\alpha}:S(\boldsymbol{\alpha},M)}\Lambda^F\boldsymbol{\Delta}^{\boldsymbol{\alpha}}/\boldsymbol{\alpha}! \tag{4.21}$$

As we already remarked, every α appears in (4.22) exactly once, so α determines an array M and, by (4.20), an array F. For the latter, we find

$$\begin{aligned} f_{13} &= \alpha_3, \\ f_{23} &= \alpha_{13} + \alpha_{23}, \\ f_{33} &= \alpha_{123}, \\ f_{12} &= \alpha_2 + \alpha_{23}, \\ f_{22} &= \alpha_{12} + \alpha_{123}, \\ f_{11} &= \alpha_1 + \alpha_{12} + \alpha_{13} + \alpha_{123}. \end{aligned} \tag{4.22}$$

Writing (4.21) with more detail on the right-hand side, we find

$$\begin{aligned} &\sum_M \Lambda^F N(M)\Gamma(M) \\ &\quad = \sum_\alpha \lambda_{13}{}^{\alpha_3}\lambda_{23}{}^{\alpha_{13}+\alpha_{23}}\lambda_{33}{}^{\alpha_{123}}\lambda_{12}{}^{\alpha_2+\alpha_{23}}\lambda_{22}{}^{\alpha_{12}+\alpha_{123}} \\ &\qquad \times \lambda_{11}{}^{\alpha_1+\alpha_{12}+\alpha_{13}+\alpha_{123}}\Delta^\alpha/\alpha! \\ &\quad = \sum_\alpha (\lambda_{11}\Delta_1)^{\alpha_1}(\lambda_{12}\Delta_2)^{\alpha_2}(\lambda_{13}\Delta_3)^{\alpha_3}(\lambda_{11}\lambda_{22}\Delta_{12})^{\alpha_{12}} \\ &\qquad \times (\lambda_{11}\lambda_{23}\Delta_{13})^{\alpha_{13}}(\lambda_{12}\lambda_{23}\lambda_{23})^{\alpha_{23}}(\lambda_{11}\lambda_{22}\lambda_{33}\Delta_{123})^{\alpha_{123}}/\alpha! \end{aligned} \tag{4.23}$$

or

$$\begin{aligned} &\sum_M \Lambda^F N(M)\Gamma(M) \\ &\quad = \exp(\lambda_{11}\Delta_1 + \lambda_{12}\Delta_2 + \lambda_{13}\Delta_3 + \lambda_{11}\lambda_{22}\Delta_{12} \\ &\qquad + \lambda_{11}\lambda_{23}\Delta_{13} + \lambda_{12}\lambda_{23}\Delta_{23} + \lambda_{11}\lambda_{22}\lambda_{33}\Delta_{123}). \end{aligned} \tag{4.24}$$

The expression in brackets follows a pattern: The determinant Δ_J is multiplied by λ_{ij} if the ith element in J is j.

The normalizing factor is found to be

$$\begin{aligned} N(M) = &[(m_{13}+2)!\,(m_{23}+1)!\,(m_{12}-m_{33}+1)! \\ &\times (m_{13}-m_{22}+1)!\,]^{1/2}\,[(m_{13}-m_{23}+1)! \\ &\times (m_{13}-m_{33}+2)!\,(m_{23}-m_{33}+1)!\,(m_{11}-m_{22})! \\ &\times (m_{12}-m_{11})!\,m_{33}!\,(m_{12}-m_{23})!\,(m_{22}-m_{33})! \\ &\times (m_{13}-m_{12})!\,(m_{23}-m_{22})!\,(m_{12}-m_{22}+1)!]^{-1/2}. \end{aligned} \tag{4.25}$$

The result (4.24) is unexpectedly simple, and capable of being generalized. Before we look ahead to $U(4)$ or $U(n)$, let us look back at the boson polynomials of $U(2)$ and $U(1)$, where the equations analogous to (4.24) are easily found from (4.7) and (4.5). We have

$$\begin{aligned} &\sum \lambda_{12}{}^{m_{12}-m_{11}}\lambda_{22}{}^{m_{22}}\lambda_{11}{}^{m_{11}}N\begin{pmatrix} m_{12} & m_{22} \\ & m_{11} \end{pmatrix}\Gamma\begin{pmatrix} m_{12} & m_{22} \\ & m_{11} \end{pmatrix} \\ &\quad = \exp(\lambda_{11}\Delta_1 + \lambda_{12}\Delta_2 + \lambda_{11}\lambda_{22}\Delta_{12}) \end{aligned}$$

where

$$\begin{aligned} N\begin{pmatrix} m_{12} & m_{22} \\ & m_{11} \end{pmatrix} = &[(m_{12}+1)!\,]^{1/2}[(m_{11}-m_{22}+1)!\,m_{22}! \\ &\times (m_{11}-m_{22})!\,(m_{12}-m_{11})!\,]^{-1/2}. \end{aligned}$$

And, of course,

$$\sum \lambda_{11}{}^{m_{11}}N(m_{11})\,\Gamma(m_{11}) = \exp(\lambda_{11}\Delta_1)$$

where

$$N(m_{11}) = (m_{11}!\,)^{-1/2}.$$

Some hint as to what is going on with these normalization coefficients may be gained from these relationships:

$$N\begin{pmatrix} m_{12} & m_{22} & 0 \\ m_{12} & m_{22} & \\ & m_{11} & \end{pmatrix} = N\begin{pmatrix} m_{12} & m_{22} \\ & m_{11} \end{pmatrix}, \tag{4.26}$$

$$N\begin{pmatrix} m_{11} & 0 \\ & m_{11} \end{pmatrix} = N(m_{11}). \tag{4.27}$$

To prove that the three cases we have seen are examples of a general pattern is the task of the next section. For more information on the boson polynomials, see Ref. 33.

C. A partial extension to $U(n)$

We shall construct a generating function, similar to (4.24), for the boson polynomials of the representations of L^m of $U(n)$ with $m = (m_1, m_2, 0, \cdots, 0)$. These representations correspond to Young diagrams of one or two rows. It will be convenient to simplify the general notation m_{ij} for the Gel'fand indices by setting

$$\begin{aligned} &m_{1j} = a_j, \quad 1 \leq j \leq n, \\ &m_{2j} = b_j, \quad 2 \leq j \leq n, \quad b_1 = 0. \end{aligned}$$

So, in most of this section, a typical Gel'fand pattern is

$$M = \begin{pmatrix} a_n & & b_n & & 0 \cdots 0 & 0 \\ & a_{n-1} & & b_{n-1} & 0 \cdots 0 & \\ & \ddots & & \ddots & & \\ & & a_2 & & b_2 & \\ & & & a_1 & & \end{pmatrix}. \tag{4.28}$$

We let M' denote the pattern (4.28) with the top row deleted. Instead of the array (f_{ij}), we use

$$\begin{aligned} &p_j = a_j - a_{j-1}, \quad 2 \leq j \leq n, \quad p_1 = a_1, \\ &q_j = b_j - b_{j-1}, \quad 3 \leq j \leq n, \quad q_2 = b_2. \end{aligned} \tag{4.29}$$

Theorem 4.1: The boson states $\Gamma(M)$ such that M is given by (4.28) have the generating function

$$\begin{aligned} &\sum_{a,b} N(M)\Gamma(M)\lambda_1{}^{p_1}\cdots\lambda_n{}^{p_n}\mu_2{}^{q_2}\cdots\mu_n{}^{q_n} \\ &\quad = \exp\left(\sum_j \lambda_j\Delta_j + \sum_{j<k}\lambda_j\mu_k\Delta_{jk}\right) \end{aligned} \tag{4.30}$$

where

$$\begin{aligned} N(M) = &\left(\prod_{j=2}^{n}(a_j - b_{j-1}+1)!\right)^{1/2} \\ &\times \left((a_n - b_n)!\prod_{j=1}^{n}(p_j!\,q_j!\,(a_{j-1}-b_j)!\,(a_j - b_j+1))\right)^{-1/2}. \end{aligned} \tag{4.31}$$

The proof of these formulas is inductive. We note that they are valid if $1 \leq n \leq 3$. Suppose the result is proved for $n-1$; we want to prove it for n. We start with the definition of $\Gamma(M)$, in the form [cf. (4.1)]

$$\Gamma(M\,;z) = (R_m^m \cdot \Gamma(M'))(z)$$

where

$$m = (a_n, b_n, 0, \ldots, 0), \quad m' = (a_{n-1}, b_{n-1}, 0, \ldots, 0);$$

applying Proposition 1.2, we have

$$\Gamma(M;z)=\int_{M(n-1)}\mathbf{R}^{m}_{m'}(z,w)\Gamma(M';w)\,d\gamma(w). \tag{4.32}$$

Now by (3.22),

$$\mathbf{R}^{m}_{m'}(z,w)=A(m)^{-1/2}A\binom{m}{m'}^{-1/2}\Delta_n(zw^*)^{p_n}\times\Delta_{1n}(zw^*)^{q_n}\Delta_1(zw^*)^{a_{n-1}-b_n}\Delta_{12}(zw^*)^{b_{n-1}}. \tag{4.33}$$

We must expand this expression into a sum of terms, each of which is a product of a factor depending on z and a factor depending on w. By Proposition 4.2, we have

$$\Delta_1(zw^*)=\sum_{j=1}^{n-1}\Delta_j(z)\bar{\Delta}_j(w) \tag{4.34}$$

$$\Delta_n(zw^*)=\Delta_n(z) \tag{4.35}$$

$$\Delta_{12}(zw^*)=\sum_{j<k\leq n-1}\Delta_{jk}(z)\bar{\Delta}_{jk}(w) \tag{4.36}$$

$$\Delta_{1n}(zw^*)=\sum_{j=1}^{n-1}\Delta_{jn}(z)\bar{\Delta}_j(w). \tag{4.37}$$

We substitute (4.35)—(4.37) in (4.34) and expand by the multinomial theorem. To express the result of this manipulation, we find it convenient to introduce indices α_j, with $1\leq j\leq n$, and α_{jk}, with $1\leq j\leq k\leq n$; we then set

$$\Delta^{\alpha}(z)=\prod_j\Delta_j(z)^{\alpha_j}\prod_{j<k}\Delta_{jk}(z)^{\alpha_{jk}}$$

and

$$\alpha!=\prod_j\alpha_j!\prod_{j<k}\alpha_{jk}!.$$

Further, we set

$$\beta_j=\alpha_j+\alpha_{jn},\quad 1\leq j\leq n-1, \tag{4.38}$$

$$S(M)=\left\{\alpha:\sum_{j=1}^{n-1}\alpha_j=a_{n-1}-b_n,\ \alpha_n=p_n,\ \sum_{j<k}\alpha_{jk}=b_{n-1},\ \sum\alpha_{jn}=q_n\right\}. \tag{4.39}$$

We find

$$\mathbf{R}^{m}_{m'}(z,w)=A(m)^{-1/2}A\binom{m}{m'}^{-1/2}\sum_{\alpha\in S(M)}p_n!\,q_n!\,(a_{n-1}-b_n)!\times b_{n-1}!\,(\alpha!)^{-1}\Delta^{\alpha}(z)\Delta^{\beta}(w)^{-}. \tag{4.40}$$

Now let us put

$$\int_{M(n-1)}\Delta^{\beta}(w)^{-}\Gamma(M';w)\,d\gamma(w)=(\beta\,|\,M'). \tag{4.41}$$

We then find from (4.33) and (4.40) that

$$\Gamma(m;z)=A(m)^{-1/2}A\binom{m}{m'}^{-1/2}p_n!\,q_n!\,(a_{n-1}-b_n)!\,b_{n-1}!\times\sum_{\alpha\in S(M)}(\beta\,|\,M')\Delta^{\alpha}(z)/\alpha!. \tag{4.42}$$

Further progress on (4.42) would seem to depend on evaluating $(\beta\,|\,M')$. Rather than do so, we shall use certain aspects of the structure of (4.42) to relate a combination of the $\Gamma(M;z)$, for fixed M', to $\Gamma(M',z)$. We shall then be able to carry out the induction step. Our manipulation depends at one step on regarding the determinants $\Delta_J(z)$ as if they were independent variables. Of course they are no such thing; so we must take care to justify that step.

First, we introduce $n(n+1)/2$ independent variables

$$\hat{\Delta}_j,\quad 1\leq j\leq n;\quad \hat{\Delta}_{jk},\quad 1\leq j\leq k\leq n.$$

We use the same indices α and β as before and set

$$\hat{\Delta}^{\alpha}=\prod_j\hat{\Delta}_j^{\alpha_j}\prod_{j<k}\hat{\Delta}_{jk}^{\alpha_{jk}}.$$

We define polynomials $\hat{\Gamma}(M)$, for M of the form (4.28), by

$$\hat{\Gamma}(M)=A(m)^{-1/2}A\binom{m}{m'}^{-1/2}p_n!\,q_n!\,(a_{n-1}-b_n)!\,b_{n-1}!\times\sum_{\alpha\in S(M)}(\beta\,|\,M')\hat{\Delta}^{\alpha}/\alpha!. \tag{4.43}$$

The reader should note the similarity of form between (4.43) and (4.42), while remaining aware of the important distinction between Δ_j, Δ_{jk} on the one hand and $\hat{\Delta}_j$, $\hat{\Delta}_{jk}$ on the other.

Consider the polynomials $\hat{\Gamma}(M)$ for a fixed choice of the subarray M' but all possible choices of the indices a_n and b_n. From (4.43) we find

$$\sum_{a_n,b_n}A(m)^{1/2}A\binom{m}{m'}^{1/2}[p_n!\,q_n!\,(a_{n-1}-b_n)!\,b_{n-1}!]^{-1}\times\hat{\Gamma}(M)\lambda_n^{p_n}\mu_n^{q_n}=\sum_{a_n,b_n}\sum_{\alpha\in S(M)}\hat{\Delta}^{\alpha}(\beta\,|\,M')\lambda_n^{p_n}\mu_n^{q_n}/\alpha!. \tag{4.44}$$

Now we must take the condition $\alpha\in S(M)$, defined by (4.39) and express it in two stages: first, a condition on the multi-index β, and then a relation between α, β, and (a_n,b_n). From

$$\sum_{j=1}^{n-1}\alpha_j=a_{n-1}-b_n\ \text{ and }\ \sum_{j=1}^{n-1}\alpha_{jn}=q_n=b_n-b_{n-1}$$

we infer

$$\sum_{j=1}^{n-1}\beta_j=a_{n-1}-b_{n-1}.$$

Therefore, if we define

$$S_1(M')=\{\beta:\sum_{j<k}\beta_{jk}=b_{n-1},\ \sum_j\beta_j=a_{n-1}-b_{n-1}\}$$

and

$$S_2(\beta;a_n,b_n)=\{\alpha:\alpha_{jk}=\beta_{jk}\text{ if }j<k<n;\ \alpha_j+\alpha_{jn}=\beta_j;\ \alpha_n=P_n;\ \sum\alpha_{jn}=q_n\},$$

then

$$\alpha\in S(M)\Leftrightarrow\beta\in S_1(M')\text{ and }\alpha\in S_2(\beta;a_n,b_n).$$

Therefore we may express the right-hand side of (4.44) as follows:

$$\sum_{a_n,b_n}\sum_{\beta\in S_1(M')}\sum_{\alpha\in S_2(\beta;a_n,b_n)}\hat{\Delta}^{\alpha}(\beta\,|\,M')\lambda_n^{\alpha_n}\mu_n^{\Sigma\alpha_{jn}}/\alpha!$$

$$=\sum_{\beta\in S_1(M')}(\beta\,|\,M')\sum_{\alpha_j+\alpha_{jn}=\beta_j}\left(\prod_{j=1}^{n-1}\frac{(\hat{\Delta}_j)^{\alpha_j}}{\alpha_j!}\right)\frac{(\lambda_n\hat{\Delta}_n)^{\alpha_n}}{\alpha_n!}\times\left(\prod_{j=1}^{n-1}\frac{(\mu_n\hat{\Delta}_{jn})^{\alpha_{jn}}}{\alpha_{jn}!}\right)\prod_{j<k}\frac{\hat{\Delta}_{jk}{}^{\beta_{jk}}}{\beta_{jk}!}$$

$$=\exp(\lambda_n\hat{\Delta}_n)\sum_{\beta\in S_1(M')}(\beta\,|\,M')\prod_{j=1}^{n-1}\frac{(\hat{\Delta}_j+\mu_n\hat{\Delta}_{jn})^{\beta_j}}{\beta_j!}\prod_{j<k}\frac{\hat{\Delta}_{jk}{}^{\beta_{jk}}}{\beta_{jk}!}.$$

The most important properties of this expression are the fact that it depends on $\hat{\Delta}_j$, on μ_n, and on $\hat{\Delta}_{jn}$ only by way of the combination

$$\hat{\Delta}_j + \mu_n \hat{\Delta}_{jn},$$

and the factor $\exp(\lambda_n \hat{\Delta}_n)$. These properties must be shared by the left side of (4.44). It follows that if we obtain any other expression for the left side of (4.44) in the special case $\lambda_n = \mu_n = 0$, then we can obtain the general case by two operations: the transformation

$$\hat{\Delta}_j \mapsto \hat{\Delta}_j + \mu_n \hat{\Delta}_{jn} \tag{4.45}$$

and multiplying by $\exp(\lambda_n \hat{\Delta}_n)$.

Now in the special case $\lambda_n = \mu_n = 0$, the right side of (4.44) is just

$$A(m')[(a_{n-1} - b_{n-1})!\, b_{n-1}!]^{-1} \hat{\Gamma}(M'^0) \tag{4.46}$$

where

$$M'^0 = \begin{pmatrix} a_{n-1} & & b_{n-1} & & 0 \cdots 0 \\ & a_{n-1} & & b_{n-1} & 0 \cdots 0 \\ & & a_{n-2} & & b_{n-2} \\ & & \cdots & & \cdots \end{pmatrix}.$$

What else can we say about (4.43)? Two observations can be made. First, if we substitute $\Delta_J(z)$ for $\hat{\Delta}_J$, then $\hat{\Gamma}(M'^0)$ becomes $\Gamma(M'^0;z)$; indeed that substitution turns $\hat{\Gamma}(M)$ into $\Gamma(M;z)$ for any M, as we remind ourselves by comparing (4.42) and (4.43). Second,

$$\Gamma(M'^0;z) = \Gamma(M';z). \tag{4.47}$$

This equation follows from (4.32) and the fact that, in (3.22), if $m_j = m'_j$ and $m_n = 0$, then $\mathbf{R}^m_{m'} = \mathbf{K}^m$.

But, by the induction hypothesis, we have a generating function for $\Gamma(M';z)$. It might seem, therefore, that we can obtain a generating function for the polynomials $\Gamma(M;z)$ by applying the transformations which were described at the end of the previous paragraph but one. There is a catch, however. The transformation (4.45) makes sense because the variables $\hat{\Delta}_J$ are independent, whereas (4.47) is an equation involving the function $\Delta_J(z)$, which are not algebraically independent. If we could assert not merely (4.47), but an equality between $\hat{\Gamma}(M'^0)$ and $\hat{\Gamma}(M')$, then we could proceed easily. As we only have (4.47), we must carry out the induction step as follows.

We have already supposed (4.30) to be true with $n-1$ in place of n. That is, if we define polynomials $\hat{\Gamma}(M')$ in the $\hat{\Delta}_J$ by

$$\sum_{M'} N(M')\, \hat{\Gamma}(M')\, \lambda_1^{p_1} \cdots \lambda_{n-1}^{p_{n-1}} \mu_2^{q_2} \cdots \mu_{n-1}^{q_{n-1}} = \exp\left(\sum_{j=1}^{n-1} \lambda_j \hat{\Delta}_j + \sum_{j<k\leq n-1} \lambda_j \mu_k \hat{\Delta}_{jk} \right),$$

then the substitution of $\Delta_J(z)$ for $\hat{\Delta}_J$ turns $\hat{\Gamma}(M')$ into $\Gamma(M';z)$. It should be emphasized that this definition of $\hat{\Gamma}(M')$, where M' is a Gel'fand pattern (4.28) with $n-1$ rows, is of a different sort from the definition (4.43) of $\hat{\Gamma}(M)$ where M has n rows. We may infer from (4.47) that

$$\hat{\Gamma}(M'^0) = \hat{\Gamma}(M') + R_{M'},$$

where $R_{M'}$ is a polynomial function of the $\hat{\Delta}_J$ that vanishes if $\hat{\Delta}_J = \Delta_J(z)$. So we have

$$\sum_{M'} N(M')(\hat{\Gamma}(M'^0) - R_{M'})\, \lambda_1^{p_1} \cdots \lambda_{n-1}^{p_{n-1}} \mu_2^{q_2} \cdots \mu_{n-1}^{q_{n-1}} = \exp\left(\sum_{j=1}^{n-1} \lambda_j \hat{\Delta}_j + \sum_{j<k\leq n-1} \lambda_j \mu_j \hat{\Delta}_{jk} \right). \tag{4.48}$$

Now let us make the transformation (4.45) and multiply by $\exp(\lambda_n \hat{\Delta}_n)$. The polynomial $R_{M'}$ is transformed to something else, say $\tilde{R}_{M'} \exp(\lambda_n \hat{\Delta}_n)$. Our reasoning that led from (4.44) to (4.45) shows us that $\hat{\Gamma}(M'^0)$ is transformed into

$$\sum_{a_n, b_n} A(m')^{-1/2} A \binom{m}{m'}^{+1/2} (a_{n-1} - b_{n-1})!\,[p_n!\, q_n!\, (a_{n-1} - b_n)!]^{-1} \times \hat{\Gamma}(M)\, \lambda_n^{p_n} \mu_n^{q_n}.$$

Therefore, if we set

$$N(M) = A(m')^{-1/2} A \binom{m}{m'}^{+1/2} \times (a_{n-1} - b_{n-1})!\,[p_n!\, q_n!\, (a_{n-1} - b_n)!]^{-1} N(M'), \tag{4.49}$$

then (4.48) becomes

$$\sum_{M} N(M)\, \hat{\Gamma}(M)\, \lambda_1^{p_1} \cdots \lambda_n^{p_n} \mu_2^{q_2} \cdots \mu_n^{q_n} = \exp\left(\sum_{j=1}^{n} \lambda_j \hat{\Delta}_j + \sum_{j<k\leq n} \lambda_j \mu_k \hat{\Delta}_{jk} \right) + \exp(\lambda_n \hat{\Delta}_n) \sum_{M'} N(M')\, \tilde{R}_{M'}\, \lambda_1^{p_1} \cdots \lambda_{n-1}^{p_{n-1}} \mu_2^{q_2} \cdots \mu_{n-1}^{q_{n-1}}. \tag{4.50}$$

If, in (4.50), we substitute $\Delta_J(z)$ for $\hat{\Delta}_J$, then we obtain (4.30), if $\tilde{R}_{M'}$ is made 0 by that substitution.

Now $R_{M'}$ was a polynomial which was annihilated by that substitution. All the relations among the $\Delta_J(z)$ and $\Delta_{kl}(z)$ are consequences of the following: If $1 \leq j < k < l \leq n-1$, then

$$\Delta_j(z)\Delta_{kl}(z) - \Delta_k(z)\,\Delta_{jl}(z) + \Delta_l(z)\,\Delta_{jk}(z) = 0. \tag{4.51}$$

Therefore, $R_{M'}$ belongs to the polynomial ideal generated by the polynomials

$$\hat{\Delta}_j \hat{\Delta}_{kl} - \hat{\Delta}_k \hat{\Delta}_{jl} + \hat{\Delta}_l \hat{\Delta}_{jk}.$$

If we made the transformation (4.45) then the above expression becomes

$$(\hat{\Delta}_j \hat{\Delta}_{kl} - \hat{\Delta}_k \hat{\Delta}_{jl} + \hat{\Delta}_l \hat{\Delta}_{jk}) + \lambda_n(\hat{\Delta}_{jn} \hat{\Delta}_{kl} - \hat{\Delta}_{kn} \hat{\Delta}_{jl} + \hat{\Delta}_{ln} \hat{\Delta}_{jk}).$$

Now we already know that the first parenthesis vanishes if $\hat{\Delta}_J = \Delta_J(z)$; and so does the second, as we can easily check. In short, the relation (4.51) is not disrupted by the transformation (4.45); the polynomial $\tilde{R}_{M'}$ vanishes if $\hat{\Delta}_J = \Delta_J(z)$; (4.50) implies (4.30); the induction step is valid.

All that remains is to verify (4.31). From (4.49), using (2.26) and (3.21), we find

$$N(M) = \left(\frac{(a_n - b_n + 1)!\,(a_{n-1} - b_{n-1})!}{p_n!\, q_n!\, (a_{n-1} - b_n)!\, (a_n - b_n + 1)!} \right)^{1/2} N(M') = \left(\frac{(a_{n-1} - b_{n-1})!}{(a_n - b_n)!} \right)^{1/2} \times \left(\frac{(a_n - b_{n+1})!}{p_n!\, q_n!\, (a_{n-1} - b_n)!\, (a_n - b_n + 1)} \right)^{1/2} N(M').$$

Therefore, (4.31) can be proved by induction. The proof of the theorem is complete.

The difficulty in this proof, related to (4.51), occurs with greater complexity if we try to carry out the argument for representations L^m with m more general than $(m_1, m_2, 0 \cdots 0)$. That part of the induction step involved in going between (4.44) and (4.46) may be generalized; but the relations among the determinants $\Delta_J(z)$ seem to be disrupted. The author has not succeeded in generalizing the proof of Theorem 4.1.

V. PRODUCTS OF REPRESENTATIONS OF SU(2)

A. Relation to $U(n)$

It is well known that (in the notation of Ref. 19, Sec. 2) the representation $\mathcal{D}_j$ of $SU(2)$ with angular momentum j is equivalent to the representation of $SU(2)$ on $\mathcal{B}^{(2j,0)}$, and the normalized state with z-component of angular momentum equal to m can be taken to be

$$\psi^j_m(z_{11}, z_{12}) = \Gamma\begin{pmatrix} 2j & 0 \\ j+m & ; z \end{pmatrix} = z_{11}{}^{j+m} z_{12}{}^{j-m}/[(j+m)!\,(j-m)!\,]^{1/2}.$$

The coupling of n angular momenta $j_1, \cdots, j_n$ can be described by the space $\mathcal{Q}(j_1 \cdots j_n)$ of polynomials in $2n$ variables $(z_{ab} : 1 \leq a \leq n,\ b = 1, 2)$, homogeneous in z_{a1} and z_{a2} of degree $2j_a$. [Thus $\mathcal{Q}(j)$ is synonymous with $\mathcal{B}^{(2j,0)}$.] If we let z stand for the $n \times 2$ matrix

$$z = \begin{pmatrix} z_{11} & z_{12} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ z_{n1} & z_{n2} \end{pmatrix} \in M(n, 2),$$

then the action of $SU(2)$ on $\mathcal{Q}(j_1, \ldots, j_n)$ given by

$$(R_u f)(z) = f(zu), \quad u \in SU(2), \quad f \in \mathcal{Q}(j_1 \cdots j_n),$$

is the Kronecker product

$$\mathcal{D}_{j_1} \otimes \cdots \otimes \mathcal{D}_{j_n}.$$

In this section we shall explicitly decompose this representation into its irreducible components.

The first step is almost a standard opening in this game: Shift attention from covariants to invariants.

Proposition 5.1: There is a one-to-one correspondence between embeddings T of $\mathcal{Q}(j)$ into $\mathcal{Q}(j_1 \cdots j_n)$ that intertwine the representations of $SU(2)$, and invariants f_T in $\mathcal{Q}(j_1 \cdots j_n j)$, given, if we set $w = (w_1, w_2)$, and $z \in M(n, 2)$, by

$$f_T\begin{pmatrix} z \\ w \end{pmatrix} = \sum_{m=-j}^{j} (-1)^{j+m} \psi^j_{-m}(w)\, T(\psi^j_m)(z). \tag{5.1}$$

Proof: Any linear map whatever from $\mathcal{Q}(j)$ to $\mathcal{Q}(j_1 \cdots j_n)$ can be associated with a polynomial in $\mathcal{Q}(j_1 \cdots j_n j)$ by (5.1); the correspondence is one-to-one both ways. We observe that

$$\sum_m (-1)^{j+m} \psi^j_{-m}(w_1, w_2)\psi^j_m(\zeta_1, \zeta_2) = \sum_m (-1)^{j+m} w_1{}^{j-m} w_2{}^{j+m} \zeta_1{}^{j+m} \zeta_2{}^{j-m}/[(j+m)!\,(j-m)!\,] = (w_1\zeta_2 - w_2\zeta_1)^{2j}/(2j)!,$$

and this is invariant under the action of $SU(2)$ on the row vectors $w = (w_1, w_2)$ and $\zeta = (\zeta_1, \zeta_2)$. Therefore, if $u \in SU(2)$,

$$\sum_m (-1)^{m+j} \psi^j_{-m}(wu)(R_u \psi^j_m) = \sum_m (-1)^{m+j} \psi^j_{-m}(w)\psi^j_m.$$

Let us apply to both sides of this equation an operator $T : \mathcal{Q}(j) \to \mathcal{Q}(j_1 \cdots j_n)$. We find

$$\sum_m (-1)^{m+j} \psi^j_{-m}(wu)\, T(R_u \psi^j_m)(z) = \sum_m (-1)^{m+j} \psi^j_{-m})(z) = f_T\begin{pmatrix} z \\ w \end{pmatrix}. \tag{5.2}$$

On the one hand, if $TR_u = R_u T$, then we have

$$f_T\begin{pmatrix} z \\ w \end{pmatrix} = \sum_m (-1)^{m+j} \psi^j_{-m}(wu) R_u (T\psi^j_m)(z) = \sum_m (-1)^m \psi^j_{-m}(wu)(T\psi^j_m)(zu) = f_T\begin{pmatrix} zu \\ wu \end{pmatrix};$$

that is, f_T is invariant. On the other hand, if f_T is invariant, then we infer from (5.2) that

$$\sum_m (-1)^{m+j} \psi^j_{-m}(wu)\, T(R_u \psi^j_m)(z) = f_T\begin{pmatrix} z \\ w \end{pmatrix} = f_T\begin{pmatrix} zu \\ wu \end{pmatrix} = \sum_m (-1)^{m+j} \psi^j_{-m}(wu)\, T(\psi^j_m)(zu).$$

But the polynomials $\psi^j_{-m}(wu)$ are linearly independent, so

$$T(R_u \psi^j_m)(z) = T(\psi^j_m)(zu) = R_u (T\psi^j_m)(z);$$

therefore $TR_u = R_u T$.

This proposition shows that we can find all the subrepresentations of $\mathcal{D}_{j_1} \otimes \cdots \otimes \mathcal{D}_{j_n}$ equivalent to $\mathcal{D}_j$ by finding invariants in $\mathcal{Q}(j_1 \cdots j_n j)$. We must consider the normalization of our maps $\mathcal{Q}(j) \to \mathcal{Q}(j_1 \cdots j_n)$, and if the multiplicity of $\mathcal{D}_j$ in $\mathcal{D}_{j_1} \otimes \cdots \otimes \mathcal{D}_{j_n}$ is greater than 1, we want to select pairwise orthogonal subspaces of $\mathcal{Q}(j_1 \cdots j_n)$. Both of these matters are covered by the following proposition.

Proposition 5.2: Let T_1 and T_2 be maps $\mathcal{Q}(j) \to \mathcal{Q}(j_1 \cdots j_n)$ which intertwine the representations of $SU(2)$, and let f_{T_1} and f_{T_2} be the invariants associated to T_1 and T_2 by Proposition 5.1. Then, if $\psi,\ \psi' \in \mathcal{Q}(j)$,

$$(T_2\psi', T_1\psi) = (2j+1)^{-1}(f_{T_2}, f_{T_1})(\psi', \psi).$$

Proof: The operator $T_2{}^* T_1$ commutes with $SU(2)$ acting on the irreducible subspace $\mathcal{Q}(j)$, so it must be a multiple of the identity; say $T_2{}^* T_1 = cI$. Equivalently,

$$(T_2\psi', T_1\psi) = c(\psi', \psi).$$

What is c? Consider f_{T_1} and f_{T_2}. From (5.1) we see

$$(f_{T_2}, f_{T_1}) = \sum_{m,m'} (\psi^j_{-m}, \psi^j_{-m'})(T_2\psi^j_m, T_1\psi^j_{m'}).$$

But the ψ^j_m are an orthonormal basis of $\textcircled{Q}(j)$, so we have

$$(f_{T_2}, f_{T_1}) = \sum_{m=-j}^{j} c = (2j+1)c.$$

This completes the proof.

We can now describe the relationship between the generalized vector coupling coefficients and the invariants. Suppose $\{f_\sigma\}$ is a complete orthonormal system of invariants in $\textcircled{Q}(j_1 \cdots j_n)$; set

$$f_\sigma(z) = \sum_{m_1 \cdots m_n} \begin{pmatrix} \sigma; m_1 \cdots m_n \\ j_1 \cdots j_n \end{pmatrix} \psi^{j_1}_{m_1}(z_{11}, z_{12}) \cdots \psi^{j_n}_{m_n}(z_{n1}, z_{n2}). \tag{5.3}$$

Then the ranges of the corresponding operators T_σ are orthogonal to one another, and span that subspace of $\textcircled{Q}(j_1 \cdots j_{n-1})$ which transforms like $\textcircled{D}(j_n)$. The normalized vectors which transform like $\psi^{j_n}_{m_n}$ are

$$\Psi^{\sigma j}_{m_n} = (2j_n+1)^{1/2} \sum_{m_1 \cdots m_{n-1}} \begin{pmatrix} \sigma; m_1 \cdots m_{n-1} j_n \\ j_1 \cdots j_{n-1} m_n \end{pmatrix} \times \psi^{j_1}_{m_1}(z_{11}, z_{12}) \cdots \psi^{j_{n-1}}_{m_{n-1}}(z_{n-1,1}, z_{n-1,2}) \tag{5.4}$$

where

$$\begin{pmatrix} \sigma; m_1 \cdots m_{n-1} j_n \\ j_1 \cdots j_{n-1} m_n \end{pmatrix} = (-1)^{m_n + j_n} \begin{pmatrix} m_1 \cdots m_{n-1} - m_n \\ j_1 \cdots j_{n-1} j_n \end{pmatrix}. \tag{5.5}$$

Our task is now to find a complete orthonormal system of invariants in $\textcircled{Q}(j_1 \cdots j_n)$. We do this by means of the following construction, which is due to Moshinsky.[9] (See also Ref. 11 for a more general result.)

Proposition 5.3: The polynomials

$$f_M(z) = \Gamma(M; \tilde{z})$$

where

$$z = \begin{pmatrix} z_{11} & z_{12} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ z_{n1} & z_{n2} \end{pmatrix}, \qquad \tilde{z} = \begin{pmatrix} z_{11} \cdots z_{n1} \\ z_{12} \cdots z_{n2} \\ 0 \cdots 0 \\ \cdot \quad \cdot \\ \cdot \quad \cdot \\ \cdot \quad \cdot \\ 0 \cdots 0 \end{pmatrix}$$

for which

$$a_n = b_n = j_1 + \cdots + j_n, \tag{5.6a}$$

$$a_k + b_k = 2(j_1 + \cdots + j_k), \quad 2 \le k \le n-1, \tag{5.6b}$$

$$a_1 = 2j_1 \tag{5.6c}$$

are a complete orthonormal system of invariants in $\textcircled{Q}(j_1 \cdots j_n)$. If $j_1 + \cdots + j_n$ is not integral, then there are no nonzero invariants in $\textcircled{Q}(j_1 \cdots j_n)$.

Note that the matrix z is transposed in the boson polynomial. This detail reflects Moshinsky's basic insight: The definition of the representation spaces $\textcircled{B}^m$ and the action of $U(n)$ on those spaces are complementary. The former uses row operations on the matrix z, that is, multiplication on the left; the latter uses multiplication on the right.

Proof: The last sentence in the statement of the proposition is easy to prove: If $2(j_1 + \cdots + j_n)$ is odd, then for $u = -I$, $R_u f = -f$ for all f in $\textcircled{Q}(j_1 \cdots j_n)$. In the case that $j_1 + \cdots + j_n$ is an integer, we begin by showing that the invariants in $\textcircled{Q}(j_1 \cdots j_n)$ are related to certain elements of $B^{(a_n b_n 0 \cdots 0)}$, where (5.6a) is true. Let f be such an invariant, and let

$$f_0(z) = f\begin{pmatrix} z_{11} & z_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ z_{1n} & z_{2n} \end{pmatrix}, \quad z \in M(n).$$

Now f is invariant under multiplication on the right, i.e.,

$$f(zu) = f(z), \quad z \in M(n, 2),$$

if $u \in SU(2)$. By Proposition 1.7, this invariance can be analytically continued to all u in $SL(2, \mathbb{C})$. It implies that f_0 is invariant on the left, i.e.,

$$f_0(gz) = f_0(z), \quad z \in M(n),$$

$$g = \begin{pmatrix} g_{11} & g_{12} & 0 \cdots 0 \\ g_{21} & g_{22} & 0 \cdots 0 \\ 0 & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & 1 \end{pmatrix}, \quad \det g = 1. \tag{5.7}$$

But $f_0(z)$ does not depend on the elements of z below the second row, so we may put arbitrary entries in the third and lower rows of g. In particular, we have

$$f_0(gz) = f_0(z) \tag{5.8a}$$

with [cf. (2.4)]

$$g = \begin{pmatrix} 1 & 0 \cdots 0 \\ g_{21} & 1 \cdots 0 \\ \cdot & \cdot \cdot \\ \cdot & \cdot \cdot \\ \cdot & \cdot \cdot \\ g_{n1} & g_{n2} \quad 1 \end{pmatrix} \in \textcircled{N}_n. \tag{5.8b}$$

Equations (5.8a)–(5.8b) imply that $f_0 \in \textcircled{B}$, which we recall is the set of all elements of $\textcircled{E}$ satisfying (2.5), and is the orthogonal direct sum of the spaces $\textcircled{B}^m$. To see which of the latter f_0 belongs to, we consider (2.6). Let $\delta = \mathrm{diag}(\delta_1, \delta_1^{-1}, \delta_3, \ldots, \delta_n)$; then because $\mathrm{diag}(\delta_1, \delta_1^{-1}) \in SL(2, \mathbb{C})$, (5.7) says that $f_0(\delta z) = f_0(z)$. Therefore, we must have $m_3 = \cdots = m_n = 0$, and $m_1 = m_2$. If $\delta = \mathrm{diag}(\delta_1, \delta_1, 1, \ldots, 1)$, then by the fact that $f \in \textcircled{Q}(j_1 \cdots j_n)$ we have $f_0(\delta z) = \delta_1^{2(j_1 + \cdots + j_n)} f_0(z)$. Therefore, $m_1 + m_2 = 2(j_1 + \cdots + j_n)$; that is,

$$f_0 \in B^{(J, J, 0, \ldots, 0)}, \quad J = j_1 + \cdots + j_n.$$

We have placed f_0 in the space spanned by Gel'fand polynomials with the first row specified by (5.6a). To show that (5.6b) and (5.6c) are appropriate, we consider the action of diagonal matrices on the right on f_0. Because $f \in \textcircled{Q}(j_1 \cdots j_n)$, we have [with $\delta = \mathrm{diag}(\delta_1, \ldots, \delta_n)$]

$$f(\delta z) = \delta_1^{2j_1} \cdots \delta_n^{2j_n} f(z)$$
$$f_0(z\delta) = \delta_1^{2j_1} \cdots \delta_n^{2j_n} f_0(z), \quad z \in M(n). \tag{5.9}$$

When we compare this statement with Proposition 4.1, we see that f_0 belongs to the subspace spanned by the

Gel'fand polynomials whose indices satisfy (5.6a)—(5.6c).

To complete the proof, we must show that each such boson polynomial actually gives rise to an $SU(2)$ invariant. But by Proposition 2.2, because $a_n = b_n$ the polynomials we are concerned with are expressible in terms of the determinants $\Delta_{jk}(z)$, and these are indeed invariant for $SU(2)$.

B. Explicit decomposition

We have obtained a solution to the problem of resolving $(D)_{j_1} \otimes \cdots \otimes (D)_{j_{n-1}}$ into irreducible subspaces on which the representation is equivalent to $(D)_{j_n}$. Indeed, (5.3)—(5.5) relate these subrepresentations to invariants in $(Q)(j_1 \cdots j_n)$; Proposition 5.3 shows that certain boson polynomials are a complete orthonormal system of such invariants, and Theorem 4.1 gives us these boson polynomials, wrapped up in a generating function. The task of this section is to unwrap these results, in an explicit formula for the coupling coefficients.

First, we can make explicit the "extra quantum numbers" denoted by σ in (5.3); the boson polynomials in which we are interested are labelled by patterns

$$M = \begin{pmatrix} a_n & & b_n & & 0 \cdots 0 \\ & a_{n-1} & & b_{n-1} & \cdot \quad \cdot \\ & \cdot & & \cdot & \cdot \quad \cdot \\ & \cdot & & \cdot & \cdot \quad \cdot \\ & & \cdot & & 0 \\ & & a_2 & b_2 & \\ & & & a_1 & \end{pmatrix}$$

satisfying (5.6a)—(5.6c). It is convenient to introduce, besides the angular momenta $j_1, \ldots, j_n$, members $r_1 \cdots r_n$ defined so that

$$a_i = j_1 + \cdots + j_i + r_i, \quad b_i = j_1 + \cdots + j_i - r_i. \tag{5.10}$$

Then (5.6b) is satisfied. To satisfy (5.6a) and (5.6c) we must have

$$r_1 = j_1, \quad r_n = 0. \tag{5.11}$$

We then have

$$p_i = j_i + (r_i - r_{i-1}), \quad q_i = j_i - (r_i - r_{i-1}); \tag{5.12}$$

since these numbers must be nonnegative integers, we must have

$$|r_i - r_{i-1}| \leq j_i, \quad r_i - r_{i-1} \equiv j_i \pmod 1. \tag{5.13}$$

The requirement on the Gel'fand pattern that $a_{i-1} \geq b_i$ implies

$$r_i + r_{i-1} \geq j_i. \tag{5.14}$$

Because $a_{n-1} = a_n$, we also have $r_{n-1} = j_n$. Conditions (5.13) and (5.14) can be summarized by saying that $r_1 = j_1$, the triples (j_2, r_1, r_2), (j_3, r_2, r_3), and so on to $(j_{n-1}, r_{n-2}, r_{n-1})$ satisfy the triangle inequalities and the condition that the sum of each triple shall be a whole number, while finally $r_{n-1} = j_n$. With these indices $r_1, \ldots, r_n$, we can replace (5.3) by the more explicit expression

$$\Gamma(M;z) = \sum_{m_1 \cdots m_n} \begin{pmatrix} m_1 \cdots m_n \\ j_1 \cdots j_n \\ r_1 \cdots r_n \end{pmatrix} \psi^{j_1}_{m_1}(z_{11}, z_{12}) \cdots \psi^{j_n}_{m_n}(z_{n1}, z_{n2}). \tag{5.15}$$

When we expand (4.30) and collect the terms with appropriate powers of $\lambda_1, \ldots, \lambda_n, \mu_2, \ldots, \mu_n$, we get an expression for the coefficients in (5.15). The number of summation indices in the expression for one such coefficient turns out to be $(n-2)^2$. The generating function for $\Gamma(M;z)$ is also a generating function for the coefficients; in the case $n = 3$, we recover essentially the generating function for the $3-j$ symbols, as obtained by Bargmann [Ref. 18, (3.21)] and Schwinger (in Ref. 22).

VI. CONCLUDING REMARKS

The theory presented here is by no means the last word. In the first place, the multiple vector coupling coefficients of Sec. V require further study. The number of summation indices, $(n-2)^2$, is too large. For invariant combinations of n angular momenta, the simple scheme of using $3-j$ symbols to couple j_1 and j_2, coupling the resulting angular momentum r_2 with j_3, and so on, would require no more than $n-2$ summations. If one could simplify the expression (5.16) for the boson polynomials by eliminating the appearance of some of the determinants Δ_{hi}, then the subsequent formulas would be simplified.

The description in Sec. IV of certain boson polynomials can be extended. First of all, one would like to explore the boson polynomials of representations L^m where $m = (m_1, m_2, m_3, \cdots)$ with $m_3 > 0$. As was mentioned at the end of Sec. IV C, the induction procedure used there does not seem to be valid in the more general setting, without some modifications.

