Representation of the entropy functional for a grand canonical ensemble in classical statistical mechanics*

B. Forte

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

C. C. A. Sastri

Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0
(Received 2 December 1976)

A representation theorem for the entropy functional of a system of an arbitrary but finite number of particles is proved. This theorem is a generalization of the main result of a previous paper of ours [J. Math. Phys. **16**, 1453 (1975)], which gives a characterization of the entropy functional on the set of all probability densities f on R_n s.t. $f\log f$ is integrable. As may be expected when n is a random variable, the expression for entropy consists of two parts; one which arises from the ignorance about n and another which is the average, over n, of the conditional entropy given the number of particles. As in the above-mentioned paper, the expression includes the term corresponding to chemical reactions (with n replaced by the average number of particles) and the continuous analog of the Hartley entropy. It is conjectured that this last term might be of some significance in physics.

1. INTRODUCTION

In a previous paper (Ref. 1), a representation theorem was proved for entropy functionals on the set of all probability densities f on R_n such that $f \log f$ is integrable. In statistical mechanics, where the notion of entropy plays a central role, n turns out to be the same, except for a factor, as the number of particles in a gas. Since the exact number of particles in a gas is never known it is necessary, for the theorem to be useful in statistical mechanics, to remove the restriction that n be known precisely. This corresponds to considering the so-called grand canonical system or ensemble, which is what we do in the present paper. However, for the sake of simplicity, we restrict ourselves to the case when n is arbitrary but finite. The extension to the case when n could be infinite is straightforward and purely technical and, we believe, does not add anything new to the basic representation theorem.

As may be expected when n varies, the expression for entropy consists of two parts: One which arises from the ignorance about n and another which is an average, over n, of the conditional entropy given the number of particles. As in Ref. 1, the representation includes a term which corresponds to chemical reactions—with n replaced by the average number of particles—and the continuous analog of the Hartley entropy, which we believe has some physical significance. More precisely, let f_n denote the conditional probability density of the system given that there are n particles. Then, the entropy functional Φ_1 has the representation

$$\Phi_1(\{p_n f_n, n=0,1,2,\cdots\}) = -a\sum_{n=1}^{\infty} p_n \int_{R_{6n}} f_n \log f_n \, d\mu + b\sum_{n=1}^{\infty} p_n n + c\sum_{n=1}^{\infty} p_n \log \mu(A_{f_n}) - d\sum_{n=0}^{\infty} p_n \log p_n, \quad (1)$$

where μ denotes the Lebesque measure on R_{6n} $(n>0)$, $A_{f_n}=\{x\in R_{6n} \mid f_n(x)>0\}$ $\forall\, n>0$, and a,b,c,d are real numbers, with $a,c,d\geq 0$; for each n, p_n is the probability that there are n particles and x_n is a six-dimensional vector denoting the position and momenta of the nth particle. If $p_n=1$ for some n and zero for all other values, the representation (1) reduces to that given in Ref. 1, namely

$$\Phi_1(f_n) = -a\int_{R_{6n}} f_n \log f_n \, d\mu + bn + c\log\mu(A_{f_n}) \quad (2)$$

for the corresponding canonical system.

The method we adopt is similar to the one in Ref. 1. In order to obtain the representation for the entropy of a single gas, we consider a mixture of two gases and impose the physically meaningful condition that the entropy of the mixture be no greater than the sum of the entropies of the components (subadditivity), with equality holding if the components are stochastically independent (additivity). Generalizing the notion that entropy should not depend on the choice of the coordinate system, we require that it be invariant under the isometries induced on the set of probability densities by invertible, measure preserving transformations from R_{6n} to itself for each n. Finally, we impose some simple regularity conditions.

2. PRELIMINARIES

Let N denote the set of all positive integers. Let the two gases in the mixture be labeled 0 and 1 respectively and, $\forall\, n\in N$, let $t_i=0$ or 1 $\forall\, i$ s.t. $1\leq i\leq n$. The value of t_i identifies the component to which the ith particle belongs. For each n, let B_n denote the σ-algebra of Borel sets in the Euclidean space R_n and let μ_n be the Lebesque measure[2] on R_n. The state of the system will be described by a sequence of functions $\{\phi_n(x_1,x_2,\ldots,x_n,t_1,t_2,\ldots,t_n)\}$—$x_i$ denotes a six-dimensional vector specifying the position and momenta of the ith particle—s.t. $\phi_n\in L_1(R_{6n},B_{6n},\mu_{6n})$, $\phi_n\geq 0$, and

$$\phi_0 + \sum_{n=1}^{\infty} \sum_{\substack{(t_1,t_2,\ldots,t_n)\\ \text{all possible}\\ \text{strings of } 0^s \text{ and } 1^s}} \times \int_{R_{6n}} \phi_n(x_1,x_2,\ldots,x_n,t_1,t_2,\ldots,t_n)\,dx_1\cdots dx_n = 1. \quad (3)$$

ϕ_0 is the probability that there are no particles. (When the number of particles is zero, we shall not distinguish between the mixture and its components.) Since the order in which the particles are counted does not

matter, ϕ_n must be symmetric in each couple (x_i, t_i):

$$\phi_n(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}, t_{\pi_1}, t_{\pi_2}, \ldots, t_{\pi_n}) = \phi_n(x_1, x_2, \ldots, x_n, t_1, t_2, \ldots, t_n),$$

where $(\pi_1, \pi_2, \ldots, \pi_n)$ is any permuation of $(1, 2, \ldots, n)$. This symmetry simplifies the description somewhat. Introduce the sequence of functions $\psi_{n,h}$ defined by

$$\psi_0 = \phi_0, \quad \psi_{n,h} = \binom{n}{h} \phi_n(x_1, x_2, \ldots, x_n, \overbrace{0, 0, \ldots, 0}^{h \text{ times}}, \overbrace{1, 1, \ldots, 1}^{n-h \text{ times}}).$$

Then (3) becomes

$$\psi_0 + \sum_{n=1}^{\infty} \sum_{h=0}^{n} \int_{R_{6n}} \psi_{n,h}(x_1, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n = 1.$$

We define the marginal of type 0 of $\psi_{n,h}$ by

$$\bar{f}_h(x_1, x_2, \ldots, x_h) = \psi_{h,h}(x_1, x_2, \ldots, x_h) + \sum_{i=1}^{\infty} \int_{R_{6i}} \psi_{h+i,h}(x_1, \ldots, x_h, \xi_1, \xi_2, \ldots, \xi_i)\, d\xi_1\, d\xi_2 \cdots d\xi_i,$$

and the marginal of type 1 of $\psi_{n,h}$ by

$$\bar{g}_k(x_1, x_2, \ldots, x_k) = \psi_{k,0}(x_1, x_2, \ldots, x_k) + \sum_{i=1}^{\infty} \int_{R_{6i}} \psi_{k+i,i}(\xi_1, \xi_2, \ldots, \xi_i, x_1, x_2, \ldots, x_k)\, d\xi_1\, d\xi_2 \cdots d\xi_i.$$

Let $\pi_{00} = \psi_0$ and, for r or $s \neq 0$, let

$$\begin{aligned} &\pi_{rs} = \int_{R_{6(r+s)}} \psi_{r+s,r}(x_1, \ldots, x_{r+s})\, dx_1 \cdots dx_{r+s}, \\ &p_0 = \bar{f}_0 = \sum_{s=0}^{\infty} \pi_{0s}, \quad q_0 = \bar{g}_0 = \sum_{r=0}^{\infty} \pi_{r0}, \\ &p_r = \int_{R_{6r}} \bar{f}_r(x_1, \ldots, x_r)\, dx_1 \cdots dx_r = \sum_{s=0}^{\infty} \pi_{rs} \quad (r \in N), \\ &q_s = \int_{R_{6s}} \bar{g}_s(x_1, \ldots, x_s)\, dx_1 \cdots dx_s = \sum_{r=0}^{\infty} \pi_{rs} \quad (s \in N). \end{aligned} \tag{4}$$

Then π_{rs} is the joint probability that there are r particles of type 0 and s particles of type 1 in the mixture. p_r is the probability that there are r particles of type 0 in the mixture, irrespective of how many particles there are of type 1; similarly for q_s. The two gases in the mixture are stochastically independent if and only if

$$\psi_0 = p_0 q_0$$

$$\psi_{nh}(x_1, x_2, \ldots, x_n) = \bar{f}_h(x_1, \ldots x_h)\bar{g}_{n-h}(x_{h+1}, \ldots, x_n)$$

for all $(x_1, x_2, \ldots, x_n) \in R_{6n}$, $n \in N$, and $h = 0, 1, 2, \ldots, n$. Finally, define the functions f_r, g_s by

$$f_0 = g_0 = 1,$$

and

$$\begin{aligned} &p_r f_r(x_1, x_2, \ldots, x_r) = \bar{f}_r(x_1, x_2, \ldots, x_r), \\ &(x_1, x_2, \ldots, x_r) \in R_{6r} \\ &q_s g_s(x_1, x_2, \ldots, x_s) = \bar{g}_s(x_1, x_2, \ldots, x_s), \\ &(x_1, x_2, \ldots, x_s) \in R_{6s}, \quad r, s \in N. \end{aligned} \tag{5}$$

f_r is the conditional probability density which describes the state of the gas of type 0, given that there are r particles of that type in the mixture; similarly for g_s.

Let

$$W^{(2)} = \{\phi_n, n = 0, 1, 2, \ldots\} \mid \forall\, n \in N: f_h, f_h \log f_h, g_h, g_h \log g_h \in L_1(R_{6h}, B_{6h}, \mu_{6h}) \text{ for all } h \text{ s.t. } 1 \leq h \leq n;\ \phi_n = 0 \text{ for sufficiently large } n\}$$

and

$$W^{(1)} = \{\{\bar{f} = p_n f_n,\ n = 0, 1, 2, \ldots\},\ f_n, p_n \text{ defined as in Eqs. (4) and (5); } \bar{f}_n = 0 \text{ for } n \text{ sufficiently large}\}.$$

Let $\bar{W}^{(2)}$ denote the set of all sequences of simple functions in $W^{(2)}$, and $\bar{W}^{(1)}$ the corresponding set for $W^{(1)}$. Denoting the entropy functionals of the mixture and a component, respectively, by $\Phi_2 : W^{(2)} \to R$ and $\Phi_1 : W^{(1)} \to R$, we shall impose the following conditions:

(1) $\Phi_2(\{\phi_n\}) \leq \Phi_1(\{\bar{f}_n\}) + \Phi_1(\{\bar{g}_n\})$ (subadditivity).

(2) If $\{\phi_n\}$ is s.t. $\psi_0 = p_0 q_0$, and

$$\psi_{nh}(x_1, x_2, \ldots x_n) = \bar{f}_h(x_1, x_2, \ldots, x_h)\, \bar{g}_{n-h}(x_{h+1}, \ldots, x_n)$$

for all $(x_1, x_2, \ldots, x_n) \in R_{6n}$, $n \in N$, and $h = 0, 1, 2, \ldots, n$, then

$$\Phi_2(\{\phi_n\}) = \Phi_1(\{f_n\}) + \Phi_1(\{g_n\}) \quad \text{(additivity)}.$$

(3) If, for a given n and $h \leq n$, $\operatorname{supp}\psi_{nh} \equiv A_{h,n-h} = \operatorname{supp}\psi_{n,n-h} \equiv A_{n-h,n}$ ($\operatorname{supp} f$ means the support of f),

$$\psi_{nh} = \pi_{h,n-h} \frac{1}{\mu_{6n}(A_{h,n-h})} \Big|_{A_{h,n-h}},$$

$$\psi_{n,n-h} = \pi_{n-h,h} \frac{1}{\mu_{6n}(A_{n-h,h})} \Big|_{A_{n-h,h}},$$

where $|_A$ stands for the characteristic function of the set A, then the interchange of ψ_{nh} and $\psi_{n,n-h}$ does not affect Φ_2.

(4) For each $n \in N$, let T_n be an invertible measure-preserving transformation from R_{6n} to itself. Let U_{T_n} be the isometry induced by T_n on $L_1(R_{6n}, B_{6n}, \mu_{6n})$. Then

$$\Phi_2(\{U_{T_n}\phi_n\}) = \Phi_2(\{\phi_n\}) \quad \text{(invariance)}.$$

(5) For $n = 0, 1, 2, \cdots$, let $\{s_{i,n},\ i \in N\}$ be a sequence of nonnegative simple functions s.t. $s_{i,n} \uparrow \bar{f}_n$ a.e., with $\{\bar{f}_n,\ n = 0, 1, \cdots\} \in W^{(1)}$. Then

$$\Phi_1\left(\left\{\frac{s_{i,n}}{s_{i,0} + \sum_{k=1}^{\infty} \| s_{i,k} \|_1}\right\}\right) \to \Phi_1(\{\bar{f}_n\}) \qquad \text{as } i \to \infty.$$

(6) Given the sequence $\{\bar{f}_n\}$ with $\bar{f}_r = (1 - q) f_r$, $\bar{f}_{r+1} = q f_{r+1}$, $q \in [0, 1]$, for some $r \in N$ with $\bar{f}_n = 0$ for $\forall\, n \neq r$ or $r + 1$, the function $q \to \Phi_1(\{\bar{f}_n\})$ is right-continuous at 0.

3. REPRESENTATION THEOREM

We are now ready to state and prove the main result of this paper.

Theorem: If $\Phi_2 : W^{(2)} \to R$ and $\Phi_1 : W^{(1)} \to R$ have properties (1)—(6), then Φ_1 has the presentation given by Eq. 1.

In order to prove this theorem, we shall need the following lemma, which is essentially a restatement of

the main result of Ref. 1:

Lemma 1: Let $W_0^{(2)} \subset W^{(2)}$ denote the set of those sequences $\{\phi_n\}$ for which $\psi_{h+k,h}=0$ $\forall\,(h,k)\neq(r,s)$ for some $r,s\in N$, and let $W_0^{(1)}$ denote the set of their marginals. Then properties (1), (2), (4), and (6), as applied to the restrictions of Φ_2 and Φ_1, respectively, to $W_0^{(2)}$ and $W_0^{(1)}$, become identical with properties (1)–(4) of the representation theorem of Ref. 1. The latter therefore yields

$$\Phi_1(f_r)=-a\int_{R_{6r}} f_r\log f_r\,d\mu+br+c\log\mu(A_{f_r}),$$

where we have identified the sequence $\{f_n\}$ s.t. $f_n=0$ $\forall$ $n\neq r$ with f_r. Here a,b,c are real numbers, and $a,c\geq 0$.

Lemma 2: Let $\{p_n s_n\}\in \bar{W}^{(1)}$, where, for each $n\in N$

$$s_n=\sum_{i=1}^{m_n}\frac{q_{ni}}{\mu(A_{ni})}\Bigg|_{A_{ni}},$$

with $q_{ni}>0$ for each i, $m_n\in N$, $\sum_{i=1}^{m_n}q_{ni}=1$, and $A_{ni}\cap A_{nj}=\phi$ $\forall$ $i\neq j$. Let $A_n=\cup_{i=1}^{m_n}A_{ni}$. Then

$$\Phi_1(\{p_n s_n\})=-a\sum_{n=1}^{\infty}p_n\sum_{i=1}^{m_n}q_{ni}\log\frac{q_{ni}}{\mu(A_{ni})}+b\sum_{n=1}^{\infty}np_n+c\sum_{n=1}^{\infty}p_n\log\mu(A_n)-d\sum_{n=0}^{\infty}p_n\log p_n, \tag{6}$$

where a,b,c,d are real numbers, and $a,c,d\geq 0$.

Proof: Let $n\in N$ and let $\{p_r f_r\}$, $\{q_r g_r\}\in W^{(1)}$. Let $q_r=0$, $r\neq n$, $\mu(\mathrm{supp}f_n)>0$, and $\mathrm{supp}f_n=\mathrm{supp}g_n$. Consider their joint sequence $\{\psi_{r+s,r}\}\in W^{(2)}$ defined by

$$\psi_{r+s,r}(x_1,x_2,\ldots,x_{r+s})=p_r q_s f_r(x_1,x_2,\ldots,x_r)g_s(x_{r+1},\ldots,x_{r+s}),$$
$$\forall\,(x_1,x_2,\ldots,x_{r+s})\in R_{6(r+s)}.$$

Construct the sequence $\{\psi^*_{r+s,r}\}\in W^{(2)}$ by setting

$$\psi^*_{r+s,r}=\psi_{r+s,r}\quad\forall\,(r,s)\neq(n,n),$$

$$\psi^*_{2n,n}(x_1,x_2,\ldots,x_{2n})=p_n g_n(x_1,x_2,\ldots,x_n)f_n(x_{n+1},x_{n+2},\ldots,x_{2n})$$
$$\forall\,(x_1,x_2,\ldots,x_{2n})\in R_{12n}.$$

Then, because of property (4), we have

$$\Phi_2(\{\psi_{r+s,r}\})=\Phi_2(\{\psi^*_{r+s,r}\}).$$

[Since ϕ_{r+s} and $\psi_{r+s,r}$ differ only by a factor, we shall denote the entropy functional by either $\Phi_2(\{\phi_{r+s},r\})$ or $\Phi_2(\{\psi_{r+s},r\})$. | Using this along with properties (1) and (2), we get

$$\Phi_1(\{p_r f_r\})+\Phi_1(\{q_r g_r\})=\Phi_2(\{\psi_{r+s,r}\})=\Phi_2(\{\psi^*_{r+s,r}\})\leq\Phi_1(\{p_r f_r^*\})+\Phi_1(\{h_r\}),$$

where $f_r^*=f_r$ and $h_r=0$ $\forall$ $r\neq n$, and $f_n^*=g_n$, $h_n=p_n f_n+(1-p_n)g_n$. Hence, by Lemma 1, we have

$$\Phi_1(\{p_r f_r\})-\Phi_1(\{p_r f_r^*\})\leq -a\int_{R_{6n}}[p_n f_n+(1-p_n)g_n]\log[p_n f_n+(1-p_n)g_n]\,d\mu+a\int_{R_{6n}}g_n\log g_n\,d\mu.$$

Interchanging f_n and g_n (but keeping everything else the same), we obtain

$$\Phi_1(\{p_r f_r^*\})-\Phi_1(\{p_r f_r\})\leq -\int_{R_{6n}}[p_n g_n+(1-p_n)f_n]\log[p_n g_n+(1-p_n)f_n]\,d\mu+a\int_{R_{6n}}f_n\log f_n\,d\mu.$$

Hence

$$-a\int_{R_{6n}}f_n\log f_n\,d\mu+a\int_{R_{6n}}[p_n g_n+(1-p_n)f_n]\times\log[p_n g_n+(1-p_n)f_n]\,d\mu\leq\Phi_1(\{p_r f_r\})-\Phi_1(\{p_r f_r^*\})\leq -a\int_{R_{6n}}[p_n f_n+(1-p_n)g_n]\log[p_n f_n+(1-p_n)g_n]\,d\mu+a\int_{R_{6n}}g_n\log g_n\,d\mu.$$

Now let f_n and g_n be simple functions with the same support

$$f_n\equiv\sum_{i=1}^{m_n}\frac{q_{ni}}{\mu(A_{ni})}\Bigg|_{A_{ni}},\qquad g_n\equiv\sum_{j=1}^{m'_n}\frac{q'_{nj}}{\mu(B_{ij})}\Bigg|_{B_{nj}},$$

where $q_{ni}>0$ $\forall$ i s.t. $1\leq i\leq m_n$, $q'_{nj}>0$ $\forall$ j s.t. $1\leq j\leq m'_n$,

$$\sum_{i=1}^{m_n}q_{ni}=\sum_{j=1}^{m'_n}q'_{nj}=1,\quad \bigcup_{i=1}^{m_n}A_{ni}=\bigcup_{j=1}^{m_n}B_{nj},$$

and

$$A_{ni}\cap A_{nj}=\emptyset=B_{ni}\cap B_{nj}\quad\forall\,i\neq j.$$

Then the above inequalities yield

$$-a\sum_{i=1}^{m_n}q_{ni}\log\frac{q_{ni}}{\mu(A_{ni})}+a\sum_{i=1}^{m_n}\sum_{j=1}^{m'_n}\left[\frac{p_n q'_{nj}}{\mu(B_{nj})}+\frac{(1-p_n)q_{ni}}{\mu(A_{ni})}\right]\times\mu(A_{ni}\cap B_{nj})\log\left[\frac{p_n q'_{nj}}{\mu(B_{nj})}+\frac{(1-p_n)q_{ni}}{\mu(A_{ni})}\right]\leq\Phi_1(\{p_r f_r\})-\Phi_1(\{p_r f_r^*\})\leq -a\sum_{i=1}^{m_n}\sum_{j=1}^{m'_n}\left[\frac{p_n q_{ni}}{\mu(A_{ni})}+\frac{(1-p_n)q'_{nj}}{\mu(B_{nj})}\right]\mu(A_{ni}\cap B_{nj})\times\log\left[\frac{p_n q_{ni}}{\mu(A_{ni})}+\frac{(1-q_n)q'_{nj}}{\mu(B_{nj})}\right]+a\sum_{j=1}^{m'_n}q'_{nj}\log\frac{q'_{nj}}{\mu(B_{nj})}.$$

From these inequalities it is not difficult to show that the function

$$f_n\to\Phi_1(\{p_r f_r\})+a\,p_n\sum_{i=1}^{m_n}q_{ni}\log\frac{q_{ni}}{\mu(A_{ni})}$$

is constant on the set of all simple probabilities f_n $\sum_{i=1}^{m_n}[q_{ni}/\mu(A_{ni})]|A_{ni}$ having the same support $A_n=\cup_{i=1}^{m_n}A_{ni}$. Indeed, let

$$q_{ni}=\frac{x_i}{\sum_{j=1}^{m_n}x_j},\quad x_i>0\ \forall\,i,$$

$$S(x_1,x_2,\ldots,x_{m_n})=-ap_n\sum_{i=1}^{m_n}\left(\frac{x_i}{\sum_{j=1}^{m_n}x_j}\right)\times\log\left[\frac{1}{\mu(A_{ni})}\cdot\frac{x_i}{\sum_{j=1}^{m_n}x_j}\right]$$

and let f denote the function

$$(x_1,x_2,\ldots,x_{m_n})\to\Phi_1(\{p_r f_r\}).$$

Then the above inequalities show that f is differentiable

and that

$$\frac{\partial f}{\partial x_i} = \frac{\partial S}{\partial x_i}, \quad i = 1, 2, \ldots, m_n,$$

which leads to the desired conclusion.

Hence

$$\Phi_1(\{p_r f_r\}) + a p_n \sum_{i=1}^{m_n} q_{ni} \log \frac{q_{ni}}{\mu(A_{ni})}$$

is some function C of p_n, $\mu(A_{n1})$, $\mu(A_{n2}), \ldots, \mu(A_{nm_n})$, f_r $(r \in N,\ r \neq n)$ and m_n. Property (4) implies that C depends on the sets A_{ni} only through their measures $\mu(A_{ni})$. Actually one can see by considering the special case $q_{ni} = \mu(A_{ni})/\mu(A_n)$ that C depends only on $\mu(A_n)$. It follows therefore that

$$\Phi_1(\{p_r f_r\}) = -a \sum_{n=1}^{\infty} p_n \sum_{i=1}^{m_n} q_{ni} \log \frac{q_{ni}}{\mu(A_{ni})} + F(\bar{p}, \overline{\mu(A)}),$$

where F is a function of the sequences $\bar{p} \equiv (p_0, p_1, p_2, \cdots)$, and $\overline{\mu(A)} = (\mu(A_1), \mu(A_2), \cdots)$, where A_i, $i \in N$, is the support of the simple function f_i. When $f_n = [1/\mu(A_n)]|A_n$, let

$$\Psi_1(\bar{p}, \overline{\mu(A)}) = \Phi_1(\{p_n f_n\}) = a \sum_{n=1}^{\infty} p_n \log \mu(A_n) + F(\bar{p}, \overline{\mu(A)}).$$

Now the restriction of Φ_2, which we shall denote by Ψ_2, to the set of sequences s.t.

$$\psi_{r+s,r} = \frac{\pi_{rs}}{\mu_{6r}(A_r)\mu_{6s}(B_s)} \Big|_{A_r \times B_s}$$

is, by virtue of property (4), a function of the probability distribution $\pi \equiv \{\pi_{rs}\}$ and the measures $\{\mu_{6r}(A_r), \mu_{6s}(B_s)\}$. On the other hand, $\{\mu_{6r}(A_r), p_r\}$ and $\{\mu_{6s}(B_s), q_s\}$ can be regarded as random variables; similarly $\{(\mu_{6r}(A_r), \mu_{6s}(B_s), \pi_{rs}\}$ can be considered a random 2-vector. Thus, Ψ_1 and Ψ_2 represent the uncertainty about these random variables and the random vector, respectively. It can be shown that because of properties (1), (2), (3), and (6) Ψ_2 and Ψ_1 satisfy properties (3), (4), 5(a), (7), and (38) of Ref. 3. They also satisfy property (39) of the same reference. Hence by Corollary 4 of that reference there exists a nonnegative constant d, a constant F_0, and functions $F_r: R \to R$, $r = 1, 2, \cdots$ s.t.

$$\Psi_1(\bar{p}, \overline{\mu(A)}) = -d \sum_{i=0}^{\infty} p_i \log p_i + \sum_{r=1}^{\infty} F_r(\mu_{6r}(A_r)) p_r + F_0 p_0. \quad (7)$$

Consider the case $p_0 = 1$. Then $\Psi_1 = F_0$. But a gas with no particles can be thought of as a mixture of two independent gases each of 0 particles. Hence by property (2)

$$F_0 + F_0 = F_0 \quad \text{or} \quad F_0 = 0.$$

Now, because of property (6), if $p_n = 1$ for some n, then

$$\Psi_1(\bar{p}, \overline{\mu(A)}) = F_n(\mu_{6n}(A_n)).$$

Combining this with Lemma 1 yields

$$\Phi_1\left(\frac{1}{\mu(A_n)} \Big|_{A_n}\right) = a \log \mu(A_n) + bn + c \log \mu(A_n) = F_n(\mu(A_n)). \quad (8)$$

Equation (6) is obtained on combining Eqs. (7) and (8) and observing that $F_0 = 0$. Note however that the constant a occurring in Eq. (6) is not the a occurring in Eqs. (7) and (8) but $a + c$. This proves the lemma.

The theorem is a straightforward consequence of Lemma 2 and property (5).

Remark: Unlike the result of Ref. 1, the theorem proved above does not characterize entropy. To do so one must find a representation not only for Φ_1 but also for Φ_2.

Note added in proof: Since the publication of Ref. 1, related results have appeared, for example, in W. Ochs, Rep. Math. Phys. 9, 331–54 (1976).

*Research supported in part by the National Research Council of Canada under Grant A-7677 and by IBM-Italia.

[1]B. Forte and C.C.A. Sastri, J. Math. Phys. 16, 1453 (1975).

[2]We shall drop the subscripts whenever writing them becomes too cumbersome and there is no room for confusion.

[3]B. Forte, "Sub-additive entropies for a random variable," to appear in Boll. Unione Mat. Ital.

Stationary axially-symmetric solutions of Einstein–Maxwell-massless scalar field equations*

A. Eriş and M. Gürses

Physics Department, Middle East Technical University, Ankara, Turkey
(Received 17 January 1977)

A procedure is presented which enables one to construct solutions to the stationary axially-symmetric gravitational field coupled to massless scalar and Maxwell fields.

1. INTRODUCTION

Since Janis, Newman, and Winicour[1] presented a static, spherically-symmetric solution to the coupled gravitational and massless-scalar field equations there has been some effort to generalize this result to more general space–times. First Penney[2] solved the coupled massless-scalar and gravitational field equations for the static, axially- symmetric geometry obtaining the solution of Janis–Newman–Winicour as a special one with spherical symmetry. The generalization of the Reissner–Nordström solution in the presence of a massless-scalar field was also obtained by Penney.[3] Later Janis, Robinson, and Winicour[4] exhibited the solutions of Einstein-scalar and Brans–Dicke field equations for static space–times and also gave a procedure to generate static solutions of the coupled Einstein–Maxwell-scalar field equations and the corresponding Brans–Dicke scalar–tensor theory. Recently, Penney[5] has given a conformally-flat solution to coupled massless-scalar and gravitational field equations.

In this work we present a procedure to obtain solutions to the stationary, axially-symmetric gravitational field coupled to massless-scalar and nonnull Maxwell fields. We show that starting from any solution to the electrovacuum field equations it is possible to generate a whole class of solutions to the coupled Einstein–Maxwell-massless scalar field equations by a suitable redefinition of one of the space–time metric coefficients.

2. SOLUTIONS TO FIELD EQUATIONS

We start by considering stationary, axially-symmetric space–times where the sources for the geometry are massless scalar and source-free, nonnull electromagnetic fields. The equations to be solved are

$$R_{\mu\nu} = -\kappa(\Phi_{;\mu}\Phi_{;\nu} + F_{\mu\alpha}F_{\nu}{}^{\alpha} - \tfrac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}), \tag{1}$$

$$g^{\alpha\beta}\Phi_{;\alpha\beta} = 0, \tag{2}$$

$$F^{\mu\nu}{}_{;\nu} = 0, \quad F_{[\mu\nu;\alpha]} = 0, \tag{3}$$

where Φ stands for the scalar field,

$$F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu}, \tag{4}$$

are components of the electromagnetic field tensor and the semicolon denotes the covariant derivative. For stationary, axially-symmetric fields the relation

$$R^{0}{}_{0} + R^{3}{}_{3} = 0, \tag{5}$$

enables us to write the space–time line element in the Weyl–Papapetrou canonical form

$$ds^2 = -\exp(2\psi)(dt - \omega d\phi)^2 + \exp(-2\psi)[\exp(2\gamma)(d\rho^2 + dz^2) + \rho^2 d\phi^2], \tag{6}$$

as in the electrovacuum case, and Eq. (2) reads

$$\nabla^2\Phi = 0, \tag{7}$$

where ∇^2 is the flat-space Laplace operator in cylindrical coordinates. The coefficients for the space–time metric ψ, ω and the nonzero components of the electromagnetic vector potential A_0 and A_3 satisfy exactly the coupled electrovacuum field equations[6] independent of the scalar field Φ. The remaining field equations

$$\begin{aligned}\gamma_{,\rho} = {} & \rho[(\psi_{,\rho})^2 - (\psi_{,z})^2] - [\exp(4\psi)/4\rho][(\omega_{,\rho})^2 - (\omega_{,z})^2] \\ & + (\kappa\rho/2)[(\Phi_{,\rho})^2 - (\Phi_{,z})^2] \\ & + (\kappa\rho/2)\exp(-2\psi)[(A_{0,\rho})^2 - (A_{0,z})^2] \\ & - [\kappa\exp(2\psi)/4\rho][(A_{3,\rho} + \omega A_{0,\rho})^2 \\ & - (A_{3,z} + \omega A_{0,z})^2],\end{aligned} \tag{8}$$

$$\begin{aligned}\gamma_{,z} = {} & 2\rho\psi_{,\rho}\psi_{,z} - [\exp(4\psi)/2\rho]\omega_{,\rho}\omega_{,z} + \kappa\rho\Phi_{,\rho}\Phi_{,z} \\ & + \kappa\rho\exp(-2\psi)A_{0,\rho}A_{0,z} \\ & - [\kappa\exp(2\psi)/\rho](A_{3,\rho} + \omega A_{0,\rho})(A_{3,z} + \omega A_{0,z}),\end{aligned} \tag{9}$$

and

$$\begin{aligned}& \gamma_{,\rho\rho} + \gamma_{,zz} + (\psi_{,\rho})^2 + (\psi_{,z})^2 + [\exp(4\psi)/4\rho^2][(\omega_{,\rho})^2 + (\omega_{,z})^2] \\ & + (\kappa/2)[(\Phi_{,\rho})^2 + (\Phi_{,z})^2] = 0\end{aligned} \tag{10}$$

enable us to define

$$\gamma = \gamma^{v} + \gamma^{\Phi}, \tag{11}$$

where γ^{v} can be evaluated by integration once a solution to ψ, ω, A_0, and A_3 is known and γ^{Φ} depends only on the scalar field Φ as

$$2\gamma^{\Phi}_{,\rho} = \kappa\rho[(\Phi_{,\rho})^2 - (\Phi_{,z})^2], \tag{12}$$

$$2\gamma^{\Phi}_{,z} = 2\kappa\rho\Phi_{,\rho}\Phi_{,z}, \tag{13}$$

which again can be evaluated by integration once a solution of Eq. (7) is specified. Thus we may state the following theorem.

Theorem: If $\psi, \omega, \gamma^{v}, A_0$, and A_3 form a solution to the electrovacuum field equations for the metric (6), then $\psi, \omega, \gamma, A_0$, and A_3, where

$$\gamma = \gamma^{v} + \gamma^{\Phi}, \tag{14}$$

$$2\gamma^{\Phi}_{,\rho} = \kappa\rho[(\Phi_{,\rho})^2 - (\Phi_{,z})^2], \tag{15}$$

$$2\gamma^{\Phi}_{,z} = 2\kappa\rho\Phi_{,\rho}\Phi_{,z}, \tag{16}$$

$$\nabla^2\Phi = 0, \tag{17}$$

is the corresponding solution to the coupled Einstein–Maxwell-massless scalar field (E.M.S.) equations for the metric given by Eq. (6).

3. CONCLUDING REMARKS

We have shown that for stationary, axially-symmetric electrovacuum space–times presence of a massless scalar field as an additional source for the geometry merely results in a redefinition of one of the metric coefficients, provided, of course, that the line element is expressed in the Weyl–Papapetrou canonical form (6). Thus starting from any known solution of the coupled Einstein–Maxwell field equations it is possible to generate solutions to the E.M.S. equations by application of the theorem stated. As a simple example, starting from flat space we obtain

$$ds^2 = -dt^2 + \exp(2\gamma^{\Phi})(d\rho^2 + dz^2) + \rho^2 d\phi^2, \tag{18}$$

solutions describing a space–time where the source is a massless scalar field only. Application of the theorem to electrovacuum solutions of the Tomimatsu and Sato[7] and Ernst[6,8] results in a class of five-parameter solutions of the (E.M.S.) equations with parameters describing mass (m), electromagnetic charge (e), scalar charge (A), rotation (a), and deformation (δ). To obtain the solution of Janis, Newman and Winicour in the limit, it will be enough to take $a=0$, $e=0$, and $\delta^{-2} - \kappa A^2 = 1$, which reduces the solution to a static, spherically-symmetric one.

Finally, we would like to add that it is also possible to generate the corresponding Einstein–Maxwell-conformal scalar field solutions[9] and solutions of the Brans–Dicke scalar–tensor theory.[4,10]

*Research supported in part by the Turkish Scientific and Technical Research Council (T. B. T. A. K.).

[1]A.I. Janis, E.T. Newman, and J. Winicour, Phys. Rev. Lett. 20, 878 (1968).
[2]R. Penney, Phys. Rev. 174, 1578 (1968).
[3]R. Penney, Phys. Rev. 182, 1383 (1969).
[4]A.I. Janis, D.C. Robinson, and J. Winicour, Phys. Rev. 186, 1729 (1969).
[5]R. Penney, Phys. Rev. D 14, 910 (1976).
[6]F.J. Ernst, Phys. Rev. 168, 1415 (1968).
[7]A. Tomimatsu and H. Sato, Phys. Rev. Lett. 29, 1344 (1972); Prog. Theor. Phys. 50, 95 (1973).
[8]F.J. Ernst, Phys. Rev. D 7, 2520 (1973).
[9]J.D. Bekenstein, Ann. Phys. (N.Y.) 82, 535 (1973).
[10]C.B.G. McIntosh, Commun. Math. Phys. 37, 335 (1974).

Fluid space-times including electromagnetic fields admitting symmetry mappings belonging to the family of contracted Ricci collineations

L. K. Norris, L. H. Green, and W. R. Davis

Department of Physics, North Carolina State University, Raleigh, North Carolina 27607
(Received 28 December 1976)

This paper investigates certain symmetry mappings belonging to the family of contracted Ricci collineations (FCRC) (satisfying $g^{ij}\pounds R_{ij}=0$) admitted by the general fluid space-times, including electromagnetic fields, that were classified and studied earlier by Stewart and Ellis (1967). Many of the results obtained are applicable to the perfect fluid models treated by Wainwright (1970) and Krasiński (1974,1975). A major part of this paper represents an extension of previous investigations (1976) of the Robertson–Walker metrics and more general perfect fluid space-times that admit FCRC symmetry mappings and concomitant conservation expressions. More specifically, these results provide a number of theorems relating to the more general fluid space-times that admit FCRC symmetry mappings (including both timelike and spacelike symmetry vectors) that lead to conservation expressions and specific conditions on the metric tensors for the given particular cases of these space-times. Also the form of the symmetry mappings induced on the electromagnetic fields (when they are present) is investigated in the case where specific symmetry mappings on the metric tensor are admitted. In particular, the results of Wainwright and Yaremovicz (1976) relating to homothetic motions admitted by given space-times, corresponding to perfect fluids including electromagnetic fields, are largely embraced by the more general results obtained in this paper.

1. INTRODUCTION

In several recent publications[1–3] an important family of symmetry mappings, called the family of contracted Ricci collineations (FCRC) (satisfying $g^{ij}\pounds R_{ij}=0$) was introduced together with considerations of certain concomitant field conservation expressions. Figure 1 of this paper provides a symmetry property inclusion diagram[4] which indicates that the FCRC embraces a large family of symmetry properties that includes as special cases more familiar symmetry properties [e.g., motions (M), affine collineations (AC), and Ricci collineations (RC)]. In particular, the FCRC has been investigated in the case of the Robertson–Walker cosmological models[2] and for more general perfect fluid space–times.[3]

In this paper we extend the work of the above mentioned papers in several ways. In Sec. 2 we investigate timelike FCRC mappings (in the direction of the timelike eigenvector of the Ricci tensor) for a large class of exact solutions of the Einstein equations which exhibit local rotational symmetry (LRS) that may be characterized as perfect fluids which include "noninteracting" electromagnetic fields.[5] Most of the fluid models that are included in the present investigation were discussed in an earlier paper by Stewart and Ellis,[6] who classified various particular cases of these models using dynamical and kinematical quantities. These models, when specialized to perfect fluids in the absence of electromagnetic fields, are particular cases of the algebraically special perfect fluid models studied by Wainwright.[7] In particular, we give a number of theorems that give the necessary and sufficient conditions that LRS perfect fluid space–times admit particular FCRC symmetry mappings. These theorems could be viewed as a partial elaboration of an invariant type of classification scheme based on the FCRC admitted by particular models.

In Sec. 3 we study spacelike FCRC symmetry mappings. Most of the results of this section will apply to

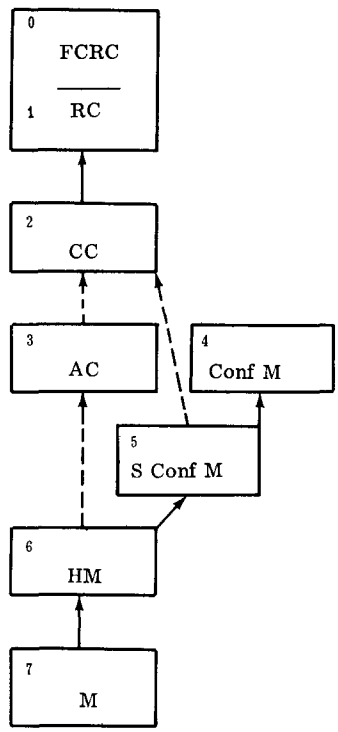


FIG. 1. Symmetry property inclusion diagram[4]. 0. FCRC-members of family of contracted Ricci collineations: $g^{ij}\pounds R_{ij}=0$. 1. RC-Ricci collineation: $\pounds R_{ij}=0$. 2. CC-curvature collineation: $\pounds R_{jkm}{}^{i}=0$. 3. AC-affine collineation: $\pounds \Gamma^{i}_{jk}=0$. 4. Conf M-conformal motion: $\pounds g_{ij}=2\sigma g_{ij}$. 5. S Conf M-special conformal motion: $\pounds g_{ij}=2\sigma g_{ij}$, $\nabla_i\nabla_j\sigma=0$. 6. HM-homothetic motion: $\pounds g_{ij}=2\sigma g_{ij}$, $\sigma=\text{const}$. 7. M-motion: $\pounds g_{ij}=0$.

LRS perfect fluid space–times which include electromagnetic fields that were studied earlier by Stewart and Ellis[6] and to the perfect fluid models (with no electromagnetic fields) recently considered by Kransiński[8] and Wainwright.[7] More specifically we give an extension of a generalized Kelvin–Helmholtz theorem found previously by Oliver and Davis.[3]

In Sec. 4 we consider symmetry mappings on the electromagnetic field tensor and related mappings on the space–time metric which can be members of the FCRC. In particular, the results of Wainwright and Yaremovicz[9] and others[10] relating to homothetic motions (HM) and motions admitted by space–times, corresponding to perfect fluids including electromagnetic fields, are largely embraced by the more general results obtained in this section.

In an appendix we give the necessary detailed results required for the considerations in this paper relating to computations of various quantities characterizing the exact solutions studied by Stewart and Ellis. Throughout this paper statements such as "Case II..." will refer to the classification of the LRS space–times given in the appendix.

2. TIMELIKE FCRC SYMMETRY MAPPINGS FOR LRS PERFECT FLUID SPACE-TIMES

The family of contracted Ricci collineations (FCRC) is defined by conditional relations of the from

$$\pounds R_{ij} = H_{ij}, \quad g^{ij}H_{ij} = 0, \tag{2.1}$$

where H_{ij} is any trace-free symmetric tensor that is not identically equal to $\pounds R_{ij}$. Here $\pounds$ denotes the operation of Lie differentiation with respect to the vector η^i where $x^i \to x^i + \epsilon\eta^i$. When, for a particular H_{ij} in a given space–time, a vector η^i satisfying (2.1) can be determined to within a multiplicative constant, then we say that the given space–time admits this FCRC symmetry property. When a particular choice of H_{ij} in a given space–time does not permit the determination of the vector η^i (e.g., when η^i is only determined to within an arbitrary multiplicative function f satisfying $\eta^i\partial_i f = 0$) then we term the mapping an FCRC quasisymmetry property. For both FCRC symmetry and quasisymmetry properties it can be shown that the defining relation (2.1) leads to the conservation expression[11]

$$\nabla_i\nabla_j[-\sqrt{-g}\,(\tfrac{1}{2}\pounds g^{ij} + g^{ij}\nabla_k\eta^k)] \equiv \nabla_i(\sqrt{-g}\,R^i_j\eta^j) \equiv \partial_i(\sqrt{-g}\,R^i_j\eta^j) = 0. \tag{2.2}$$

Throughout the rest of this section we restrict our attention to LRS perfect fluid space–times in the absence of electromagnetic fields. We shall also assume that $2\rho_0 = \mu + 3p > 0$.