A broader extension of the boson polynomials can be contemplated. Let us consider the matrix elements of L^m with respect to the Gel'fand basis. It is convenient to use the abbreviation m for $(m_1, \ldots, m_n)$ and to use the letters M', N', $\cdots$ for Gel'fand patterns of $n-1$ rows satisfying the "betweenness" relations with respect to m as the top row. We denote the matrix elements by

$$L\begin{pmatrix} N' \\ m \\ M' \end{pmatrix};u),$$

defined by the equation:

$$L^m_u \Gamma\begin{pmatrix} m \\ M' \end{pmatrix} = \sum_N L\begin{pmatrix} N' \\ m \\ M' \end{pmatrix};u)\Gamma\begin{pmatrix} m \\ N' \end{pmatrix}. \tag{6.1}$$

From (2.1) we see that we may write (6.1) as:

$$\Gamma\begin{pmatrix} m \\ M' \end{pmatrix};zu) = \sum_{N'} L\begin{pmatrix} N' \\ m \\ M' \end{pmatrix};u)\,\Gamma\begin{pmatrix} m \\ N' \end{pmatrix};z). \tag{6.2}$$

Or, because of the orthonormality of the $\Gamma(M)$,

$$L\begin{pmatrix} N' \\ m \\ M' \end{pmatrix};u) = \int \Gamma\begin{pmatrix} m \\ N' \end{pmatrix};z)^* \Gamma\begin{pmatrix} m \\ N' \end{pmatrix};zu)\, d\gamma(z). \tag{6.3}$$

From (6.3) it is clear that

$$L\begin{pmatrix} N' \\ m \\ M' \end{pmatrix}$$

is a polynomial. In fact, this system of polynomials was introduced by Louck in 1965 [17]; their identification with

matrix elements of the representation is found in Ref. 35, Eq. (5.2.3).

A recently published encyclopedic survey of the theory of group representations (Ref. 39, pp. 155, 6) notes the fact that complete descriptions of the finite-dimensional representations of the simple Lie groups, that is, explicit matrix elements, have remained largely unknown. The further study of the polynomials

$$L\begin{pmatrix} N' \\ m \\ M' \end{pmatrix}$$

is thus a concern of both pure mathematicians and theoretical physicists. The result of Sec. IV of this paper is a modest contribution to determining those polynomials; it raises the hope that a simple general description is possible.

Beyond the groups $U(n)$, one can look to the other compact Lie groups. The Borel–Weil theorem applies to all of them; thus in principle the method of this paper could be generalized. The technical difficulties would be considerable.

[1]H. Weyl, "The structure and representations of continuous groups," Institute of Advanced Studies, Princeton, 1935 (unpublished lecture notes).
[2]I.M. Gel'fand and M.L. Zeitlin, Dokl. Akad. Nauk **71**, 825 (1950).
[3]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1949).
[4]I.M. Gel'fand and M.I. Graev, Izv. Akad. Nauk. SSSR Ser. Mat. **29**, 1329 (1965) [Amer. Math. Soc. Transl. Ser. 2 **64**, 116 (1967)].
[5]G.E. Baird and L.C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[6]M. Ciftan and L.C. Biedenharn, J. Math. Phys. **10**, 221 (1969).
[7]C.-S. Huang and A.C.T. Wu, J. Math. Phys. **14**, 263 (1973).
[8]V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960); **23**, 177 (1961).
[9]M. Moshinsky, Nucl. Phys. **31**, 384 (1962).
[10]M. Ciftan, J. Math. Phys. **14**, 1176 (1973).
[11]J.D. Louck, J. Math. Phys. **6**, 1786 (1965).
[12]J.G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).
[13]J.G. Nagel and M. Moshinsky, Rev. Mex. Phys. **14**, 29 (1965).
[14]A.C.T. Wu, J. Math. Phys. **12**, 437 (1971).
[15]M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[16]L.C. Biedenharn, J. Math. Phys. **8**, 691 (1967).
[17]J.D. Louck and L.C. Bidenharn, J. Math. Phys. **11**, 2368 (1970).
[18]V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962) (reprinted in Ref. 22).
[19]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[20]H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, 1946).
[21]P. Jordan, Z. Phys. **94**, 531 (1935).
[22]L.C. Biedenharn and H. Van Dam, editors, *Quantum Theory of Angular Momentum* (Academic, New York, 1965).
[23]I.M. Gel'fand and M.A. Naimark, *Unitäre Darstellung der Klassischen Gruppen* (Akademie Verlag, Berlin, 1957).
[24]D.P. Zhelobenko, Russ. Math. Surv. **17**, 1 (1962).
[25]G.W. Mackey, Bull. Am. Math. Soc. **69**, 628 (1963).
[26]R. Godement, Trans. Am. Math. Soc. **73**, 496 (1952).
[27]A. Borel and A. Weil, "Representations linearies et espaces homogenes Kählériens des groupes de Lie compacts," Séminaire Bourbaki, May 1954 (exposé by J.-P. Serre).
[28]R. Bott, Ann. Math. **66**, 203 (1957).
[29]N.R. Wallach, Trans. Am. Math. Soc. **136**, 181 (1969).
[30]N.R. Wallach, *Harmonic Analysis on Homogenous Spaces* (Dekker, New York, 1973).
[31]M.A. Lohe and C.A. Hurst, J. Math. Phys. **14**, 1955 (1973).
[32]R. Weitzenbock, *Invariantentheorie* (Noordhoff, Groningen, 1923).
[33]J.D. Louck and L.C. Biedenharn, J. Math. Phys. **14**, 1336 (1973).
[34]F.D. Murnaghan, *The Unitary and Rotation Groups* (Spartan, Washington, D.C., 1962).
[35]J.D. Louck, Am. J. Phys. **38**, 3 (1970).
[36]G. Kowalewski, *Einführung in die Determinantentheorie einschliesslich der Freholmschen Determinanten* (de Gruyter, Berlin, 1954).
[37]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Nat. Bur. Stand., Washington, D.C., 1964).
[38]E. Hewitt and K.A. Ross, *Abstract Harmonic Analysis* (Springer, New York, 1970), Vol. II.

On finite mass renormalizations in the two-dimensional Yukawa model*

Erhard Seiler†

Institute for Advanced Study, Princeton, New Jersey 08540

Barry Simon‡

Departments of Mathematics and Physics, Princeton University, Princeton, New Jersey 08540

(Received 16 May 1975)

In the Mathews–Salam formulas for the (space–time cutoff) Schwinger functions of Y_2, no restriction on finite mass renormalizations for the boson is necessary.

1. INTRODUCTION

In an earlier paper[1] it was shown that the Matthews–Salam formulas[2] can be used to construct the Schwinger functions for Y_2, at least in the presence of a space–time cutoff. Unfortunately, this could be shown only under a certain restriction on the finite renormalization of the boson mass: It had to be essentially nonnegative. But in Glimm[3,4] and Glimm and Jaffe[5] it is shown that for the semiboundedness of the Y_2 Hamiltonian such a restriction is unnecessary. This is achieved by separating out contributions coming from low fermion momenta and estimating them in a way different from the estimate for the high momentum contribution. Since only the latter need renormalization the counterterms can be made smaller by choosing the "lower momentum cutoff" high enough. In this paper we carry through the same idea in the Euclidean framework invented by Matthews and Salam.[2]

2. IMPROVED INTEGRABILITY ESTIMATES

We use the notation of Ref. 1. In particular, we write for $A\in C_{n+1}$ [that is, $\mathrm{Tr}(A^*A)^{(n+1)/2}<\infty$]

$$\det{}_{(n)}(1+A)=\det\left[(1+A)\exp\left(\sum_{k=1}^{n}\frac{(-1)^k}{k}A^k\right)\right]. \tag{2.1}$$

If A is a (bounded) linear operator on a Hilbert space $\mathcal{H}$, we denote by $\Lambda^m(A)$ the operator induced by A on $\Lambda^m(\mathcal{H})$, the m-fold antisymmetric tensor product of $\mathcal{H}$.

Our main result is

Theorem 2.1:

$$u=\|\Lambda^m\left(\frac{1}{1+\lambda K}\right)\det{}^{(0)}_{\mathrm{ren}}(1+\lambda K)\|_{\mathcal{H}},$$

$$\times\exp\left(M^2\int:\phi^2(x):g^2(x)d^2x\right)\in\bigcap_{1\le p<\infty}L^p(d\mu_0)$$

for all $M^2\in\mathbb{R}$, $\lambda\in\mathbb{R}$ ($\det^{(0)}_{\mathrm{ren}}$ denotes the renormalized determinant defined in Ref. 1 with the finite mass renormalization parameter M^2 appearing there put equal to 0).

From Theorem 2.1 we get immediately

Corollary 2.2: The finite volume Schwinger functions exist and fulfill

$$|S_g^{(M)}(h_1,\ldots,h_n;f_1,\ldots,f_m;g_1,\ldots,g_m)|$$

$$\le c_1\exp(nc_1+mc_2)\,\Gamma(n+\tfrac{1}{2})^{1/2}\prod_{r=1}^{n}\|h_r\|_{-1}\prod_{i=1}^{m}\|f_i\|_{\mathcal{H}^*}\|g_i\|_{\mathcal{H}^*} \tag{2.2}$$

where the constants c_1, c_2, c_3 may depend on g and M.

Proof of the Corollary: By the Matthews–Salam formula [cf. (4.15) of Ref. 1] the left-hand side of (2.1) is

$$\left|\int\left(\frac{1}{\sqrt{p^2+m^2}}f_1\wedge\cdots\wedge\frac{1}{\sqrt{p^2+m^2}}f_m,\ \Lambda^m\left(\frac{1}{1+\lambda K}\right)\right.\right.$$

$$S_Fg_1\wedge\cdots\wedge S_Fg_m\Big)_{\Lambda^m(\mathcal{H})}\prod_{r=1}^{n}\phi(h_r)$$

$$\left.\det{}^{(0)}_{\mathrm{ren}}(1+\lambda K)\exp[M^2\int:\phi^2:(x)g^2(x)\,d^2x]\,d\mu_0(\phi)\right|$$

$$\le\int u\left|\prod_{r=1}^{n}\phi(h_r)\right|d\mu_0\prod_{i=1}^{m}\|f_i\|_{\mathcal{H}^*}\|g_i\|_{\mathcal{H}^*}. \tag{2.3}$$

To prove the theorem we have to split the operator K into two parts:

$$K=S_F\Gamma\phi g=L_\zeta+H_\zeta \tag{2.4}$$

where

$$L_\zeta=S_{F,\zeta}\Gamma\phi g, \tag{2.5}$$

$$S_{F,\zeta}=\frac{1}{(2\pi)^2}\int_{|p|\le\zeta}\frac{\not p+m}{p^2+m^2}\exp(ipx)\,d^2p. \tag{2.6}$$

The crucial estimate to separate the contributions from L_ζ and H_ζ is contained in the following:

Lemma 2.3: Let $L\in C_1$, $H\in C_3$. Then

$$\left\|\Lambda^m\frac{1}{1+L+H}\det\{(1+L+H)\exp[-(L+H)+H^2/2]\}\right\|^2$$

$$\le\det{}_{(2)}(1+O_{H_+})\exp[8\|L\|_1-\tfrac{1}{2}\mathrm{Tr}(H^*H)^2-2\mathrm{Re}\,\mathrm{Tr}H^2H^*]$$

$$\times\exp(3m/2), \tag{2.7}$$

where $O_H=H+H^*+H^*H$ and O_{H_+} its nonnegative part. The proof is given in Sec. 3.

Corollary 2.4:

$$u\le\det{}_{(2)}(1+O_{H_\zeta^+})^{1/2}$$

$$\times\exp\left[-\frac{\lambda^4}{4}\mathrm{Tr}(H_\zeta^*H_\zeta)^2-\lambda^3\,\mathrm{Re}\,\mathrm{Tr}H_\zeta^*\exp\left(-\frac{\lambda^2}{2}\mathrm{Tr}_{\mathrm{reg}}:H_\zeta^2:\right)\right]$$

$$\times \exp\left(4\lambda\|L_\zeta\|_1 + \frac{\lambda^2}{2}\langle \mathrm{Tr}\, L_\zeta^2\rangle \exp[M^2 \int :\phi^2(x): g^2(x)\, d^2x]\right). \tag{2.8}$$

where $\mathrm{Tr}_{\mathrm{reg}} :H_\zeta^2:$ is defined as in Ref. 1, that is, with subtraction of the full formal counterterm $\int d^2p/(p^2+m^2)$; hence

$$\mathrm{Tr}_{\mathrm{reg}} :K^2: = \mathrm{Tr}_{\mathrm{reg}} :H_\zeta^2: + 2\mathrm{Tr}:H_\zeta L_\zeta: + \mathrm{Tr}:L_\zeta^2:. \tag{2.9}$$

Next we deal with the low momentum part:

Lemma 2.5:

$$\exp(4\lambda\|L_\zeta\|_1) \in \bigcap_{1\le p<\infty} L^p(d\mu_0).$$

Proof: (a) $L_\zeta \in \mathcal{C}_1$ because it can be factored into two Hilbert–Schmidt operator. Let θ_ζ be the projection onto momenta $|p| \le \zeta$; χ a function which is 1 on supp g and fulfills

$$\int |p|^3 |\tilde{\chi}(p)|^2 d^2p < \infty, \tag{2.10}$$

$$A = \frac{1}{p+m}\,\theta_\zeta \chi(p^2+m^2), \tag{2.11}$$

$$B = \frac{1}{p^2+m^2}\,\phi_g. \tag{2.12}$$

Then $L_\zeta = A \cdot B$,

$$A^*A = \sqrt{p^2+m^2}\,\chi \frac{\theta_\zeta}{(p^2+m^2)^{1/2}}\,\chi(p^2+m^2), \tag{2.13}$$

$$B^*B = \frac{1}{(p^2+m^2)^{1/2}}\,\phi_g \frac{1}{(p^2+m^2)^{3/2}}\,\phi_g \tag{2.14}$$

(recall that we are working in $\mathcal{H} = \mathcal{H}_{1/2} \oplus \mathcal{H}_{1/2}$).

Then it is easy to see that

$$\mathrm{Tr}A^*A < K < \infty \quad (K \text{ independent of } \phi), \tag{2.15}$$

$$\mathrm{Tr}B^*B = (\phi, C\phi)_1, \tag{2.16}$$

where C is trace class in $\mathcal{H}_1$ (actually

$$C = \frac{1}{k^2+\mu^2}\, gEg\Big), \tag{2.17}$$

where E is a multiplication operator in momentum space:

$$E(k) = \int d^2p \frac{1}{[(p+k/2)^2+m^2]^{1/2}[(p-k/2)^2+m^2]^{3/2}}. \tag{2.18}$$

By using the numerical inequality

$$x \le \tfrac{1}{2}(1/\delta + \delta x^2) \quad (x \in \mathbf{R},\ \delta > 0), \tag{2.19}$$

we obtain

$$\exp(4\lambda\|L_\zeta\|_1) \le \exp(4\lambda\|A\|_2\|B\|_2) \le \exp[2\lambda\|A\|_2(1/\delta + \delta(\phi, C\phi)_1)]. \tag{2.20}$$

The right-hand side is in L^p for small enough δ.

The high momentum part is estimated essentially as in Ref. 1. But there is a modification because our estimate (2.7) involves $\det_{(2)}(1+O_{H_+})$ instead of $\det_{(3)}(1+O_{H_+})$. If we estimate

$$w_\zeta(\phi) = \det_{(2)}(1+O_{H_+})^{1/2} \exp\left(-\frac{\lambda^4}{4}\mathrm{Tr}(H_\zeta H_\zeta)^2\right) \times \exp(\lambda^3\, \mathrm{Re}\, \mathrm{Tr}H_\zeta^2 H_\zeta^*) \tag{2.21}$$

by introducing a cutoff κ in ϕ and consider

$$\ln w_\zeta(\phi) - \ln w_\zeta(\phi_\kappa), \tag{2.22}$$

we get, using Lemmas 3.1, 3.2 of Ref. 1,

$$\begin{aligned}\ln w_\zeta(\phi) - \ln w_\zeta(\phi_\kappa) &\le \frac{\lambda^2}{4}[\mathrm{Tr}(H_\zeta^*(\phi)H_\zeta(\phi))^2 - \mathrm{Tr}(H_\zeta^*(\phi_\kappa)H_\zeta(\phi_\kappa))^2] \\ &+ \lambda^3[\mathrm{Tr}H_\zeta(\phi)^2 H_\zeta^*(\phi) - \mathrm{Tr}H_\zeta(\phi_\kappa)^2 H_\zeta^*(\phi_\kappa)] \\ &+ \|O_{H_\zeta(\phi)} - O_{H_\zeta(\phi_\kappa)}\|_4 \sum_{k=0}^{3} c_k \|H_\zeta(\phi)\|_4^k \|H_\zeta(\phi_\kappa)\|_4^{3-k} \\ &+ \tfrac{1}{3}\,\mathrm{Tr}(O^3_{H_\zeta(\phi)-} - O^3_{H_\zeta(\phi_\kappa)-});\end{aligned} \tag{2.23}$$

The only new term is the last one. We have to show that

$$\int |\mathrm{Tr}(O^3_{H_\zeta(\phi)-} - O^3_{H_\zeta(\phi_\kappa)-})|^2\, d\mu_0 = O(\kappa^{-\epsilon}). \tag{2.24}$$

This follows from the following two lemmas.

Lemma 2.6: Let $A, B \in \mathcal{C}_3$ be self-adjoint. Then

$$|\mathrm{Tr}(A_+^3 - B_+^3)| \le \|A-B\|_4(\|A\|_{8/3}^2 + \|A\|_{8/3}\|B\|_{8/3} + \|B\|_{8/3}^2).$$

Proof: Denote by λ_i, μ_i the eigenvalues of A_+, B_+, respectively (ordered decreasingly):

$$\begin{aligned}|\mathrm{Tr}(A_+^3 - B_+^3)| &= |\textstyle\sum(\lambda_i - \mu_i)(\lambda_i^2 + \lambda_i\mu_i + \mu_i^2)| \le (\sum|\lambda_i - \mu_i|^4)^{1/4} \\ &\times (\textstyle\sum(\lambda_k^2 + \lambda_k\mu_k + \mu_k^2)^{4/3})^{3/4} \\ &\le \|A-B\|_4(\|A\|_{8/3}^2 + \|A\|_{8/3}\|B\|_{8/3} + \|B\|_{8/3}^2)\end{aligned}$$

by Lemma 3.2 of Ref. 1.

We next require a general interpolation theorem for the spaces $\mathcal{C}_p$; explicitly, the following three-lines theorem:

Proposition: Let K_z be an analytic operator-valued function in the strip $S = \{z \in \mathbb{C} \mid a < \mathrm{Re}z < b\}$, weakly continuous on the closure of S with $(\varphi, K_z\psi)$ bounded for a dense set of φ and ψ. Suppose that $K_{a+iy} \in \mathcal{C}_{p_0}$ for all real y, $K_{b+iy} \in \mathcal{C}_{p_1}$ for all real y with $\alpha = \sup_y \|K_{a+iy}\|_{p_0} < \infty$ and $\beta = \sup_y \|K_{b+iy}\|_{p_1} < \infty$. Then for any $z \in S$, $K_z \in \mathcal{C}_{p_t}$ with $\ln\|K_z\|_{p_t} \le t\ln\beta + (1-t)\ln\alpha$ where $t = (\mathrm{Re}z - a)b - a$ and $p_t^{-1} = tp_1^{-1} + (1-t)p_0^{-1}$.

Interpolation theorems fall into three closely related types: three line lemmas, Riesz–Thorin theorems, and Stein theorems. Kunze[6] proved a general Riesz–Thorin theorem for $\mathcal{C}_p$ spaces and Calderon[7] made a general analysis of interpolation spaces. By combining these works, one gets the Proposition above (see, e.g., Reed and Simon[8], Appendix to Sec. IX.4). The proposition has been independently discovered by Gohberg, Krein, and Krein and the reader can find a self-contained proof on pp. 137-139 of Ref. 9.

Lemma 2.7: For every $\epsilon > 0$, K, H_ζ, O_k, O_{H_ζ} are in $C_{2+\epsilon}$ a.e.

Proof: We give the proof for K; for the other cases, the proof is analogous. We apply the proposition with

$$z \mapsto K_z = \frac{p+m}{(p^2+m^2)^z}\phi_g \tag{2.25}$$

on the strip

$$S = \{z \in \mathbb{C} : \tfrac{3}{4} + \delta \le \mathrm{Re}z \le 1 + \delta\} \quad (0 < \delta < \tfrac{1}{4}). \tag{2.26}$$

Since it is easy to see that $K_z \in C_4$ a.e. for $\mathrm{Re}z = \frac{3}{4} + \delta$ and $K_z \in C_2$ a.e. for $\mathrm{Re}z = 1 + \delta$ it follows that $K = K_1 \in C_{2/(1-2\delta)}$ a.e. and

$$\ln\|K\|_{2/(1-2\delta)} \le 4\delta \ln\|K_{3/4+\delta}\|_4 + (1 - 4\delta)\ln\|K_{1+\delta}\|_2. \tag{2.27}$$

Lemma 2.6 and 2.7 allow to estimate the non-Gaussian part of $w_\zeta(\phi)$ in the same way as this is done in Ref. 1 (essentially Nelson's argument). The Gaussian parts coming from H_ζ are as in Ref. 1

$$u_\zeta = \exp\left(\frac{\lambda^2}{2}\mathrm{Tr}_{reg} : H_\zeta^* H_\zeta :\right).$$

Lemma 2.8:

$$u_\zeta = \exp\left(\frac{\lambda^2}{2}\mathrm{Tr}_{reg} : H_\zeta^* H_\zeta :\right) = \exp\left[-(\lambda^2/2)(\phi, B_\zeta \phi)_1\right] \tag{2.28}$$

where B_ζ is a positive Hilbert–Schmidt operator on H_{+1} and

$$B_\zeta \ge \frac{1}{k^2+\mu^2}\pi \ln(1 + \zeta^2/m^2)g^2 \tag{2.29}$$

Proof:

$$H_\zeta = (S_F - S_{F,\zeta})\phi_g, \tag{2.30}$$

$$B_\zeta = \frac{1}{k^2+\mu^2} g G_\zeta g, \tag{2.31}$$

$$G_\zeta = -\int d^2p \left(\frac{\theta(p_+^2 - \zeta^2)}{(p_+^2+m^2)^{1/2}} \frac{1}{(p_-^2+m^2)^{1/2}} - \frac{1}{p^2+m^2}\right) \tag{2.32}$$

($p_\pm = p \pm k/2$). Using $ab \le \frac{1}{2}(a^2 + b^2)$ we get

$$G_\zeta \ge \tfrac{1}{2}\int d^2p \left(\frac{1}{p^2+m^2} - \frac{\theta(p_+^2-\zeta^2)}{p_+^2+m^2}\right) + \tfrac{1}{2}\int d^2p \left(\frac{1}{p^2+m^2} - \frac{1}{p_-^2+m^2}\right) = \tfrac{1}{2}\int_{|p|\le\zeta} d^2p \frac{1}{p^2+m^2} = \pi \ln(1+\zeta^2/m^2). \tag{2.33}$$

(The last equality follows because

$$\int d^2p \left(\frac{1}{p^2+m^2} - \frac{1}{p_-^2+m^2}\right) = \int d^2p \left(\frac{1}{(p+k/4)^2+m^2} - \frac{1}{(p-k/4)^2+m^2}\right)$$

$= 0$ by symmetry.)

Now it is clear that we only have to choose ζ such that

$$\pi\lambda^2 \ln(1 + \zeta^2/m^2) \ge M^2 \tag{2.34}$$

to make $u_\zeta \cdot \exp(M^2 \int : \phi^2 : g^2\,dx) \in \cap_{1 \le p < \infty} L^p$. This completes the proof of Theorem 2.1.

3. DETERMINANT INEQUALITIES

In this section we make use of ideas of Lieb[10] and Kato (private communication) in order to prove the crucial lemma 2.3.

Lemma 3.1: Let A, B be linear operators from a M-dimensional (real or complex) Hilbert space H_M to N-dimensional Hilbert space H_N. Then

$$\|\Lambda^m(1/A^*B)\det(A^*B)\|^2 \le \|\Lambda^m(1/A^*A)\det(A^*A)\|\,\|\Lambda^m(1/B^*B)\det(B^*B)\|. \tag{3.1}$$

Remark: Note that $\Lambda^m(A^{-1})\det A$ is a polynomial in the matrix elements of A, so (3.1) makes sense (and is true) also for singular A or B.

Proof: We use the polar decompositions

$$A = U|A|, \quad B = V|B| \text{ where } |A| = (A^*A)^{1/2}, \quad B = (B^*B)^{1/2};$$

U, V are partial isometries. Then $A^*B = |A|C|B|$ where $C = U^*V$ is a contraction in H_M. The left-hand side of (3.1) is then bounded by

$$\|\Lambda^m(1/|A|)\det|A|\|^2 \|\Lambda^m(1/|B|)\det|B|\|^2 \times \|\Lambda^m(1/C)\det C\|^2 \tag{3.2}$$

and the last factor is ≤ 1 as can be seen by replacing C by $|C|$ (unitaries do not matter); then $\|\Lambda^m(1/|C|) \times \det|C|\|$ is the product of the m largest eigenvalues of $|C|$. The first two factors in (3.2) give the right-hand side of (3.1).

Lemma 3.2: Let A_i, B_i $(i = 1, \ldots, n)$ be linear operators from H_M to H_M. Then

$$\|\Lambda^m(1/\textstyle\sum_i A_i^* B_i)\det\sum_k A_k^* B_k\|^2 \le \|\Lambda^m(1/\sum_i A_i^* A_i)\det\sum_k A_k^* A_k\|\,\|\Lambda^m(1/\textstyle\sum_i B_i^* B_i) \times \det\sum_k B_k^* B_k\|. \tag{3.3}$$

Proof: This is a special case of Lemma 3.1 ($H_N = \oplus_{k=1}^n H_M$).

Remark: Lemma 3.2 has been brought to our attention by Lieb who first proved it for the special case $m = 0$ and then proved a general result[10] from which (3.3) follows easily. The idea of the proof given here (the reduction to Lemma 3.1) is due to Kato (private communication).

Lemma 3.3: Let A, B be trace class operators on a Hilbert space H. Then

$$\left\|\Lambda^m\left(\frac{1}{1+A+B}\right)\det(1+A+B)\right\|^2 \le \left\|\Lambda^m\left(\frac{1}{|1+A|+|B|}\right)\det(|1+A|+|B|)\right\| \times \left\|\Lambda^m\left(\frac{1}{|1+A|+W|B|W^{-1}}\right)\det(|1+A|+W|B|W^{-1})\right\| \tag{3.4}$$

with a unitary W. (For $m = 1$ this result also can be found in Lieb.[10])

Proof: If $\mathcal{H}$ has finite dimension, this is a special case of Lemma 3.2 with $A_1 = B_1 = 1$, $A_2 = U|A|^{1/2}$, $B_2 = |A|^{1/2}$, $A_3 = V|B|^{1/2}$, $B_3 = |B|^{1/2}$ ($A = U|A|$, $B = V|B|$, $W = U^*V$). The infinite-dimensional case follows from an easy limiting argument (see Appendix, Proposition 2).

Lemma 3.4: If $0 \leq A \leq 1$ is a $\mathcal{C}_{n+1}$ operator on $\mathcal{H}$ (i.e., $\mathrm{Tr}|A|^{n+1} < \infty$), then $\|\Lambda^m[1/(1-A)]\det_{(n)}(1-A)\| \leq \exp[m(1+\frac{1}{2}+\cdots+1/n)]$. (3.5)

Proof: If $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$ are the eigenvalues of A, the left-hand side is

$$\prod_{i\geq 1}(1-\alpha_i)\exp(\alpha_i+\alpha_i/2+\cdots+\alpha_i^n/n)\prod_{i\leq m}\frac{1}{1-\alpha_i}$$

$$\leq \prod_{i\leq m}\exp(\alpha_i+\alpha_i^2/2+\cdots+\alpha_i^n/n)$$

$$\leq \exp[m(1+1/2+\cdots+1/n)].$$

Lemma 3.5 (= Lemma 2.3): For $A\in\mathcal{C}_3$, $B\in\mathcal{C}_1$

$$\left\|\Lambda^m\left(\frac{1}{1+A+B}\right)\det((1+A+B)\exp(-A+A^2/2))\right\|^2 \leq \det{}_{(2)}(1+O_{A+})\exp(6\|B\|_1+\tfrac{3}{2}m-\tfrac{1}{2}\|A\|_4^4-2\mathrm{Re}\,\mathrm{Tr}A^2A^*) \quad (3.6)$$

where $O_{A+}=(A+A^*+A^*A)_+$.

Proof: It is sufficient to give the proof for $A, B\in\mathcal{C}_1$ (see Appendix, Proposition 4). By Lemma 3.3 it suffices to estimate

$$X=\|\Lambda^m\left(\frac{1}{1+C+|B|}\right)\det(1+C+|B|)\| \quad (3.7)$$

where $C=|1+A|-1=C_+-C_-$ ($\Rightarrow 0\leq C_-\leq 1$). With

$$D=\frac{1}{\sqrt{1+C_+}}|B|\frac{1}{\sqrt{1+C_+}},\quad E=\frac{1}{\sqrt{1+D}}C_-\frac{1}{\sqrt{1+D}}$$

we have (since $C_+C_-=C_-C_+=0$)

$$X\leq\det(1+C_+)\|\Lambda^m\left(\frac{1}{1+D-C_-}\right)\det(1+D-C_-)\|$$

$$\leq\det(1+C_+)\det(1+D),$$

$$\|\Lambda^m\left(\frac{1}{1-E}\right)\det(1-E)\|\leq\det(1+C_+)\det(1+D)\times\exp(\tfrac{3}{2}m)\exp(-\mathrm{Tr}E-\mathrm{Tr}E^2/2) \quad (3.8)$$

by Lemma 3.4.

Now

$$\mathrm{Tr}E=\mathrm{Tr}(1+D)^{-1}C_-\geq\mathrm{Tr}(1-D)C_-\geq\mathrm{Tr}C_--\mathrm{Tr}D\|C_-\|\geq\mathrm{Tr}C_--\mathrm{Tr}|B|, \quad (3.9)$$

$$\mathrm{Tr}E^2=\mathrm{Tr}\left[\left(1-\frac{D}{1+D}\right)C_-\right]^2=\mathrm{Tr}C_-^2-2\,\mathrm{Tr}\frac{D}{1+D}C_-^2+\mathrm{Tr}\left(\frac{D}{1+D}C_-\right)^2\geq\mathrm{Tr}C_-^2-2\,\mathrm{Tr}D\geq\mathrm{Tr}C_-^2-2\,\mathrm{Tr}|B|. \quad (3.10)$$

Because of $1+O_{A+}=|1+A|^2P_++(1-P_+)=(1+C)^2P_+ +(1-P_+)=(1+C_+)^2$ (where P_+ is the projection associated with C_+) we get

$$X\leq\det(1+O_{A+})^{1/2}\det(1+|B|)\exp(2\,\mathrm{Tr}|B|+3m/2)\times\exp(-\mathrm{Tr}C_--\mathrm{Tr}C_-^2/2)=\det{}_{(2)}(1+O_{A+})^{1/2}\det(1+|B|)\times\exp(2\,\mathrm{Tr}|B|+3m/2)\times\exp(-\mathrm{Tr}C_--\tfrac{1}{2}\mathrm{Tr}C_-^2+\tfrac{1}{2}\mathrm{Tr}O_{A+}-\tfrac{1}{4}\mathrm{Tr}O_{A+}^2) \quad (3.11)$$

We claim

$$\mathrm{Tr}C_-+\tfrac{1}{2}\mathrm{Tr}C_-^2\geq\tfrac{1}{2}\mathrm{Tr}O_{A-}+\tfrac{1}{4}\mathrm{Tr}O_{A-}^2. \quad (3.12)$$

Proof: Since $(1-C_-)^2=1-O_{A-}$,

$$\mathrm{Tr}(C_-+\tfrac{1}{2}C_-^2)=\mathrm{Tr}(1-\sqrt{1-O_{A-}})+\tfrac{1}{2}\mathrm{Tr}(1-\sqrt{1-O_{A-}})^2=2\mathrm{Tr}(1-\sqrt{1-O_{A-}})-\tfrac{1}{2}\mathrm{Tr}O_{A-}\geq\tfrac{1}{2}\mathrm{Tr}O_{A-}+\tfrac{1}{4}\mathrm{Tr}O_{A-}^2$$

($\sqrt{1-x}\leq 1-x/2-x^2/8$ for $0\leq x\leq 1$).

Therefore,

$$X\leq\det{}_{(2)}(1+O_{A+})^{1/2}\exp(3\|B\|_1+\tfrac{3}{2}m)\times(\tfrac{1}{2}\mathrm{Tr}O_A-\tfrac{1}{4}\mathrm{Tr}O_A^2). \quad (3.13)$$

The lemma now follows from

$$\tfrac{1}{2}\mathrm{Tr}O_A-\tfrac{1}{4}\mathrm{Tr}O_A^2-\mathrm{Re}\,\mathrm{Tr}A+\tfrac{1}{2}\mathrm{Re}\,\mathrm{Tr}A^2=-\tfrac{1}{4}\|A\|_4^4-\mathrm{Re}\,\mathrm{Tr}A^2A^*. \quad (3.14)$$

ACKNOWLEDGMENT

We are grateful to E. Lieb for making the results of Ref. 10 available to us before publication and to T. Kato for communicating to us the idea for the simple proof of Lemma 3.2. One of us (E.S.) wishes to thank the Institute for Advanced Study for its hospitality.

Remark: When this work was essentially completed we received a paper by O.A. McBryan[11] on the same subject which also proves Theorem 2.1; the method used there, an expansion of the low momentum interaction, is quite different.

APPENDIX: CONTINUITY OF CERTAIN OPERATOR FUNCTIONS

Proposition 1: For $A\in\mathcal{C}_1$

$$\|\Lambda^m\left(\frac{1}{1+A}\right)\det(1+A)\|\leq\exp(\|A\|_1+m). \quad (A1)$$

(This estimate is not best possible: The factor e^m can be eliminated.)

Proof:

$$\|\Lambda^m\left(\frac{1}{1+A}\right)\det(1+A)\|=\sup_{\{e_i,f_k\}}|(e_1\wedge\cdots\wedge e_m,\Lambda^m\frac{1}{1+A}f_1\wedge\cdots\wedge f_m)\times\det(1+A)| \quad (A2)$$

where the sup is over orthonormal systems of vectors e_i, f_i ($i=1,\ldots,m$). If we denote by C_k the operator which maps $u\in\mathcal{H}$ into $e_k(f_k,u)$ ($k=1,\ldots,m$), then

$$e_1 \wedge \cdots \wedge e_m, \Lambda^m\left(\frac{1}{1+A}\right) f_1 \wedge \cdots \wedge f_m \ \det(1+A)$$
$$= \frac{\partial^m}{\partial\lambda_1 \cdots \partial\lambda_m} \det\left(1 + A + \sum_{k=1}^{m} \lambda_k C_k\right)$$
$$= \left(\frac{1}{2\pi i}\right)^m \oint_{|\lambda_1|=1} \left\{\frac{d\lambda_1}{\lambda_1^2} \cdots \oint_{|\lambda_m|=1} \left\{\frac{d\lambda_m}{\lambda_m^2} \det\left(1 + A + \sum_{k=1}^{m} \lambda_k C_k\right)\right. \right. \quad \text{(A3)}$$

Using the well-known inequality

$$|\det(1+B)| \leq \exp(\|B\|_1) \quad \text{(A4)}$$

we obtain(A1).

Proposition 2: For $A, B \in C_1$

$$\|\Lambda^m\left(\frac{1}{1+A}\right) \det(1+A) - \Lambda^m\left(\frac{1}{1+B}\right) \det(1+B)\|$$
$$\leq \|A-B\|_1 \exp(\|A\|_1 + \|B\|_1 + m + 1). \quad \text{(A5)}$$

Proof: Without loss of generality we can assume $A \neq B$. Consider then

$$F(t) = \Lambda^m\left(\frac{1}{1+A+t(B-A)}\right) \det(1+A+t(B-A)). \quad \text{(A6)}$$

The left-hand side of (A5) is bounded by

$$\int_0^1 \|F'(t)\| \, dt \leq \sup_{t\in[0,1]} \|F'(t)\|. \quad \text{(A7)}$$

By Cauchy's formula we have

$$\|F'(t)\| = \|\frac{1}{2\pi i} \oint_{|\tau|=\xi} \frac{dt}{\tau^2} F(t+\tau)\| \leq \frac{1}{2\pi} \oint_{|\tau|=\xi} \left|\frac{dt}{\tau^2}\right| \|F(t+\tau)\|. \quad \text{(A8)}$$

Choosing $|\tau| = \xi = (\|A-B\|_1)^{-1}$ and using proposition 1, we obtain

$$\|F(t+\tau)\| \leq \exp[\|(1-t)A + tB + \tau(B-A)\|_1 + m]$$
$$\leq \exp(\|A\|_1 + \|B\|_1 + m + 1) \quad \text{(A9)}$$

and, therefore,

$$\|F'(t)\| \leq \|A-B\|_1 \exp(\|A\|_1 + \|B\|_1 + m + 1) \quad \text{(A10)}$$

which proves the assertion by (A6).

Proposition 3: The function

$$R_n : C_n \to C_1$$
$$A \mapsto R_n(A) = (1+A) \exp\left(-\sum_{k=1}^{n-1} \frac{(-A)^k}{k}\right) - 1 \quad \text{(A11)}$$

is continuous.

Proof: $R_n(A) = A^n G(A)$ where G is an entire function. Therefore

$$\|R_n(A) - R_n(B)\|_1 \leq \|A^n - B^n\|_1 \|G(A)\| + \|B^n\|_1 \|G(A) - G(B)\|. \quad \text{(A12)}$$

Repeated use of Hölder's inequality for operators gives

$$\|A^n - B^n\|_1 \leq \|A-B\|_n \sum_{k=0}^{n-1} \|A\|_n^k \|B\|_n^{n-1-k}. \quad \text{(A13)}$$

Proposition 4: The function

$$L_n^m : C_n \to \mathcal{L}(\Lambda^m \mathcal{H}) \text{ (bounded operators on } \Lambda^m \mathcal{H}),$$
$$A \mapsto L_n^m(A) = \Lambda^m\left(\frac{1}{1+A}\right) \det_{(n-1)}(1+A)$$

is continuous.

Proof: $L_n^m(A) = \Lambda^m\left(\exp \sum_{k=1}^{n} \frac{(-A)^k}{k}\right)$,

$$\Lambda^m\left(\frac{1}{1+R_n(A)}\right) \det(1+R_n(A)). \quad \text{(A14)}$$

The first factor is obviously continuous; Proposition 5 then follows from Proposition 3 and Proposition 4.

Remark: For special cases ($m = 0$ or 1) most results of this appendix can be found in Refs. 9 and 12.

*This work was partially supported by USNSF grants GP-40768X and GP-39048.

†On leave from the Max Planck Institute für Physik und Astrophysik, München, Germany.
‡A Sloan Foundation Fellow.

[1] E. Seiler, "Schwinger functions for the Yukawa model in two-dimensions with space-time cutoff," Commun. Math. Phys. **42**, 163 (1975).
[2] P.T. Matthews and A. Salam, Nuovo Cimento **12**, 563 (1954); **2**, 120 (1955).
[3] J. Glimm, Commun. Math. Phys. **5**, 343 (1967).
[4] J. Glimm, Commun. Math. Phys. **6**, 61 (1967).
[5] J. Glimm and A. Jaffe, in *Statistical Mechanics and Quantum Field Theory*, edited by C. de Witt and R. Stora (Gordon and Breach, New York, 1971).
[6] R.A. Kunze, Trans. Am. Math. Soc. **89**, 519 (1958).
[7] A. Calderon, Studia Math. **24**, 113 (1964).
[8] M. Reed and B. Simon, *Methons of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[9] I.C. Gohberg, and M.G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators*, Translations of Mathematical Monographs Vol. 18 (Am. Math. Soc., Providence, R.I., 1969).
[10] E. Lieb, "Inequalities for Some Operator and Matrix Functions," Adv. Math., to appear.
[11] O.A. McBryan, "Finite Mass Renormalizations in the Yukawa$_2$ Theory," Commun. Math. Phys., to appear.
[12] N. Dunford and J. Schwartz, *Linear Operators* (Interscience, New York, 1963), Vol. II.

Average multiplicity and the zeros of multiparticle generating function*

Toichiro Kinoshita†

Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853

Jnanadeva Maharana‡

Department of Physics, University of Tokyo, Tokyo 113, Japan
(Received 19 August 1974; revised manuscript received 6 June 1975)

We have obtained several theorems exhibiting relations between the asymptotic behavior of the average multiplicity $\langle n\rangle$ and the distribution of zeros of a generating function for multiparticle production cross sections.

As is well known there is a close analogy between the multiplicity distribution at high energies of multiparticle production processes and the statistical mechanics of a one-dimensional gas.[1] The analog of the grand partition function in statistical mechanics is the multiparticle generating function[2,3] such as

$$\Omega(z,y)=\sum_{n=2}^{N(s)}\sigma_n(s)z^n, \tag{1}$$

where z is the analog of fugacity, $y=\ln s$ (s being the square of c.m. energy) is the maximum rapidity available in the collision, and $N(s)$ is the largest integer not exceeding $\sqrt{s/m}$ (all particles are assumed to have the same mass m for simplicity). One may then define an analog of thermodynamical quantities such as "pressure" p and "density" ρ:

$$\begin{aligned} p(z)&=\lim_{y\to\infty}\frac{1}{y}\ln\Omega(z,y),\\ \rho(z)&=\lim_{y\to\infty}z\frac{\partial}{\partial z}\frac{1}{y}\ln\Omega(z,y), \end{aligned} \tag{2}$$

y playing the role of volume in statistical mechanics.

When generating functions are constructed based on some specific models, their characteristics might reveal themselves through these thermodynamical quantities. In fact, some dynamical models, such as the dual model and the multiperipheral-type models,[3] have been used recently to study their high energy behavior and the problem of "critical phenomena" in hadron–hadron collisions.[4]

One of the effective approaches to the problem of phase transition is to study the distribution of zeros of the grand partition function.[5] Phase transition takes place if zeros close in onto the real fugacity axis as the volume V tends to ∞. Analogously one may ask whether the knowledge of the distribution and energy dependence of zeros of a multiparticle generating function provides information on the asymptotic behavior of experimentally observed quantities.

In this article we report some results of our study concerning the relation between the asymptotic behavior of the average multiplicity $\langle n\rangle$ and the distribution of zeros around the unit circle $|z|=1$ of a generating function for multiparticle production defined by

$$F_s(z)=\frac{1}{2}\sum_{n=0}^{N(s)}a_n(s)z^n. \tag{3}$$

Here $a_0(s)=1$, $a_1(s)=0$, and

$$a_n(s)=\frac{\sigma_n(s)}{\sigma_{\rm tot}(s)},\quad n\geq 2,\quad \sigma_{\rm tot}(s)=\sum_{n=2}^{N(s)}\sigma_n(s). \tag{4}$$

The quantities $\langle n\rangle$, $\langle n(n-1)\rangle$, etc. are related to $F_s(z)$ by

$$\begin{aligned} \langle n\rangle&=2\frac{d}{dz}F_s(z)\Big|_{z=1},\\ \langle n(n-1)\rangle&=2\frac{d^2}{dz^2}F_s(z)\Big|_{z=1},\ \text{etc.} \end{aligned} \tag{5}$$

Thus they will be influenced by some local properties of $F_s(z)$ near $z=1$, such as the distribution of its zeros. The question is: To what extent and how? In order to avoid unnecessary complications, we shall assume that $\langle n\rangle$, $\langle n(n-1)\rangle$, $\cdots$, are continuous, monotonically increasing functions of s. Most of our results will hold with slight modifications if we weaken this assumption. For instance, $\langle n\rangle$, $\langle n(n-1)\rangle$, $\cdots$, may have some oscillating s dependence.

We have chosen to study $F_s(z)$ instead of $\Omega(z,y)=\sigma_{\rm tot}(s)(2F_s(z)-1)$ because of simpler mathematical property. Of course, the zeros of $F_s(z)$ and $\Omega(z,y)$ do not coincide, although most zeros of $\Omega(z,y)$ will be close to those of $F_s(z)$. [A notable exception is the double zero of $\Omega(z,y)$ at $z=0$.] If we define the "pressure" p' and "density" ρ' for $F_s(z)$ by

$$\begin{aligned} p'(z)&=\lim_{y\to\infty}\frac{1}{y}\ln F_s(z),\\ \rho'(z)&=\lim_{y\to\infty}z\frac{\partial}{\partial z}\frac{1}{y}\ln F_s(z), \end{aligned} \tag{6}$$

we find

$$\rho'(z)\Big|_{z=1}=\tfrac{1}{2}\rho(z)\Big|_{z=1}. \tag{7}$$

We are now ready to examine the properties of $F_s(z)$. By definition $F_s(z)$ is a polynomial and hence is regular for all z. Unitarity (in the weakest sense) requires that

$$0\leq a_n(s)\leq 1. \tag{8}$$

We also have

$$\sum_{n=2}^{N(s)}a_n(s)=1 \tag{9}$$

from (4). Thus $|F_s(z)|$ is uniformly bounded by 1 for $|z|<1$.