For perfect fluid space–times with matter tensors of the form

$$T_{ij} = \mu u_i u_j - p\gamma_{ij}; \tag{2.3}$$

we will consider timelike symmetry mapping vectors of the form $\eta^i = \varphi u^i$, i.e., symmetry mappings along the direction of the matter flow. In a recent paper it was shown[2] that for mapping vectors of the form $\eta^i = \varphi u^i$, $\pounds R_{ij}$ may be expressed in the form

$$\pounds R_{ij} = \psi(u_i u_j - \tfrac{1}{4}g_{ij}) + 2\rho_0 u_{(i}\gamma^k_{j)}(a_k\varphi + \partial_k\varphi) + 2\rho\varphi\sigma_{ij} + \tfrac{1}{2}g_{ij}\nabla_k(\rho_0\varphi u^k), \tag{2.4}$$

where it has been assumed that $\rho_0 = (\mu + 3p)/2 > 0$, where $R_{ij} = \rho_0 u_i u_j + \rho\gamma_{ij}$ and

$$\psi \equiv 2\nabla_k[\rho_0\varphi u^k] - \tfrac{4}{3}(2\rho_0)^{1/2}\varphi\nabla_k[(2\rho_0)^{1/2}u^k]. \tag{2.5}$$

We observe that no sum of terms on the right-hand side of (2.4) can vanish unless each individual term in the given sum vanishes. Note that if $\eta^i = \varphi u^i$ is an FCRC mapping vector then the last term on the right-hand side of (2.4) will vanish by (2.2). These remarks suggest the following three distinct types of choices for the tensor H_{ij}[12]:

$$\overset{1}{H}_{ij} = \Lambda(u_i u_j - \tfrac{1}{4}g_{ij}) + 2\rho\varphi\sigma_{ij}, \tag{2.6}$$

$$\overset{2}{H}_{ij} = 2\rho_0 u_{(i}\gamma^k_{j)}[a_k\varphi + \partial_k\varphi] + 2\rho\varphi\sigma_{ij}, \tag{2.7}$$

$$\overset{3}{H}_{ij} = \Lambda(u_i u_j - \tfrac{1}{4}g_{ij}) + 2\rho_0 u_{(i}\gamma^k_{j)}[a_k\varphi + \partial_k\varphi], \tag{2.8}$$

where

$$\Lambda \equiv -\tfrac{8}{3}(\rho_0^{1/2}\varphi)\nabla_i[\rho_0^{1/2}u^i]. \tag{2.9}$$

We now proceed to investigate in turn each of these types of symmetry demands in the case of LRS perfect fluid space–times.

Theorem 2.1: *A LRS perfect fluid space–time with* $\rho_0 > 0$ *admits the timelike FCRC symmetry property* $\pounds R_{ij} = \overset{1}{H}_{ij}$ *with* $\eta^i = \varphi u^i$ *if and only if* (i) $\varphi = f(x^1)/(XY^2\rho_0)$ *and* (ii) $\partial_1(\varphi F) = 0$.

Proof: First, assume that this symmetry property is admitted. From the identity (2.4), setting $\pounds R_{ij} = \overset{1}{H}_{ij}$ with $\eta^i = \varphi u^i$ and $\rho_0 > 0$, implies (a) $\nabla_i(\rho_0\varphi u^i) = 0$ and (b) $a_k\varphi + \gamma^i_k\partial_i\varphi = 0$. Condition (a) can be solved for φ in the form $\varphi = f(x^1, x^2, x^3)/(XY^2\rho_0)$, f arbitrary. Using the results given in the Appendix we find (b) equivalent to the three conditions: (i) $\partial_1(\ln\varphi F) = 0$, (ii) $\partial_2(\ln\varphi) = 0$, and (iii) $y\partial_0(\ln\varphi) - h\partial_1(\ln F) + \partial_3\ln\varphi = 0$. However, because of the properties of the functions y, h, F, and $XY^2\rho_0$ in the three main cases, (iii) reduces to $\partial_3\ln\varphi = 0$. The proof of the converse is straightforward.

Using the properties listed in the Appendix for the major cases it is easy to show that in Cases I and III this timelike FCRC symmetry property is admitted, where $\varphi = (\text{const})/F$ and $\varphi = (\text{const})/(XY^2\rho_0)$ in Case I and Case III, respectively. In fact, because of the special properties of the solutions, in Case I $\overset{1}{H}_{ij} \equiv 0$ and hence this type of symmetry would correspond to Ricci collineations. Moreover, it is not difficult to show that this case I RC degenerates to a motion. Applying the conditions of the above theorem to Case II we find $\varphi = f(x^1)/(XY^2\rho_0)$ where φ must satisfy the additional condition $\partial_1(\varphi F) = 0$. If we assume an equation of state of the form $p = (\gamma - 1)\mu$, $1 < \gamma \leq 2$, and put $a_k = 0$ then we may consider the homogeneous solutions in subcase IIa considered by Stewart and Ellis.[13] All of these homogeneous solutions (containing for example a generalized Einstein–de Sitter universe) admit the timelike FCRC symmetry property of Theorem 2.1 with $f = \text{const}$. Subcase IIb contains the Robertson–Walker solutions with $a_k = \sigma_{ij} = 0$. For these solutions we find that the above timelike symmetry reduces to the symmetry property

found earlier[2] with $\varphi = (\text{const})/\ddot{\mathcal{J}}\mathcal{J}^2$, where $\mathcal{J}(x^0)$ is the "cosmic scale factor"[14] and where $\ddot{\mathcal{J}} \equiv \partial_0\partial_0\mathcal{J}$.

Theorem 2.2: *A LRS perfect fluid space–time, with* $\rho_0 > 0$, *admits the timelike FCRC quasisymmetry property* $\mathcal{L}R_{ij} = \overset{2}{H}_{ij}$, *with* $\eta^i = \varphi u^i$, *if and only if* (i) $\varphi = f(x^1, x^2, x^3)/(XY^2\rho_0)$, f *arbitrary and* (ii) $\partial_0[X^2Y^4\rho_0] = 0$.

Proof: Assume first that the symmetry $\mathcal{L}R_{ij} = \overset{2}{H}_{ij}$ with $\eta^i = \varphi u^i$ is admitted. From (2.4) we observe that setting $\mathcal{L}R_{ij} = \overset{2}{H}_{ij}$, $\eta^i = \varphi u^i$, implies (a) $\nabla_i[\rho_0^{1/2}u^i] = 0$ and (b) $\nabla_i[\rho_0\varphi u^i] = 0$. These two conditions together imply (c) $u^i\partial_i\varphi = \varphi\theta$, while condition (b) can be solved for φ in the form $\varphi = f(x^1, x^2, x^3)/(XY^2\rho_0)$, f arbitrary. Using this expression for φ and the expression for θ given in the Appendix in (c) we find $\partial_0[X^2Y^4\rho_0] = 0$.

Conversely, assuming conditions (i) and (ii) for $\eta^i = \varphi u^i$ one can reverse the steps of the above proof and show $\mathcal{L}R_{ij} = \overset{2}{H}_{ij}$.

We observe that Case I solutions satisfy condition (ii) of this theorem and hence admit this quasisymmetry. There is a subcase IIa solution, given by Stewart and Ellis[13] with an equation of state of the form $\mu = p$ that satisfies condition (ii) of this theorem and thus also admits this FCRC quasisymmetry. None of the special Case III solutions presented by Stewart and Ellis admit this quasisymmetry property.

Theorem 2.3: *A LRS perfect fluid space–time with* $\rho_0 > 0$ *admits the timelike FCRC quasisymmetry property* $\mathcal{L}R_{ij} = \overset{3}{H}_{ij}$ *with* $\eta^i = \varphi u^i$, *if and only if* (i) $\varphi = f(x^1, x^2, x^3)/(XY^2\rho_0)$ *and either* (ii) $\sigma_{ij} = 0$ *or* (iii) $\rho = 0$.

Proof: The proof of this theorem follows from the identity (2.4) in a manner similar to the proofs of the preceding theorems.

We observe that because of the special properties of Case I solutions $\overset{2}{H}_{ij} = \overset{3}{H}_{ij}$ (in Case I). Hence Theorems (2.2) and (2.3) are equivalent in Case I. Case IIb solutions contain the Robertson–Walker models, which are shear free, and hence admit this FCRC quasisymmetry. Stewart and Ellis list, for subcase IIa, a generalization of the Einstein–deSitter universe.[13] This model is also shear free and hence satisfies condition (ii) of Theorem 2.3 and thus admits the FCRC quasisymmetry of this theorem. For futher discussion of the physical interpretation of the conservation laws that follow in consequence of timelike members of the FCRC being admitted we refer the reader to the relevant papers[1–3] mentioned in the Introduction.

3. SPACELIKE FCRC SYMMETRY MAPPINGS FOR LRS PERFECT FLUID SPACE-TIMES

In terms of the definitions, Ricci collineations are evidently the simplest proper members of the FCRC. Here we will consider first under what conditions space–times, corresponding to LRS perfect fluids that include electromagnetic fields, admit RC which have symmetry vectors η^i of the form $\eta^i = \varphi\overset{1}{h}{}^i$.

Theorem 3.1: *A LRS perfect fluid space–time admits a RC with* $\eta^i = \varphi\overset{1}{h}{}^i$ *if and only if* (i) $\partial_1(\rho_0/F^2) = 0$, (ii) $\partial_1(\varphi^2\rho_1) = 0$, (iii) $\partial_1(\rho_2Y^2) = 0$ *and* (iv) *either* $\rho_1 = 0$ *or* $\partial_0(\varphi/X) = 0$, $\partial_2\varphi = 0$, $\partial_3\varphi = 0$.

Proof: Assume a RC is admitted by the given space–time. The Ricci tensor can be expressed in the form $R_{ij} = \sum_{\alpha=0}^{3} e_\alpha\rho_\alpha\overset{\alpha}{h}{}_i\overset{\alpha}{h}{}_j$. Therefore, the RC condition now becomes

$$\sum_{\alpha=0}^{3} e_\alpha\{\overset{\alpha}{h}{}_j\overset{\alpha}{h}{}_i\mathcal{L}\rho_\alpha + \rho_\alpha[\overset{\alpha}{h}{}_i\mathcal{L}\overset{\alpha}{h}{}_j + \overset{\alpha}{h}{}_j\mathcal{L}\overset{\alpha}{h}{}_i]\} = 0.$$

Forming the independent projections of the above equation yields: (a) $\mathcal{L}\rho_\alpha + 2e_\alpha\rho_\alpha\overset{\alpha}{h}{}^i\mathcal{L}\overset{\alpha}{h}{}_i = 0$, $\alpha = 0,1,2,3$ and (b) $\rho_\alpha\overset{\beta}{h}{}^i\mathcal{L}\overset{\alpha}{h}{}_i + \rho_\beta\overset{\alpha}{h}{}^i\mathcal{L}\overset{\beta}{h}{}_i = 0$, $\alpha \neq \beta$, $\alpha, \beta = 1,2,3$. Taking $\alpha = 0,1$, and 2 (or 3) in (a) gives $\overset{1}{h}{}^i\partial_i(\rho_0/F^2) = 0$, $\overset{1}{h}{}^i\partial_i(\varphi^2\rho_1) = 0$, and $\overset{1}{h}{}^i\partial_i(\rho_2Y^2) = 0$, respectively. Taking $\alpha = 1$, $\beta = 0$ in (b) gives $\rho_1 = 0$ or $\overset{1}{h}{}^i\overset{1}{h}{}^j\sigma_{ij} - \frac{1}{3}\theta = -u^j\partial_j\ln\varphi$. Using the results given in the Appendix, the last equation can be re-expressed as $\partial_0(\varphi/X) = 0$. Letting $\alpha = 1$, $\beta = 2$ in (b) gives $\rho_1 = 0$ or $\overset{2}{h}{}^j\partial_j\varphi = 0$. Letting $\alpha = 1$, $\beta = 3$ in (b) gives $\overset{3}{h}{}^j\partial_j\varphi = 0$ or $\rho_1 = 0$. If $\rho_1 \neq 0$ then $\overset{3}{h}{}^j\partial_j\varphi = 0$ implies $y\partial_0\varphi + h\partial_1\varphi + \partial_3\varphi = 0$. From the classification scheme given in the Appendix, if $y \neq 0$ then $X = 1$ and using $\partial_0(\varphi/X) = 0$ we see that $\partial_0\varphi = 0$. Similarly if $h \neq 0$ then $\rho_1 = \rho_1(x^0)$ and using $\partial_1(\varphi^2\rho_1) = 0$ we see that $\partial_1\varphi = 0$. Therefore, the condition $\overset{3}{h}{}^i\partial_i\varphi = 0$ reduces to $\partial_3\varphi = 0$. The rest of the proof follows in a straightforward manner.

Neither subcase Ib nor subcase Ic admits a RC with a symmetry vector of the form $\eta^i = \varphi\overset{1}{h}{}^i$. In subcase Ia all of the conditions of the above theorem are met, however, the RC degenerates to a motion. In Case II the conditions of the theorem are satisfied only in special cases. All Case III space–times satisfy the above theorem, however, the RC degenerates to a motion.

Next we consider a more general type of FCRC mapping which is admitted by some of the exact solutions given by Stewart and Ellis.

Theorem 3.2: *A matter fluid space–time, that is* $R_{ij} = \rho_0u_iu_j + S_{ij}$ $(u^jS_{ij} = 0)$, *admits the FCRC quasisymmetry mapping* $g^{ij}\mathcal{L}R_{ij} = 0$, $u^iu^j\mathcal{L}R_{ij} = 0$ *with* $\eta^i = \varphi\omega^i$ *if and only if* (i) $\nabla_i(\rho_0\omega^i) = 0$ *and* (ii) $\nabla_i(\varphi S^i_j\omega^j) = 0$.

Proof: One can easily show that $u^iu^j\mathcal{L}R_{ij} = 0$ is equivalent to $\eta^k\partial_k\rho_0 - 2\rho_0\eta^ka_k = 0$. Using $\eta^k = \varphi\omega^k$ and the identity $\nabla_k\omega^k + 2a_k\omega^k = 0$ we find $\nabla_i(\rho_0\omega^i) = 0$. The second condition follows from conservation expression (2.2).

This theorem[15] will be discussed presently within the context of a generalized Kelvin–Helmholtz theorem.

We now restrict ourselves to perfect fluid space–times in the absence of electromagnetic fields. For the case of a perfect fluid with $\eta^ku_k = 0$ one finds the general result

$$\mathcal{L}R_{ij} = \psi(u_iu_j - \tfrac{1}{4}g_{ij}) + 2(\rho_0 - \rho)(u_i\omega_{kj} + u_j\omega_{ki})\eta^k + \rho(L\eta)_{ij} + \tfrac{1}{2}g_{ij}\nabla_k(\rho\eta^k), \tag{3.1}$$

where $R_{ij} = \rho_0u_iu_j + \rho\gamma_{ij}$, $\psi = \eta^k\partial_k(\rho_0 - \rho) - \eta^ka_k(\rho_0 - \rho)$, and $(L\eta)_{ij} \equiv \nabla_i\eta_j + \nabla_j\eta_i - \frac{1}{2}(\nabla_k\eta^k)g_{ij}$. This decomposition of $\mathcal{L}R_{ij}$ suggests several possible distinct types of spacelike members of the FCRC that could be of interest. We observe that if $\mathcal{L}R_{ij} = \psi(u_iu_j - \frac{1}{4}g_{ij}) + 2(\rho_0 - \rho)\times(u_i\omega_{kj} + u_j\omega_{ki})\eta^k$, then $(L\eta)_{ij} = 0$, hence the FCRC degenerates to a conformal motion, where η^i

satisfies $\nabla_j\nabla^j(\nabla_i\eta^i)=0$. We also observe that if $\pounds R_{ij}=\psi(u_iu_j-\frac{1}{4}g_{ij})+\rho(L\eta)_{ij}$ then $\omega^k=\lambda\eta^k$ for some λ.

Theorem 3.3: *A perfect fluid space–time admits the FCRC quasisymmetry property* $\pounds R_{ij}=\rho(L\eta)_{ij}$ *with* $\eta^i=\varphi\omega^i$ *if and only if* (i) $\nabla_k[(\rho_0-\rho)^2\omega^k]=0$ *and* (ii) $\nabla_k(\varphi\rho\omega^k)=0$.

Proof: Assume $\pounds R_{ij}=\rho(L\eta)_{ij}$. Comparing this relation with the identity mapping (3.1) we find $\psi(u_iu_j-\frac{1}{4}g_{ij})+\frac{1}{2}g_{ij}\nabla_k(\rho\eta^k)=0$. If we take the trace of this equation we find $\nabla_k(\rho\eta^k)=0$. Since $\nabla_k(\rho\eta^k)=0$ either $\psi=0$ or $(u_iu_j-\frac{1}{4}g_{ij})=0$. However, $(u_iu_j-\frac{1}{4}g_{ij})$ cannot be zero, thus $\psi=0$. If $\psi=0$, then $\eta^k\partial_k(\rho_0-\rho)-\eta^k a_k(\rho_0-\rho)=0$. Using the identity $\nabla_k\omega^k+2a_k\omega^k=0$ we find $\nabla_k[(\rho_0-\rho)^2\omega^k]=0$. The rest of the proof follows in a similar manner.

We now interpret the previous two theorems using some results obtained by Greenberg.[16] In particular Greenberg demonstrated that

$$\frac{1}{A}\frac{DA}{D\tau}=-\frac{1}{\omega}\frac{D\omega}{D\tau}-\frac{a_i\omega^i}{\omega}, \tag{3.2}$$

where A is the proper area subtended by the vortex lines as they pass through the screen which is the 2-surface dual to the surface formed by u_i and ω_i and where $DA/D\tau=(\omega^i/\omega)\partial_iA$. If the space–time under consideration admits a FCRC symmetry mapping with symmetry vector η^i of the form $\eta^i=\varphi\omega^i$ and ω^i is an eigenvector of the Ricci tensor (with eigenvalue ρ_ω) we have from the field conservation expression (2.2) $\nabla_k(\varphi\rho_\omega\omega^k)=0$. Using $\nabla_i\omega^i+2a_i\omega^i=0$ along with Greenberg's result we find

$$\frac{D}{D\tau}(\varphi^{1/2}\rho_\omega^{1/2}\omega A)=0. \tag{3.3}$$

Thus we have that $\varphi^{1/2}\rho_\omega^{1/2}\,\omega A$ is a constant along the vortex flow. This result may be regarded as a generalization of the Kelvin–Helmholtz theorem[17] of Newtonian fluid theory which essentially states that ωA is constant along the vortex flow.[16]

Next we examine the above results in the context of certain Case I solutions considered by Stewart and Ellis. In particular, we note that the FCRC quasisymmetry mapping $g^{ij}\pounds R_{ij}=2\nabla_i(R^i_j\eta^j)=0$ with $\eta^i=\varphi\omega^i$ is admitted in all Case I solutions if φ is of the form $\varphi=f(x^0,x^2,x^3)F/\rho_1\omega Y^2$ (where $\rho_1\neq0$ and f is an arbitrary function of x^0, x^2, and x^3 not determined by the symmetry demand). Therefore, in accord with the previous discussion we find $\varphi^{1/2}\rho_1^{1/2}\omega A=[f(x^0,x^2,x^3)F\omega/Y^2]^{1/2}A$ is conserved along the vortex flow. Stewart and Ellis consider a particular exact solution which corresponds to a Case I barotropic fluid which has acceleration and rotation. This particular solution is given by:[18] $\omega=\omega_0/F(x^1)$ $(\omega_0\neq0)$; $p=\Lambda-r+\tau-\omega_0^2/F^2$; $\mu=-\Lambda+r-\tau+3\omega_0^2/F^2$, and $x^1=\int[cF^4+(2\tau-r)F^2-\omega_0^2]^{-1/2}dF$ where c,τ,r are constant. Thus, in this particular case the quantity $\varphi^{1/2}\rho_1^{1/2}\omega A=[f(x^0,x^2,x^3)F\omega/Y^2]^{1/2}A=(f\omega_0)^{1/2}A$ is conserved along the vortex flow.

If we specialize the above model by setting $2\tau-r=0$ and $2\omega_0^2=c$ then the conditions of Theorem 3.2 are satisfied. This then gives the following two conservation expressions: (i) $\nabla_k(\rho_0\omega^k)=0$ and (ii) $\nabla_k(\varphi\rho_1\omega^k)=0$. For this special case one finds $\rho_0=0$ $(\mu=-\Lambda+\tau+3\omega_0^2/F^2$, $p=\Lambda-\tau-\omega_0^2/F^2)$ hence, the conservation expression $\nabla_k(\rho_0\omega^k)=0$ is trivially satisfied and no Kelvin–Helmholtz type theorem is obtained. The conservation expression $\nabla_k(\varphi\rho_1\omega^k)$ is the same expression as discussed above which does provide a type of generalized Kelvin–Helmholtz conservation expression.

4. SYMMETRY MAPPINGS OF THE ELECTROMAGNETIC FIELD

It is of interest, in connection with a study of symmetry mappings admitted by perfect fluids including electromagnetic fields, to consider the explicit form of symmetry mappings of the electromagnetic field tensor. Several aspects of this problem have received considerable attention in the recent literature.[9,10,19] Here we are particularly interested in considering symmetry mappings of the electromagnetic field as they relate to the family of contracted Ricci collineations.

Wainwright and Yaremovicz[9] have studied perfect fluid space–times with null and nonnull electromagnetic fields present, under the assumption that the space–time admits a homothetic motion (HM, i.e., $\pounds g_{ij}=2\sigma g_{ij}$, $\sigma=$ const). In accord with this assumption, together with the field equations, they show that the electromagnetic field tensor F_{ij} and its dual ${}^*F_{ij}$ satisfy

$$\pounds F_{ij}=\sigma F_{ij}+\tilde{\sigma}{}^*F_{ij} \tag{4.1}$$

and

$$\pounds {}^*F_{ij}=\sigma{}^*F_{ij}-\tilde{\sigma}F_{ij} \tag{4.2}$$

with $\tilde{\sigma}=\pounds\alpha$ where α is the complexion[20] of the electromagnetic field. We observe that the pair of symmetry mappings (4.1) and (4.2) have the interesting property that they imply $g^{ij}\pounds\tau_{ij}=0$, independent of the form of $\pounds g_{ij}$, where $\tau_{ij}=-\frac{1}{2}(F_{ir}F_j{}^r+{}^*F_{ir}{}^*F_j{}^r)$ is the electromagnetic stress-energy–momentum tensor. Hence the symmetry mappings (4.1) and (4.2) on the electromagnetic field tensor and its dual induce an FCRC symmetry mapping for electrovac space–times. This suggests that a generalized form of the mappings (4.1) and (4.2) together with the corresponding generalized symmetry mapping on the space–time metric could lead to more general FCRC symmetry mappings in the case of matter space–times including electromagnetic fields. In order to find such a generalization we first note that (4.1) and (4.2) are special cases of mappings $\pounds F_{ij}$ and $\pounds{}^*F_{ij}$ which satisfy

$$\pounds F_{ij}=-\tfrac{1}{2}\eta_{ijkm}g^{ka}g^{mb}\pounds{}^*F_{ab}. \tag{4.3}$$

In particular if $\pounds F_{ij}$ and $\pounds{}^*F_{ij}$ satisfy (4.3) then we again have $g^{ij}\pounds\tau_{ij}=0$ independent of the form of $\pounds g_{ij}$. Of course, if (4.3) holds in a given space–time, conditions will be placed on the metric.

Theorem 4.1: *The electromagnetic field tensor* F_{ij} *and its dual* ${}^*F_{ij}$ *satisfy* (4.3) *if and only if the space–time metric salisfies* $\pounds g_{ij}=2\sigma g_{ij}+\Lambda_{ij}$, *where* Λ_{ij} *is a trace-free symmetric tensor that satisfies* ${}^*F^i_{\cdot k}\Lambda^{kj}={}^*F^j_{\cdot k}\Lambda^{ki}$.

Proof: By definition we have $F_{ij}=-\frac{1}{2}\eta_{ijkm}g^{kr}g^{ms}{}^*F_{rs}$. By taking the Lie derivative of both sides of this equation and assuming (4.3) we obtain ${}^*F_{rs}\pounds(\eta_{ijkm}g^{kr}g^{ms})=0$. Using $\pounds\eta_{ijkm}=\eta_{ijkm}\nabla_r\eta^r=\frac{1}{2}\eta_{ijkm}g^{rs}\pounds g_{rs}$ we may rewrite

this last equation as (a) $\eta_{ijkm}{}^*F^k_{\cdot r}(\pounds g^{mr} + \frac{1}{4}g^{mr}g^{ab}\pounds g_{ab}) = 0$. Defining $\Lambda^{ij} = -\pounds g^{ij} - \frac{1}{4}g^{ij}g^{km}\pounds g_{km}$ we obtain from (a) that ${}^*F^{ik}_{\cdot j}\Lambda^{mj} = 0$ as desired. The converse is proved by reversing the steps of the above proof.

We note that by starting with ${}^*F_{ij} = \frac{1}{2}\eta_{ijkm}g^{kr}g^{ms}F_{rs}$, assuming $\pounds {}^*F_{ij} = \frac{1}{2}\eta_{ijkm}g^{kr}g^{ms}\pounds F_{rs}$ [which is equivalent to (4.3)] and repeating the steps in the above proof one can derive the equivalent conditions $F^{ik}_{\cdot j}\Lambda^{mj} = 0$ on Λ^{ij}.

This theorem demonstrates that the general relationship (4.3) between $\pounds F_{ij}$ and $\pounds {}^*F_{ij}$ is induced by the mapping

$$\pounds g_{ij} = 2\sigma g_{ij} + \Lambda_{ij} \tag{4.4}$$

on the space–time metric, where Λ_{ij} must satisfy certain conditions. Only in the case $\Lambda_{ij} = 0$ does this mapping on the metric correspond to a conformal motion. An example of a nonzero symmetric tensor[21,22] Λ_{ij} that can satisfy the conditions of the above theorem is $\Lambda_{ij} = ak_ik_j$, where a is an arbitrary function and k_i is a null eigenvector of F_{ij} and ${}^*F_{ij}$.

Having determined the general mapping (4.4) on the metric that induces the relationship (4.3), we can now in turn use (4.4) to determine the explicit form of the mappings on F_{ij} and ${}^*F_{ij}$. We formalize these results for the nonnull case in the following theorem.

Theorem 4.2: *A space–time containing a nonnull electromagnetic field admits the symmetry mapping* $\pounds g_{ij} = 2\sigma g_{ij} + \Lambda_{ij}$, *where* Λ_{ij} *is a trace-free symmetric tensor that satisfies* $F^i_{\cdot j}\Lambda^{jk} = F^k_{\cdot j}\Lambda^{ji}$, *if and only if*

$$\pounds F_{ij} = \sigma F_{ij} + \tilde{\sigma}{}^*F_{ij} + \Delta^{km}_{ij}\pounds \tau_{km} \tag{4.5}$$

and

$$\pounds {}^*F_{ij} = \sigma {}^*F_{ij} - \tilde{\sigma}F_{ij} + {}^*\Delta^{km}_{ij}\pounds \tau_{km}, \tag{4.6}$$

where $\tilde{\sigma} = \pounds \alpha$, $\Delta^{km}_{ij} = -(1/8f^4)F_{ij}\tau^{km} - 2(\tau^{ab}\tau_{ab})^{-1/2} \times \delta^{(k}_{(i}\{F^{m)}_{j)\cdot}\cos(2\alpha) - {}^*F^{m)}_{j)\cdot}\sin(2\alpha)\}$ *and* ${}^*\Delta^{km}_{ij} \equiv \frac{1}{2}\eta_{ijpq}g^{pa}g^{qb}\Delta^{km}_{ab}$. *Here* $\pounds \tau_{ij}$ *is to be evaluated using the field equations*, i.e., $\pounds \tau_{ij} = \pounds R_{ij} - \pounds (T_{ij} - \frac{1}{2}g_{ij}T) = -2\nabla_i\nabla_j\sigma - g_{ij}(\nabla_k\nabla^k\sigma) + \nabla_k\nabla_{(i}\Lambda^k_{j)} - \frac{1}{2}\nabla_k\nabla^k\Lambda_{ij} - \pounds (T_{ij} - \frac{1}{2}g_{ij}T)$ *where* T_{ij} *is the nonelectromagnetic part of the total matter tensor.*

Proof: Assume the space–time admits the symmetry mapping $\pounds g_{ij} = 2\sigma g_{ij} + \Lambda_{ij}$ where Λ_{ij} is a symmetric trace-free tensor which satisfies $F^i_{\cdot r}\Lambda^{rj} = F^j_{\cdot r}\Lambda^{ri}$. We now want to calculate $\pounds F_{ij}$. F_{ij} can be expressed in the form[20] $F_{ij} = f_{ij}\cos\alpha + {}^*f_{ij}\sin\alpha$ with $f_{ij} = f(k_in_j - k_jn_i)$ where f_{ij} is the so-called extremal field, α is the complexion, and n_i and k_j are the null eigenvectors of τ_{ij} ($n_ik^i = 1$). In order to evaluate $\pounds F_{ij}$ we need to evaluate $\pounds f_{ij}$. We do this by expressing $\pounds f$, $\pounds k_i$, $\pounds n_i$ in terms of $\pounds \tau_{ij}$. In the case of nonnull electromagnetic fields $\tau_{ij} = -f^2[g_{ij} - 2(k_in_j + n_ik_j)]$. From the relation $F^i_{\cdot r}\Lambda^{rj} = F^j_{\cdot r}\Lambda^{ri}$ one can show $\Lambda^i{}_r n^r = \epsilon k^i$, $\Lambda^i{}_r k^r = \gamma n^i$ and therefore $\tau^j_i\tau^{ik}\Lambda_{jk} = 0$. We find $\pounds f = -\sigma f + (1/8f^3)\tau^{ij}\pounds \tau_{ij}$, $\pounds k_j = [2\sigma - (1/8f^4)\tau^{mn}\pounds \tau_{mn} - k^m\pounds n_m]k_j + (k^i/2f^2)\pounds \tau_{ij}$, and $\pounds n_j = [2\sigma - (1/8f^4)\tau^{mn}\pounds \tau_{mn} - n^m\pounds k_m]n_j + (n^i/2f^2)\pounds \tau_{ij}$. Using these three equations we obtain $\pounds f_{ij} = [\sigma - (1/8f^4)\tau^{mn} \times \pounds \tau_{mn}]f_{ij} + (1/2f)[k_in^k\pounds \tau_{jk} + n_jk^k\pounds \tau_{ik} - k_jn^k\pounds \tau_{ik} - n_ik^k\pounds \tau_{jk}]$ and a similar result for $\pounds {}^*f_{ij}$. We now calculate $\pounds F_{ij}$ and find

$$\pounds F_{ij} = \sigma F_{ij} + (\pounds \alpha){}^*F_{ij} - [(\tau^{mn}\pounds \tau_{mn})/8f^4]F_{ij}$$
$$+\left(\frac{\cos\alpha}{2f}\delta^r_i\delta^s_j + \frac{\sin\alpha}{4f}\eta_{ij\cdot\cdot}{}^{rs}\right)(k_rn^m\pounds \tau_{sm} + n_sk^m\pounds \tau_{rm} - k_sn^m\pounds \tau_{mn} - n_rk^m\pounds \tau_{sm}).$$

Using the definitions of f_{ij} and F_{ij} given above we find Eq. (4.5). A similar result can be obtained for $\pounds {}^*F_{ij}$. We emphasize that $\pounds \tau_{mn}$ is to be evaluated using the field equations. In particular if $R_{ij} = (T_{ij} - \frac{1}{2}g_{ij}T) + \tau_{ij}$ where T_{ij} is the nonelectromagnetic part of the matter tensor then $\pounds \tau_{ij} = \pounds R_{ij} - \pounds (T_{ij} - \frac{1}{2}g_{ij}T)$. Since $\pounds R_{ij}$ can be found from $\pounds g_{ij} = 2\sigma g_{ij} + \Lambda_{ij}$ the only indeterminacy is in the specification of T_{ij}. The rest of the theorem follows from Theorem 4.1.

Theorem 4.2 gives the explicit form of the mappings on F_{ij} and ${}^*F_{ij}$ when the space–time metric admits the symmetry mapping (4.4). When the mapping (4.4) is consistent with $\pounds \tau_{ij} = 0$ then we have $\pounds F_{ij} = \sigma F_{ij} + \tilde{\sigma}{}^*F_{ij}$ and $\pounds {}^*F_{ij} = \sigma {}^*F_{ij} - \tilde{\sigma}F_{ij}$. These mappings specialize to the mappings found by Wainwright and Yaremovicz[9] when $\Lambda_{ij} = 0$ and $\sigma = \text{const}$, i.e., in the case of homothetic motions.[23]

If the symmetry mapping (4.4) is to lead to an FCRC symmetry mapping then it must satisfy $g^{ij}\pounds R_{ij} = 0$. By inserting the mapping (4.4) into $g^{ij}\pounds R_{ij}$ [or by using (2.2)] we may express the necessary and sufficient conditions for (4.4) to be a member of the FCRC in terms of σ and Λ_{ij}.

Theorem 4.3: *The symmetry mapping* (4.4) [*which implies* (4.5) *and* (4.6) *when* Λ_{ij} *satisfies* $\Lambda^i_jF^{jk} = \Lambda^k_jF^{ji}$] *is an FCRC symmetry mapping if and only if* $\nabla_i\nabla_j\Lambda^{ij} - 6\nabla_i\nabla^i\sigma = 0$.

When the conditions of this theorem are satisfied for a space–time that includes an electromagnetic field, then the conservation expression (2.2) follows. Furthermore there are certain space–times for which the symmetry mapping (4.4) effects a splitting of the conservation expression (2.2) into two parts.

Consider those space–times which may be characterized as uncharged fluids including nonnull electromagnetic fields (e.g., the uncharged LRS models considered by Stewart and Ellis), and which admit symmetry mappings (4.4), (4.5), and (4.6). For those space–times it follows from $g^{ij}\pounds \tau_{ij} = 0$ and $\nabla_i\tau^{ij} = 0$ that $\nabla_i(\tau^i_j\eta^j) = 0$. It should be noted that this conservation expression holds if (4.4) is admitted by the given (uncharged) space–time, independent of whether or not $g^{ij}\pounds R_{ij} = 0$. If the mapping (4.4) is also an FCRC symmetry mapping then it follows that the FCRC conservation expression (2.2) splits into the two expressions $\nabla_i(\tau^i_j\eta^j) = 0$ and $\nabla_i[(T^i_j - \frac{1}{2}\delta^i_jT)\eta^j] = 0$, where T_{ij} is the uncharged fluid part of the total matter tensor.

So far in this section we have not made any assumptions about the nature of the symmetry vector beyond it being a solution of the defining symmetry relation (4.4). We now briefly consider two choices for the symmetry vector that are suggested by the electromagnetic fields.

When the matter fluid is charged ($\nabla_iF^{ki} = J^k \neq 0$) one may consider symmetry vectors in the direction of the electromagnetic current vector, i.e., $\eta^i = \varphi J^i$. In this case we again find $g^{ij}\pounds \tau_{ij} = 2\nabla_i(\tau^i_j\varphi J^j) = 0$ if $\eta^i = \varphi J^i$ is

a solution of the symmetry mapping (4.4) in the given space–time. Thus, if (4.4) is also an FCRC symmetry mapping then we again obtain a splitting of the conservation expression (2.2).

In the case of null fields one can consider a mapping vector that is in the direction of the null eigenvector (k^i) of the null electromagnetic field tensor, i.e., $\eta^i = \varphi k^i$. Thus, if $F_{ij}k^j = 0$ and the source free Maxwell equations hold, it is not difficult to show that[24-25] $\mathcal{L}F_{ij} = 0$ and $\mathcal{L}^*F_{ij} = 0$. We conclude on the basis of Theorem 4.1 that any null electromagnetic field space–time admits a symmetry mapping of the form $\mathcal{L}g_{ij} = 2\sigma g_{ij} + \Lambda_{ij}$ with symmetry vector $\eta^i = \varphi k^i$ where Λ_{ij} is a trace free-symmetric tensor that satisfies $\Lambda_{ir}F^r_{\cdot j} = \Lambda_{jr}F^r_{\cdot i}$.

APPENDIX

Stewart and Ellis[6] have classified the solutions of the Einstein gravitational equations (with cosmological constant) for a space–time with local rotational symmetry (LRS), near a point P, which contains a perfect fluid including a "noninteracting" electromagnetic field.[5] The matter tensor takes the form $T_{ij} = \mu u_i u_j - p\gamma_{ij} + \tau_{ij}$ where u_i is the average 4-velocity of the matter and τ_{ij} is the stress-energy–momentum tensor of the electromagnetic field. In the rest frame of u^i, μ is the energy density of the fluid and p is the pressure.

Theorem (Stewart and Ellis): If a space–time containing a perfect fluid and an electromagnetic field has LRS near a point P, then the coordinate freedom can be used to set the metric in the form

$$\begin{aligned} ds^2 = {} & [(dx^0)^2/F^2(x^0,x^1)] - X^2(x^0,x^1)(dx^1)^2 \\ & - Y^2(x^0,x^1)[(dx^2)^2 + t^2(x^2)(dx^3)^2] \\ & - [y(x^2)/F^2(x^0,x^1)][2dx^0 - y(x^2)dx^3]dx^3 \\ & + X^2(x^0,x^1)h(x^2)[2dx^1 - h(x^2)dx^3]dx^3, \end{aligned}$$

where $d^2t(x^2)/(dx^2)^2 + Kt(x^2) = 0$, $dy/dx^2 = -2ct(x^2)$, $dh/dx^2 = -2Ct(x^2)$, where K, c, C are arbitrary constants. The field equations place no extra demands on the form of the metric except in Case II where the following condition must be satisfied:

$$FX\partial_0\partial_1 Y + X(\partial_1 F)(\partial_0 Y) - F(\partial_0 X)(\partial_1 Y) = 0.$$

The solution has three major cases and Stewart and Ellis have subclassified each of these special cases as listed below:

Case I: $Y = Y(x^1)$, $F = F(x^1)$, $X = 1$, $h \equiv 0$; $\sigma_{ij} = 0$, $\theta = 0$

subcase Ia: $Y = \text{const}$, $F = \text{const}$,

subcase Ib: $Y = \text{const}$, $\partial_1 F \neq 0$,

subcase Ic: $F = \text{const}$, $\partial_1 Y \neq 0$,

subcase Id: $\partial_1 Y \neq 0$, $\partial_1 F \neq 0$.

Case II: $h = y \equiv 0$; $\omega^i = 0$

subcase IIa: $Y = Y(x^0)$, $F = F(x^0)$,

subcase IIb: $F = F(x^0)$, $\partial_1 Y(x^0,x^1) \neq 0$,

subcase IIc: $\partial_1 Y(x^0,x^1) \neq 0$, $\partial_1 F(x^0,x^1) \neq 0$.

Case III: $F = 1$, $X = X(x^0)$, $Y = Y(x^0)$, $y \equiv 0$; $\omega^i = 0$, $a^i = 0$

subcase IIIa: $Y = \text{const}$,

subcase IIIb: $\partial_0 Y(x^0) \neq 0$.

For this class of space–time models the Ricci tensor can be expressed in the form $R_{ij} = \sum_{\alpha=0}^{3} e_{(\alpha)}\rho_{(\alpha)}\underset{\alpha}{h}{}^i\underset{\alpha}{h}{}^j$ where $e_{(0)} = +1$, $e_{(\mu)} = -1$, $\mu = 1,2,3$, $\rho_{(\alpha)}$ is the αth eigenvalue of the Ricci tensor ($\rho_2 = \rho_3$) and $\underset{\alpha}{h}{}^i$ is the αth eigenvector of the Ricci tensor. We now list the nonzero components of the Ricci tensor, its eigenvalues and eigenvectors:

	Case I	Case II	Case III
R_{00}	$[A_0/F^2]$	$[B_0/F^2]$	C_0
R_{03}	$[-yA_0/F^2]$	0	0
R_{11}	$-A_1$	$-X^2B_1$	$-X^2C_1$
R_{13}	0	0	hX^2C_1
R_{22}	$-Y^2A_2$	$-Y^2B_2$	$-Y^2C_2$
R_{33}	$(y^2A_0/F^2) - T^2Y^2A_2$	$-Y^2t^2B_2$	$-h^2X^2C_1 - T^2Y^2C$

$A_0(B_0,C_0)$, $A_1(B_1,C_1)$, $A_2(B_2,C_2)$ are the eigenvalues of the Ricci tensor in Case I (Case II, Case III):

$$A_0 = \frac{-\partial_1\partial_1 F}{F} + \frac{2(\partial_1 F)(\partial_1 F)}{F^2} - \frac{2(\partial_1 F)(\partial_1 Y)}{FY} + \frac{c^2}{F^2Y^4},$$

$$A_1 = \frac{-\partial_1\partial_1 F}{F} + \frac{2(\partial_1 F)(\partial_1 F)}{F^2} + \frac{2\partial_1\partial_1 Y}{Y},$$

$$A_2 = \frac{\partial_1\partial_1 Y}{Y} - \frac{(\partial_1 F)(\partial_1 Y)}{FY} + \frac{(\partial_1 Y)(\partial_1 Y)}{Y^2} - \frac{K}{Y^2} - \frac{2c^2}{F^2Y^4},$$

$$\begin{aligned} B_0 = {} & \frac{-\partial_1\partial_1 F}{FX^2} + \frac{2(\partial_1 F)(\partial_1 F)}{F^2X^2} + \frac{(\partial_1 F)(\partial_1 X)}{FX^3} - \frac{F(\partial_0 F)(\partial_0 X)}{X} \\ & - \frac{F^2\partial_0\partial_0 X}{X} - \frac{2F^2\partial_0\partial_0 Y}{Y} - \frac{2F(\partial_0 F)(\partial_0 Y)}{Y} - \frac{2(\partial_1 F)(\partial_1 Y)}{FX^2Y}, \end{aligned}$$

$$\begin{aligned} B_1 = {} & \frac{-\partial_1\partial_1 F}{FX^2} + \frac{2(\partial_1 F)(\partial_1 F)}{F^2X^2} + \frac{(\partial_1 F)(\partial_1 X)}{FX^3} - \frac{F(\partial_0 X)(\partial_0 F)}{X} \\ & - \frac{F^2\partial_0\partial_0 X}{X} - \frac{2F^2(\partial_0 X)(\partial_0 Y)}{XY} + \frac{2\partial_1\partial_1 Y}{X^2Y} + \frac{2(\partial_1 Y)(\partial_1 X)}{X^3Y}, \end{aligned}$$

$$\begin{aligned} B_2 = {} & \frac{-F^2\partial_0\partial_0 Y}{Y} - \frac{F(\partial_0 F)(\partial_0 Y)}{Y} - \frac{(\partial_1 F)(\partial_1 Y)}{FX^2Y} - \frac{F^2(\partial_0 X)(\partial_0 Y)}{XY} \\ & + \frac{\partial_1\partial_1 Y}{X^2Y} - \frac{(\partial_1 Y)(\partial_1 X)}{X^3Y} - \frac{F^2(\partial_0 Y)(\partial_0 Y)}{Y^2} + \frac{(\partial_1 Y)(\partial_1 Y)}{X^2Y^2} - \frac{K}{Y^2}, \end{aligned}$$

$$C_0 = \frac{-\partial_0\partial_0 X}{X} - \frac{2\partial_0\partial_0 Y}{Y},$$

$$C_1 = \frac{-\partial_0\partial_0 X}{X} - \frac{2(\partial_0 X)(\partial_0 Y)}{XY} - \frac{2C^4X^2}{Y^4},$$

$$C_2 = \frac{-\partial_0\partial_0 Y}{Y} - \frac{(\partial_0 X)(\partial_0 Y)}{XY} - \frac{(\partial_0 Y)^2}{Y^2} - \frac{K}{Y^2} + \frac{2C^2X^2}{Y^4}.$$

Corresponding to these eigenvalues we find a set of normalized eigenvectors of the Ricci tensor of the form

$$\begin{aligned} u^i &= \underset{0}{h}{}^i = (F,0,0,0), \\ \underset{1}{h}{}^i &= (0, X^{-1}, 0, 0), \\ \underset{2}{h}{}^i &= (0, 0, Y^{-1}, 0), \\ \underset{3}{h}{}^i &= (y/Yt, h/Yt, 0, 1/Yt). \end{aligned}$$

Most of the cases that will concern us will be LRS perfect fluids in the absence of electromagnetic fields.

For these cases we place additional restrictions on the solutions: in Case I, $A_1 = A_2$; in Case II, $B_1 = B_2$; in Case III, $C_1 = C_2$ and $R_{ij} = \rho_0 u_i u_j + \rho\gamma_{ij}$ where $\rho = \rho_1 = \rho_2 = \rho_3$.

Using the orthonormal tetrad formed from the eigenvectors of the Ricci tensor: $\omega^{(\alpha)} = \overset{\alpha}{h}_i \omega^i = \omega\delta_1^{(\alpha)}$, $a^{(\alpha)} = \overset{\alpha}{h}_i a^i = a\delta_1^{(\alpha)}$, $\tau_{(\alpha)(\beta)} = \mathrm{diag}(\tau, -\tau, \tau, \tau)$, $\theta = (\alpha + 2\beta)$, $\sigma_{(\alpha)(\beta)} = \mathrm{diag}[0, -\frac{2}{3}(\alpha - \beta), \frac{1}{3}(\alpha - \beta), \frac{1}{3}(\alpha - \beta)]$, where $\omega = \omega_0/Y^2F$ (ω_0—nonzero constant in Case I—zero in Case II and Case III), $a = -(1/X)\partial_1 \log F$, $\alpha = F\partial_0(\log X)$, $\beta = F\partial_0(\log Y)$, $\tau = \frac{1}{2}(E^2 + B^2)$, and $(0, E, 0, 0)$ and $(0, B, 0, 0)$ are the electric and magnetic fields in the rest frame of u^i, respectively.

[1]W. R. Davis, L. H. Green, and L. K. Norris, Nuovo Cimento B 34, 256 (1976).

[2]L. H. Green, L. K. Norris, D. R. Oliver, Jr., and W. R. Davis, "The Robertson—Walker Metric and the Symmetries Belonging to the Family of Contracted Ricci Collineations," to appear in Gen. Rel. Grav.

[3]D. R. Oliver, Jr. and W. R. Davis, J. Math. Phys. **17**, 1790 (1976).

[4]For a more complete symmetry property inclusion diagram and references relating to the various symmetry properties it embraces, etc., see Ref. 1. In accord with the notations and definitions used by J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954), here and throughout this paper we use (i) ∂_k for $\partial/\partial x^k$, (ii) ∇_k for the operation of covariant differentiation, (iii) $\pounds$ for the operation of Lie differentiation with respect to the vector η^i, and (iv) round and square brackets on indices for the operations of symmetrization and antisymmetrization respectively. The following definitions of quantities associated with a timelike congruence defined by the 4-velocity u^i ($u^i u_i = 1$ with signature of metric -2) will be needed: (i) acceleration $a^i = u^j \nabla_j u^i$; (ii) expansion, $\theta = \nabla_j u^j$; (iii) projection tensor, $\gamma_{ij} = g_{ij} - u_i u_j$; (iv) shear tensor, $\sigma_{ij} = \nabla_{(j} u_{i)} - a_{(i} u_{j)} - (\frac{1}{3})\theta\gamma_{ij}$; (v) rotation tensor, $\omega_{ij} = \nabla_{[j} u_{i]} - a_{[i} u_{j]}$; (vi) rotation vector, $\omega^i = \frac{1}{2}\eta^{ijkm} u_j \nabla_m u_k$, where η^{ijkm} is the permutation tensor with $\eta^{0123} = -(-g)^{-1/2}$; and (vii) rotation scalar, $2\omega^2 = \omega^{ij}\omega_{ij} = -2\omega_i\omega^i$. Also, the following identity will prove to be useful: $\nabla_i \omega^i + 2a_i\omega^i = 0$.

[5]Stewart and Ellis[6] give the following definition: "Space—time is said to be locally rotationally symmetric (LRS) in the neighborhood $N(P_0)$ of a point P_0 if at each point P in $N(P_0)$ there exists a nondiscrete subgroup g of the Lorentz group in the tangent space T_p which leaves invariant u^a, the curvature tensor, and their derivatives up to third order. ... Thus g operates in the subspace of T_p orthogonal to u^a and so g is a one- or three-dimensional group of rotations in T_p." In addition, Stewart and Ellis define a "noninteracting" electromagnetic field: "A noninteracting electromagnetic field can be included by adding to T_{ij} [matter tensor] the extra term $\tau_{ij} = \frac{1}{4}g_{ij}(F_{km}F^{km}) - F_{ik}F^k_j$, $\tau^i_i = 0$, where F_{ij} is the electromagnetic field tensor which satisfies Maxwell's equations. For a charged fluid, $\nabla_j F^{ij} = \epsilon u^i$, $\nabla_{[i}F_{jk]} = 0$. ϵ is the charge density (possibly zero)." Here we note that the conductivity is taken to be zero. A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967), calls this type of system "a charged [or uncharged] perfect fluid with a null conductivity in an electromagnetic field ... without induction."

[6]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968). In this context see also G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).

[7]J. Wainwright, Commun. Math. Phys. **17**, 42 (1970).

[8]A. Kransiński, Acta Phys. Pol. B **5**, 411 (1974); **6**, 223, 239 (1975); J. Math. Phys. **16**, 125 (1975).

[9]J. Wainwright and P. E. A. Yaremovicz, Gen. Rel. Grav. **7**, 345 (1976); **7**, 595 (1976).

[10]H. Michalski and J. Wainwright, Gen. Rel. Grav. **6**, 289 (1975); J. R. Ray and E. L. Thompson, J. Math. Phys. **16**, 345 (1975).

[11]See Refs. 1—3 in relation to the physical interpretation of this conservation expression. In particular, in Refs. 1 and 2 it is shown that for timelike FCRC (with $\eta^i = \varphi u^i$) this expression can be interpreted in terms of particle number conservation.

[12]While the decomposition (2.4) suggests the three particularly simple choices for H_{ij} which are listed it certainly does not exclude other choices of H_{ij}.

[13]See Stewart and Ellis,[6] p. 1078.

[14]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), p. 413.

[15]A somewhat more restrictive form of this theorem was discussed by Oliver and Davis in Ref. 3.

[16]P. J. Greenberg, J. Math. Anal. Appl. **30**, 128 (1970).

[17]We observe that whenever one has an expression of the form $\nabla_i(\lambda^2\omega^i) = 0$, then a generalized Kelvin—Helmholtz type theorem is obtained with $\lambda\omega A$ constant along the vortex flow.

[18]See Stewart and Ellis,[6] p. 1075.

[19]K. P. Singh and D. N. Sharma, J. Phys. A: Math. Gen. **8**, 1875 (1975); C. D. Collinson, Gen. Rel. Grav. **1**, 137 (1970).

[20]L. Witten, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), pp. 382—411.

[21]It is of interest to note that C. D. Collinson, J. Math. Phys. **11**, 818 (1970), in studying curvature collineations (CC) admitted by Petrov type N space—times has obtained a similar mapping on the metric. He has found a mapping of the form $\pounds g_{ij} = \frac{1}{2}\varphi g_{ij} + 2l_i l_j$ where l_i is the principal vector of the Weyl tensor satisfying $C_{jkm}{}^i l_i = 0$.

[22]For nonnull fields it is not difficult to show that the conditions placed on Λ^{ij} in Theorem 4.1 imply that the general form of Λ^{ij} is $\Lambda^{ij} = ak^ik^j + bn^in^j + cp^{(i}q^{j)} + d(p^ip^j - q^iq^j)$, where a, b, c, and d are arbitrary functions, k^i and n^i are the null eigendirections of F_{ij} ($k_i n^i = 1$), and p^i and q^i are two spacelike eigenvectors[20] of τ_{ij} such that $k^ip_i = k^iq_i = 0$, $n^iq_i = n^ip_i = 0$.

[23]Most of the results found by Wainwright and Yaremovicz in Ref. 9, including the mappings (4.1) and (4.2), can be shown to hold when the space—time symmetry is generalized from homothetic motions to special conformal motions (S Conf M) which are defined by $\pounds g_{ij} = 2\sigma g_{ij}$, $\nabla_i\nabla_j\sigma = 0$. This follows from the fact that a S Conf M is a RC (i.e., $\pounds R_{ij} = 0$).

[24]We note that R. Sigal, J. Math. Phys. **14**, 1434 (1973), has $\pounds_\pi F^+_{ij} \neq 0$, where $F^+_{ij} = F_{ij} + i^*F_{ij}$ and $F^+_{ij}\eta^j = 0$ which, for source-free null fields that obey Maxwell's equations, appears to be incorrect.

[25]C. D. Collinson[19] has studied the vanishing Lie derivative of the electromagnetic field tensor, $\pounds F^i_j = 0$, and termed this symmetry condition a Maxwell collineation.

Energy-momentum tensors in the theory of electromagnetic fields admitting electric and magnetic charge distributions*

Hanno Rund

Department of Mathematics, University of Arizona, Tucson, Arizona 85721
(Received 9 December 1976)

When the field tensor of an electromagnetic field admitting both electric and magnetic charge distributions is expressed in terms of a Clebsch representation, the extended Maxwell equations in the presence of a given gravitational field are derivable from an invariant variational principle in which the Clebsch potentials play the role usually assumed by the classical 4-potentials. The corresponding Lagrange density gives rise in a unique manner to a symmetric tensor density T^{hj}, which displays some of the properties normally associated with the energy-momentum tensor density of the electromagnetic field. However, this interpretation may be in conflict with the generally accepted expression for the modified Lorentz force. Accordingly an alternative energy-momentum tensor density θ^{hj} is derived which does not suffer from this drawback. However, when a generalized variational principle for the simultaneous determination of the behavior of both the electromagnetic and the dynamical gravitational fields is introduced, the resulting Euler-Lagrange equations give rise to extended Einstein-Maxwell equations which involve the density T^{hj}. On the other hand, the alternative Einstein-Maxwell equations, obtained by the replacement of T^{hj} by θ^{hj}, are not derivable from a variational principle. The solutions of the two Einstein-Maxwell equations, for the case of a spherically symmetric metric and static electromagnetic field, predict distinctly different effects of the magnetic charges on the gravitational field.

I. INTRODUCTION

The classical theory of electromagnetic fields admitting magnetic as well as electric charge distributions, as described by Schwinger[1] and others, has recently been examined[2] from the point of view of the Clebsch representations of arbitrary skew-symmetric type (0,2) tensors in the presence of gravitational fields.[3] In this treatment the Clebsch potentials play the role of the 4-potentials of classical electrodynamics, which allows for the formulation, against the background of a given gravitational field, of an invariant variational principle, from which the field equations may be derived in a rigorous manner. Since the corresponding Lagrangian is required to be a scalar density, it gives rise in a unique manner to a symmetric type (2,0) tensor density, which may be interpreted as the energy—momentum density of the field. However, as will be indicated below, this particular density suffers from several drawbacks, and accordingly in Sec. II an alternative energy—momentum tensor density which does not display such undesirable features is derived. As a direct consequence of its definition, the original energy—momentum tensor density must inevitably appear in the extended Einstein—Maxwell equations if the latter are to be the Euler—Lagrange equations of an invariant variational principle for the simultaneous determination of both the electromagnetic and the dynamical gravitational fields; however, as will be seen in Sec. III, the alternative Einstein—Maxwell equations based on the new energy—momentum tensor density do not appear to be derivable from a variational principle within this framework. Moreover, it is found that, for the case of a spherically symmetry metric and a static electromagnetic field, the two Einstein—Maxwell equations expressed respectively in terms of these distinct energy—momentum tensors, predict different effects of magnetic charges on the gravitational field.

The underlying manifold[4] is assumed to be a pseudo-Riemannian space V_4 endowed with a metric tensor g_{hj}, it being assumed that $g \equiv |\det(g_{hj})| \neq 0$. The entire theory is based on two assumptions, namely that the electromagnetic field is represented by a skew-symmetric tensor field F_{hj}, and that the behavior of the latter (in the presence of gravitational effects) is governed by the Euler—Lagrange equations resulting from the Lagrange density

$$L = \tfrac{1}{4}\sqrt{g}F^{hj}F_{hj} - \psi_h J^h + \phi_h S^h, \tag{1.1}$$

where J^h, S^h denote the electric and magnetic current densities respectively, while ψ_h, ϕ_h are the aforementioned Clebsch potentials which appear in the Clebsch representation of F_{hj}. This representation is determined uniquely by the Euler—Lagrange equations associated with (1.1), namely as

$$F_{hj} = f_{hj} + ib_{hj}, \tag{1.2}$$

where

$$f_{hj} = \psi_{j|h} - \psi_{h|j}, \quad b^{hj} = g^{-1/2}\epsilon^{hjlk}\phi_{k|l}, \tag{1.3}$$

while the remaining Euler—Lagrange equations give rise to the extended Maxwell equations

$$\sqrt{g}F^{jh}{}_{|j} = -J^h, \quad \sqrt{g}F^{*jh}{}_{|j} = S^h, \tag{1.4}$$

where F^{*jh} denotes the dual of F^{jh}, that is,

$$F^{*jh} = \tfrac{1}{2}ig^{-1/2}\epsilon^{jhlk}F_{lk}. \tag{1.5}$$

Since the Lagrangian (1.1) is a scalar density, the expressions

$$T^{hj} = -2c^{-2}\frac{\partial L}{\partial g_{hj}}, \tag{1.6}$$

are the components of a type (2,0) tensor density field,[5] whose explicit form is given by

$$c^2 T^j_h = \sqrt{g}\,[F^{jl}(\psi_{l|h} - \psi_{h|l}) + F^{*jl}(\phi_{l|h} - \phi_{h|l}) - \tfrac{1}{4}(F\cdot F)\delta^j_h]. \tag{1.7}$$

This field has the following properties: (i) T^{hj} is symmetric; (ii) $T^j_j = 0$; (iii) $T^{hj}{}_{|j} = 0$ whenever Maxwell's equations with $J^h = 0$, $S^h = 0$ are satisfied; (iv) T^{hj} reduces to the energy—momentum tensor of classical

electrodynamics in the absence of magnetic charges. Accordingly the tensor density (1.6) was interpreted in Ref. 2 as the *energy–momentum tensor density of the electromagnetic field*. However, this interpretation suffers from the following drawbacks: (v) Although T^{hj} is invariant under all gauge transformations of the Clebsch potentials, its explicit expression does not involve only F_{hj}, but also the derivatives of the Clebsch potentials; (vi) when the rate of work done by the field on a test particle carrying both electric and magnetic charges is evaluated in terms of the modified Lorentz force (the latter having been obtained from a separate single integral variational principle), the resulting expression is not, in general, consistent with the 4-divergence of T_4^j. Because of property (iv), these difficulties do not arise in the absence of magnetic charges.

One is therefore confronted with two alternatives. On the one hand, one might abandon the modified Lorentz force (which, in flat space–time, reduces to the Lorentz force postulated by Schwinger[6] and others); this would necessitate the replacement of the aforementioned single integral variational principle, which is independent of (1.1), by a more appropriate one. On the other hand, one might seek an alternative energy–momentum tensor θ^{hj} which possesses the desirable properties (i)–(iv), but which does not display the drawbacks (v) and (vi). In view of the physical feasibility as well as the mathematical simplicity of of the modified Lorentz force, the second of these alternatives will be explored here: This is the objective of the present note.

II. THE ALTERNATIVE ENERGY-MOMENTUM TENSOR

In a flat space–time the modified Lorentz force due to electric and magnetic field strengths $\mathbf{E}$ and $\mathbf{H}$ is assumed to possess the form

$$\mathbf{F}=[\rho\mathbf{E}+(1/c)\mathbf{j}\times\mathbf{H}]+[\sigma\mathbf{H}-(1/c)\mathbf{s}\times\mathbf{E}], \tag{2.1}$$

where ρ and σ represent the electric and magnetic charge densities respectively, the 4-vector representation of J^h and S^h being taken as

$$J^h=((1/c)\mathbf{j}, i\rho), \quad S^h=((1/c)\mathbf{s}, i\sigma). \tag{2.2}$$

In terms of the usual identification[7]

$$iE_\alpha=F_{4\alpha}, \quad iH_\alpha=F^*_{\alpha 4}, \tag{2.3}$$

the relation (2.1) may be shown to be equivalent to

$$F_\alpha=F_{\alpha l}J^l-F^*_{\alpha l}S^l \quad (\alpha=1,2,3), \tag{2.4}$$

to which we adjoin a fourth component, namely

$$F_4\equiv F_{4l}J^l-F^*_{4l}S^l=ic^{-1}[\mathbf{E}\cdot\mathbf{j}+\mathbf{H}\cdot\mathbf{s}], \tag{2.5}$$

which represents the rate at which work is done by the field. Since the Maxwell equations (1.4) imply that

$$\mathbf{j}=c\nabla\times\mathbf{H}-\frac{\partial\mathbf{E}}{\partial t}, \quad \mathbf{s}=-c\nabla\times\mathbf{E}-\frac{\partial\mathbf{H}}{\partial t}, \tag{2.6}$$

it follows that (2.5) is equivalent to

$$ic(F_{4l}J^l-F^*_{4l}S^l)=\mathbf{E}\circ\frac{\partial\mathbf{E}}{\partial t}+\mathbf{H}\cdot\frac{\partial\mathbf{H}}{\partial t}+c(\mathbf{H}\cdot\nabla\times\mathbf{E}-\mathbf{E}\circ\nabla\times\mathbf{H})$$
$$=\frac{1}{2}\frac{\partial}{\partial t}(E^2+H^2)+c\,\mathrm{div}(\mathbf{E}\times\mathbf{H}). \tag{2.7}$$

In regions devoid of electric and magnetic charges this gives rise to an equation of continuity, that is, a conservation law. Thus, in accordance with our program, we shall now construct a tensor density θ_h^j which is such that, in flat space–time, the right-hand side of (2.7) is proportional to the 4-divergence of θ_4^j.

The relations (2.4) and (2.5) clearly indicate that this purpose will have been attained quite generally if the required tensor density θ_h^j is such that

$$-c^2\theta^j_{h|j}=F_{hl}J^l-F^*_{hl}S^l, \tag{2.8}$$

as a consequence of Maxwell's equations (1.4), and this condition is certainly satisfied if

$$c^2\theta^j_{h|j}=\sqrt{g}\,(F_{hl}F^{jl}{}_{|j}+F^*_{hl}F^{*jl}{}_{|j}). \tag{2.9}$$

Now, from the definition (1.5) it follows that

$$\epsilon_{hlpq}F^{*jl}{}_{|j}=\tfrac{1}{2}ig^{-1/2}\epsilon^{jlmk}\epsilon_{hlpq}F_{mk|j}$$
$$=\tfrac{1}{2}ig^{-1/2}\delta^{jmk}_{hpq}F_{mk|j}$$
$$=ig^{-1/2}(F_{pq|h}+F_{qh|p}+F_{hp|q}). \tag{2.10}$$

This is multiplied by F^{pq}, which gives

$$\epsilon_{hlpq}F^{pq}F^{*jl}{}_{|j}=ig^{-1/2}[\tfrac{1}{2}(F^{pq}F_{pq})_{|h}+2F^{pq}F_{qh|p}], \tag{2.11}$$

where repeated use has been made of the skew-symmetry of F^{pq}. It is now observed that the inverse of (1.5) is given by

$$F^*_{hl}=\tfrac{1}{2}ig^{1/2}\epsilon_{hlpq}F^{pq},$$

and accordingly (2.11) yields

$$F^*_{hl}F^{*jl}{}_{|j}=F^{jl}F_{hl|j}-\tfrac{1}{4}(F^{lm}F_{lm})_{|h}. \tag{2.12}$$

This is merely an identity which depends solely on the skew-symmetric character of F_{hj} and the definition of its dual. However, when (2.12) is substituted in the condition (2.9), the latter becomes

$$c^2\theta^j_{h|j}=\sqrt{g}\,[F_{hl}F^{jl}{}_{|j}+F_{hl|j}F^{jl}-\tfrac{1}{4}(F^{lm}F_{lm})_{|h}]$$
$$=\sqrt{g}\,[F_{hl}F^{jl}-\tfrac{1}{4}\delta^j_h(F^{lm}F_{lm})]_{|j},$$

which may be integrated to yield

$$c^2\theta^j_h=\sqrt{g}\,[F_{hl}F^{jl}-\tfrac{1}{4}\delta^j_h(F^{lm}F_{lm})]. \tag{2.13}$$

This, then, is the required form of the *alternative energy–momentum tensor density*. It is remarkable that its formal structure in terms of F_{hj} is identical with that of the energy–momentum tensor of classical electrodynamics, despite the fact that the representation (1.2) of F_{hj} involves 2- (rather than 1-) vector fields.

It is obvious that the tensor field θ_h^j as defined by (2.13) satisfies the conditions (i)–(iv) of the previous section, while, as a result of our construction, the objections (v) and (vi) do not apply.

III. THE EXTENDED EINSTEIN-MAXWELL EQUATIONS

In the above theory it is tacitly assumed that the metric tensor of V_4 assumes preassigned values. However, if the Lagrangian (1.1) is augmented by an additive term $a\sqrt{g}R$, where a is a suitably chosen constant and R denotes the scalar curvature of V_4, the variational principle[8] based on the resulting Lagrangian L not only

yields the Clebsch representations (1.2), (1.3) of F_{hj}, the Maxwell equations (1.4), but also the field equations *in vacuo*

$$\sqrt{g}R^{hj} = -8\pi\kappa c^{-2}T^{hj}, \tag{3.1}$$

in which R^{hj} is the Ricci tensor and κ is the gravitational constant. The presence of T^{hj} on the right-hand side of (3.1) is a direct consequence of its definition (1.6) and the appearance of the term $-\partial\hat{L}/\partial g_{hj}$ in the Euler–Lagrange equation for g_{hj}.

Again one is confronted with two alternatives. First, it would seem to be more consistent with the analysis above to replace T^{hj} by θ^{hj} in (3.1). But, under these circumstances, the resulting Einstein–Maxwell equations could be derivable from a variational principle only if it is possible to construct a Lagrangian L^* which is such that (1.2), (1.3), (1.4) remain unchanged, the term $-\partial L^*/\partial g_{hj}$ giving rise to θ^{hj} rather than to T^{hj}. The second alternative is simply the acceptance of (3.1) without change.

Let us briefly consider the first possibility, setting

$$\mathcal{L} = L^* - \hat{L}. \tag{3.2}$$

Our requirements would be met if the Euler–Lagrange equations of $\mathcal{L}$ for the Clebsch potentials ψ_h, ϕ_h are satisfied identically, and if

$$-2\frac{\partial\mathcal{L}}{\partial g_{hj}} = c^2(\theta^{hj} - T^{hj}). \tag{3.3}$$

But since (1.7) can be expressed in the form

$$c^2T^j_h = \sqrt{g}[F^{jl}f_{hl} + iF^{*jl}b^*_{lh} - \tfrac{1}{4}\delta^j_h(F\circ F)], \tag{3.4}$$

it follows with the aid of (2.13) that (3.2) is equivalent to the condition that

$$-2\frac{\partial\mathcal{L}}{\partial g_{hj}} = i\sqrt{g}g^{mh}[F^{jl}b_{ml} + F^{*jl}b^*_{ml}]. \tag{3.5}$$

With the aid of (1.2) and (1.3) it is seen by inspection that the right-hand side of (3.5) is a quadratic polynomial in $\psi_{h|j}$, $\phi_{h|j}$, which is also homogeneous of degree -1 in g_{hj}. Hence $\mathcal{L}$ must be a density which is homogeneous of degree 0 in g_{hj} while being quadratic in $\psi_{h|j}$, $\phi_{h|j}$. Now, it is readily verified that the only density satisfying these requirements for which the Euler–Lagrange equations for ψ_h and ϕ_h are satisfied identically possesses the form

$$A\epsilon^{hklj}\psi_{h|k}\psi_{l|j} + B\epsilon^{hklj}\psi_{h|k}\phi_{l|j} + C\epsilon^{hjlk}\phi_{h|j}\phi_{l|k},$$

where A, B, C are arbitrary constants. But since this expression is actually independent of g_{hj}, it cannot possibly satisfy the condition (3.5). It is therefore concluded that there does not exist a density $\mathcal{L}$ of the required kind. Thus it would appear that, within the present framework, the alternative to the field equations (3.1), namely

$$\sqrt{g}R^{hj} = -8\pi\kappa c^{-2}\theta^{hj}, \tag{3.6}$$

cannot be derived from a variational principle.

Nevertheless, Eqs. (3.1) and (3.6) lead to distinctly different physical conclusions, which are best illustrated by the following special case. Let us suppose that our metric is represented by a spherically symmetric line element

$$ds^2 = -(e^\lambda dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2) + c^2e^\nu dt^2, \tag{3.7}$$

with $\lambda = \lambda(r)$, $\nu = \nu(r)$, and that a static electromagnetic field is given, whose Clebsch representation is characterized by

$$\psi_\alpha = 0, \quad \psi_4 = iV(r), \quad \phi_\alpha = 0, \quad \phi_4 = iU(r). \tag{3.8}$$

Under these circumstances the Maxwell equations (1.4) imply[9] that the derivatives of V and U are given by

$$V' = \epsilon r^{-2}\exp[(\lambda+\nu)/2], \quad U' = \gamma r^{-2}\exp[(\lambda+\nu)/2], \tag{3.9}$$

where ϵ and γ are constants of integration, which are interpreted as the electric and magnetic charge respectively of the gravitating mass which gives rise to the metric (3.7). These deductions are independent of the choice of the energy–momentum tensor. Also, for the field (3.8) the electromagnetic field tensor assumes the form

$$F^j_h = \begin{pmatrix} 0 & 0 & 0 & -ie^{-\lambda}V' \\ 0 & 0 & g^{-1/2}r^2\sin^2\theta U' & 0 \\ 0 & -g^{-1/2}r^2U' & 0 & 0 \\ ie^{-\nu}V' & 0 & 0 & 0 \end{pmatrix}, \tag{3.10}$$

and it may be verified with the aid of (3.9) that, under these circumstances,

$$T^j_h = \tfrac{1}{2}c^{-2}(\epsilon^2 - \gamma^2)g^{1/2}r^{-4}\,\mathrm{diag}[-1,1,1,-1], \tag{3.11}$$

while

$$\theta^j_h = \tfrac{1}{2}c^{-2}(\epsilon^2 + \gamma^2)g^{1/2}r^{-4}\,\mathrm{diag}[-1,1,1,-1], \tag{3.12}$$

where $\mathrm{diag}[a,b,c,d]$ denotes the 4×4 diagonal matrix with (ordered) diagonal entries a, b, c, d. When (3.11) is substituted in the field equation (3.1), the resulting solution is given by

$$e^\nu = e^{-\lambda} = 1 - 2m/r + (4\pi\kappa/c^4r^2)(\epsilon^2 - \gamma^2), \tag{3.13}$$

where m is a constant of integration which is identified as usual with the mass of the gravitating body. However, as a result of (3.12), it is found that the corresponding solution of the alternative field equation (3.6) is given by

$$e^\nu = e^{-\lambda} = 1 - 2m/r + (4\pi\kappa/c^4r^2)(\epsilon^2 + \gamma^2), \tag{3.14}$$

thus predicting a different effect of the magnetic charge on the gravitational field. It is remarkable that the generalized Reissner–Nordström metric generated by (3.14) is identical with a line element postulated very recently by Adler,[10] who showed that such metrics exhibit some physically significant features, particularly with regard to the quantization of charge and angular momentum. Clearly the ultimate choice between the field equations (3.1) and (3.6) must await further developments.

ACKNOWLEDGMENTS

It is a pleasure to thank R. Baumeister, D. Lovelock, and M. O. Scully of the University of Arizona for valuable discussions.

*This research was sponsored in part by NSF GP43070.
[1]J. Schwinger, (a) Phys. Rev. D 12, 3105 (1975); (b) Science 165, 757 (1969).
[2]H. Rund, J. Math. Phys. 18, 84 (1977).
[3]H. Rund, *Topics in Differential Geometry*, edited by H. Rund and W. Forbes (Academic, New York, 1976), p. 111.
[4]Latin indices j, h, k, $\cdots$ range from 1 to 4, the summation convention being operative throughout. In the local reference frame of an observer $x^4 = ict$. A vertical bar followed by a subscript denotes covariant differentiation.
[5]D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles* (Wiley-Interscience, New York, 1975), pp. 302–04. The general equivalence of this derivation of the energy–momentum tensor with the more commonly used prescriptions for the determination of the Belinfante tensor was established by E.A. Lord [J. Math. Phys. 17, 37 (1976)]. This may be done even in flat space–time by means of the introduction of curvilinear coordinate systems.
[6]Ref. 1(b).
[7]Greek indices α, β, $\cdots$ range from 1 to 3; they represent spatial components.
[8]Reference 2, Eq. (4.1).
[9]Reference 2, Eqs. (4.16), (4.17).
[10]R.J. Adler, Phys. Rev. D 14, 392 (1976).

On the macroscopic equivalence of descriptions of fluctuations for chemical reactions

Joel Keizer

Chemistry Department, University of California, Davis, California 95616
(Received 3 February 1977)

Three descriptions of spontaneous fluctuations in macroscopic systems have been proposed: One uses generalized Fokker–Planck equations and treats fluctuations as a stochastic diffusion process; another uses a connection between fluctuations and dissipation and generalizes the Langevin method; the third is the master equation theory which treats fluctuations as arising from a birth and death process. For a variety of systems it is known that the master equation theory is identical to the fluctuation–dissipation theory in the macroscopic limit. For chemical reactions it is shown here that the appropriate diffusion process also becomes identical with the fluctuation–dissipation theory in the macroscopic limit.

I. INTRODUCTION

The random fluctuation of a particle undergoing Brownian motion can be described in a variety of ways. The earliest description was the diffusion equation approach of Einstein[1] and Smoluchowski,[2] which lead to the more complete Fokker—Planck description.[3] A second way of describing Brownian motion is the fluctuating force approach of Langevin[4] and Ornstein—Uhlenbeck.[5] Finally the use of the urn models[6] provides a description of Brownian motion as a birth and death process. Although each of these descriptions of spontaneous fluctuations has a different mathematical structure, all three methods are equivalent when applied near equilibrium.[7]

Recently these approaches for describing Brownian motion have been generalized so as to apply to fluctuations for a wide variety of macroscopic variables. Generalized Fokker—Planck equations, which describe a stochastic diffusion process, have been introduced to describe fluctuations in optical systems.[8] A generalization of the Langevin approach based on a connection between fluctuations and dissipation has been used to describe fluctuations associated with macroscopic transport equations.[9,10] Finally the master equation theory, which generalizes the urn model, has been used to describe fluctuations in a number of physical processes.[11,12] All three theories purport to describe spontaneous fluctuations and are based on the form of the transport equations. Since the transport equations are valid only in the limit of large system, these theories are relevant only for describing fluctuations in the macroscopic limit. As a consequence, it is only this limit of these theories that is physically important. Because none of the alternative theories has a clear foundation in statistical mechanics, it is not possible to say *a priori* that one approach is more fundamental than the others. On the other hand, it is the purpose of this note to argue that all three theories are equivalent in the macroscopic limit. In this sense, the three descriptions of fluctuations are equally valid in the macroscopic domain, and the real test of their validity lies in comparison with experiment.

It is already known for a number of examples[13–15] that the macroscopic limit of the master equation theory coincides with the random force theory based on fluctuation—dissipation postulates. This correspondence shows for a uniform system of coupled chemical reactions that, in the limit of a large system, the conditional probability density obtained from the birth and death theory approaches the Gaussian form given by the fluctuation—dissipation postulates.[13,14] The results below show that a similar result holds for the conditional probability for stochastic diffusion processes modeled after the mass action law for chemical reactions. Although only chemical reactions are explicitly considered here, the fact that macroscopic transport laws have a canonical form similar to the mass action law,[10] suggests that the correspondence is more general.

The outline of this work is as follows. First, Fokker—Planck equations for a stochastic diffusion process are introduced for chemical reactions and this approach is compared to the fluctuating force theory given by fluctuation—dissipation postulates. To motivate the equivalence of these theories, the stationary equilibrium distribution for the chemical reaction $X + Y \rightleftharpoons 2X$ is obtained analytically for the two theories. These distributions are different, but in the macroscopic limit the diffusion theory approaches the Gaussian form obtained from the fluctuation—dissipation postulates. It is then shown how to systematically expand the Fokker—Planck equations for the conditional probability of the diffusion process in terms of the inverse volume. The terms which survive the infinite volume limit yield the equation solved by the Gaussian form obtained from the fluctuation—dissipation postulates. Finally the mean convergence of the two processes is proven. Estimates of the validity of the approximate Gaussian solution suggests that it is valid at typical densities and for volumes containing more than a few thousand molecules.

II. STOCHASTIC DIFFUSION PROCESSES FOR REACTIONS

In this section a stochastic diffusion process is associated with a system undergoing chemical reactions. The system considered here is uniform at fixed temperature and fixed volume and consists of k different kinds of molecules—symbolically $C_1, C_2, \ldots, C_k$—undergoing m distinct elementary chemical reactions. In addition to being changed by the chemical reactions, it is supposed that the number concentration, n_i, of the

molecule C_i changes due to an external source, whose strength K_i is independent of all the n_i. Thus the rate of change of the number densities is given by the expression

$$\frac{dn_i}{dt} = \sum_{\kappa=1}^{m} \nu_{\kappa i}(V_\kappa^+ - V_\kappa^-) + K_i \equiv V_i(\mathbf{n}, t), \tag{1}$$

where $\nu_{\kappa i}$ is the stoichiometric coefficient for C_i in the κth reaction. The forward and reverse rates of the kth reaction are written as $V_\kappa^\pm$ and are given by the mass action law

$$V_\kappa^\pm = K_\kappa^\pm \prod_{j=1}^{k} n_j^{l_{\kappa j}^\pm}, \tag{1'}$$

where $K_\kappa^\pm$ is the rate constant and $l_{\kappa j}^\pm$ is the number of C_j molecules involved in the forward or reverse step of the κth reaction. Since the variables n_j represent concentrations, the rates V_i are *independent of the volume at fixed concentration*.

To associate a stochastic diffusion process with the reactions in Eq. (1) requires the specification of the Kolomogorov forward (Fokker–Planck) equation for this process.[16] This equation is satisfied by all single time probability densities, $W(\mathbf{n}, t)$, where

$$W(\mathbf{n}, t)d\mathbf{n} = \text{probability of number densities in the interval } d\mathbf{n} \text{ around } \mathbf{n} \text{ at time } t.$$

The Fokker–Planck equation for the diffusion process is taken to be

$$\frac{\partial W(\mathbf{n}, t)}{\partial t} + \frac{\partial V_i(\mathbf{n}, t)W(\mathbf{n}, t)}{\partial n_i} - \frac{1}{2}\frac{\partial^2 \gamma_{ij}(\mathbf{n})W(\mathbf{n}, t)}{\partial n_i \partial n_j} = 0. \tag{2}$$

In Eq. (2) the Einstein summation convention for repeated indices is used and

$$\gamma_{ij}(\mathbf{n}) = \Omega^{-1} \sum_\kappa \nu_{\kappa i}(V_\kappa^+ + V_\kappa^-)\nu_{\kappa j}, \tag{3}$$

where Ω is the volume of the system. The second term in Eq. (2) is the drift term and it is natural that the macroscopic rate function $V_i(\mathbf{n}, t)$ appear there [cf. Eqs. (20) and (21) below]. The form of the diffusion matrix $\gamma_{ij}(\mathbf{n})$ in Eq. (3) is patterned after the terms which cause fluctuations in the fluctuation–dissipation theory.[14] *Equations* (1)–(3) *completely define the Markov stochastic process and determine* $W(\mathbf{n}, t)$ *once* $W(\mathbf{n}, 0)$ *is given*. When the initial condition is perfectly sharp, i.e., $W(\mathbf{n}, 0) = \delta(\mathbf{n} - \mathbf{n}^0)$, the solution to Eq. (2) is the conditional probability density, $P_2(\mathbf{n}^0 | \mathbf{n}t)$.

An entirely different description of fluctuations a for chemical reaction is given by the fluctuation–dissipation theory.[9,10] In that theory the conditional probability can be shown to be the Gaussian[9]

$$P_2^{fd}(\mathbf{n}^0 | \mathbf{n}t) = ((2\pi)^k \det\sigma(\mathbf{n}^0, t))^{-1/2} \times \exp[-\tfrac{1}{2}(\mathbf{n} - \bar{\mathbf{n}}(\mathbf{n}^0, t))^T \sigma^{-1}(\mathbf{n}^0, t)(\mathbf{n} - \bar{\mathbf{n}}(\mathbf{n}^0, t))], \tag{4}$$

where the superscript *fd* denotes the fluctuation–dissipation theory result. In Eq. (4) the conditional average is $\bar{\mathbf{n}}(\mathbf{n}^0, t)$ and is found by solving the deterministic Eq. (1) subject to the condition $\bar{\mathbf{n}}(\mathbf{n}^0, 0) = \mathbf{n}^0$. The conditional covariance matrix is[9,10]

$$\sigma(\mathbf{n}^0, t) = \int_0^t P\exp[-\int_s^t H(\mathbf{n}^0, \tau)d\tau]\gamma(\mathbf{n}^0, s) \times (P\exp[-\int_s^t H(\mathbf{n}^0, \tau)d\tau])^T ds \tag{5}$$

with $P\exp$ the time ordered exponential,[17] the matrix H is

$$H_{ij}(\mathbf{n}^0, t) = \frac{\partial V_i(\bar{\mathbf{n}}(\mathbf{n}^0, t), t)}{\partial \bar{n}_j} \tag{6}$$

and $\gamma(\mathbf{n}^0, s) \equiv \gamma(\bar{\mathbf{n}}(\mathbf{n}^0, s))$.

The conditional probability in Eq. (4) results from the fluctuation–dissipation postulates which assert[10] (i) that the conditional average satisfies Eq. (1) and (ii) that deviations from the conditional average, $\delta\mathbf{n} = \mathbf{n} - \bar{\mathbf{n}}(\mathbf{n}^0, t)$, solve the stochastic differential equation

$$\frac{d\delta\mathbf{n}}{dt} = H(\mathbf{n}^0, t)\delta\mathbf{n} + \tilde{\mathbf{f}}(t), \tag{7}$$

where $\langle \tilde{\mathbf{f}}(t) \rangle = 0$ and

$$\langle \tilde{f}_i(t)\tilde{f}_j(s) \rangle = \gamma_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t))\,\delta(t - s), \tag{8}$$

with γ obtained from Eq. (3), and $\tilde{\mathbf{f}}(t)$ is Gaussian.

Actually the diffusion process satisfying the Fokker–Planck equation (2) can also be associated with a stochastic differential equation, although this equation is different from Eq. (7). Indeed every stochastic diffusion is generated by an Itô stochastic differential equation.[17] For Eq. (2) the associated Itô equations are

$$dn_i = V_i(\mathbf{n}, t)dt + g_{ij}(\mathbf{n})dw_j, \tag{9}$$

where the w_j are independent Wiener processes and the matrix g is the square root of γ, i.e.,

$$g_{ij}(\mathbf{n})g_{jk}(\mathbf{n}) = \gamma_{ik}(\mathbf{n}). \tag{10}$$

As is known,[8,18] this simple correspondence between a stochastic diffusion and a stochastic differential equation does not hold true if one chooses Stratonovich's definition of the stochastic integral; however, this correspondence is generally true when Itô's definition is used. Since, loosely speaking, dw_j/dt is white noise, the form of Eqs. (9) and (10) suggests a similarity between the diffusion process and the fluctuation–dissipation process. They are *not*, however, identical since Eqs. (7) and (8) imply the Gaussian conditional density in Eq. (4) whereas Eqs. (9) and (10) give a Gaussian form only in the limit of a large system. Furthermore a diffusion process is a Markov process,[16] whereas the Gaussian conditional density in Eq. (4) is known not to satisfy the Smoluchowski equation as an identity[14] and so it is not a Markov process. These distinctions will be clarified in the following sections.

Another similarity between the diffusion process and the fluctuation–dissipation process can be seen by calculating the time derivative of P_2^{fd} using Eqs. (1), (4), and (5). It is verified, after a short calculation, that $p^{fd}(\delta\mathbf{n}, t) \equiv P_2^{fd}(\mathbf{n}^0 | \mathbf{n}t)$ satisfies

$$\frac{\partial P^{fd}}{\partial t} + \partial H_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t))\frac{\delta n_j P^{fd}}{\partial \delta n_i} - \frac{1}{2}\frac{\partial^2 \gamma_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t))P^{fd}}{\partial \delta n_i \partial \delta n_j} = 0 \tag{11}$$

with $P^{fd}(\delta\mathbf{n}, 0) = \delta(\delta\mathbf{n})$ and $\delta n_j = n_j - \bar{n}_j(\mathbf{n}^0, t)$. Equation (11) is closely related to the Fokker–Planck equation (2) for the diffusion process. An important difference is that the drift and diffusion terms in Eq. (11) depend on time and the initial condition. As a consequence an unconditional density, $W^{fd}(\mathbf{n}, t)$, for a fluctuation–dissipa-

tion process does not identically satisfy an equation like Eq. (2).