If $F_s(z)$ is uniformly bounded in a larger circle $|z|<r_c$, $r_c>1$, $dF_s(z)/dz$ is also bounded for $|z-1|<r_c-1$. Then, in the limit $s\to\infty$, the density $\rho'(z)$ of (6) vanishes there and no "phase transition" occurs at $z=1$. Since we have nothing more to say about it, we shall consider in the following only the case where $\lim_{s\to\infty}F_s(z)$ is singular at $z=1$, or more specifically the case where $\langle n\rangle(=2\,dF_s/dz|_{z=1})$ diverges for $s\to\infty$.

As is readily seen from (8) and (9), $F_s(z)$ is free from zero for $|z|<1$ and is positive on the positive real axis. Thus all zeros $z_i(s)$, $i=1,2,\cdots$, of $F_s(z)$ must satisfy

$$|z_i(s)|\geq 1. \tag{10}$$

If $a_n(s)$ are continuous functions of s, which we may assume without loss of generality, the zeros $z_i(s)$ are also continuous functions of s. [Note that if $z_i(s)$ is a complex zero of $F_s(z)$ so is $z_i^*(s)$. $F_s(z)$ may also have negative real zeros, which we shall not write explicitly.]

For our following considerations it is crucial to have some knowledge of the large s behavior of $F_s(z)$. We know from (3) already that F_s is bounded for large s as

$$|F_s(z)|\leq \tfrac{1}{2}R^{N(s)}$$

for any $|z|<R$, $R>1$. In order to have a closer control of the asymptotic behavior of $F_s(z)$, however, let us introduce $g(s,R)$ by $|F_s(R)|=R^{g(s,R)}$. This function is not necessarily monotonic in s. For any fixed $R>1$, however, we can construct a positive monotonically increasing function $h(s,R)$ by

$$h(s,R)=\max_{s'\leq s} g(s',R),$$

which satisfies

$$h(s,R)\leq N(s). \tag{11}$$

From now on we shall regard R as large ($\gg 1$) but fixed, and denote $h(s,R)$ simply as $h(s)$. Then, because of the positivity (8),

$$|F_s(z)|\leq R^{h(s)},\quad s>s_0\gg 4m^2, \tag{12}$$

holds not only for $z=R$ but also for all $|z|\leq R$.

An immediate consequence of (12) is the inequality

$$a_n(s)\leq 2R^{h(s)-n}, \tag{13}$$

which is a stronger constraint than (8) for $n>h(s)$.

Unlike $a_n(s)$ and $\langle n\rangle$ it is difficult to associate $h(s)$ with directly observable quantities. As is shown in the following, however, $h(s)$ is closely related to the number and distribution of zeros of $F_s(z)$ near $|z|=1$ under certain circumstances.

Let us begin by noting that $h(s)$ gives an upper bound for the number $N_\delta(s)$ of zeros of $F_s(z)$ in $|z|<1+\delta$, $\delta>0$. In fact, applying Jensen's theorem[6] to the circle $|z|<R$, and taking account of (12) and $F_s(0)=1/2$, we obtain

$$N_\delta(s)\leq \frac{1}{\ln(R/1+\delta)}\,[h(s)\ln R+\ln 2].$$

Since $F_s(z)$ has no zeros in $|z|<1$, this also gives an upper bound for the number of zeros in $1\leq|z|<1+\delta$.

It is not so straightforward, however, to find a useful lower bound for $N_\delta(s)$. Let us derive it in several steps, exhibiting additional assumptions clearly where needed.

(i) Suppose $\langle n\rangle\to\infty$ for $s\to\infty$. Then, for any given $\delta>0$, the number of $a_n(s)$ satisfying

$$a_n(s)\geq \frac{a_0(s)}{(1+\delta)^n}\quad [a_0(s)=1] \tag{14}$$

increases without bound as $s\to\infty$. In fact, if $a_n(s)<a_0(1+\delta)^{-n}$ holds for all $n>n_0$, n_0 being independent of s, we find

$$\langle n\rangle=\sum_{n=0}^{n_0} na_n+\sum_{n=n_0+1}^{N} na_n < n_0+\frac{1+n_0\delta}{\delta^2(1+\delta)^{n_0}}<\infty, \tag{15}$$

which contradicts our assumption.

(ii) Let $K\equiv K(s,\delta)$ be the largest n satisfying (14) for given s and δ. Then, for very large s, we obtain

$$\langle n\rangle\leq K\leq \frac{1}{\ln(R/1+\delta)}\,[h(s)\ln R+\ln 2]. \tag{16}$$

The upper bound is trivially satisfied if $K\leq h(s)$. For $K>h(s)$, it follows readily from

$$a_0(1+\delta)^{-K}\leq a_K\leq 2R^{h(s)-K}.$$

To obtain the lower bound, note that

$$a_n(s)<\frac{a_0}{(1+\delta)^n},\quad n\geq K+1.$$

Thus, replacing n_0 in (15) by K, we have

$$\langle n\rangle<K+(1+K\delta)/[\delta^2(1+\delta)^K].$$

Obviously K must increase indefinitely as $\langle n\rangle\to\infty$. In particular, for sufficiently large s [such that $K>(c/\delta)\ln(1/\delta)$, $c>2$, holds], we find $\langle n\rangle\leq K$.

(iii) Next, consider the partial sum

$$S_K(z)=\sum_{n=0}^{K} a_n(s)z^n, \tag{17}$$

whose zeros we denote as $y_1,y_2,\ldots,y_K$. From (14) we get

$$|y_1y_2\cdots y_K|=a_0/a_K\leq(1+\delta)^K.$$

Let $y_{K-p+1},\ldots,y_K$ be the zeros for which $|y_i|>1+\epsilon$, $\epsilon>0$. Then, noting that $|y_i|>1$ for all i, we obtain

$$(1+\epsilon)^p<|y_{K-p+1}\cdots y_K|<|y_1y_2\cdots y_K|\leq(1+\delta)^K,$$

and hence

$$\frac{p}{K}<\frac{\ln(1+\delta)}{\ln(1+\epsilon)}\approx\frac{\delta}{\epsilon}.$$

Thus one can make p/K arbitrarily small by choosing first ϵ and then δ. Recalling that $K=K(s,\delta)$, one can therefore find, for very large s, positive constants δ,η such that the number of zeros of $S_K(z)$ in $|z|<1+\delta$ is greater than $K(1-\eta)$.

(iv) By an adaptation of Jentsch's theorem,[7] it is easy to show that every point of the circle $|z|=1$ has a

neighborhood in which a zero of $S_K(z)$ is found for sufficiently large s. It can be shown further that the distribution of these zeros is nearly unform in $\theta = \arg z$. To be more precise, let $\nu(\beta, \Delta\beta)$ be the number of zeros of $S_K(z)$ in the sector $\beta \leq \arg z \leq \beta + \Delta\beta$. Then we have[8]

$$\left|\nu(\beta, \Delta\beta) - \frac{\Delta\beta}{2\pi} K\right| < 16\left[K \ln\left(\sum_{i=0}^{K} |a_i| / \sqrt{a_0 a_K}\right)\right]^{1/2}. \quad (18)$$

The steps (iii) and (iv) have been introduced in the hope that the distribution of zeros of $F_s(z)$ may be inferred from that of $S_K(z)$. This will be the case if the remainder $R_K(z) \equiv 2F_s(z) - S_K(z)$ satisfies the inequality

$$|R_K(z)| < |S_K(z)| \quad (19)$$

on a small closed contour surrounding a zero of $S_K(z)$. By Rouché's theorem[9] $F_s(z)$ will then have a zero within the same contour. Of course, (19) will not be valid unless $R_K(z)$ is sufficiently constrained. As is shown below, the assumption

$$a_n \leq \frac{a_K}{(1+\alpha)^{n-K}}, \quad n = K+1, K+2, \cdots, \quad (20)$$

which is perhaps the simplest of its kind, does indeed provide such a constraint on $R_K(z)$, where K is defined in the step (ii) and α is an s-independent positive number. Since this is one of our main results, let us formulate it as

Theorem 1: Suppose $\langle n \rangle \to \infty$ for $s \to \infty$. Suppose also that a_n decreases exponentially for very large n according to (20). Then $F_s(z)$ has a large number (at least of order $\langle n \rangle$) of zeros lying in the ring $1 < |z| < 1 + \delta$, where δ can be chosen arbitrarily small as $s \to \infty$. Furthermore, every point, with possibly a few exceptions, of the circle $|z| = 1$ has a neighborhood in which a zero of $F_s(z)$ is found for sufficiently large s.

Proof: Let us assume for simplicity that all zeros of $S_K(z)$ are simple. According to (18), for any fixed $\Delta\beta$ ($\gg \sqrt{\delta}$) such that $2\pi/\Delta\beta$ is an integer, $\nu(\beta, \Delta\beta)$ has a minimum when β goes over a set of discrete values separated by $\Delta\beta$. Let this minimum be $(\Delta\beta/2\pi)(K - \Delta K)$. Let A be the set consisting of the $(\Delta\beta/2\pi)(K - \Delta K)$ zeros of smallest moduli in each sector, and B be the set of remaining zeros. More precisely, let $\nu^A(\beta, \Delta\beta)$ and $\nu^B(\beta, \Delta\beta)$ be the number of zeros of the sets A and B, respectively, in the sector $\beta \leq \arg z \leq \beta + \Delta\beta$. Then

$$\nu^A(\beta, \Delta\beta) = \frac{\Delta\beta}{2\pi}(K - \Delta K) \quad \text{independent of } \beta, \quad (21a)$$

$$\nu^B(\beta, \Delta\beta) = \nu(\beta, \Delta\beta) - \nu^A(\beta, \Delta\beta) = O(\Delta K), \quad (21b)$$

where $\Delta K = \nu^B(\beta, 2\pi)$, the total number of zeros belonging to the set B, is subject to the constraint

$$\Delta K < 16K(\delta/2)^{1/2}$$

according to (18).

Let $y_i = \rho_i \exp(i\theta_i)$ be a zero belonging to the set A. Then, making use of the inequality

$$|a\exp(i\theta) - b\exp(i\phi)|^2 = ab\left(|\exp(i\theta) - \exp(i\phi)|^2 + \frac{(a-b)^2}{ab}\right) \geq ab|\exp(i\theta) - \exp(i\phi)|^2, \quad (a, b > 0),$$

we obtain

$$|S_K'(y_i)| = \frac{1}{|y_i|}\prod_{j\neq i}\left|1 - \frac{y_i}{y_j}\right| \geq \frac{1}{|y_i|}(a_K|y_i|^K)^{1/2}\prod_{\substack{j\neq i \\ j\in A}}|\exp(i\theta_i) - \exp(i\theta_j)|\prod_{k\in B}|\exp(i\theta_i) - \exp(i\theta_k)|, \quad (22)$$

where we may assume that y_i is located in the center of one sector and that the factor $|\exp(i\theta_i) - \exp(i\theta_j)|$ is replaced by the exact expression $(|\exp(i\theta_i) - \exp(i\theta_j)|^2 + (\rho_i - \rho_j)^2/\rho_i\rho_j)^{1/2}$ whenever y_j, $j \in A \cup B$, belongs to the same sector as y_i. From this construction and (21a) we find

$$\prod_{\substack{j\neq i \\ j\in A}}|\exp(i\theta_i) - \exp(i\theta_j)| = O(K). \quad (23a)$$

Note that, if each sector contains one zero in its center, the right-hand side of (23a) is exactly equal to $K - \Delta K$.

It is more difficult to estimate the last factor of (22). If we average it geometrically over all y_i, $i \in A$, however, we obtain

$$\left(\prod_{i\in A}\prod_{k\in B}|\exp(i\theta_i) - \exp(i\theta_k)|\right)^{1/(K-\Delta K)} = O(1).$$

Of course, $\prod_{k\in B}|\exp(i\theta_i) - \exp(i\theta_k)|$ itself can be much smaller than this average for some y_i. However,

$$\prod_{k\in B}|y_i - y_k| = O(1) \quad (23b)$$

will hold for the majority of y_i ($i \in A$), the only exceptions being those that happen to be very close to a cluster of zeros of the set B. If we exclude such zeros, we obtain from (22) and (23) that

$$|S_K'(y_i)| \geq cK|y_i|^{-1}(a_K|y_i|^K)^{1/2}$$

where $c = O(1)$.

Now, since the average separation of zeros is at least $2\pi/K$, one can find a radius r_i [$= O(K^{-1})$] such that $S_K(z)/(z - y_i)$ is zero-free in the circle C_i: $|z - y_i| < r_i$. The integral

$$\int_0^z S_K(y_i + r_i w)\frac{dw}{w}$$

is then univalent in $|z| < 1$ and $S_K(z)$ satisfies the bound[10]

$$|S_K(z)| > \sigma(1-\sigma)(1+\sigma)^{-3} r_i |S_K'(y_i)|$$

on the circle $C_{i,\sigma}$: $|z - y_i| = \sigma r_i$, $0 < \sigma < 1$. On the other hand, we obtain from (20)

$$|R_K(z)| < (1 + \alpha - |z|)^{-1} a_K |z|^{K+1}.$$

Thus, for most y_i satisfying $|y_i| < 1 + \delta$, the inequality (19) holds on $C_{i,\sigma}$ if we choose

$$\sigma r_i < ((1+\delta)|y_i|)^{1/2} - |y_i|.$$

By Rouché's theorem $F_s(z) = \frac{1}{2}[S_K(z) + R_K(z)]$ must therefore have a zero in $C_{i,\sigma}$. QED

Remarks: (1) With minor modifications the above proof applies also to nonsimple zeros. (2) Theorem 1 will not be valid if the assumption (20) is weakened too much. A counterexample is provided by the function[11]

$$\frac{1}{2}+\frac{1}{2}\sum_{n=2}^{N} n^{-\epsilon n} z^n \Big/ \left(\sum_{n=2}^{N} n^{-\epsilon n}\right), \quad \epsilon > 0,$$

which satisfies all requirements for $F_s(z)$ except (20).

Formula (16) (see also Appendix, Lemma 2) shows that $\langle n\rangle$ is bounded from above by $h(s)$. To see the relation of $h(s)$ and $\langle n\rangle$ more closely, let us write $F_s(z)$ as

$$F_s(z)=\prod_i \frac{(z-z_i)(z-z_i^*)}{(1-z_i)(1-z_i^*)},$$

where the product is over all zeros $z_i = z_i(s)$ with $\mathrm{Im} z_i > 0$. Differentiating $\ln F_s(z)$ and setting $z=1$, we obtain

$$\langle n\rangle = 2F_s'(1) = 2\sum_i \frac{2(1-\mathrm{Re} z_i)}{|1-z_i|^2}. \tag{24}$$

According to Theorem 1 the total number of zeros of $F_s(z)$ in the ring $1 < |z| < 1+\delta$ is greater than $K(1-\eta')$ where $\eta'(>\eta)$ is a small positive constant. Most of these zeros lie also in the circle Z_δ: $(\mathrm{Re} z+\delta)^2+(\mathrm{Im} z)^2=(1+\delta)^2$, and hence satisfy

$$\frac{1}{1+\delta} \le \frac{2(1-\mathrm{Re} z_i)}{|1-z_i|^2} \le 1.$$

Let the number of these zeros be $K(1-\eta')(1-\xi)$, ξ being a small positive constant. Then (24) may be written as

$$\langle n\rangle = K(1-\eta')(1-\xi)(1-O(\delta)) + \sum_{\substack{\mathrm{Re} z_i<1 \\ \text{outside} Z_\delta}} \frac{4(1-\mathrm{Re} z_i)}{|1-z_i|^2} + \sum_{\mathrm{Re} z_i>1} \frac{4(1-\mathrm{Re} z_i)}{|1-z_i|^2}. \tag{25}$$

The contributions from the first and second terms are positive whereas that of the last term is negative. Thus, if the last term is finite for $s\to\infty$, we may conclude from (16) and (25) that $\langle n\rangle \approx K$ for very large s. Unfortunately, there is no guarantee that this is the case. In fact, a zero in the Re $z>1$ region, if it lies at a distance $\sim 2/(eh(s))$ from $z=1$, may give a negative contribution as large as $-h(s)$. [No zero exists in $|z-1| < 2/(eh(s))$ according to Lemma 1.] Just a few such zeros are sufficient to render the right-hand side of (25) much smaller than K.

To see how close to $z=1$ a zero of $F_s(z)$ can be, let us assume that $F_s(z)$ has no zero in $|z-1|<r$, r being independent of s, and that $\langle n\rangle$ behaves as

$$\langle n\rangle = C[h(s)]^\alpha, \quad 0<\alpha\le 1, \tag{26}$$

where the constant C may be replaced by a very slowly varying function of s. Then it follows from (26) and (A15) of Lemma 3 that

$$r < \frac{64}{C^2}[h(s)]^{1-2\alpha}, \tag{27}$$

where $r \ll 1$ is assumed. For $\alpha > \frac{1}{2}$ this is in contradiction with the s-independence of r. In other words, when $\frac{1}{2} < \alpha \le 1$, $F_s(z)$ must have at least one pair of zeros in the circle

$$|z-1| < \frac{64}{C^2}[h(s)]^{1-2\alpha} \tag{28}$$

whose radius tends to zero for $s\to\infty$. Furthermore, some of these zeros must be in the region Re $z>1$ if $\langle n\rangle \ll K$ is to hold.

In the particular case where zeros of $F_s(z)$ are restricted to the region Re $z<1$, or more precisely to the region satisfying

$$\mathrm{Re} z < 1-\epsilon|1-z|^2, \tag{29}$$

ϵ being a small positive constant, it is possible to draw a much stronger conclusion:

Theorem 2: Suppose all zeros of $F_s(z)$ lie in the region defined by (29). Suppose further that $\langle n\rangle$ satisfies (26) with α restricted to $\frac{3}{4} < \alpha \le 1$. Then the multiplicity distribution $a_n(s)$, $n=2,3,\cdots$, is peaked at $n\approx\langle n\rangle$ in the sense that $\langle n^2\rangle/\langle n\rangle^2 = 1+O(\langle n\rangle^{-1})$, $\langle n^3\rangle/\langle n\rangle^3 = 1+O(\langle n\rangle + O(\langle n\rangle^{-1})$.

Proof: The function

$$\tilde F_s(z) = \prod_i{}' \frac{(1-z_i)(1-z_i^*)}{(z-z_i)(z-z_i^*)} F_s(z) \quad [\tilde F_s(1)=1] \tag{30}$$

is zero-free in $|z-1|<r$, if the product $\prod_i'$ is over all zeros of $F_s(z)$ in $|z-1|<r$ with $\mathrm{Im} z_i > 0$. Taking the logarithms of both sides, differentiating them n times with respect to z, and setting $z=1$, we obtain

$$\tilde c_n(s) = c_n(s) + \frac{1}{n}\sum_i{}'\left(\frac{1}{(z_i-1)^n}+\frac{1}{(z_i^*-1)^n}\right), \tag{31}$$

where

$$c_n(s) = \frac{1}{n!}\frac{d^n}{dz^n}\ln F_s(z)\Big|_{z=1}, \qquad \tilde c_n(s) = \frac{1}{n!}\frac{d^n}{dz^n}\ln F_s(z)\Big|_{z=1}. \tag{32}$$

In particular, we need an explicit form of (31) for $n=3$:

$$\tilde F_s^{(3)} - 3\tilde F_s''\tilde F_s' + 2(\tilde F_s')^3 = F_s^{(3)} - 3F_s''F_s' + 2(F_s')^3 + 4\sum_i{}' \frac{(1-\mathrm{Re} z_i)[(\mathrm{Im} z_i)^2-(1-\mathrm{Re} z_i)^2]}{|1-z_i|^6}, \tag{33}$$

where $F_s^{(n)}$, $\tilde F_s^{(n)}$ are the values of $F_s^{(n)}(z)$, $\tilde F_s^{(n)}(z)$ at $z=1$. Since all zeros are in the region satisfying (29), we obtain

$$0 < 4\sum_i{}' \frac{(1-\mathrm{Re} z_i)[(\mathrm{Im} z_i)^2-(1-\mathrm{Re} z_i)^2]}{|1-z_i|^6} = O([h(s)]^{4\alpha-2}) \tag{34}$$

from (28). On the other hand, we have from Lemma 4

$$|\tilde F_s^{(3)} - 3\tilde F_s''\tilde F_s' + 2(\tilde F')^3| = O(\tilde h(s)), \tag{35}$$

which is smaller than (34) for $\alpha > \frac{3}{4}$. Thus $F_s^{(3)} - 3F_s''F_s' + 2(F_s')^3$ must be *negative* and of order $[h(s)]^{4\alpha-2}$ in magnitude in (33). Rewriting it in the form

$$\tfrac{1}{2}[\textstyle\sum n^3 a_n - (\sum n^2 a_n)^2/(\sum n a_n)] + [\sum n^2 a_n - \frac{1}{2}(\sum n a_n)^2][\sum n^2 a_n - (\sum n a_n)^2]/(\sum n a_n) - \frac{3}{2}\sum n^2 a_n + \frac{3}{4}(\sum n a_n)^2 + \sum n a_n, \tag{36}$$

we see that it can take negative values if and only if

$$\sum n^3 a_n = (\sum n a_n)^3 + O((\sum n a_n)^2), \tag{37}$$

$$\sum n^2 a_n = (\sum n a_n)^2 + O(\sum n a_n), \tag{38}$$

where (38) actually follows from (37) and Schwarz's inequality. This result may be rewritten in a more trans-

parent form:

$$\frac{\langle n^2\rangle-\langle n\rangle^2}{\langle n\rangle^2}=O\left(\frac{1}{\langle n\rangle}\right),\quad \frac{\langle n^3\rangle-\langle n\rangle^3}{\langle n\rangle^3}=O\left(\frac{1}{\langle n\rangle}\right), \tag{39}$$

which proves the theorem.

Actually Theorem 2 will hold even if some zeros are in the region $\mathrm{Re}\,z>1$ insofar as they do not upset the relation (34).

Note also that the distribution of zeros obtained in Theorem 2 is qualitatively not dissimilar to that of the extreme model

$$F_s(z)=\tfrac{1}{2}+\tfrac{1}{2}z^{h(s)},\quad \text{i.e.,}\quad a_n(s)=\begin{cases}1 & \text{if } n=h(z),\\ 0 & \text{otherwise,}\end{cases} \tag{40}$$

whose zeros $z_k=\exp[i\pi k/h(s)]$, $k=1,3,5,\cdots$, are uniformly distributed on the unit circle $|z|=1$. This model gives $c_1(s)=h(s)/2$, $c_2(s)=h(s)(h(s)-2)/8$, $c_3(s)=-h(s)(h(s)-\frac{4}{3})/8$, etc.

Let us conclude this report with a few general remarks:

(a) In this article we have examined the relation between the asymptotic behavior of $\langle n\rangle$, $\langle n(n-1)\rangle$, $\cdots$, and the distribution of zeros of $F_s(z)$. We have found, under certain restrictions, that $F_s(z)$ has a large number of zeros accumulating on nearly every point of the circle $|z|=1$ if the average multiplicity $\langle n\rangle$ is divergent for $s\to\infty$. Such an accumulation will lead to the discontinuity of the "density" $\rho'(z)$ at $z=1$ if $\langle n\rangle\geq C\ln s$, a situation quite similar to that encountered in statistical mechanics of the lattice gas. We note that one finds $\langle n\rangle\geq C\ln s$, where C is a positive constant independent of s, in a planar dual model[12] as well as the scaling theory of Koba, Nielsen, and Olsen.[13]

(b) Information on the asymptotic behavior of the second moment $f_2(s)=\langle n(n-1)\rangle-\langle n\rangle^2$ would be very useful in constructing models of multiparticle production processes. If $f_2(s)$ is nonzero, one can infer the presence of long range correlations in hadron–hadron collisions, while the sign of $f_2(s)$ is crucial in characterizing more salient features of the models. In principle one should be able to obtain the properties of $f_2(s)$ knowing that of $\langle n\rangle$, $\langle n(n-1)\rangle$, etc. We have thus far been unable, however, to correlate the magnitude and sign of $f_2(s)$ with the position and number of zeros of $F_s(z)$ near $|z|=1$. [In the example (40) we have $f_2(s)=-h(s)$.]

(c) Equation (39) has some similarity with the KNO scaling,[13] which implies $\langle n^q\rangle/\langle n\rangle^q=C_q+O(1/\ln s)$, C_q being a constant independent of s. The underlying assumptions of this scaling law are: (i) The scaling functions

$$f^{(q)}(x_1,p_{1\perp},\cdots,x_q,p_{1q})=\frac{1}{\sigma_{\rm tot}}\frac{d\sigma}{d^3p_1/\omega_1\cdots d^3p_q/\omega_q}$$

are nonsingular at $x_1=\cdots=x_q=0$. (ii) For large s the function $f^{(q)}$ approaches a constant limit rapidly and does not even have variations of the form $C\ln s$. Our result (39) appears to imply that some consequences of the KNO scaling may be valid under less specific assumptions than those of Ref. 13.

(d) Recent experiments indicate that the s dependence of $\langle n\rangle$ is somewhere between $\ln s$ and $s^{1/4}$. If $\langle n\rangle=O(s^{1/4+\epsilon})$, $\epsilon>0$, $\alpha>\frac{1}{2}$ holds in (26) and $F_s(z)$ will have zeros tending to $z=1$. We note that some recent theories[14] based on Landau's hydrodynamical model predict $\langle n\rangle\sim s^{1/4}$.

(e) Recently Khuri[15] has obtained bounds on $\langle n\rangle$ and $\langle n^p\rangle$, p fixed and finite, under the following assumptions: (1) The cross sections $\sigma_n(s)$ have convergent power series expansions in some coupling constant λ. (2) σ_n are polynomially bounded in s for all complex values of λ within a circle of convergence. (3) The Jin–Martin lower bound is satisfied for all physical values of λ. His results are

$$\langle n\rangle\leq C\ln s,\quad \langle n^p\rangle\leq C^p(\ln s)^p, \tag{41}$$

where C is a constant independent of s. They may be regarded as special cases of Lemma 2 with $h(s)\propto\ln s$. In view of the stronger assumptions of Ref. 15, it would be interesting to see whether it is possible to obtain more detailed results than (41).

(f) It has been pointed out by Khuri[16] that the Froissart bound for $\sigma_{\rm tot}(s)$ can be improved by $\ln s$ if the zeros of his generating function [similar to (1)] do not collapse on the origin as $s\to\infty$. It may be of interest to study whether the method developed here is applicable for such a problem, too.

ACKNOWLEDGMENTS

The authors are grateful to the members of the high energy physics group of the University of Tokyo for their kind hospitality. We would like to thank Dr. H. Miyazawa for his interest and helpful comments. We are indebted to Dr. H. Cornille for pointing out an error in the preliminary version of this article. One of us (T.K.) wishes to thank Dr. W.H.J. Fuchs for very helpful discussions.

APPENDIX

Lemma 1: The number $n_r(s)$ of zeros of $F_s(z)$ within the circle $|z-1|<r$ is bounded from above by

$$n_r(s)\leq erh(s)\quad \text{for } s>s_0, \tag{A1}$$

where $e=2.718\cdots$, $er<R$, and $R\gg 1$.

Proof: According to Jensen's theorem[6] $n_r(s)$ satisfies the inequality

$$n_r(s)\leq\frac{1}{\ln(1/\delta)}\ln F_s(1+r/\delta),\quad \frac{r}{(R-1)}<\delta<1. \tag{A2}$$

On the other hand, an upper bound of $F_s(z)$ for $z=1+\rho<R$ is obtained by applying Hadamard's theorem[17] to the three circles $|z|=1$, $|z|=1+\rho$, $|z|=R$:

$$F_s(1+\rho)\leq(1+\rho)^{h(s)}, \tag{A3}$$

taking account of (12) and $F_s(1)=1$. From (A2) and (A3) we obtain

$$n_r(s)\leq\frac{\ln(1+r/\delta)}{\ln(1/\delta)}h(s)\quad (\text{for } r/\delta<R-1)$$

$$<\frac{r/\delta}{\ln(1/\delta)}h(s). \tag{A4}$$

Optimization with respect to δ yields the upper bound (A1).

Lemma 2: The quantities

$$b_n(s)=\frac{1}{n!}\frac{d^n}{dz^n}F_s(z)\Big|_{z=1} \tag{A5}$$

satisfy the inequalities

$$b_n(s)\leq \begin{cases}(eh(s)/n)^n, & n\ll h(s), \quad (A6)\\ \min[2^{h(s)}, R^{h(s)}(R-1)^{-n}], & n\leq N(s). \quad (A7)\end{cases}$$

In particular, for $n=1$ we have the stronger result

$$\langle n\rangle = 2b_1(s)\leq eh(s). \tag{A8}$$

Proof: From (A3) and Cauchy's inequality for a circle $|z-1|=\rho$ we obtain

$$b_n(s)\leq (1+\rho)^{h(s)}/\rho^n, \quad 0<\rho<R-1. \tag{A9}$$

Optimizing this with respect to ρ we get

$$b_n(s)\leq \left(\frac{h(s)}{h(s)-n}\right)^{h(s)-n}\left(\frac{h(s)}{n}\right)^n \text{ for } n<h(s). \tag{A10}$$

For $n\ll h(s)$ this can be approximated by (A6). In particular,

$$b_1(s)\leq eh(s), \tag{A11}$$

$$b_2(s)\leq \tfrac{1}{4}(eh(s))^2. \tag{A12}$$

Actually (A11) can be improved. Noting the positivity $a_n\geq 0$, we obtain

$$b_1(s)(b_1(s)-\tfrac{1}{2})\leq b_2(s), \tag{A13}$$

as is seen by applying Schwarz's inequality to $b_2(s)=\frac{1}{4}\sum n(n-1)a_n$. From (A12) and (A13) we find, for $h(s)\gg 1$,

$$b_1(s)\leq \tfrac{1}{4}+\tfrac{1}{4}(1+(2eh(s))^2)^{1/2} \approx \tfrac{1}{2}eh(s), \tag{A14}$$

an improvement of factor 2 over (A11).

Lemma 3: If $F_s(z)$ has no zero in $|z-1|<r$ for all $s>s_0\gg 4m^2$, and if $h(s)\to\infty$ for $s\to\infty$, then we have

$$|c_1(s)|\leq \frac{4}{r}(h(s)\ln(1+r))^{1/2}, \tag{A15}$$

$$|c_n(s)|\leq \frac{8(n-1)}{r^n}h(s)\ln(1+r), \quad n=2,3,\cdots, \tag{A16}$$

where $c_n(s)$ is defined by (32).

Proof: Under the assumption, $\ln F_s(z)$ is holomorphic in $|z-1|<r$. Cauchy's inequality gives for $0<\rho<r$

$$|c_n(s)|\leq \frac{M_\rho}{\rho^n}, \quad M_\rho=\max_{|z-1|=\rho}|\ln F_s(z)|. \tag{A17}$$

[Recall $F_s(1)=1$.] According to the Borel–Carathéodory theorem,[18] M_ρ is bounded as

$$M_\rho\leq \frac{2\rho}{r-\rho}\max_{|z-1|=r}\ln|F_s(z)|. \tag{A18}$$

From (A3), (A17), and (A18) we obtain

$$|c_n(s)|\leq \frac{2h(s)}{\rho^{n-1}(r-\rho)}\ln(1+r). \tag{A19}$$

Optimizing this with respect to ρ, we find

$$c_1(s)\leq \frac{2}{r}h(s)\ln(1+r), \tag{A20}$$

$$|c_n(s)|\leq \frac{8(n-1)}{r^n}h(s)\ln(1+r), \quad n=2,3,\cdots. \tag{A21}$$

(The factor 8 can be replaced by $2e$ for large n.)

Noting that

$$c_2(s)=b_2(s)-\tfrac{1}{2}b_1^2(s) \tag{A22}$$

is positive, we can obtain a better bound for $c_1(s)$ than (A20). From (A22) and (A13) we obtain

$$c_2(s)\geq \tfrac{1}{2}b_1(s)(b_1(s)-1). \tag{A23}$$

Combining this with (A21) for $n=2$, we find the improved bound (A15). However, we cannot improve the bound on $c_1(s)$ making use of (A21) for $n\geq 3$. This is because $c_n(s)$ is not positive definite for $n\geq 3$.

If $F_s(z)$ is defined by $\beta+(1-\beta)\sum_{n=2}^N a_n z^n$, $0<\beta<1$, instead of (3), the inequalities (A13) and (A23) become

$$\tfrac{1}{2}b_1(s)\left(\frac{b_1(s)}{1-\beta}-1\right)\leq b_2(s)$$

and

$$c_2(s)\geq \tfrac{1}{2}b_1(s)\left(\frac{\beta b_1(s)}{1-\beta}-1\right),$$

respectively. Thus the improved formula (A15) is critically dependent on the property that $\beta=F_s(0)\neq 0$. Formula (A20) holds for any regular function free of zeros in $|z-1|\leq r$, even for functions such as $\exp[h(s)(z-1)]$ if we replace $\ln(1+r)$ by r. This example fails to satisfy (A15), however, since its value at $z=0$, $\exp[-h(s)]$, tends to zero for $h(s)\to\infty$.

Lemma 4. $\tilde{c}_n(s)$ defined by (32) satisfy the inequalities

$$|\tilde{c}_1(s)|\leq \frac{2}{r}\tilde{h}(s)\ln(1+r), \tag{A24}$$

$$|\tilde{c}_n(s)|\leq \frac{2e(n-1)}{r^n}\tilde{h}(s)\ln(1+r), \quad n=2,3,\cdots, \tag{A25}$$

where

$$\tilde{h}(s)\equiv h(s)-n_r'(s)+\frac{n_r'(s)}{\ln(1+r)}\ln\left(\frac{e\,rh(s)}{n_r'(s)}\right), \tag{A26}$$

$n_r'(s)$ being the number of zeros of $F_s(z)$ satisfying

$$r/2<|z-1|<r. \tag{A27}$$

Proof: Since $\tilde{F}_s(z)$ defined by (30) has no zero for $|z-1|\leq r$, $\ln\tilde{F}_s(z)$ is holomorphic in $|z-1|\leq r$. Thus we find

$$|\tilde{c}_n(s)|\leq \frac{2}{\rho^{n-1}(r-\rho)}\max_{|z-1|=r}\ln|\tilde{F}_s(z)|, \quad \rho<r, \tag{A28}$$

in parallel with (A19). To evaluate an upper bound of the right-hand side, we note that

$$\max_{|z-1|=r+\epsilon}\left|\Pi'\frac{(1-z_i)(1-z_i^*)}{(z-z_i)(z-z_i^*)}\right|\leq\left(\frac{r}{\epsilon}\right)^{n_r'(s)}, \quad \epsilon>0, \tag{A29}$$

where the product is over the zeros satisfying (A27). For zeros in $|z-1|<r/2$, we obtain

$$\max_{|z-1|=r+\epsilon}\left|\Pi''\frac{(1-z_i)(1-z_i^*)}{(z-z_i)(z-z_i^*)}F_s(z)\right|\leq$$

$$\leq\left(\frac{r/2}{r/2+\epsilon}\right)^{n_r(s)-n_r'(s)}(1+r+\epsilon)^{h(s)}\lesssim(1+r+\epsilon)^{h(s)} \tag{A30}$$

taking account of (A3). From (A29) and (A30) we find

$$\max_{|z-1|=r+\epsilon} |\tilde{F}_s(z)| \leq \left(\frac{r}{\epsilon}\right)^{n_r'(s)} (1+r+\epsilon)^{h(s)}. \tag{A31}$$

Optimizing this with respect to ϵ we obtain

$$\max_{|z-1|=r} |\tilde{F}_s(z)| \leq (1+r)^{h(s)-n_r'(s)} \left(\frac{e\,r h(s)}{n_r'(s)}\right)^{n_r'(s)}. \tag{A32}$$

Substituting this in (A28) we obtain (A24) and (A25) for $n_r'(s) \ll h(s)$. Unlike (A15) it is not possible to improve (A24) since we cannot use (A13) here.

*Work supported in part by the National Science Foundation.
†John Simon Guggenheim Foundation Fellow, 1973-74.
‡Research Fellow, Ministry of Education, Japan. Present address: Institute of Physics, A/105 Saheed Nagar, Bhubaneswar-751007, India.

[1]R. P. Feynman, in *Particle Physics* (Irving Conference-1971), A. I. P. Conference Proceeding No. 6, p. 183; K. G. Wilson, Cornell University Preprint CLNS-131, 1971 (unpublished).
[2]An extensive literature exists on this topic. For a detailed list of references, see R. C. Arnold, ANL/HEP 7317 (unpublished); H. Harari, Lectures at Scottish Universities Summer School (1973).
[3]T. D. Lee, Phys. Rev. D 6, 3617 (1972); D. K. Campbell and S.-J. Chang, Phys. Rev. D 8, 2996 (1973); S.-J. Chang and T.-M. Yan, Phys. Rev. D 7, 3698 (1973).
[4]R. C. Arnold and G. H. Thomas, Phys. Lett. 47B, 371 (1973). See also R. C. Arnold, Ref. 2.
[5]C. N. Yang and T. D. Lee, Phys. Rev. 87, 404, 410 (1952).
[6]E. C. Titchmarsh, *The Theory of Functions* (Oxford U. P., London, 1939), p. 171.
[7]Ref. 6, p. 240.
[8]P. Erdös and P. Turán, Ann. Math. 51, 105 (1950).
[9]Ref. 6, p. 116.
[10]W. K. Hayman, *Multivalent Functions* (Cambridge U. P., London, 1958), p. 5.
[11]Private communication from Dr. W. H. J. Fuchs.
[12]C. Rosenzweig and G. Veneziano, Phys. Lett. 52B, 335 (1974).
[13]Z. Koba, H. B. Nielsen, and P. Olesen, Nucl. Phys. B 40, 317 (1972).
[14]P. Carruthers and M. Duong-Van, Phys. Lett. 41B, 597 (1972); F. Cooper and E. Schonberg, Phys. Rev. Lett. 30, 880 (1973); E. Suhonen, J. Enkenberg, K. E. Lassila, and S. Sohlo, Phys. Rev. Lett. 31, 1567 (1973).
[15]N. N. Khuri, Phys. Rev. D 9, 1802 (1974).
[16]N. N. Khuri, Phys. Rev. D 8, 2702 (1973).
[17]Ref. 6, p. 172.
[18]Ref. 6, p. 174.

Resonantly coupled nonlinear evolution equations

Mark J. Ablowitz

Department of Mathematics, Clarkson College of Technology, Potsdam, New York 13676

Richard Haberman

Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903
(Received 23 May 1975)

A differential matrix eigenvalue problem is used to generate systems of nonlinear evolution equations. They model triad, multitriad, self-modal, and quartet wave interactions. A nonlinear string equation is also recovered as a special case. A continuum limit of the eigenvalue problem and associated evolution equations are discussed. The initial value solution requires an investigation of the corresponding inverse-scattering problem.

1. INTRODUCTION

In recent years, a method of solution for the initial value problem of certain nonlinear evolution equations which describe physically important cases of nonlinear dispersive wave propagation has been developed. The inverse-scattering method was first ingeniously applied to the Korteweg–de Vries equation by Gardner, Greene, Kruskal, and Miura.[1] Lax[2] demonstrated an alternative formulation, which was successfully utilized by Zakharov and Shabat[3] to solve the nonlinear Schrödinger equation. Subsequent workers used these ideas in solving other interesting equations.[4–8] Recently, the coupled system of partial differential equations describing resonant triads has been considered by Zakharov and Manakov.[9] Ablowitz, Kaup, Newell, and Segur,[10,11] using an eigenvalue problem similar to that of Ref. 3, isolated a class of physically interesting nonlinear evolution equations tractable by the inverse-scattering method.

In this paper, we propose a class of higher order differential matrix eigenvalue problems motivated by Refs. 9 and 10. The resulting system of nonlinear evolution equations describes resonantly interacting wave envelopes. As special cases we: (i) rederive Zakharov and Manakov's result[9]; (ii) find a coupled system of nonlinear partial differential equations which describe the interaction of "multi-triad" resonant wave envelopes; (iii) discuss an additional example, which yields evolution equations describing the simultaneous interaction of triad, multitriad, self-modal, and quartet resonances; (iv) show how a nonlinear string equation[12–15] fits into this formulation; (v) as a limit of (iii), derive a differential–integral eigenvalue problem and an associated evolution equation. The evolution equations solvable by second order eigenvalue problems (nonlinear Schrödinger equation, etc.) are also recoverable. We do not attempt to perform the inverse-scattering analysis for the higher order eigenvalue problems.

2. EIGENVALUE PROBLEM

Consider the differential matrix eigenvalue problem

$$\frac{\partial \mathbf{v}}{\partial x} = i\zeta \mathbf{D}\mathbf{v} + \mathbf{N}\mathbf{v}, \tag{2.1}$$

where $\mathbf{v}$ is an n-dimensional vector eigenfunction and ζ is the eigenvalue. $\mathbf{D}$ and $\mathbf{N}$ are $n\times n$ matrices. The time dependency of the vector eigenfunction is chosen such that

$$\frac{\partial \mathbf{v}}{\partial t} = \mathbf{Q}\mathbf{v}, \tag{2.2}$$

where $\mathbf{Q}$ is also an $n\times n$ matrix. The eigenvalue problem of Zakharov and Shabat,[3] generalized by Ablowitz, Kaup, Newell, and Segur,[10] is equivalent to (2.1) for the 2×2 case. The eigenvalue problem of Zakharov and Manakov[9] may be put into this form. By cross-differentiation of (2.1) and (2.2) the eigenvalues ζ are constant in time t if

$$\mathbf{Q}_x = \mathbf{N}_t + i\zeta[\mathbf{D},\mathbf{Q}] + [\mathbf{N},\mathbf{Q}], \tag{2.3}$$

where $[\mathbf{A},\mathbf{B}]$ is the commutator of $\mathbf{A}$ and $\mathbf{B}$,

$$[\mathbf{A},\mathbf{B}] = \mathbf{AB} - \mathbf{BA}.$$

We show that simple matrix operators $\mathbf{Q}$ and $\mathbf{N}$ exist satisfying (2.3), such that (2.3) is equivalent to the "generalized potentials" N_{ij}, $i\neq j$, evolving in time according to a coupled system of nonlinear partial differential equations. Here, as in the corresponding 2×2 case by Ablowitz, Kaup, Newell, and Segur,[10] solutions to (2.3) can be obtained by expanding $\mathbf{Q}$ in powers[16] or inverse powers of ζ. There are many possibilities.