What is shown in Sec. IV is that the conditional density for the diffusion process satisfies Eq. (11) asymptotically as the volume of the system gets large. Furthermore it is shown that the conditional density for the diffusion actually converges (L_1) to the Gaussian density of the fluctuation–dissipation process. In this sense the two descriptions of fluctuations become identical in the macroscopic limit.

III. AN EXAMPLE

To motivate the macroscopic equivalence of the processes generated by the Fokker–Planck equation (2) and the fluctuation–dissipation postulates, the equivalence of the equilibrium distributions is demonstrated for a simple example. Consider the chemical reaction

$$X + Y \rightleftharpoons 2X$$

with $K^+ = K^- = 1$. If no external sources are present, then $n_x + n_y = n = \text{const}$, and Eq. (1) simplifies to

$$\frac{dn_x}{dt} = n_x(n - n_x) - n_x^2. \tag{12}$$

Only fluctuations in n_x need be considered since $\delta n_x = -\delta n_y$, and so Eq. (3) becomes

$$\gamma(n_x) = \Omega^{-1}[n_x(n - n_x) + n_x^2] = \Omega^{-1} n n_x, \tag{13}$$

where Ω is the volume. For the single independent variable n_x, the Fokker–Planck equation (2) becomes

$$\frac{\partial W}{\partial t} + \frac{\partial n_x(n - 2n_x)W}{\partial n_x} - \frac{1}{2}\frac{\partial^2 \Omega^{-1} n n_x W}{\partial n_x^2} = 0. \tag{14}$$

The difficulty in solving Eq. (14) is evident because of the nonconstant coefficients. However, an exact solution for the equilibrium density can be found since the equilibrium density solves

$$\frac{\partial n_x(n - 2n_x)W}{\partial n_x} = \frac{n}{2\Omega}\frac{\partial^2 n_x W}{\partial n_x^2}. \tag{15}$$

Equation (15) can be integrated twice in a straightforward fashion to yield

$$W^d(n_x) = (\Omega A / n n_x)\exp[-2\Omega(n_x - n^e)^2/n] \times [1 + B(n/2\Omega)^{1/2}\int_0^{(2\Omega/n)^{1/2}(n^e - n_x)} \exp(y^2)\,dy].$$

The superscript "d" denotes the result for the diffusion process, A and B are constants, and $n^e = n/2$ is the equilibrium value of n_x. In order that the density can be normalized, the singularity at $n_x = 0$ is removed by setting[19]

$$B(n/2\Omega)^{1/2} = -\int_0^{(2\Omega/n)^{1/2} n^e} \exp(y^2)\,dy.$$

Thus the equilibrium density for the diffusion process is

$$W^d(n_x) = (W^0/n_x)\exp[-2\Omega(n_x - n^e)^2/n] \times (1 - \int_0^{(2\Omega/n)^{1/2}(n^e - n_x)} \exp(y^2)\,dy / \int_0^{(2\Omega/n)^{1/2} n^e} \times \exp(y^2)\,dy) \tag{16}$$

with W^0 a normalization constant.

For comparison, the equilibrium density for this reaction can be gotten from the fluctuation–dissipation postulates by taking the limit $t \to \infty$ in Eq. (4). According to previous work,[9] this gives

$$W^{fd}(n_x) = (\pi n/2\Omega)^{1/2}\exp[-2\Omega(n_x - n^e)^2/n]. \tag{17}$$

The expressions for the equilibrium probability densities in Eqs. (16) and (17) are exact for the diffusion and fluctuation–dissipation processes and for arbitrary Ω and n are obviously not the same. However, it is easy to see that in the limit $\Omega \to \infty$ with n fixed, the density in Eq. (16) approaches the Gaussian form in Eq. (17). This follows since $0 \leq |n_x - n^e| \leq n^e$, so that as $\Omega \to \infty$ with n fixed, the integral ratio in Eq. (16) vanishes unless $n_x = 0$ or n. Also in this limit the density $W^d(n_x)$ becomes sharply peaked around n^e and so has as a limiting form the Gaussian in Eq. (17).

Incidentally, it is easy to check, for this example, that the birth and death process associated with this reaction[12] is different from these two stochastic process. Using standard results,[20] the birth and death theory gives a binomial distribution for the equilibrium probability that the number of X molecules is N_x, i.e.,

$$W^{bd}(N_x) = N!/2^N(N - N_x)!\,N_x!, \tag{18}$$

where $n = N/\Omega$ and $n_x = N_x/\Omega$. This result is different from both Eq. (18) and Eq. (17), but according to the Laplace–De Moivre central limit theorem,[21] the fluctuation–dissipation result in Eq. (17) is recovered in the limit $\Omega \to \infty$ with n and n_x fixed.

IV. THE LIMIT $\Omega \to \infty$

The example of the previous section can be generalized by looking at the conditional probability densities for the diffusion process in Eq. (2). Although Eq. (2) cannot be solved directly, the fact that the diffusion term vanishes in the limit $\Omega \to \infty$ permits the introduction of systematic approximations which allow the asymptotic density to be obtained for delta-function initial distributions.

The complete Fokker–Planck equation for the conditional density $P_2^d(\mathbf{n}^0|\mathbf{n}t)$ for the diffusion process in Sec. II is

$$\frac{\partial P_2^d}{\partial t} + \frac{\partial V_i P_2^d}{\partial n_i} - \frac{1}{2}\frac{\partial^2 \gamma_{ij} P_2^d}{\partial n_i \partial n_j} = 0, \qquad P_2^d(\mathbf{n}^0|\mathbf{n}0) = \delta(\mathbf{n} - \mathbf{n}^0). \tag{19}$$

To clarify the expansion which follows, the literal limit $\Omega \to \infty$ is taken on both sides of Eq. (19). Since Eq. (3) shows that γ_{ij} is proportional to Ω^{-1}, and V_i is volume independent, this gives for $\bar{P}_2^d = \lim_{\Omega\to\infty} P_2^d$

$$\frac{\partial \bar{P}_2^d}{\partial t} + \frac{\partial V_i \bar{P}_2^d}{\partial n_i} = 0, \qquad \bar{P}_2^d(\mathbf{n}^0|\mathbf{n}0) = \delta(\mathbf{n} - \mathbf{n}^0). \tag{20}$$

In this equation only a drift term remains and the unique solution to Eq. (20) is the delta function

$$\bar{P}_2^d(\mathbf{n}^0|\mathbf{n}t) = \delta(\bar{\mathbf{n}} - \bar{\mathbf{n}}(\mathbf{n}^0, t)), \tag{21}$$

where $\mathbf{n}(\mathbf{n}^0, t)$ is the deterministic solution to the rate Eq. (1). Thus to lowest order in Ω^{-1}, the solution is perpetually sharp around the deterministic solution.

The persistence of a sharp deterministic solution in the infinite volume limit suggests that the asymptotic form of the probability density can be obtained by looking at the deviations around the deterministic solution, $\delta n_j = n_j - \bar{n}_j(\mathbf{n}^0, t)$. These deviations are small, and a proper scaling is required to see their distribution in the limit that $\Omega \to \infty$. The usual central limit theorem type scaling, $\delta n_j = \Omega^{-1/2} q_j$, is appropriate. This leads to the Taylor expansions for the streaming and diffusion terms,

$$V_i(\mathbf{n}, t) = V_i(\bar{\mathbf{n}}(\mathbf{n}^0, t), t) + \sum_{k=1}^{K} \bar{V}_i^{j_1,\ldots,j_k} q_{j_1} \cdots q_{j_k} \Omega^{-k/2}, \tag{22}$$

$$\gamma_{ij}(\mathbf{n}) = \gamma^*_{ij}(\bar{\mathbf{n}}, (\mathbf{n}^0, t))\Omega^{-1} + \sum_{j=1}^{K} \bar{\gamma}^{*\, j_1,\ldots,j_k}_{ij} q_{j_1} \cdots q_{j_k} \Omega^{-(2+k)/2}. \tag{23}$$

The volume dependence of γ [see Eq. (3)] has been made explicit in Eq. (23) by writing $\gamma_{ij} = \gamma^*_{ij}\Omega^{-1}$, the overbars represent evaluation at $\bar{\mathbf{n}}$, and the superscripts represent partial derivatives. For the mass action laws in Eq. (1′) these expansions terminate at some small integer K which represents the highest order of the reaction occurring. In realistic examples K is no larger than 3.

The density of the scaled variables $\mathbf{q}$ is

$$P^d(\mathbf{q}, t) \equiv P_2^d(\mathbf{n}^0 \,|\, \bar{\mathbf{n}}(\mathbf{n}^0, t) + \Omega^{-1/2}\mathbf{q}, t).$$

The equation solved by this function can be gotten from Eq. (19) and the identity

$$\frac{\partial P^d}{\partial t} = \frac{\partial P_2^d}{\partial t} + \frac{\partial P_2^d}{\partial n_i}\frac{dn_i}{dt} = \frac{\partial P_2^d}{\partial t} + \frac{\partial P^d}{\partial q_i} V_i(\bar{\mathbf{n}}, t)\Omega^{1/2}. \tag{24}$$

Substituting this result and the Taylor expansions in Eq. (19) yields

$$\frac{\partial P^d}{\partial t} + \partial \bar{V}_i^j q_j + \sum_{k=2}^{K} V_i^{j_1,\ldots,j_k} q_{j_1} \cdots q_{j_k} \Omega^{(1-k)/2} P^d \Big/ \partial q_i - \tfrac{1}{2}\partial^2 \bar{\gamma}^*_{ij} + \sum_{k=1}^{K} \bar{\gamma}^{*\, j_1,\ldots,j_k}_{ij} q_{j_1} \cdots q_{j_k} \Omega^{-k/2} P^d \Big/ \partial q_i \partial q_j = 0. \tag{25}$$

As long as q_j stays the order of unity in Ω, this is a systematic expansion of the differential equation into powers of $\Omega^{-1/2}$. Only the first terms inside the derivatives survive the limit $\Omega \to \infty$ and asymptotically the equation becomes

$$\frac{\partial P^d}{\partial t} + \frac{\partial \bar{V}_i^j q_j P^d}{\partial q_i} - \frac{1}{2}\frac{\partial^2 \bar{\gamma}^*_{ij} P^d}{\partial q_i \partial q_j} = 0.$$

Returning now to the unscaled variables δn_j, the limiting equation becomes

$$\frac{\partial P^d}{\partial t} + \partial H_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t))\frac{\delta n_j P^d}{\partial \delta n_i} - \frac{1}{2}\frac{\partial^2 \gamma_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t))P^d}{\partial \delta n_i \partial \delta n_j} = 0, \qquad P^d(\delta\mathbf{n}, 0) = \delta(\delta\mathbf{n}), \tag{26}$$

where the identity $\bar{V}_i^j = H_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t))$ has been used. Equation (26) is identical to Eq. (11), which is the equation solved by $P_2^{fd}(\mathbf{n}^0 \,|\, \mathbf{n}t)$, the conditional density obtained from the fluctuation–dissipation postulates. Thus the unique solution to Eq. (26) is the Gaussian density in Eq. (4).

Although these considerations are suggestive, they do not really constitute a derivation of the limiting equivalence of the stochastic diffusion in Eqs. (1)–(3) and the fluctuation–dissipation process in Eqs. (7) and (8). The difficulty is in justifying the neglect of the higher order terms in q_j. An identical problem arises in van Kampen's expansion[11] which has been used to treat the $\Omega \to \infty$ limit of master equations.[14,15] In both cases what is needed is a proof that the coefficient of $\Omega^{-k/2}$ in Eq. (25) is of order $\Omega^{(k-1)/2}$ or smaller. Although this problem has been ignored in most treatments of master equations,[11] the expansion procedure is justified by Kurtz's work.[13] For the diffusion process introduced here, this problem is confronted in the following section.

V. A LIMIT THEOREM

The derivation of the $\Omega \to \infty$ limit in the previous section suffers from the limitation that the coefficients of the powers of $\Omega^{-1/2}$ in Eq. (25) have not been shown to be of negligible order. On the other hand, it is easy to show that these terms are negligible for the Gaussian solution which results from this assumption. Consequently the neglect of these terms is at least self-consistent. Actually more than this is true, and the mean convergence of the exact solution to the Gaussian solution is shown below.

The results in this section are based on the following L_1 estimate of the difference between the exact solution to the Fokker–Planck equation (19), P_2^d, and the Gaussian P_2^{fd} which is derived in the Appendix:

$$\| P_2^d(t) - P_2^{fd}(t) \| \equiv \int \left| P_2^d(\mathbf{n}^0 \,|\, \mathbf{n}t) - P_2^{fd}(\mathbf{n}^0 \,|\, \mathbf{n}t) \right| d\mathbf{n} \leq \int_0^T d\tau \int d\mathbf{n} \left| L_R P_2^{fd}(\mathbf{n}^0 \,|\, \mathbf{n}\tau) \right|, \tag{27}$$

where $0 \leq t \leq T$ and L_R is the differential operator involving the higher order terms in Eq. (24), i.e.,

$$L_R \bullet \equiv \left(\partial \sum_{k=2}^{K} \bar{V}_i^{j_1,\ldots,j_k} \delta n_{j_1} \cdots \delta n_{j_k} \bullet\right) \Big/ \partial n_i - \frac{1}{2}\left(\partial^2 \sum_{k=1}^{K} \bar{\gamma}_{ij}^{j_1,\ldots,j_k} \delta n_{j_1} \cdots \delta n_{j_k} \bullet\right) \Big/ \partial n_i \partial n_j. \tag{28}$$

The inequality in Eq. (27) can be used to show that the Gaussian density P_2^{fd} remains a good approximation to P_2^d for appreciable times if the volume is large. To see this, the integrals over concentration $\mathbf{n}$ are written out more completely taking advantage of the *exact* scaling $\delta n_j = \Omega^{-1/2} q_j$ for the Gaussian solution, i.e., from Eqs. (3) and (5), $\sigma(\mathbf{n}^0, t) \propto \Omega^{-1}$. This variable change gives

$$\int d\mathbf{n} \left| L_R P_2^{fd} \right| = \int d\mathbf{q} \left| \left(\partial \sum_{k=2}^{K} \bar{V}_i^{j_1,\ldots,j_k} q_{j_1} \cdots q_{j_k} \Omega^{-(k-1)/2} P^{fd}(\mathbf{q})\right) \Big/ \partial q_i + \left(\partial^2 \sum_{k=1}^{K} \bar{\gamma}^{*\, j_1,\ldots,j_k}_{ij} q_{j_1} \cdots q_{j_k} \Omega^{-k/2} P^{fd}(\mathbf{q})\right) \Big/ \partial q_i \partial q_j \right|, \tag{29}$$

where $P^{fd}(\mathbf{q})$ is the Gaussian density resulting from Eq. (4) for the change of variable $n_j - \bar{n}_j = \Omega^{-1/2} q_j$. Because $\sigma(\mathbf{n}^0, t)$ is proportional to Ω^{-1}, $P^{fd}(\mathbf{q})$ is *independent of* Ω. Since neither the derivatives of V or γ^* depend on Ω, this explicitly shows that the right-hand side of Eq. (27) is no greater than order $\Omega^{-1/2}$. Equation (29) also shows

that the coefficients of $\Omega^{-k/2}$ are bounded above by Gaussian moment integrals over positive values of $\mathbf{q}$ and so are finite as long as $\sigma(\mathbf{n}^0, t)$ is invertible. Thus Eq. (27) can be written as

$$\| P_2^d(t) - P_2^{fd}(t) \| < \sum_{k=1}^{K} \int_0^T I_k(\tau)\, d\tau \Omega^{-k/2}, \tag{30}$$

where the form of the integrals $I_k(\tau)$ can be obtained from the moment expressions discussed above. If these integrals are finite, then for fixed T the right-hand side vanishes as $\Omega \to \infty$. The finiteness of these time intergrals is quaranteed by Eq. (5) which shows that at short times $\sigma(\mathbf{n}^0, \tau) \propto \tau$. Hence a scaling argument using $q_j \propto \tau^{1/2}$ for short times shows that $I_k(\tau) \propto \tau^{l/2}$ with l an integer no smaller than -1. This means that all the time integrals in Eq. (30) are finite because the only potential singularity is at $l = 0$ (where σ vanishes) if dynamical constraints among the variables have been removed.

These considerations establish the following result: Fix the initial state $\mathbf{n}^0$. Then given any $T > 0$,

$$\lim_{\Omega \to \infty} \| P_2^d(t) - P_2^{fd}(t) \| = 0, \tag{31}$$

for all $0 \leq t \leq T$. With this mean sense of convergence, the stochastic diffusion process and the fluctuation–dissipation process become identical in the macroscopic limit.

VI. ERROR ESTIMATES

The preceding limit theorem is useful for obtaining error estimates. A simple illustration is provided by the second order reaction $X + Y \rightleftarrows 2X$ of Sec. III. For this reaction the remainder operator L_R is first evaluated and found to be

$$L_R \cdot = -2 \frac{\partial (\delta n_x)^2}{\partial n_x} \cdot - \frac{1}{2} \frac{\partial^2 \Omega^{-1} n \delta n_x}{\partial n_x^2} \cdot \tag{32}$$

Using the properties of the Gaussian solution $P_2^{fd}(n_x^0 | n_x t)$ [see Eqs. (4) and (5)], the inequality in Eq. (27) becomes

$$\| P^d(t) - P_2^{fd}(t) \| < \int_0^T (16[\sigma^0(\tau)/2\pi]^{1/2} + 5n/[2\pi\Omega^2 \sigma^0(\tau)]^{1/2}) d\tau \tag{33}$$

for all $0 \leq t \leq T$, where $\sigma^0(\tau) \equiv \sigma(n_x^0, \tau)$ as defined in Eq. (5). For small τ, Eq. (3) gives $\sigma^0(\tau) \approx \gamma(n_x^0)\tau$ so that the second integral on the right-hand side of Eq. (33) is finite. Furthermore $\sigma^0(\tau) \propto \Omega^{-1}$, so that the right-hand side of Eq. (33) is proportional to $\Omega^{-1/2}$, which is an example of the expansion in Eq. (30).

Although it would be possible to use the explicit expression for $\sigma^0(\tau)$ to obtain an estimate in Eq. (33), a different property of the reaction will be used. For the reaction in Eq. (12), as t gets large the variance approaches that of the equilibrium distribution,[9] i.e., $\lim_{t\to\infty} \sigma^0(t) = \sigma^e = n/4\Omega$. Thus for T large enough, Eq. (33) implies that

$$\| P_2^d(t) - P_2^{fd}(t) \| < T(16(\sigma^e/2\pi)^{1/2} + 5n/(2\pi\Omega^2\sigma^e)^{1/2}) < 9T(n/\Omega)^{1/2} \tag{34}$$

for $0 \leq t \leq T$. Since the probability densities P_2^{fd} and P_2^d are normalized to unity, Eq. (34) suggests that for a given T, a criterion for the validity of the Gaussian approximation is

$$9T(n/\Omega)^{1/2} \ll 1. \tag{35}$$

If T is scaled in terms of "relaxation times," $\tau_0 \equiv |\partial V/\partial n_x^0|^{-1} = |n - 4n_x^0|^{-1}$, then $T = N_0 \tau_0$. Thus the criterion in Eq. (35) can be expressed in terms of N_0, the number of elapsed relaxation times, as

$$N_0 \ll (n\Omega)^{1/2}/3. \tag{36}$$

Now $n\Omega$ is just the number of molecules in the volume Ω, so Eq. (36) asserts that the solution to the diffusion equation will agree with the fluctuation–dissipation postulates as long as the number of half-lives elapsed is much smaller than the square root of the number of molecules in the volume. For example, if a volume contains as few as 10 000 molecules, then the Gaussian approximation would hold for at least five half-lives. However, equilibrium is essentially complete within five half-lives, so even for such a small system the agreement between the diffusion process and the fluctuation–dissipation process persists effectively forever.

ACKNOWLEDGMENTS

The author would like to thank Professor Robert Mazo for drawing his attention to Graham's work and the National Science Foundation for research support through Research Grant No. CHE 74-00483 A03.

APPENDIX

In order to obtain the estimate in Eq. (2), notice that $P_2^d(\mathbf{n}^0 | \mathbf{n}t)$ satisfies Eq. (19) and that $P_2^{fd}(\mathbf{n}^0 | \mathbf{n}t)$ satisfies Eq. (26). Consequently, using Eq. (24), their difference $\Delta \equiv P_2^d - P_2^{fd}$ can be shown to satisfy the partial differential equation

$$\frac{\partial \Delta}{\partial t} + \frac{\partial V_i \Delta}{\partial n_i} - \frac{1}{2} \frac{\partial^2 \gamma_{ij} \Delta}{\partial n_i \partial n_j} = -L_R P_2^{fd}, \qquad \Delta(\mathbf{n}^0 | \mathbf{n}0) = 0, \tag{A1}$$

where

$$L_R \cdot = \frac{\partial R_i}{\partial n_i} \cdot - \frac{1}{2} \frac{\partial^2 R_{ij}}{\partial n_i \partial n_j} \cdot,$$
$$R_i = V_i(\mathbf{n}, t) - V_i(\bar{\mathbf{n}}(\mathbf{n}^0, t), t) - H_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t))\delta n_j, \tag{A2}$$
$$R_{ij} = \gamma_{ij}(\mathbf{n}) - \gamma_{ij}(\bar{\mathbf{n}}(\mathbf{n}^0, t)).$$

The solution to Eq. (A1) involves the Green's function $P_2^d(\mathbf{n}' | \mathbf{n}t)$ and is

$$\Delta(\mathbf{n}^0 | \mathbf{n}t) = -\int_0^t d\tau \int d\mathbf{n}' P_2^d(\mathbf{n}' | \mathbf{n}t - \tau) L_R P_2^{fd}(\mathbf{n}^0 | \mathbf{n}'\tau), \tag{A3}$$

as can be verified by differentiation. Since $P_2^d(\mathbf{n}' | \mathbf{n}t)$ is positive and a normalized probability density, Eq. (A3) leads directly to

$$\begin{aligned} \| P_2^d(t) - P_2^{fd}(t) \| &\equiv \int | \Delta(\mathbf{n}^0 | \mathbf{n}t) | d\mathbf{n} \\ &\leq \int_0^t d\tau \int d\mathbf{n}' \, | L_R P_2^{fd}(\mathbf{n}^0 | \mathbf{n}'\tau) | \\ &\leq \int_0^T d\tau \int d\mathbf{n}' \, | L_R P_2^{fd}(\mathbf{n}^0 | \mathbf{n}'\tau) | \end{aligned} \tag{A4}$$

for all $0 \leq t \leq T$, since the integrand of the time integral is nonnegative.

[1]A. Einstein, *Investigations on the Theory of the Brownian Movement* (Dover, New York, 1965).
[2]M. von Smoluchowski, Ann. Phys. **21**, 756 (1906).
[3]A.D. Fokker, Ann. Phys. **43**, 810 (1914); M. Planck, Sitzungsber. Preuss. Akad. Wiss., 324 (1917).
[4]P. Langevin, C.R. Acad. Sci. Paris **146**, 530 (1908).
[5]G.E. Uhlenbeck and L.S. Ornstein, Phys. Rev. **36**, 823 (1930).
[6]P. and T. Ehrenfest, Phys. Z. **8**, 311 (1907); M. Kac, Am. Math. Mon. **54**, 369 (1947).
[7]M.C. Wang and G.E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).
[8]R. Graham, in *Quantum Statistics in Optics and Solid State Physics*, Springer Tracts in Modern Physics No. 66 (Springer, Berlin, 1973), p. 1; work related to Graham's approach is reviewed in H. Haken, Rev. Mod. Phys. **47**, 67 (1975); Z. Phy. B **24**, 321 (1976).
[9]J. Keizer, J. Chem. Phys. **63**, 398 (1975).
[10]J. Keizer, J. Chem. Phys. **64**, 1679 (1976).
[11]N.G. van Kampen, Can. J. Phys. **39**, 551 (1961); R. Kubo, K. Matsuo, and K. Kitahara, J. Stat. Phys. **9**, 51 (1973); H. Mori, Prog. Theor. Phys. **53**, 1617 (1975); N.G. van Kampen, Adv. Chem. Phys. **34**, 246 (1976).
[12]D. McQuarrie, J. Appl. Prob. **4**, 413 (1967).
[13]T. Kurtz, J. Appl. Prob. **7**, 49 (1970); **8**, 344 (1971); T. Kurtz, J. Chem. Phys. **57**, 2976 (1972).
[14]J. Keizer, J. Chem. Phys. **63**, 5037 (1975).
[15]J. Logan and M. Kac, Phys. Rev. A **13**, 458 (1976); R. Fox and M. Kac, *Biosystems*, to be published (1977).
[16]L. Arnold, *Stochastic Differential Equations: Theory and Applications* (Wiley, New York, 1974). Secs. 2.5 and 2.6.
[17]In Eq. (5) the time ordered exponential involves Dyson's chronological ordering operator, P. See F. Dyson, Phys. Rev. **75**, 486 (1949).
[18]Ref. 16, Secs. 9.3 and 10.2.
[19]In this example the singularity at $n_x = 0$ is caused by the fact that when no X is in the system, the mechanism in Eq. (12) causes no changes in the amount of X or Y. This singular point has been removed for all three stochastic processes in this section. Otherwise each equilibrium distribution could be partly concentrated on $n_x = 0$.
[20]P.G. Hoel, S.C. Port, and C.J. Stone, *Introduction to Stochastic Processes* (Houghton Mifflin, Boston, Massachusetts, 1972), Sec. 3.3.1. In this regard, note that the isolated point $n_x = 0$ has been eliminated from the domain of the birth and death process.
[21]W. Feller, *An Introduction to Probability and Its Application* (Wiley, New York, 1957), 2nd ed., Vol. I, p. 168.

Dispersive properties and observables at infinity for classical KMS systems

J. De Cannière,* J. Pulè,† P. Vanheuverzwijn,* and A. Verbeure

Universiteit Leuven, Belgium
(Received 12 July 1976)

For infinite classical dynamical systems, satisfying the KMS condition, relations between asymptotic dispersive and cluster properties are proved. The local structure of the algebra of observables is explicitly characterized by the Poisson bracket commutant, and it is proved that the algebra of observables at infinity are constants of the motion.

1. INTRODUCTION

As in Ref. 1, 2, and 3, we describe an infinite classical dynamical system by a triplet $Y=(\mathcal{A}, \omega, \alpha_t)$ where $\mathcal{A}$ is an algebra of bounded functions on the phase space K, ω a state on $\mathcal{A}$, and α_t stands for the time evolution. (For exact definitions see Sec. 2.)

This work consists of two parts. The first was motivated by Ref. 3 where the classical KMS condition is derived from a notion of stability under some conditions of clustering and dispersiveness. In particular we study some properties of what we call asymptotic dispersiveness in accordance with the notion introduced in Ref. 3. It corresponds to asymptotic abelianess in the quantum case. In comparison with the quantal asymptotic abelianess where the commutator $[A, B_t]$ of some observables A and B_t is studied, in the classical case we have to consider the Poisson bracket $\{f, g_t\}$ of the observables f and g_t for large times t. The essential difference is that the commutator is a bounded linear map on the algebra of bounded observables, but the Poisson bracket an unbounded one. In Ref. 2, respectively Ref. 3, it is assumed that the pair (K, ω) is a Lebesgue space and that the Poisson bracket of any pair of observables is L_1, respectively L_2. We give a purely algebraic description of the classical KMS condition, i.e., at no place do we refer to the time evolution as to be induced by a flow on the phase space. It was done in this way, because we study only properties following from the algebraic structure. It is indicated how it is related to Refs. 2 and 3. The main result of Sec. 3 is the link between cluster properties and asymptotic dispersiveness for classical KMS states, due to the self-adjointness of the Liouville operator. This result makes it possible to translate these properties in terms of properties on the spectrum of the time evolution. In the second part, Sec. 4, we consider in detail the local structure of $\mathcal{A}$. This enables us to introduce the algebra of observables at infinity[4] which should correspond to the center of the von Neumann algebra for the representation induced by the state, at least for locally normal states in the quantum case. We give, as far as the algebra $\mathcal{A}$ is concerned, an explicit construction of the Poisson bracket commutant of a local algebra. To extend this property to the weak closures, i.e., to develop a more complete "von Neumann algebra"-like theory with respect to the Poisson bracket, another condition is needed, again because of the unboundedness of this operation. Finally we prove that, as in the quantum case, the observables at infinity are all constants of the motion.

2. PRELIMINARIES

Let K be the set of sequences in $\mathbf{R}^d \times \mathbf{R}^d$ with $d \in \mathbf{N}$, i.e., $x \in K$; $x(i) = (q_i, p_i) \in \mathbf{R}^{2d}$ subject to the locality condition $\lim_{i\to\infty} |q_i| = \infty$ and such that $q_i \neq q_j$ if $i \neq j$.

Denote by K the equivalence classes of such mappings, two mappings being equivalent if they differ only by a permutation. We call K the set of configurations.

Denote by $\tilde{A}$ the algebra of (unbounded) observables consisting of functions $f: K \to \mathbf{R}$ which are described by a sequence $(f^{(m)})_0^\infty$ of C^∞ functions $f^{(m)}: (\mathbf{R}^d \times \mathbf{R}^d)^m \to \mathbf{R}$ which are symmetric, have compact support and with a finite number of components $f^{(m)}$ different from zero.

Then for any $x \in K$ and $f \in \tilde{A}$

$$f(x) = \sum_{m=0}^{\infty} \sum_{i_1 < \cdots < i_m} f^{(m)}(x_{i_1}, \ldots, x_{i_m}) \equiv \sum_{m=0}^{\infty} (\textstyle\sum f^{(m)})(x).$$

Denote by $\mathcal{A}$ the complex $*$-algebra of bounded local observables as the algebra generated by the real-valued functions $\mathcal{G}$ on K of the form

$$\mathcal{G}(x) = \Psi(f(x)),$$

where $f \in \tilde{A}$ and Ψ is a bounded C^∞ function from $\mathbf{R}$ to $\mathbf{R}$, with all derivatives bounded.

A state ω is, as usual, a normalized positive linear functional on the involutive algebra $\mathcal{A}$ with unit, the unit constant on K.

Denote by $\{\cdot, \cdot\}$ the bilinear map from $\mathcal{A} \times \mathcal{A}$ into the *unbounded* functions on K, for all $\Psi, \Phi \in \mathcal{A}$

$$\{\Psi, \Phi\}(x) = \sum_i \frac{\partial \Psi}{\partial q_i} \frac{\partial \Phi}{\partial p_i}(x) - \frac{\partial \Psi}{\partial p_i} \frac{\partial \Phi}{\partial q_i}(x)$$

with $(\partial/\partial q_i)(\partial/\partial p_i)$ the scalar product of the gradients. Let π_ω be the GNS representation of A into the linear operators on H_ω, induced by a state ω; let Ω_ω be the cyclic vector, and ϕ the canonical embedding of $\mathcal{A}$ into H_ω.

Definition 2.1 Let $\mathcal{A}_r$ be the real part of $\mathcal{A}$, then we call the state ω admissible if there exists a set $\mathcal{D}$ in $\mathcal{A}_r$ such that $\phi(\mathcal{D})$ is dense in $\phi(\mathcal{A}_r)$ and for all $f, g \in \mathcal{D}$,

$$\lim_{\lambda \to 0} \omega\left(\frac{\exp(i\lambda\{f, g\}) - 1}{i\lambda}\right)$$

exists. We denote this limit by $\omega(\{f, g\})$. It is clear that $\omega(\{\cdot, \cdot\})$ is a bilinear form on $\mathcal{D}$. By linearity we extend it to $\mathcal{E} \equiv \mathcal{D} + i\mathcal{D}$.

In Refs. 2 and 3 where H_ω is taken to be $L_2(K, \omega)$, it was assumed that for all $f, g \in \mathcal{A}$, $\{f, g\} \in L_1(K, \omega)$. If now $f, g \in \mathcal{A}_r$, then

$$\left|\frac{\exp(i\lambda\{f, g\}) - 1}{i\lambda}\right| \leq |\{f, g\}|$$

and hence by the bounded convergence theorem, if $\{f, g\} \in L_1(K, \omega)$, then

$$\int \{f, g\}\, d\omega = \lim_{\lambda \to 0} \int \left(\frac{\exp(i\lambda\{f, g\}) - 1}{i\lambda}\right) d\omega.$$

This motivates Definition 2.1.

If furthermore $f, g \in \mathcal{A}$, $\{f, g\} \in L_2(K, \omega)$ as in Ref. 3, then, as for $f, g \in A_r$,

$$\left|\frac{\exp(i\lambda\{f, g\}) - 1}{i\lambda} - \{f, g\}\right|^2 \leq 4|\{f, g\}|^2,$$

again by the bounded convergence theorem

$$\frac{\exp(i\lambda\{f, g\}) - 1}{i\lambda} \text{ converges to } \{f, g\} \text{ in } L_2(K, \omega).$$

This suggests another way of defining $\omega(\{f, g\})$. Suppose that $f, g \in \mathcal{D}$, the sequence

$$n\phi[\exp((i/n)\{f, g\}) - 1], \quad n \in \mathbb{N}$$

is Cauchy. Then define $\phi(\{f, g\})$ by

$$\phi(\{f, g\}) = -i \lim_{n \to \infty} n\phi[\exp((i/n)\{f, g\}) - 1]$$

and

$$\omega(\{f, g\}) = (\Omega_\omega, \phi(\{f, g\})\Omega_\omega).$$

The advantage of this is that it identifies $\{f, g\}$ with an element of H_ω. In this work we only use Definition 2.1, which is implied by the ones in Refs. 2 and 3.

Definition 2.2: We call the one-parameter group $(\alpha_t)_{t \in \mathbb{R}}$ of $*$-automorphisms of $\mathcal{A}$, a time evolution of the physical system $(\mathcal{A}, \omega)$ if there exists a strongly continuous group of unitaries $(U_t)_{t \in \mathbb{R}}$ on H_ω such that $\pi(\alpha_t f) = U_t \pi(f) U_{-t}$. Furthermore we suppose:

(i) if L is the infinitesimal generator of U_t, i.e., $U_t = \exp itL$ then the domain of L, denoted by $\mathcal{D}(L)$ contains $\phi(\mathcal{A})$

(ii) the subset $\mathcal{E}$ of $\mathcal{A}$ is invariant under α_t

(iii) α_t leaves ω invariant, i.e., $\omega \cdot \alpha_t = \omega$.

It must be realized that most realistic time evolutions are not described by automorphisms of this algebra, but of a larger one. The content of Sec. 3 is however independent of the choice of the algebra.

Motivated by the conditions of the main theorem of Ref. 3, we now introduce the following notions: Let $Y = (\mathcal{A}, \omega, \alpha_t)$ be any dynamical classical system, where $\mathcal{A}$ is the algebra as defined above, ω an admissible state, and α_t satisfies definition 2.2.

Definition 2.3: The dynamical system Y is called

(i) weakly asymptotic dispersive, if for all $f, g \in \mathcal{E}$

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \omega(\{\alpha_t(f), g\})\, dt = 0,$$

(ii) absolutely weakly asymptotic dispersive if for all $f, g \in \mathcal{E}$

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} |\omega(\{\alpha_t(f), g\})|\, dt = 0,$$

and

(iii) strongly asymptotic dispersive if for all $f, g \in \mathcal{E}$

$$\lim_{|t| \to \infty} \omega(\{\alpha_t(f), g\}) = 0.$$

It is easily seen that (iii) implies (ii), which implies (i).

Definition 2.4: The dynamical system Y is called, weakly clustering if for all $f, g \in \mathcal{A}$

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \omega(\alpha_t(f)g)\, dt = \omega(f)\omega(g),$$

absolutely weakly clustering if for all $f, g \in \mathcal{A}$

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} dt\, |\omega(\alpha_t(f)g) - \omega(f)\omega(g)| = 0,$$

and strongly clustering if for all $f, g \in \mathcal{E}$

$$\lim_{|t| \to \infty} \omega(\alpha_t(f)g) = \omega(f)\omega(g).$$

Definition 2.5: Let $Y = (\mathcal{A}, \omega, \alpha_t)$ be as above, then the classical dynamical system is said to satisfy the KMS condition at inverse temperature $\beta > 0$, if for all $f, g \in \mathcal{E}$ and all $t \in \mathbb{R}$, Y satisfies

$$\beta \frac{d}{dt} \omega(f\alpha_t g) = -\omega(\{f, \alpha_t g\}).$$

In the following we put $\beta = 1$.

3. DISPERSIVENESS

The main result in the following theorem is the relation between clustering and dispersive properties of KMS states. Denote by E_0 the orthogonal projection on the U_t invariant subspace of H_ω.

Theorem 3.1: If Y is a KMS system then

(i) Y is weakly asymptotic dispersive;

(ii) if Y is strongly clustering, respectively absolutely weakly clustering, then Y is strongly asymptotic dispersive, respectively absolutely weakly asymptotic dispersive. If, in addition E_0 is one-dimensional, then also the converse holds.

Proof: (i) Using the KMS condition and the invariance of ω for $f, g \in \mathcal{E}$,

$$\omega(\{\alpha_t(f)g\}) = -i(L\, \pi(f)^* \Omega, U_{-t}\pi(g)\Omega).$$

Let us denote

$$M(f(t)) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)\, dt.$$

Then[5]

$$(\mathcal{L}\pi(f)^*\Omega, U_t\pi(g)\Omega) = (\mathcal{L}\pi(f)^*\Omega, E_0\pi(g)\Omega).$$

As $\mathcal{L}E_0 = 0$ (i) follows.

Now we prove (ii): Suppose first that Y is strongly clustering, i.e., for all $f, g \in \mathcal{A}$,

$$\lim_{|t|\to\infty} (\pi(f)^*\Omega, U_t\pi(g)\Omega) = (\pi(f)^*\Omega, \Omega)(\Omega, \pi(g)\Omega).$$

Since Ω is cyclic for $\pi(\mathcal{A})$ we have for all $\phi, \psi \in H_\omega$

$$\lim_{|t|\to\infty} (\phi, U_t\psi) = (\phi, \Omega)(\Omega, \psi).$$

Take ϕ equal to $\mathcal{L}\pi(f)^*\Omega$ and ψ equal to $\pi(g)\Omega$ with f and g in $\mathcal{E}$, then

$$\lim_{|t|\to\infty} (\mathcal{L}\pi(f)^*\Omega, U_t\pi(g)\Omega) = 0$$

which, using the KMS condition, is equivalent with

$$\lim_{|t|\to\infty} \omega(\{f, \alpha_t g\}) = 0.$$

Similarly if Y is absolutely weakly clustering then it is absolutely weakly dispersive.

Suppose now that E_0 is one-dimensional and that Y is strongly asymptotic dispersive; then again using the KMS condition, this is equivalent to, for all f and g in $\mathcal{E}$,

$$\lim_{|t|\to\infty} (\mathcal{L}\pi(f)^*\Omega, U_t\pi(g)\Omega) = 0.$$

Let $N(\mathcal{L})$ be the kernel of $\mathcal{L}$. Since $\mathcal{L}$ is self-adjoint $N(\mathcal{L})$ is closed and $N(\mathcal{L}) = R(\mathcal{L})^\perp$ where $R(\mathcal{L})$ is the range of $\mathcal{L}$ (see Ref. 6, p. 267). Therefore, $N(\mathcal{L})^\perp = \overline{R(\mathcal{L})}$.

Let $\psi \in N(\mathcal{L})^\perp$ and $\phi \in H_\omega$. Since $\phi(\mathcal{E})$ is dense in H_ω, for all $\epsilon > 0$, there exists a $g \in \mathcal{E}$ such that

$$\|\phi - \pi(g)\Omega\| < \epsilon/3\|\psi\|.$$

As $R(\mathcal{L})$ is dense in $N(\mathcal{L})^\perp$, there is an element $\psi' \in D(\mathcal{L})$ such that

$$\|\psi - \mathcal{L}\psi'\| < \epsilon/3\|\pi(g)\Omega\|.$$

Finally, there exists an $f \in \mathcal{E}$ such that

$$\|\psi' - \pi(f)\Omega\| < \epsilon/3\|\mathcal{L}\pi(g)\Omega\|.$$

Therefore,

$$\begin{aligned}
|(\psi, U_t\phi| &\le |(\mathcal{L}\pi(f)\Omega, U_t\pi(g)\Omega)| \\
&\quad + |(\psi, U_t\phi) - (\mathcal{L}\pi(f)\Omega, U_t\pi(g)\Omega)| \\
&\le |(\mathcal{L}\pi(f)\Omega, U_t\pi(g)\Omega)| \\
&\quad + |(\psi, U_t\phi) - (\psi, U_t\pi(g)\Omega)| \\
&\quad + |(\psi, U_t\pi(g)\Omega) - (\mathcal{L}\psi', U_t\pi(g)\Omega)| \\
&\quad + |(\psi', U_t\mathcal{L}\pi(g)\Omega) - (\pi(f)\Omega, U_t\mathcal{L}\pi(g)\Omega)| \\
&\le |(\mathcal{L}\pi(f)\Omega, U_t\pi(g)\Omega)| + \epsilon.
\end{aligned}$$

As this is true for any $\epsilon > 0$ uniformly in t we get

$$\lim_{|t|\to\infty} (\psi, U_t\phi) = 0 \quad \text{for all } \psi \in N(\mathcal{L})^\perp \text{ and } \phi \in H_\omega.$$

Now let ϕ and ϕ' be arbitrary in H_ω, since E_0 is one-dimensional,

$$\phi' - (\Omega, \phi')\Omega \in N(\mathcal{L})^\perp.$$

Therefore,

$$\lim_{|t|\to\infty} (\phi' - (\Omega, \phi')\Omega, U_t\phi) = 0,$$

i.e.,

$$\lim_{|t|\to\infty} (\phi', U_t\phi) = (\phi', \Omega)(\Omega, \phi)$$

which is equivalent with Y being strongly clustering.

Similarly, if E_0 is one-dimensional and Y is absolutely weakly asymptotic dispersive then it is absolutely weakly clustering. Q.E.D.

For quantum systems the implications of cluster properties on the spectrum of the Hamiltonian have been studied before.[7] As a consequence of Theorem 3.1 we are in a position to make analogous statements about the relation between dispersiveness and the spectrum properties of the operator $\mathcal{L}$. However, as the proof of Ref. 7, Lemma II. 2.1 seems unclear to us we give an alternative proof, consisting mainly of a different proof of the following lemma.