3. RESONANT TRIADS

As the first example, we will show that the choice

$$\mathbf{Q} = \mathbf{Q}^{(1)} + \zeta\mathbf{Q}^{(0)} \tag{3.1}$$

yields the coupled system of nonlinear partial differential equations describing resonant triads (in the 3×3 case) and multitriad resonance (in the $n\times n$ case, $n>3$). The substitution of (3.1) into the commutator equation (2.3) yields the following equations in powers of ζ:

$$O(\zeta^2):\quad 0 = i[\mathbf{D},\mathbf{Q}^{(0)}], \tag{3.2a}$$

$$O(\zeta^1):\quad \mathbf{Q}^{(0)}_x = i[\mathbf{D},\mathbf{Q}^{(1)}] + [\mathbf{N},\mathbf{Q}^{(0)}], \tag{3.2b}$$

$$O(\zeta^0):\quad \mathbf{Q}^{(1)}_x = \mathbf{N}_t + [\mathbf{N},\mathbf{Q}^{(1)}]. \tag{3.2c}$$

$\mathbf{D}$ and $\mathbf{Q}^{(0)}$ are chosen to be diagonal matrices, $D_{ij} = \delta_{ij}a_j$ and $Q^{(0)}_{ij} = \delta_{ij}c_j$, so that (3.2a) is automatically satisfied, where a_j and c_j are all assumed constant. Equation (3.2b) is simplified since $\mathbf{Q}^{(0)}_x = 0$ and can be

shown to yield

$$Q^{(1)}_{ik} = [(c_i - c_k)/i(a_i - a_k)]\, N_{ik} \quad \text{for} \quad i \neq k, \tag{3.3}$$

where we assume $a_i \neq a_k$ for $i \neq k$. The diagonal entries of (3.2c) imply that

$$Q^{(1)}_{ii_x} = N_{ii_t}. \tag{3.4}$$

The nondiagonal entries yield a system of $n(n-1)$ coupled nonlinear partial differential equations which the potentials N_{ik} $(i \neq k)$ must satisfy

$$\alpha_{ik} N_{ik_x} = N_{ik_t} + \sum_{j \neq i,k} N_{ij} N_{jk} (\alpha_{jk} - \alpha_{ij}) + (Q^{(1)}_{kk} - Q^{(1)}_{ii}) N_{ik} + \alpha_{ik} N_{ik} (N_{ii} - N_{kk}), \tag{3.5}$$

for $i \neq k$, where

$$\alpha_{ik} = (c_k - c_i)/i(a_k - a_i) = \alpha_{ki}.$$

The ikth entry N_{ik} has a linearized velocity α_{ik}, which we thus insist is real. Since $\alpha_{ik} = \alpha_{ki}$, N_{ik}'s velocity is the same as that corresponding to N_{ki}. In most physical problems different modes are propagated at different velocities, and hence this suggests N_{ik} depends on N_{ki}. Furthermore, for system (3.5) to represent a resonance phenomena,

$$N_{jk} = \sigma_{jk} N^*_{kj} \quad \text{for } j > k, \tag{3.6}$$

where the σ_{jk} are real normalizing coefficients and where $*$ indicates the complex conjugate. In this case, the ikth equation will be equivalent to the kith equation if

$$\sigma_{jk}\sigma_{ij} = -\sigma_{ik} \quad \text{for all} \quad i > j > k \tag{3.7}$$

and the diagonal elements $Q^{(1)}_{ii}$ and N_{ii} are all pure imaginary. This last requirement allows us to take $Q^{(1)}_{ii} = N_{ii} = 0$ without loss of generality, as they effect only the phase of N_{ik}. A further reduced form of (3.5) can therefore be written as

$$\alpha_{ik} N_{ik_x} = N_{ik_t} + \sum_{j>k>i} \sigma_{jk} N_{ij} N^*_{kj} (\alpha_{jk} - \alpha_{ij}) + \sum_{k>j>i} N_{ij} N_{jk} (\alpha_{jk} - \alpha_{ij}) + \sum_{k>i>j} \sigma_{ij} N^*_{ji} N_{jk} (\alpha_{jk} - \alpha_{ij}) \tag{3.8}$$

for $k > i$.

In the 2×2 case, the nonlinear interaction terms vanish and the uncoupled and uninteresting linear system results,

$$\alpha_{ik} N_{ik_x} = N_{ik_t}, \tag{3.9}$$

for $k > i$.

In the 3×3 case, the nonlinear interaction terms do not vanish. Instead (3.8) is equivalent to the result of Zakharov and Manakov.[9] In particular, the potentials N_{ik} $(i \neq k)$ represent the amplitudes of a triad of resonantly interacting wave envelopes. For example, by letting $N_{12} = A_1$, $N_{13} = A_2$, and $N_{23} = A_3$, Eq. (3.8) becomes

$$\begin{aligned} A_{1_t} &= \alpha_{12} A_{1_x} + \sigma_{32} (\alpha_{13} - \alpha_{32}) A_2 A^*_3, \\ A_{2_t} &= \alpha_{13} A_{2_x} + (\alpha_{12} - \alpha_{23}) A_1 A_3, \\ A_{3_t} &= \alpha_{23} A_{3_x} + \sigma_{21} (\alpha_{21} - \alpha_{13}) A^*_1 A_2. \end{aligned} \tag{3.10}$$

By scaling the A_i's, and by noting that (3.7) becomes $\sigma_{21}\sigma_{32} = -\sigma_{31}$ in the 3×3 case, two distinct systems are allowed. The "decay interaction" occurs when all the σ_{mn} are negative; otherwise it is the "explosive case." The "decay interaction" case has the existence of a positive definite conserved "energy."

In the general $n \times n$ case, $n > 3$, the system is more complicated. (3.8) describes resonantly interacting sets of triads, what we call "multitriads," as described below. There are $n(n-1)/2$ interacting waves. The following notation for these waves is employed:

k_{ij} is the wavenumber and $\omega_{ij} = \omega(k_{ij})$, where $\omega(k)$ is the real dispersion relation. Then (3.8) describes the nonlinear interaction between the $n(n-1)/2$ wave envelopes that obey the following resonance conditions:

$$\begin{aligned} k_{ik} &= \underbrace{k_{ij} - k_{kj}}_{j>k>i} = \underbrace{k_{ij} + k_{jk}}_{k>j>i} = \underbrace{k_{jk} - k_{ji}}_{k>i>j}, \\ \omega_{ik} &= \underbrace{\omega_{ij} - \omega_{kj}}_{} = \underbrace{\omega_{ij} + \omega_{jk}}_{} = \underbrace{\omega_{jk} - \omega_{ji}}_{}, \end{aligned} \tag{3.11}$$

for $k > i$. Again, if all the σ_{ij} are negative, then there is a positive definite conserved "energy."

The case in which there are six waves $(n = 4)$ may be described as follows. Three waves are nonlinearly coupled when

$$\mathbf{k}_1 = \mathbf{k}_2 + \mathbf{k}_3$$

and

$$\omega_1 = \omega_2 + \omega_3,$$

$\omega_i = \omega(\mathbf{k}_i)$. The partial differential equations governing such a resonant triad were obtained in the field of nonlinear optics by Bloembergen[18] and for a general nonlinear dispersive media by Benney and Newell.[19] If the waves k_1, k_2, and k_3 form a resonant triad, then frequently one of the waves of the triad (for example, k_1) also simultaneously is a part of another resonant triad. In this case

$$k_1 = k_2 + k_3 = k_4 + k_5$$

and

$$\omega_1 = \omega_2 + \omega_3 = \omega_4 + \omega_5.$$

Examples of this are known to occur in nonlinear optics.[20,21] Other waves will be quadratically generated by the waves k_1, k_2, k_3, k_4, and k_5. Consider the wave $k_4 - k_3$ which we define to be k_6,

$$k_6 \equiv k_4 - k_3.$$

If, in addition to the above resonances, k_6 forms a resonant triad with k_4 and k_3, i.e., *if*

$$\omega_6 = \omega_4 - \omega_3,$$

then the entire system of six waves is resonantly coupled since $k_6 = k_4 - k_3 = k_2 - k_5$ and $\omega_6 = \omega_4 - \omega_3 = \omega_2 - \omega_5$. Each wave forms part of two distinct triads. An interaction of the above type, involving two coupled resonant triads (including six waves), we call "multitriad" resonant.

4. HIGHER ORDER RESONANCES

Consider

$$\mathbf{Q}=\mathbf{Q}^{(2)}+\zeta\mathbf{Q}^{(1)}+\zeta^2\mathbf{Q}^{(0)}. \tag{4.1}$$

In the $n\times n$ case, using the methods we have described in Sec. 3, the following system of evolution equations results:

$$\begin{aligned}
&\beta_{ij}N_{ij_{xx}}+\epsilon_{ij}N_{ij_x}-\sum_{\substack{k\neq i\\ k\neq j}}\gamma_{ijk}\frac{\partial}{\partial x}(N_{ik}N_{kj})\\
&=N_{ij_t}+\theta_{ij}N_{ij}+\sum_{\substack{k\neq i\\ k\neq j}}N_{ik}N_{kj}(\epsilon_{kj}-\epsilon_{ik})\\
&\quad+N_{ij}\Big(2\beta_{ij}N_{ji}N_{ij}+\sum_{\substack{k\neq i\\ k\neq j}}[(\beta_{kj}+\gamma_{ikj})N_{jk}N_{kj}\\
&\quad-(\beta_{ki}+\gamma_{kji})N_{ik}N_{ki}]\Big)+\sum_{\substack{k\neq i\\ k\neq j}}(\beta_{kj}N_{ik}N_{kj_x}-\beta_{ik}N_{kj}N_{ik_x})\\
&\quad+\sum_{\substack{k\neq i\\ k\neq j}}\sum_{\substack{m\neq i\\ m\neq j}}(\gamma_{ikm}N_{kj}N_{im}N_{mk}-\gamma_{kjm}N_{ik}N_{km}N_{mj})\\
&\qquad\qquad(i\neq j),
\end{aligned} \tag{4.2}$$

where

$$\begin{aligned}
&\alpha_{ij}=(c_i-c_j)/i(a_i-a_j)=\alpha_{ji},\\
&\beta_{ij}=\alpha_{ij}/i(a_i-a_j)=-\beta_{ji},\\
&\gamma_{ijk}=(\alpha_{kj}-\alpha_{ik})/i(a_i-a_j)=\gamma_{jik}=\gamma_{kij},\\
&\epsilon_{ij}=(Q_{ii}^{(1)}-Q_{jj}^{(1)})/i(a_i-a_j)=\epsilon_{ji},\\
&\theta_{ij}=\bar{Q}_{jj}^{(2)}-\bar{Q}_{ii}^{(2)}
\end{aligned} \tag{4.3a}$$

and

$$\begin{aligned}
&D_{ij}=a_i\delta_{ij},\\
&Q_{ij}^{(0)}=c_i\delta_{ij},\\
&Q_{ij}^{(1)}=\alpha_{ij}N_{ij},\quad i\neq j,
\end{aligned}$$

$$Q_{ij}^{(2)}=\begin{cases}\beta_{ij}N_{ij_x}+\epsilon_{ij}N_{ij}-\sum_{\substack{k\neq i\\ k\neq j}}\gamma_{ijk}N_{ik}N_{kj}, & i\neq j,\\ \bar{Q}_{ii}^{(2)}+\sum_{k\neq i}\beta_{ki}N_{ik}N_{ki}, & i=j,\end{cases} \tag{4.3b}$$

where a_i, c_i, $Q_{ii}^{(1)}$, $\bar{Q}_{ii}^{(2)}$ are all constant. For convenience, it has been assumed that $N_{ii}=0$. We again let $N_{ij}=\sigma_{ij}N_{ji}^*$ $(i>j)$. Then the ijth equation is equivalent to the jith equation if a_i is real, c_i is imaginary (thus α_{ij} is real, β_{ij} is imaginary, and γ_{ijk} is imaginary), $Q_{ii}^{(1)}$ is imaginary (thus ϵ_{ij} is real), $\bar{Q}_{ii}^{(2)}$ is imaginary, and

$$\sigma_{jk}\sigma_{ij}=-\sigma_{ik}\quad\text{for all } i>j>k.$$

In the 2×2 case, the nonlinear Schrödinger equation is found (since there are no values of k such that $k\neq i$ and $k\neq j$).

In the general 3×3 case, the last expression in (4.2) vanishes since k must equal m. Dispersion is represented by the terms $iN_{ij_{xx}}$ and N_{ij_t}. The terms of the form $N_{ij}[\cdots]$ in (4.2) represent cubic self-modal interactions; the first term in the bracket being a cubic self-self interaction. In physical problems, these terms are usually of smaller order of magnitude than the triad term. The terms $(\partial/\partial x)(N_{ik}N_{kj})$ and $N_{ik}N_{kj_x}$ for example represent a higher order type of triad resonance. Newell[22] discusses the effect of these terms. In the 3×3 case, by letting $A_1=N_{12}$, $A_2=N_{13}$, $A_3=N_{23}$ and by using the relations (4.3) and $N_{ij}=\sigma_{ij}N_{ji}^*$, we obtain the evolution equations

$$\begin{aligned}
A_{1_t}&=\beta_{12}A_{1_{xx}}+\epsilon_{12}A_{1_x}-\theta_{12}A_1+(\epsilon_{13}-\epsilon_{23})\sigma_{32}A_2A_3^*\\
&\quad-\gamma_{123}\sigma_{32}(A_2A_3^*)_x+\beta_{13}\sigma_{32}A_{2_x}A_3^*\\
&\quad+\beta_{23}\sigma_{32}A_2A_{3_x}^*-A_1[2\beta_{12}\sigma_{21}A_1A_1^*\\
&\quad+\sigma_{32}(\gamma_{123}-\beta_{23})A_3A_3^*-\sigma_{31}(\gamma_{123}-\beta_{13})A_2A_2^*],
\end{aligned} \tag{4.4a}$$

$$\begin{aligned}
A_{2_t}&=\beta_{13}A_{2_{xx}}+\epsilon_{13}A_{2_x}-\theta_{13}A_2+(\epsilon_{12}-\epsilon_{23})A_1A_3\\
&\quad-\gamma_{123}(A_1A_3)_x+\beta_{12}A_{1_x}A_3-\beta_{23}A_1A_{3_x}\\
&\quad-A_2[2\beta_{13}\sigma_{31}A_2A_2^*+\sigma_{32}(\gamma_{123}+\beta_{23})A_3A_3^*\\
&\quad-\sigma_{21}(\gamma_{123}-\beta_{12})A_1A_1^*],
\end{aligned} \tag{4.4b}$$

$$\begin{aligned}
A_{3_t}&=\beta_{23}A_{3_{xx}}+\epsilon_{23}A_{3_x}-\theta_{23}A_3+(\epsilon_{12}-\epsilon_{13})\sigma_{21}A_1^*A_2\\
&\quad-\gamma_{123}\sigma_{21}(A_1^*A_2)_x-\beta_{13}\sigma_{21}A_1^*A_{2_x}\\
&\quad-\beta_{12}\sigma_{21}A_2A_{1_x}^*-A_3[2\beta_{23}\sigma_{32}A_3A_3^*\\
&\quad+\sigma_{31}(\gamma_{123}+\beta_{13})A_2A_2^*-\sigma_{21}(\gamma_{123}+\beta_{12})A_1A_1^*].
\end{aligned} \tag{4.4c}$$

In this form it is clear which terms represent linear dispersion, cubic interactions, and contributions from the triad resonance. Since all of these terms arise in physical problems, (4.4) should make an excellent model system for further study.

It should also be mentioned that a very interesting nonlinear partial differential equation falls into this formulation. The equation is,

$$w_{tt}-w_{xx}-6(w^2)_{xx}-w_{xxxx}=0. \tag{4.5}$$

(4.5) was obtained by Boussinesq[12] in his study of long water waves [although (4.5) is not generally referred to as the "Bossinesq equation" by fluid dynamicists], and independently considered by Kruskal and Zabusky (see, for example, Ref. 13) in their study of one Fermi–Pasta–Ulam problem. Zakharov[15] has proposed an associated eigenvalue problem. We find similar results via our formulation.

The reduction proceeds as follows. In (4.2) for $n=3$, let $\epsilon_{ij}=\theta_{ij}=0$ and

$$\mathbf{N}=\begin{bmatrix}0 & 0 & 1\\ N_{21} & 0 & (1+\Omega_3)N_{31}\\ N_{31} & 1 & 0\end{bmatrix}, \tag{4.6}$$

where $\Omega_3=\exp(-2\pi i/3)$ is a cube root of unity. If (4.2) is to be consistent with $\mathbf{N}$ chosen via (4.6) we are forced to choose:

$$\beta_{13}=\beta_{21}=-\beta_{23},\quad \gamma_{123}=\beta_{21}(1+2\Omega_3) \tag{4.7}$$

in which case the a_i are arbitrary as long as

$$a_1+\Omega_3a_3+\Omega_3^2a_2=0.$$

The system (4.2) then reduces to

$$N_{31_t}+\beta_{21}N_{31_{xx}}-(2\beta_{21}/\Omega_3)N_{21_x}=0, \tag{4.8a}$$

$$N_{21_t}-\beta_{21}N_{21_{xx}}-2\beta_{21}(1-\Omega_3)N_{31}N_{31_x}=0. \tag{4.8b}$$

We may further reduce (4.8a, b) to a single equation by letting

$$N_{31}=\mu\phi_x+\nu, \tag{4.9}$$

whereupon (4.8a) yields

$$N_{21}=\mu[(\Omega_3/2\beta_{21})\phi_t+(\Omega_3/2)\phi_{xx}]. \tag{4.10}$$

Upon choosing

$$w=\phi_x,$$
$$\mu=-(1-\Omega_3),$$
$$\nu=\frac{1-\Omega_3}{-12}, \tag{4.11}$$
$$\beta_{21}=1,$$

Eq. (4.8b) reduces to the nonlinear string equation (4.5).

It is known[14] that (4.5) has soliton solutions. However, it is to be expected that the inverse problem associated with (2.1), where $\mathbf{N}$ is chosen via (4.6), must be studied in order to do the general initial value problem. The inverse question is presumably best studied using the single equation

$$v_{1_{xxx}}=(\lambda+N_{31_x}+N_{21})v_1+(2+\Omega_3)N_{31}v_{1_x} \tag{4.12}$$

(resulting from $a_1=\Omega_3^2$, $a_2=\Omega_3$, $a_3=1$, $\lambda=-i\xi^3$), rather than in the system form. This is also true for the Korteweg–de Vries equation, and is discussed in Refs. 10 and 11.

Finally, it should be noted that in the case $n=3$, where N_{ij} tends to zero at infinity, the inverse problem has been recently considered by Kaup[23] and Zakharov and Manakov.[24]

In the $n\times n$ case ($n>3$), the last expression in (4.2) no longer vanishes. However, it represents resonant quartets. Whenever multitriad resonance physically occurs, then multiquartet resonance must also occur since, from (3.11),

$$k_{ik}=k_{im}-k_{jm}-k_{kj}, \quad m>j>k>i,$$

and

$$\omega_{ik}=\omega_{im}-\omega_{jm}-\omega_{kj}, \quad m>j>k>i.$$

(4.2) therefore describes the balance between dispersion and many kinds of resonant interactions. In physical problems, we note that these resonant processes are usually of different orders of magnitude and hence (4.2) is only a model system of evolution equations.

5. CONTINUUM RESONANCE

Finally, we note an interesting generalization of the above ideas. The following eigenvalue problem and associated time dependence,

$$\frac{\partial v}{\partial x}(x,y;t)=i\xi d(y)v(x,y;t)+\int_{-\infty}^{\infty}N(x,y,z;t)v(x,z;t)\,dz, \tag{5.1}$$

$$\frac{\partial v}{\partial t}(x,y;t)=\int_{-\infty}^{\infty}Q(x,y,z;t)v(x,z;t)\,dz, \tag{5.2}$$

may be viewed as a continuum ($n\to\infty$) version of (2.1) and (2.2). Setting $v_{xt}=v_{tx}$ and $\xi_t=0$ yields

$$\begin{aligned}Q_x(x,y,z;t)&=N_t(x,y,z;t)+i\xi[d(y)-d(z)]Q(x,y,z;t)\\&\quad+\int_{-\infty}^{\infty}[Q(x,z',z;t)N(x,y,z';t)-N(x,z',z;t)\\&\quad\times Q(x,y,z';t)]\,dz',\end{aligned} \tag{5.3}$$

as the analogy of (2.3). Expansion in powers of ξ yield nonlinear evolution equations. Letting, for example,

$$Q=Q^{(1)}+\xi Q^{(0)}$$

results in the evolution equation,

$$\begin{aligned}&\alpha(y,z)N_x(x,y,z;t)\\&\quad=N_t(x,y,z;t)+\int_{-\infty}^{\infty}[\alpha(z',z)-\alpha(y,z')]\\&\qquad\times N(x,y,z';t)N(x,z',z;t)\,dz',\end{aligned} \tag{5.4}$$

where

$$\alpha(y,z)=[c(z)-c(y)]/i[d(z)-d(y)]=\alpha(z,y),$$
$$Q^{(0)}=\delta(y-z)c(y),$$
$$Q^{(1)}=-\alpha(y,z)N(x,y,z;t),$$
$$N(x,y,y)=0. \tag{5.5}$$

The symmetry condition

$$N(x,y,z;t)=\sigma(y,z)N^*(x,z,y;t) \tag{5.6}$$

for $y>z$ is consistent if σ satisfies

$$\sigma(y,z')\sigma(z',z)=-\sigma(y,z) \tag{5.7}$$

for $y>z'>z$. (5.4)–(5.7) are a continuum version of the multitriad resonance discussed in Sec. 3. Indeed, (5.6) says we need solve only for $N(x,y,z;t)$, $z>y$. Hence Eq. (5.4) [with condition (5.7)] may be thought of as a resonance equation resulting from an appropriate interaction of a continuous spectrum of modes.

Although the analysis is largely formal, it nevertheless suggests that differential–integral eigenvalue problems such as (5.1) may lead to the solution of higher variable equations such as (5.4).

6. CONCLUSION

We have provided a procedure to generate isospectral flows which are equivalent to systems of nonlinear partial differential equations. The method of solution for the initial value problem in the general $n\times n$ case requires an investigation of the corresponding inverse-scattering problem. We have not done that here.

ACKNOWLEDGMENTS

Portions of this work were supported by NSF Grants GP32839X and GA27727A1.

Note added in proof: We have become aware that V. E. Zakharov and A. B. Shabat [Func. Anal. Appl. **8**, 226 (1974)] have obtained, by a different scheme, (3.8) and (4.5).

[1] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967); Comm. Pure Appl. Math. **27**, 97 (1974).
[2] P. D. Lax, Comm. Pure Appl. Math. **21**, 467 (1968).
[3] V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[4] G. L. Lamb, Jr., Phys. Rev. Lett. **31**, 196 (1973).
[5] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973).
[6] M. Wadati, J. Phys. Soc. Japan **32**, 1681 (1972).
[7] H. Flaschka, Phys. Rev. B **9**, 1924 (1974).
[8] M. J. Ablowitz and J. F. Ladik, J. Math. Phys. **16**, 598 (1975).

[9]V. E. Zakharov and S. V. Manakov, Zh. Eksp. Teor. Fiz. Pis. Red. **18**, 413 (1973) [JETP Lett. **18**, 243 (1973)].
[10]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973).
[11]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[12]Boussinesq, J. Math. Pures Appl. **17**, 55 (1872).
[13]N. J. Zabusky, *Proceedings of the Symposium on Nonlinear Partial Differential Equations*, edited by W. F. Ames (Academic, New York, 1967).
[14]R. Hirota, J. Math. Phys. **14**, 810 (1973).
[15]V. E. Zakharov, Zh. Eksp. Teor. Fiz. **65**, 219 (1973) [Sov. Phys. JETP **38**, 108 (1974)].
[16]Haberman[17] has shown that expanding $\mathbf{Q}$ in powers of ζ is usually equivalent to procedures developed by Lax[2] and used by Zakharov, *et al.*[3,9]
[17]R. Haberman (to be published).
[18]N. Bloembergen, *Nonlinear Optics* (Benjamin, New York, 1965).
[19]D. J. Benney and A. C. Newell, J. Math. & Phys. **46**, 133 (1967).
[20]J. M. Yarborough and E. O. Ammann, Appl. Phys. Lett. **18**, 145 (1971).
[21]S. A. Akhmanov and R. V. Khokhlov, *Problems of Nonlinear Optics* (Gordon and Breach, New York, 1972).
[22]A. C. Newell, J. Fluid Mech. **35**, 255 (1969).
[23]D. J. Kaup (preprint).
[24]V. E. Zakharov and S. V. Manakov (preprint).

A new solution of the Einstein–Maxwell equations*

Raymond G. McLenaghan and Nessim Tariq

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
(Received 2 July 1975)

A space–time is determined which is a solution of the Einstein–Maxwell equations for a nonsingular electromagnetic field and for which the electromagnetic field tensor is weakly parallelly propagated along its principal null directions. A coordinate system is given in which the metric depends upon one essential arbitrary constant. The space–time admits a four-parameter simply transitive group of motions and its Weyl tensor is of Petrov type I.

1. INTRODUCTION AND HYPOTHESES

In this paper we consider space-times[1] V_4 satisfying the following conditions:

I. The Einstein—Maxwell equations are satisfied, that is,

$$R_{ab}=F_{ac}F_{b}{}^{c}-\tfrac{1}{4}g_{ab}F_{cd}F^{cd}, \tag{1.1}$$

$$F_{ab;}{}^{b}=0, \tag{1.2a}$$

$$F_{[ab;c]}=0. \tag{1.2b}$$

II. The electromagnetic field tensor F_{ab} is nonsingular. This implies that there exists a null tetrad[2] $(l_a, m_a, \bar{m}_a, n_a)$ such that

$$\overset{+}{F}_{ab}=\phi(l_{[a}n_{b]}-m_{[a}\bar{m}_{b]}), \tag{1.3}$$

where[3]

$$\overset{+}{F}_{ab}=F_{ab}-i*F_{ab}. \tag{1.4}$$

The null tetrad is determined by F_{ab} up to the transformation

$$l'_a=e^{a}l_a, \quad m'_a=e^{ib}m_a, \quad n'_a=e^{-a}n_a. \tag{1.5}$$

III. The tensor F_{ab} is "weakly" parallelly propagated along its two principal null directions defined by l_a and n_a. By this we mean that

$$\overset{+}{F}_{ab;c}l^{c}=f\overset{+}{F}_{ab}, \tag{1.6}$$

$$\overset{+}{F}_{ab;c}n^{c}=g\overset{+}{F}_{ab}. \tag{1.7}$$

IV. The Weyl tensor and Ricci tensor satisfy

$$2C_{abcd}l^{a}n^{b}(l^{c}n^{d}-m^{c}\bar{m}^{d})+R_{ab}(l^{a}n^{b}+m^{a}\bar{m}^{b})=0. \tag{1.8}$$

V. The null congruence associated to the principal null vector l^a is expansion free, that is,

$$l^{a}{}_{;a}=0. \tag{1.9}$$

The above conditions are invariant under the transformations (1.5).

The main result of this paper is given in the following theorem:

Theorem 1. In a space—time satisfying the conditions I—V there exists a system of coordinates (u, x, y, v) for which the metric and the electromagnetic field have the following form:

$$ds^2=du^2+2du\,dv-2kx\,dy(du+dv)$$
$$+(k^2x^2-e^{kv})\,dy^2-e^{-kv}\,dx^2, \tag{1.10}$$

$$\overset{+}{F}=(1/\sqrt{2})\,k\exp(ikv)(dv\wedge du+kxdy\wedge dv+i\,dx\wedge dy), \tag{1.11}$$

where k is any positive real number.

A special case of this metric corresponding to $k=4$ has been found by Tariq and Tupper.[4] They were searching for solutions of the Einstein—Maxwell equations for nonsingular electromagnetic fields under the assumption[5] that a null tetrad associated with F_{ab} is parallelly propagated along the null congruences defined by l_a and n_a. The above metric is included in the case when the expansion of the null congruence defined by l_a is zero. In the same paper they found the general solution corresponding to the case when the twist of the null congruence defined by l_a is zero.

Our conditions III and IV are equivalent to the parallel propagation assumption of Tariq and Tupper. This equivalence is made precise in the following theorem:

Theorem 2. The conditions III and IV imply that there exists in the family of null tetrads defined by F_{ab} a null tetrad whose vectors are parallelly propagated along the null congruences defined by the vectors l_a and n_a. Conversely, if a null tetrad associated with F_{ab} is parallelly propagated along the null congruences de- by l_a and n_a, the conditions III and IV are satisfied. The proof of this theorem is given in Sec. 3.

The advantage of employing the conditions III and IV rather than the parallel propagation condition is that the conditions III and IV are invariant under the complete group defined by (1.5) whereas the parallel propagation condition is not. It also seems preferable to impose, as far as possible, any supplementary conditions directly on the electromagnetic field tensor rather than on the spin coefficients.

The proof of Theorem 1 is given in Sec. 3, 4, and 5. We shall employ the vectorial formalism of Cahen, Debever, and Defrise[6,7] for our calculations. A brief description of this formalism together with the correspondence[8] with the equivalent Newman—Penrose[9] (NP) spin coefficient formalism is given in Sec. 2. In Sec. 6 we discuss some of the properties of the metric.

2. THE VECTORIAL FORMALISM

Let l_a and n_a denote two future pointing real vector fields on V_4 and let m_a denote a complex vector field such that metric on V_4 has the form

$$g_{ab}=2l_{(a}n_{b)}-2m_{(a}\bar{m}_{b)}. \tag{2.1}$$

Then the vectors $h^i{}_a = \{l_a, m_a, \bar{m}_a, n_a\}$ define a null tetrad on V_4. If we introduce the 1-forms

$$\theta^0 = l_a\, dx^a, \quad \theta^1 = m_a\, dx^a, \quad \theta^3 = n_a\, dx^a, \tag{2.2}$$

the relation (2.1) can be written as

$$ds^2 = 2\theta^0\theta^3 - 2\theta^1\theta^2 \equiv g_{ij}\,\theta^i\theta^j. \tag{2.3}$$

A basis for the space of self-dual 2-forms (bivectors) is given by

$$Z^1 = \theta^2 \wedge \theta^3, \quad Z^2 = \theta^0 \wedge \theta^1, \quad Z^3 = \tfrac{1}{2}(\theta^0 \wedge \theta^3 - \theta^1 \wedge \theta^2). \tag{2.4}$$

The metric in this space is

$$\gamma^{\alpha\beta} = 2\delta_1^{(\alpha}\delta_2^{\beta)} - \tfrac{1}{2}\delta_3^\alpha\delta_3^\beta. \tag{2.5}$$

The absence of torsion of the pseudo-Riemannian connection on V_4 is expressed by Cartan's first structure equation

$$d\theta^i + \omega^i{}_j \wedge \theta^j = 0, \tag{2.6}$$

where the $\omega^i{}_j$ denote the 1-form valued components of the connection. The relationship between the connection and the rotation coefficients $\gamma^i{}_{jk}$ of the null tetrad is

$$\omega^i{}_j = -\gamma^i{}_{jk}\theta^k, \tag{2.7}$$

where

$$\gamma^i{}_{jk} = h^i{}_{a;b}\, h_j{}^a\, h_k{}^b. \tag{2.8}$$

In the vectorial formalism Eq. (2.6) has the form

$$dZ^\alpha + \sigma^\alpha{}_\beta \wedge Z^\beta = 0. \tag{2.9}$$

The relationship between $\sigma^\alpha{}_\beta$ and $\omega^i{}_j$ will now be given. Firstly, we define

$$\sigma_{\alpha\beta} = \gamma_{\alpha\gamma}\sigma^\gamma{}_\beta. \tag{2.10}$$

The matrix $\sigma_{\alpha\beta}$ is skew-symmetric and one has

$$\sigma_{12} = \omega_{12} - \omega_{03}, \quad \sigma_{13} = -2\omega_{23}, \quad \sigma_{23} = 2\omega_{01}, \tag{2.11}$$

where

$$\omega_{ij} = g_{ik}\omega^k{}_j. \tag{2.12}$$

Associated with $\sigma^\alpha{}_\beta$ is the 1-form valued vector

$$\sigma^\alpha = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma}\gamma_{\beta\delta}\sigma^\delta{}_\gamma, \tag{2.13}$$

where $\epsilon^{\alpha\beta\gamma}$ is the three-dimensional Levi-Civita symbol. Writing σ^α in terms of the basis, one gets

$$\sigma^\alpha = \sigma^\alpha{}_i\theta^i. \tag{2.14}$$

The quantities $\sigma^\alpha{}_a$ are related to the NP spin coefficients as follows:

$$\sigma^\alpha{}_i = 2\begin{pmatrix} -\nu & \lambda & \mu & -\pi \\ -\tau & \rho & \sigma & -\kappa \\ -\gamma & \alpha & \beta & -\epsilon \end{pmatrix}. \tag{2.15}$$

The curvature 2-form $\Theta^i{}_j$ is given in terms of the connection by Cartan's second structure equation

$$d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j = \Theta^i{}_j. \tag{2.16}$$

The tetrad components of the Riemann tensor are determined by the relation

$$\Theta^i{}_j = -\tfrac{1}{2}R^i{}_{jkl}\theta^k \wedge \theta^l. \tag{2.17}$$

The equations (referred to as the field equations) which correspond to (2.16) in the vectorial formalism are

$$d\sigma^\alpha{}_\beta + \sigma^\alpha{}_\gamma \wedge \sigma^\gamma{}_\beta = \Sigma^\alpha{}_\beta. \tag{2.18}$$

If one defines

$$\Sigma_{\alpha\beta} = -\Sigma_{\beta\alpha} = \gamma_{\alpha\gamma}\Sigma^\gamma{}_\beta, \tag{2.19}$$

then the relationship between the components of Σ and the components of Θ is given by

$$\Sigma_{12} = \Theta_{12} - \Theta_{03}, \quad \Sigma_{13} = -2\Theta_{23}, \quad \Sigma_{23} = 2\Theta_{01}. \tag{2.20}$$

Associated with $\Sigma_{\alpha\beta}$ is the 2-form valued vector

$$\Sigma_\alpha = \tfrac{1}{2}\gamma_{\alpha\beta}\epsilon^{\beta\gamma\delta}\Sigma_{\gamma\delta}. \tag{2.21}$$

If one expands Σ_α in the basis, one gets

$$\Sigma_\alpha = C_{\alpha\beta}Z^\beta - \tfrac{1}{6}R\gamma_{\alpha\beta}Z^\beta + E_{\alpha\bar{\beta}}\bar{Z}^\beta, \tag{2.22}$$

where $C_{\alpha\beta}$ is a complex valued trace-free symmetric tensor which corresponds to the Weyl tensor, $E_{\alpha\bar{\beta}}$ is a Hermitian tensor corresponding to the trace-free Ricci tensor, and R is the curvature scalar. The quantities $C_{\alpha\beta}$ and $E_{\alpha\bar{\beta}}$ are related to the NP quantities as follows:

$$C_{\alpha\beta} = 2\begin{pmatrix} -\Psi_0 & -\Psi_2 & 2\Psi_1 \\ -\Psi_2 & -\Psi_4 & 2\Psi_3 \\ 2\Psi_1 & 2\Psi_3 & -4\Psi_2 \end{pmatrix}, \tag{2.23}$$

$$E_{\alpha\bar{\beta}} = 2\begin{pmatrix} -\Phi_{00} & \Phi_{02} & \Phi_{01} \\ \Phi_{20} & -\Phi_{22} & \Phi_{21} \\ \Phi_{10} & -\Phi_{12} & -4\Phi_{11} \end{pmatrix}. \tag{2.24}$$

In the vectorial formalism the Bianchi identities have the form

$$d\Sigma^\alpha{}_\beta - \Sigma^\alpha{}_\gamma \wedge \sigma^\gamma{}_\beta + \sigma^\alpha{}_\gamma \wedge \Sigma^\gamma{}_\beta = 0. \tag{2.25}$$

The commutation relations arise as follows: If $A = A_i\theta^i$ is an exact 1-form, then it is closed, that is, $dA = 0$. Using the expression for A and (2.6), we get

$$(dA_i - A_j\omega^j{}_i) \wedge \theta^i = 0. \tag{2.26}$$

When this equation is written out in full, one gets the four complex commutation relations.

We now give the relations between the Pfaffian derivatives of the vectorial formalism defined by

$$B_{,i} = B_{,a}\, h_i{}^a \tag{2.27}$$

and the NP differential operators. Since $h_i{}^a = (n^a, -\bar{m}^a, -m^a, l^a)$, we have

$$\begin{aligned} &B_{,0} = B_{,a}n^a = \Delta B, \quad B_{,1} = -B_{,a}\bar{m}^a = -\bar{\delta}B, \\ &B_{,2} = -B_{,a}m^a = -\delta B, \quad B_{,3} = B_{,a}l^a = DB. \end{aligned} \tag{2.28}$$

The self-dual part of the electromagnetic field F_{ab} has the form

$$\overset{+}{F} = F_\alpha Z^\alpha. \tag{2.29}$$

In view of this the Einstein–Maxwell equations (1.1) and (1.2) may be written as follows:

$$E_{\alpha\bar{\beta}} = -2F_\alpha\bar{F}_\beta, \tag{2.30}$$

$$d\overset{+}{F} = 0. \tag{2.31}$$

We conclude our summary of the vectorial formalism

by remarking that Eqs. (2.18) correspond to the NP field equations while Eq. (2.6) is equivalent to the NP metric equations. Equations (2.6), (2.9), (2.18), (2.25), (2.26), and (2.31) are written out in full in the article of Debever.[10] In this paper we shall write most equations using the NP quantities as defined in Eqs. (2.15), (2.23), (2.24), and (2.28).

3. PROOF OF THEOREM 2

If we insert the canonical form for $\overset{+}{F}_{ab}$ into (1.6) and contract successively with $l^a m^b$ and $\bar{m}^a n^b$, we obtain the conditions

$$\kappa = \pi = 0. \tag{3.1}$$

By a similar procedure we obtain from (1.7) the conditions

$$\nu = \tau = 0. \tag{3.2}$$

The only further consequences of contractions on (1.6) and (1.7) are

$$f = \phi^{-1} D\phi, \tag{3.3}$$

$$g = \phi^{-1} \Delta\phi. \tag{3.4}$$

It is easy to verify that the conditions (3.1) and (3.2) are invariant[11] under the tetrad transformations (1.5). These transformations induce the following transformations for the spin coefficients ϵ and γ:

$$\epsilon' = e^a(\epsilon + \tfrac{1}{2} Dw), \tag{3.5}$$

$$\gamma' = e^{-a}(\gamma + \tfrac{1}{2}\Delta w), \tag{3.6}$$

where

$$w = a + ib. \tag{3.7}$$

In order to set $\epsilon' = \gamma' = 0$, we must have

$$Dw = -\tfrac{1}{2}\epsilon, \quad \Delta w = -\tfrac{1}{2}\gamma. \tag{3.8}$$

The integrability condition for (3.8), obtained by applying the commutator[12] $[D, \Delta]$, is

$$D\gamma - \Delta\epsilon = -(\epsilon + \bar{\epsilon})\gamma - (\gamma + \bar{\gamma})\epsilon. \tag{3.9}$$

In view of the field equations[13] (2.18) this condition is satisfied if and only if

$$\Psi_2 + \Phi_{11} = 0, \tag{3.10}$$

which is just condition IV in terms of the tetrad components. To complete the proof of Theorem 2, it is only necessary to remark that the conditions

$$\kappa = \pi = \epsilon = \nu = \tau = \gamma = 0 \tag{3.11}$$

are the necessary and sufficient conditions in order that the tetrad vectors $\{l_a, m_a, \bar{m}_a, n_a\}$ be parallelly propagated along both the null congruences defined by l^a and n^a.

In view of (3.5) and (3.6) it is clear that the conditions (3.11) are not invariant under a transformation (1.5) unless

$$Dw = \Delta w = 0. \tag{3.12}$$

This justifies our remark following the statement of Theorem 2.

4. CONSEQUENCES OF CONDITIONS I AND II

We choose as a null tetrad on V_4 a null tetrad defined by F_{ab} (see condition II) which is parallelly propagated along the null congruences defined by l^a and n^a. Thus the conditions (3.11) are satisfied. This is permitted by virtue of the conditions III and IV and Theorem 2. The self-dual 2-form associated with the electromagnetic field hence may be written as

$$\overset{+}{F} = \tfrac{1}{2} F_3(\theta^0 \wedge \theta^3 - \theta^1 \wedge \theta^2) = F_3 Z^3, \tag{4.1}$$

where

$$F_3 = \phi. \tag{4.2}$$

Einstein's equations (1.1) or (2.30) are now equivalent to the statement

$$E_{1\bar{1}} = E_{1\bar{2}} = E_{1\bar{3}} = E_{2\bar{2}} = E_{2\bar{3}} = 0, \quad E_{3\bar{3}} = -2F_3\bar{F}_3. \tag{4.3}$$

Maxwell's equations (1.2) or (2.31) become[14]

$$D\phi = 2\rho\phi, \quad \delta\phi = \bar{\delta}\phi = 0, \quad \Delta\phi = 2\mu\phi. \tag{4.4}$$

The conditions (3.11) and (4.3) together with the field equations (2.18) imply

$$\Psi_0 = \Psi_1 = 0. \tag{4.5}$$

in view of (3.11), (4.3), and (4.5) Bianchi's identities[15] (2.25) become

$$D\Phi_{11} = 2(\rho + \bar{\rho})\Phi_{11}, \quad \delta\Phi_{11} = \bar{\delta}\Phi_{11} = 0,$$

$$\Delta\Phi_{11} = -2(\mu + \bar{\mu})\Phi_{11}, \tag{4.6a}$$

$$\bar{\delta}\Psi_0 = 4\alpha\Psi_0, \quad \Delta\Psi_0 = -\mu\Psi_0 + \sigma\Psi_2, \tag{4.6b}$$

$$D\Psi_2 = \rho\Psi_2 - \lambda\Psi_0, \quad \delta\Psi_2 = \bar{\delta}\Psi_2 = 0, \quad \Delta\Psi_2 = -\mu\Psi_2 + \sigma\Psi_4, \tag{4.6c}$$

$$\delta\Psi_4 = -4\beta\Psi_4, \quad D\Psi_4 = -\lambda\Psi_2 + \rho\Psi_4. \tag{4.6d}$$

The commutator $[\delta, \bar{\delta}]$ applied to (4.6a) yields

$$\rho\bar{\mu} = \bar{\rho}\mu. \tag{4.7}$$

This is the so-called "coupling theorem" of Debney and Zund.[16] It is obtained here as a consequence of Einstein's equations whereas Debney and Zund derived it from Maxwell's equations.

With the help of Eqs. (3.10) and (4.6a, b) we obtain the relation

$$\sigma\Psi_4 = -(\mu + 2\bar{\mu})\Psi_2. \tag{4.8}$$

If we operate on both sides of the above with D and use (4.6c, d) along with the field equations (2.18), we obtain the further relation

$$(4\bar{\rho}\mu + 2\bar{\rho}\bar{\mu} + 2\bar{\sigma}\bar{\lambda} + 3\Psi_2)\lambda\sigma + (\rho + 2\bar{\rho})(\mu + 2\bar{\mu})\Psi_2 = 0. \tag{4.9}$$

We now invoke condition V, which may be written as

$$\rho + \bar{\rho} = 0. \tag{4.10}$$

In view of (4.7) we also have

$$\mu + \bar{\mu} = 0. \tag{4.11}$$

If we subtract from (4.9) its complex conjugate and use the fact that Ψ_2 is real, there results

$$(\bar{\sigma}\bar{\lambda} - \sigma\lambda)(2\rho\mu - 3\Psi_2) = 0. \tag{4.12}$$

Since $2\rho\mu + 3\Psi_2 = 0$ leads to a contradiction, the above

equation implies

$$\sigma\lambda = \bar{\sigma}\bar{\lambda}. \tag{4.13}$$

We shall now show that the combination $\alpha + \bar{\beta}$ may be made zero by a tetrad transformation (1.5) which preserves the conditions $\epsilon = \gamma = 0$. We have $\alpha' + \bar{\beta}' = \exp(-i\bar{b})(\alpha + \bar{\beta} + \bar{\delta}a)$. Thus $\alpha' + \bar{\beta}'$ may be annulled if we are able to set

$$\bar{\delta}a = -\alpha - \bar{\beta}. \tag{4.14}$$

The commutator $[\delta, \bar{\delta}]$ applied to (4.14) yields, on account of (2.18) and (3.12), the integrability condition

$$\overline{\rho\mu} - \rho\mu + \sigma\lambda - \bar{\sigma}\bar{\lambda} = 0, \tag{4.15}$$

which is satisfied on account of (4.10), (4.11), and (4.13). Thus, dropping the primes, we have

$$\alpha + \bar{\beta} = 0. \tag{4.16}$$

We now proceed to the determination of the spin coefficients ρ, μ, σ, and λ. On taking account of (3.11) the relevant field equations are

$$D\rho = \rho^2 + \sigma\bar{\sigma}, \quad \Delta\rho = -\rho\bar{\mu} - \sigma\lambda - \Psi_2, \tag{4.17a, b}$$

$$D\mu = \bar{\rho}\mu + \sigma\lambda + \Psi_2, \quad \Delta\mu = -\mu^2 - \lambda\bar{\lambda}. \tag{4.17c, d}$$

It follows easily from (4.10), (4.11), (4.13), and (4.17) that

$$D\rho = \Delta\rho = D\mu = \Delta\mu = 0, \tag{4.18}$$

$$\rho^2 + \sigma\bar{\sigma} = 0, \tag{4.19}$$

$$\mu^2 + \lambda\bar{\lambda} = 0, \tag{4.20}$$

$$\Psi_2 = \rho\mu - \sigma\lambda. \tag{4.21}$$

To complete our determination of ρ and μ, we appeal to Maxwell's equations. By applying the commutators to (4.4) we find on account of (4.16)

$$\delta\rho = \delta\mu = 0. \tag{4.22}$$

Equations (4.18) and (4.22) imply that

$$\rho = ik, \quad \mu = il, \tag{4.23}$$

where k and l are real constants. It follows from (4.13) that $\sigma = |\sigma| \exp(i\theta)$ and $\lambda = |\lambda| \exp(-i\theta)$. Equations (4.19) and (4.20) hence imply that $|\sigma| = |k|$ and $|\lambda| = |l|$. Thus

$$\sigma = |k| \exp(i\theta), \tag{4.24}$$

$$\lambda = |l| \exp(-i\theta). \tag{4.25}$$

On account of (4.23), (4.24), and (4.25) Eq. (4.21) becomes

$$|kl| + kl = -\Psi_2. \tag{4.26}$$

We note by (3.10) and (4.3) that $\Psi_2 < 0$ if the electromagnetic field is not too vanish. Thus (4.26) implies that

$$kl > 0. \tag{4.27}$$

We may thus set $k = l$ by a tetrad transformation. The induced transformations for ρ and μ are $\rho' = \exp(a)\rho$ and $\mu' = \exp(-a)\mu$. In order to have $\rho' = \mu'$, we must set

$$\exp(2a) = \mu/\rho = l/k. \tag{4.28}$$

This is possible since k and l have the sign because of (4.27). Thus dropping the primes we have

$$\rho = \mu = ik, \tag{4.29}$$

$$\sigma = k\exp(i\theta), \quad \lambda = k\exp(-i\theta). \tag{4.30}$$

The remaining tetrad freedom is $\theta^{0'} = \theta^0$, $\theta^{1'} = \exp(i\bar{b})\,\theta^1$, $\theta^{3'} = \theta^3$. We use this freedom to make $\sigma = \lambda$ real. The induced transformations for σ and λ are $\sigma' = \exp(2i\bar{b})\,\sigma$ and $\lambda' = \exp(-2i\bar{b})\,\lambda$. Thus if we choose $2\bar{b} = -\theta$, we have $\sigma' = \lambda' = k$. Dropping the primes gives

$$\sigma = \lambda = k. \tag{4.31}$$

The above transformation preserves the values of ρ and μ given by (4.29). If we now substitute the values of ρ, μ, σ, and λ given by (4.29) and (4.31) into the field equations, taking account as always of (4.3), we find that the spin coefficients have the following values:

$$\kappa = \pi = \nu = \tau = \alpha = \beta = 0$$
$$\rho = \mu = ik, \quad \sigma = \lambda = k, \quad \epsilon = \gamma = (i/2)\,k. \tag{4.32}$$

The nonzero components of the curvature are

$$\Psi_0 = -2ik^2, \quad -\Psi_2 = \Phi_{11} = 2k^2, \quad \Psi_4 = -2ik^2. \tag{4.33}$$

5. THE DETERMINATION OF THE METRIC

The values of the spin coefficients given in (4.32) allow the determination of the metric of V_4 satisfying conditions I–V in a system of canonical coordinates. In view of (4.32) Cartan's first structure equation (2.6) yields the following:

$$d\theta^0 = 2ik\,\theta^1 \wedge \theta^2, \tag{5.1a}$$

$$d\theta^1 = k(\theta^2 \wedge \theta^3 - i\theta^0 \wedge \theta^1 + i\theta^1 \wedge \theta^3 + \theta^0 \wedge \theta^3), \tag{5.1b}$$

$$d\theta^3 = 2ik\,\theta^1 \wedge \theta^2. \tag{5.1c}$$

We obtain from (5.1b) the equation

$$d\theta^1 \wedge \theta^1 \wedge \theta^2 = 0. \tag{5.2}$$

Thus by virtue of the Frobenius integration theorem[18] there exist a complex coordinate z and two complex valued functions C and D of 4-variables such that

$$\theta^1 = C\,dz + D\,d\bar{z}. \tag{5.3}$$

In order that the metric be nondegenerate,

$$\Delta \equiv C\bar{C} - D\bar{D} \neq 0. \tag{5.4}$$

We now let u and v denote the two remaining real coordinates. The 1-forms θ^0 and θ^3 may thus be written as

$$\theta^0 = a\,du + B\,du + \bar{B}\,dz + B\,d\bar{z} + e\,dv, \tag{5.5}$$

$$\theta^3 = f\,du + G\,dz + \bar{G}\,d\bar{z} + h\,dv, \tag{5.6}$$

where a, e, f, and h denote real valued functions and B and G complex valued functions of u, z, $\bar{z}$, and v satisfying the inequality[19]

$$ah - ef \neq 0. \tag{5.7}$$

It follows from (5.1a, c) that

$$d\theta^0 \wedge \theta^1 = d\theta^3 \wedge \theta^1 = 0. \tag{5.8}$$

These conditions imply, by virtue of (5.3), (5.5), and (5.6), that

$$a_v = e_u, \tag{5.9}$$

$$f_v = h_u. \tag{5.10}$$

The above equations permit us to make a change of coordinates so that θ^0 and θ^3 can be written in the form

$$\theta^0 = du + A\,dz + \bar{A}\,d\bar{z}, \tag{5.11}$$

$$\theta^3 = dv + B\,dz + \bar{B}\,d\bar{z}, \tag{5.12}$$

where A and B are complex valued functions of the four coordinates. If one substitutes for θ^0, θ^1 and θ^3 from (5.3), (5.11), and (5.12) into (5.8) there results

$$DA_u - C\bar{A}_u = DA_v - C\bar{A}_v = 0, \tag{5.13}$$

$$DB_u - C\bar{B}_u = DB_v - C\bar{B}_v = 0. \tag{5.14}$$

These equations imply, on account of (5.4), that

$$A_u = A_v = B_u = B_v = 0. \tag{5.15}$$

In other words A and B are functions of z and $\bar{z}$ only.

The most general coordinate transformation which preserves the form of the θ^i given by (5.3), (5.11), and (5.12) is

$$u = u' + a(z'\bar{z}'), \quad z = Z(z', \bar{z}'), \quad v = v' + b(z', \bar{z}'). \tag{5.16a, b, c}$$

The transformation laws for A, B, C, and D are

$$A' = a_{z'} + Az_{z'} + \bar{A}\bar{z}_{z'}, \tag{5.17a}$$

$$B' = b_{z'} + Bz_{z'} + \bar{B}\bar{z}_{z'}, \tag{5.17b}$$

$$C' = Cz_{z'} + D\bar{z}_{z'}, \tag{5.17c}$$

$$D' = Cz_{z'} + D\bar{z}_{\bar{z}'}. \tag{5.17d}$$

It follows that the conditions (5.15) are preserved by the transformation (5.16).

We now compute the nonzero spin coefficients arising from the tetrad θ^i given in (5.3), (5.11), and (5.12) and which satisfies (5.15). This is effected by calculating in the natural basis the exterior derivatives of the Z^α defined in (2.4) and comparing with the formula[20] (2.9). The results are the following:

$$\mu - 2\gamma = \Delta^{-1}(C\bar{C}_u - D\bar{D}_u), \tag{5.18}$$

$$\lambda = \Delta^{-1}(\bar{C}\bar{D}_u - \bar{D}\bar{C}_u), \tag{5.19}$$

$$\rho - 2\epsilon = \Delta^{-1}(\bar{D}D_v - \bar{C}C_v), \tag{5.20}$$

$$\sigma = \Delta^{-1}(DC_v - CD_v), \tag{5.21}$$

$$\alpha = -\bar{\beta} = \tfrac{1}{2}\Delta^{-1}(\bar{C}_z - \bar{D}_{\bar{z}} - B\bar{C}_v + \bar{B}\bar{D}_v - A\bar{C}_u + \bar{A}\bar{D}_u), \tag{5.22}$$

$$\mu = \tfrac{1}{2}\Delta^{-1}(\bar{B}_z - B_{\bar{z}} + \bar{C}C_u - \bar{D}D_u + C\bar{C}_u - D\bar{D}_u), \tag{5.23}$$

$$\rho = \tfrac{1}{2}\Delta^{-1}(\bar{A}_z - A_{\bar{z}} - \bar{C}C_v + \bar{D}D_v - C\bar{C}_v + D\bar{D}_v). \tag{5.24}$$

In order to determine the metric functions, we compare the values of the spin coefficients given above with those in (4.32). The resulting equations are

$$\bar{C}C_u - \bar{D}D_u = 0, \tag{5.25}$$

$$DC_u - CD_u = -k\Delta, \tag{5.26}$$

$$\bar{C}C_v - \bar{D}D_v = 0, \tag{5.27}$$

$$DC_v - CD_v = k\Delta, \tag{5.28}$$

$$A_{\bar{z}} - \bar{A}_z = -2ik\Delta, \tag{5.29}$$

$$B_{\bar{z}} - \bar{B}_z = -2ik\Delta, \tag{5.30}$$

$$D_z - C_{\bar{z}} + \bar{A}C_u - AD_u + \bar{B}C_v - BD_v = 0. \tag{5.31}$$

The first step in the integration of these equations is to use some of the remaining coordinate freedom to set $A = B$. From (5.29) and (5.30) we have

$$(A - B)_{\bar{z}} = (\bar{A} - \bar{B})_z. \tag{5.32}$$

It follows that there exists a real-valued function r of z and $\bar{z}$ such that

$$B = A + r_z. \tag{5.33}$$

From (5.17a, b) we have on account of (5.33)

$$B' - A' = (b - a)_{z'} + (B - A)z_{z'} + (\bar{B} - \bar{A})\bar{z}_{\bar{z}'} = (b - a + r)_{z'}. \tag{5.34}$$

Thus $B' = A'$ if we choose $a - b = r$. By (5.34) the relation

$$B = A \tag{5.35}$$

is preserved by all transformations (5.16) with $b = a + b_0$, where b_0 is constant:

$$u = u' + a(z', \bar{z}'), \quad z = Z(z', \bar{z}'), \quad v = v' + a(z', \bar{z}') + b_0. \tag{5.36}$$

We next integrate Eqs. (5.25), (5.26), (5.27), and (5.28) to determine the u and v dependence of C and D. To achieve this, we solve the first two equations for C_u and D_u and the last two for C_v and D_v obtaining

$$C_u = k\bar{D}, \quad D_u = k\bar{C}, \tag{5.37}$$

$$C_v = k\bar{D}, \quad D_v = -k\bar{C}. \tag{5.38}$$

The system (5.37) has the general solution

$$C = C_1 \exp(ku) + C_2 \exp(-ku), \tag{5.39}$$

$$D = \bar{C}_1 \exp(ku) - \bar{C}_2 \exp(-ku), \tag{5.40}$$

where C_1 and C_2 are complex valued function of z, $\bar{z}$, and v. From (5.38) we obtain the v dependence of C_1 and C_2 which is

$$C_1 = P \exp(-kv), \quad C_2 = Q \exp(kv), \tag{5.41}$$

where P and Q are complex valued functions of z and $\bar{z}$. Hence C and D have the following form:

$$C = P \exp[k(u - v)] + Q \exp[-k(u - v)], \tag{5.42}$$

$$D = \bar{P} \exp[k(u - v)] - \bar{Q} \exp[-k(u - v)]. \tag{5.43}$$

When these expressions for C and D are substituted into (5.31) and (5.35) is taken into account, we obtain

$$P_{\bar{z}} = \bar{P}_z, \quad Q_{\bar{z}} = -\bar{Q}_z. \tag{5.44}$$

These equations imply the existence of two real valued functions m and n of z and $\bar{z}$ such that

$$P = m_z, \quad Q = in_z. \tag{5.45}$$

From (5.17c, d) together with (5.42) and (5.43) we find the following transformation laws:

$$P' = \exp(-kb_0)(Pz_{z'} + \bar{P}\bar{z}_{z'}), \tag{5.46a}$$

$$Q' = \exp(kb_0)(Qz_{z'} - \bar{Q}\bar{z}_{z'}). \tag{5.46b}$$

By combining these we obtain $P' = \exp(-kb_0)\, m_{z'}$ and $Q' = i \exp(kb_0)\, n_{z'}$. Thus if we choose $z' = \frac{1}{2}[\exp(-kb_0)\, m + i \exp(kb_0)\, n]$, we obtain

$$P' = Q' = 1. \tag{5.47}$$

The conditions (5.47) are preserved by the coordinate transformations

$$u=u'+a(z',\bar{z}'),\quad v=v'+a(z',\bar{z}')+b_0,$$
$$z=z'\cosh(kb_0)+\bar{z}'\sinh(kb_0)+z_0, \tag{5.48}$$

where z_0 is a complex constant. On account of (5.42), (5.43), and (5.47) the final form of C and D is, on dropping primes,

$$C=\exp[k(u-v)]+\exp[-k(u-v)], \tag{5.49}$$

$$D=\exp[k(u-v)]-\exp[k(u-v)]. \tag{5.50}$$

It remains to solve Eq. (5.29) which now becomes

$$A_{\bar{z}}-\bar{A}_z=-8ik. \tag{5.51}$$

This equation has the general solution

$$A=-4ik(z+\bar{z})+t_z, \tag{5.52}$$

where t is a real valued function of z and $\bar{z}$. The transformation law for A under the remaining coordinate freedom (5.48) is, on account of (5.17a),

$$A'=-4ik(z'+\bar{z}')-4ik(z_0+\bar{z}_0)\exp(-kb_0)+(a+t)_{z'}. \tag{5.53}$$

Hence, if we choose $t=-a$ and $z_0=0$, we obtain after dropping the primes

$$A=-4ik(z+\bar{z}). \tag{5.54}$$

The group of coordinate transformations which preserves (5.54) is given by

$$\begin{aligned}u&=u'+4ik\exp(-kb_0)(z_0+\bar{z}_0)(z'-\bar{z}')+a_0,\\ z&=z'\cosh(kb_0)+\bar{z}'\sinh(kb_0)+z_0,\\ v&=v'+4ik\exp(-kb_0)(z_0+\bar{z}_0)(z'-\bar{z}')+a_0+b_0,\end{aligned} \tag{5.55}$$

where a_0 and b_0 are arbitrary real constants and z_0 is a complex constant.

We summarize our results by giving the expression for θ^i and the metric of V_4 satisfying conditions I–V:

$$\theta^0=du-4ik(z+\bar{z})(dz-d\bar{z}),\quad \theta^3=dv-4ik(z+\bar{z})(dz-d\bar{z}), \tag{5.56a, b}$$

$$\theta^1=2\cosh[k(u-v)]\,dz+2\sinh[k(u-v)]\,d\bar{z}, \tag{5.56c}$$

$$\begin{aligned}ds^2=&\,2\,du\,dv-8ik(z+\bar{z})(du+dv)(dz-d\bar{z})\\ &-4(4k^2(z+\bar{z})^2+\sinh[2k(u-v)]\,(dz^2+d\bar{z}^2)\\ &+8\{8k^2(z+\bar{z})^2-\cosh[2k(u-v)]\}dz\,d\bar{z}.\end{aligned} \tag{5.57}$$

Since the θ^i are invariant under the transformation (5.55), the same is true for the metric (5.57). Thus the space–time admits at least a four-parameter group of motions. In Sec. 6 we shall see that the maximal group of motions is, in fact, a G_4. It should be remarked at this point that all the second order differential invariants of the metric are functions of k^2. Thus $|k|$ is an essential constant since it cannot be eliminated from the metric by a coordinate transformation. However, the sign of k is irrelevant.

In order to obtain the form of the metric given in (1.10), we make the following coordinate transformation and replace $2\sqrt{2}k$ by k:

$$u=-[(1/\sqrt{2})u'+\sqrt{2}\,v'],\quad z=(1/2\sqrt{2})(x+iy),$$
$$v=-(1/\sqrt{2})u'. \tag{5.58}$$

The 1-forms θ^i in these coordinates are after suppressing the primes

$$\theta^0=-(1/\sqrt{2})\,du-\sqrt{2}\,dv+2kx\,dy, \tag{5.59a}$$

$$\theta^1=(1/\sqrt{2})[\exp(-\sqrt{2}kv)\,dx+i\exp(\sqrt{2}kv)\,dy]. \tag{5.59b}$$

$$\theta^3=-(1/\sqrt{2})\,du+(1/\sqrt{2})\,kx\,dy, \tag{5.59c}$$

To complete the proof of Theorem 1, Maxwell's equations (4.4) must be integrated. In view of (4.32) they become

$$\phi_v=2ik\phi,\quad \phi_x=\phi_{\bar{z}}=0,\quad \phi_u=-2ik\phi. \tag{5.60}$$

By virtue of (4.3) and (4.33) the solution is

$$\phi=2k\exp[-2ik(u-v)]. \tag{5.61}$$

The 2-form given in (1.11) is obtained when one takes into consideration (4.1), (5.56), (5.60), and the transformation (5.58).

6. PROPERTIES OF THE SOLUTION

We shall first discuss the symmetry properties of the space–time with metric (1.10). The integration of Killing's equations yields the following four Killing vectors:

$$X_1=\xi_1{}^a\frac{\partial}{\partial x^a}=\frac{\partial}{\partial u}, \tag{6.1a}$$

$$X_2=\xi_2{}^a\frac{\partial}{\partial x^a}=\frac{\partial}{\partial y}, \tag{6.1b}$$

$$X_3=\xi_3{}^a\frac{\partial}{\partial x^a}=ky\frac{\partial}{\partial u}+\frac{\partial}{\partial x}, \tag{6.1c}$$

$$X_4=\xi_4{}^a\frac{\partial}{\partial x^a}=-\frac{\partial}{\partial u}+\frac{\partial}{\partial v}+\tfrac{1}{2}kx\frac{\partial}{\partial x}-\tfrac{1}{2}ky\frac{\partial}{\partial y}. \tag{6.1d}$$

Thus the maximal group of motions is a G_4. It is also easy to verify, by integrating the equations for the 1-parameter subgroups, that one recovers the transformations (5.55).

From (6.1) we find the commutation relations

$$\begin{aligned}&[X_1,X_2]=[X_1,X_2]=[X_1,X_3]=0,\\ &[X_2,X_3]=kX_1,\quad [X_2,X_4]=-\tfrac{1}{2}kX_2,\quad [X_3,X_4]=\tfrac{1}{2}kX_3.\end{aligned} \tag{6.2}$$

These relations imply that the G_4 is of Bianchi[21] type I. Further since $\det(\xi_i{}^a)=-1\neq 0$, G_4 is a simply transitive group[22] of motions on V_4. We also note that the solution is stationary because X_1 is timelike.

The contravariant tetrad vectors $h_i{}^a$ defined by (5.59) form a set of invariant vectors[23] of the group since it can be shown that they satisfy the equations

$$\xi_i{}^b h_j{}^a{}_{,b}-h_j{}^b\xi_i{}^a{}_{,b}=0, \tag{6.3}$$

$$\det(h_i{}^a)\neq 0. \tag{6.4}$$

This elucidates the previously discovered fact[24] that all the spin coefficients associated with this tetrad are constant.

The self-dual electromagnetic field bivector $\overset{+}{F}{}^{ab}$ may

be expressed in terms of the tetrad as

$$\overset{+}{F}{}^{ab} = (1/\sqrt{2})\, k \exp(ikv)(h_3{}^{[a} h_0{}^{b]} - h_2{}^{[a} h_1{}^{b]}). \tag{6.5}$$

It is easily verified using (6.1) and (6.3) that

$$\mathcal{L}_{\xi_i} \overset{+}{F}{}^{ab} = 0 \quad (i = 1, 2, 3), \quad \mathcal{L}_{\xi_4} \overset{+}{F}{}^{ab} = ik \overset{+}{F}{}^{ab}. \tag{6.6}$$

Thus the electromagnetic field is not invariant[25] under the full group of motions G_4 admitted by the space–time.

From the property (6.6) we conclude that the class of solutions presented do not belong to Ozsváth's[26] class of homogeneous solutions of the Einstein–Maxwell equations, since he assumes that both the metric of space–time and the electromagnetic field are invariant under the transformations of a four-parameter simply transitive group.

We conclude the discussion of properties of the space–time by determining the Petrov type and by giving a set of principal null vectors of the Weyl tensor. On account of (4.33) the biquadratic equation[27]

$$\Psi_0 \lambda^4 + 6\Psi_2 \lambda^2 \mu^2 + \Psi_4 \mu^2 = 0 \tag{6.7}$$

has four distinct roots. Hence the Weyl tensor is of Petrov type I.

The tetrad components of the principal null vectors are

$$l_a{}^i = \begin{pmatrix} \alpha & \exp(-3i\pi/4) & \exp(3i\pi/4) & \alpha^{-1} \\ \alpha & \exp(i\pi/4) & \exp(-i\pi/4) & \alpha^{-1} \\ \alpha^{-1} & \exp(-i\pi/4) & \exp(i\pi/4) & \alpha \\ \alpha^{-1} & \exp(3i\pi/4) & \exp(-3i\pi/4) & \alpha \end{pmatrix} (i = 1, 2, 3, 4) \tag{6.8}$$

where $\alpha = (\sqrt{10} + 3)^{1/2}$. It is clear from (6.8) that the principal null vectors of the electromagnetic field, l_a and n_a, are not principal null vectors of Weyl tensor. However, they possess an interesting geometrical property in that they define the axis of the involution[28] which permutes the principal null vectors l^1 and l^2 and l^3 and l^4. This property may be better understood if the quasi-orthonormal tetrad[29] (t, x, y, z) is introduced. Null directions can then be represented as points on a sphere which lies in the three-dimensional space orthogonal to the timelike vector t. This sphere may be considered as the field of vision of an observer whose 4-velocity is t. The principal null vectors l^i thus yield four points on the sphere, $A = (\beta, -1, 1)$, $B = (\beta, 1, -1)$, $C = (\beta, 1, 1)$, $D = (-\beta, -1, -1)$, where $\beta = (1/\sqrt{2})(\alpha - \alpha^{-1})$. These points form the vertices of a tetrahedron with opposite edges equal in pairs and in addition with the edges AC and BC equal. The points corresponding to the null directions defined by l_a and n_a are $E = ((1 + \beta^2/2)^{1/2}, 0, 0)$ and $F = (-(1 + \beta^2/2)^{1/2}, 0, 0)$. These points lie on the line joining the midpoints of the edges AB and CD. A rotation through π about this line transforms the tetrahedron into itself and hence is the axis of the involution mentioned above. We also note that the y axis joins the midpoints of the edges AD and BC and that the z axis joins the midpoints of the edges AC and BD. Hence the y axis is the axis of the involution which interchanges the vertices A and D and B and C. Similarly the z axis is the axis of the involution which interchanges the vertices A and C and B and D. Thus the t, x, y, and z axes determine the Riemann principal directions[30] of the Weyl tensor.

Note added in proof: Dr. B. O. J. Tupper kindly informed us that he has found essentially the same solution; his result is to appear in Gen. Relativ. Gravit.

*This work was supported in part by a grant from the National Research Council of Canada.

[1]The conventions of our paper are the same as those of R. Debever, Cah. Phys. **168–169**, 303 (1964).

[2]l_a and n_a are future pointing real null vectors and m_a is a complex vector whose only nonzero inner products are $l_a n_a = -m_a m^a = 1$.

[3]$*F_{ab} = \frac{1}{2}\epsilon_{abcd} F^{cd}$, where $\epsilon_{abcd} = \sqrt{-g}\, e_{abcd}$ is the Levi-Civita density and e_{abcd} is the permutation symbol.

[4]N. Tariq and B. O. J. Tupper, Gen. Relativ. Gravit. **6**, 777 (1975).

[5]This assumption has also been employed by G. C. Debney and J. D. Zund, Tensor, N.S., **25**, 53 (1972).

[6]R. Debever, Cah. Phys. **168–169**, 303 (1964).

[7]M. Cahen, R. Debever, and L. Defrise, J. Math. Mech. **16**, 761 (1967).

[8]In a recent paper R. Debever, Bull. Cl. Sci. Acad. Roy. Belg. **60**, 998 (1974), has redefined his formalism in order to give a simpler correspondance with the NP formalism.

[9]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[10]Ref. 6.

[11]Ref. 6, Eq. (19.19).

[12]Ref. 6, Eq. (A.2).

[13]Ref. 6, Eq. (A.8).

[14]Ref. 6, Eq. (20.24).

[15]Ref. 6, Eqs. (A.16)–(A.27)

[16]Ref. 5.

[17]Ref. 6, Eqs (A.6), (A.7), and (A.8).

[18]See, for example, H. Flanders, *Differential with Applications to the Physical Sciences* (Academic, New York, 1963), p. 97.

[19]This follows from the fact that the 1-forms θ^i are linearly independent.

[20]See also Ref. 6, Eq. (A.3).

[21]See A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1964), p. 63. To obtain the exact correspondence with the commutation relations given by Petrov it is necessary to make the correspondence $X_1 \to 2X_1$, $X_2 \to \sqrt{2}k^{-1/2}X_2$, $X_3 \to \sqrt{2}k^{-1/2}X_3$, $X_4 \to \sqrt{2}k^{-1/2}X_4$.

[22]L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961), p. 72.

[23]I. Ozsváth, J. Math. Phys. **6**, 1255 (1965).

[24]See Eq. (4.32).

[25]This result has been proved for the special case $k = 4$ by H. Michalski and J. Wainwright, Gen. Relativ. Gravit. (to appear).

[26]Ref. 23.

[27]Ref. 6, Eq. (18.4).

[28]Defined by $L\theta^0 = \theta^0$, $L\theta^1 = -\theta^1$, $L\theta^3 = \theta^3$. See also Ref. 6, Sec. 19.

[29]Defined by $\sqrt{2}\, t = \theta^0 + \theta^3$, $\sqrt{2}\, x = \theta^0 - \theta^3$, $\sqrt{2}\, y = \theta^1 + \theta^2$, $i\sqrt{2}\, z = \theta^1 - \theta^2$.

[30]R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960).

Macroscopic and microscopic, coherent and incoherent variables

H. Ekstein

UER Marseille-Luminy and Centre de Physique Théorique, C.N.R.S., Marseille, France
(Received 27 May 1975; revised manuscript received 23 July 1975)

Quantitative figures of merit for microscopicity (resolving power) and for coherence (the necessity of simultaneous observations of large volumes) are defined operationally. Such indices are not assignable to observables (which are equivalence classes of observation procedures) but to observation procedures. As an application, the problem of reversibility is reconsidered. Known qualitative arguments explaining the difficulty of creating certain processes are made quantitative. A conjectured theorem states that the difficulty of preparing a state at $t = 0$ so that a particular target situation is observed at $t = \tau$ increases monotonically with τ.

1. INTRODUCTION

Some of the most commonly used terms in physics, such as macroscopic or microscopic variables or quantities, and coherent and incoherent states or variables, have not been defined in a precise and general way. Illustrative examples are in abundance, but one does not have a quantitative index of microscopicity or of coherence. It is not enough to distinguish between objects having or not having these properties, because there is a gradual transition between small and large.

Pressure and temperature of a gas are macroscopic; momenta and positions of individual molecules are commonly considered as microscopic quantities. Incoherent beams add intensities when combined; coherent ones may combine to have larger or smaller intensities. If these words were used only as a summary characterization of some phenomena understood quantitatively, precision would be an unnecessary luxury. When experience suggests some general regularities in connection with these loosely defined concepts, and when theory is challenged to produce the corresponding predictions, precision and generality become imperative. For instance, if macroscopic measuring instruments are said (by *N.* Bohr) to have some special properties not shared by smaller systems, and when theory attempts to formulate this statement precisely, a quantitative measure of microscopicity becomes imperative. If one tries to understand why it is difficult to produce a convergent coherent spherical wavefunction, qualitative descriptive words are in abundance. However, theory ought to give a precise meaning to such words as "controlling phases at distant points" and to derive from prime principles theorems that embody and imply the empirical circumstances.

Associating a well-defined mathematical object with a nonmathematical procedure appears often deceptively simple. The oldest fallacy of this kind is the association of an integer multiple of a small positive number to the physical concept of distance. The Pythagoreans proved conclusively that the vertices of some triangles had no distance under this definition. Modern physics shows many cases where a loosely defined physical concept was associated with a well-defined but ill-chosen mathematical object. Von Neumann's quantum mechanical ergodic theorem, for instance, had, on closer examination, little relevance to the intuitively and somewhat loosely perceived ergodic experience of physicists.[1]

To avoid such pitfalls, it seemed prudent to approach the problem in a somewhat unfamiliar manner. Usually one seeks the mathematical counterpart of a nonmathematical object or experience in an established mathematical framework. A well-known example is the association of particles with irreducible representation spaces of the Poincaré group in the Hilbert space of quantum mechanics, discovered by Wigner. The validation of such an association lies in the experimental verification of the mathematical predictions implied by it; in this case, e.g., the integer or half-integer spin values of particles.

Unfortunately, in the matter of concern to us, mathematical deductions are not readily available, and it seems more prudent to begin by a precise operational statement of what is suggested by experience.

As a model of this approach, one may think of Max Planck's[2] reconstruction of thermodynamics. The first and second principles are stated first in an operational form: "It is impossible to construct a machine which ... ." The mathematical theory is then constructed so that its physical interpretation includes the principles. In our case, adequate theories, classical and quantum mechanical, exist, and the task consists of two steps: the precise operational definition of macroscopicity and coherence, and the mapping of these more or less nonmathematical objects into the theory or theories.

It is a somewhat unexpected result of this search that an index of microscopicity or coherence cannot be attributed to an *observable* but rather to an *observation procedure*. By the first, we mean a function on phase space for classical, and a self-adjoint operator for quantum mechanics. The algebra of observation procedures, on the other hand, is a theoretical structure more directly related to laboratory hardware.

As an application, the old problem of (apparent) irreversibility is considered. Why is it not possible to arrange the state of a uniformly distributed gas so that it will, in an hour, be all concentrated in the left upper corner of the vessel? The qualitative answer: "because the necessary preparation procedure is too microscopic" is made quantitative. A conjectured theorem is formulated which asserts the increase of difficulty in preparing

a state at $t=0$ as the time $t=\tau$ of the target situation is increased.

2. OPERATIONAL MICROSCOPICITY

An observation procedure—which, by a slight abuse of language, we identify with an instruction or a computer program for performing it—has two types of instructions: the blueprint for the construction of the hardware and instructions concerning the location and time at which marks are to be positioned, switches to be thrown or levers to be actuated. By "time" we mean the time elapsed after a starting signal for the experiment, not calendar time.

An element α of the collection O of nonmathematical observation procedures is a pair $(b, \{P_n\})$, consisting of blueprints b and an ordered set of space–time points $\{P_n\}$, e.g., points on the world line of a mark on a telescope. A transformation

$$g: V_4 \to V_4, \quad (V_4 \ni P_n \mapsto gP_n \in V_4) \tag{2.1}$$

of space–time (the affine four-dimensional space V_4) induces a change of instructions through

$$(b, \{P_n\}) \vdash (b, \{gP_n\}). \tag{2.2}$$

The particular transformations considered here are space translations

$$g_{\mathbf{a}}: (\mathbf{x}, t) \vdash (\mathbf{x}+\mathbf{a}, t). \tag{2.3}$$

It is assumed that such a change in the instructions creates a feasible other procedure, i.e., that g induces a permutation $V_{\mathbf{a}}$

$$\begin{aligned} &V_{\mathbf{a}}: O \to O, \\ &[(b, \{P_n\}) \vdash (b, \{g_{\mathbf{a}} P_n\})]. \end{aligned} \tag{2.4}$$

A similar collection of nonmathematical procedures is the collection S of state-preparing procedures. If objects prepared by a procedure $s \in S$ are repeatedly made to be observed by a procedure $\alpha \in O$, the mean value of numerical outcomes $s_n(\alpha)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_1^N s_n(\alpha) \equiv \mathcal{E}(s, \alpha) \tag{2.5}$$

is the expectation value $\mathcal{E}$. The expectation is a map

$$\mathcal{E}: S \times O \to \mathbb{R} \tag{2.6}$$

of the Cartesian product of S and O into the reals. It depends on the external field in which the experiment is performed. We consider the complete set $\{\mathcal{E}\}$ of such expectations.

The particular property of observation procedures that we want to define is a generalization of the resolving power of a microscope. The resolving power is high if a large gradient of luminosity of the object causes rapid variations of the light signal received as the instrument is translated. Let $g_{\mathbf{a}}$ be a space translation by the 3-vector $\mathbf{a}$. Then the translation $\mathbf{a}$ induces the permutation

$$\alpha \vdash V_{\mathbf{a}} \alpha \in O.$$

This translation induces a change of dial readings or printouts of α:

$$\mathcal{E}(s, \alpha) \vdash \mathcal{E}(s, V_{\mathbf{a}} \alpha).$$

A procedure has a high resolving power if there exists a state-preparing procedure s such that

$$|\mathcal{E}(s, V_{\mathbf{a}}\alpha) - \mathcal{E}(s, \alpha)|\,|\mathbf{a}|^{-1}$$

is large.

Assume that α has a supremum

$$\|\alpha\| = \sup_{\substack{s\in S \\ \mathcal{E}\in\{\mathcal{E}\}}} |\mathcal{E}(s, \alpha)| \tag{2.7}$$

of all expectation values of α in all external fields. In the following, we will be concerned only with this class of procedures. Then, a figure of merit—the microscopicity $r(\alpha)$ of an observation procedure α—can be defined as

$$r(\alpha) \equiv \lim_{|\mathbf{a}|\to 0} |\mathbf{a}|^{-1} \sup_{\substack{s\in S \\ \mathcal{E}\in\{\mathcal{E}\}}} |\mathcal{E}(s, V_{\mathbf{a}}\alpha) - \mathcal{E}(s, \alpha)|\,\|\alpha\|^{-1} \tag{2.8}$$

if the limit exists.

To see the meaning of the definition (2.8), consider the quantum mechanical expression

$$\mathcal{E}(s, \alpha) = \mathrm{Tr}\rho_s^{\mathcal{E}} A_\alpha^{\mathcal{E}}$$

for the expectation value $\mathcal{E}(s, \alpha)$. Here, $\rho_s^{\mathcal{E}}$ is the density operator which is the image of the state-preparing procedure s induced by the Hamiltonian corresponding to $\mathcal{E}$, and $A_\alpha^{\mathcal{E}}$ the self-adjoint operator that is the image of the observation procedure α induced by $\mathcal{E}$.

A family $\{A_\alpha^{\mathcal{E}}(\mathbf{a})\}$ is obtained by the translation operators $\mathbf{P}$ through

$$A_\alpha^{\mathcal{E}}(\mathbf{a}) = \exp(i\mathbf{P}\cdot\mathbf{a}) A_\alpha^{\mathcal{E}}(0) \exp(-i\mathbf{P}\cdot\mathbf{a}).$$

If all operations are considered as interchangeable, Eq. (2.8) becomes

$$r(\alpha) = \sup_{\substack{s\in S \\ \mathcal{E}\in\{\mathcal{E}\}}} \mathrm{Tr}\,\rho_s^{\mathcal{E}} \frac{\partial}{\partial|\mathbf{a}|} A_\alpha^{\mathcal{E}}(\mathbf{a}) \cdot \|A\|^{-1}. \tag{2.9}$$

Our definition of microscopicity does not agree with some of the common connotations of the word. For instance, the question what is the number of particles in a cube of 1 cm^3 is very microscopic in our sense, because the precise location of particles near the boundary of the cube requires a high resolving power of the observation instrument. A more macroscopic question would add some qualification such as "within an error of . . . ."

3. COHERENCE

If observation instruments examine conditions at several points P_i one has to distinguish between those that combine the results of independent observations at $P_1, P_2, \cdots$ and those in which an observation cannot be subdivided into local ones, i.e., where an examination of a large region cannot be replaced by successive observations of included smaller regions. Examples are (1) observations of the mean velocity or the positions of particles and (2) the measurement of the phase difference between the values of a wavefunction at different points. An observation may be termed coherent or incoherent if it belongs to one or the other of these types.

To make these heuristic remarks precise, we need an operational definition of the decomposition of observations. We consider a sequence $[s_n(\alpha)]$ of observation results $s_n(\alpha)$ obtained by using an observation procedure α on samples prepared by a given state-preparing procedure s. Let $[s_n(\alpha)]$ and $[s_n(\beta)]$ be two such sequences obtained by the same state-preparing procedure, but by two distinct observation procedures α and β, each sample being observed only once and then discarded, and assume that the sequences are random.[3] Consider the use of two observation procedures α and β on the same sample prepared by the procedure s. One obtains two new sequences $[s'_n(\alpha)]$ and $[s'_n(\beta)]$ as results. If the sequences $[s_n(\alpha)]$ and $[s'_n(\alpha)]$ are stochastically indistinguishable and the sequences $[s_n(\beta)]$ and $[s'_n(\beta)]$ are also stochastically indistinguishable for all state-preparing procedures, and for all expectations $\mathcal{E} \in \{\mathcal{E}\}$, we say that the two observation procedures are compatible (or mutually not interfering), because the performance of an observation α on a sample observed by procedure β has no relevant effect on the outcome sesequence.[4,5]

A special case, most frequently considered, is the simultaneous performance of the two observations on the same sample, but our definition holds as well for the observation of the momentum at $t=0$ (α) and $t=\tau(\beta)$.

Consider a new procedure γ which is obtained by adding the outcomes $s'_n(\alpha)$ and $s'_n(\beta)$ of the two compatible procedures α and β, performed each on the same nth sample. One can properly call this procedure the sum $\alpha+\beta=\gamma$.

Similarly, one can define a procedure α^2 by the modified outcome sequences $[s_n^2(\alpha)]$ and, for a large class of functions f, procedures with sequences $[f(s_n(\alpha))]$. By combining these operations, one can define functions of the two procedures α and β, e.g.,

$$2\alpha\cdot\beta=2\beta\cdot\alpha=(\alpha+\beta)^2-\alpha^2-\beta^2. \tag{3.1}$$

One obtains a large class $\{F(\alpha,\beta)\}$ of real-valued functions F of the two compatible procedures α and β.

Conversely, given a procedure γ, one can ask whether it can be decomposed, i.e., replaced by compatible procedures α and β. This will be the case if there exists a function F such that the procedure $F(\alpha,\beta)$ generates the same mean values $\mathcal{E}[s,F(\alpha,\beta)]$ as γ, for all state-preparing procedures and all expectations $\mathcal{E}$, i.e., all Hamiltonians. This effective identify of procedures will be denoted by the symbol $\approx$:

$$F(\alpha,\beta)\approx\gamma. \tag{3.2}$$

It will be assumed that observation procedures are associated with regions R of space–time in which the observation is performed. The apparatus of the procedure is not necessarily confined to R, but the procedure must be insensitive to events outside of R (e.g., a microscope focused on a microscopic volume). For every procedure α, there exists a unique smallest region among those regions R with which it is associated, which we will denote by $R(\alpha)$. This definition of such sets $\mathcal{O}(R)$ of local observation procedures differs from the assignment of space–time regions to *observables* by the requirement that the intersection $\mathcal{O}(R_1)\cap\mathcal{O}(R_2)$ of two subsets associated with two regions vanishes if the two regions do not intersect:

$$[R_1\cap R_2=0]\Rightarrow[\mathcal{O}(R_1)\cap\mathcal{O}(R_2)=0]$$

for observation procedures, while no such requirement can be imposed on observables.

It will be assumed, as an extrapolation of experience, that two procedures belonging to spacelike regions of space–time do not mutually interfere and are compatible in the precise sense described above.

We consider simultaneous space volumes V_i. If $V_1+V_2=V_3$ and $\gamma\in\mathcal{O}(V_3)$, one can ask whether there exist compatible observation procedures $\alpha_i\in\mathcal{O}(V_i)$ $(i=1,2)$ such that γ can be decomposed as $\gamma\approx F(\alpha_1,\alpha_2)$. Clearly, a procedure γ that has such a decomposition can be considered as less coherent than one in which the observation of events in V_3 cannot be replaced by subsequent observations in smaller volumes.

More generally, call V' [such that $R(\gamma)\supset V'$] indecomposable for $\gamma\in\mathcal{O}(V)$ if there exists no decomposition such such that

$$\gamma\approx F(\alpha_1,\alpha_2,\ldots,\alpha_n),\quad \bigcup_{i=1}^{n} R(\alpha_i)\subseteq V, \tag{3.3}$$

and $R(\alpha_i)\subset V'$ for some $i=1,\ldots,n$.

Then,

$$c(\gamma)\equiv\sup\{\mathrm{Vol}(V')\,|\,V'\text{ indecomposable for }\gamma\} \tag{3.4}$$

may be called the coherence of γ because at least a region of volume $c(\gamma)$ must be observed *in toto*, and this observation cannot be replaced by a successive examination of smaller volumes.

Thus far no theoretical (in a narrow sense) but only operational assumptions have been made which are directly falsifiable if not verifiable. We want to show that within the framework of classical physics there is no coherence in the sense defined ($c=0$ for all procedures). The main assumption of classical physics can be stated in operational form: All observation procedures are compatible. (More precisely: Given two equivalence classes of observation procedures, there exists a procedure in one class that is compatible with one procedure in the other class.)

Hence, all real-valued functions of observation procedures are observation procedures, and one can assign to $\mathcal{O}$ the structure of an Abelian algebra.

If, in the spirit of classical field theory, it is further assumed that the algebra $\mathcal{O}(R)$ can be generated by the family $\{\mathcal{O}_x\}_{x\in R}$ of point-observation procedures associated with points $x\in R$ in 4-space, then each procedure associated with a spacelike hyperplane h can be generated by the family $\{\mathcal{O}_x\}_{x\in h}$ of point procedures $\mathcal{O}_x$ associated with points $x\in h$. Hence, the irreducible region V_{irr} is a point.

4. APPLICATIONS

One can ask the question: Are more than 90% of the gas molecules in a vessel concentrated in the left upper corner at the time $t=0$?

The theoretical counterpart of this question is an ele-

ment Q in the algebra of observables—a projection for quantum mechanics, and a characteristic function of an observable in phase space for classical physics. To answer the question, one can measure the number of molecules outside the corner, at the time $t=0$. One could also answer the question by measurements performed at an instant τ, if one knows the dynamics and can solve the equations of motion. This alternative procedure is obviously more difficult to perform. If the time τ is large as compared to the relaxation time, all macroscopic measurements fail, because they merely indicate close-to-equilibrium conditions. It will be necessary to determine within small limits of error, the positions and velocities of all molecules (for classical physics). For quantum mechanics, the requirements are even more stringent. Not only the amplitude of the many-body wavefunction but also its phase has to be determined with a high accuracy. Hence, the difficulty is due to an increase in the microscopicity and of the coherence of the observation procedure. These heuristic considerations can be made precise by saying that the observation procedure $\beta(\tau)$, performed at time τ has a higher microscopicity and coherence than the procedure $\alpha(0)$ performed at time 0, if both procedures are in the equivalence class of procedures that constitutes (or maps onto) the question Q. (The appendix summarizes the assumptions concerning the observation procedures and their mapping onto the observables.)

This problem may seem contrived, but it leads up to the problem of irreversibility. One of the familiar illustrative forms of this problem is the question: Why do we never see the molecules of an uniformly distributed gas converge into one of the corners of the vessel, although such a process should be possible by the laws of physics? The qualitative answer is widely accepted[6]: Such states exist, but it is very difficult or practically impossible to produce them. It is also obvious that the difficulty increases as the time interval between the state preparation and the target situation increases. (It is easy to realize the exceptional situation by pumping gas into an enclosure, in the vessel, and suddenly removing the walls of the enclosure. For a microsecond, the high concentration persists.) One would like to (i) state this answer in a precise and general form and (ii) derive it from known laws of physics. We shall try to do only the first.

The set of states in question (the target situation) is defined by the requirement that certain observables have mean values within some numerical tolerance. For example, the observable is the question Q concerning the presence of N molecules in a (not too sharply defined) region, and the problem may call for a 90% scoring in subsequent trials. Then, the set of states is given by

$$0.9 \leq \langle Q \rangle \leq 1,$$

where $\langle Q \rangle$ is the expectation value of Q with respect to the states to be prepared.

Such a set of states is called a weak neighborhood in quantum mechanics, and, in classical physics, a volume in phase space defined by the intersection of characteristic functions. (We use the Heisenberg or time-independent definition of states.) A state-preparing procedure is usually terminated (adjusted) by monitoring experiments. Thus, the angular and energy width of a particle beam are adjusted by monitoring measurements. The difficulty of preparing a state can be measured by the difficulty of monitoring. Hence, the microscopicity of a weak neighborhood of states could be equated to some monotonic function of the microscopicities associated to the defining observables. An observable is an equivalence class of observation procedures, and it is only for the latter that the microscopicity and coherence are well defined. Thus, if the observables defining the weak neighborhood are associated with a region R, there exist observation procedures in the equivalence classes of these observables associated with R, but also with many other space–time regions, and the microscopicity then depends not only on the observables but on the region in which it is measured.