Lemma 3.2: Let U be a unitary operator on a Hilbert space H and $V = U \otimes U^*$ a unitary on $H \otimes H$; let Ω be an eigenvector of U such that $U\Omega = \Omega$, then Ω is the unique eigenvector of U if and only if $\Omega \otimes \Omega$ is the only invariant vector of V in $H \otimes H$.

Proof: (a) If $\Omega \otimes \Omega$ is the only invariant eigenvector of V, then Ω is the only eigenvector of U since if $\Omega' \neq \Omega$ is another eigenvector such that $U\Omega' = \lambda\Omega$, then $V\Omega' \otimes \Omega' = |\lambda|^2\Omega' \otimes \Omega' = \Omega' \otimes \Omega'$.

(b) Conversely, suppose that Ω is the only eigenvector of U. Let T be the real linear isomorphism of $H \otimes H$ onto the set of Hilbert–Schmidt operators on H, defined by

$$T(x \otimes y)z = (y, z)x.$$

Then

$$\begin{aligned}
T(V(x \otimes y))z &= T(Ux \otimes U^*y)z = (U^*y, z)Ux \\
&= U(y, Uz)x = UT(x \otimes y)Uz
\end{aligned}$$

or

$$T(V(x \otimes y)) = UT(x \otimes y)U.$$

By linearity and continuity for all $\xi \in H \otimes H$

$$T(V(\xi)) = UT(\xi)U.$$

Suppose now that ψ is a normalized invariant vector of V, then

$$T(V\psi) = T(\psi)$$

or

$$UT(\psi)U = T(\psi).$$

Also

$$U^*T(\psi)^*U^* = T(\psi)^*$$

and therefore

$$UT(\psi)T(\psi)^*U^* = T(\psi)T(\psi)^*$$

or

$$[U, T(\psi)T(\psi)^*] = 0.$$

Therefore, the eigenspaces of $T(\psi)T(\psi)^*$ are also invariant subspaces of U. Since it is trace class all its nonzero eigenvalues are of finite multiplicity. Therefore, $T(\psi)T(\psi)^*$ has only one nonzero eigenvalue, which is simple, as this is true for U. Since

$$\mathrm{Tr}T(\psi)T(\psi)^* = 1,$$

it follows that

$$T(\psi)T(\psi)^* = E_\Omega,$$

where E_Ω is the orthogonal projection on Ω. The lemma is proved if we show that $\psi = \Omega \otimes \Omega$. Let $\{e_i\}_i$ be an orthonormal basis of H such that $e_0 = \Omega$, then ψ can be put in the form

$$\psi = \sum_i e_i \otimes f_i$$

for some sequence of vectors f_i in H_ω. Now

$$T(\psi)T(\psi)^* = \sum_{i,j} T(e_i \otimes e_j)(f_i, f_j)$$

hence

$$(\Omega, T(\psi)T(\psi)^*\Omega) = (f_0, f_0) = 1.$$

But

$$\|\psi\|^2 = \sum_i \|f_i\|^2 = 1.$$

Therefore,

$$f_i = 0 \quad \text{for all } i \neq 0$$

and $\psi = e_0 \otimes f_0 = \Omega \otimes f_0$. But $V\psi = \psi$ yields $Uf_0 = f_0$, hence $f_0 = e_0 = \Omega$. Q.E.D.

Now we formulate the following theorem.

Theorem 3.3: Let $Y = (\mathcal{A}, \omega, \alpha_t)$ be a KMS system. Then

(i) if E_0 is one-dimensional, then Y is absolutely weakly asymptotic dispersive if and only if zero is the only discrete point of the spectrum of L.

(ii) if E_0 is one-dimensional and if the spectral measure of L is absolutely continuous except for the vector Ω, then Y is strongly asymptotic dispersive.

(iii) if Y is strongly asymptotic dispersive then the point spectrum of L is reduced to the point zero.

Proof: (i) By Theorem 3.1 absolutely weakly asymptotic dispersiveness is equivalent with absolutely weakly clustering. As in Ref. 7 this is equivalent with

$$(\psi \otimes \psi', F_0\phi \otimes \phi') = 0$$

for all $\psi, \psi' \in N(L)^\perp$; $\phi, \phi' \in H_\omega$, where F_0 is the orthogonal projection on the invariant subspace of $V_t = U_t \otimes U_t^*$ in $H_\omega \otimes H_\omega$ for all $t \in \mathbb{R}$;

(i) follows now from Lemma 3.2,

(ii) is immediate from Theorem 3.1 and Ref. 7,

(iii) by Theorem 3.1, Y is strongly asymptotic dispersive yields that for all $\Psi \in N(L)^\perp$ and $\phi \in H_\omega$

$$\lim_{|t| \to \infty} (\psi, U_t\phi) = 0.$$

Suppose that there exists a vector $\psi \in N(L)^\perp$ such that $U_t\psi = \exp(i\lambda t)\psi$ where $\lambda \neq 0$. Then $\lim_{|t| \to \infty}(\psi, \exp(i\lambda t)\psi)$ does not exist. Hence (iii) follows. Q.E.D.

4. LOCAL STRUCTURE AND OBSERVABLES AT INFINITY

In the algebra of observables $\mathcal{A}$ we introduce a local structure. For any bounded open set V in $\mathbb{R}^d$, define the local algebra $\mathcal{A}_V$ as the subalgebra of $\mathcal{A}$ generated by the real valued functions $\mathcal{G}$ on K satisfying $\mathcal{G}(x) = \mathcal{G}'(\{x\} \cap (V \times \mathbb{R}^d))$, where $\mathcal{G}'$ is a function with domain $\{\{x\} \cap (V \times \mathbb{R}^d) \mid x \in K\}$. Then of course $\mathcal{A} = \cup_{V \subset \mathbb{R}^d} \mathcal{A}_V$.

As usual[4] let Z be defined by

$$Z = \bigcap_V \pi(\tilde{\mathcal{A}}_V)'',$$

where

$$\tilde{\mathcal{A}}_V = \bigcup_{W \cap V = \phi} \mathcal{A}_W$$

and $\pi(\tilde{\mathcal{A}}_v)''$ is the von Neumann algebra generated by $\pi(\tilde{\mathcal{A}}_v)$. Analogously we define Z' by

$$Z' = \bigcap_V \pi(\mathcal{A}_V^c)''$$

where $\mathcal{A}_V^c = \{\mathcal{G} \in \mathcal{A} \mid \{\mathcal{G}, F\} = 0 \text{ for all } F \in \mathcal{A}_V\}$.

Lemma 4.1: With the notations of above, $\mathcal{A}_V^c = \tilde{\mathcal{A}}_V$.

Proof: For simplicity we give the proof for V a connected set, but the argument holds for a general open V.

Let ϕ be an element of $\mathcal{A}_V^c$. The subset of $(V \times \mathbb{R}^d)^n$ defined by the condition $q_i \neq q_j$ for $i \neq j$ is not connected. Let D be any connected component which is open.

Let $x \in K$ such that $x \cap (V \times \mathbb{R}^d) = \phi$ and let F_x be the complex-valued function on D defined by

$$F_x(q_1, p_1, \ldots, q_n, p_n) = \phi(\{(q_1, p_1), \ldots, (q_n p_n)\} \cup x).$$

Take any $((q_1, p_1), \ldots, (q_n p_n)) \in D$; there are open subsets V_1, V_2 of $\mathbb{R}^d$ such that $q_1 \in V_1$, $q_i \notin V_1$ for $i \neq 1$ and $V_1 \subset V$, also $p_1 \in V_2$. Let g be a real C^∞ function with support in $V_1 \times V_2$.

Choose ψ a bounded C^∞ function from $\mathbb{R}$ to $\mathbb{R}$ such that $\psi'(g(q_1, p_1)) = 1$ where ψ' is the derivative of ψ and define $\mathcal{G}$ an element of $\mathcal{A}_V$ by

$$\mathcal{G}(y) = \psi\left(\sum_i g(y_i)\right)$$

for all $y \in K$. Since $\mathcal{G} \in \mathcal{A}_V$ we have $\{\phi, \mathcal{G}\} = 0$. Hence

$$0 = \{\phi, \mathcal{G}\}(\{(q_1, p_1), \ldots, (q_n, p_n)\} \cup x)$$

$$= \psi'(g(q_1 p_1))\left[\frac{\partial F_x}{\partial q_1}((q_1, p_1), \ldots, (q_n, p_n))\frac{\partial g}{\partial p_1}(q_1 p_1) - \frac{\partial F_x}{\partial p_1}((q_1 p_1), \ldots, (q_n p_n))\frac{\partial g(q_1 p_1)}{\partial q_1}\right].$$

If g is chosen such that $\partial g(q_1, p_1)/\partial q_1 = 0$ and $\partial g(q_1, p_1)/\partial_{p_1 k} = \delta_{kl}$ for $l = 1, \ldots, d$ in turn, then

$$\frac{\partial F_x}{\partial q_1}((q_1 p_1), \ldots, (q_n p_n)) = 0.$$

Similarly

$$\frac{\partial F_x}{\partial p_1}((q_1 p_1), \ldots, (q_n p_n)) = 0.$$

By repeating the argument $\partial F_x/\partial q_i = \partial F_x/\partial p_i = 0$ on D, $\forall i,\ i = 1, \ldots, n$.

Therefore, F_x is constant on D and letting the p_i's tend to infinity, gives $F_x \equiv \phi(x)$. Hence for any $x \in K$,

$$\phi(x) = \phi(x \cap C(V) \times \mathbf{R}^d)$$

and by continuity

$$\phi(x) = \phi(x \cap C(\bar{V}) \times \mathbf{R}^d).$$

Since $\phi \in \mathcal{A}$ there is a $V' \subset \mathbf{R}^d$ such that $\phi(x) = \phi'(x \cap V' \times \mathbf{R}^d)$ for some function ϕ'. Therefore,

$$\begin{aligned}\phi(x) &= \phi(x \cap C(\bar{V}) \times \mathbf{R}^d) \\ &= \phi'(x \cap (V' \cap C(\bar{V})) \times \mathbf{R}^d)\end{aligned}$$

and

$$\phi \in \mathcal{A}_{C(\bar{V}) \cap V'} \subset \tilde{\mathcal{A}}_V. \qquad \text{Q.E.D.}$$

Theorem 4.2: With the notations of above, $Z = Z'$.

Proof: Immediate from the definitions and Lemma 4.2. Q.E.D.

Finally we prove in the following theorem that, as in the quantum case, the set of observables at infinity consists of constants of the motion.

Theorem 4.3: If $Y = (\mathcal{A}, \omega, \alpha_t)$ is a KMS system and if

(i) the set $\{X \in Z \mid X^* \Omega_\omega \in D(\mathcal{L})\}$ is weakly dense in Z,

(ii) the set $\mathcal{E} \cap \mathcal{A}_V$ is dense in $\mathcal{A}_V$,

then $U_t X U_{-t} = X$ for all $X \in Z$.

Proof: By condition (i) it is sufficient to prove that $U_t X U_{-t} = X$ for all $X \in Z$ such that

$$X^* \Omega_\omega \in D(\mathcal{L}).$$

For any f and g in $\mathcal{A}$

$$\begin{aligned}&\frac{d}{dt}(\pi(g)\Omega_\omega, U_t X U_{-t} \pi(f)\Omega_\omega) \\ &\quad = \frac{d}{dt}(\Omega_\omega, U_t X U_{-t} \pi(\bar{g}f)\Omega_\omega) \\ &\quad = -i(\mathcal{L} X^* \Omega_\omega, \pi(\alpha_{-t}(\bar{g}f))\Omega_\omega).\end{aligned}$$

As $\mathcal{E}$ is dense in $\mathcal{A}$, for each $\epsilon > 0$ there exists an element $k \in \mathcal{E}$ such that

$$\left|\frac{d}{dt}(\pi(g)\Omega_\omega, U_t X U_{-t} \pi(f)\Omega_\omega)\right| \leq |(\mathcal{L} X^* \Omega_\omega, \pi(k)\Omega_\omega)| + \epsilon.$$

Since $k \in \mathcal{E}$ there is a V such that $k \in \mathcal{A}(V)$. As $X \in \pi(\mathcal{A}_V)'$, by condition (ii) there exists an element l in $\mathcal{E} \cap \mathcal{A}_V$ such that

$$|(\mathcal{L} X^* \Omega_\omega, \pi(k)\Omega_\omega)| < |(\Omega_\omega, \pi(l)\mathcal{L}\pi(k)\Omega_\omega)| + \epsilon.$$

Hence

$$\begin{aligned}&\left|\frac{d}{dt}(\pi(g)\Omega_\omega, U_t X U_{-t} \pi(f)\Omega_\omega)\right| \\ &\quad \leq 2\epsilon + |(\Omega_\omega, \pi(l)\mathcal{L}\pi(k)\Omega_\omega)| \\ &\quad = 2\epsilon + \left|\frac{d}{d\tau}\omega(l\,\alpha_\tau(k))_{|\tau=0}\right| \\ &\quad = 2\epsilon + |\omega(\{l, k\})|\end{aligned}$$

by using the KMS condition at $\tau = 0$. Now $\omega(\{l, k\}) = 0$ by Lemma 4.1 and the theorem follows. Q.E.D.

We note that the proof of Theorem 4.3 is performed under the weaker KMS condition, namely the condition at $t = 0$ or the so-called static KMS condition.

As in Ref. 4 for lattice systems, it can also be proved that Z consisting of the multiples of the identity is equivalent to uniform clustering or, expressed otherwise, is equivalent with short range correlations.

ACKNOWLEDGMENTS

The authors thank Professor H. Borchers and Professor R. Weder for useful remarks. Also, one of us (A. V.) thanks Professor G. Gallavotti and Professor A. Van Daele for stimulating discussions.

*Navorsingsstagiair NFWO, Belgium.
†On leave of absence from the Royal University of Malta.
‡Postal address: Instituut voor Theoretische Fysica, K.U. Leuven, Celestijnenlaan 200 D, B-3030 Heverlee.

[1]G. Gallavotti and E. Verboven, Nuovo Cimento 28, 274 (1975).
[2]G. Gallavotti and M. Pulvirenti, Commun. Math. Phys. 46, 1 (1976).
[3]M. Aizenman, G. Gallavotti, S. Goldstein, and J. L. Lebowitz, "Stability and Equilibrium States of Infinite Classical Systems," preprint.
[4]O.E. Lanford III and D. Ruelle, Commun. Math. Phys. 13, 194 (1969).
[5]R. Godement, Trans. Amer. Math. Soc. 63, 1 (1948).
[6]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[7]H.J.F. Knops, thesis, Nijmegen, Holland (1969).

Quantum system subject to random pulses

Henryk Gzyl

Escuela de Física y Matemáticas, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela
(Received 11 November 1976)

We "solve" the Schrödinger equation for a system subject to random pulses, and show how to compute ensemble averages of observables or their Laplace transforms.

1. INTRODUCTION

In this paper we show how to integrate the Schrödinger equation of a system subject to random pulses, i.e., a system described by a Schrödinger equation of the form

$$id\psi_t = H\psi_t\,dt + dV_t\psi_{t-}, \quad \hbar = 1, \tag{1.1}$$

where H is a time independent Hamiltonian and

$$dV_t = \sum_1^\infty V_k \epsilon_{T_k}(dt). \tag{1.2}$$

The operators V_k are assumed to be random variables whose values are Hermitian operators (see Bharucha-Reid[1]) for definitions of such notions). We shall assume the V_k's to be independent and identically distributed. We shall assume the T_k's to be the arrival times of a Poisson process with parameter λ, the T_k's are also assumed to be independent of the V_k's. Also $\epsilon_{T_k}(dt)$ is usually written as $\delta(t - T_k)dt$.

An equation like (1.1) can be useful for modeling the effect of random environments on a system. It has already been used by Blume and Clausser [see Eq. (1) in Ref. 2] to describe changes in line shape, but they do not integrate it correctly. For more applications see Chap. XI of Ref. 3. In the next section we show how to integrate (1.1) and how to obtain averaged expected values of observables and their Laplace transforms.

Let us briefly describe what is behind the calculations leading to (2.4) below. In probability theory (see Ref. 1) random variables are thought as functions on a triple $(\Omega, \mathcal{J}, P)$, where Ω is called the sample space, $\mathcal{J}$ is a σ-algebra of subsets of Ω whose elements are called events and correspond to the questions we can ask about the system, and P is a measure on $\mathcal{J}$ such that $P(\Omega) = 1$.

Also, integrals with respect to P are denoted by E, i.e., $\int H\,dP = E(H)$, and are called *mean (or averaged) values*.

The following facts follow from our assumptions on the V_k's and the T_k's.

If we denote by $\mathcal{H}$ the Hilbert space associated with our system and if $\psi_0, \ldots, \psi_n$ are arbitrary unit vectors in $\mathcal{H}$, then for any $k_1 < k_2 < \cdots < k_n$

$$E\{\langle\psi_n | V_{k_n} | \psi_{n-1}\rangle \cdots \langle\psi_1 | V_{k_1} | \psi_0\rangle\}$$
$$= \langle\psi_n | EV_{k_n} | \psi_{n-1}\rangle \cdots \langle\psi_1 | EV_1 | \psi_0\rangle. \tag{1.3}$$

Also, $P(T_{k+1} - T_k > t) = \exp(-\lambda t)$ and for any complex measurable function $f(t_1, \ldots, t_n)$, the quantity $E\{f(T_1, \ldots, T_n);\ T_n \le t < T_{n+1}\}$ is interpreted as the average value of $f(T_1, \ldots, T_n)$ over the set $\{T_n \le t < T_{n+1}\}$, and the average value can be easily seen to be given by

$$\lambda^n \exp(-\lambda t) \int_0^t dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1 f(t_1, \ldots, t_n). \tag{1.4}$$

2. INTEGRATION OF (1.1)

Notice that dV_t does not charge (T_k, T_{k+1}), i.e., $\int_{(T_k, T_{k+1})} dVt = 0$, which implies that if $T_k \le t < T_{k+1}$ then

$$\psi(t) = \exp[-iH(t - T_k)\psi(T_k +)].$$

From Eq. (1.2) it follows that $\psi_{t+} = \lim_{s \downarrow t} \psi_s = \psi_t$ for all t. Now, integrating (1.1) over $(T_k - \epsilon, T_k + \epsilon)$ and letting $\epsilon \to 0$, we obtain

$$\psi(T_k) - \psi(T_k -) = -iV_k\psi(T_k -).$$

From this we see the reason for the ψ_{t-} in (1.1), namely, the instantaneous change in the state of the system, caused by V_k at $t = T_k$ depends on the state of the system right before T_k, i.e., on $\psi(T_k -)$.

From $\psi(t) = \exp[iH(t - T_k)\psi(T_k)]$ on $T_k \le t < T_{k+1}$ we obtain, by letting $t \to T_{k+1}$, $\psi(T_{k+1} -) = \exp[-iH(T_{k+1} - T_k)\psi(T_k)]$. We can conclude that if $T_k \le t < T_{k+1}$,

$$\psi(t) = \exp[-iH(t - T_k)](1 - iV_k)\exp[-iH(T_k - T_{k-1})] \times (1 - iV_{k-1}) \cdots (1 - iV_1)\exp(-iHT_1)\psi_0,$$

and, introducing $T_0 = 0$, $I_{[T_k, T_{k+1})}$ the indicator function of the interval $[T_k, T_{k+1})$, we can write

$$|\psi(t)\rangle = \sum_{k=0}^\infty \exp[-iH(t - T_k)](1 - iV_k) \cdots (1 - iV_1) \times \exp(-iHT)|\psi_0\rangle I_{[T_k, T_{k+1})}(t). \tag{2.1}$$

This solution differs considerably from Eq. (2) in Ref. 2.

We just saw that right after a pulse the state of the system is $\psi(T_k) = (1 - iV_k)\psi(T_k -)$. Therefore, $\langle\psi(T_k)|\psi(T_k)\rangle = \langle\psi(T_k -)|(1 + V_k^2)\psi(T_k -)\rangle$, i.e., the evolution in the presence of pulses does not conserve probability. This has nothing to do with randomness, and it is entirely due to the nature of the time dependence of the perturbation. We can attribute the nonconservation of probability to the fact that a system in a time dependent environment is only part of a larger system. It may also be a defect of the model.

At this point we should mention that if H were time dependent, we would replace $\exp[-iH(t-T_k)]$ by

$$U(t,T_k)=1+\sum_{n=1}^{\infty}(-1)^n\int_{T_k}^{t}H(t_n)\,dt_n \times\int_{T_k}^{t_n}H(t_{n-1})\,dt_{n-1}\cdots\int_{T_k}^{t_2}H(t_1)\,dt_1$$

and analogously $\exp[-iH(T_k-T_{k-1})]$ by $U(T_k,T_{k-1})$.

Now, $\psi(t)$ is a random variable defined on $(\Omega,\mathcal{F},P)$ and if A represents an observable, $\langle\psi_t|A|\psi_t\rangle$ is also a random variable. By taking a basis for H it is very easy to see, exchanging sums with integrals, that the mean value of $\langle\psi_t|A|\psi_t\rangle$ satisfies

$$E\langle\psi_t|A|\psi_t\rangle=\mathrm{tr}(AE|\psi_t\rangle\langle\psi_t|), \tag{2.2}$$

which enables us to compute average expected valves of observables associated with the system. We shall mention that the appearance of expressions like (2.2) does not imply that we are working in a density-matrix formalism; it is just a randomized Schrödinger formalism.

Now we will imitate Blume and Clausser in their way of computing expressions like (2.2).

Let us put $H^x(A)=[H,A]$ and $\hat{W}(A)=WAW^*$, where A and W are operators on H, and let us put $W_k=(1-iV_k)$. From (2.1) and using the fact that $I_{[T_k,T_{k+1})}I_{[T_e,T_{e+1})}=\delta_{ke}I_{[T_k,T_{k+1}]}$ we can obtain (2.3)

$$|\psi_t\rangle\langle\psi_t|=\sum_{0}^{\infty}\exp[-iH^x(t-T_k)]\hat{W}_k\cdots\hat{W}_1\exp(-iH^xT_1)\times|\psi_0\rangle\langle\psi_0|I_{[T_{k_1}T_{k+1}^{(t)})}. \tag{2.3}$$

We note that each operator to the left of $|\psi_0\rangle\langle\psi_0|$ in (2.3) is applied to whatever is on its right side. Also taking into account the fact that the W_k's are independent, identically distributed and independent of the T_k's, putting $-iH^x=L$ and with (1.3) and (1.4), we can write for $E|\psi_{-t}\rangle\langle\psi_t|$

$$\sum_{k=0}^{\infty}\lambda^k\exp(-\lambda t)\int_0^t-\exp[L(t-t_k)]\langle\hat{W}\rangle\int_0^{t_k}\exp[L(t_k-t_k)]\langle\hat{W}\rangle\times\cdots\langle\hat{W}\rangle\int_0^{t_2}\exp(Lt_1)\,dt_1\cdots dt_k|\,\psi_0\rangle\langle\psi_0|, \tag{2.4}$$

where $\langle\hat{W}\rangle(A)=EW_kAW_k^*$, $k=1,2,\cdots$, for any operator A.

An equally bad-looking expression may be obtained by taking the Laplace transform of (1.4). Let s be any complex number with $\mathrm{Re}\,s>0$. Then it is easy to see that

$$\int_0^{\infty}\exp(-st)E|\psi_t|\,dt=\frac{1}{s+\lambda-L}\sum_{k=0}^{\infty}\left\langle W\right\rangle\frac{1}{s+\lambda-L}\Bigg)^k|\psi_0\rangle\langle\psi_0|. \tag{2.5}$$

To see it, consider the following identities:

$$\lambda^k\int_0^{\infty}\exp[-(\lambda+s)t]\,dt\int_0^t\exp[L(t-t_k)]\,dt_k\langle\hat{W}\rangle\int_0^{t_k}\times\cdots\langle\hat{W}\rangle\int_0^{t_2}\exp(Lt_1)\,dt_1$$

$$=\lambda\int_0^{\infty}d\frac{-\exp[-(\lambda+\rho-L)t]}{\lambda+\rho-L}\int_0^t\exp(Lt_k)dt_k\langle\hat{W}\rangle\times\int_0^{t_k}\exp[L(t_k-t_{k-1})]\cdots\langle\hat{W}\rangle\int_0^{t_2}\exp(Lt_1)\,dt_1$$

$$=\lambda^k\frac{1}{\lambda+s-L}\int_0^{\infty}\exp[-(\lambda+s)t]dt\langle\hat{W}\rangle\times\int_0^{t}\exp[L(t-t_{k-1})]\,dt_{k-1}\langle\hat{W}\rangle\times\cdots W\int_0^{t_2}\exp(Lt_1)\,dt_1=\cdots$$

$$=\lambda^k\frac{1}{\lambda+s-L}\langle\hat{W}\rangle\frac{1}{\lambda+s-L}\langle\hat{W}\rangle\cdots\langle\hat{W}\rangle\frac{1}{\lambda+s-L}$$

$$=\frac{1}{\lambda+s-L}\left\langle\hat{W}\right\rangle\frac{\lambda}{\lambda+s-L}\Bigg)^k.$$

If in (2.5) we substitute L by $-iH^x$, we obtain

$$\int_0^{\infty}\exp(-st)E|\psi_t\rangle\langle\psi_t|\,dt=\frac{1}{s+\lambda+iH^x}\sum_{k=0}^{\infty}\left(\langle\hat{W}\rangle\frac{1}{s+\lambda+iH^x}\right)^k|\psi_0\rangle\langle\psi_0|, \tag{2.6}$$

which may be used in perturbative calculations, as can be done with (2.4).

If we put $\int_0^{\infty}\exp(-st)E\int\psi_t\rangle\langle\psi_t|\,dt=R(s)|\psi_0\rangle\langle\psi_0|$, then Eq. (2.6) can be rewritten as

$$R(s)|\psi_0\rangle\langle\psi_0|=\left(\frac{1}{s+\lambda+iH^x}-\frac{1}{s+\lambda+iH^x}(-\lambda\langle\hat{W}\rangle)R(s)\right)|\psi_0\rangle\langle\psi_0| \tag{2.7}$$

or even better as

$$R(s)|\psi_0\rangle\langle\psi_0|=\left(\frac{1}{s+iH^x}-\frac{\lambda}{s+iH^x}(1-\langle\hat{W}\rangle)R(s)\right)|\psi_0\rangle\langle\psi_0|$$

and if we take into account the definition of $\langle\hat{W}\rangle$ it is easy to see that

$$R(s)|\psi_0\rangle\langle\psi_0|=\left(\frac{1}{s+iH^x}(i\langle V^x\rangle-\langle\tilde{V}\rangle)R(s)\right)|\psi_0\rangle\langle\psi_0|,$$

where, for any operator A, $\langle V^x\rangle(A)=E([V_k,A])$ and $\langle\tilde{V}\rangle(A)=E(V_kAV_k^*)$ (which do not depend on k since the V_k's have the same distribution). From the last identity for $R(s)$ if follows, after a symbolic calculation, that

$$R(s)=\{s+i(H^x+\lambda\langle V^x\rangle)-\lambda\langle\tilde{V}\rangle\}^{-1} \tag{2.8}$$

Now, we can invert the Laplace transform and obtain

$$E|\psi_t\rangle\langle\psi_t|=\{\exp[-i(H^x+\lambda\langle V^x\rangle)+i\lambda\langle\tilde{V}\rangle]t\}|\psi_0\rangle\langle\psi_0|. \tag{2.9}$$

At this point we should mention that similar symbolic calculation would yield

$$E|\psi_t\rangle=\{\exp[-i(H+\lambda\langle V\rangle)t]\}|\psi_0\rangle. \tag{2.10}$$

Comparing (2.10) with (2.9) we see that even if "average states" evolve according to an "average Hamiltonian" without changing their norm, there is a "dissipative-like effect" present in the time evolution of the expected values of the observables. This effect is *not statistical in nature*, but it is due only to the fact that the perturbation is instantaneous. Thus even if an

observable A commutes with H and each of the V_k's, its average expected value will change in time according to $dE\langle\psi_t|A|\psi_t\rangle\,dt = \mathrm{tr}(\langle\hat{V}\rangle AE|\psi_t\rangle\langle\psi_t|)$.

3. EXAMPLES

In his section we will consider two simple situations and we shall develop more advanced and interesting problems in forthcoming work.

Case 1

Consider a particle moving freely and assume that at instants T_k a constant random potential taking values V^1 with probability p and V^2 with probability $q = 1-p$ is "shut on." Since each value of V is constant, it is easy to see that $|\psi_t\rangle = \prod_{k=1}^{N_t}(1 - iV_k)[\exp(-iHt)]|\psi_0\rangle$ and $|\psi_t\rangle\langle\psi_t| = \prod_{k=1}^{N_t}(1+V_k^2)\exp(-iHt)|\psi_0\rangle\langle\psi_0|\exp(-iHt)$, where H is the free Hamiltonian and N_t is the number of pulses up to t, from which it follows that

$$E|\psi_t\rangle\langle\psi_t| = \exp(\lambda t\langle V^2\rangle)\exp(-iHt)|\psi_0\rangle\langle\psi_0|\exp(iHt), \tag{3.1}$$

where $\langle V^2\rangle = pV_1^2 + qV_2^2$ is a multiplicative operator. If A denotes an observable, we will have that

$$E\langle\psi_t|A|\psi_t\rangle = \exp(\lambda\langle V^2\rangle)\langle\psi_f(t)|A|\psi_f(t)\rangle,$$

where $|\psi_f(t)\rangle = [\exp(-iHt)]|\psi_0\rangle$. If $H|\psi_0\rangle = \epsilon_0|\psi_0\rangle$, this last identity yields $E\langle\psi_t|H|\psi_t\rangle = \epsilon_0\exp(\lambda t\langle V^2\rangle)$.

This example shows that if the potential is changed instantaneously at least once, not only the phase of the state vector changes but also its modulus.

Case 2

Let us consider now a system with two states whose Hamiltonian is given by $\begin{pmatrix}\epsilon_1 & 0\\ 0 & \epsilon_2\end{pmatrix}$ and has eigenstates $|1\rangle = \begin{pmatrix}1\\0\end{pmatrix}$, $|2\rangle = \begin{pmatrix}0\\1\end{pmatrix}$. Let us assume that each of the V_k's equals $\alpha\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$ with probability p and $\beta\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ with probability $q = 1 - p$. In this case $R(s) = \{s + i(H^x + \lambda\langle V^x\rangle) - \lambda\langle\hat{V}\rangle\}^{-1}$ is an operator acting on observables A which we shall identify with the $\sum_{ij}a_{ij}|i\rangle\langle j|$.

Let us put $F(s) = s + i(H + \lambda\langle V^x\rangle) - \lambda\langle\hat{V}\rangle$. Then $F(s)|i\rangle\langle j| = \sum_{(n,m)}f^{(i,j)}_{(n,m)}|n\rangle\langle m|$, and we can obtain $R(s)$ by inverting the matrix corresponding to $F(s)$. From the definitions it follows rather easily that the f^{ij}_{nm} are given by the following table:

	(1, 1)	(1, 2)	(2, 1)	(2, 2)
(1, 1)	$s - \lambda q\beta^2$	$-i\lambda p\alpha$	$i\lambda p\alpha$	$-\lambda p\alpha^2$
(1, 2)	$-i\lambda p\alpha$	$s + i\Delta\epsilon - \lambda q\beta^2$	$-\lambda p\alpha^2$	$i\lambda p\alpha$
(2, 1)	$i\lambda p\alpha$	$-\lambda p\alpha^2$	$s - i\Delta\epsilon - \lambda q\beta^2$	$-i\lambda p\alpha$
(2, 2)	$-\lambda p\alpha^2$	$i\lambda p\alpha$	$-i\lambda p\alpha$	$s - \lambda q\beta^2$

where $\Delta = \epsilon_1 - \epsilon_2$. Standard techniques would yield $R(s)$.

If we put $\alpha_t(1\to 2) = E|\langle 2|\psi_t\rangle|^2$ for $|\psi_0\rangle = |1\rangle$, and $\beta_t(2-1) = E|\langle 1|\psi_t\rangle|^2$ for $|\psi_0\rangle = |2\rangle$, then it is easy to see that $\int_0^\infty \exp(-st)\alpha_t(1\to 2)\,dt = \int_0^\infty \exp(-st)\beta_t(2\to 1)\,dt = -D^{1,1}_{2,2}/D$, where D is the determinant of $(f^{i,j}_{n,m})$ and $D^{1,1}_{2,2}$ is the minor of $F(s)$ at $((1,1),(2,2))$. Also it takes no effort to see that if $|\psi_0\rangle = |i\rangle$, $i = 1, 2$, then $E|\langle 1|\psi_t\rangle|^2 + E|\langle 2|\psi_t\rangle|^2 \neq 1$.

ACKNOWLEDGMENT

I want to thank the referee for his comments which resulted in an improvement in the presentation of the paper.

[1] A. T. Bharucha-Reid, *Random Integral Equations* (Academic, New York, 1972).

[2] M. Blume and M. J. Clausser, Phys. Rev. B 3, 583—91 (1971).

[3] C. V. Heer, *Statistical Mechanics, Kinetic Theory and Stochastic Processes* (Academic, New York, 1972).

Approximate symmetry groups of inhomogeneous metrics

Alan Spero*

Department of Physics, Wesleyan University, Middletown, Connecticut
and Department of Physics, Northern Michigan University, Marquette, Michigan

Ralph Baierlein

Department of Physics, Wesleyan University, Middletown, Connecticut
(Received 29 November 1976)

A useful step toward understanding inhomogeneous space–times would be to classify them, perhaps in a fashion analogous to that used for spatially homogeneous space–times. To that end, a technique for determining an approximate simply-transitive three-parameter symmetry group of a three-dimensional positive-definite Riemannian metric is developed. The technique employs a variational principle to find a set of three orthonormal vectors whose commutation coefficients are as close as possible to a set of structure constants. The Bianchi classification of the structure constants of three-parameter groups is then used to classify these inhomogeneous metrics. Application of this technique to perturbed homogeneous metrics is discussed in detail. We find that only four types of symmetry groups can be considered generic in the space of all perturbed homogeneous metrics.

1. INTRODUCTION

Much work has been done on the dynamics of spatially homogeneous but anisotropic cosmological models in the past ten years. In his pioneering work, Taub[1] applied a classification scheme of Bianchi to the symmetry groups of these models. The classes of models thus defined could then be studied using tetrads of vectors defined via the symmetry groups.[2] The study of these models was a first step in the investigation of more complicated cosmological solutions to the gravitational field equations.

Spatially inhomogeneous cosmological models have, on the other hand, resisted most attempts to deal with them. The work so far falls into three categories: (i) perturbation calculations; (ii) study of exact solutions (of which there are few); and (iii) singularity theorems. In comparison to spatially homogeneous cosmologies, almost nothing is known about the dynamics of these models.

These inhomogeneous cosmologies are a subject demanding attention at the present time, primarily due to Misner's chaotic cosmology.[4] This is the idea that the present state of the universe—its surprising isotropy and spectrum of inhomogeneities—would result no matter what initial conditions were chosen. Thus the problem of cosmogeny becomes irrelevant to cosmology.[5] Collins and Hawking[6] have shown, however, that in the space of all spatially homogeneous initial data, there is no open set which asymptotically approaches isotropy. This means that in the space of all initial data there can be no open set which approaches isotropy and intersects the space of spatially homogeneous initial data. They also show that the Robertson–Walker spaces are unstable to anistropic perturbations. These results are apparently damaging to Misner's conception. However, they do not exclude the possibility that there may be some type of coupling between anisotropic and inhomogeneous modes,[7] particularly on a large scale, which would lead to the isotropization of an initially anisotropic inhomogeneous universe. Furthermore, chaotic cosmology involves the generic behavior of models rather than their stability.[8] Therefore a more detailed understanding of inhomogeneous cosmological models is needed before chaotic cosmology can be properly judged.

We propose, as a first step toward dealing with spatially inhomogeneous models, to determine a metric that is the "best" spatially homogeneous approximant. The symmetry group of the approximant provides a means of classifying the inhomogeneous space. Indeed, the structure constants of the group are found during the first stage of a two-stage process of determining the metric approximant. Thus classification, as a practical matter, can proceed prior to full construction of the homogeneous approximant. Here our primary result is that only certain symmetries can be considered generic, a not unexpected conclusion.

Once the group has been found, we may construct a homogeneous metric that best approximates the original metric and has the group as its symmetry group. If the initial data are completed in a suitable way, the evolution of the original metric may be studied (i) by comparing that evolution with the evolution of the approximant and (ii) by studying the approximate group as a function of time.

Approximate symmetries have been discussed before. In particular, Matzner[9] proposed a way of finding vectors which were "almost Killing vectors" of a given metric. His method bears little connection to ours since we are interested in finding three vectors which give an approximate symmetry but together generate a simply transitive group.

In Sec. 2 we outline the standard classification of spatially homogeneous space—times and our generalization of it. In Sec. 3 our technique is formulated as a variational principle. Section 4 discusses the resulting differential equation and Sec. 5, the algebraic equations. The theorems stated in Sec. 5 make up the stability theorem which is discussed in Sec. 6. Our conclusions are in Sec. 7.

2. OUTLINE AND MOTIVATION: SPATIALLY HOMOGENEOUS MODELS

A space—time is said to be spatially homogeneous if

it is invariant under a three-parameter abstract Lie group[10] G_3 which acts simply transitively on a family of spacelike hypersurfaces $S(t)$. Given a space–time metric[11] ${}^4g_{\alpha\beta}$, one determines whether it is spatially homogeneous or not by solving for its Killing vectors (if any). If it is spatially homogeneous it will possess three Killing vector fields ξ^a_A which are everywhere nonzero and linearly independent and which span the $S(t)$. These vectors will generate a group, isomorphic to G_3, which is the symmetry group of $S(t)$.[12] We then choose the space–time coordinates so that the hypersurfaces of homogeneity, $S(t)$, are parameterized by the time, as indicated.

If ${}^4g_{\alpha\beta}$ is spatially homogeneous, we may choose a set of invariant orthonormal basis vectors $\{\mathbf{E}_A\}$ spanning $S(t)$ such that[15]

$$[\boldsymbol{\xi}_A, \mathbf{E}_B] \equiv \xi^a_A \partial_a \mathbf{E}_B - E^a_B \partial_a \xi_A = 0, \tag{2.1}$$

$$[\mathbf{E}_A, \mathbf{E}_B] = C^C_{AB} \mathbf{E}_C, \tag{2.2}$$

$$C^C_{AB} = C^C_{AB}(t), \tag{2.3}$$

$$g_{ab} E^a_A E^b_B = \delta_{AB}. \tag{2.4}$$

The triad vectors $\{\mathbf{E}_A\}$ generate the group reciprocal to the symmetry group. The structure constants C^C_{AB} satisfy

$$C^C_{AB} = -C^C_{BA} \tag{2.5}$$

and the Jacobi condition

$$C^D_{AB} C^C_{DE} + C^D_{BE} C^C_{DA} + C^D_{EA} C^C_{DB} = 0. \tag{2.6}$$

Spatially homogeneous space–times may be classified by classifying the possible[16] G_3. This was first done by Taub, using an algebraic classification of the C^C_{AB} due to Bianchi, and later modified and developed by Schücking, Kundt, Behr, and Ellis and MacCallum. One first decomposes the structure constants

$$C^C_{AB} = \epsilon_{ABF} n^{FC} + a_A \delta^C_B - a_B \delta^C_A, \tag{2.7}$$

$$n^{AB} = \tfrac{1}{2} C^{(A}_{GH} \epsilon^{B)GH}, \quad a_B = \tfrac{1}{2} C^D_{BD}. \tag{2.8}$$

Under position-independent rotations of the triad,

$$\mathbf{E}'_A = \mathcal{R}^B_A \mathbf{E}_B,$$

n^{AB} and a_B transform as a tensor density and a vector, respectively.[17] In terms of n^{AB} and a_B, the Jacobi conditions (2.6) reduce to

$$n^{AB} a_B = 0. \tag{2.9}$$

Defining

$$\mathrm{Tr}n \equiv n^A_A, N \equiv \tfrac{1}{2}[n^{AB} n_{AB} - (\mathrm{Tr}n)^2], \tag{2.10}$$

one can then classify the possible n^{AB} and a_B as in Table I. Since the G_3 acts on a family of nonintersecting hypersurfaces, the Bianchi type (BT) of a metric is preserved in time and the metric remains spatially homogeneous so long as the $S(t)$ are spacelike.

Collins and Hawking[6] have noted that each Bianchi type can be assigned a "dimension" determined by the minimum number of parameters needed to specify the most general (n^{AB}, a_B) for that type. These dimensions are given in Table I.

TABLE I. Classification and dimension of the groups G_3 as in Ellis and King.[37] Here $a^2 \equiv a^A a_A$. The group types VI_h and VII_h are one parameter families of groups. The parameter h is a space–time constant and $h^h = a^2/n_2 n_3$, where n_2 and n_3 are the two nonzero eigenvalues of n^{AB}; equivalently, $h = -a^2/N$.

Group class	Group type		Dimension
Class A $a^2 = 0$	$\det n^{AB} > 0$	IX: $\det n^{AB} > 0$	6
		VIII: $\det n^{AB} < 0$	6
	$\det n^{AB} = 0$	VI_0: $N > 0$	5
		VII_0: $N < 0$	5
		II: $N = 0$, $\mathrm{Tr}\, n > 0$	3
		I: $N = 0 = \mathrm{Tr}\, n$	0
Class B $a^2 \neq 0$	$\det n^{AB} = 0$	VI_h: $N > 0$	6
		VII_h: $N < 0$	6
		IV: $N = 0$, $\mathrm{Tr}\, n > 0$	5
		V: $N = \mathrm{Tr}\, n = 0$	3

Of course, if ${}^4g_{\alpha\beta}$ is not spatially homogeneous, then none of this can be carried out. The necessary Killing vectors do not exist and the $\mathbf{E}_A$ cannot be defined. Consider, however, metrics constructed by perturbing a spatially homogeneous metric. A set of triad vectors $\{\mathbf{e}_A\}$ orthonormal with respect to the perturbed metric and spanning the hypersurface t = constant may be formed by slightly changing the $\{\mathbf{E}_A\}$:

$$e^a_A = E^a_A + \delta e^a_A. \tag{2.11}$$

The commutation coefficients γ^C_{AB}, defined by

$$[\mathbf{e}_A, \mathbf{e}_B] = \gamma^C_{AB} \mathbf{e}_C, \tag{2.12}$$

which in general are functions of position and time, can be written (provided the first derivatives of the perturbation are not too large) as

$$\gamma^C_{AB} = C^C_{AB} + \delta\gamma^C_{AB}, \tag{2.13}$$

where C^C_{AB} are the structure constants of the unperturbed metric's symmetry group and $\delta\gamma^C_{AB}$ are functions of at most first order in the perturbation. Although the metric is inhomogeneous, we would like to say it is approximately spatially homogeneous on the basis of (2.13) and classify it by applying the Bianchi scheme to C^C_{AB} in (2.13). The technique for doing this will be discussed in detail in the following sections. Here we give only an outline of what is done.