Thus, the problem of assigning numbers of merit to various methods (instants) for preparing a given weak neighborhood of states is reduced to the problem of assessing the difficulty of various observation procedures that belong to the same observable.

The intuitive experience says: It is more and more difficult to prepare states at an instant $t=\tau$ if they are to lead to given observations at $t=0$, as $|\tau|$ increases. The precise form of this statement can now be given.

Conjectured theorem: Given an observable $A(0)$ associated with the instant $t=0$, and a family of observation procedures $\alpha(t) \in A(0)$ associated with instants t and members of the equivalence class $A(0)$, the coherence $c[\alpha(t)]$ and the microscopicity $r[\alpha(t)]$ increase monotonically with $|t|$.

This conjectured theorem should take the place of the conjectured (and disproved) H-theorem. It states not that that disorder increases with time, but that our difficulty in creating order at a time t, after our manipulation has ceased, increases with the time delay t.

The difficulty of preparing the states in the weak neighborhood may be defined as some monotonically increasing function

$$f\{\sum_i r[\alpha_i(t)],\ \sum_i c[\alpha_i(t)]\}$$

of the sums $\sum_i r[\alpha_i(t)]$ of the microscopicities r and $\sum_i c[\alpha_i(t)]$ of the coherences c associated with the observation procedures α_i performed at the instant t which define the weak neighborhood (the target situation) through

$$|\mathcal{E}(s, \alpha_i) - \mathcal{E}(s_0, \alpha_i)| \leq \epsilon_i.$$

ACKNOWLEDGMENTS

The author is indebted for valuable discussions to F. Coester, A Grossmann, and A.J.F. Siegert.

APPENDIX

The relationship between the nonmathematical procedures $\mathcal{O}$ and $\mathcal{S}$ and their theoretical representatives is made in two steps: First, the sets $\mathcal{O}$ and $\mathcal{S}$ are assumed to have certain algebraic and topological properties; second, morphisms $\mathcal{O} \to \mathfrak{A}$ and $\mathcal{S} \to S(\mathfrak{A})$ onto the

algebra $\mathfrak{A}$ of observables and onto the set $S(\mathfrak{A})$ of states on $\mathfrak{A}$, respectively, are introduced.[7,8] The postulates depend, of course, on the choices between classical and quantum mechanical and between Galilean and relativistic theories.

For convenience, some of the assumptions valid for Galilean quantum mechanics, are reproduced here:

(A1) The set O of observation procedures is a normed star algebra with unit I, the set S of state-preparing procedures a convex linear set, and $\{\mathcal{E}\}$ a collection of mappings (expectations)

$$\mathcal{E}: S \times O \to \mathbb{R}$$

of the Cartesian product $S \times O$ into the complex numbers.

(A2) $\mathcal{E}(s, \cdot)$ is a linear, positive, normalized form on O $[\mathcal{E}(s, I) = 1]$.

(A3) $\mathcal{E}(\cdot, \alpha)$ is a linear function on S.

(A4) S and O are mutually separating in the following sense: For every pair $s_1, s_2 \in S$, the statement that $\mathcal{E}(s_1, \alpha) = \mathcal{E}(s_2, \alpha)$ for all $\alpha \in O$ and all $\mathcal{E} \in \{\mathcal{E}\}$ implies $s_1 = s_2$. For every $\alpha \in O$, the statement that $\mathcal{E}(s, \alpha) = 0$ for all $\mathcal{E} \in \{\mathcal{E}\}$ and all $s \in S$ implies $\alpha = 0$.

(A5) Every expectation $\mathcal{E} \in \{\mathcal{E}\}$ determines a kernel

$$K(\mathcal{E}) = \{\alpha \in O \mid \mathcal{E}(s, \alpha) = 0 \text{ for all } s \in S\}.$$

Every kernel $K(\mathcal{E})$, $\mathcal{E} \in \{\mathcal{E}\}$, is an ideal in O.

(A6) All quotient algebras

$$O/K(\mathcal{E}) = \mathfrak{A}$$

are isomorphic to each other for all $\mathcal{E} \in \{\mathcal{E}\}$. The algebra of observables is a discrete von Neumann algebra.

(A7) A given $\mathcal{E} \in \{\mathcal{E}\}$ induces an equivalence relation $E(\mathcal{E})$ of elements $s_i \in S$:

$$s_i E s_j \text{ iff } \mathcal{E}(s_i, \alpha) = \mathcal{E}(s_j, \alpha) \text{ for all } \alpha \in O.$$

Then, the quotient set $S/E(S) \approx S(\mathfrak{A})$ is isomorphic to the convex set $S(\mathfrak{A})$ of normal, positive, linear, normalized forms on $\mathfrak{A}$ (i.e., the normal states on $\mathfrak{A}$).

(A8) Weak causality: There is a (nonlinear) representation $T_\tau \mapsto V_\tau$ of the time-translation group $\{T_\tau\}$ by automorphisms $V_\tau | \{O_t\}$ restricted to a family $\{O_t\}$ of subalgebras of observation procedures at instants t:

$$V_\tau : O_t \to O_{t-\tau} \, [\alpha(t) \in O_t \mapsto \alpha(t - \tau) \in O_{t-\tau}]$$

such that for every $\mathcal{E} \in \{\mathcal{E}\}$ and for every pair s, $r \in S$, the statement that

$$\mathcal{E}(s, \alpha) = \mathcal{E}(r, \alpha) \text{ for all } \alpha \in O_t$$

implies

$$\mathcal{E}(s, \beta) = \mathcal{E}(r, \beta) \text{ for all } \beta \in O.$$

(A9) Strong causality: There is a (nonlinear) representation $T_\tau \mapsto V_\tau$ of the time-translation group $\{T_\tau\}$ by automorphisms $V_\tau | \{O_{tc}\}$ restricted to a family $\{O_{tc}\}$ of intrinsic (canonical) subalgebras O_{tc}:

$$V_\tau : O_{tc} \to O_{t-\tau, c} \, [\alpha(t) \in O_{tc} \mapsto \alpha(t - \tau) \in O_{t-\tau, c}],$$

and a similar representation $T_\tau \to W_\tau$ by automorphisms $W_\tau | \{S\}_t$ restricted to a family $\{S_\tau\}$ of convex subsets (instant state-preparing procedures) $S_t \subset S$:

$$W_\tau : S_t \to S_{t-\tau} \, [s(t) \in S_t \mapsto s(t - \tau) \in S_{t-\tau}],$$

such that:

(a) identical state-preparing procedures are defined by the statement that, for every pair $\mathcal{E}_1$, $\mathcal{E}_2 \in \{\mathcal{E}\}$ and for fixed t, one finds

$$\mathcal{E}_1(s, \alpha) = \mathcal{E}_2(s, \alpha) \text{ for } \alpha \in O_{tc} \text{ and } s \in S_t.$$

(b) the statement of strong causality is that for every pair r, $s \in S_t$ and every $\mathcal{E} \in \{\mathcal{E}\}$,

$$\mathcal{E}(r, \alpha) = \mathcal{E}(s, \alpha) \text{ for all } \alpha \in O_{tc}$$

implies

$$\mathcal{E}(r, \beta) = \mathcal{E}(s, \beta) \text{ for all } \beta \in O.$$

(c) the union $\cup_t O_{tc}$ of the intrinsic subalgebras generates O. The union $\cup_t S_t$ of instant state-preparing procedures generates, by convex linear combinations, the convex set S.

(A10) For every $\mathcal{E} \in \{\mathcal{E}\}$ and $\alpha \in O$, the statement that

$$\mathcal{E}(s, \alpha) = 0 \text{ for all } s \in S_t$$

implies

$$\mathcal{E}(r, \alpha) = 0 \text{ for all } r \in S.$$

(That is, all states can be produced at one instant.)

Another assumption to be added, in the context of the present paper, is the commutativity of observation procedures which have the operational relationship of compatibility. It is shown in Ref. 5 that this assumption agrees with what one may call quantum mechanical intuition.

[1] I.E. Farquhar, *Ergodic Theory in Statistical Mechanics* (Wiley, New York, 1964).
[2] M. Planck, *Vorlesungen über Thermodynamik* (de Gruyter, Berlin, 1954).
[3] A. precise definition of randomness is given by P. Benioff, Phys. Rev. D 7, 3603 (1973).
[4] A more precise form of this statement is: All four sequences are (τ) random in the sense of Ref. 3 and the probability measure constructed from $[s_n(\alpha)]$ equals that constructed from $[s'_n(\alpha)]$, and the same is true for $[s_n(\beta)]$ and $[s'_n(\beta)]$, respectively. The construction is carried out explicitly in Ref. 5.
[5] P. Benioff, J. Math. Phys. 15, 522 (1974).
[6] J. Rothstein, Found. Phys. 4, 83 (1974).
[7] H. Ekstein, Phys. Rev. 184, 1315 (1969).
[8] Y. Avishai and H. Ekstein, Phys. Rev. D 7, 983 (1973).

Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics

D. Bessis, P. Moussa, and M. Villani*

Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, BP n° 2-91190-Gif-sur-Yvette, France
(Received 2 May 1975)

We show, by making use of the functional integral technique, that, for a large class of useful quantum statistical systems, the partition function is, with respect to the coupling constant, the Laplace transform of a positive measure. As a consequence, we derive an infinite set of monotonicly converging upper and lower bounds to it. In particular, the lowest approximation appears to be identical to the Gibbs–Bogolioubov variational bound, while the next approximations, for which we give explicit formulas for the first few ones, lead to improve the previous bound. The monotonic character of the variational successive approximations allows a new approach towards the thermodynamical limit.

I. INTRODUCTION AND HISTORICAL BACKGROUND

Our aim is to extend to quantum statistical mechanical systems a variational and perturbative method introduced in a previous paper,[1] leading to monotonic converging bounds for the eigenvalues of a semibounded N particle Hamiltonian H.

More precisely, we showed that each pole of the Padé approximation (PA) constructed on the resolvent of H was providing, for the discrete part of the spectrum of H, an upper bound to the corresponding exact eigenvalue, and that this upper bound can be used as a variational bound in the pivot $|\varphi\rangle$ (the test vector, for which the mean value of the resolvant is computed). These variational upper bounds were clear generalizations of the Rayleigh—Ritz variational principle, which include more moments than the first one: the nth moment being $\langle\varphi|H^n|\varphi\rangle$.

In particular, when the moments do not exist (for instance, in the case of ultraviolet divergence), but admit a regularization, while the PA cannot have variational properties in the regulator (because they do not remain bounded in the vicinity of their poles), it was shown that the arctan of the PA admits nice variational properties in the regulator, which allow one to reconstruct the spectrum through an extended Padé—Rayleigh—Ritz variational principle which includes the regulator as variational parameter.

However, in physics, one is not only interested in reconstructing the spectrum, but also in computing, either the evolution operator $\exp(itH)$ or its "Euclidean version" $\exp(-\beta H)$. $\exp(itH)$ represents the evolution operator of the system between time zero and time t, while $\exp(-\beta H)$ is related to the Gibbs density matrix of the system at equilibrium and temperature $kT=1/\beta$.

In the previous paper, all mathematical properties, were based on the fact that the resolvent appears as a Stieltjes function in the energy or in the coupling constant parameter. Stieltjes functions are special types of Laplace transform of positive measure. Similarly, from the spectral decomposition of H, $\exp(-\beta H)$ is, in β, the Laplace transform of a positive valued measure:

$$\exp(-\beta H)=\int_{E_0}^{\infty}\exp(-\beta E)\,dP_E, \tag{I.1}$$

where dP_E is the projector onto the eigenvalue E of H and dP_E is a positive operator valued measure.

The positivity of the measure allows us to construct, for the trace of (I.1), approximations based on the Gaussian integration method.[2] This approximation is also known under the name of generalized Padé approximants (GPA), because the weights of the Gaussian integration method are simply the residues of the ordinary PA associated with the positive measure, while the zeroes of the Gaussian method are the poles of the same PA.

More precisely it was shown[2,3] that the GPA were providing, for a finite number of particles N, monotonic decreasing sequences of converging upper bounds for the diagonal GPA, while the subdiagonal ones were giving monotonic increasing sequences of converging lower bounds to the partition function. The exact solution is constrained between these bounds.

Furthermore,[3] these GPA can have variational properties in the number of particles N. By making use of this remark, it is possible to obtain, for $N\to\infty$ (the thermodynamical limit), a monotonic sequence of decreasing or increasing converging bounds.

This method converges clearly for any temperature and any density. However, one would like to start from an exactly solvable Hamiltonian H_0, for which, for example, the thermodynamical quantities are exactly known, and perturb it, to see how these quantities are changed. It is therefore of interest to study the following:

Question: What are the positivity properties of $\exp(-\beta H)$ in the coupling constant λ

$$(\text{where } H=H_0+\lambda H_I)? \tag{I.2}$$

We propose the following

$$Z(\lambda)=\mathrm{Tr}\exp[-\beta(H_0+\lambda H_I)]=\int\exp(-\lambda\tau)\,d\mu(\tau), \tag{I.3}$$

where $d\mu(\tau)$ is a *positive measure* the support of which is contained in the convex hull of the eigenvalues of H_I (that is, between the inf and the sup of the spectrum of H_I).

Up to now, we are able to prove this statement in the three following cases:

A—H_0 and H_I are Hilbert space commuting operators.

B—H_0 and H_I are 2×2 matrices, and more generally when $\exp(-\beta H_0)$ has all its matrix elements positive in a basis where H_I is diagonal.

C—When

$$H_0=\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i}+V(\mathbf{r}_1,\ldots,\mathbf{r}_N),\quad \mathbf{p}_k=-i\nabla_k, \tag{I.4}$$

$$H_I=V'(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N).$$

Case A corresponds to classical statistical mechanics. Case C corresponds to an N-body quantum mechanical system of N spinless particles with the most general bounded from below *local* interaction. We could not find any counterexample to the general statement (I. 3). Whether (I. 3) stands in general or is subject to restrictions is an open question. A consequence of the theorem of case C is that it is possible to construct, for a system, with a finite number of particles N, a monotonic sequence of converging upper and lower bounds, to the free energy per particle of the system:

$$F_N(\rho,\beta,\lambda)=-(1/\beta N)\log[Z_N(\rho,\beta,\lambda)], \tag{I.5}$$

where ρ is the density of the system. In particular, it is interesting to investigate the behavior $N\to\infty$ of these bounds: the thermodynamical limit.

The lowest approximation, which is constructed from the GPA [0/1] to the partition function gives rise to the bound

$$F_N(\lambda)<[0/1]_N(\lambda)=F_N(0)+(1/N)\,\mathrm{Tr}(\rho_0\lambda H_I), \tag{I.6}$$

where ρ_0 is the density matrix for the Hamiltonian H_0:

$$\rho_0=\exp(-\beta H_0)/\mathrm{Tr}\exp(-\beta H_0)\quad\text{and}$$

$$\rho=\exp(-\beta H)/\mathrm{Tr}\exp(-\beta H). \tag{I.7}$$

(I. 6) is generally derived from the Gibbs—Bogolioubov variational bound:

$$\mathrm{Tr}(\rho_0\log\rho_0)>\mathrm{Tr}(\rho_0\log\rho). \tag{I.8}$$

Due to the extensive properties of the terms in Eq. (I. 6), the $N\to\infty$ limit is straightforward and meaningful. The next approximations [1/1] and [1/2] do not enjoy the same extensivity property as the lowest approximation [0/1]. In particular, in the thermodynamical limit, all our bounds have the same limit as the lowest approximation [0/1] or [0/0] depending if one takes the $[P\text{-}1/P]$ or the $[P/P]$ approximation.

Therefore, if one wants to work directly with $N=\infty$ on the bounds, the information coming from the knowledge of more than the first two terms in the perturbation expansion in λ of the partition function seems to be lost.

However, it is easy to circumvent this difficulty by changing the traditions: One works with finite N (number of particles) which provides a natural cutoff for the perturbation terms of the expansion in λ of the partition function. Then, it is not difficult to show following the method described in Ref. 3 that the bounds are variational in N.

Then one obtains, for the true free energy per particle, in the thermodynamical limit, a succession of embedded variational converging bounds, the variational parameter being the number of particles itself.

Finally, we want to point out that, as will be shown in the sequel, our method for treating the operator $\exp[-\beta(H_0+\lambda H_I)]$ is based actually on Wiener's functional approach.

The reader will find:

– In Sec. II, the present status of the Laplace transform theorems.

– In Sec. III, the derivation of monotonic sequences of lower and upper bounds to the free energy of a quantum system of N particles.

– In Sec. IV, the analysis of the thermodynamical limit of the previous bounds, as well as the extension of the method to the case of singular interactions.

– In Sec. V, the conclusion and a general outlook for the Euclidean field theory.

II. THE LAPLACE TRANSFORM THEOREMS

We shall discuss under which conditions the following conjecture holds:

Conjecture ($\mathcal{C}$): Let A and B be two bounded from below selfadjoint operators, and $|\varphi\rangle$ be an eigenvector of B. Then

$$\langle\varphi|\exp[-(A+\lambda B)]|\varphi\rangle=\int\exp(-\lambda\tau)\,d\mu(\tau). \tag{II.1}$$

$d\mu(\tau)$ is a positive measure with support contained in the convex hull of the spectrum of B.

We shall prove the following theorems:

Theorem 1: ($\mathcal{C}$) holds when A and B commute.

Theorem 2: ($\mathcal{C}$) holds when, for all real positive number ρ, $\exp(-\rho A)$ has nonnegative matrix elements in a basis where B is diagonal.

Corollary to Theorem 2: ($\mathcal{C}$) holds when:

(a) A, B are any two dimensional matrices.

(b) A is a tridiagonal matrix [$A_{ij}=0$ if $j\neq(i;i-1$ or $i+1)$] in a basis where B is diagonal (A bounded).

(c) Off-diagonal matrix elements of A are negative in a basis where B is diagonal (A bounded).

Proof: —Theorem 1 is a trivial consequence of the spectral decomposition of B.

—Theorem 2 makes use of the Trotter's formula[4] for bounded from below auto-adjoint operators. We have

$$\langle\varphi|\exp[-(A+\lambda B)]|\varphi\rangle=\lim\langle\varphi|\{\exp(-A/n)\times\exp(-\lambda B/n)\}^n|\varphi\rangle. \tag{II.2}$$

Consider

$$\langle\varphi|\{\exp(-A/n)\exp(-\lambda B/n)\}^n|\varphi\rangle=X_n \tag{II.3}$$

and

$$X_n=\int d\varphi_1\cdots d\varphi_{n-1}\exp\{-\lambda/n[b(\varphi_1)+\cdots+b(\varphi_{n-1})+b(\varphi)]\}$$

$$\times\langle\varphi\,|\exp(-A/n)|\,\varphi_1\rangle\langle\varphi_1|\exp(-A/n)|\,\varphi_2\rangle\cdots$$
$$\times\langle\varphi_{n-1}|\exp(-A/n)|\,\varphi\rangle, \tag{II.4}$$

where we have introduced $(2n-1)$ times the closure relation for the spectral decomposition of B:

$$B=\int |\varphi\rangle\, b(\varphi)\langle\varphi|\, d\varphi, \tag{II.5}$$

$$I=\int |\varphi\rangle\, d\varphi\langle\varphi|. \tag{II.6}$$

Therefore we have

$$X_n=\int \exp(-\lambda\tau)\, d\mu_n(\tau) \tag{II.7}$$

with

$$d\mu_n(\tau)=\int d\varphi_1\cdots d\varphi_{n-1}\,\delta\left(\tau-\frac{1}{n}\sum_{i=1}^{n} b(\varphi_i)\right)$$
$$\langle\varphi|\exp(-A/n)|\,\varphi_1\rangle$$
$$\times\langle\varphi_1|\exp(-A/n)|\,\varphi_2\rangle\cdots\langle\varphi_{n-1}|\exp(-A/n)|\,\varphi\rangle, \tag{II.8}$$

where we have set

$$\varphi_n=\varphi. \tag{II.9}$$

Since all matrix elements occurring in (II.8) are positive $d\mu_n(\tau)$ is a positive measure. Furthermore, the support of $d\mu_n(\tau)$ is clearly contained in the convex hull of the spectrum of B. The limit $n\to\infty$ of (II.8) is a positive measure[5] (see the Appendix for the proof). Theorem 2 is then proved.

For the corollary first prove (c). Consider $\exp(-\rho A/n)$ for n integer big enough such that $(\exp(-\rho A/n))_{ij}\sim\delta_{ij}-\rho A_{ij}/n$ is positive for all i and j. This is possible, if for $i\neq j$, A_{ij} is strictly negative and bounded. The positivity of $(\exp(-\rho A))_{ij}$ is then obtained by using $\exp(-\rho A)=(\exp(-\rho A/n))^n$. The case where some A_{ij} are zero can be dealt with by continuity. Conditions of Theorem 2 are fulfilled without any hypothesis on the diagonal elements of A, and therefore (c) is obtained.

The nondiagonal elements of a tridiagonal matrix can always be given real negative values by multiplying the basis vectors by a suitable phase factor. Then (b) becomes a particular case of (c). It is clear, on the other hand, that 2×2 matrices are all tridiagonal, which proves (a).

Remark on the 2×2 *matrix case*: In this case the direct computation of the measure is possible. The measure appears as the sum of δ functions the argument of which are the eigenvalues of B, plus an entire and positive Bessel function spread out between these eigenvalues. It would be interesting to investigate in the general case of $d\times d$ matrices the nature of the measure.

Finally we want to point out that conjecture ($\mathcal{C}$) does *not* extend to case where the mean value of $\exp(-(A+\lambda B))$ is taken in a general vector, instead of $|\varphi\rangle$ eigenvector of B: Explicit counterexamples can be worked out.

Conjecture ($\mathcal{C}$) holds in quantum statistical mechanics when we consider spinless particles. More precisely we have the fundamental theorem:

Theorem 3: Let T, V, $T+V=H_0$, V', and $H_0+\lambda V'$ be self-adjoint Hilbert space operators, bounded from below such that

$$T=\sum_{i=1}^{N}\mathbf{p}_i^2;\ \mathbf{p}_K=-i\nabla_k \tag{II.10}$$

$$V=V(\mathbf{r}_1,\ldots,\mathbf{r}_N) \tag{II.11}$$

$$V'=V'(\mathbf{r}_1,\ldots,\mathbf{r}_N) \tag{II.12}$$

where V includes eventually the wall potential which describes the finite volume in which particles are confined. Then, if we represent by $\mathbf{r}$ the collection $\{\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N\}$, we can write

$$\mathcal{J}(\beta,\lambda)=\langle\mathbf{r}|\exp[-\beta(T+V+\lambda V')]|\,\mathbf{r}\rangle=\int\exp(-\lambda\tau)\,d\mu(\tau). \tag{II.13}$$

the support of the positive measure $d\mu(\tau)$ extends from the inf to the sup of the spectrum of V'. The proof goes along the same lines as that for Theorem 2, using the positivity of

$$\langle\mathbf{r}|\exp(-\beta/n\,T)|\,\mathbf{r}'\rangle=\left(\frac{\beta}{4\pi n}\right)^{3N/2}\exp\left(-\frac{n}{4\beta}\sum_{i=1}^{n}(\mathbf{r}_i-\mathbf{r}_i')^2\right). \tag{II.14}$$

We insert the closure relation for the position operator in Trotter's formula and set

$$\mathcal{J}(\beta,\lambda)=\langle\mathbf{r}|\exp[-\beta(T+V+\lambda V')]|\,\mathbf{r}\rangle$$
$$=\lim_{n\to\infty}\langle\mathbf{r}|(\exp(-\beta T/n)\exp(-\beta V/n)$$
$$\times\exp(-\beta\lambda V'/n))^n|\,\mathbf{r}\rangle \tag{II.15}$$
$$=\lim_{n\to\infty}\left(\frac{\beta}{4\pi n}\right)^{3Nn/2}\int d\mathbf{r}^{(1)}\cdots d\mathbf{r}^{(n-1)}$$
$$\times\exp\left[-\frac{\beta}{n}\left(\sum_{i=1}^{n}V(\mathbf{r}^{(i)})+V'(\mathbf{r}^{(i)})\right)\right]$$
$$-\frac{n}{4\beta}\sum_{j=1}^{N}\sum_{i=0}^{n-1}(\mathbf{r}_j^{(i)}-\mathbf{r}_j^{(i+1)})^2\Bigg], \tag{II.16}$$

where $\mathbf{r}^{(i)}$ stands symbolically for the set $\{\mathbf{r}_1^{(i)},\mathbf{r}_2^{(i)},\ldots,\mathbf{r}_N^{(i)}\}$ and $\mathbf{r}^{(n)}=\mathbf{r}=\mathbf{r}^{(0)}$.

$\mathcal{J}(\beta,\lambda)$ can be rewritten as

$$\mathcal{J}(\beta,\lambda)=\lim_{n\to\infty}\int\exp(-\lambda\tau)\,d\mu_n(\tau) \tag{II.17}$$

$$d\mu_n(\tau)=\left(\frac{\beta}{4\pi n}\right)^{3Nn/2}\int d\mathbf{r}^{(1)}\cdots d\mathbf{r}^{(n-1)}\,\delta\left(\tau-\frac{\beta}{n}\sum_{i=1}^{n}V'(\mathbf{r}^{(i)})\right)$$
$$\times\exp\left(-\frac{n}{4\beta}\sum_{i=0}^{n-1}\sum_{j=1}^{N}(\mathbf{r}_j^{(i)}-\mathbf{r}_j^{(i+1)})^2-\frac{\beta}{n}\sum_{i=1}^{n}V(\mathbf{r}^{(i)})\right). \tag{II.18}$$

Clearly $\mathcal{J}(\beta,\lambda)$ is the limit of the Laplace transform of positive measures $d\mu_n$. Trotter's theorem tells us that the limit in (II.17) exists and $d\mu_n(\tau)$ itself has a limit which is a positive measure $d\mu(\tau)$ (see the Appendix for the proof) and

$$\mathcal{J}(\beta,\lambda)=\int\exp(-\lambda\tau)\,d\mu(\tau). \tag{II.19}$$

Of course, the support of $d\mu(\tau)$ is obtained as the convex hull of the spectrum of V', which means the interval going from inf spectrum of V' to sup spectrum of V'.

It is clear that this proof of Theorem 3 could have been derived using the Wiener functional integration instead of Trotter's formula.[4] It is evident that, in all these theorems from 1 to 4, one can replace everywhere the mean value by the traces.

Extend our results at least for the trace to more general situations, such as the finite dimensional case, which can be thought to describe a discrete fermion field, could be of a great interest for the theory of functional integration itself.

III. UPPER AND LOWER BOUNDS TO THE FREE ENERGY OF A QUANTUM SYSTEM OF N PARTICLES

We are now faced with the problem of approximating [cf. (I. 2) and (I. 3)]:

$$Z(\lambda)=\int_a^{+\infty}\exp(-\lambda\tau)\,d\mu(\tau),\quad d\mu(\tau)>0. \tag{III. 1}$$

In general, a, the lower bound of the spectrum of the perturbation, will depend on the number of body N. We shall give to ourselves the "perturbative" expansion of $Z(\lambda)$

$$Z(\lambda)=\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}\,\mu_n, \tag{III. 2}$$

where the moments are

$$\mu_n=\int_a^{+\infty}\tau^n\,d\mu(\tau). \tag{III. 3}$$

It is always possible when a is finite (finite number of bodies) to transform (III. 1) into an analogous problem with $a=0$ by setting $\tau=a+\xi$. In (III. 1) we get

$$Z(\lambda)=\exp(-\lambda a)\int_0^{+\infty}\exp(-\lambda\xi)\,d\mu(a+\xi). \tag{III. 4}$$

The moments $\bar{\mu}_n$ of

$$\bar{Z}(\lambda)=\int_0^{\infty}\exp(-\lambda\xi)\,d\mu(a+\xi) \tag{III. 5}$$

being simply related to the moments $\bar{\mu}_n$:

$$\mu_n=\sum_{p=0}^{n}C_n^p\,a^{n-p}\bar{\mu}_p \tag{III. 6}$$

The interesting fact is that it is possible knowing a *finite* number of μ_n (or $\bar{\mu}_n$) to get lower and upper bounds to $Z(\lambda)$. In fact much better can be achieved: Monotonic sequences of lower or upper bounds can be constructed which constrain the solution $Z(\lambda)$.[2] We have the following set of inequalities:

$$Z^{[0/1]}(\lambda)<Z^{[1/2]}(\lambda)<\cdots<Z^{[P-1/P]}(\lambda)<\cdots<Z(\lambda),$$

and

$$Z(\lambda)<\cdots<Z^{[P/P]}(\lambda)<\cdots<Z^{[2/2]}(\lambda)<Z^{[1/1]}(\lambda)<Z^{[0/0]}(\lambda), \tag{III. 7}$$

where

$$Z^{[P-1/P]}(\lambda)=\sum_{i=1}^{P}w_i^{(P)}\exp(-\lambda\xi_i^{(P)}), \tag{III. 8}$$

$$Z^{[P/P]}(\lambda)=\sum_{i=0}^{P}\tilde{w}_i^{(P)}\exp(-\lambda\tilde{\xi}_i^{(P)})\quad(\tilde{\xi}_0^{P}=0). \tag{III. 9}$$

These formulas are just the traditional Gaussian integration approximations. It is, however, simpler to interpret (III. 8) and (III. 9) as generalized Padé approximations.[2,6]

In fact, to build up the ξ's and w's, one has to introduce the Stieltjes function associated with the Laplace transform (III. 5):

$$R(\lambda)=\int_0^{\infty}\frac{d\mu(a+\xi)}{1+\lambda\xi}, \tag{III. 10}$$

for which we can construct the PA to $R(\lambda)$ in terms of a finite number of moments μ_n. They read

$$R^{[P-1/P]}(\lambda)=\sum_{i=1}^{P}\frac{w_i^{(P)}}{1+\xi_i^{(P)}\lambda},\quad R^{[P/P]}(\lambda)=\sum_{i=0}^{P}\frac{\tilde{w}_i^{(P)}}{1+\tilde{\xi}_i^{(P)}\lambda}, \tag{III. 11}$$

$$\tilde{\xi}_0^{(P)}=0.$$

The ξ's are simply connected to the poles of the PA to $R(\lambda)$ and w's to their residues. On the other hand the denominators of PA are set of orthogonal polynomials with respect to the measure[6] $d\mu$. This allows us to understand how the Gaussian integration is a particularly simple generalized Padé approximation. For more details on generalized PA see Ref. 6, on the Gaussian integration see Ref. 2, and on the orthogonal polynomials, Ref. 7.

If we consider the partition function for a quantum system of N particles interacting via local N-body potentials, then

$$H_N=\sum_{i=1}^{i=N}\mathbf{p}_i^2+V(\mathbf{r}_1,\ldots,\mathbf{r}_N)+\lambda V'(\mathbf{r}_1,\ldots,\mathbf{r}_N) \tag{III. 12}$$

$$=H_0+\lambda H_I. \tag{III. 12'}$$

Using Theorem 3 and the inequalities (III. 7), we get, introducing the free energy per particle,

$$F_N(\lambda)=-(1/\beta N)\log[Z(\lambda,N)], \tag{III. 13}$$

as well as the free energy per particle given by the $[P/Q]$th approximation,

$$F_N^{[P/Q]}(\lambda)=-(1/\beta N)\log[Z^{[P/Q]}(\lambda,N)], \tag{III. 14}$$

the following inequalities:

$$F_N^{[0/0]}(\lambda)<F^{[1/1]}(\lambda)<\cdots<F^{[P/P]}(\lambda)<\cdots<F_N(\lambda)<\cdots$$

$$<F_N^{[P-1/P]}(\lambda)<\cdots<F_N^{[0/1]}(\lambda). \tag{III. 15}$$

In this set of inequalities we recall that $F_N(\lambda)$ is the true free energy per particle (N finite) and $F_N^{[P/Q]}(\lambda)$ is $[P/Q]$ generalized Padé approximation to it:

$$F_N^{[P-1/P]}(\lambda)=-(1/\beta N)\log\exp[-\lambda a(N)]\sum_{i=1}^{P}w_i^{(P)}\times\exp(-\lambda\xi_i^{(P)}) \tag{III. 16}$$

and an analogous formula for $F_N^{[P/P]}(\lambda)$.

Before going into the explicit formulas for the approximation let us remark that the $[P\text{-}1/P]$ approximation does not depend explicitly on the lower bound a of the integral in (III. 1), this property is due to the well-known homographical transformation properties of the $[P\text{-}1/P]$ PA under a translation on the variable of integration, property which is not true for the $[P/P]$ approximation. As a consequence, for the $[P\text{-}1/P]$ GPA one does not need to make the a translation as in (III. 6).

In general we shall give our approximations, in terms of the "moments":

$$\mu_k=(-)^k \frac{d^k}{d\lambda^k} Z_N(\lambda)\Big|_{\lambda=0}, \tag{III. 17}$$

where $Z_N(\lambda)$ is the true partition function for N particles. However, it appears to be more convenient to introduce the cumulants defined by

$$C_k=(-1)^k \frac{d^k}{d\lambda^k} \log[Z_N(\lambda)]\Big|_{\lambda=0}, \quad k>1. \tag{III. 18}$$

The first three cumulants read

$$C_1=\frac{\mu_1}{\mu_0}, \quad C_2=\frac{\mu_2}{\mu_0}-\left(\frac{\mu_1}{\mu_0}\right)^2>0,$$
$$C_3=\frac{\mu_3}{\mu_0}-3\frac{\mu_1}{\mu_0}\frac{\mu_2}{\mu_0}+2\left(\frac{\mu_1}{\mu_0}\right)^3 \tag{III. 19}$$

This allows one to have a simple understanding of the lowest approximations, which make use only of the first three moments or cumulants.

With these remarks in mind, one can now consider:

— The [0/0] approximation which reads

$$F_N^{[0/0]}(\lambda)=-\frac{1}{\beta N}\log(\exp(-\lambda a)\mu_0) \tag{III. 20}$$

$$=F_N(0)+\frac{\lambda}{\beta}\frac{a(N)}{N}<F_N(\lambda), \tag{III. 21}$$

which provides a lower bound to the exact free energy.

— The [0/1] approximation which reads

$$F_N^{[0/1]}(\lambda)=-\frac{1}{\beta N}\log\left[\mu_0\exp\left(-\lambda\frac{\mu_1}{\mu_0}\right)\right] \tag{III. 22}$$

$$=F_N(0)+\frac{\lambda}{N}\mathrm{Tr}(\rho_0 H_I)>F_N(\lambda), \tag{III. 23}$$

which provides an upper bound to the exact free energy, where we have introduced the unperturbed density matrix

$$\rho_0=\exp(-\beta H_0)/\mathrm{Tr}\,\exp(-\beta H_0). \tag{III. 24}$$

If we introduce also the exact density matrix

$$\rho=\exp[-\beta(H_0+\lambda H_I)]/\mathrm{Tr}\,\exp[-\beta(H_0+\lambda H_I)] \tag{III. 25}$$

We can derive (III. 23) from the Gibbs—Bogolioubov inequality:

$$\mathrm{Tr}\rho_0\log\rho_0>\mathrm{Tr}\rho_0\log\rho. \tag{III. 26}$$

One recognizes in (III. 23) the traditional Gibbs—Bogolioubov inequality of quantum statistical mechanics for the free energy, which appears to be nothing but the [0/1] GPA in our scheme.

— The [1/1] approximation reads

$$F_N^{[1/1]}(\lambda)=F_N^{[0/0]}(\lambda)$$
$$-\frac{1}{\beta N}\log\left(\frac{C_2+(C_1-a)^2\exp\{-\lambda[C_2+(C_1-a)^2]/C_1-a\}}{C_2+(C_1-a)^2}\right). \tag{III. 27}$$

This approximant clearly provides a better lower bound to $F_N(\lambda)$ than the previous $F_N^{[0/0]}(\lambda)$.

And the [1/2] approximation reads

$$F_N^{[1/2]}(\lambda)=F_N^{[0/1]}+\frac{\lambda}{\beta N}\frac{C_3}{2C_2}-\frac{1}{\beta N}\log\cosh\left[\left(\frac{C_3}{2C_2}\right)^2+C_2\right]^{1/2}$$
$$-\frac{1}{\beta N}\log\left\{1+\frac{(C_3/2C_2)}{[(C_3/2C_2)^2+C_2]^{1/2}}\right.$$
$$\left.\times\tanh\left[\left(\frac{C_3}{2C_2}\right)^2+C_2\right]^{1/2}\right\}. \tag{III. 28}$$

$F_N^{[1/2]}(\lambda)$ provides a better upper bound to $F_N(\lambda)$ than $F_N^{[0/1]}(\lambda)$ (which is the Gibbs—Bogolioubov bound).

To understand clearly the content of these new bounds, it is necessary to consider the thermodynamical limit; this is the object of the next paragraph.

IV. UPPER AND LOWER BOUNDS FOR THE FREE ENERGY OF A QUANTUM SYSTEM OF N PARTICLES IN THE THERMODYNAMICAL LIMIT

Before we take the limit $N\to\infty$ (number of particles going to infinity) on the set of bounds (III. 15), we want to point out that, for fixed N, if we let the order of approximations $P\to\infty$, the two bounds

$$F_N^{[P/P]}(\lambda)<F_N(\lambda)<F_N^{[P-1/P]}(\lambda) \tag{IV. 1}$$

tend to each other and therefore to the true $F_N(\lambda)$, when the moment problem is determinate. For a finite number of particles and a bounded perturbation the moment problem is always determined because then, by Theorem 3, $Z_N(\lambda)$, the partition function, is an entire function of λ, the support of the measure being bounded. The radius of convergence of the Taylor expansion of $Z_N(\lambda)$ in λ being infinite, we have for the moments the inequality

$$|\mu_n|<\epsilon^n n!, \quad \epsilon>0, \tag{IV. 1'}$$

and therefore the Carleman[6] condition is fulfilled and the moment problem determined.

The limit we have to take is

$$\lim_{N\to\infty}\lim_{P\to\infty} F_N^{[P/P]}(\lambda)=F(\lambda). \tag{IV. 2}$$

It is not possible to invert these two limits even with the use of PA instead of the perturbation series. However, we shall see that we can *connect* N and P *by a variational principle*, in such a way that P becomes a well-defined function of N, such that

$$P\to\infty, \quad N(P)\to\infty \tag{IV. 3}$$

and

$$F(\lambda)=\lim_{P\to\infty} F_{N(P)}^{[P/P]}(\lambda). \tag{IV. 4}$$

Before explaining this technique, let us see what are the limits for fixed P of the approximations:

$$\lim_{N\to\infty} F_N^{[P/P]}(\lambda) \quad \text{or} \quad \lim_{N\to\infty} F_N^{[P-1/P]}(\lambda).$$

— The thermodynamical limit of the [0/0] approximation: From (III. 21) taking the limit $N\to\infty$, we get

$$F(0)+(\lambda/\beta)a<F(\lambda), \tag{IV. 5}$$

where a is the limit $a(N)/N$ which exists for physical situation.[8] We therefore obtain our lower bound for the true free energy, which is linear in the coupling λ.

— The thermodynamical limit of the [0/1] approximation: From (III. 23) we get, in the limit $N \to \infty$,

$$F(0) + \lambda F'(0) > F(\lambda). \tag{IV. 6}$$

This bound is equivalent to the Gibbs–Bogolioubov bound and gives rise to the concavity property of $F(\lambda)$ in λ. One can see making use of the extensitivity property of the cumulants that the other PA will give in the limit $N \to \infty$ either the bound (IV. 5) for the $[P/P]$ or the bound (IV. 6) for the $[P-1/P]$. This is not surprising, because it is due to the nonuniform convergence of the approximations in the limit $N \to \infty$. To set useful sets of converging bounds, we must now use the *variational* properties of our bounds in the number of particles N.

Those variational properties are deduced from the following theorem [9]:

Theorem 5: Let

$$A^{(1)}(N) < A^{(2)}(N) < \cdots < A^{(P)}(N) < \cdots < A(N) \tag{IV. 7}$$

be a monotonic sequence of converging lower bounds to $A(N)$ for any $N > 0$.

Suppose, for $N > N_0$, $A(N)$ reaches its sup for $N = \infty$,

$$A(\infty) = A.$$

Then, for $N > N_0$, $\bar{A}^{(P)}(N)$ has a sup in N: $\bar{A}^{(p)}$ which

$$\bar{A}^{(1)} < \bar{A}^{(2)} < \cdots < \bar{A}^{(P)} < \cdots < A \tag{IV. 8}$$

and

$$\lim_{P \to \infty} \bar{A}^{(P)} = A. \tag{IV. 9}$$

Therefore, we are able due to the monotonicity of the set of approximations to extract a variational subsequence which converges towards the exact ($N \to \infty$) limit.

A corresponding theorem can be derived also, obviously, for monotonic sequences of converging upper bounds provided $A(N)$, this time, reaches its inf for $N = \infty$, when $N > N_0$.

For a certain number [8,3] of physical problems the free energy appears to be a monotonic function of the number of particles. For such systems we can directly apply the previous technique and look for extremal values of the approximated free energy per particle $F_N^{[P-1/P]}(\lambda)$ in N, this will give an upper bound to the true $F(\lambda)$ for the case where $F_N(\lambda)$ reaches its inf at $N = \infty$. For the case, where $F_N(\lambda)$ reaches its sup at $N = \infty$, we shall use the extremal value of $F_N^{[P/P]}(\lambda)$ in N to get a lower bound to $F(\lambda)$. The interest of this method is that it is a *convergent* algorithm for *any temperature, density or coupling*. The price to pay to have these nice convergence properties in the coupling constant, temperature, and density is to work in a framework of a finite number of particles. In fact, if one uses periodical boundary conditions for the statistical system (such as to consider the system on a torus) it is very likely that $F_N(\lambda)$ as a function of N is extremely flat round $N = \infty$ (all derivatives in N equal to zero) and therefore extremely good numerical results can be achieved even with low approximations, that is, with low values of N.

It will be the object of forthcoming papers to illustrate by physical example the usefulness of such methods.

Extension of the method to singular interactions

In the previous section, we have treated the case where the perturbation series of the partition function for a finite number of particles N was existing and the moment problem determinate, that is, for bounded perturbations. In practice one has often to face the problem of an unbounded perturbation operator associated with hard cores or arbitrary singular interactions. In such cases the moment problem could become undeterminate or the perturbation series itself may not exist any more.

We shall show how the method we propose adapts itself to such a case.

Let us consider an N-body local interaction via a two body singular potential; for instance,

$$V_I(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \sum_{i<j} V_{(2)}(|\mathbf{r}_i - \mathbf{r}_j|), \tag{IV. 10}$$

where $V_2(r)$ is an arbitrarily singular potential in the origin bounded from below (which corresponds to the physically interesting case).

Let us regularize $V_2(r)$ in the following way:

$$V_{(2)}^{(\epsilon)}(r) = \begin{cases} V_2(\epsilon), & r \leq \epsilon, \\ V_2(r), & r \geq \epsilon. \end{cases} \tag{IV. 11}$$

It is easy to see that for $0 < \epsilon_1 < \epsilon_2 < \epsilon_0$

$$V_{(2)}^{(\epsilon_1)} > V_{(2)}^{(\epsilon_2)}. \tag{IV. 12}$$

The inequality being taken in the sense of operators. Then defining the regularized N-body perturbation interaction by

$$V_1^{(\epsilon)}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \sum_{i<j} V_2^{(\epsilon)}(|\mathbf{r}_i - \mathbf{r}_j|), \tag{IV. 13}$$

we see that in the sense of operators

$$V_I^{(\epsilon_1)} > V_I^{(\epsilon_2)} \quad 0 < \epsilon_1 < \epsilon_2 < \epsilon_0. \tag{IV. 14}$$

We have then [10]

$$\begin{aligned} Z_N(\lambda, \epsilon_1) &= \mathrm{Tr} \exp[-\beta(H_0 + \lambda V_I^{(\epsilon_1)})] \\ &< Z_N(\lambda, \epsilon_2) = \mathrm{Tr} \exp[-\beta(H_0 + \lambda V_I^{(\epsilon_2)})], \\ &0 < \epsilon_1 < \epsilon_2 < \epsilon_0. \end{aligned} \tag{IV. 15}$$

That is, the partition function appears as an *increasing* function of the ultraviolet cutoff ϵ, for ϵ sufficiently small.

We can now treat the "ultraviolet" cutoff ϵ on the same footing as the number of particles N in the previous section.