We assume we are given a space–time metric ${}^4g_{\alpha\beta}$ and a spacelike hypersurface S in this space–time. The metric induces a positive-definite metric g_{ab} on S. Using a variational technique, we then search for an everywhere nonzero triad vector field[18] $\{\mathbf{e}_A\}$ which is orthonormal with respect to g_{ab} and whose commutation coefficients γ^C_{AB} are as close as possible to some structure constants C^C_{AB}. The metric g_{ab} is then classified by applying the Bianchi scheme to C^C_{AB} and is said to belong to a *generalized Bianchi type* (GBT). The simply-transitive three-parameter Lie group to which C^C_{AB} corresponds[19] is said to be the approximate symmetry group of g_{ab}, and $\{\mathbf{e}_A\}$ is said to be the "best fit" triad.

Once the structure constants of the approximate symmetry group have been determined, one can construct the metric that is the best spatially homogeneous approximant. That development would—logically—be placed directly after Sec. 5, but it is not central to the

primary focus of the paper and has been relegated to Appendix B.

We have chosen to search for triad vectors with nearly constant commutation coefficients rather than look for vectors which were almost Killing vectors[9] because the best fit triad vectors lead directly to a classification system. A second advantage to finding best fit triad vectors—rather than approximate Killing vectors—is that, once found, they may be used as a basis for expressing the Einstein equations. The geometric techniques using tetrads may then be employed.[3]

Our technique is a three-dimensional one and presently takes no account of the hypersurface's embedding in the space–time. Thus the approximate symmetry group of g_{ab} need not give approximate symmetries of the second fundamental form.[20] Furthermore, the results obtained using our technique are dependent upon the hypersurface chosen. For example, if ${}^4g_{\alpha\beta}$ were spatially homogeneous and if the given hypersurface were an orbit of a point under the G_3, then we would of course be able to choose a triad whose γ^C_{AB} would be exactly constant. However, if a hypersurface not a group orbit were chosen, a constant γ^C_{AB} could not necessarily be found and we might conclude that ${}^4g_{\alpha\beta}$ was not spatially homogeneous.

The extension of our classification scheme to space–times depends upon having a way of choosing families of hypersurfaces, since two different slicings of the same space–time may result in two different classifications. We will not deal with this problem here, restricting ourselves to three-dimensional considerations.

3. THE VARIATIONAL PRINCIPLE

Given a subset $\mathcal{U}$ of a hypersurface $\mathcal{S}$ with metric g_{ab}, we wish to find a triad of vector fields $\{\mathbf{e}_A\}$ in $\mathcal{U}$ such that

$$g_{ab}e^a_A e^b_B = \delta_{AB} \tag{3.1}$$

and the commutation coefficients

$$\gamma^C_{AB} \equiv g_{ab}[\mathbf{e}_A, \mathbf{e}_B]^a e^{bC} = 2e^a_{[A}\nabla_a e^b_{B]}e^C_b \tag{3.2}$$

are as close as possible to some set of structure constants C^C_{AB}. This is done by requiring that $\{\mathbf{e}_A\}$ and C^C_{AB} give a global minimum of I where[21]

$$I \equiv (1/V)\int_{\mathcal{U}} \Delta^C_{AB}\Delta^C_{AB}\,dV + 8\lambda_A n^{AB}a_B + \lambda_{[AB]}(n^{AB} - n^{BA}), \tag{3.3}$$

$$\Delta^C_{AB} \equiv \gamma^C_{AB} - C^C_{AB} = \gamma^C_{AB} - \epsilon_{ABF}n^{FC} - (a_A\delta^C_B - a_B\delta^C_A), \tag{3.4}$$

$$V \equiv \int_{\mathcal{U}} dV = \int_{\mathcal{U}} (\det g_{ab})^{1/2} dx^1\,dx^2\,dx^3. \tag{3.5}$$

The necessary conditions for a minimum are given by

$$\delta I = 0, \tag{3.6}$$

where the variation is with respect to the e^a_A [subject to the orthonormality constraint (3.1)], n^{AB}, a_B, λ_A, and $\lambda_{[AB]}$. The Lagrange multipliers λ_A and $\lambda_{[AB]}$ are necessary to ensure that the C^C_{AB} are structure constants, i.e., that they satisfy (2.5) and (2.6).[22]

For a homogeneous[23] g_{ab}, one gets $I=0$. Also note that if $\{\mathbf{e}_A\}$ gives the minimum value of I, then any triad related to this one by a position-independent rotation will also give the minimum value. This is consistent with the fact that $\{\mathbf{E}_A\}$ is determined only up to a constant roation when g_{ab} is homogeneous.

We would ideally like to choose $\mathcal{U}=\mathcal{S}$. When this choice gives a finite V, we encounter no problems. However, if an infinite V results, instead of choosing $\mathcal{U}=\mathcal{S}$ we must construct a sequence of sets $\mathcal{U}_i$ such that $\mathcal{U}_i \subset \mathcal{U}_{i+1}$ and $\lim_{i\to\infty}\mathcal{U}_i = \mathcal{S}$. We then perform a sequence of variational calculations, each one using a $\mathcal{U}_i$ and yielding the triad $\{\mathbf{e}_A\}_i$. The limit

$$\{\mathbf{e}_A\} = \lim_{i\to\infty}\{\mathbf{e}_A\}_i,$$

if it exists, is the triad we desire.

When $\mathcal{U}$ is chosen to be a proper subset of $\mathcal{S}$, then the group resulting from our variational calculation must be considered a locally approximate symmetry group. If $\mathcal{U}=\mathcal{S}$ then it is a globally approximate symmetry group. Difficulties may be encountered when $\mathcal{U}=\mathcal{S}$ since the topology of $\mathcal{S}$ may not admit certain groups.[24–27]

The variation with respect to the e^a_A must preserve the orthonormality condition (3.1). Hence the only variations permitted are position-dependent rigid rotations of the triad through some angle $\delta\theta^A$, given with respect to the unvaried triad,

$$\delta\mathbf{e}_A = \epsilon_{ABC}\mathbf{e}^B\delta\theta^C. \tag{3.7}$$

Carrying out the variations in (3.6) results in the following equations:

$$\delta\mathbf{e}_M\cdot\nabla\times[\mathbf{e}^A\Delta^C_{AB}]\epsilon_{BCM} + [\tfrac{1}{2}C^D_{AB}\epsilon_{DCM} + C^C_{AD}\epsilon_{DBM}]\Delta^C_{AB} = 0, \tag{3.8}$$

$$\delta n_{AB}: -4\bar{\eta}^{AB} + 4n^{AB} + 8\lambda^A a^B + 2\lambda^{[AB]} = 0, \tag{3.9}$$

$$\delta a^B: -8\bar{\alpha}_B + 8a_B + 8\lambda^A n_{AB} = 0, \tag{3.10}$$

$$\delta\lambda_A: n^{AB}a_B = 0, \tag{3.11}$$

$$\delta\lambda_{[AB]}: n^{AB} - n^{BA} = 0, \tag{3.12}$$

surface term:

$$(1/V)\int_{\partial\mathcal{U}}[\Delta^C_{AB}\epsilon_{BCM}\delta\theta^M\mathbf{e}^A]\cdot dS = 0, \tag{3.13}$$

for arbitrary $\delta\theta^M$.

$\bar{\eta}^{AB}$ and $\bar{\alpha}_B$ are defined by

$$\gamma^C_{AB} = \epsilon_{ABK}\eta^{CK} + \alpha_A\delta^C_B - \alpha_B\delta^C_A, \tag{3.14}$$

$$\bar{\eta}^{AB} \equiv (1/V)\int_{\mathcal{U}}\eta^{AB}\,dV, \quad \bar{\alpha}_B \equiv (1/V)\int_{\mathcal{U}}\alpha_B\,dV. \tag{3.15}$$

The surface term (3.13) appears because we have discarded a total divergence in deriving (3.8).

The symmetric part of $C^C_{AD}\epsilon_{DBM}$ in the differential equation (3.8) does not contribute due to the antisymmetry of Δ^C_{AB}. Therefore we rewrite (3.8) as

$$\nabla\cdot[\mathbf{e}^A\Delta^C_{AB}]\epsilon^{BCM} + \Lambda^{CM}_{AB}\Delta^C_{AB} = 0, \tag{3.16}$$

where

$$\Lambda^{CM}_{AB} \equiv \tfrac{1}{2}[C^D_{AB}\epsilon_{DCM} + C^C_{AD}\epsilon_{DBM} - C^C_{BD}\epsilon_{DAM}]. \tag{3.17}$$

This equation will be discussed in detail in the next section. The section following that will deal with the algebraic equations (3.9)–(3.12). Here we make a few comments about the surface integral, (3.13).

If $\mathcal{U} = \mathcal{S}$ and $\mathcal{S}$ is compact then of course (3.13) is identically satisfied. However if $\mathcal{U}$ has a boundary then (3.13) is an additional restriction on the solutions. In fact (3.13) then provides the natural boundary conditions for the differential equation.[28]

4. THE DIFFERENTIAL EQUATION

The differential equation (3.16) is a system of three coupled nonlinear partial differential equations for the three Euler angles which express the orientation of the triad $\{\mathbf{e}_A\}$ with respect to some fiducial triad.[29] The Euler angles result from satisfying the orthonormality constraint on $\{\mathbf{e}_A\}$. The structure constants C^C_{AB} enter (3.16) as undetermined parameters. Once (3.16) is solved for the $\{\mathbf{e}_A\}$, one fixes the values of the C^C_{AB}, or equivalently the pair (n^{AB}, a_B) by computing $\overline{\eta}^{AB}$ and $\overline{\alpha}_B$ from $\{\mathbf{e}_A\}$, and then solving the algebraic equations (3.9)–(3.12) for (n^{AB}, a_B). In general, $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ will depend upon (n^{AB}, a_B). The logic is illustrated in Fig. 1.

We would like to give a proof of the existence of solutions to the variational problem or the differential equation. However we have not been able to construct one. It does seem reasonable that solutions exist when g_{ab} is a perturbed homogeneous metric. This will be discussed later. In regard to the uniqueness of solutions, there is no reason to expect them to be unique. In fact one can easily find examples where they are not. Consider the case where g_{ab} is a homogeneous metric with a G_4 or G_6. Then the number of simply transitive three-parameter subgroups that may be constructed is in general greater than one.[30] Each of these subgroups will generate a triad that is a solution to $\delta I = 0$ and gives $I = 0$. Uniqueness of solutions when g_{ab} is inhomogeneous is an open question. Note, however, the following discussion.

When the metric g_{ab} differs from a homogeneous metric $\tilde{g}_{ab}$ with symmetry group $\tilde{G}_3$ by terms of first order in a small quantity ϵ, then (3.16) may be linearized. We will assume that the $\{\mathbf{e}_A\}$ and C^C_{AB} differ from the corresponding quantities in $(\tilde{g}_{ab}, \tilde{G}_3)$, i.e., $\{\tilde{\mathbf{E}}_A\}$ and $\tilde{C}^C_{AB}$, by terms of at most first order also. We write

$$g_{ab} = \tilde{g}_{ab} + \epsilon\delta g_{ab}, \tag{4.1}$$

$$\mathbf{e}_A = \tilde{\mathbf{E}}_A + \epsilon\delta\mathbf{e}_A + O(\epsilon^2), \tag{4.2}$$

where

$$\tilde{g}_{ab}\tilde{E}^a_A\tilde{E}^b_B = \delta_{AB}, \tag{4.3}$$

$$g_{ab}e^a_A e^b_B = \delta_{AB}. \tag{4.4}$$

Also

$$C^C_{AB} = \tilde{C}^C_{AB} + \epsilon\delta C^C_{AB} + O(\epsilon^2), \tag{4.5}$$

where

$$[\tilde{\mathbf{E}}_A, \tilde{\mathbf{E}}_B] = \tilde{C}^C_{AB}\tilde{\mathbf{E}}_C. \tag{4.6}$$

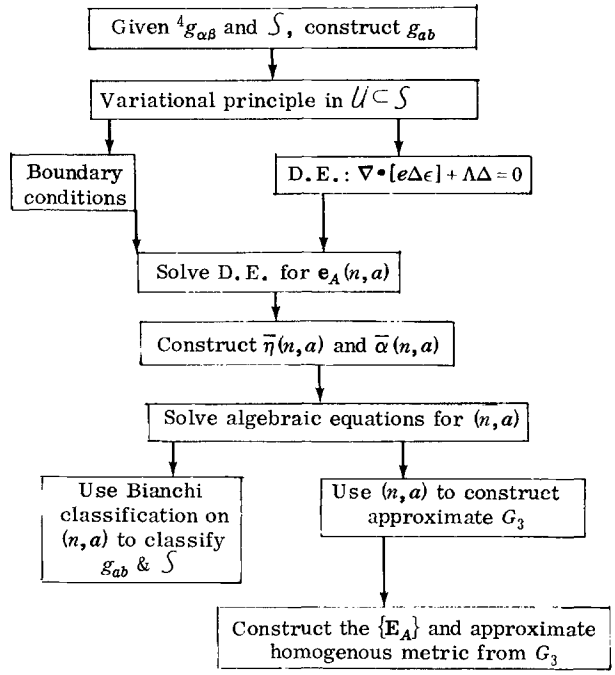


FIG. 1. A flow chart for the logic of the variational technique and classification scheme.

All calculations will be done to first order in ϵ. Coordinate indices are raised and lowered with $\tilde{g}_{ab}$.

Orthonormality, (4.4), to first order implies

$$\delta e^a_A = -\tfrac{1}{2}\tilde{g}^{ab}\delta g_{bc}\tilde{E}^c_A + \epsilon_{ABC}\tilde{E}^{aB}\delta\theta^C \tag{4.7}$$

with $\delta\theta^C$ arbitrary. Defining $\delta\gamma^C_{AB}$ by

$$\gamma^C_{AB} = \tilde{C}^C_{AB} + \epsilon\delta\gamma^C_{AB} + O(\epsilon^2), \tag{4.8}$$

we can calculate that

$$\begin{aligned}\delta\gamma^C_{AB} &= 2\tilde{E}^q_{[A}\epsilon_{B]CL}\tilde{\nabla}_a\delta\theta^L - 2\Lambda^{CL}_{AB}(\tilde{C})\delta\theta_L \\ &\quad - \tilde{g}^{ab}\delta g_{bd}\tilde{E}^d_{[A}\tilde{\nabla}_a\tilde{E}^f_{B]}\tilde{E}^C_f - (\tilde{\nabla}_a\delta g^b_d)\tilde{E}^a_{[A}\tilde{E}^d_{B]}\tilde{E}^C_b \\ &= 2E^a_{[A}\epsilon_{B]CL}\tilde{\nabla}_a\delta\theta^L - 2\Lambda^{CL}_{AB}(\tilde{C})\delta\theta_L \\ &\quad - E^a_{[A}\tilde{E}^b_{B]}(\delta g^d_b\tilde{\nabla}_d\tilde{E}^C_a + \tilde{E}^C_d\tilde{\nabla}_a\delta g^d_b).\end{aligned} \tag{4.9}$$

Since

$$\Delta^C_{AB} = \gamma^C_{AB} - C^C_{AB} = \epsilon\delta\gamma^C_{AB} - \epsilon\delta C^C_{AB} \equiv \epsilon\delta\Delta^C_{AB} \tag{4.10}$$

is of first order, the differential equation becomes

$$\tilde{\nabla}\cdot(\tilde{E}^A\delta\Delta^C_{AB})\epsilon^{BCM} + \Lambda^{CM}_{AB}(\tilde{C})\delta\Delta^C_{AB} = 0. \tag{4.11}$$

The second derivative terms in $\delta\theta^A$ in (4.11) have the form

$$(\delta_{AB}\tilde{\nabla}^2 + \tilde{E}^a_A\tilde{E}^b_B\tilde{\nabla}_a\tilde{\nabla}_b)\delta\theta^B, \tag{4.12}$$

showing that (4.11) is a strongly elliptic system of differential equations.[31] From the theory of strongly elliptic systems, strong solutions to (4.11) in a compact $\mathcal{U}$ can be shown to exist provided the operators are sufficiently continuous and satisfy a specific inequality. Since our results here are inconclusive, we refer the reader to Refs. 32 and 33.

The variational problem leading to (4.11) may be shown to have unique solutions when $\mathcal{U} = \mathcal{S}$. Varying the

$\delta\theta^A$ in

$$J \equiv (1/V)\int_U (\delta\Delta^C_{AB}[\delta\theta])^2 dV \tag{4.13}$$

and setting $\delta J = 0$ yields (4.11) and a linearized version of the boundary condition (3.13). Note that J is just I after (4.1)–(4.5) have been substituted into Δ^C_{AB} and terms of order ϵ^3 and higher have been dropped. The terms involving λ_A and $\lambda_{[AB]}$ in I have been dropped since they have no bearing on this argument.

Consider a variation $\delta\theta^A \to \delta\theta^A + \rho\phi^A$, with $0 < \rho < 1$. The new value of J is

$$\begin{aligned} J[\delta\theta^A + \rho\phi^A] = (1/V)&\int_U (\delta\Delta^C_{AB}[\delta\theta])^2 dV \\ &+ \frac{2\rho}{V}\int_U \delta\Delta^C_{AB}[\delta\theta]\,\delta\Delta^{*C}_{AB}[\phi]dV \\ &+ \frac{\rho^2}{V}\int_U (\delta\Delta^{*C}_{AB}[\phi])^2 dV, \end{aligned} \tag{4.14}$$

where

$$\delta\Delta^{*C}_{AB}[\phi] \equiv 2\tilde{E}_{[A}\epsilon_{B]CF}\partial_a\phi^F - 2\Lambda^{CF}_{AB}(\tilde{C})\phi_F. \tag{4.15}$$

The last term in (4.14) is proportional to the second variation. For those $\delta\phi^A$ which solve the differential equation (4.11) and boundary condition, the first variation vanishes and we have

$$J[\delta\theta + \rho\phi] \geq J[\delta\theta]. \tag{4.16}$$

Thus those $\delta\theta^A$ give a minimum or inflection point. To have an inflection point, the second variation must vanish, so

$$\delta\Delta^{*C}_{AB}[\phi] = 0. \tag{4.17}$$

This implies that

$$\begin{aligned} \mathbf{E}^C \circ \partial\phi^M &= -\epsilon^{MAB}\Lambda^{CF}_{AB}(\tilde{C})\phi_F \\ &= -[\epsilon_{FL(C}\tilde{n}_{M)}L + \delta_{F[C}\tilde{a}_{M]}]\phi^F. \end{aligned} \tag{4.18}$$

Regarding this equation as giving the change of ϕ^M along the integral curves of $\mathbf{E}_C$, we can show, by considering every Bianchi type to which $\tilde{C}^C_{AB}$ may belong, that ϕ^M must equal zero or a constant on the integral curves if it is required to be bounded, real, and single valued. Since the integral curves are space-filling, we conclude that the only variations which give inflection points are spatially constant changes in $\delta\theta^A$. This is not surprising since the variational integral is invariant under constant rotations of the triad.

Hence, for nonconstant variations ϕ^A about a stationary point, we have

$$J[\delta\theta + \rho\phi] > J[\delta\theta].$$

Thus every stationary point is a minimum and, given the continuity of J, there can therefore be only one minimum, i.e., the solutions to (4.11) are unique.

We have not shown that the full equation (3.16) has a unique solution if g_{ab} is a perturbed homogeneous metric. If $\tilde{g}_{ab}$ admits more than one $\tilde{G}_3$, then it is possible to construct one linearized equation for each $\tilde{G}_3$. Each one of these equations will have a unique solution (if a solution exists). The full equation will therefore have multiple solutions. However, it seems likely that in general only one solution will give the lowest value for I.

5. THE ALGEBRAIC EQUATIONS

In this section we discuss the behavior of the algebraic equations, (3.9)–(3.12), for n^{AB} and a_B, assuming that the pair $(\bar{\eta}^{AB}, \bar{\alpha}_B)$ are a given set of constants. Of course $(\bar{\eta}^{AB}, \bar{\alpha}_B)$ will in fact depend upon (n^{AB}, a_B) in some complicated fashion. The effects of this dependence will be considered at the end of Sec. 5. After some preliminary algebra and an example, we present a series of theorems (whose proofs are in Appendix A) which lead to a stability theorem stated in Sec. 6.

To begin, we note that (3.9) may be solved for $\lambda^{[AB]}$ to yield

$$\lambda^{[AB]} = -4\lambda^{[A}a^{B]}. \tag{5.1}$$

The remaining equations are

$$n^{AB} = n^{BA}, \tag{5.2}$$

$$n^{AB}a_B = 0, \tag{5.3}$$

$$a_B = \bar{\alpha}_B - n_{BC}\lambda^C, \tag{5.4}$$

$$n_{AB} = \bar{\eta}_{AB} - (\lambda_A a_B + a_A\lambda_B). \tag{5.5}$$

These immediately imply the two useful results[34]:

$$a^2 = a\circ\bar{\alpha}, \tag{5.6}$$

$$n^2 = n^{AB}\bar{\eta}_{AB}. \tag{5.7}$$

The role of λ_A is, of course, to ensure that (n^{AB}, a_B) satisfy the Jacobi condition, (5.3). Since we wish to obtain structure constants, (5.3) must be exactly satisfied, even if the differential equation was solved only in the linear approximation. We will deal with the algebraic equations without approximation in this section. In passing, we note that the effect of λ_A is to reduce the size of $n^2 + 2a^2$ from its maximum value of $\bar{\eta}^2 + 2\bar{\alpha}^2$. This can be seen by deriving from (5.2)–(5.5) the relations:

$$\bar{\eta}^2 - n^2 = 2\lambda\circ\bar{\eta}\circ a = 2\lambda^2 a^2 + 2(a\cdot\lambda)^2 \geq 0 \tag{5.8}$$

and

$$\bar{\alpha}^2 - a^2 = \lambda\cdot n\circ\bar{\alpha} = \lambda\cdot n\circ n\circ\lambda \geq 0. \tag{5.9}$$

Given $(\bar{\eta}^{AB}, \bar{\alpha}_B)$, the algebraic equations can in principle be solved as follows. Substitution of (5.5) into (5.4) and a subsequent contraction to find $a\circ\lambda$ yields

$$a_B(1-\lambda^2) = \bar{\alpha}_B - \bar{\eta}_{BC}\lambda^C + \zeta\lambda_B, \tag{5.10}$$

where

$$\zeta \equiv a\circ\lambda = (\bar{\alpha}\circ\lambda - \lambda\circ\bar{\eta}\circ\lambda)/(1-2\lambda^2). \tag{5.11}$$

We now have a_B in terms of $(\bar{\eta}^{AB}, \bar{\alpha}_B)$ and λ_A. Putting (5.10) into (5.5) gives n^{AB} in terms of these quantities.

We need now only solve for λ_A. The necessary equation, constructed using (5.3), (5.5), (5.6), and (5.10), is

$$\begin{aligned} \bar{\eta}_{AB}\bar{\alpha}^B = \lambda_A&[\bar{\alpha}^2 + \zeta^2 + \zeta\bar{\alpha}\circ\lambda - \bar{\alpha}\circ\bar{\eta}\circ\lambda] \\ &+ \zeta\bar{\alpha}_A - 2\zeta\bar{\eta}_{AB}\lambda^B + \bar{\eta}_{AB}\bar{\eta}^{BC}\lambda_C. \end{aligned} \tag{5.12}$$

This equation is not very helpful. It does suggest that there may be multiple solutions for λ_A and consequently for (n^{AB}, a_B). That is to say, given $(\bar{\eta}^{AB}, \bar{\alpha}_B)$, there may be several ways, consistent with (5.2)–(5.5), of form-

ing structure constants. The existence of this multiplicity is confirmed by an example. Consider the case when

$$\overline{\eta}_{11} \neq 0 \neq \overline{\alpha}_1$$

and the rest are zero. The two possible solutions are

$$\lambda_1 = \overline{\alpha}_1/\overline{\eta}_{11}, \quad \lambda_2 = \lambda_3 = a_B = 0, \quad n^{AB} = \overline{\eta}^{AB}$$

and

$$\lambda_1 = \overline{\eta}_{11}/2\overline{\alpha}_1, \quad \lambda_2 = \lambda_3 = n^{AB} = 0, \quad a_B = \overline{\alpha}_B.$$

The first solution would be classified GBT II and the second, GBT V.

When $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ satisfy the Jacobi condition, the algebraic equations give only one solution:

Theorem 1: $n^{AB} = \overline{\eta}^{AB}$ and $a_B = \overline{\alpha}_B$ if and only if $\overline{\eta}^{AB}\overline{\alpha}_B = 0$.

For many $(\overline{\eta}^{AB}, \overline{\alpha}_B)$, though, the algebraic equations will give multiple (n^{AB}, a_B), as can be seen from the following two theorems:

Theorem 2: There exists a Class A solution (with $n^{AB} = \overline{\eta}^{AB}$ and $a_B = 0$) if and only if there exists a λ_B such that $\overline{\eta}_{AB}\lambda^B = \overline{\alpha}_A$.

Theorem 3: A given $\overline{\eta}^{AB}$ and $\overline{\alpha}_B$ admit a Class B solution if and only if $\overline{\alpha}^2 \neq 0$.

When these troublesome multiple solutions do occur, we choose among them by picking the solution which gives the smallest value of I. The value of I for a particular solution is

$$I = (1/V)\int_U 2(\eta^2 + 2\alpha^2)dV - 2(n^2 + 2a^2). \tag{5.13}$$

From (5.13), (5.8), and (5.9) it is clear that I is minimized when (n^{AB}, a_B) are as close to $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ as possible. The size of λ^2 will also be as small as possible in that case.

When $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ differ only slightly from true structure constants, one expects to be able to determine the relative sizes of I for the various solutions. Therefore, consider the following.

If g_{ab} differs from a homogeneous metric $\overline{g}_{ab}$ by terms of at most first order in a small quantity ϵ, as assumed in the previous section, then we expect $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ to differ from $(\tilde{n}^{AB}, \tilde{a}_B)$ of the $\tilde{G}_3$ belonging to $\tilde{g}_{ab}$ by terms of at most first order. Hence we write

$$\overline{\eta}^{AB} = \tilde{n}^{AB} + \delta\overline{\eta}^{AB}, \tag{5.14a}$$

$$\overline{\alpha}_B = \tilde{a}_B + \delta\overline{\alpha}_B. \tag{5.14b}$$

All terms in $\overline{\eta}^{AB}$ and $\overline{\alpha}_B$ of order ϵ or higher have been included in $\delta\overline{\eta}^{AB}$ and $\delta\overline{\alpha}_B$. The perturbation is chosen so that

$$|\delta\overline{\eta}^{AB}| \text{ and } |\delta\overline{\alpha}_B| < \text{smallest, nonzero quantity in } \{\tilde{a}^2, |\text{eigenvalues of } \tilde{n}^{AB}|\}. \tag{5.15}$$

In cases where $\tilde{g}_{ab}$ and $\tilde{G}_3$ are BT VIII or BT IX, so that $\det\tilde{n}^{AB} \neq \tilde{a}^2$, we can explicitly construct a λ_B satisfying the conditions of Theorem 2. This λ_B gives the smallest value of I among the possible solutions and the resulting (n^{AB}, a_B) are of the same GBT as the original $\tilde{G}_3$.

Theorem 4: $\overline{\eta}^{AB}$ and $\overline{\alpha}_B$ admit a solution, n^{AB} and a_B, of GBT VIII or GBT IX if they arise from a perturbation of a homogeneous metric with a $\tilde{G}_3$ of BT VIII or BT IX, respectively. Furthermore, the solutions give global minima of I.

This type of result will not be true if we begin with a homogeneous metric whose $\tilde{G}_3$ belongs to one of the lower-dimensional Class A groups. For example, if $(\tilde{n}^{AB}, \tilde{a}_B)$ are BT II, so $\tilde{n}_{11} \neq 0 = \tilde{a}_A$ and the other $\tilde{n}_{AB}$ are zero, we can imagine a perturbation such that $\overline{\eta}_{AB} = \tilde{n}_{AB}$, $\overline{\alpha}_A = \overline{\alpha}_2\delta_{A2}$. In this case $\overline{\eta}^{AB}\overline{\alpha}_B = 0$ and Theorem 1 applies. The resulting G_3 is BT IV. A theorem slightly stronger than Theorem 4 is obtained if the homogeneous metric's $\tilde{G}_3$ belongs to Class B.

Theorem 5: $\overline{\eta}^{AB}$ and $\overline{\alpha}_B$ admit at least one solution, n^{AB} and a_B, of Class B if they arise from the perturbation of a homogeneous metric with a $\tilde{G}_3$ of Class B. The solution giving a global minimum of I will be in Class B. Furthermore, this solution will be GBT VII_h or GBT VI_h if the $\tilde{G}_3$ is BT VII_h or BT VI_h, respectively.

TABLE II. The behavior of the solutions to the algebraic equations is summarized here. The metric g_{ab} is assumed to arise from a perturbation of a homogeneous metric $\tilde{g}_{ab}$ with the symmetry group $\tilde{G}_3$.

	Class A						Class B			
BT of unperturbed $\tilde{G}_3$ of $\tilde{g}_{ab}$.	IX	VIII	VI_0	VII_0	II	I	VI_h	VII_h	IV	V
Do solutions (n, a) belonging to the same class always exist?	yes	yes	no	no	no	no Counter examples to existence can be found in these cases by taking $\delta\overline{\eta}^{AB} \neq 0$ & $\delta\overline{\alpha}^2 \neq 0$. If $\det\overline{\eta} \neq 0$ these become yes.	yes	yes	yes	yes
Does the solution giving the minimum I belong to the same class?	yes	yes	depends upon perturbation — cannot say in general				yes	yes	yes	yes
Do solutions of the same type exist?	yes	yes	—not in general				yes	yes	depends upon perturbation	
Does the solution giving the minimum I belong to the same type?	yes	yes	—not in general				yes	yes	depends upon perturbation	

Theorems 4 and 5 are the principal results of this section. Table II exhibits these results in graphic form for each Bianchi type.

Earlier we remarked that $(\bar{\eta}^{AB}, \bar{\alpha}_B)$ had a parametric dependence upon (n^{AB}, a_B). Symbolically, then

$$n^{AB} = n^{AB}(\bar{\eta}(n,a), \bar{\alpha}(n,a))$$

and similarly for a_B. Because of the manner in which they are stated, Theorems 1—3 are unaffected by this dependence. Theorems 4 and 5 could potentially be changed. However, the proofs of these last two theorems depend only upon (i) the zeroth order parts of $(\bar{\eta}^{AB}, \bar{\alpha}_B)$, i.e., $\tilde{n}^{AB}$ and $\tilde{a}_B$, which are independent of (n^{AB}, a_B), and (ii) the assumption that the remainders $\delta\bar{\eta}^{AB}(n,a)$ and $\delta\bar{\alpha}_B(n,a)$ are small. It is unlikely that consideration of the parametric dependence would invalidate this assumption. Therefore, we feel that a proper treatment of the equations (3.8)—(3.13) which takes all interrelations into account would not change the results in Theorems 4 and 5.

6. THEOREMS ON STABLE AND GENERIC SYMMETRIES

Together, Theorems 4 and 5 constitute a stability theorem for symmetry groups.

Theorem 6—*Stable symmetries*: The G_3 in Bianchi types VI_h, VII_h, VIII, and IX are stable symmetry groups. That is to say, a homogeneous metric with one of these symmetry group types will preserve this group type as its approximate symmetry group type when perturbed by a small but otherwise arbitrary metric perturbation. Groups belonging to the other types are unstable.

The instability of G_3 in the other group types follows from the existence of perturbations of a $\tilde{g}_{ab}$ with one of these G_3 which produces a $(\bar{\eta}^{AB}, \bar{\alpha}_B)$ yielding a group of some GBT other than the Bianchi type of the original one.

This suggests that the metrics of GBT's other than the four listed above are a set of measure zero in the space of all perturbed homogeneous metrics.

Theorem 7: If a homogeneous metric is subjected to a *general* perturbation, the approximate symmetry group of the resultant metric will belong to one of the four GBT's: VI_h, VII_h, VIII, or IX.

Proof: A general perturbation δg_{ab} is one in which all Fourier components are nonzero. If we assume that the differential equation (3.16) behaves in the expected fashion, this perturbation will result in the most general $(\bar{\eta}^{AB}, \bar{\alpha}_B)$, which require the maximum of nine parameters to be specified. If the resulting (n^{AB}, a_B) are Class A, then $n^{AB} = \bar{\eta}^{AB}$, since $a_B = 0$, and we must have GBT VIII or IX because $\det\bar{\eta} \neq 0$. If they are Class B, then axes may be chosen such that $a_A = \delta_{A1}a_1$ and $n_{AB} = \mathrm{diag}(0, n_{22}, n_{33})$. A contradiction arises if either n_{22} or n_{33} is zero, since then the maximum number of parameters available to specify $(\bar{\eta}^{AB}, \bar{\alpha}_B)$ is eight, three to give the orientation of the axes, a_1, n_{22} or n_{33}, and the three λ_A. Hence both n_{22} and n_{33} must be nonzero and the metric is GBT VI_h or VII_h.

If a perturbed metric is to be classified as a GBT other than the four above, the perturbation must be such that $\bar{\eta}^{AB}$ and $\bar{\alpha}_B$ satisfy certain constraints (if V is finite). Hence this cannot be a general perturbation and therefore these GBT must represent a lower-dimensional subspace of the space of all perturbed homogeneous metrics.

Since the proof of Theorem 6 does not depend upon the detailed behavior of $(\bar{\eta}^{AB}, \bar{\alpha}_B)$, we expect Theorem 6 to remain true after the dependence of $(\bar{\eta}^{AB}, \bar{\alpha}_B)$ upon (n^{AB}, a_B) is taken into account.

7. CONCLUSIONS AND COMMENTS

We have in a natural way generalized the notion of simply-transitive symmetry groups for three-dimensional Riemannian metrics by defining an approximate symmetry group for a given metric. These approximate symmetry groups may then be classified using the now-standard Bianchi classification.

We find that a metric may possess one of two kinds of inhomogeneity in a set $\mathcal{U}$ of finite volume V. Once the global minimum of I is found, the possibilities are:

(a) η^{AB} and α_B not constant but $\bar{\eta}^{AB}\bar{\alpha}_B = 0$. This gives $I > 0$, $\lambda^2 = 0$.

(b) η^{AB} and α_B not constant and $\bar{\eta}^{AB}\bar{\alpha}_B \neq 0$. This gives $I > 0$, $\lambda^2 > 0$.

I and λ^2 provide a measure of the metric's inhomogeneity since

$$\text{homogeneity} \iff I = 0 = \lambda^2,$$

$$\text{inhomogeneity} \iff I > 0.$$

Consider the subspace of superspace consisting of all homogeneous three-dimensional metrics in a compact subset $\mathcal{U}$ on a manifold $\mathcal{M}$.[35] This subspace is partitioned into sets, all of whose members have $\tilde{G}_3$ belonging to the same Bianchi type. These sets are not disjoint since some metrics may have more than one $\tilde{G}_3$. The subspace may be extended to the space of perturbed homogeneous metrics on $\mathcal{U}$ by using a modification of Hawking's[8] topology for Lorentz metrics. Define the neighborhood of a metric g_{ab} on $\mathcal{U}$ by

$$N(\epsilon_r, g_{ab}, \mathcal{U}) \equiv \{g'_{ab} \text{ on } \mathcal{U} \cap \mathcal{M} \mid \text{the } i\text{th derivatives } (0 \leq i \leq r) \text{ of } g'_{ab} \text{ differ from those of } g_{ab} \text{ by less than } \epsilon_i\}.$$

The derivatives are taken with respect to a Euclidean metric on $\mathcal{U}$. The metrics in $N(\epsilon_r, \tilde{g}_{ab}, \mathcal{U})$ can be considered as arising from perturbations of the homogeneous $\tilde{g}_{ab}$. The space of all perturbed $\tilde{g}_{ab}$ on $\mathcal{U}$ can be defined as the

$$\bigcup_{\text{all } \tilde{g}_{ab} \text{ on } \mathcal{U}} N(\epsilon_r, \tilde{g}_{ab}, \mathcal{U})$$

for some fixed ϵ_r.

The space of perturbed $\tilde{g}_{ab}$ can be partitioned according to approximate symmetry groups by using (n^{AB}, a_B) and may be further divided according to degree of inhomogeneity by using I. That is, we can define a neighborhood of the set of $\tilde{g}_{ab}$ on $\mathcal{U}$ with a $\tilde{G}_3$ of Bianchi

type Y by

$$B(\delta, Y, \mathcal{U}) = \{g_{ab} \text{ on } \mathcal{U} \subset \mathcal{M} \mid I[g_{ab}] < \delta \text{ and}$$

$$(n^{AB}, a_B)[g_{ab}] \text{ are of GBT Y}\},$$

where δ is some positive number greater than zero. Each member of the space of perturbed $\tilde{g}_{ab}$ (for some ϵ_r) will lie in one of the $B(\delta, Y, \mathcal{U})$ if δ is chosen sufficiently large.

If $\tilde{g}_{ab}$ is of BT Y and if, for some ϵ_r and δ, every $g_{ab} \epsilon N(\epsilon_r, \tilde{g}_{ab}, \mathcal{U})$ is also in $B(\delta, Y, \mathcal{U})$, then Y can be considered a stable symmetry type. The collection of stable symmetry types is generic in the space of perturbed $\tilde{g}_{ab}$.

We find that some group types do not belong to this generic set. Not unexpectedly, those types that do belong, i.e., Types VI_h, VII_h, VIII, and IX, are those of the highest dimension, as Collins and Hawking[6] anticipated in their study of the dynamics of homogeneous metrics.

It is the dynamical behavior of metrics with approximate symmetry groups belonging to these generic types that should be studied to determine, for example, the viability of chaotic cosmology. Initial data might be completed by requiring that the tensor χ_{ab} on $\mathcal{S}$, which gives the second fundamental form of $\mathcal{S}$, have the same approximate symmetry group[20] as g_{ab}. Some things are known. The dynamics of a $\tilde{g}_{ab}$ belonging to a generic type should preserve the classification for small increments in time since small changes in the metric can be regarded as a perturbation to which Theorems 4–6 apply. (These theorems also imply that the classification is stable to small changes in the choice of hypersurface $\mathcal{S}$.) Also, some cosmological metrics are known which are inhomogeneous and have timelike Killing vectors.[36] These will of course preserve their generalized Bianchi type throughout their evolution. However, much more work must be done before we can say that the behavior of even slightly inhomogeneous cosmological models is understood.

APPENDIX A: PROOFS OF THE THEOREMS IN SEC. 5

The proofs are in the order 2, 4, 3, 5, and 1.

Proof of Theorem 2: If λ_B satisfies $\overline{\eta}_{AB}\lambda^B = \overline{\alpha}_A$, then $\zeta = 0$ provided $\lambda^2 = \frac{1}{2}$ and therefore $a_B = 0$ provided $\lambda^2 \neq 1$. Equation (5.5) then gives $n^{AB} = \overline{\eta}^{AB}$. If $\lambda^2 = 1$, (5.10) does not determine a_B. The only restriction upon a_B is $\zeta = a^\circ \lambda = 0$. Hence $a_B = 0$ is a solution. If $\lambda^2 = \frac{1}{2}$, $a^\circ \lambda$ cannot be determined from

$$a_B(1 - \lambda^2) = \overline{\alpha}_B - \overline{\eta}_{BC}\lambda^C + (a^\circ \lambda)\lambda_B.$$

This implies that a_B satisfies

$$a_B = 2\lambda_B(a^\circ \lambda).$$

Again $a_B = 0$ is a solution.

If $a^2 = 0$, then (5.5) gives $n^{AB} = \overline{\eta}^{AB}$ and this together with (5.4) shows that $\overline{\eta}_{AB}\lambda^B = \overline{\alpha}_A$. QED

Proof of Theorem 4: If $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ satisfy (5.14a, b)–(5.15) with $\det\tilde{n}^{AB} \neq 0$, a λ_B such that $\overline{\eta}_{AB}\lambda^B = \overline{\alpha}_A$ can be constructed. Since

$$\det\overline{\eta}^{AB} \approx \det\tilde{n}^{AB} \neq 0,$$

the necessary λ_A is

$$\lambda_A = (\overline{\eta}^{-1})_{AB}\overline{\alpha}^B. \tag{A1}$$

By Theorem 2, then, a Class A solution must exist. This solution has $n^{AB} = \overline{\eta}^{AB}$ and therefore must be GBT IX or VIII. In fact, n^{AB} is of the same type as $\tilde{n}^{AB}$ since the perturbation cannot change the sign of the determinant.

The question remains whether these solutions give the smallest possible value of I. To show they do, we assume the contrary and obtain a contradiction. Since (A1) gives the only Class A solution, any other solution must belong to Class B. Assume that

$$I|_{\text{Class A}} > I|_{\text{Class B}}.$$

Then

$$(n^2 + 2a^2)|_{\text{Class A}} < (n^2 + 2a^2)|_{\text{Class B}}$$

and, invoking (5.8),

$$\overline{\eta}^2 < \overline{\eta}^2 - 2\lambda^2 a^2 - 2(a^\circ \lambda)^2 + 2a^2$$

$$a^2 > \lambda^2 a^2 + (a^\circ \lambda)^2. \tag{A2}$$

From (5.2)–(5.5), one can show that

$$\lambda^A = [\overline{\eta}^{AB}a_B - a^A(a^\circ \overline{\eta}^\cdot a)/2a^2]/a^2, \tag{A3}$$

assuming $a^2 \neq 0$. Using (A3), (A2) becomes

$$2a^4 + (a^\cdot \overline{\eta}^\circ a)^2/a^2 > 2a^\cdot \overline{\eta}^\circ \overline{\eta}^\circ a. \tag{A4}$$

Equations (5.6), (5.14b), and $\tilde{a}^2 = 0$ imply that a^2 is at most $\mathcal{O}(\epsilon^2)$. Define

$$\sigma^A \equiv \overline{\eta}^{AB}a_B/|a|.$$

Since $\det\overline{\eta}^{AB} \approx \det\tilde{n}^{AB} \neq 0$, $|\sigma|$ must be nonzero and of $\mathcal{O}(\epsilon^\circ)$. In terms of σ^A, (A4) becomes

$$2a^4 + (a^\circ \sigma)^2 > 2a^2\sigma^2. \tag{A5}$$

But due to the size of a^2 and σ^2,

$$2a^4 + (a^\circ \sigma) \leq 2a^4 + a^2\sigma^2 < 2a^2\sigma^2.$$

Thus (A5) cannot be true and we must have

$$I|_{\text{Class A}} < I|_{\text{Class B}}. \qquad \text{QED}$$

To prove Theorems 1, 3, and 5, more preliminary work must be done. Substituting (A3) into (5.10) gives

$$a_A[1 - (a^\circ \overline{\eta}^\circ \overline{\eta}^\circ a)/a^4 + (a^\circ \overline{\eta}^\cdot a)^2/a^6]$$

$$= \overline{\alpha}_A - \overline{\eta}_{AB}\overline{\eta}^{BC}a_C/a^2 + \overline{\eta}_{AB}a^B(a^\circ \overline{\eta}^\circ a)/a^4. \tag{A6}$$

Here we regard a_B as the funadamental quantity. Given $(\overline{\eta}^{AB}, \overline{\alpha}_B)$, (A6) is solved for a_B. Choosing λ_A as in (A3) will then guarantee that $n^{AB}a_B = 0$. Proving the existence of a Class B solution for some $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ now amounts to proving that there are solutions to (A6).