The regularized free energy per particle $F_N(\lambda, \epsilon)$ depends now on two cutoffs, N the number of particles and ϵ the ultraviolet cutoff. The true free energy is the limit

$$\lim_{N \to \infty} \lim_{\epsilon \to 0} F_N(\lambda, \epsilon) = F(\lambda). \tag{IV. 16}$$

However, $F(\lambda)$ is *also* by the previous argument

$$F_N = \lim_{N\to\infty} \sup_{\epsilon\to 0} F_N(\lambda,\epsilon). \tag{IV.17}$$

Let us suppose $F_N(\lambda,\epsilon)$ reaches its sup for $N=\infty$; then

$$F(\lambda) = \sup_N \sup_\epsilon F_N(\lambda,\epsilon), \quad N>N_0,\ \epsilon<\epsilon_0, \tag{IV.18}$$

and, by Theorem 5, we are able to extract from the sequence of $[P/P]$ generalized Padé approximations built on the knowledge of only a finite number of regularized terms of the perturbative expansion in λ of the partition function, a variational set of monotonously converging lower bounds to $F(\lambda)$.

It is not very difficult to treat the case where $F_N(\lambda,\epsilon)$ reaches its inf for $N=\infty$, in an analogous way, by multiplying the partition function by a suitable ϵ dependent factor.

Conclusion

In this paper, we have shown how the positivity properties of the partition function in the coupling constant lead to a new method for approximating it. This method enjoys the remarkable properties of converging for all temperatures, all densities, and all values of the coupling, still being built up from the standard perturbation series in the coupling, or its regularized version when the interaction is singular and the perturbation expansion does not exist. Furthermore, the approximation is achieved through monotonous sequences of increasing and decreasing bounds to the partition function.

To reach the thermodynamical limit, it is necessary to consider the variational properties that these monotonic bounds exhibit in the number of particles N (and in the cutoff ϵ when the interaction is singular). We then extract from the previous bounds monotonic sequences of converging upper (or lower) bounds to the free energy in the thermodynamical limit, which converge for any temperature, density, and value of the coupling.

Outlook

If one considers now the case of the Euclidean field theory, one deals with a double regularization: one for the volume divergences and the other for the ultraviolet ones. As we have shown for the statistical mechanical systems, it may be possible to construct PA which will give rise to variational properties both in the volume and the UV divergencies. This has been done explicitly for the UV divergencies coming from the theory of singular potentials in Ref. 9.

However, due to the lack of any existence theorem, it is not possible, for the moment, to use this technique in a rigourous way to achieve renormalization. We can, however, describe, shortly, what would be the procedure. In this approach, one first compute, PA for physical quantities at a given order P, in terms of the cutoff

$$G^{(P)}(\lambda, C),$$

where (P) is the order of PA, λ are the bare parameters, C the cutoffs. Then the cutoffs are fixed by requiring

$$\frac{\partial G^{(P)}(\lambda, C)}{\partial C} = 0. \tag{IV.19}$$

This fixes the "best" cutoff for the given order of approximations (P). Then the bare parameters λ, and the cutoffs C are eliminated with the help of the set of equations (IV. 19) among sufficient physical quantities G. This scheme would lead to have a renormalization procedure *simultaneous* with the summations of the perturbation series, in contradistinction with the ordinary process in which one renormalizes first and then is faced with the problem of summing up a divergent series. Furthermore, the difference between renormalizable and nonrenormalizable theories disappears in this scheme. Such scheme has been numerically tested for the four fermions interaction with zero mass.[11]

ACKNOWLEDGMENTS

We are grateful to Dr. R. Balian, Dr. J. Bros, Dr. C. Itzykson, and Dr. M. L. Mehta for fruitful discussions and to Professor M. Froissart for deep considerations on the problem. We are indebted to Dr. des Cloizeaux for a careful reading of the manuscript. One of us (M. V.) would like to thank Professor C. de Dominicis for the kind hospitality he provided him in the Saclay Theoretical Division.

APPENDIX

We want to prove the following:

Theorem: Let $d\mu_1, d\mu_2, \ldots, d\mu_n$ be positive measures having Laplace transforms $Z_1(\lambda), Z_2(\lambda), \ldots, Z_n(\lambda)$ Suppose that, for any positive or null λ, $Z_n(\lambda)\to Z(\lambda)$. (simple convergence) and that the support of all the measures $d\mu_n$ are contained in $(0,\infty)$. Then $Z(\lambda)$ is the Laplace transform of a positive measure $d\mu$ which is the limit of the $d\mu_n$ and the support of $d\mu$ is contained in $(0,\infty)$.

Proof: We first prove that $Z(\lambda)$ is the Laplace transform of a positive measure $d\mu$: $Z_n(\lambda)$ is holomorphic in $\mathrm{Re}\lambda>0$. Furthermore, $Z_n(\lambda)$ is uniformly bounded in $\mathrm{Re}\lambda>0$ because

$$|Z_n(\lambda)| < \int_0^\infty \exp(-\tau\,\mathrm{Re}\lambda)\,d\mu_n < \int_0^\infty d\mu_n = Z_n(0). \tag{A1}$$

But for $n>n_0$

$$Z(0)-\epsilon < Z_n(0) < Z(0)+\epsilon. \tag{A2}$$

Then for $n>n_0$

$$|Z_n(\lambda)| < C = Z(0)+\epsilon, \quad \mathrm{Re}\lambda>0. \tag{A3}$$

By Vitali's theorem, it follows due to the simple convergence for $\lambda>0$ of the $Z_n(\lambda)$ towards $Z(\lambda)$ that:

$Z_n(\lambda)\to Z(\lambda)$ uniformly on arbitrary compacts in $\mathrm{Re}\lambda>0$ and that $Z(\lambda)$ is holomorphic in $\mathrm{Re}\lambda>0$.

As a consequence

$$\frac{d^P Z(\lambda)}{d\lambda^P} - \frac{d^P Z_n(\lambda)}{d\lambda^P} = \frac{p!}{2i\pi}\oint_C \frac{[Z(u)-Z_n(u)]}{(u-\lambda)^{P+1}}\,du. \tag{A4}$$

Taking as circuit of integration a circle of radius R around λ, we have

$$\left|\frac{d^P Z(\lambda)}{d\lambda^P} - \frac{d^P Z_n(\lambda)}{d\lambda^P}\right| < \frac{p!}{R^P}\sup_C |Z(u)-Z_n(u)|, \tag{A5}$$

which in $\mathrm{Re}\lambda>0$ tends to zero for $n\to\infty$.

Therefore, the derivatives of $Z_n(\lambda)$ tend to the derivatives of $Z(\lambda)$ in $\mathrm{Re}\lambda > 0$.

Now, by using the Bernstein theorem[12] which states that the necessary and sufficient condition for a function to be the Laplace transform of a positive measure with support contained in $(0,\infty)$ is to be C_∞ on the interval $[0,\infty]$ with derivatives having alternating signs, it results that $Z(\lambda)$ is actually the Laplace transform of a positive measure $d\mu$.

We now proceed to prove that the $d\mu_n$ tend towards $d\mu$. It is clear that if the $d\mu_n$ converge, it is towards $d\mu$ because the inverse Laplace transform is unique up to trivial measure zero changes.

Let us introduce the Fourier transforms of the positive distributions $\exp(-\lambda\tau)\,d\mu_n(\tau)$ (λ real positive fixed):

$$Z_n(\lambda+i\rho)=\int_{-\infty}^{\infty}\exp(-i\rho\tau)\exp(-\lambda\tau)\,d\mu_n(\tau),\quad \rho>0 \tag{A6}$$

(the support of $d\mu_n$ is $0,\infty$).

The *uniform* convergence of $Z_n(\lambda+i\rho)$ on arbitrary compacts in ρ, together with the bound (A3) implies the convergence of this sequence viewed as tempered distributions (convergence in the sense of S'). By Fourier transformation this gives the convergence of the $\exp(-\lambda\tau)\,d\mu_n(\tau)$ towards a limit distribution positive, therefore a positive measure. This convergence, proved for the test functions in S, can be extended to the continuous test functions, by using once again bound (A3).

We have used in the text a slightly generalized version of this theorem. In fact the support of $d\mu_n$ and $d\mu$ are all contained in an interval (b,∞), where b can be negative and finite. The convergence of $d\mu_n$ to $d\mu$ is obtained by an obvious change of variable to bring back the support to $(0,\infty)$.

Notes added in proof: M. Froissart has found an explicit example showing that the conjecture of paragraph II does not hold for the most general 3×3 matrix. However the conjecture seems to remain valid for the trace.

Theorem 3 obviously extends to nondiagonal elements in the position representation. This allows us to prove (I.3) in the boson case. Thanks to R. Balian for this remark.

*On leave from absence from Bari University. Permanent address: Istituto di Fisica, Università di Bari, Italy.

[1]D. Bessis and M. Villani, J. Math. Phys. **16**, 462 (1975).
[2]J. C. Wheeler and R. G. Gordon, "Bounds for averages using moment," in *Padê Approximant in Theoretical Physics*, edited by G. Baker, Jr., and J. Gammel (Academic, New York and London, 1970), Vol. 71 of Mathematics and Science and Engineering.
[3]G. Baker Jr., J. Math. Phys. **13**, 1862 (1972).
[4]E. Nelson, J. Math. Phys. **5**, 332 (1964).
[5]H. H. Stone, *Linear Transformations in Hilbert Space* (American Mathematical Society, New York, 1932).
[6]G. Baker, in *Cargèse Lectures in Physics*, edited by D. Bessis (Gordon and Breach, New York, 1970).
[7]D. Bessis, "Topics in the Theory of Padê Approximations," in *Padê Approximations*, edited by P.R. Graves-Morris (The Institute of Physics, London and Bristol, 1973).
[8]D. Ruelle, Helvetica Physica Acta **36**, 183 (1963); M.E. Fisher, in *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, 1964), Vol. VIIC.
[9]D. Bessis, L. Epele, and M. Villani, J. Math. Phys. **15**, 2071 (1974).
[10]D. Ruelle, Helvetica Physica Acta **36**, 789 (1963).
[11]L. Nitti, M. Pellicoro, and M. Villani, Lett. Nuovo Cimento **3**, 541 (1972).
[12]W. F. Donoghue, Jr., *Monotone Matrix Functions and Analytic Continuations* (Springer-Verlag, Berlin, 1974).

Normal coordinates and quantization of nonlinear fields by Feynman's path integral*

C. S. Hsue

Institute of Physics, National Tsing Hua University, Hsin-Chu, Taiwan, Republic of China
(Received 12 May 1975)

A simple derivation for the quantization of nonlinear fields by Feynman's path integral method is given in terms of normal coordinates. The origins of the additional term obtained by DeWitt can be clearly seen in this formalism.

I. INTRODUCTION

Feynman's path integral[1] has been used for the quantization of nonlinear fields. The result differs from the "usual Schrödinger equation" in that there is an additional term proportional to the total curvature R of the coordinate space defined with a geometry given by the kinetic energy tensor. In a curved space or in cases of constraints where R is not a numerical constant, the presence of this additional term would change the energy spectrum of the whole system. This result has been given by De Witt[2] and K.S. Cheng.[3] However, their derivations are tedious and lengthy, and it seems amazing that such a simple result is obtained by a miracuous cancellation of complicated expressions. In this report we present a simple derivation by using normal coordinates which also exhibits clearly the geometrical origin of this beautiful result.

In Sec. III we review the Hamiltonian path integral method[4] and give a derivation for the manifestly covariant formulation used. Discussion and conclusions follow in Sec. IV.

II. DERIVATION OF THE SCHRÖDINGER EQUATIONS

Let us consider a given mechanical system described by a set of coordinates $q=(q^1,\ldots,q^N)$, with the Lagrangian

$$L(\dot{q}(t),q(t))=\tfrac{1}{2}g_{ij}(q(t))\dot{q}^i\dot{q}^j. \tag{1}$$

Following Ref. 1, the quantization of the system is achieved through the formula

$$\psi(q(t+\epsilon),t+\epsilon)=\frac{1}{A}\int\exp(i/\hbar)S(g(t+\epsilon),q(t))\times\psi(q(t),t)\sqrt{g(q(t))}\,dq(t) \tag{2}$$

where $\psi(q(t+\epsilon),t+\epsilon)$ and $\psi(q(t),t)$ are, respectively, wavefunctions at time $t+\epsilon$ and t, $S(q(t+\epsilon),q(t))$ is the classical action, that is,

$$S(q(t+\epsilon),q(t))=\text{minimum of }\int_t^{t+\epsilon}L(\dot{q}(t'),q(t'))dt' \tag{3}$$

with the boundary conditions

$$q(t')\big|_{t'=t}=q(t),\quad q(t')\big|_{t'=t+\epsilon}=q(t+\epsilon). \tag{4}$$

A is the normalization factor to be determined later and g is the determinant of (g_{ij}). The introduction of the factor $g(q(t))$ guarantees that the integral is invariant under any transformation of the generalized coordinates $q^i(t)$. This property is very important. It enables us to use normal coordinates[5] and simplify the whole calculation. In Sec. III we review the Hamiltonian path integral method and give a derivation for it. From Eq. (2) we can derive the Schrödinger equation by taking the limit $\epsilon\to 0$. Now as $\epsilon\to 0$, the factor $\exp[(i/\hbar)S(q(t+\epsilon),q(t))]$ oscillates very rapidly. Only the vicinity of the stationary point

$$q(t)=q(t+\epsilon) \tag{5}$$

of $S(q(t+\epsilon),q(t))$ contributes to the integral in Eq. (2). Let us now introduce the normal coordinates at $q(t+\epsilon)$ and expand in powers of

$$\Delta q^i(t)=q^i(t)-q^i(t+\epsilon). \tag{6}$$

We have

$$g_{ij}=\delta_{ij}+\tfrac{1}{3}R_{hijk}\Delta q^h\Delta q^k+\cdots \tag{7}$$

where the Riemann symbols R_{hijk} are independent of Δq. To calculate the classical action $S(q(t+\epsilon),q(t))$ we observe that the equations of motion for the paths $q(t')$ which minimize the classical action S,

$$\ddot{q}^h+\begin{Bmatrix}h\\ij\end{Bmatrix}\dot{q}^i\dot{q}^j=0, \tag{8}$$

are geodesics. Therefore the Lagrangian (1) is a constant on the paths.[6] From Eq. (7) we easily obtain for the classical action

$$S(q(t+\epsilon),q(t))=\epsilon\dot{q}^i\dot{q}^i=\frac{\Delta q^i\Delta q^i}{\epsilon}+O(\epsilon^3). \tag{9}$$

We also need the expansions

$$\sqrt{g(q(t))}=1+\tfrac{1}{6}R_{hk}\Delta q^h\Delta q^k+\cdots, \tag{10}$$

$$\psi(q(t),t)=\psi(q(t+\epsilon)t)-\Delta q^i\frac{\partial\psi}{\partial t}+\tfrac{1}{2}\Delta q^i\Delta q^j\frac{\partial^2\psi}{q^iq^j}+\cdots \tag{11}$$

where we have introduced the Rici tensor

$$R_{ij}=R^h_{ijh}. \tag{12}$$

Equation (2) now becomes, after we perform the Gaussian integration,

$$\psi(q(t+\epsilon),t)+\epsilon\frac{\partial\psi}{\partial t}+\cdots=\frac{(i\pi\hbar\epsilon)^{N/2}}{A}\,\psi(q(t+\epsilon),t)+i\hbar\epsilon\left[\frac{1}{2}\frac{\partial}{\partial q^m}\left(\frac{\partial\psi}{\partial q^m}\right)+\frac{\hbar^2R}{6}\psi\right]+\cdots \tag{13}$$

where

$$R=g^{ij}R_{ij}. \tag{14}$$

Comparing the coefficients up to order ϵ in Eq. (13), we obtain

$$A=(i\pi\hbar\epsilon)^{N/2} \tag{15}$$

and

$$i\hbar\frac{\partial\psi}{\partial t}=-\frac{\hbar^2}{2}\nabla^2\psi-\frac{\hbar^2}{6}R\psi. \tag{16}$$

Equation (16) is the "Schrödinger equation" using Feynman's path integration formulation of quantum mechanics. Its covariant form is

$$ih\frac{\partial\psi}{\partial t}=-\frac{1}{2}\frac{\hbar^2}{\sqrt{g}}\frac{\partial}{\partial q^m}\left(\sqrt{g}\,g^{mn}\frac{\partial\psi}{\partial q^n}\right)-\frac{\hbar^2R}{6}\psi. \tag{17}$$

III. HAMILTONIAN PATH INTEGRAL

In this section we derive the covariant formulation (2) from the Hamiltonian path integral method. This method is related to the following variational principle in classical dynamics.

Let us consider a system described by N coordinates $q_1,\ldots,q_N$ and their canonical momenta $p_1,\ldots,p_N$. The system has a Hamiltonian $H(q,p)$. The variational principle is as follows: Given that the coordinates of the system at times t' and t'' are, respectively, $q_1',\ldots,q_N'$ and $q_1'',\ldots,q_N''$ we consider the set of all phase-space trajectories $q(t)$, $p(t)$ which satisfy the given boundary conditions with no restrictions on the energy and the momenta. The solution of Hamilton's equations of motion are the paths which makes

$$S=\int_{t'}^{t''}\left(\sum_j p_j\dot{q}_j-H(q(t),p(t))\right)dt \tag{18}$$

an extremum. The quantization of the system is then achieved by

$$\psi(q(t+\epsilon),t+\epsilon)=\int\exp\left[\left(\frac{i}{\hbar}\right)\int_t^{t+\epsilon}[p_j dq_j-H(q(t)p(t))\,dt]\times\psi(q(t),t)\,dp(t)\,dq(t)\right]. \tag{19}$$

For the nonlinear system (1) the Hamiltonian is given by

$$H(q,p)=\tfrac{1}{2}g^{ij}(q)p^ip^j \tag{20}$$

where g^{ij} is the inverse of g_{ij}. Inserting this into Eq. (19) and carrying through the integrations in p, we obtain the desired Eq. (2). As the phase space volume element is invariant under any canonical transformations, this formulation is manifestly covariant.

IV. DISCUSSION

(a) The introduction of normal coordinates makes the whole calculation extremely simple. It also shows that the phase factor has nothing to do with the additional term. The term proportional to R comes from the geometry of the curved space, i.e., the determinant g.

(b) In quantization of nonlinear relativistic fields such as the gravitational field and non-Abelian gauge fields, similar techniques should be taken into account.

(c) As the quantities involved are covariant scalars, the quantization of nonlinear fields is unambiguous and unique. It agrees with the "usual Schrödinger equation" only when the curvature R vanishes.

*Supported by National Council of Science of the Republic of China.

[1]R. P. Feynman, Rev. Mod. Phys. **20**, 327 (1948).
[2]B. S. DeWitt, Rev. Mod. Phys. **29**, 377 (1957).
[3]K. S. Cheng, J. Math. Phys. **13**, 1723 (Princeton U. P., Princeton, N.J., 1943), 2nd ed. (1972).
[4]C. Garrod, Rev. Mod. Phys. **38**, 483 (1966). See also Ref. 1.
[5]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N. J., 1946), 2nd ed.
[6]This can be checked directly by using (7).

Spectral representation and decomposition of self-adjoint operators*

Bernard M. de Dormale and Henri-François Gautrin

Department of Mathematics, Université de Montréal, Montréal 101, Canada
(Received 16 May 1975)

This paper is an extension of results established by Jauch and Misra [Helv. Physica Acta **38**, 30 (1965)] concerning finite or countable sets of commuting self-adjoint operators. We have obtained the following results: let $\mathcal{A} = \{A_i\}_{i\in I}$ be a set of commuting self-adjoint operators on a separable Hilbert space $\mathcal{H}$. Then (i) for any I (possibly noncountable), there exists a spectral representation for $\mathcal{A}$ iff $\mathcal{A}''$ is maximal Abelian. (ii) If I is finite or countable, $\cap_{i\in I, n\in\mathbb{N}} \mathcal{D}(A_i^n)$ is dense in $\mathcal{H}$. As a corollary of a theorem of Maurin [Bull. Acad. Pol. Sci., Sér. Sci. Math. Astron. Phys. **7**, 471 (1959)], this implies the existence of a common complete set of generalized eigenvectors.

I. INTRODUCTION

In this paper, we intend to study the families of strongly commuting selfadjoint operators on a separable Hilbert space $\mathcal{H}$. More precisely, we establish the complete equivalence of two notions related to such a family: completeness and existence of a spectral representation.

The interest of such a work has been put forward by Jauch and Misra[1]: The notion of completeness is of prime importance in quantum mechanics and spectral representation is of great interest in the study of families of commuting operators (see Ref. 1 for advantages on Gelfand representation; note also that spectral representations are related to the study of commutation relations[2,3]).

In the second section, we set up the proper mathematical framework for the problem and summarise the results obtained mainly by Jauch and Misra.[1,4] In the third part, we establish the complete equivalence between the three notions. Then, we apply these results in the last part to prove the existence of a common spectral decomposition.

II. DEFINITIONS AND GENERALITIES

In the whole paper, we shall adopt the following conventions: β is the σ-algebra of Borel sets of $\mathbb{R}$, $\mathcal{H}$ is a *separable* Hilbert space, A_i are self-adjoint operators on $\mathcal{H}$ with spectral measure E_i: $B \to E_i(B)$, $B\in\beta$. We recall that A_i and A_j are said to be (strongly) commuting if

$$\forall B_1, B_2 \in \beta,\ E_i(B_1)E_j(B_2) = E_j(B_2)E_i(B_1).$$

Let $\mathcal{A} = \{A_i\}_{i\in I}$ be a family of commuting selfadjoint operators (unless otherwise stated, the index set I may be noncountable). The commutant $\mathcal{A}'$ of $\mathcal{A}$ is the set of all bounded operators C on $\mathcal{H}$ such that $CA_i \subseteq A_i C$ for all A_i. $\mathcal{A}'' = (\mathcal{A}')'$ is called the von Neuman algebra generated by $\mathcal{A}$. One may prove[5] that for any bounded operator C

$$CA_i \subseteq A_i C \Leftrightarrow CE_i(\Delta) = E_i(\Delta)C, \quad \forall \Delta =]a, b] (-\infty \leqslant a < b \leqslant \infty).$$

Thus, $\mathcal{A}''$ is also the von Neuman algebra generated by the set $\mathcal{E}$ of linear combinations of finite products of operators $E_i(\Delta)$. $\mathcal{E}$ being a $*$-algebra, $\mathcal{A}''$ coincides with both the strong and weak closure of $\mathcal{E}$ in $\mathcal{L}(\mathcal{H})$.[6]

Let us define the various notions enumerated in the introduction:

Definition 1. (Jauch[4]): We say that the family of self-adjoint operators $\mathcal{A}$ is complete iff the von Neumann algebra $\mathcal{A}''$ generated by $\mathcal{A}$ is maximal Abelian in $\mathcal{L}(\mathcal{H})$.

In the sequel, we shall need two different definitions of a cyclic vector:

Definition 2a: $g\in\mathcal{H}$ is a cyclic vector for the von Neuman algebra $\mathcal{A}''$ iff $\mathcal{A}''(g) = \{Cg \mid C\in\mathcal{A}''\}$ is dense in $\mathcal{H}$.

Definition 2b: $g\in\mathcal{H}$ is a cyclic vector for the set of operators $\mathcal{A}$ iff $\mathcal{E}(g)$ is dense in $\mathcal{H}$, i. e., if vectors such that

$$\prod_{i\in\alpha} E_i(\Delta_i)g, \quad \Delta_i =]a_i, b_i], \quad \alpha \text{ finite subset of } I$$

generate a dense subspace of $\mathcal{H}$.

We recall that if I is any set of indices, the product $\mathbb{R}^I$ is the set of all the functions X: $I \to \mathbb{R}$. The canonical projections π_i ($i\in I$) are defined by $\pi_i(x) = x(i)$. Let $\otimes_{i\in I}\beta$ be the smallest σ-algebra such that π_i be measurable for every i, i. e., the σ-algebra generated by the sets $\pi_i^{-1}(B)$ where i runs over I and B over β. We call $\otimes_{i\in I}\beta$ the product σ-algebra on $\mathbb{R}^I$. We may then define a spectral representation by:

Definition 3: We say that we have a spectral representation of $\mathcal{A} = \{A_i\}_{i\in I}$ if there is a unitary isomorphism U between $\mathcal{H}$ and $L^2(\mathbb{R}^I, \mu)$ where μ is a measure on $\otimes_{i\in I}\beta$, such that if $f\in\mathcal{D}(A_i)$, then

$$U(A_i f)(x) = \pi_i(x)U(f)(x).$$

Note that the only difference between the definition of Jauch and Misra[1] and ours lies in the fact that they use a measure defined on the σ-algebra generated by the "Borel rectangles" $\Pi_{i\in I}B_i$. But they did not investigate the possibility of a noncountable set of operators. One may verify that if I is finite or countable, the product σ-algebra and the one generated by the Borel rectangles are both identical to the σ-algebra of Borel sets. Our definition is thus a natural generalisation of theirs.

Let us now summarise the results obtained by Jauch and others as well as some obvious facts.

It is well known[4] that $\mathcal{A}''$ is maximal Abelian iff there exists a cyclic vector for it. There is thus a complete equivalence between the completeness of $\mathcal{A}=\{A_i\}_{i\in I}$ and the existence of a cyclic vector for $\mathcal{A}''$.

Since $\mathcal{E}\subset\mathcal{A}''$, it is trivial that if $g\in\mathcal{H}$ is cyclic for $\mathcal{A}$, it is also cyclic for $\mathcal{A}''$. The converse is also true: $\mathcal{A}''$ being the closure of $\mathcal{E}$ in the strong topology, $\overline{\mathcal{E}(g)}\supseteq\mathcal{A}''(g)$ for every $g\in\mathcal{H}$. Thus, Definitions 2a and 2b are fully equivalent.

Morever, Prugovečki[2,7] has shown that, in the finite case, $\mathcal{A}$ has a spectral representation iff it has a cyclic vector. In the countable case, Jauch and Misra[1] have proven that the completeness implies the existence of a spectral representation, under a certain restrictive assumption. We shall see that this assumption is unnecessary. (It comes from the fact that they tacitly assumed that if g is cyclic for $\mathcal{A}$, then $g\in\mathcal{D}(A_i)$ for every A_i: Most of the time, this is false.)

In order to prove the complete equivalence between the various notions, we shall now establish that there is a spectral representation for $\mathcal{A}=\{A_i\}_{i\in I}$ if $\mathcal{A}$ has a cyclic vector, whatever is the cardinality of I.

III. EXISTENCE OF A SPECTRAL REPRESENTATION

We shall now prove the equivalence between the existences of a cyclic vector and of a spectral representation.

Proposition 1: Suppose $\mathcal{A}$ has a cyclic vector. Then, there is a spectral representation for $\mathcal{A}$.

In order to prove this proposition, we need the following lemmas:

Lemma 1: Let $(X,\mathcal{C})$ be a measurable space and let μ be a finite measure on $\mathcal{C}$. If $\mathcal{D}$ is an algebra of sets generating $\mathcal{C}$, (i.e., if $A,B\in\mathcal{D}$, then $A\cap B\in\mathcal{D}$; if $A\in\mathcal{D}$, then $A^c\in\mathcal{D}$ and $\mathcal{C}$ is the smallest σ-algebra containing $\mathcal{D}$) then $F=\{\sum_{i=1}^n c_i\delta_{D_i}\,|\,D_i\in\mathcal{D}\}$ is dense in $L^2(X,\mu)$.

Proof: Let $D\in\mathcal{C}$, by a theorem due to Carathedory,[8,9] we know that for every $\epsilon>0$, there exists a sequence of sets $D_i\in\mathcal{D}$ such that

$$D\subset\bigcup_{i=1}^{\infty}D_i,\quad D_i\cap D_j=\phi\quad\text{if } i\neq j,$$

$$\sum_{i=1}^{\infty}\mu(D_i)-\mu(D)<\epsilon^2.$$

One verifies that $\delta_{\cup_{i=1}^{\infty}D_i}=\lim\sum_{i=1}^{n}\delta_{D_i}$ is in $\overline{F}$ and that $\|\delta_D-\delta_{\cup_{i=1}^{\infty}D_i}\|<\epsilon$. Thus, $\delta_D\in\overline{F}$. Since simple functions (i.e., linear combinations of characteristic functions of measurable sets) are dense in $L^2(X,\mu)$,[8] we have $\overline{F}=L^2(X,\mu)$.

Lemma 2: Let α be a finite subset of I:

(i) For every $g\in\mathcal{H}$ with $\|g\|=1$, there is a normed Borel measure μ_α on $\mathbb{R}^\alpha$ which extends the set function

$$\prod_{i\in\alpha}\Delta_i\mapsto\langle g,\prod_{i\in\alpha}E_i(\Delta_i)g\rangle,\quad\Delta_i=]a_i,b_i].$$

(ii) The correspondence

$$\prod_{i\in\alpha}E_i(\Delta_i)g\mapsto\delta_{\prod_{i\in\alpha}\Delta_i}$$

defines a unitary isomorphism U_α between a closed subspace $\mathcal{H}_\alpha$ of $\mathcal{H}$ and $L^2(\mathbb{R}^\alpha,\mu_\alpha)$.

Proof: (i) The collection of sets $\Pi_{i\in\alpha}\Delta_i$ is a semi-algebra $\mathcal{J}_\alpha$, i.e., the intersection of any two sets in $\mathcal{J}_\alpha$ is a set in $\mathcal{J}_\alpha$ and the complement of any set in $\mathcal{J}_\alpha$ is a finite disjoint union of sets in $\mathcal{J}_\alpha$. $\mathcal{J}_\alpha$ is contained in an algebra $\mathcal{G}_\alpha$ whose sets are finite disjoint unions of sets in $\mathcal{J}_\alpha$[8] and which trivialy generates the σ-algebra of Borel sets of $\mathbb{R}^\alpha$. All we have to check is that μ_α is σ-additive on $\mathcal{J}_\alpha$: μ_α will then be extendable in a unique manner to $\mathcal{G}_\alpha$[8] and then to the generated σ-algebra.[7,8,9] But, if $\Delta_i=\Delta_i'=\Delta_i''$ for $i\neq k$ and $\Delta_k=\Delta_k'\cup\Delta_k''$ with $\Delta_k'\cap\Delta_k''=\phi$, we have

$$\mu_\alpha\left(\prod_{i\in\alpha}\Delta_i\right)=\mu_\alpha\left(\prod_{i\in\alpha}\Delta_i'\right)+\mu_\alpha\left(\prod_{i\in\alpha}\Delta_i''\right)$$

and if $\Delta_i^{(m)}=\Delta_i$ for $i\neq k$, $\Delta_k^{(m)}=]a_k,b_k^{(m)}]$ where a_k is the limit of the decreasing sequence $b_k^{(m)}$, then

$$\mu_\alpha\left(\prod_{i\in\alpha}\Delta_i^{(m)}\right)\to 0.$$

We may then claim that μ_α is σ-additive on $\mathcal{J}_\alpha$.[10]

(ii) It is not difficult to verify that the function defined on $\mathcal{J}_\alpha$ by $\Pi_{i\in\alpha}\Delta_i\to\Pi_{i\in\alpha}E_i(\Delta_i)g$ is additive: We may thus define a linear transformation U_α of a subspace of $\mathcal{H}$ into $L^2(\mathbb{R}^\alpha,\mu_\alpha)$ by

$$\sum_{k=1}^{n}c_k\prod_{i\in\alpha}E_i(\Delta_{i,k})g\mapsto\sum_{k=1}^{n}c_k\delta_{\prod_{i\in\alpha}\Delta_{i,k}}$$

where the products $\Pi_{i\in\alpha}\Delta_{i,k}$ may always be chosen disjoint: all we need to check is $U_\alpha(0)=0$. But it is easy to see that

$$\prod_{i\in\alpha}\Delta_{i,1}\cap\prod_{i\in\alpha}\Delta_{i,2}=\phi\Rightarrow\prod_{i\in\alpha}E_i(\Delta_{i,1})\prod_{i\in\alpha}E_i(\Delta_{i,2})=0$$

Then

$$\left\|\sum_{k=1}^{n}c_k\prod_{i\in\alpha}E_i(\Delta_{i,k})g\right\|^2=\sum_{k=1}^{n}|c_k|^2\mu_\alpha\left(\prod_{i\in\alpha}\Delta_{i,k}\right)=\left\|\sum_{k=1}^{n}c_k\delta_{\prod_{i\in\alpha}\Delta_{i,k}}\right\|^2$$

which assures that zero is sent into zero. U_α is trivialy isometric and by the use of Lemma 1, we may extend it to a unitary isomorphism of $\mathcal{H}_\alpha$ to $L^2(\mathbb{R}^\alpha,\mu_\alpha)$ where $\mathcal{H}_\alpha$ is the closure of the subspace $\{\sum c_k\Pi_{i\in\alpha}E_i(\Delta_{i,k})g\}$.

Let now π_α be the canonical projection

$$\begin{array}{ccc}\pi_\alpha:\mathbb{R}^I&\to&\mathbb{R}^\alpha\\ \cup&&\cup\\ X&\to&\{X(i)\}_{i\in\alpha}.\end{array}$$

We have the following lemma:

Lemma 3: There is a unique normed measure μ on $(\mathbb{R}^I,\otimes_{i\in I}\mathcal{B})$ such that $\mu_\alpha=\pi_\alpha(\mu)$ where all the μ_α are defined as in Lemma 2 with the same $g\in\mathcal{H}$. Moreover, there is a isometric imbedding J_α of $L^2(\mathbb{R}^\alpha,\mu_\alpha)$ in $L^2(\mathbb{R}^I,\mu)$ defined by $J_\alpha(f)=f\circ\pi_\alpha$.

Proof: If $\beta\subseteq\alpha$, we may define a projection $\pi_{\beta\alpha}:\mathbb{R}^\alpha\to\mathbb{R}^\beta$ which send $\{X(i)\}_{i\in\alpha}$ on $\{X(i)\}_{i\in\beta}$. These projections obey the equations

$\pi_{\beta\alpha}\circ\pi_\alpha=\pi_\beta$ if $\beta\subseteq\alpha$,

$\pi_{\gamma\beta}\circ\pi_{\beta\alpha}=\pi_{\gamma\alpha}$ if $\gamma\subseteq\beta\subseteq\alpha$

and are such that if $\beta\subseteq\alpha$, $\mu_\beta=\pi_{\beta\alpha}(\mu_\alpha)$, i.e.,

$\mu_\beta(B)=\mu_\alpha(\pi_{\beta\alpha}^{-1}(B))$, B: any Borel set of $\mathbb{R}^\beta$

The collection of all $(\mathbb{R}^\alpha,\mu_\alpha)$ for all finites parts α of I together with the $\pi_{\beta\alpha}$ for $\beta\subseteq\alpha$ is thus a projective system of measures. Using a theorem due to Kolmogorov,[9] we may claim that there is a unique measure on $(\mathbb{R}^I,\otimes_{i\in I}\beta)$ such that $\mu_\alpha=\pi_\alpha(\mu)$ for every α. For every linear combination f of characteristic functions of disjoint sets in $\mathcal{J}_\alpha$, put $J_\alpha(f)=f\circ\pi_\alpha: J_\alpha$ is trivialy linear and isometric. It follows from Lemma 1 that we can extend it to an isometric imbedding of $L^2(\mathbb{R}^\alpha,\mu_\alpha)$ in $L^2(\mathbb{R}^I,\mu)$.

Lemma 4: Let g be a cyclic vector for $\mathcal{A}$. Then,

$$\sum_k c_k\prod_{i\in\alpha_k}E_i(\Delta_{i,k})g\mapsto\sum_k c_k\delta_{\pi_{\alpha_k}^{-1}}\left(\prod_{i\in\alpha_k}\Delta_{i,k}\right)$$

defines a unitary isomorphism U between $\mathcal{H}$ and $L^2(\mathbb{R}^I,\mu)$ where μ is the measure occurring in Lemma 3; verifies the equality

$$(UE_i(\Delta)f)(x)=\delta_{\pi_i^{-1}(\Delta)}(x)(Uf)(x),\quad\forall\Delta=]a,b].$$

Proof: Let $f=\sum_k c_k\prod_{i\in\alpha_k}E_i(\Delta_{i,k})g$: It is easy to check that $Uf=J_\alpha U_\alpha f$ for every α such that $\alpha\subseteq\cup_k\alpha_k$. U is thus a linear isometric mapping defined on a dense subspace of $\mathcal{H}$ (because g is cyclic) whose range is dense in $L^2(\mathbb{R}^I,\mu)$ (because of Lemma 1). One may then extend U to a unitary isomorphism.

Lemma 5: Let $F_i(\Delta)$ be the operator of multication by $\pi_i^{-1}(\Delta)$ in $L^2(\mathbb{R}^I,\mu)$: The unique extension of F_i to β is the spectral measure of the self-adjoint operator A_i' defined by

$$\mathcal{D}(A_i')=\{f\in L^2(\mathbb{R}^I,\mu)\mid\int\lambda^2 d\|F_i(\lambda)f\|^2<\infty\},$$
$$(A_i'f)(x)=\pi_i(x)f(x).$$

Proof: $F_i(\Delta)=UE_i(\Delta)U^{-1}$ is trivialy σ-additive on the semialgebra of the Δ's because E_i has this property. F_i is thus extendable to β (Ref. 7) and it is not difficult to verify that this extension is given by

$$(F_i(B)f)(x)=\delta_{\pi_i^{-1}(B)}(x)f(x).$$

F_i is thus the spectral measure of a selfadjoint operator A_i' of domain[77]

$$\mathcal{D}(A_i')=\{f\mid\int\lambda^2 d\|F_i(\lambda)f\|^2<\infty\}.$$

In order to prove that $(A_i'f)(x)=\pi_i(x)f(x)$, we must show that for every f, $g\in\mathcal{D}(A_i')$, we have

$$\int_{\mathbb{R}^I}\pi_i(x)f(x)\bar{g}(x)\,d\mu(x)=\int_{\mathbb{R}}\lambda\,d\nu_i(\lambda)$$

with

$$\nu_i(B)=\langle F(B)f,g\rangle=\int_{\pi_i^{-1}(B)}f(x)\bar{g}(x)\,d\mu(x).$$

But we need only to prove this for $f=g$ because one may always write

$$f\bar{g}=\frac{1}{4}\sum_{k=1}^{4}i^k(f+i^kg)\overline{(f+i^kg)}\quad(i=\sqrt{-1}).$$

In this case, ν_i is a finite positive Borel measure on $\mathbb{R}$ and λ being ν_i integrable, there exists a sequence of a simple functions converging to λ at the same time in $L^1(\mathbb{R},\nu_i)$ and ν_i almost everywhere [i.e., $\exists\,C\in\beta$ with $\nu_i(C)=0$ and $\sum\lambda_j\delta_{B_j}(\lambda)\to\lambda$ in $\mathbb{R}\sim C$].

Furthermore, the application from $L^1(\mathbb{R},\nu_i)$ to $L^1(\mathbb{R}^I,\mu)$ defined by

$$\sum\lambda_j\delta_{B_j}\to\sum\lambda_j|f|^2\delta_{\pi_i^{-1}(B_j)}$$

is an isometry. Then, the image of the sequence of simple functions will be a Cauchy sequence in $L^1(\mathbb{R}^n,\mu)$. This sequence converges almost everywhere to $\pi_i|f|^2$. Indeed, if $X\in\mathbb{R}^I\sim\pi_i(C)$,

$$X(i)\in\mathbb{R}\sim C\Rightarrow\sum\lambda_j\delta_{B_j}(X(i))\to X(i).$$

$$\Rightarrow\sum\lambda_j\delta_{\pi_i^{-1}(B_j)}(X)|f(x)|^2\to X(i)\,|f(x)|^2=\pi_i(x)|f(x)|^2$$

and

$$\nu_i(C)=0\qquad\int_{\pi_i^{-1}(C)}|f(x)|^2\,d\mu(x)=0,$$

i.e., $\pi_i^{-1}(C)$ is the union of a μ-negligeable set and of a set on which $f(x)=0$. We have thus also

$$\sum\lambda_j\delta_{\pi_i^{-1}(B_j)}|f|^2\to\pi_i|f|^2$$

in $L^1(\mathbb{R}^I,\nu_i)$ and this proves that

$$\int\pi_i|f|^2\,d\mu=\int\lambda\,d\nu_i.$$

Proof of Proposition 1: This proposition follows from Lemmas 4 and 5:

Because $F_i=UE_iU^{-1}$, we have trivially $A_i=UA_iU^{-1}$, both operators having the same domain.

We shall now prove the converse of Proposition 1:

Proposition 2: Suppose that there is a spectral representation for $\mathcal{A}$. Then, $\mathcal{A}$ has a cyclic vector.

Proof: Let U: $\mathcal{H}\to L^2(\mathbb{R}^I,\mu)$ be a spectral representation: It is equivalent to prove that there is a cyclic vector for $\mathcal{A}=\{A_i\}_{i\in I}$ or for $U(\mathcal{A})=\{A_i'\}_{i\in I}$ with $A_i'=UA_iU^{-1}$ i.e., that there is a g in $L^2(\mathbb{R}^I,\mu)$ such that the functions $\prod_{i\in\alpha}\delta_{\pi_i^{-1}(\Delta_i)}g$ generate a dense subspace of $L^2(\mathbb{R}^I,\mu)$.

Now, because $\mathcal{H}$ and thus $L^2(\mathbb{R}^I,\mu)$ are separable, μ must be finite or σ-finite and if μ is finite, Lemma 1 tells us that $g(x)=1$ is cyclic. Let then μ be σ-finite: There is a covering of $\mathbb{R}^I$ by disjoint sets B_n of finite measure. We may define a finite measure μ' on $\otimes_{i\in I}\beta$ by

$$\mu'(B)=\sum_{n=1}^{\infty}\frac{1}{2^n}\,\frac{\mu(B\cap B_n)}{\mu(B_n)}.$$

μ and μ' being equivalent, there is a unitary isomorphism U' between $L^2(\mathbb{R}^I,\mu)$ and $L^2(\mathbb{R}^I,\mu')$ given by

$$U'(g)=g\sqrt{d\mu/d\mu'}$$

which is such that

$$U'(\delta_{\pi_i^{-1}(B)}g)=\delta_{\pi_i^{-1}(B)}U(g),$$

i.e., we just have a spectral representation for $\mathcal{A}$ with finite measure μ', for which the function equal to one is cyclic.

We may summarise the previous results in the theorem:

Theorem 1: Let $\mathcal{A}$ be any set of commuting self-adjoint operators; then the following assertions are equivalent.

(i) $\mathcal{A}$ is complete (i.e., $\mathcal{A}''$ is maximal Abelian).

(ii) There is a cyclic vector for $\mathcal{A}$ (or for $\mathcal{A}''$)

(iii) There is a spectral representation for $\mathcal{A}$.

IV. SPECTRAL DECOMPOSITION OF A COUNTABLE SET OF OPERATORS

We shall now restrict ourselves to the case of a finite or countable set of indices I. One may then use the theorem of Prokhorov[12,13] to prove that if $U:\mathcal{H}\to L^2(\mathbb{R}^I, \mu)$ is a spectral representation then μ is a Radon measure (we know yet that for finite or countable I, μ is a Borel measure, cf. Sec. II), i.e., μ has the property[12]:

$$K \text{ compact} \Rightarrow \mu(K) < \infty,$$

$$\forall B \in \bigotimes_{i\in I} \mathcal{B},\ \mu(B) = \sup_{\substack{K\subset B \\ K \text{ compact}}} \mu(K).$$

As a corollary to this, we may claim that the set of all functions f such that $\operatorname{supp} f$ is compact in the product topology on $\mathbb{R}^I$ and $f|_{\operatorname{supp} f}$ is continuous is dense in $L^2(\mathbb{R}^I, \mu)$ [or, equivalently, their equivalence classes under the equality μ a.e. is a dense subset of $L^2(\mathbb{R}^I, \mu)$].[12] But if f is such a function, for every $p\in\mathbb{N}$ and $i\in I$, $\pi_i^p f$ is also of compact support and continuous when restricted to it. This of course implies that f is in $\mathcal{D}(A_i'^p)$ for every $i\in I$ and $p\in\mathbb{N}$. We have thus proven that for any finite or countable complete set of operators $\mathcal{A}=\{A_i\}_{i\in I}$, $\cap_{\substack{i\in I \\ p\in\mathbb{N}}}\mathcal{D}(A_i^p)$ is a dense subset of $\mathcal{H}$ [it contains the image by U^{-1} of a dense subset of $L^2(\mathbb{R}^I, \mu)$]. But any finite or countable set of operators is contained in a finite or coutable complete set of operators: There are many ways to complete such a set by just one operator. One of these is: Take any maximal Abelian von Neuman algebra containing $\mathcal{A}''$. It may be generated by a single operator with simple spectrum[4,5] and if you add this operator to $\mathcal{A}$ you have a complete set. Thus, the property we have obtained is independant of the completeness:

Proposition 3: Let $\mathcal{A}=\{A_i\}_{i\in I}$ be any finite or countable set of commuting self-adjoint operators on a separable Hilbert space $\mathcal{H}$: $\cap_{\substack{i\in I \\ p\in\mathbb{N}}}\mathcal{D}(A_i^p)$ is a dense subis a dense subspace of $\mathcal{H}$. (A proof of Proposition 3 for the finite case was given in Ref. 14 and implicitly contained in Ref. 2.)