Multiplying (A6) by $a^4 = (a^\circ \overline{\alpha})^2$ gives

$$a_A(a^\circ X^\circ a) = X_{AB}a^B a^2, \tag{A7}$$

where

$$X_{AB} \equiv \overline{\alpha}_A\overline{\alpha}_B - \overline{\eta}_{AC}\overline{\eta}^C_B + \overline{\eta}_{AB}(a^\cdot \overline{\eta}^\circ a)/a^2. \tag{A8}$$

The magnitude of a_B may now be eliminated from the problem. Define

$$u_A \equiv a_A/|a|, \quad u^2 = 1, \tag{A9}$$

so (A7) and (A8) become

$$X_{AB}u^B = \chi u_A, \tag{A10}$$

$$X_{AB} = \overline{\alpha}_A\overline{\alpha}_B - \overline{\eta}_{AC}\overline{\eta}^C_B + \overline{\eta}_{AB}(u\cdot\overline{\eta}\cdot u), \tag{A11}$$

$$\chi \equiv u\circ X\circ u. \tag{A12}$$

A necessary and sufficient condition for a solution a_A to (A6) is the existence of a unit eigenvector u_A of $X_{AB}(u)$. Once a u_A is found, a_A is given by $a_A = (\overline{\alpha}\circ u)\,u_A$ where the direction of u_A is chosen to make $\overline{\alpha}\cdot u$ nonnegative.

When $\overline{\alpha}_A$ is an eigenvector of $\overline{\eta}_{AB}$, (A10) is easily solved. Assume

$$\overline{\eta}_{AB}\overline{\alpha}^B = \tau\overline{\alpha}_A. \tag{A13}$$

Then

$$u_A = \overline{\alpha}_A/|\overline{\alpha}|$$

solves (A10) with $\chi = \overline{\alpha}^2$. The solution is then $a_A = \overline{\alpha}_A$ and (A3) gives $\lambda_A = \tau\overline{\alpha}_A/2\overline{\alpha}^2$. Whenever $\tau \neq 0$, another $\lambda'_A = \overline{\alpha}_A/\tau$ can be found so that $\overline{\eta}_{AB}\lambda'^B = \overline{\alpha}_A$. Thus by Theorem 2 these $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ also give Class A solutions.

An explicit solution to (A9)—(A12) cannot be found in general.

Proof of Theorem 3: By (5.6) $a^2 \neq 0 \Rightarrow \overline{\alpha}^2 \neq 0$. Define the symmetric matrix

$$X^*_{AB}(T) \equiv \overline{\alpha}_A\overline{\alpha}_B - \overline{\eta}_{AC}\overline{\eta}^C_B + \overline{\eta}_{AB}T. \tag{A14}$$

For every real T, X^*_{AB} will have three orthogonal unit eigenvectors $u_{iA}(T)$, $i = 1, 2, 3$. Define

$$W_i(T) \equiv u_i\circ\overline{\eta}\circ u_i. \tag{A15}$$

Each of the W_i will be continuous functions of T. Finding a solution u_A to (A9)—(A12) is equivalent to finding a T such that

$$W_i(T) = T \tag{A16}$$

for some i. Since W_i is bounded from above and below by the eigenvalues of $\overline{\eta}^{AB}$, there must be at least three T satisfying (A16). The resulting $u_i(T)$ will give a nonzero $a_A = (u_i\circ\overline{\alpha})u_{iA}$ so long as $u_i\circ\overline{\alpha} \neq 0$. If $\overline{\alpha}^2 \neq 0$, then we need only show that one u_i gives $u_i\circ\overline{\alpha} \neq 0$ to complete the proof. Suppose u^A_1 is an eigenvector of $X^*_{AB}(T)$ for $T = T_1$ and $u_1\cdot\overline{\alpha} = 0$. Then u^A_1 must also be an eigenvector of $\overline{\eta}^{AB}$ and hence of $X^*_{AB}(T)$ for all T, since it will give

$$X^*_{AB}(T)u^B_1 = (\overline{\eta}_{AC}\overline{\eta}^C_B - T\overline{\eta}_{AB})u^B_1$$

$$\propto u_{1A}$$

for all values of T. If a second u^A_i, u^A_2 say, also gives $u_2\circ\overline{\alpha} = 0$, then it too must be an eigenvector of $X^*_{AB}(T)$ for all T. The third eigenvector of X^*_{AB} is then determined for all values of T and it must give $u_3\circ\overline{\alpha} \neq 0$.

QED

Proof of Theorem 5: If $(\overline{\eta}^{AB}, \overline{\alpha}_B)$ satisfy (5.14a,b)—(5.15), then certainly $\overline{\alpha}^2 \neq 0$ and therefore they admit a Class B solution by Theorem 3.

We must now show that a Class B solution gives the minimum value of I. For a Class B solution:

$$n^2 + 2a^2 = \overline{\eta}^2 + 2\chi - T^2, \tag{A17}$$

using (A9)—(A12) and $T = u\circ\overline{\eta}\circ u$. The condition that a Class B solution give a lower value of I than a Class A solution is

$$\overline{\eta}^2 < \overline{\eta}^2 + 2\chi - T^2$$

or

$$2\chi > T^2. \tag{A18}$$

Among the possible Class B solutions the one giving the largest $2\chi - T^2$ will have the smallest I.

For the unperturbed metric

$$\tilde{X}^*_{AB} = \tilde{a}_A\tilde{a}_B - \tilde{n}_{AC}\tilde{n}^C_B + T\tilde{n}_{AB}, \tag{A19}$$

$$\tilde{W}_i(T) = u_i\cdot\tilde{n}\circ u_i. \tag{A20}$$

Since $\tilde{n}^{AB}\tilde{a}_B = 0$, $\tilde{X}^*_{AB}$ may be diagonalized for all values of T by choosing axes such that

$$\tilde{a}_A = \tilde{a}\delta_{A1}, \quad \tilde{n}_{AB} = \mathrm{diag}(0, \tilde{n}_{22}, \tilde{n}_{33}). \tag{A21}$$

Therefore the directions of the u_{iA} are independent of T and are simply

$$u_{iA} = \delta_{iA}. \tag{A22}$$

The u_{iA} gives $\tilde{a}^2 \neq 0$. The W_i are also independent of T. Table III gives the values of W_i and χ_i. For $i = 1$, $2\chi_i > T_1^2 = W_1^2$. Now perturb $(\tilde{n}^{AB}, \tilde{a}_B)$ as in (5.14a,b)—(5.15). For T near 0, $\tilde{n}_{22}$, or $\tilde{n}_{33}$ the principal axes of the new X^*_{AB} can differ only slightly from those of $\tilde{X}^*_{AB}$. So the T which give $T = W_i(T)$ must differ from the unperturbed values by terms of order ϵ. Therefore $2\chi_1 > T_1^2$ for the perturbed u^A_1 and this must give the smallest I.

Finally we note that if the original metric and $\tilde{G}_3$ were BT VII$_h$ or BT VI$_h$ then the perturbed metric and G_3 will be GBT VII$_h$ or GBT VI$_h$ respectively, since T can change by $O(\epsilon)$ only. QED

Proof of Theorem 1: If $\overline{\eta}^{AB}\overline{\alpha}_B = 0$ and $\overline{\alpha}^2 = 0$, then by (5.6) $a^2 = 0$; so $n^{AB} = \overline{\eta}^{AB}$ and $a_B = \overline{\alpha}_B$.

If $\overline{\eta}^{AB}\overline{\alpha}_B = 0$ and $\overline{\alpha}^2 \neq 0$, there can be no λ_B such that $\overline{\eta}_{AB}\lambda^B = \overline{\alpha}_A$ since that would imply $\overline{\alpha}^2 = \overline{\alpha}\cdot\overline{\eta}\circ\lambda = 0$. By Theorem 2 any solution must have $a^2 \neq 0$. X^*_{AB} in (A14) may be diagonalized for all values of T by choosing axes that diagonalize $\overline{\eta}^{AB}$, since $\overline{\alpha}_B$ lies along a principal axis of $\overline{\eta}^{AB}$. The u^A_i are thus eigenvectors of $\overline{\eta}^{AB}$ and are independent of T. The only u^A_i giving $a^2 \neq 0$ is then the one parallel to $\overline{\alpha}_A$ which is also an eigenvector of $\overline{\eta}^{AB}$. Therefore $a_A = (u\circ\overline{\alpha})u_A = \overline{\alpha}_A$. (A3) implies $\lambda_A = 0$ and the result follows.

If $n^{AB} = \overline{\eta}^{AB}$ and $a_B = \overline{\alpha}_B$, (5.3) implies $\overline{\eta}^{AB}\overline{\alpha}_B = 0$. QED

TABLE III.

i	W_i when $W_i = T$	χ_i
1	0	$\tilde{a}^2$
2	$\tilde{n}_{22}$	0
3	$\tilde{n}_{33}$	0

APPENDIX B: CONSTRUCTION OF AN APPROXIMANT

We assume that the variational problem discussed in Sec. 3 has been solved on some open set $\mathcal{U}$ of $\mathcal{S}$ with the metric g_{ab}, and both the best fit triad $\{\mathbf{e}_A\}$ and the structure constants C^C_{AB} of the approximate symmetry group are known. Using some appropriate criteria, we wish to define on $\mathcal{U}$ the approximant, a homogeneous metric g^*_{ab} whose G_3 has the structure constants C^C_{AB} (defined as in the beginning of Sec. 2). One possible construction is the following. Consider the set of all triples of linearly independent vector fields $\{\mathbf{E}^*_A\}$ in $\mathcal{U}$ which satisfy

$$[\mathbf{E}^*_A, \mathbf{E}^*_B] = C^C_{AB}\mathbf{E}^*_C. \tag{B1}$$

Choosing the triple which gives the smallest value of

$$K \equiv (1/V)\int_{\mathcal{U}} g_{ab}(E^{*a}_A - e^a_A)(E^{*b}_B - e^b_B)\delta^{AB}\,dV, \tag{B2}$$

where V is the proper volume of $\mathcal{U}$ as in (3.5), we may define g^*_{ab} as the matrix inverse of

$$g^{*ab} \equiv \delta^{AB}E^{*a}_A E^{*b}_B. \tag{B3}$$

The vectors $\{\mathbf{E}^*_A\}$ are then orthonormal basis vectors for g^*_{ab} and generate the group reciprocal to the symmetry group of g^*_{ab}, as in Sec. 2.

The problem of finding the triple yielding the minimum of K can be converted to the problem of solving a set of differential equations with boundary conditions by using the calculus of variations in the standard fashion. To restrict the variations of $\{\mathbf{E}^*_A\}$ to the set of triples solving (B1), we make use of the result that any two solutions of (B1), $\{\mathbf{E}^*_A\}$ and $\{\mathbf{E}'_A\}$, with the same C^C_{AB} are related by a "coordinate transformation" (Ref. 14, Sec. 22).

$$E^{*a}_A(x) = \frac{\partial x^a}{\partial f^b}\,E'^b_A(f), \quad f^I = f^I(x). \tag{B4}$$

If $\{\mathbf{E}'_A(f)\}$ is a particular, fixed solution to (B1), in the sense of a fixed functional form, we may use (B4) to write the variation of $\{\mathbf{E}^*_A(x)\}$ in terms of a variation of $f^I(x)$ and thereby ensure that the triples being considered all solve (B1). Thus the explicit form of the variation is

$$\delta E^{*a}_A(x) = E^{*a}_B E'^B_I\,\frac{\partial E'^I_A}{\partial f^k}\,\delta f^k - E^{*I}_A E^{*a}_B E'^B_k\,\frac{\partial}{\partial x^I}(\delta f^k), \tag{B5}$$

where E'^B_I is the matrix inverse of E'^a_A. With this variation, $\delta K = 0$ implies, after one uses (B4) and the commutation relations for $\{\mathbf{E}'_B(f)\}$:

$$\nabla\circ[\mathbf{E}^*_B(\mathbf{D}^B\cdot\mathbf{E}^*_A)] + C^C_{AB}\mathbf{E}^*_C\circ\mathbf{D}^B = 0, \tag{B6}$$

where

$$\mathbf{D}^B \equiv \delta^{BC}(\mathbf{E}^*_C - \mathbf{e}_C) \tag{B7}$$

and the inner products and covariant derivative refer to g_{ab}. Boundary conditions for the differential equation (B6) come from the surface integral discarded in deriving (B6). That integral is

$$\frac{1}{V}\int_{\mathcal{U}}(\mathbf{D}^A\circ\mathbf{E}^*_B)\delta^{BC}E^{*i}_C\,\frac{\partial f^I}{\partial x^I}\,g_{Ik}\delta f^k\mathbf{E}^*_A\circ\mathbf{dS}. \tag{B8}$$

Requiring that the integral be zero independently of $\delta f^k(x)$ gives the boundary conditions for (B6).

Given g_{ab} and $\{\mathbf{e}_A\}$ on $\mathcal{U}$, the approximant g^*_{ab} may be found by solving (B1) and (B6), for $\{\mathbf{E}^*_A\}$ subject to the boundary conditions from (B8), and then using (B3). It is worth noting that the structure constants C^C_{AB} in (B1) alone and already determine the scalar invariants of the three-dimensional Riemann tensor associated with the homogeneous approximant. This occurs because we stipulate that the C^C_{AB} are structure constants of a reciprocal group with orthonormal generators. The problem that we have been solving in this appendix by a minimization criterion is—at least locally—nothing more than the problem of how best to "orient" an already-determined intrinsic geometry and fix the coordinate form of its metric, as one superimposes it on the inhomogeneous three-space.

As mentioned in Sec. 3, we may wish to find a global approximant, that is, a homogeneous metric which approximates g_{ab} everywhere on $\mathcal{S}$. If $\mathcal{S}$ requires several coordinate patches to cover it or has a complicated topology, this may be done by breaking $\mathcal{S}$ up into a collection of simply connected open sets whose union is $\mathcal{S}$, solving for the best-fit triad and approximant in each open set, and then piecing the sets and solutions together to obtain a global best-fit triad and approximant. It may occur, however, that this cannot be done. The possible topologies of a homogeneous metric are restricted by the symmetry group (see Ref. 27 for a partial list of permitted topologies). If the topology of $\mathcal{S}$ is not among those permitted by the approximate symmetry group of g_{ab}, then it will not be possible to define an approximant, g^*_{ab}, on $\mathcal{S}$ with three globally defined Killing vector fields. In a neighborhood of any point, g^*_{ab} will have the necessary Killing vector fields, but there will be no extension of these fields which covers all of $\mathcal{S}$.[26]

*Present address: Department of Physics, Northern Michigan University, Marquette, Mich. 49855.

[1]A.H. Taub, Ann. Math. 53, 472 (1951).
[2]For an excellent review of the necessary group theory, geometric techniques, and results, see MacCallum, Ref. 3.
[3]M.A.H. MacCallum, "Cosmological Models from a Geometric Point of View," in *Cargese Lectures in Physics*, edited by E. Schatzman (Gordon and Breach, New York, 1973), Vol. 6.
[4]C.W. Misner, Astrophys. J. 151, 431 (1968).
[5]For a discussion of other possible approaches, see: E.R. Harrison, Comm. Astrophys. Space Phys. 4, 29 (1974).
[6]C.B. Collins and S.W. Hawking, Ap. J. 180, 317 (1973).
[7]See Ref. 6, p. 318.
[8]For a discussion of the difference between stable and generic behavior in General Relativity, see: S.W. Hawking, Gen. Rel. Grav. 1, 393 (1971).
[9]R.A. Matzner, J. Math. Phys. 9, 1063, 1657 (1968).
[10]This definition excludes a class of space-times found by Kantowski[13] which have a G_4 but no simply-transitive G_3. See MacCallum, Ref. 3, Section IV for a fuller discussion of spatially homogeneous space–times. In this paper we shall reserve the symbol G_3 to mean a simply transitive group.
[11]Our conventions are these: $^4g_{\alpha\beta}$ has a signature of +2. Lower case Greek indices run from 0 to 3 and refer to coordinate components. Lower case Latin indices run from 1 to 3 and refer to spatial coordinate components. Triad vectors and

components are given using upper case Latin letters which run from 1 to 3. These are raised and lowered using the Kronecker delta. The three-dimensional metric in S is denoted by g_{ab}. Partial derivatives with respect to the coordinates are given by ∂_α. Covariant derivatives with respect to g_{ab} by ∇_a. The summation convention is used except where stated otherwise.

[12]If there are more than three Killing vectors, then the full symmetry group of S will have more than three parameters. The possibilities are a G_4 or G_6; a G_5 cannot be the complete group of motions, i.e., a G_6 must exist (Ref. 14, pp. 215 and 229). In all cases except one,[13] the G_4 and G_6 will contain at least one simply transitive G_3 subgroup.

[13]R. Kantowski, Ph.D. thesis, University of Texas, Austin (1966); and R. Kantowski and R.K. Sachs, J. Math. Phys. **7**, 443 (1966).

[14]L.P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1961).

[15]The vectors ξ_A and $\mathbf{E}_A$ lie entirely within S and therefore have a zero temporal component. To simplify comparison with what is done later, we will treat them as three-dimensional vectors.

[16]When the full symmetry group is G_4 or G_6, then there may exist two or more subgroups G_3, each belonging to different classes. Thus the classification of g_{ab} may not be unique. See Ref. 3, pp. 119–120.

[17]The vector a_B has the interpretation: $a_B = -\frac{1}{2}\nabla_l E^l_B$.

[18]In the remainder of this paper, vectors such as $\mathbf{e}_A$ will be three-dimensional vectors in S.

[19]This always exists. See Ref. 14.

[20]A suitable definition for the second fundamental form's approximate symmetry group would have to be found. See Ref. 3, pp. 108–109, for a similar situation in spatially homogeneous space–times.

[21]In many of the following calculations it becomes awkward to maintain the summation convention for triad indices with one index of a repeated pair up and the other down. Therefore we shall drop that restriction and sum over any pair of repeated indices in a product. For example, then

$$\Delta_{AB}^{C}\Delta_{AB}^{C} = \sum_{A,B,C=1}^{3}\Delta_{AB}^{C}\Delta_{AB}^{C}.$$

[22]Equations (2.5) and (2.6) or equivalently (2.9) and $n^{[LAB]} = 0$ are, by Lie's third fundamental theorem, necessary and sufficient for C_{AB}^{C} to be the structure constants of a Lie group. See Ref. 14.

[23]A positive-definite three-dimensional Riemannian metric will be said to be homogeneous if it is invariant under a simply transitive G_3.

[24]B.G. Schmidt, Comm. Math. Phys. **15**, 329 (1969).

[25]S.W. Hawking and G.F. R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U.P., Cambridge, 1973), p. 136.

[26]A.E. Fischer, "The Theory of Superspace," in *Relativity*, edited by M. Carmelli, S. Fickler, and L. Witten (Plenum, New York, 1970).

[27]G.F. R. Ellis, Gen. Rel. Grav. 2, 7 (1971).

[28]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. 1.

[29]In two dimensions there is only one Euler angle and (3.16) reduces simply to Poisson's equation. A forthcoming paper will treat the two-dimensional problem in its entirety along with some three-dimensional examples.

[30]Reference 3, 119.

[31]See Ref. 32, p. 44.

[32]G. Fichera, *Linear Elliptic Differential Systems and Eigenvalue Problems*, Lect. Notes in Math. No. 8 (Springer-Verlag, Berlin, 1965).

[33]L. Bers, *Contributions to the Theory of Partial Differential Equations, Ann. of Math. Studies No.* 33 (Princeton University, Princeton, N.J., 1954).

[34]Henceforth we use an abbreviated notation in which contraction on triad indices is signfied by a centered dot, e.g., $a\cdot\bar{\alpha} \equiv a^B\bar{\alpha}_B$. Also $n^2 \equiv n^{AB}n_{AB}$.

[35]Superspace is the space of all three-geometries. A point in superspace represents the equivalence class of metrics related by a coordinate transformation. See Ref. 26 for details.

[36]G.F.R. Ellis, J. Math. Phys. 8, 1171 (1967).

[37]G.F.R. Ellis and A.R. King, Comm. Math. Phys. **31**, 209 (1973).

Localized solutions of a nonlinear scalar field with a scalar potential*

Luis Vazquez†

Department of Mathematics, Brown University, Providence, Rhode Island 02912
(Received 14 February 1977)

The exact localized solutions for a nonlinear scalar field with a scalar potential are studied. In particular, we compare the stability of the above solutions and those obtained by Rosen in absence of the scalar potential.

I. INTRODUCTION

We study in this paper the static and spherically symmetric solutions, with finite energy, for the nonlinear scalar field, considered by G. Rosen,[1] in the presence of a scalar potential. Also in this case the field equation is solvable, and it has a continuous biparametric set of localized solutions which are metastable.

When the scalar potential is attractive, the localized solutions are more stable and have less energy than the solutions in absence of the potential. The opposite behavior occurs when the scalar potential is repulsive

II. GENERAL DESCRIPTION OF THE MODEL

The Lagrangian density is

$$\mathcal{L} = \left(\frac{\partial\phi}{\partial t}\right)^2 - (\nabla\phi)^2 + g\phi^6 + hr^n\phi^2, \tag{1}$$

where g is a positive constant and h and n are constants to be fixed later.

The field equation is

$$-\phi_{tt} + \Delta\phi + 3g\phi^5 + hr^n\phi = 0. \tag{2}$$

We consider static and spherically symmetric solutions $\phi = \phi(r)$. Then the field equation becomes

$$\frac{d^2\phi}{dr^2} + \frac{2}{r}\frac{d\phi}{dr} + 3g\phi^5 + hr^n\phi = 0. \tag{3}$$

Using an argument of Rosen,[2] we can see that n must be -2 in order to get solutions with finite energy. In effect the Eq. (3) is associated with a variational principle. In particular, its corresponding Lagrangian is

$$L = \int\left[\left(\frac{d\phi}{dr}\right)^2 - g\phi^6 - hr^n\phi^2\right] r^2\,dr.$$

Then we have the global condition

$$\left\{\frac{dL}{d\lambda}\left[\phi(\lambda r)\right]\right\}_{\lambda=1} = 0. \tag{4}$$

Invoking Eq. (3), we eliminate the derivative term in Eq. (4), getting the relation

$$(n+2)\int_0^\infty hr^{n+2}\,\phi^2\,dr = 0 \tag{5}$$

which implies that a nontrivial localized solution can exist only if $n = -2$. Thus the equation to study is

$$\frac{d^2\phi}{dr^2} + \frac{2}{r}\frac{d\phi}{dr} + 3g\phi^5 + \frac{h}{r^2}\,\phi = 0. \tag{6}$$

The solution obtained by Rosen[1] when $h = 0$ suggests that we can try to find solutions for (6) of the form $\phi = Ar^\alpha/(B + r^\beta)^{1/2}$. In fact here we found the biparametric continuous set of solutions

$$\phi = \phi_0 \equiv \frac{Zr^\alpha}{\left[(4/\beta^2)Z^4g + r^\beta\right]^{1/2}} \tag{7}$$

with the conditions

$$\beta = \pm 2(1 - 4h)^{1/2}, \quad \alpha = (\beta - 2)/4 \tag{8}$$

and where Z is an arbitrary constant. So we can see that h must be smaller than $1/4$ in order to get β real. When $h = 0$, Eq. (6) admits the Rosen solution and also the solution $\phi = Z/(Z^4gr^2 + 1)^{1/2}$. We note that the dimension of Z depends on β.

The energy associated with the solution (7) is

$$E = 4\pi\int_0^\infty\left[\left(\frac{d\phi}{dr}\right)^2 - g\phi^6 - \frac{h}{r^2}\,\phi^2\right] r^2\,dr, \tag{9a}$$

$$E = \tfrac{1}{4}E_R\beta^2 = E_R(1 - 4h), \tag{9b}$$

where $E_R = M^2/2g^{1/2}$ is the energy obtained by Rosen in the case $\beta = 2$. The energy is independent of the parameter Z, and it is the same for $\pm\beta$. When the scalar potential is attractive ($h > 0$), we have $E > E_R$. Also E is analytic about $h = 0$, so that the potential term in (6) is amenable to a rigorous perturbation-theory treatment

Let us now consider the dynamical stability of the solution (7). With the perturbed general solution about ϕ_0 given by

$$\phi = \phi_0 + \phi_1 e^{i\omega t}, \tag{10}$$

the linearization of (2) with (10) and the above restrictions on the parameters produces the following eigenvalue equation for the ϕ_1:

$$\frac{d^2\phi_1}{dr^2} + \frac{2}{r}\frac{d\phi_1}{dr} + \left(\omega^2 + 15g\phi_0^4 + \frac{h}{r^2}\right)\phi_1 = 0. \tag{11}$$

We can regard Eq. (11) as the equation

$$\Delta\phi_1 + [\omega^2 - V(r)]\phi_1 = 0,$$

where $V(r) = -(15g\phi_0^4 + h/r^2)$, with $\lim_{r\to\infty} rV(r) = 0$: thus by applying the results of Kato,[3] Eq. (11) has only the quadratically integrable trivial solution $\phi_1 = 0$ if $\omega^2 > 0$. Thus for the eigenfunctions of (11), ω is either zero or purely imaginary, and the associated perturbation term in (10) grows exponentially with time.

By substituting (7) into Eq. (11) and setting

$$r = \left(\frac{4}{\beta^2}Z^4g\right)^{1/8}\rho, \qquad \gamma^2 = \left(\frac{4Z^4g}{\beta^2}\right)^{2/8}(-\omega^2), \tag{12}$$

TABLE I.

β	γ_0
20	12.74
10	7.23
6	4.46
4	3.13
2	1.92
1.8	1.81
1.4	1.60
1	1.39
0.9	1.33
0.7	1.16

we obtain the dimensionless eigenvalue equation

$$\frac{d^2\phi_1}{d\rho^2}+\frac{2}{\rho}\frac{d\phi_1}{d\rho}+\left(\frac{\epsilon\rho^{\beta-2}}{(1+\rho^\beta)^2}+\frac{h}{\rho^2}\right)\phi_1=\gamma^2\phi_1, \tag{13}$$

where $\epsilon=\frac{15}{4}\beta^2$.

We computed numerically the first eigenvalue γ_0 corresponding to (13) (the Appendix) for different $\beta(h)$ and we represent them in Table I. We have that

$$\tau=\frac{1}{\gamma_0}\left(\frac{4}{\beta^2}Z^4g\right)^{1/\beta} \tag{14}$$

gives a measure of the lifetime of ϕ_0.

It follows from Equation (13) that $\gamma_0(\beta)=\gamma_0(-\beta)$. But $\tau(Z,\beta)$ and $\tau(Z,-\beta)$ have opposite behavior when Z and β change. In particular, for fixed β, $\tau(Z,\beta)$ increases when $|Z|$ increases, because the localized solution becomes more extended.

Since the dimension of Z depends on β, it only makes sense to compare the time τ for different β, when $(4Z^4g/\beta^2)^{1/\beta}$ is given. The above quantity plays the role of "particle radius." Thus we can say that τ decreases when β increases, the solution ϕ_0 becoming more and more unstable. So the localized solutions are more stable when the scalar potential h/r^2 is attractive $(0<h<\frac{1}{4})$. The opposite behavior occurs when the scalar potential is repulsive $(h<0)$.

ACKNOWLEDGMENTS

It is my pleasure to thank Professor W. Strauss for valuable advice and encouragement.

APPENDIX: FIRST EIGENVALUE FOR EQ.(13)

With the change $\phi_1=\rho^{(\beta-2)/4}\varphi_1$ we obtain the equation

$$\frac{d^2\varphi_1}{d\rho^2}+\frac{2+\beta}{2\rho}\frac{d\varphi_1}{d\rho}+\frac{\epsilon\rho^{\beta-2}}{(1+\rho^\beta)^2}\varphi_1=\gamma^2\varphi_1. \tag{A1}$$

(*a*) $\beta\geq 2$: For a localized perturbation φ_1 the boundary conditions are $\varphi_1(0)=c$, $d\varphi_1(0)/d\rho=0$ and an exponential decay at infinity. In order to find the first eigenvalue γ_0^2, numerically, we integrate Eq. (A1) with the above conditions at $\rho=0$ (by the linearity in φ_1 we can choose $c=1$) for different γ^2. In general, we get, for $\rho\to\infty$, $\varphi_1\to\pm\infty$, which gives two types of behavior. If we represent the solution as a point in the axis γ^2, the eigenvalues of Eq. (A1) correspond to the values of γ^2 which separate the two kinds of behavior above. We recognize the first eigenvalue γ_0^2 because the solution φ_1 is without nodes.

(*b*) $1<\beta<2$: In this case the boundary conditions at $\rho=0$ are as before: $\varphi_1(0)=1$; $d\varphi_1(0)/d\rho=0$, but $d^2\varphi_1(0)/d\rho^2=\infty$. Thus, to integrate numerically Eq. (A1) we cannot start at the origin. To avoid this problem, we use the following expansion of φ_1 near $\rho=0$:

$$\varphi_1\sim 1-\tfrac{5}{2}\rho^\beta+[\gamma^2/(4+\beta)]\rho^2+\cdots \tag{A2}$$

which is compatible with Eq. (A1) at $\rho=0$. Also we start the numerical integration at $\rho=\Delta\rho$, where $\Delta\rho$ is the interval of integration, with $\varphi_1(\Delta\rho)$, $d\varphi_1(\Delta\rho)/d\rho$, and $d^2\varphi_1(\Delta\rho)/d\rho^2$ given by (A2).

(*c*) $\beta=1$: We do not have the singularity in the second derivative at $\rho=0$, and the initial conditions are

$$\varphi_1(0)=1,\quad d\varphi_1(0)/d\rho=-\tfrac{5}{2},\quad \text{and } d^2\varphi_1(0)/d\rho^2=\tfrac{15}{4}+\tfrac{2}{5}\gamma^2.$$

(*d*) $0<\beta<1$: In this case we have $\varphi_1(0)=1$, but $d\varphi_1(0)/d\rho=d^2\varphi_1(0)/d\rho^2=\infty$. Thus we proceed as in (b), using the following expansion of φ_1 near the origin:

$$\varphi_1\sim 1+\sum_{n=1}^{N}a_n\rho^{n\beta}+f\rho^2+\cdots, \tag{A3}$$

with N such that $N\beta-2\geq 0$. The coefficients a_n, f are obtained as in (b) by substituting (A3) in (A1). For instance, when $\beta>\frac{2}{3}$ we get $a_1=-\frac{5}{2}$ $a_2=\frac{15}{8}$, $f=\gamma^2/(4+\beta)$.

For the numerical integration we used Hamming's predictor-corrector method of fourth order with an interval $\Delta\rho=0.01$. When $\beta=2$, we obtained $\gamma_0=1.92$, which is in agreement with the value obtained by Rosen.

*Research supported by a fellowship of the Program of Cultural Cooperation between the United States and Spain.
†On leave from Departamento de Física Teórica, Universidad de Zaragoza, Zaragoza, Spain.
[1]G. Rosen, J. Math. Phys. **6**, 1269 (1965).
[2]G. Rosen, Quarterly of Appl. Math. **27**, 133 (1969).
[3]T. Kato, Comm. Pure Appl. Math. **12**, 403 (1959).

Stationary localized solutions in nonlinear classical fields*

Luis Vazquez†

Department of Mathematics, Brown University, Providence, Rhode Island 02912
(Received 20 December 1976)

The existence of stationary localized solutions for the Dirac field interacting with Maxwell and pseudoscalar fields is studied.

INTRODUCTION

For nonlinear classical fields the stationary localized solutions are without radiation such that the physical quantities—energy, charge, spin and magnetic moment—associated with the field are finite, and they can be used as classical representations of extended particles.

Since the work of Rosen,[1] several authors[2–6] looking for localized solutions have considered the interaction of classical fields. We study in this paper the existence of stationary localized solutions for the Dirac field interacting with the electromagnetic field (Sec. I) and the pseudoscalar field (Sec. II).

I. CLASSICAL DIRAC AND MAXWELL FIELDS

The general Lagrangian is

$$\mathcal{L} = \mathcal{L}_D + \mathcal{L}_{EM} + \mathcal{L}_I, \tag{1}$$

$$\mathcal{L}_D = (i/2)(\bar\psi\gamma^\mu\partial_\mu\psi - \partial_\mu\bar\psi\gamma^\mu\psi) - M\bar\psi\psi, \tag{2}$$

$$\mathcal{L}_{EM} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{3}$$

$$\mathcal{L}_I = -e\delta\bar\psi\gamma^\mu\psi A_\mu - k\bar\psi\sigma^{\mu\nu}\psi F_{\mu\nu}, \tag{4}$$

our notation will be

$$g_{\mu\mu} = (1, -1, -1, -1), \quad \gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{5}$$

$$\gamma^\kappa = \begin{pmatrix} 0 & \sigma^\kappa \\ -\sigma^\kappa & 0 \end{pmatrix} \quad (\sigma^\kappa \text{ the Pauli matrices}),$$

$$\sigma^{\mu\nu} = \frac{1}{2i}[\gamma^\mu, \gamma^\nu],$$

$$A^\mu = (A^0, \mathbf{A}), \quad F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu, \tag{6}$$

e is the electromagnetic constant, δ is a parameter taking the values 0 and 1, and k is the coupling constant for the Pauli term.

We consider stationary localized solutions of the form

$$\psi = e^{-i\omega t}\psi(\mathbf{x}), \quad \psi(\mathbf{x}) = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}, \quad A_\mu = A_\mu(\mathbf{x}) \quad (\partial_\mu A^\mu = 0) \tag{7}$$

such that $\{u_i\}_{i=1,4}$ and A_μ are functions of class $C^2(\mathbb{R}^3)$ which are bounded. Also, $\{u_i\}$ and the first derivatives of A_μ are quadratically integrable.

The field equations are

$$i\gamma^\kappa\partial_\kappa\psi - M\psi + \omega\gamma^0\psi - e\delta\gamma^\mu A_\mu\psi - k\sigma^{\mu\nu}F_{\mu\nu}\psi = 0, \tag{8a}$$

$$\Delta A^\mu + e\delta\bar\psi\gamma^\mu\psi - k\partial_\nu(\bar\psi\sigma^{\mu\nu}\psi) = 0. \tag{8b}$$

We consider the following cases:

(A) $\delta = 1$, $k = 0$, $A^\mu = (A^0, \mathbf{0})$

When $(\omega/M)^2 < 1$, Wakano[5] numerically, found spherically symmetric localized solutions with negative energy. In the general case it is possible to prove the nonexistence of localized solutions, when $\omega \le -M$, by using a generalization of the pseudovirial theorem conconsidered by Rosen[3] in the case of time-independent solutions. In our situation the spinor field is time dependent but the Lagrangian is static, so we get the following global condition:

$$\left\{\frac{dI}{d\lambda}(\psi(\lambda\mathbf{x}, t), A_\mu(\lambda\mathbf{x}))\right\}_{\lambda=1} = 0, \tag{9}$$

where $I(\psi(\mathbf{x}, t), A_\mu(\mathbf{x})) = \int\mathcal{L}\,d^3x$. Eliminating all spatial derivatives of the fields which appear in (9), by evoking the field equations (8), we get an integral condition which must be satisfied by the localized solutions. That condition can be used in order to prove the nonexistence of localized solutions, to test the accuracy of the numerical localized solutions, and to find variational solutions.[7]

In the present case we get

$$\int(\omega\psi^*\psi - M\bar\psi\psi - \frac{\mathbf{E}^2}{2})\,d^3x = 0, \tag{10}$$

where $\mathbf{E}$ is the electric field. Using the expression for $\psi(\mathbf{x})$ given in (7), we obtain

$$\int((\omega - M)(|u_1|^2 + |u_2|^2) + (\omega + M)(|u_3|^2 + |u_4|^2) - \mathbf{E}^2/2) \times d^3x = 0.$$

Since $M > 0$, the integrand is definite negative if $\omega \le -M$. Thus (10) implies $\psi = 0$ if $\omega \le -M$.

Remark: The Dirac equation with an electrostatic potential $A_0 = q/r$ (q arbitrary) does not have localized solutions if $(\omega/M)^2 \ge 1$. In effect the global condition (9) for Eq. (8a) with $A_\mu = (q/r, \mathbf{0})$ is $\int(\omega\psi^*\psi - M\bar\psi\psi)d^3x = 0$ and so we have the above conclusion.

(B) $\delta = 1$, $k = 0$, $A^\mu = (0, \mathbf{A})$

Multiplying Eq. (8a) by ψ^* and its adjoint equation by $\gamma^0\psi$ and subtracting, we get

$$2\omega(\bar\psi\psi) - 2M(\psi^*\psi) + i\partial_\kappa(\psi^*\gamma^\kappa\psi) = 0.$$

If ψ is a localized solution,

$$\lim_{|\mathbf{x}|\to\infty} |\mathbf{x}|^3 u_i^2 = 0,$$

then

$$\int(M\psi^*\psi - \omega\bar\psi\psi)d^3x = 0 \Rightarrow \psi \equiv 0 \tag{11}$$

if $(\omega/M)^2 \le 1$, so that there is no localized solution. This case has been treated numerically by Wakano[5] considering a multipole expansion of the fields.

The pseudovirial theorem gives us the integral

condition

$$\int(\omega\psi^*\psi - M\bar{\psi}\psi - \mathbf{H}^2/2 + \sum_{\kappa=1}^{3}(\nabla A^\kappa)^2)d^3x = 0, \tag{12}$$

where H is the magnetic field. Using (11) and since $\mathbf{H}^2/2 < \sum_{\kappa=1}^{3}(\nabla A^\kappa)^2$ we get

$$\frac{\omega^2 - M}{\omega}\int\psi^*\psi\, d^3x < 0,$$

so there is no localized solution if $\omega > M$. When $\omega < -M$, proving the nonexistence of localized solutions is equivalent to proving that the Pauli equation in a magnetic field has no eigenvalues imbedded in the continuous part of the spectrum. In effect we can transform Eq. (8a), with $k = 0$, into an equation like the Klein–Gordon equation.

$$\Delta\psi + (\omega^2 - M^2)\psi + (e^2A_0^2 - 2e\omega A_0 - e^2\mathbf{A}^2 + 2ie\mathbf{A}\cdot\nabla + ei\alpha\cdot\mathbf{E} - e\Sigma\cdot\mathbf{H})\psi = 0, \tag{13}$$

where $\alpha^\kappa = \gamma^0\gamma^\kappa$, $\Sigma^\kappa = \begin{pmatrix}\sigma^\kappa & 0\\ 0 & \sigma^\kappa\end{pmatrix}$, and **E** and **H** are the electric and magnetic fields. In the present case $A_0 = 0$, $\mathbf{E} = 0$, and since Σ is diagonal we get

$$e^2A^2V_\kappa - \Delta V_\kappa - 2ie\mathbf{A}\cdot\nabla V_\kappa + e\boldsymbol{\sigma}\cdot\mathbf{H}V_\kappa = (\omega^2 - M^2)V_\kappa, \tag{14}$$

where $V_1 = \begin{pmatrix}u_1\\ u_2\end{pmatrix}$ and $V_2 = \begin{pmatrix}u_3\\ u_4\end{pmatrix}$. Thus we can regard V_κ as an eigenvector associated with the eigenvalue $\omega^2 - M^2 > 0$ of the Pauli equation in a magnetic field **H**. In this case the continuous part of the spectrum is $[0,\infty)$[10] since H is a localized solution and then

$$H(\mathbf{x}) = \underset{|\mathbf{x}|\to\infty}{O}(|\mathbf{x}|^{3/2+\theta}) \text{ for a certain } \theta.$$

So there are no localized solutions if there are no points of the discrete spectrum in the continuous spectrum in the Pauli equation.

(C) $\delta = 0$, $k \neq 0$, $A^\mu = (A^0, 0)$

In the same way as in (B) we again obtain Eq. (11). So in this case there are no localized solution if $(\omega/M)^2 \leq 1$.