We need to recall some facts about spectral decomposition before going further; and first of all, the complete spectral theorem[7,15]:

Theorem 2 (*von Neuman*): Let $\mathcal{A}=\{A_i\}_{i\in I}$ be any set of commuting selfadjoint operators on a separable Hilbert space $\mathcal{H}$. Then, there is

(i) a direct integral

$$\hat{\mathcal{H}} = \int_\Lambda \hat{\mathcal{H}}(\lambda)\, d\nu(\lambda)$$

which is a certain family of square integrable vector fields

$$\Lambda \ni \lambda \mapsto \hat{u}(\lambda) \in \hat{\mathcal{H}}(\lambda)$$

on a locally compact space Λ the scalar product in $\hat{\mathcal{H}}$ being

$$(\hat{u},\hat{v})_{\hat{\mathcal{H}}} = \int_\Lambda \sum_{k=1}^{\dim\hat{\mathcal{H}}(\lambda)} \hat{u}_k(\lambda)\,\overline{\hat{v}_k(\lambda)}\, d\nu(\lambda),$$

(ii) a unitary map

$$\mathcal{H} \ni u \mapsto Fu = \hat{u} \in \mathcal{H}$$

such that

$$(FA_i u)_k(\lambda) = \hat{A}_i(\lambda)\,\hat{u}_k(\lambda), \quad 1 \le k \le \dim\hat{\mathcal{H}}(\lambda),$$

where the function $\hat{A}_i$ maps Λ onto the spectrum of the operator A_i.

We shall say that we have a spectral decomposition of the family $\mathcal{A}$ or that we have a complete set of generalized eigenvectors of $\mathcal{A}$ if there is a nuclear space Φ contained and dense in $\mathcal{H}$ (i.e., a dense subspace of $\mathcal{H}$ which is nuclear when equipped with a certain locally convex topology stronger than the one of $\mathcal{H}$) such that the F of Theorem 2 could be written

$$\forall \varphi \in \Phi,\ (F\varphi)_k(\lambda) = \hat{\varphi}_k(\lambda) = \langle \varphi, e_k(\lambda)\rangle, \quad 1 \le k \le \dim\hat{\mathcal{H}}(\lambda)$$

with $e_k(\lambda)\in\Phi'$.

There is a very useful theorem concerning such a spectral decomposition[15,16,17]:

Theorem 3 (*Maurin*): Let the family of operators $\mathcal{A}$ be finite or countable. If $\cap_{\substack{i=I \\ p=\mathbb{N}}}\mathcal{D}(A_i^p)$ is dense in $\mathcal{H}$, then $\mathcal{A}$ has a spectral decomposition.

If we put together Proposition 3 and Theorem 3, we obtain the following theorem:

Theorem 4: Let $\mathcal{A}$ be any finite or countable set of commuting self-adjoint operators on a separable Hilbert space $\mathcal{H}$; then there is always a spectral decomposition of $\mathcal{A}$.

V. CONCLUDING REMARKS

The main result of our paper concerning spectral representation is to extend the work of Jauch and Misra. Even in the countable case, we improve their results by showing that there is no need for a supplementary assumption.

In Sec. IV, we obtain Theorem 4 as a corollary of a theorem of Maurin and of Proposition 3; note that this proposition is interesting in itself for the study of commuting selfadjoint operators. As an example, consider the following proposition:

P: A and B are symmetric operators defined on a dense subspace D of $\mathcal{H}$ which is a common domain of essential self-adjointness for every $aA+bB$ $(a,b\in\mathbb{R})$. Moreover, $ABf = BAf$ for every $f\in D$.

In a well-known counter example, Nelson[18] has shown that P is not a sufficient condition for the commutatively of A and B; but Proposition 3 allows us to say that P is a

necessary condition. Indeed, we may consider A and B as operators of multiplication by x and y in $L^2(\mathbb{R}^3, \mu)$ where μ is finite and it is not difficult to see that the linear combinations of functions of compact support, continuous when restricted to it, constitute a common domain D of self-adjointness for every operator of multiplication by $ax+by$ $(a, b \in \mathbb{R})$, whose closure has domain

$$\{f \mid \int (ax+by)^2 |f(x,y,z)|^2 \, d\mu(x,y,z,) < \infty\}.$$

Moreover, it is trivial that $xyf(x,y,z) = yxf(x,y,z)$ for every $f \in D$.

ACKNOWLEDGMENTS

We want to thank MM. Michel Boyer and Alain Boulanger for a very useful discussion concerning projective systems of measures and the theorem of Kolmogorov. We want also to thank Professor E. Prugovečki for reading the manuscript and pointing out some errors.

*Partially supported by grants from the National Research Council of Canada and the Quebec Government.

[1]J. M. Jauch and B. Misra, Helv. Phys. Acta 38, 30 (1965).
[2]E. Prugovečki, Can. J. Phys. 47, 1083 (1969).
[3]I. M. Gelfand and N. Y. Vilenkin, *Les distributions* (Dunod, Paris, 1967), Vol. 4.
[4]J. M. Jauch, Helv. Phys. Acta 33, 711 (1960).
[5]F. Riesz and B. S. Sz.-Nagy, *Leçons d'analyse fonctionelle* (Gauthier-Villars, Paris, 1972), 6th ed.
[6]J. Dixmier, *Les algèbres d'operateur dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1969), 2nd ed.
[7]E. Prugovečki, *Quantum Mechanic in Hilbert Space* (Academic, New York, 1971).
[8]H. L. Royden, *Real Analysis* (MacMillan, London, 1968), 2nd ed.
[9]J. Neveu, *Calcul des probabilitês* (Masson, Paris, 1970), 2nd ed.
[10]H. G. Garnir, M. De Wilde, and J. Schmets, *Analyse fonctionnelle* (Birkhauser Verlag, Basel, 1972), Vol. 2.
[11]K. Yosida, *Functional Analysis* (Springer-Verlag, Berlin, 1971), 3rd ed.
[12]N. Bourbaki, *Elêments de mathêmatique* (Hermann, Paris, 1969), Vol. 5, book 6.
[13]L. Schwartz, seminar, Ecole Polytechnique, Paris, 1969-1970.
[14]B. M. de Dormale and H. F. Gautrin, "Spectral representation of commuting self-adjoint operators," Preprint, Université de Montréal.
[15]K. Maurin, Methods of Hilbert Spaces (Polish Scientific, Warsaw, 1972), 2nd ed.
[16]K. Maurin, Bull Acad. Pol. Sci. Ser. Sci. Math. Astr. Phys. 7, 471 (1959).
[17]K. Maurin, *General Eigenfunction Expansion and Unitary Representations of Topological Groups* (Polish Scientific, Warsaw, 1968).
[18]E. Nelson, Ann. Math. 70, 572 (1959).

Expansions for simple cubic lattice Green's functions

S. Jorna

Physical Dynamics, Inc., La Jolla, California 92038
(Received 16 June 1975)

Expansions in terms of elementary functions are given for the diagonal elements of the simple cubic Green's function $I(l;s)=(1/\pi^3)\int_0^\pi\int_0^\pi\int_0^\pi [\cos(l\phi_1)\cos(l\phi_2)\cos(l\phi_3)/(s+i\epsilon-\cos\phi_1-\cos\phi_2-\cos\phi_3)]\,d\phi_1 d\phi_2 d\phi_3$ in the ranges $0<s<1$, $1<s<3$, and $s>3$ for arbitrary l. The fast convergence of these expansions renders them extremely useful for the rapid determination of numerical values for $\mathrm{Re}\,I(l;s)$ and $\mathrm{Im}\,I(l;s)$ of high accuracy. More slowly converging expansions are given for $s=1$ and $s=3$. The expansion for $s=1$ is particularly useful for large l.

1. GENERAL FORMULATION

Work on the scattering of waves by periodic discrete lattices leads to the following integral for the simple cubic lattice Green function[1]:

$$I(l_1,l_2,l_3;s)=\frac{1}{\pi^3}\lim_{\epsilon\to 0}\int_0^\pi\int_0^\pi\int_0^\pi \frac{\cos(l_1\phi_1)\cos(l_2\phi_2)\cos(l_3\phi_3)\,d\phi_1 d\phi_2 d\phi_3}{s+i\epsilon-(\cos\phi_1+\cos\phi_2+\cos\phi_3)}, \tag{1}$$

where the limiting process is required to avoid singularities at $s=1$ and 3. Since

$$\alpha^{-1}=\int_0^\infty \exp(-\alpha x)\,dx, \tag{2}$$

Eq. (1) reduces after integration over ϕ_1, ϕ_2 and ϕ_3 to

$$I(l_1,l_2,l_3;s)=\frac{1}{i^{l_1+l_2+l_3+1}}\int_0^\infty J_{l_1}(x)J_{l_2}(x)J_{l_3}(x)\times\exp[i(s+i\epsilon)x]\,dx. \tag{3}$$

In this paper we will confine attention to the evaluation of I for the diagonal elements. The off-diagonal elements can then be calculated from its known recurrence relation.[2] With $l\equiv l_1=l_2=l_3$, Eq. (3) becomes

$$I(l;s)=i^{l-1}\int_0^\infty J_l{}^3(x)\exp[i(s+i\epsilon)x]\,dx. \tag{4}$$

Replacing $J_l{}^2(x)$ with its Mellin–Barnes integral representation,[3] we obtain

$$I(l;s)=\frac{i^{l-1}}{2\pi i}\int_0^\infty J_l(x)\exp(isx)\,dx\times\int_{c-i\infty}^{c+i\infty}\frac{\Gamma(-\mu)\,\Gamma(2\mu+2l+1)}{[\Gamma(\mu+l+1)]^2\,\Gamma(\mu+2l+1)}(\tfrac14x^2)^{\mu+l}\,d\mu, \tag{5}$$

with s in the lower half plane and the usual restriction on c that the integration contour must separate the poles of $\Gamma(-\mu)$ from those of $\Gamma(2\mu+2l+1)$. The integral over x is a Gauss hypergeometric function so that for Eq. (5) we may write

$$I(l;s)=\frac{(2s)^{-3l}}{2\pi i s l!}\int_{c-i\infty}^{c+i\infty}\frac{\Gamma(-\mu)\,\Gamma(2\mu+2l+1)\,\Gamma(2\mu+3l+1)}{[\Gamma(\mu+l+1)]^2\,\Gamma(\mu+2l+1)(-4s^2)^\mu}\times F(\mu+\tfrac32l+\tfrac12,\mu+\tfrac32l+1;l+1;s^{-2})\,d\mu. \tag{6}$$

This is the basic expression from which will be calculated expansions convergent for $0\le s<1$, $1<s<3$, $s>3$, $s=1$, and $s=3$.

2. EXPANSION FOR $s>3$

A double series valid for $s>3$ follows on replacing F with its infinite series which converges absolutely for $s>3$, interchanging integration and summation, and evaluating the μ-integral at the poles of $\Gamma(-\mu)$. Thus,

$$\begin{aligned}I(l;s)&=\frac{(2s)^{-3l}}{2\pi i s l!}\int_{c-i\infty}^{c+i\infty}\frac{\Gamma(-\mu)\,\Gamma(2\mu+2l+1)\,\Gamma(2\mu+3l+1)}{[\Gamma(\mu+l+1)]^2\,\Gamma(\mu+2l+1)(-4s^2)^\mu}\\&\times\frac{l!}{\Gamma(\mu+\frac32l+\frac12)\,\Gamma(\mu+\frac32l+1)}\\&\times\sum_{n=0}^\infty\frac{\Gamma(n+\mu+\frac32l+\frac12)\,\Gamma(n+\mu+\frac32l+1)}{n!\,(n+l)!}s^{-2n}\,d\mu\\&=\frac{(2s)^{-3l}}{s}\sum_{n=0}^\infty\sum_{m=0}^\infty\frac{(2m+2l)!\,(2m+2n+3l)!}{m!n!\,[(m+l)!]^2\,(m+2l)!\,(n+l)!}\\&\times(4s^2)^{-m-n},\end{aligned} \tag{7}$$

which converges for $s>3$. The imaginary part of I is zero for $s>3$ as will be evident from inspection of Eq. (1).

3. EXPANSIONS FOR $0<s<1$

The usual infinite series representation for the hypergeometric function in Eq. (6) diverges when $0<s<1$. An expansion valid in this range is readily obtained from its Mellin–Barnes integral representation [or from Kummer's relation linking $F(\ldots;z)$ and $F(\ldots;1/z)$]. Substitution in Eq. (6) leads to

$$I(l;s)=\frac{2^{2l-1}}{s(2s)^{3l}\pi^{1/2}(2\pi i)^2}\int_{-i\infty}^{i\infty}\int_{-i\infty}^{i\infty}\frac{\Gamma(-\mu)\,\Gamma(\mu+l+\frac12)\,\Gamma(t+2\mu+3l+1)}{\Gamma(\mu+l+1)\,\Gamma(\mu+2l+1)}\times\frac{\Gamma(-\frac12t)\exp(i\pi\mu)\exp(i\pi t/2)}{\Gamma(\frac12t+l+1)s^{2\mu+t}2^t}\,d\mu\,dt, \tag{8}$$

where the integration paths have the usual restrictions on the separation of poles and $|\arg(-1/s)|<\pi$. The required expansion in ascending powers of s is obtained by first evaluating the residues in the t-integral at the poles of $\Gamma(t+2\mu+3l+1)$, i.e., at $t=-m-2\mu-3l-1$ with $m=0,1,2,\ldots$, and then in the μ-integral at the simple poles of $\Gamma(\mu+l+\frac12)$, i.e., $\mu=-n-l-\frac12$, $n=0,1,\ldots,\frac12(m+l)-1$, and for even $(m+l)$ at the double

poles $\mu = -n - \frac{1}{2}m - \frac{3}{2}l - \frac{1}{2}$, $n = 0, 1, \ldots$. This yields the expansions

$(m+l)$ odd

$$\mathrm{Re}I(l;s) = \tfrac{1}{2}\pi^{-3/2}(-1)^l \times \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\Gamma(n+l+\frac{1}{2})\Gamma(n-l+\frac{1}{2})\,\Gamma(n+\frac{1}{2})}{\Gamma(n-[\frac{1}{2}m]+\beta+\frac{1}{2})\Gamma(n-[\frac{1}{2}m]-\alpha+\frac{1}{2})} \times \frac{(2s)^m}{m!n!4^n}, \tag{9}$$

where α is the largest integer not exceeding $\frac{1}{2}l$, i.e., $\alpha = [\frac{1}{2}l]$, and $\beta = [\frac{1}{2}(l+1)]$, and

$$\begin{aligned}\mathrm{Im}I(l;s) = &-\tfrac{1}{2}\pi^{-5/2}\sum_{m=0}^{\infty}\sum_{n=0}^{(m+l)/2-1}(-1)^{n-(m-l)/2} \\ &\times \frac{\Gamma(n+l+\frac{1}{2})\,\Gamma(n+\frac{1}{2})\,\Gamma(n-l+\frac{1}{2})\,\Gamma(-n+\frac{1}{2}m+\frac{1}{2}l)}{m!n!\,\Gamma(n-\frac{1}{2}m+\frac{1}{2}l+1)} \\ &\times \frac{(2s)^m}{4^n} \\ &-(\tfrac{1}{2})^{l+1}\pi^{-5/2}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \\ &\frac{\Gamma(n+\frac{1}{2}m+\frac{3}{2}l+\frac{1}{2})\,\Gamma(n+\frac{1}{2}m+\frac{1}{2}l+\frac{1}{2})\,\Gamma(n+\frac{1}{2}m-\frac{1}{2}l+\frac{1}{2})}{m!n!\,(n+l)!\,(n+\frac{1}{2}m+\frac{1}{2}l)!} \times \frac{s^m}{4^n} \\ &\times[\ln 4 - \psi(n+\tfrac{1}{2}m+\tfrac{3}{2}l+\tfrac{1}{2}) - \psi(n+\tfrac{1}{2}m+\tfrac{1}{2}l+\tfrac{1}{2}) \\ &-\psi(n+\tfrac{1}{2}m-\tfrac{1}{2}l+\tfrac{1}{2}) \\ &+\psi(n+l+1)+\psi(n+\tfrac{1}{2}m+\tfrac{1}{2}l+1)+\psi(n+1)],\end{aligned} \tag{10}$$

with $(m+l)$ even.

Equation (9) and (10) converge rapidly in the range $0 < s < 1$ for s not too close to 1.

4. EXPANSIONS FOR $1 < s < 3$

To treat the cases l even and l odd simultaneously, we write Eq. (6) in the form

$$\begin{aligned}I(l;s) = &\frac{1}{l!s(2s)^{3l}} \\ &\times \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(-\mu)\,\Gamma(2\mu+2l+1)\,\Gamma(2\mu+3l+1)\exp(i\pi\mu)}{\Gamma(\mu+2l+1)[\Gamma(\mu+l+1)]^2(4s^2)^{\mu}} \\ &\times F(\mu+l+\beta+\tfrac{1}{2},\ \mu+\alpha+l+1;\ l+1;\ 1/s^2)\,d\mu,\end{aligned} \tag{11}$$

with the abbreviations introduced above: $\alpha = [\frac{1}{2}l]$, $\beta = [\frac{1}{2}(l+1)]$. We now force an expansion in terms of $\frac{1}{4}(s^2-5)$, whose magnitude is always less than unity for $1 < s < 3$, by applying the substitution

$$\Gamma(a)(1-z)^{-a} = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Gamma(p+a)\,\Gamma(-p)(-z)^p\,dp. \tag{12}$$

By the formula connecting $F(\ldots;z)$ and $F(\ldots;1-z)$, Eq. (11) becomes

$$\begin{aligned}I(l;s) = &\frac{2^{2l}}{\pi s^{3l+1}(2\pi i)^2}\int_{c-i\infty}^{c+i\infty}\int_{c-i\infty}^{c+i\infty}\Gamma(-\mu) \\ &\times \frac{\Gamma(\mu+l+\frac{1}{2})\,\Gamma(\mu+l+\alpha+1)\,4^{\mu}(-s^2)^{l+\beta+1/2}}{\Gamma(\mu+2l+1)\,\Gamma(\mu+l+1)} \\ &\times \frac{\Gamma(-t)\,\Gamma(t+\mu+l+\beta+\frac{1}{2})\,\Gamma(t-\mu-\alpha)\exp(i\pi t)}{\Gamma(-\mu-\alpha)\Gamma(t+l+1)} \\ &\times (1-s^2)^{-t-\mu-l-\beta-1/2}\,d\mu\,dt,\end{aligned} \tag{13}$$

with the familiar restrictions on c, i.e., $c = 0^-$. Now write $(1-z)$ in Eq. (12) as $(1-b)[1-(b-z)/(b-1)]$ and set $b=5$ and $z=s^2$. Substitution in Eq. (13) and shifting μ by $t-\alpha$ yields the triple integral

$$\begin{aligned}I(l;s) = &\frac{s^{2\beta-l}(-1)^{\alpha}}{\pi 2^{2\beta+1}}(1/2\pi i)^3\int_{c_1-i\infty}^{c_1+i\infty}\int_{c-i\infty}^{c+i\infty}\int_{c-i\infty}^{c+i\infty} \\ &\frac{\Gamma(\mu+t+1)\Gamma(\mu+t+l-\alpha+\frac{1}{2})\Gamma(\mu+t+l+1)\Gamma(-t)\Gamma(-\mu)}{\Gamma(\mu+t+2l-\alpha+1)\,\Gamma(\mu+t+l-\alpha+1)} \\ &\times \frac{\Gamma(p+2t+\mu+l+\beta-\alpha-\frac{1}{2})\,\Gamma(-p)}{\Gamma(\mu+t-\alpha+1)\,\Gamma(t+l+1)}\ \exp[i\pi(\mu+t)] \\ &\times 4^{-t}\left(\frac{s^2-5}{4}\right)^p d\mu\,dt\,dp;\end{aligned} \tag{14}$$

c_1 must be chosen to maintain the pole separation in the μ-plane. The integration over μ can be carried out if we first apply Barnes' first lemma and write[4]

$$\begin{aligned}&\frac{\Gamma(\mu+t+l-\alpha+\frac{1}{2})\,\Gamma(\mu+p+2t+l+\beta-\alpha+\frac{1}{2})}{\Gamma(\mu+t+2l-\alpha+1)} \\ &= \frac{1}{2\pi i}\int\frac{\Gamma(q+t+l-\alpha+\frac{1}{2})\Gamma(q+p+2t+l+\beta-\alpha+\frac{1}{2})}{\Gamma(-t-p+l-\beta+\frac{1}{2})\Gamma(l+\frac{1}{2})} \\ &\quad\times\Gamma(-q+\mu)\Gamma(-q-p-2t+\alpha-\beta)\,dq.\end{aligned} \tag{15}$$

Thus,

$$\begin{aligned}&\frac{1}{2\pi i}\int\frac{\Gamma(\mu+t+1)\,\Gamma(\mu+t+l+1)\,\Gamma(\mu-q)\,\Gamma(-\mu)}{\Gamma(\mu+t+l-\alpha+1)\,\Gamma(\mu+t-\alpha+1)}\,d\mu \\ &= \frac{\Gamma(-q)\,\Gamma(t+1)\,\Gamma(t+l+1)}{\Gamma(t-\alpha+1)\Gamma(t+l-\alpha+1)}\ {}_3F_2(t+l+1, t+1, -q; \\ &\quad t-\alpha+1, t+l-\alpha+1; 1).\end{aligned} \tag{16}$$

The convergence criterion for this ${}_3F_2$ function ($q > 2\alpha$) is too restrictive for our purposes as we shall shortly want to interchange summation and integration. Therefore, we adopt a generalization of Dixon's theorem (Ref. 5, Eq. 2.3.3.7)

$$\begin{aligned}{}_3F_2(a,b,c;d,e;1) = &\frac{\Gamma(d)\Gamma(e)\Gamma(s)}{\Gamma(a)\,\Gamma(b+s)\,\Gamma(c+s)} \\ &\times {}_3F_2(d-a, e-a, s; s+b, s+c; 1),\end{aligned} \tag{17}$$

which transforms the rhs of Eq. (16) into

$$\frac{\Gamma(-q)\Gamma(t+1)}{\Gamma(-l-\alpha)\Gamma(-\alpha)}\sum_{n=0}^{\alpha}\frac{\Gamma(n-l-\alpha)\,\Gamma(n-\alpha)\,\Gamma(n+q-2\alpha)}{\Gamma(n+q+t-2\alpha+1)\Gamma(n-2\alpha)}, \tag{18}$$

when we substitute the (finite!) series for ${}_3F_2$. The q-integral formed by combining Eqs. (15) and (18) can be carried out to yield

$$\int \frac{\Gamma(q+t+l-\alpha+\tfrac{1}{2})\Gamma(q+p+2t+l+\beta-\alpha+\tfrac{1}{2})\,\Gamma(-q-p-2t+\alpha-\beta)\,\Gamma(q+n-2\alpha)\,\Gamma(-q)\,dq}{\Gamma(q+n+t-2\alpha+1)}$$

$$=\frac{\Gamma(n-2\alpha)\Gamma(t+l-\alpha+\tfrac{1}{2})\,\Gamma(p+2t+l+\beta-\alpha+\tfrac{1}{2})\,\Gamma(l+\tfrac{1}{2})\,\Gamma(-p-t+l-\beta+\tfrac{1}{2})}{\Gamma(n+t-2\alpha+1)\Gamma(t+2l-\alpha+1)}$$

$$\times{}_3F_2(t+1,p+2t+l+\beta-\alpha+\tfrac{1}{2},t+l-\alpha+\tfrac{1}{2};t+2l-\alpha+1,t+n-2\alpha+1;1). \tag{19}$$

The convergence criterion for this ${}_3F_2$ function ($p+2t-n+\alpha+\beta<0$) is too restrictive for our purpose and so we apply the generalized Dixon theorem to obtain the function

$${}_3F_2(2l-\alpha,n-2\alpha,-p-2t+n-\alpha-\beta;n+l-2\alpha+\tfrac{1}{2},-p-t+n-2\alpha-\beta+l+\tfrac{1}{2};1), \tag{20}$$

whose series representation terminates because $n\leq\alpha$. Substituting this series for ${}_3F_2$, we find that Eq. (14) reduces to

$$I(l;s)=\frac{s^{2\beta-l}(-1)^{l+\alpha}}{2\pi}$$

$$\times\sum_{n=0}^{\alpha}\sum_{m=0}^{2\alpha-n}\frac{(n-l-\alpha-1)!\,(n-\alpha-1)!}{m!\,n!\,(n-2\alpha-1)!\,(-l-\alpha-1)!\,(-\alpha-1)!}$$

$$\times\frac{(m+2l-\alpha-1)!\,(m+n-2\alpha-1)!}{\Gamma(m+n+l-2\alpha+\tfrac{1}{2})(2l-\alpha-1)!}$$

$$\times(1/2\pi i)^2\int_{\mathrm{Re}\,p=0^-}\int_{\mathrm{Re}\,t=\beta+0^-}$$

$$\frac{\Gamma(t+\tfrac{1}{2})\Gamma(p+2t+\tfrac{1}{2})\Gamma(-p-2t+m+n+\gamma)\Gamma(-p)}{\Gamma(t+\alpha+1)\Gamma(-p-t+m+n+\gamma+\tfrac{1}{2})4^t}$$

$$\times\Gamma(-t+\beta)\exp(i\pi t)[(s^2-5)/4]^p\,dp\,dt. \tag{21}$$

The integration over p is routine. Adding the residues at the poles $p=p$ and $p-2t+m+n+\gamma$ we obtain (p is henceforth an integer)

$$I(l;s)=\frac{s^{2\beta-l}(-1)^l}{2\pi}\sum_{n=0}^{\alpha}\sum_{m=0}^{2\alpha-n}\sum_{p=0}^{\infty}$$

$$\frac{(n-l-\alpha-1)!\,(n-\alpha-1)!\,(m+2l-\alpha-1)!\,(m+n-2\alpha-1)!}{m!\,n!\,p!\,(n-2\alpha-1)!\,(-l-\alpha-1)!\,(-\alpha-1)!}$$

$$\times\frac{(-1)^p[(s^2-5)/4]^p}{\Gamma(m+n+\gamma+\tfrac{1}{2})(2l-\alpha-1)!}\,(T_1+T_2) \tag{22}$$

where

$$T_1=1/(2\pi i)\int_{\mathrm{Re}\,t=\beta+0^-}$$

$$\frac{\Gamma(t+\tfrac{1}{2})\,\Gamma(2t+p+\tfrac{1}{2})\,\Gamma(-2t-p+m+n+\gamma)\,\Gamma(-t+\beta)}{\Gamma(t+\alpha+1)\,\Gamma(-t-p+m+n+\gamma+\tfrac{1}{2})4^t}$$

$$\times\exp(i\pi t)\,dt \tag{23}$$

and

$$T_2=(1/2\pi i)[(s^2-5)/4]^{m+n+\gamma}\,\Gamma(p+m+n+\gamma+\tfrac{1}{2})$$

$$\times\int_{\mathrm{Re}\,t=\beta+0^-}\frac{\Gamma(t+\tfrac{1}{2})\,\Gamma(2t-p-m-n-\gamma)\,\Gamma(-t+\beta)}{\Gamma(t+\alpha+1)\,\Gamma(t-p+\tfrac{1}{2})4^t}$$

$$\times\exp(i\pi t)[(s^2-5)/4]^{-2t}\,dt. \tag{24}$$

In T_1 we must sum over the residues of poles in the positive half-plane such that $t\geq\beta+0^-$ while in T_2 the poles for $t<\beta+0^-$ are to be taken. It is convenient to introduce the abbreviations, in addition to the α and β of Eq. (9),

$$\gamma=l-2\alpha,\ \delta=[\tfrac{1}{2}(m+n-p+\gamma+1)],\ \epsilon=[\tfrac{1}{2}(m+n-p+\gamma)],$$
$$\mu=[\tfrac{1}{2}(p+m+n+\gamma+1)],\ \nu=[\tfrac{1}{2}(p+m+n+\gamma)], \tag{25}$$

so that the required poles of $\Gamma(-2t-p+m+n+\gamma)\times\Gamma(-t+\beta)$ in T_1 are

double poles at $t=q,\ q=\beta,\ldots,$

simple poles at $t=q+\tfrac{1}{2},\ q=\beta,\ldots,$

and the required poles of $\Gamma(2t-p-m-n-\gamma)$ in T_2 are

simple poles at $t=-q-\tfrac{1}{2},\quad q=-\min(\beta,\mu),\ldots,$

simple poles at $t=-q,\quad q=-\min(\beta-1,\nu),\ldots.$ (26)

Adding the contributions from these poles, and separating the result into its real and imaginary parts, we finally obtain that

$$\mathrm{Re}I(l;s)=-\frac{s^{2\beta-l}(-1)^{\beta}}{4\pi^2}\sum_{n=0}^{\alpha}\sum_{m=0}^{(2\alpha-n)}\sum_{p=0}^{\infty}\frac{(n-l-\alpha-1)!\,(n-\alpha-1)!\,(m+2l-\alpha-1)!\,(m+n-2\alpha-1)!}{m!\,n!\,p!\,(n-2\alpha-1)!\,(-l-\alpha-1)!\,(-\alpha-1)!\,(2l-\alpha-1)!\,\Gamma(m+n+\gamma+\tfrac{1}{2})}$$

$$\times\Bigg(\sum_{q=\beta}^{\infty}(-1)^{p+q+\beta}\,\frac{\Gamma(q+\frac12)\Gamma(2q+p+\frac12)\,\Gamma(q+p-m-n-\gamma+\frac12)}{(q-\beta)!\,(q+\alpha)!\,(2q+p-m-n-\gamma)!\,4^q}\,[(s^2-5)/4]^p$$

$$\times[\psi(q+\tfrac12)+2\psi(2q+p+\tfrac12)+\psi(q+p-m-n-\gamma+\tfrac12)-\psi(q+\alpha+1)-\ln 4-\psi(q-\beta+1)-2\psi(2q+p-m-n-\gamma+1)]$$

$$-\pi[(s^2-5)/4]^{m+n+\gamma}\sum_{q=-\min(\beta-1,\nu)}^{\alpha}(-1)^{p+q+m+n+\gamma}\,\frac{\Gamma(p+m+n+\gamma+\frac12)\,\Gamma(p+q+\frac12)\,(q+\beta-1)!}{(-q+\alpha)!\,(p+2q+m+n+\gamma)!\,\Gamma(q+\frac12)}\,4^q[(s^2-5)/4]^{2q}\Bigg), \tag{27}$$

$\mathrm{Im}I(l;s)$

$$=\frac{s^{2\beta-l}(-1)^{\beta}}{4\pi^2}\sum_{n=0}^{\alpha}\sum_{m=0}^{(2\alpha-n)}\frac{(n-l-\alpha-1)!\,(n-\alpha-1)!\,(m+2l-\alpha-1)!\,(m+n-2\alpha-1)!}{m!\,n!\,(n-2\alpha-1)!\,(-l-\alpha-1)!\,(-\alpha-1)!\,(2l-\alpha-1)!\,\Gamma(m+n+\gamma+\frac12)}$$

$$\times\Bigg(\tfrac12\pi^2\sum_{p=0}^{(m+n+\gamma-1)}\sum_{q=\beta}^{(m+n-p+\gamma-1)}\frac{(-1)^{m+n+\gamma+\beta}\,q!\,(p+2q+\frac12)!\,[(s^2-5)/4]^p}{p!\,\Gamma(q+\alpha+\frac32)(-q-p+m+n+\gamma-1)!\,(p+2q-m-n-\gamma+1)!\,\Gamma(q-\beta+\frac32)4^q}$$

$$-\pi\sum_{p=0}^{\infty}\sum_{q=\beta}^{\infty}(-1)^{p+q+\beta}\,\frac{\Gamma(q+\frac12)\Gamma(p+2q+\frac12)\Gamma(p+q-m-n-\gamma+\frac12)}{p!\,(q-\beta)!\,(q+\alpha)!\,(p+2q-m-n-\gamma)!\,4q}\,[(s^2-5)/4]^p$$

$$+2[(s^2-5)/4]^{m+n+\gamma}\sum_{p=0}^{\infty}\sum_{q=-\min(\beta,\mu)}^{\infty}(-1)^{p+m+n+\gamma+\alpha}\,\frac{\Gamma(p+m+n+\gamma+\frac12)\Gamma(q-\alpha+\frac12)(q+p)!\,\Gamma(q+\beta+\frac12)}{p!\,q!\,(p+2q+m+n+\gamma+1)!}\,4^q[(s^2-5)/4]^{2q+p+1}\Bigg). \tag{28}$$

The convention to be used in the summation is that a sum is zero if the upper summation index is smaller than the lower index.

5. EXPANSION FOR $s=1$

To obtain an expansion valid at the singular point $s=1$ we go back to Eq. (21). The p-integral with $p\to -p$

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\Gamma(p-2t+m+n+\gamma)\Gamma(p)\Gamma(-p+2t+\frac12)}{\Gamma(p-t+m+n+\gamma+\frac12)}\times\exp(-i\pi p)\,dp \tag{29}$$

is just the Gauss hypergeometric function

$$-i\frac{\Gamma(m+n+\gamma+\frac12)\Gamma(2t+\frac12)}{\Gamma(t+m+n+\gamma+1)}\times F(m+n+\gamma+\tfrac12,2t+\tfrac12;t+m+n+\gamma+1;1)\exp(-2\pi i t), \tag{30}$$

which sums to

$$-i\frac{\Gamma(m+n+\gamma-\frac12)\Gamma(2t+\frac12)\Gamma(-t)}{\Gamma(t+\frac12)\Gamma(-t+m+n+\gamma+\frac12)}\exp(-2\pi i t). \tag{31}$$

The t-integral is

$$\int_{\mathrm{Re}t<0}\frac{\Gamma(-t+\beta)\Gamma(t+\frac34)\Gamma(t+\frac14)\Gamma(-t)}{\Gamma(t+\alpha+1)\Gamma(-t+m+n+\gamma+\frac12)}\exp(-i\pi t)\,dt, \tag{32}$$

which by the general theorem for ${}_AF_B(Z)$ given by Slater[5] (Sec. 4.6.2) integrates to

$$\frac{i}{\sqrt2}\Bigg((1-i)\frac{\Gamma(-\frac12)\Gamma(\beta+\frac34)\Gamma(\frac34)}{\Gamma(\alpha+\frac14)\Gamma(m+n+\gamma+\frac54)}\times{}_3F_2(\beta+\tfrac34,\tfrac34,-\alpha+\tfrac34;\tfrac32,m+n+\gamma+\tfrac54;1)$$

$$-(1+i)\frac{\Gamma(\frac12)\Gamma(\beta+\frac14)\Gamma(\frac14)}{\Gamma(\alpha+\frac34)\Gamma(m+n+\gamma+\frac34)}\times{}_3F_2(\beta+\tfrac14,\tfrac14,-\alpha+\tfrac14,\tfrac12,m+n+\gamma+\tfrac34;1)\Bigg). \tag{33}$$

The series for these ${}_3F_2$ functions will be seen to converge as $q^{-m-n-1/2}$, where q is the summation integer. This convergence which is quite slow can be substantially improved by utilizing the generalized Dixon's theorem (cf. Ref. 5, Eq. 2.3.3.7) to transform the ${}_3F_2$ functions in Eq. (33). Substituting the result into Eq. (21), we then obtain

$$\left.\begin{matrix}\mathrm{Re}I(l;1)\\ \mathrm{Im}I(l;1)/i\end{matrix}\right\}=\frac{(-1)^{\beta}}{4\sqrt{\pi}}\sum_{n=0}^{\alpha}\sum_{m=0}^{2\alpha-n}\frac{(n-l-\alpha-1)!\,(n-\alpha-1)!\,(m+2l-\alpha-1)!\,(m+n-2\alpha-1)!\,\Gamma(m+n+\frac12)}{m!\,n!\,(n-2\alpha-1)!\,(-l-\alpha-1)!\,(-\alpha-1)!\,(2l-\alpha-1)!}$$

$$\times\Bigg(-\frac{\Gamma(\frac34)\,{}_3F_2(-\beta+\frac34,m+n+\gamma-\beta+\frac12,m+n+\frac12;m+n+\frac54,m+n-\alpha+\frac54;1)}{\Gamma(\alpha+\frac14)\Gamma(m+n+\frac54)\Gamma(m+n-\alpha+\frac54)}$$

$$\pm\frac{\Gamma(\frac14)\,{}_3F_2(-\beta+\frac14,m+n+\gamma-\beta+\frac12,m+n+\frac12;m+n+\frac34,m+n-\alpha+\frac34;1)}{\Gamma(\alpha+\frac34)\Gamma(m+n+\frac34)\Gamma(m+n-\alpha+\frac34)}\Bigg). \tag{34}$$

The series for ${}_3F_2(-\beta+\frac{3}{4},\cdots)$ in Eq. (34) will be seen to converge as $q^{-\beta-3/4}$, while that for ${}_3F_2(-\beta+\frac{1}{4},\cdots)$ converges as $q^{-\beta-1/4}$. Equation (34) is thus particularly useful for large values of l.

For $l=0$, we regain the well-known result given by Katsura *et al.*,[1] i.e.,

$$\frac{\mathrm{Re}I(0;1)}{\mathrm{Im}I(0;1)/i} = -\frac{1}{8}\frac{[\Gamma(\frac{3}{4})]^2}{[\Gamma(\frac{7}{8})]^4} \pm \frac{1}{4}\frac{[\Gamma(\frac{1}{4})]^2}{[\Gamma(\frac{5}{8})]^4}, \tag{35}$$

after substitution of the known sums for the ${}_3F_2$ functions in Eq. (33).

6. EXPANSIONS FOR $s=3$

Again starting from Eq. (21), we obtain for the p-integral

$$\frac{1}{2\pi i}\int \frac{\Gamma(p+2t+\frac{1}{2})\Gamma(-p-2t+m+n+\gamma)\Gamma(-p)}{\Gamma(-p-t+m+n+\frac{1}{2})}\,dp \tag{36}$$

$$= \frac{\Gamma(m+n+\gamma+\frac{1}{2})\Gamma(2t+\frac{1}{2})}{\Gamma(t+m+n+\gamma+1)}$$

$$\times {}_2F_1(2t+\tfrac{1}{2}, m+n+\gamma+\tfrac{1}{2}; t+m+n+\gamma+1; -1), \tag{37}$$

which by analytic continuation transforms to

$$\frac{\Gamma(m+n+\gamma+\frac{1}{2})\Gamma(2t+\frac{1}{2})}{\Gamma(t+m+n+\gamma+1)}2^{-m-n-\gamma-1/2}$$

$$\times {}_2F_1\,(m+n+\gamma+\tfrac{1}{2}, -t+m+n+\gamma+\tfrac{1}{2}; t+m+n+\gamma+1; \tfrac{1}{2}). \tag{38}$$

The full expression for $I(l;3)$ follows by substitution into Eq. (21) so that

$$I(l;3) = \frac{3^{2\beta-l}(-1)^\beta}{2\pi}\sum_{n=0}^{\alpha}\sum_{m=0}^{2\alpha-n}\sum_{p=0}^{\infty}\frac{(n-l-\alpha-1)!\,(n-\alpha-1)!\,(m+2l-\alpha-1)!\,(m+n-2\alpha-1)!}{m!\,n!\,p!\,(n-2\alpha-1)!\,(-l-\alpha-1)!\,(-\alpha-1)!\,(2l-\alpha-1)!}$$

$$\times\frac{1}{2\pi i}\frac{\Gamma(p+m+n+\gamma+\frac{1}{2})}{\Gamma(m+n+l-2\alpha+\frac{3}{2})}2^{-p-m-n-\gamma-1/2}\int_{\mathrm{Re}t>0}\frac{\Gamma(t+\frac{1}{2})\Gamma(2t+\frac{1}{2})\Gamma(-t+p+m+n+\gamma+\frac{1}{2})\Gamma(-t+\beta)\exp(i\pi t)}{\Gamma(t+\alpha+1)\Gamma(-t+m+n+\gamma+\frac{1}{2})\Gamma(t+p+m+n+\gamma+1)4^t}\,dt. \tag{39}$$

The only poles in the positive t half-plane are at $t=q+\beta$, and summing the residues there we obtain the expansion

$$I(l;3) = -\frac{3^{2\beta-l}}{2^{2\beta+\gamma}2\pi\sqrt{2}}$$

$$\times\sum_{n=0}^{\alpha}\sum_{m=0}^{2\alpha-n}\sum_{p=0}^{\infty}\frac{(n-l-\alpha-1)!\,(n-\alpha-1)!\,(m+2l-\alpha-1)!\,(m+n-2\alpha-1)!\,\Gamma(p+m+n+\gamma+\frac{1}{2})}{m!\,n!\,p!\,(n-2\alpha-1)!\,(-l-\alpha-1)!\,(-\alpha-1)!\,(2l-\alpha-1)!\,\Gamma(m+n+l-2\alpha+\frac{1}{2})}$$

$$\times\frac{(-1)^p}{2^{p+m+n}}\sum_{q=0}^{\infty}\frac{\Gamma(q+\beta+\frac{1}{2})\Gamma(2q+2\beta+\frac{1}{2})\Gamma(q+\beta-m-n-\gamma+\frac{1}{2})}{q!\,(q+l)!\,\Gamma(q+\beta-p-m-n-\gamma+\frac{1}{2})(q+\beta+p+m+n+\gamma)!\,4^q}, \tag{40}$$

where inspection shows that the series in q converges as $q^{-m-n-1/2}$. The imaginary part is, of course, zero for $s\geq 3$.

7. CONCLUSIONS

We have obtained expansions for the diagonal elements of the cubic lattice Green function $I(l;s)$ for the ranges $0\leq s\leq 1$ [Eqs. (9) and (10)], $1<s<3$ [Eqs. (27) and (28)], and at the singular points $s=1$ (Eq. 34) and $s=3$ (Eq. 40) for arbitrary l. Eqs. (9) and (10), (27) and (28) converge rapidly through most of the range in s, so that typically ten terms in the q-summation are sufficient to provide ten digit accuracy. The expansion at $s=1$ converges more slowly for small values of l, but becomes extremely useful for large l. The expansion for $s=3$ is perhaps least useful as it contains a series in q which only falls off as $q^{-1/2}$ in the worst case ($m=n=0$). Generalized expansions around $s=1$ and $s=3$ would therefore be extremely useful. Work towards this end is currently in progress. It should also be noted that the expansions given here avoid the use of the recondite Kampé de Fériet functions which are central to the recent work of Abe and Katsura,[6] and are therefore more directly amenable to programming on a computer. The computing time necessary to provide any given accuracy is readily determined from our convergent expansions. For example, on a CDC-7600 $I(l;s)$ accurate to twelve significant figures is obtained for $s=0.4$ and $l=4$ in 0.02 seconds from Eqs. (9) and (10), and for $s=2$ and $l=4$ in 0.1 seconds from Eqs. (27) and (28). Comparison values for selected cases were obtained from Joyce[7] for $l=0$, and by direct evaluation of the integral in Eq. (4) for $l\neq 0$. It now becomes quite feasible, therefore, to construct rapidly extensive tables of the cubic lattice Green functions from their recurrence relations and our expressions.

ACKNOWLEDGMENTS

The author is grateful to Dr. E.W. Montroll for suggesting the problem and to Mr. R.S. Janda for expert programming assistance. This paper was sponsored in part by the Defense Advanced Research Projects Agency,

ARPA Order No. 2953, under contract DAAH01-75-C-0803.

[1]S. Katsura, S. Inawashiro, and Y. Abe, J. Math. Phys. **12**, 895 (1971).
[2]T. Horiguchi, J. Phys. Soc. Japan **30**, 1261 (1971).
[3]G.N. Watson, *Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1944).
[4]We shall henceforth omit the limits on the integrals, it being understood that the integrations run parallel to the imaginary axis with due allowance for the pole separation.
[5]L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966).
[6]Y. Abe and S. Katsura, Ann. Phys. **75**, 348 (1973).
[7]G.S. Joyce, Phil. Trans. Roy. Soc. **273**A, 583 (1973).