From the pseudovirial theorem we obtain

$$\int(\omega(\psi^*\psi) - M(\bar{\psi}\psi) + \mathbf{E}^2/2)\, d^3x = 0 \tag{15}$$

and using relation (11),

$$\frac{\omega^2 - M^2}{\omega}\int\psi^*\psi\, d^3x + \int\frac{\mathbf{E}^2}{2}d^3x = 0,$$

so there are no localized solutions if $\omega > M$. When $\omega < -M$ we find a *semilocalized solution* in which the electromagnetic and interaction energies are finite but not the spinorial energy. In fact, in the present case Eq. (8) admits stationary solutions which are separable in spherical coordinates

$$\psi = e^{-i\omega t}\begin{pmatrix} h(r)\begin{pmatrix}1\\0\end{pmatrix} \\ if(r)\begin{pmatrix}\cos\theta\\ \sin\theta e^{i\varphi}\end{pmatrix}\end{pmatrix}, \quad A^0 = A(r). \tag{16}$$

The radial equations are

$$f' + \frac{2}{r}f + (M - \omega)h + 2kA'f = 0, \tag{17a}$$

$$h' + (M + \omega)f - 2kA'h = 0, \tag{17b}$$

$$A'' + \frac{2}{r}A' = 4k((fh)' + 2fh/r). \tag{17c}$$

From (17c) we get $A' = 4kfh$ and so we reduce system (17) to the equations

$$f' + \frac{2}{r}f + (M - \omega)h + 8k^2f^2h = 0,$$
$$h' + (M + \omega)f - 8k^2h^2f = 0. \tag{18}$$

Actually these radial equations correspond to a spinor field with a pseudoscalar coupling $(\bar{\psi}\gamma^5\psi)^2$, and every solution of (18) tends asymptotically to zero when $\omega < -M$. We can see this from the following phase space analysis of (18): These differential equations describe a nonconservative, one-dimensional motion, since "time" appears explicitly. The energy for the corresponding conservative motion [defined by Eqs. (18) after the $1/r$ term has been deleted] is

$$K = \tfrac{1}{2}(M - \omega)h^2 - \tfrac{1}{2}(M + \omega)f^2 + 4k^2f^2h^2. \tag{19}$$

For the nonconservative motion which corresponds to our actual problem, we have

$$\frac{dK}{dr} = 2(M + \omega)\frac{f^2}{r} - \frac{16}{r}f^2h^2 \tag{20}$$

so that if $\omega < -M$, then $K > 0$ in the whole plane (f, h), $K = 0$ at the origin and $dK/dr < 0$. Thus every regular solution of (18) for which $f(0) = 0$ and $h(0) \neq 0$ goes to zero as $r \to \infty$. Its asymptotic behavior is

$$f \sim \frac{1}{r}\sin(\beta_1 + r(\omega^2 - M^2)^{1/2}),$$
$$h \sim \frac{1}{r}\sin(\beta_2 + r(\omega^2 - M^2)^{1/2}), \tag{21}$$

where β_1 and β_2 are constants. In this way we can see that $A' \sim 1/r^2$ as $r \to \infty$ being finite everywhere. The electrostatic and interaction energies are finite since the energy momentum tensor is

$$T^{\alpha\beta} = F^{\nu\alpha}F^{\beta}{}_{\nu} + \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}g^{\alpha\beta} + \frac{i}{4}[\bar{\psi}\gamma^\alpha\partial^\beta\psi + \bar{\psi}\gamma^\beta\partial^\alpha\psi - (\partial^\alpha\bar{\psi})\gamma^\beta\psi - (\partial^\beta\bar{\psi})\gamma^\alpha\psi] - \frac{e\delta}{2}(\bar{\psi}\gamma^\alpha\psi A^\beta + \bar{\psi}\gamma^\beta\psi A^\alpha) - k(\bar{\psi}\sigma^{\nu\alpha}\psi F_\nu{}^\beta + \bar{\psi}\sigma^{\nu\beta}\psi F_\nu{}^\alpha) \tag{22}$$

and in our case we get

$$E = \int T^{00}d^3x = \int\omega(f^2 + h^2)d^3x + \int\tfrac{1}{2}A'^2d^3x - \int kA'fhd^3x. \tag{23}$$

The value of the two last integrals is $16\pi k^2\int_0^\infty f^2h^2r^2dr$, which is finite.

(D) $\delta = 0$, $k \neq 0$, $A^\mu = (0, \mathbf{A})$

In this case from the virial theorem we obtain the integral condition

$$\int(\omega\psi^*\psi - M\bar{\psi}\psi - \tfrac{1}{2}\mathbf{H}^2)d^3x = 0 \tag{24}$$

so that when $\omega \leq -M$ there are no localized solutions. When $\omega > -M$ the existence of those is an open question. In this case Eqs. (8) admit no stationary solutions which are separable in spherical coordinates. We have to make a multipole expansion and write down equations for each partial wave. It is, however, much simpler and equivalent to substitute a multipole approximation in the Lagrangian, integrate over the angles and make variations of the radial functions. Taking the first term

in the expansion of ψ and $\mathbf{A}$,

$$\psi = e^{-i\omega t}\begin{pmatrix} h(r)\begin{pmatrix}1\\0\end{pmatrix} \\ if(r)\begin{pmatrix}\cos\theta \\ \sin\theta e^{i\varphi}\end{pmatrix}\end{pmatrix}, \quad \begin{aligned} A_x &= V(r)\sin\theta\sin\varphi, \\ A_y &= -V(r)\sin\theta\cos\varphi, \\ A_z &= 0 \end{aligned} \tag{25}$$

(A^μ automatically satisfies the Lorentz condition $\partial_\mu A^\mu = 0$), and making the following changes in functions and variables

$$(f,h) = \frac{1}{4k}(3M)^{1/2}(F,H), \quad V = \frac{3}{4k}\mathcal{V},$$

$$r = \rho/M, \quad \Lambda = \omega/M,$$

we obtain the radial equations

$$F' + \frac{2}{\rho}F + (1-\Lambda)H - (\mathcal{V}' + 2\mathcal{V}/\rho)H = 0, \tag{26a}$$

$$H' + (1+\Lambda)F + \mathcal{V}'F = 0, \tag{26b}$$

$$\mathcal{V}'' + \frac{2}{\rho}\mathcal{V}' - \frac{2}{\rho^2}\mathcal{V} = (HH' + FF' + F^2/\rho). \tag{26c}$$

For the localized solutions the regularity at the origin implies $F(0)=0$, $\mathcal{V}(0)=0$; and from (26c) we get $H^2(0) = 6\mathcal{V}'(0) + 2\int_0^\infty F^2/\rho\, d\rho$. So given Λ the localized solutions of (26) depend on two parameters $H(0)$ and $\mathcal{V}'(0)$. Since the equations are invariant under the change $(F,H,\mathcal{V}) \to (-F,-H,\mathcal{V})$, we only have to study the region of the plane $(H(0), \mathcal{V}'(0))$ for which $H(0) \geq 0$ and $\mathcal{V}'(0) < \frac{1}{6}H^2(0)$. We numerically explored[8] the above region and we did not find localized solutions. In particular we always found that $\mathcal{V}/\rho \sim_{\rho\to\infty} \eta$ $(\eta > 0)$ and that F and H have an oscillatory behavior. The above computations have been made with $\Lambda = 0.2,\ 0.5,\ 0.8$.

We can explain the nonexistence of localized solutions for (26) observing that (a) in these equations $-\mathcal{V}'$ is like an electrostatic potential which changes its sign at least one time for a localized solution [since $\mathcal{V}(0)=0$] so this potential is attractive in one region and repulsive in the other one. (b) From Eq. (26c) and using the Green function associated we get

$$\mathcal{V} = \frac{1}{6\rho^2}\int_0^\rho (3H^2 + F^2)\rho^2 d\rho - \frac{\rho}{3}\int_\rho^\infty \frac{F^2}{\rho}\, d\rho.$$

In this way if $\mathcal{V}$ is a localized solution $\mathcal{V} \sim \epsilon/\rho^2$ $(\epsilon > 0)$ as $\rho \to \infty$, then F and H tend to the solution of the equations

$$\begin{aligned} &F' + \frac{2}{\rho}F + (1-\Lambda)H = 0, \\ &H' + (1+\Lambda)F - \frac{2\epsilon}{\rho^3}F = 0, \end{aligned} \tag{27}$$

when $\rho \to \infty$, but (27) has only the trivial localized solution $F = H = 0$. In effect for these equations we have the following integral conditions:

$$\int_0^\infty \left[(1-\Lambda)H^2 + (1+\Lambda)F^2 - \frac{2\epsilon}{\rho^3}F^2\right]\rho^2 d\rho = 0, \tag{28a}$$

obtained by direct integration on (27) and

$$\int_0^\infty [(1+\Lambda)F^2 + (\Lambda-1)H^2 + 4\epsilon F^2/\rho^3]\rho^2 d\rho = 0 \tag{28b}$$

obtained by applying the pseudovirial theorem to the Lagrangian

$$\mathcal{L} = \int [(1+\Lambda)F^2 + (\Lambda-1)H^2 + FH' - HF' - 2\frac{FH}{\rho} - 2\epsilon F^2/\rho^3]\rho^2 d\rho$$

corresponding to (27).

Thus from (28) we get

$$\int_0^\infty [2(1+\Lambda)F^2 + 2\epsilon F^2/\rho^3]\rho^2 d\rho = 0, \tag{28c}$$

which implies $F = 0$ $(H=0)$ if $\Lambda = \omega/M \geq -1$. In this way only the trivial solution of (26) satisfies the above asymptotic behavior.

Also we tried to find localized solutions for the equations (26a), (26b) with $\mathcal{V} = \epsilon\rho/(\beta + \rho^3)$ which could correspond to a localized solution of (26) (if $\epsilon > 0$), and we did not find them. In particular, we used the values $(\epsilon,\beta) = (10,2)$, $(300,16)$, $(-40,2)$, $(-300,16)$. For the first two values, the "electrostatic potential" $\mathcal{V}'$ is attractive near the origin and repulsive far away; for the last two values the opposite behavior occurs. $\mathcal{V}'$ changes sign at $\rho = 1, 2$. The parameters (ϵ,β) must satisfy $\epsilon/\beta > 1 - \Lambda/3$, $|\epsilon|/\beta > 1 + \Lambda$, the condition required by the relation

$$\int_0^\infty [H^2(1-\Lambda) + F^2(1+\Lambda) + \mathcal{V}'F^2 - (\mathcal{V}' + \frac{2}{\rho}\mathcal{V})H^2]\rho^2 d\rho = 0$$

obtained by integrating (26a) and (26b)

In the general case we can expect the nonexistence of localized solutions for Eq. (8) because in this case the interaction term in the Dirac equation is given by a diagonal matrix $\sum^\kappa = \begin{pmatrix}\sigma^\kappa & 0\\ 0 & \sigma^\kappa\end{pmatrix}$ with the first derivatives of $\mathbf{A}$; that is, like an electrostatic potential attractive and repulsive in different regions, and the repulsive effect cannot be compensed by the spinor field.

II. CLASSICAL DIRAC AND PSEUDOSCALAR FIELDS

A. Pseudoscalar coupling

The Lagrangian is

$$\mathcal{L} = \mathcal{L}_D + \mathcal{L}_{KG} + \mathcal{L}_I, \tag{29}$$

$$\mathcal{L}_D = \frac{i}{2}[\bar\psi\gamma^\mu\partial_\mu\psi - (\partial_\mu\bar\psi)\gamma^\mu\psi] - M\bar\psi\psi, \tag{30}$$

$$\mathcal{L}_{KG} = \tfrac{1}{2}[\partial_\mu\phi\partial^\mu\phi - m^2\phi^2] + \frac{\alpha}{4}\phi^4, \tag{31}$$

$$\mathcal{L}_I = g\bar\psi\gamma^5\psi\phi, \tag{32}$$

where $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$.

The field equations are

$$i\gamma^\mu\partial_\mu\psi - M\psi + g\phi\gamma^5\psi = 0, \tag{33a}$$

$$\partial_\nu\partial^\nu\phi + m^2\phi - \alpha\phi^3 - g\bar\psi\gamma^5\psi = 0, \tag{33b}$$

and we look for stationary localized solutions as in Sec. 1 where now $\phi = \phi(\mathbf{x})$ is a function of class $C^2(\mathbb{R}^3)$, bounded, quadratically integrable as are its first derivatives.

Multiplying Eq. (32a) by ψ^* and its adjoint equation by $\gamma^0\psi$ and subtracting, we get

$$2\omega(\bar\psi\psi) - 2M(\psi^*\psi) + i\partial_\kappa(\psi^*\gamma^\kappa\psi) = 0$$

and for localized solutions $\int(M\psi^*\psi - \omega\bar\psi\psi)d^3x = 0$ which implies $\psi \equiv 0$ if $(\omega/M)^2 \leq 1$ and there are no localized

solutions. The above result is independent of the pseudoscalar self-coupling $\alpha\phi^4$. Thus the effect of this term in Eqs. (33) is weaker than a positive $(\bar{\psi}\psi)^2$ self-interaction, tion, since in this case we have localized solutions.[9]

The pseudovirial term gives us the integral condition

$$\int(\omega(\psi^*\psi) - M(\bar{\psi}\psi) + \tfrac{1}{2}(\nabla\phi)^2 - m^2\phi^2 - (\alpha/4)\phi^4)d^3x = 0 \tag{34}$$

which makes no statement about the existence (or nonexistence) of localized solutions when $(\omega/M)^2 > 1$. In this case to prove the nonexistence of those solutions is equivalent to proving the nonexistence of eigenvalues in the continuous part of the spectrum for a certain operator. In effect, as in Sec. I(B), we can reduce Eq. (33a) to an equation like the Klein—Gordon one.

$$\Delta\psi + (\omega^2 - M^2)\psi - g^2\phi^2\psi - g\gamma^0\Sigma\cdot\nabla\phi\psi = 0 \tag{35}$$

and since $\gamma^0\Sigma$ is diagonal, the above equation is equivalent to two independent systems of equations.

$$-\Delta V_\kappa + g^2\phi^2 V_\kappa - \epsilon_\kappa g\sigma\cdot\nabla\phi V_\kappa = (\omega^2 - M^2)V_\kappa \tag{36}$$

where

$$V_1 = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad V_2 = \begin{pmatrix} u_3 \\ u_4 \end{pmatrix}, \quad \text{and } \epsilon_1 = -1, \quad \epsilon_2 = 1.$$

Thus we can regard V_κ as an eigenvector associated with the eigenvalue $(\omega^2 - M^2) > 0$, corresponding to the operator $T = P + Q(\mathbf{x})$,

$$P = -\Delta,$$

$$Q(\mathbf{x}) = \begin{pmatrix} g^2\phi^2 - \epsilon_\kappa g\partial_z\phi & -\epsilon_\kappa g(\partial_x\phi + i\partial_y\phi) \\ -\epsilon_\kappa g(\partial_x\phi - i\partial_y\phi) & g^2\phi^2 + \epsilon_\kappa g\partial_z\phi \end{pmatrix}.$$

The continuous part of the spectrum of P is the semiaxis $\lambda \geq 0$. The largest and smallest eigenvalues of the matrix $Q(\mathbf{x})$ are

$$\nu_1 = g^2\phi^2 + |\epsilon_\kappa g|\,\mu, \quad \nu_2 = g^2\phi^2 - |\epsilon_\kappa g|\,\mu,$$

where $\mu = (\nabla\phi\cdot\nabla\phi)^{1/2}$. Suppose $\lim_{|\mathbf{x}|\to\infty}\phi^2 = \lim\mu = 0$ (condition satisfied if ϕ is a localized solution), then ν_1, ν_2 tend to zero for $|\mathbf{x}| \to \infty$ and the norm of the matrix $Q(\mathbf{x})$ also tends to zero. Then by the theorems about differential operators on vector functions[10] we conclude that the continuous part of the spectrum of the above operator T coincides with the semiaxis $\lambda \geq 0$.

Remark: If $g = 0$ and $\alpha > 0$, Eq. (33b) has at least a countably infinite number of localized solutions as has been proved by Berger.[11]

B. Pseudovectorial coupling

We have the same Lagrangian as before assuming that $\mathcal{L}_I$ has the form

$$\mathcal{L}_I = ig\bar{\psi}\gamma^\mu\gamma^5\psi\partial_\mu\phi \tag{37}$$

and the field equations are

$$i\gamma^\mu\partial_\mu\psi - M\psi + ig\gamma^\mu\gamma^5\partial_\mu\phi\psi = 0, \tag{38a}$$

$$\partial_\nu\partial_\nu\phi + m^2\phi - \alpha\phi^3 + ig\partial_\mu(\bar{\psi}\gamma^\mu\gamma^5\psi) = 0. \tag{38b}$$

From the pseudovirial theorem we get the integral condition

$$\int(\omega(\psi^*\psi) - M(\bar{\psi}\psi) - \tfrac{1}{2}(\nabla\phi)^2 - \tfrac{3}{2}m^2\phi^2 + \alpha\phi^4)d^3x = 0 \tag{39}$$

which implies the nonexistence of localized solutions if $\omega \leq -M$ and $\alpha \leq 0$.

Equations (38) admit no stationary solutions which are separble in spherical coordinates, but as in Sec. I I(D) we get for the lowest terms in the multipole expansion

$$\psi = e^{i\omega t}\begin{pmatrix} h(r)\begin{pmatrix}1\\0\end{pmatrix} \\ if(r)\begin{pmatrix}\cos\theta \\ \sin\theta e^{i\varphi}\end{pmatrix}\end{pmatrix}, \quad \phi = \phi(r)\cos\theta, \tag{40}$$

and making the following changes in functions and variables

$$(f,h) = \frac{1}{g}(\tfrac{3}{2}M)^{1/2}(F,H), \quad \phi = \frac{3}{g}\Phi,$$
$$r = \rho/M, \quad \Lambda = \omega/M, \quad \nu = m/M, \quad \beta = \frac{108}{5}\frac{\alpha}{g^2M^2}. \tag{41}$$

The radial equations are

$$F' + \frac{2}{\rho}F + (1 - \Lambda)H - \Phi'H - \frac{2}{\rho}\Phi H = 0, \tag{42a}$$

$$H' + (1 + \Lambda)F + \Phi'F - \frac{2}{\rho}\Phi F = 0, \tag{42b}$$

$$\Phi'' + \frac{2}{\rho}\Phi' - \frac{2}{\rho^2}\Phi - \nu^2\Phi + \beta\Phi^3 - \left(HH' + FF' + \frac{2}{\rho}F^2\right) = 0. \tag{42c}$$

Let us consider the case $\beta = 0$. For the lcoalized solutions the regularity at the origin implies $F(0) = 0$, $\Phi(0) = 0$, and from (42c) with the help of the Green function we get

$$H^2(0) = 6\Phi'(0) + 2\int_0^\infty\left(2(1+\nu)\frac{F^2}{\rho} + \frac{\nu^2\rho}{2}(H^2F^2)\right)e^{-\nu\rho}d\rho.$$

So given Λ and ν, the localized solutions of (42) depend on two parameters $H(0)$ and $\Phi'(0)$, and since the equations are invariant under the change $(F,H,\Phi) \to (-F, -H, \Phi)$, we only have to study the region of the plane $(H(0), \Phi'(0))$ for which $H(0) \geq 0$ and $\Phi'(0) < \frac{1}{6}H^2(0)$. We explored numerically[8] the above region and we did not find localized solutions. The above computations have been made with $\Lambda = 0.2$, 0.5, 0.8 and $\nu = m_\pi/m_p$, m_π/M_e (where m_π, M_p and M_e are the mass of the pion, proton, and electron respectively).

We can explain the above result by observing that in Eqs. (42a), $-\Phi'$ is like an electrostatic potential and $-(2/\rho)\Phi$ is like a scalar potential, and when $\rho \to \infty$ then $\Phi \sim (A/\rho)e^{-\nu\rho}$ and the sum of both "potentials" is like repulsive electrostatic potential if $\Phi > 0$ $(A > 0)$ and an attractive electrostatic potential in the other case. On the other hand, when there is only a scalar potential it has to be negative in some region in order to have localized solutions. Here the situation is analogous to Eqs. (26), because if $\Phi'(0) > 0$ then we have in (42a) and (42b) attractive electrostatic potential near the origin and a repulsive electrostatic potential far away. The scalar potential is attractive near the origin, but is

TABLE I.

Fields*	Interaction Term	Localized solutions $(\omega/M)^2>1$	Localized solutions $(\omega/M)^2<1$
ψ, A^0	$-e\bar{\psi}\gamma^0\psi A_0$	No if $\omega \leqslant -M$ (Virial)	Yes (numerically found[5])
ψ, $\mathbf{A}$	$-e\bar{\psi}\gamma^\kappa\psi A_\kappa$	No if $\omega > M$ (Virial)	No (particular integral condition)
ψ, A^0	$-k\bar{\psi}\sigma^{\mu\nu}\psi F_{\mu\nu}$	No (Virial and numerically)	No (particular integral condition)
ψ, $\mathbf{A}$	$-k\bar{\psi}\sigma^{\mu\nu}\psi F_{\mu\nu}$	No, if $\omega \leqslant -M$ (Virial)	Virial theorem inconclusive but numerically: no
ψ, ϕ	$g\bar{\psi}\gamma^5\psi\phi + \frac{\alpha}{4}\phi^4$	Virial theorem inconclusive	No (particular integral condition)
ψ, ϕ	$ig\bar{\psi}\gamma^\mu\gamma^5\psi\partial_\mu\phi + \frac{\alpha}{4}\phi^4$	No, if $\omega \leqslant -M$ and $\alpha \leqslant 0$ (Virial)	Virial theorem inconclusive but numerically: no

*ψ: Spinor field
A^0: Electrostatic potential
$\mathbf{A}$: Electromagnetic vector potential
ϕ: Pseudoscalar field

dominated far away. The scalar potential is attractive near the origin, but is dominated far away by the repulsive electrostatic potential. The situation is the opposite when $\Phi'(0) < 0$ (we are thinking in the simplest localized solution without nodes) so in both cases there is a repulsive region without compensation in order to have the localized solution.

We also tried to find localized solutions for Eqs. (42a) and (42b), Φ being the lowest order solution of Eq. (42c) with $\beta > 0$ and without the spinorial source, but we did not find them.

For the general solution we have a situation as in I(D) because the interaction term in the Dirac equation is given by a diagonal matrix $\gamma^5\gamma^\kappa$ with the first derivatives of ϕ. That is like an electrostatic potential attractive and repulsive in different regions, and the repulsive effect cannot be compensed by the spinor field.

III. CONCLUSIONS

A summary of the results is given in Table I. When $(\omega/M)^2 > 1$, in some cases, proving the nonexistence of localized solutions is related to proving the nonexistence of eigenvalues in the continuous part of the spectrum for a certain operator.

It is interesting that in all the cases considered here, there are stationary localized solutions with positive energy, when $(\omega/M)^2 < 1$, if the spinor field has a positive $(\bar{\psi}\psi)^2$ self-interaction.[6,9,12]

ACKNOWLEDGMENT

I should like to express my sincere thanks to Professor W. A. Strauss for his very useful suggestions.

*Research supported by a fellowship of the Subdireccion General de Promocion de la Investigacion of Spain.

†On leave from Departamento de Fisica Teorica Universidad de Zaragoza, Zaragoza, Spain.

[1]N. Rosen, Phys. Rev. 55, 94 (1939).
[2]R. Finkelstein, C. Fronsdal, and P. Kaus, Phys. Rev. 103, 1571 (1956).
[3]G. Rosen, J. Math. Phys. 7, 2066 (1966).
[4]G. Rosen, J. Math. Phys. 9, 999 (1968).
[5]M. Wakano, Prog. Theor. Phys. 35, 1117 (1966).
[6]A. F. Rañada, M. F. Rañada, M. Soler, and L. Vazquez, Phys. Rev. D 10, 517 (1974).
[7]D. D. Betts, H. Schiff, and W. B. Strickfaden, J. Math. Phys. 4, 334 (1963).
[8]The method used is the one indicated in Ref. 9, p. 316.
[9]A. F. Rañada and L. Vazquez, Prog. Theor. Phys. 56, 311 (1976).
[10]I. Glazman, *Direct Methods of Qualitative Spectral Analysis of Singular Differential Operators* (Moscow, 1963) (English translation, Davey, Hartford, Connecticut, 1965).
[11]M.S. Berger, J. Funct. Anal. 9, 249 (1972).
[12]The localized solutions corresponding to the spinor field with a positive $(\bar{\psi}\psi)^2$ self-interaction and interacting with the pseudoscalar field (pseudovectorial coupling) have been found by A. F. Rañada (private communication).

Noncentral potentials: The generalized Levinson theorem and the structure of the spectrum*

Roger G. Newton

Physics Department, Indiana University, Bloomington, Indiana 47401
(Received 28 December 1976)

Fredholm theory is applied to the Lippmann–Schwinger equation for noncentral potentials. For a specified wide class of potentials it is proved that the modified Fredholm determinant cannot vanish for real $k \neq 0$. The point $k = 0$ is examined and the analog of the distinction between zero-energy bound states and zero-energy resonances for central potentials is found. A generalized Levinson theorem is proved.

1. INTRODUCTION AND SUMMARY

Nonrelativistic scattering theory for noncentral local potentials is worked out in much less detail than for central potentials. This is regrettable, both from the point of view of principle and from that of physics. There are important mathematical issues that we cannot really understand unless we know if they are special to the central case. As a matter of physical application, the scattering of large molecules and heavy nuclei is often approximated by a description in terms of local noncentral potentials. There is therefore good reason to study the scattering by such potentials in as much detail, if possible, as that by central ones.

Apart from the proof of the generalized Levinson theorem itself, there are two principal new results of this paper. The first is the discovery of the exact analog, in the case of noncentral potentials, of the well-known distinction between s-wave zero-energy resonances and the zero-energy bound states of higher angular momenta in the case of central potentials. In view of the role played by the zero-energy resonances in producing the Efimov effect in three-particle systems, this may turn out to be of more than passing interest. The second is a proof of the absence of real singular points $k \neq 0$ at which the modified Fredholm determinant vanishes for potentials in the class defined by (2.1).

This paper approaches the subject from a time-independent point of view, in the coordinate representation. The primary tool used will be the modified form of the theory of Fredholm equations for L^2 kernels. The use of this method in scattering theory was initiated by Jost and Pais.[1] Simon[2] attacked the same general problem with more powerful tools. The results of the present paper touch on his but go beyond them in certain directions. Specific generalizations of Levinson's theorem were recently given by Dreyfus[3] and by Osborn and Bollé.[4]

The restriction we impose on the potential, given by (2.1), is more severe than that of Simon,[2] who usually uses the intersection of L^1 with the Rollnik class (2.3). I found (2.1) very convenient for some specific purposes that will be noted.

In Sec. 2 the Lippmann–Schwinger equation is discussed. It mainly establishes the setting and notation; the methods and results are standard. Section 3 establishes the connection between the bound states of negative energy and the zeros of the modified Fredholm determinant D of the Lippmann–Schwinger equation, including the connections between the multiplicity of a zero and the degeneracy of the corresponding bound state.

In Sec. 4 we discuss the exceptional points on the real axis, not at the origin. These are points at which D vanishes. We distinguish between exceptional points of the first and second kind. The first are positive-energy bound states, and they are shown to have no effect on the scattering cross section. They are known not to exist for potentials in the class (2.1). The second produce an infinite cross section and a divergence of the canonical completeness relation. It is proved that for potentials in the class (2.1) such points cannot exist. This implies that the singular continuous spectrum of the Hamiltonian is empty and we have "strong asymptotic completeness."

Section 5 deals with the point $k = 0$, if it is exceptional. We show that three, and only three, cases are possible: In the first, the origin is an exceptional point of the first kind and there exists an n-fold degenerate bound state. In that case D goes exactly as k^{2n}, and the scattering cross section is finite at $k = 0$. In the second case $k = 0$ is an exceptional point of the second kind, and there exists a "half-bound" state, but no bound states. In that case D is shown to go to zero exactly linearly at $k = 0$, and the cross section tends to infinity. The third case is a combination of the first two: D goes exactly like k^{2n+1}, and the cross section tends to infinity. We thus have the precise analog of the situation in the central case.

In Sec. 6 the Fredholm determinant of the S matrix is defined, shown to exist, and expressed in terms of D. If there is a half-bound state at $k = 0$, $\det S$ is shown to have the value -1 there. Otherwise it has the value $+1$ at $k = 0$. Eigenphase shifts and their sum δ are introduced, and the latter is expressed in terms of the phase of D. The connection with the known facts in the central case is also established.

Section 7 contains the proof of the generalized Levinson theorem on the basis of the properties of D established earlier. The method is the same as in the central case for one angular momentum, and the result is analogous, except that δ now generally tends to infinity as $k \to \infty$. Again, an extra $\frac{1}{2}\pi$ appears if there is a zero-energy half-bound state.

There are four appendices that provide mathematical details.

2. THE INTEGRAL EQUATION AND THE S MATRIX

We shall assume that the potential is real and that there exist constants[5] $a>0$ and $C<\infty$ such that for all $\mathbf{y}$

$$\int (d\mathbf{x})\,|V(\mathbf{x})|^2+\int (d\mathbf{x})\,|V(\mathbf{x})|\left(\frac{|\mathbf{x}|+|\mathbf{y}|+a}{|\mathbf{x}-\mathbf{y}|}\right)^2\leq C. \tag{2.1}$$

It follows that $V\in L^1(R^3)\cap L^2(R^3)$, that for all $\mathbf{y}$

$$\int (d\mathbf{x})\,|V(\mathbf{x})|/|\mathbf{x}-\mathbf{y}|^2<C/a^2<\infty, \tag{2.2}$$

and that V is in the Rollnik class

$$\int (d\mathbf{x})(d\mathbf{y})\,|V(\mathbf{x})\,V(\mathbf{y})|/|\mathbf{x}-\mathbf{y}|^2<\infty. \tag{2.3}$$

For example, if there exist $0<a<\infty$, $C<\infty$, and $\epsilon>0$ such that

$$|V(\mathbf{x})|<C(a+|\mathbf{x}|)^{-3-\epsilon}, \tag{2.4}$$

then (2.1) holds.

Our starting point is the Lippmann–Schwinger equation[6]

$$\psi(\mathbf{k},\mathbf{x})=\exp(i\mathbf{k}\cdot\mathbf{x})+\int (d\mathbf{y})\,G(k;\mathbf{x},\mathbf{y})\,V(\mathbf{y})\,\psi(\mathbf{k},\mathbf{y}), \tag{2.5}$$

where

$$G(k;\mathbf{x},\mathbf{y})=-\exp(ik|\mathbf{x}-\mathbf{y}|)/4\pi|\mathbf{x}-\mathbf{y}|. \tag{2.6}$$

In the usual manner, we multiply (2.5) by $|V|^{1/2}$ and define $\varphi(\mathbf{k},\mathbf{x})=|V(\mathbf{x})|^{1/2}\psi(\mathbf{k},\mathbf{x})$, so that $\varphi(\mathbf{k},\mathbf{x})$ is subject to the equation

$$\varphi(\mathbf{k},\mathbf{x})=\varphi_0(\mathbf{k},\mathbf{x})+\int (d\mathbf{y})\,K(k;\mathbf{x},\mathbf{y})\,\varphi(\mathbf{k},\mathbf{y}), \tag{2.7}$$

in which

$$\varphi_0(\mathbf{k},\mathbf{x})=|V(\mathbf{x})|^{1/2}\exp(i\mathbf{k}\cdot\mathbf{x}),$$

and the kernel is given by

$$K(k;\mathbf{x},\mathbf{y})=|V(\mathbf{x})|^{1/2}\,G(k;\mathbf{x},\mathbf{y})\,V^{1/2}(\mathbf{y}). \tag{2.8}$$

We have written here as a shorthand

$$V^{1/2}(\mathbf{x})=V(\mathbf{x})/|V(\mathbf{x})|^{1/2}.$$

It is well known that if V is in the Rollnik class (2.3), then for each k in the open upper half of the complex plane, or on the real axis, $\mathrm{Im}k\geq 0$, $K(k;\mathbf{x},\mathbf{y})$ is the kernel of a Hilbert–Schmidt operator K, and[7,8]

$$\lim_{|k|\to\infty}\mathrm{tr}K^2=\lim_{|k|\to\infty}\mathrm{tr}(KK^\dagger)^2=0 \tag{2.9}$$

for all $\mathrm{Im}k\geq 0$. In the same region we have

$$\lim_{|k|\to\infty}\mathrm{tr}K^3=0. \tag{2.9'}$$

This is shown in Appendix D.

It follows that modified Fredholm theory[9] is applicable to (2.7). The modified Fredholm determinant[10] is well defined for each k, $\mathrm{Im}k\geq 0$:

$$D(k)=\det{}_2(1-\lambda K)\big|_{\lambda=1}$$

as an absolutely convergent power series in λ, for $\lambda=1$. Furthermore, for $\mathrm{Im}k\geq 0$

$$\lim_{|k|\to\infty}D(k)=1 \tag{2.10}$$

because of (2.9) and (2.9') (see Appendix D).

The kernel $K(k;\mathbf{x},\mathbf{y})$ defines an analytic, operator-valued function of k, regular in the open upper half-plane. Since

$$\mathrm{tr}\,\frac{dK}{dk}\,\frac{dK^\dagger}{dk}\leq\frac{1}{(4\pi)^2}\left(\int (d\mathbf{x})\,|V(\mathbf{x})|\right)^2<\infty$$

if $V\in L^1$, each term in the power series of $\det_2(1-\lambda K)$ is an analytic function of k, and hence $D(k)$ is an analytic function of k, regular in the open upper half-plane. Its boundary values on the real axis are continuous. (If V decreases exponentially, then D has an analytic continuation to $\mathrm{Im}k<0$.)

The form (2.6) of G shows explicitly that for real k $G^*(k;\mathbf{x},\mathbf{y})=G(-k;\mathbf{x},\mathbf{y})$. Therefore,

$$D(-k)=D^*(k) \tag{2.11}$$

for real k.

The Fredholm alternative assures us that if for $k=k_0$

$$D(k_0)=0, \tag{2.12}$$

then the homogeneous form of (2.7)

$$\varphi(\mathbf{x})=\int (d\mathbf{y})\,K(k_0;\mathbf{x},\mathbf{y})\,\varphi(\mathbf{y}) \tag{2.13}$$

has a solution $\varphi\in L^2$. In that case (2.7) has a solution in L^2 if and only if φ_0 (which, for real k, is in L^2 if V is in L^1) is orthogonal to all solutions φ' of the equation

$$K^\dagger\varphi'=\varphi'.$$

This condition reads explicitly[11]

$$\int (d\mathbf{x})\exp(ik_0\hat{k}\cdot\mathbf{x})\,|V(\mathbf{x})|^{1/2}\,\varphi(\mathbf{x})=0. \tag{2.14}$$

Only for directions $\hat{k}$ for which (2.14) holds, has (2.7) a solution $\varphi(k_0\hat{k},\mathbf{x})\in L^2$. We shall call points k_0 with $\mathrm{Im}k_0\geq 0$ for which (2.12) holds *exceptional points*.

For real values of k that are not exceptional, (2.7) has a unique solution $\varphi(\mathbf{k},\mathbf{x})\in L^2$. Because of (2.2) it follows that $\psi(\mathbf{k},\mathbf{x})$, defined by

$$\psi(\mathbf{k},\mathbf{x})=\exp(i\mathbf{k}\cdot\mathbf{x})+\int (d\mathbf{y})\,G(k;\mathbf{x},\mathbf{y})\,V^{1/2}(\mathbf{y})\,\varphi(\mathbf{k},\mathbf{y})$$

exists pointwise for all $\mathbf{x}$, is continuous in $\mathbf{k}$, and satisfies (2.5) as well as, at least in a generalized sense, the Schrödinger equation

$$[-\Delta+V]\psi=k^2\psi.$$

Its asymptotic form for large $|\mathbf{x}|$ is

$$\psi(\mathbf{k},\mathbf{x})=\exp(i\mathbf{k}\cdot\mathbf{x})+(\exp(ik|\mathbf{x}|)/|\mathbf{x}|)\,T(\hat{x},\mathbf{k})+o(|x|^{-1}), \tag{2.15}$$

where, with $\mathbf{k}'=k\hat{k}$ and $v(\mathbf{x})=V(\mathbf{x})/|V(\mathbf{x})|$,

$$\begin{aligned}T(k;\hat{k}',\hat{k})&=-\frac{1}{4\pi}\int (d\mathbf{x})\exp(-i\mathbf{k}'\cdot\mathbf{x})\,V(\mathbf{x})\,\psi(\mathbf{k},\mathbf{x})\\&=-\frac{1}{4\pi}\int (d\mathbf{x})\,\varphi_0^*(\mathbf{k}',\mathbf{x})\,v(\mathbf{x})\,\varphi(\mathbf{k},\mathbf{x}).\end{aligned} \tag{2.16}$$

Thus the scattering amplitude, or T matrix, $T(k;\hat{k}',\hat{k})$ is well-defined and continuous for all $\hat{k}$ and $\hat{k}'$ for each real nonexceptional value of k.

The differential scattering cross section is given by

$$\frac{d\sigma}{d\Omega}(k;\hat{k}',\hat{k})=|T(k;\hat{k}',\hat{k})|^2 \tag{2.17}$$

and is thus well defined too and finite for each nonexceptional real k and all $\hat{k}$ and $\hat{k}'$.

The S matrix is defined in terms of the scattering operator so that for all f in L^2 and almost all $\mathbf{k}$

$$(Sf)(\mathbf{k}) = \int d\hat{k}'\, S(k;\hat{k},\hat{k}')f(k\hat{k}'), \tag{2.18}$$

and it is expressed in terms of T as[12]

$$S(k;\hat{k},\hat{k}') = \delta(\hat{k},\hat{k}') + (ik/2\pi)\,T(k;\hat{k},\hat{k}'). \tag{2.19}$$

If the potential is both in L^1 and in the Rollnik class (2.3), then the scattering operator is known to be unitary.[13] We may write this statement in the form that for all $f \in L^2$

$$\begin{aligned}\int (d\mathbf{k})\,|f(\mathbf{k})|^2 &= \int_0^\infty dk\,k^2 \int d\hat{k}\,\Big|\int d\hat{k}'\,S(k;\hat{k},\hat{k}')f(k;\hat{k}')\Big|^2 \\ &= \int_0^\infty dk\,k^2 \int d\hat{k}\,\Big|\int d\hat{k}'\,S^*(k;\hat{k}',\hat{k})f(k;\hat{k}')\Big|^2,\end{aligned} \tag{2.20}$$

and it implies that the S matrix is unitary for almost all k. Consequently, T must obey the generalized optical theorem for almost all k:

$$\begin{aligned}&T(k;\hat{k}',\hat{k}) - T^*(k;\hat{k},\hat{k}') \\ &\quad = (ik/2\pi)\int d\hat{k}''\,T(k;\hat{k}',\hat{k}'')\,T^*(k;\hat{k},\hat{k}'') \\ &\quad = (ik/2\pi)\int d\hat{k}''\,T^*(k;\hat{k}'',\hat{k}')\,T(k;\hat{k}'',\hat{k}).\end{aligned} \tag{2.21}$$

Continuity implies that these equations must hold for all real nonexceptional values of k.

3. NEGATIVE-ENERGY BOUND STATES

For $\mathrm{Im}k > 0$ the operator GV is Hilbert–Schmidt (because $V \in L^2$). Hence, if k_0 with $\mathrm{Im}k_0 > 0$ is an exceptional point (2.12), then there exists a nontrivial solution $\psi \in L^2$ of the equation

$$\psi(\mathbf{x}) = -\int (d\mathbf{y})\,\frac{\exp(ik_0|\mathbf{x}-\mathbf{y}|)}{4\pi|\mathbf{x}-\mathbf{y}|}\,V(\mathbf{y})\,\psi(\mathbf{y}). \tag{3.1}$$

Thus k_0^2 is an eigenvalue, or bound state. It follows that k_0 must lie on the imaginary axis.

Conversely, if $k_0^2 < 0$ is an eigenvalue of $H = -\Delta + V$, then each eigenfunction ψ must satisfy the homogeneous form (3.1) of (2.5). It follows that $D(k_0) = 0$. Thus there is a one-to-one correspondence between the negative-energy bound states and the zeros of D in the upper half-plane.

The multiplicity of the zero of $D(k)$ at $k = k_0$, $\mathrm{Im}k_0 > 0$, equals the degeneracy of the eigenvalue k_0^2. This is shown as follows[14]:

$$\begin{aligned}\frac{d}{dk}\ln\det{}_2(1-K) &= \frac{d}{dk}\ln\det{}_2(1-GV) \\ &= -\mathrm{tr}\left[(1-GV)^{-1}\frac{dG}{dk}\,VGV\right] = 2k\,\mathrm{tr}\mathcal{G}(GV)^2,\end{aligned}$$

where

$$\mathcal{G} = (1-GV)^{-1}G = (k^2 - H_0 - V)^{-1}$$

is the resolvent of $H = H_0 + V$ and all differentiations are justified by the absolute convergence of the series and integrals. Now let $D(k)$ have a zero of order p at $k = k_0$, and let P be the projection onto the n-dimensional eigenspace of H at k_0^2, so that $GVP = P$. Then

$$p = \lim_{k\to k_0}(k-k_0)\frac{d}{dk}\ln D(k) = \mathrm{tr}P(GV)^2 = \mathrm{tr}P = n$$

because $\mathcal{G}$, as the resolvent of the self-adjoint operator H, has a simple pole at $k^2 = k_0^2$ whose residue is P.

4. REAL EXCEPTIONAL $k_0 \neq 0$

Suppose that $k_0 \neq 0$, with $\mathrm{Im}k_0 = 0$, is an exceptional point (2.12). Then (2.13) has a nontrivial solution $\varphi \in L^2$ of the form $\varphi = |V|^{1/2}\psi$, where ψ obeys Eq. (3.1) pointwise for all $\mathbf{x}$. On the assumption that V satisfies (2.1), the asymptotic form of ψ for large $|\mathbf{x}|$ is

$$\begin{aligned}\psi(\mathbf{x}) = &-(\exp(ik_0|\mathbf{x}|)/4\pi|\mathbf{x}|)\int (d\mathbf{y})\exp(-ik_0\hat{x}\cdot\mathbf{y})\,V(\mathbf{y})\,\psi(\mathbf{y}) \\ &+ h(\mathbf{x}),\end{aligned} \tag{4.1}$$

where $h(\mathbf{x}) \in L^2(R^3)$, as is shown in Appendix A. Therefore, $\psi(\mathbf{x})$ is normalizable if and only if for almost all directions $\hat{n}$

$$\int (d\mathbf{y})\exp(ik_0\hat{n}\cdot\mathbf{y})\,V(\mathbf{y})\,\psi(\mathbf{y}) = 0. \tag{4.2}$$

We shall call a real exceptional point k_0 for which a nontrivial solution of (3.1) exists such that (4.2) holds for almost all $\hat{n}$ an *exceptional point of the first kind*. If a solution of (3.1) exists for which (4.2) fails for a set of directions $\hat{n}$ of positive Lebesgue measure, we call k_0 an *exceptional point of the second kind*.[15] We see that if $k_0 \neq 0$ is a real exceptional point of the first kind, then k_0^2 is a positive-energy bound state.

Conversely, if $k_0^2 > 0$ is an eigenvalue, then (3.1) must have a nontrivial solution and (2.12) must hold, as well as (4.2). Hence k_0^2 is a positive-energy bound state if and only if k_0 is an exceptional point of the first kind.

Now note that (4.2) is identical with (2.14). Therefore, if the exceptional point is a bound state then (2.5) still has a well-defined solution for almost all $\hat{k}$. Of course, this solution is not unique since it may be augmented by any multiple of a solution of (3.1). However, if the solution of (2.5) is inserted in (2.16) for the calculation of the scattering amplitude, then (4.2) assures that the ambiguity of $\psi(\mathbf{k},\mathbf{x})$ causes an ambiguity in T in almost no direction. We are therefore free to define T continuously even at $k = k_0$.

It was proved by Kato[16] that potentials in the class (2.4) produce no positive energy bound states. Hence, if there are real exceptional points $k_0 \neq 0$, then they must be of the second kind.

Suppose now that $k_0 \neq 0$ is a real exceptional point of the second kind, so that there are solutions of (3.1) that violate (4.2) for all directions $\hat{n}$ in a set Ω of positive Lebesgue measure. Then, for all $\hat{k} \in \Omega$, (2.5) has no solution, and the scattering amplitude for those values of $\hat{k}$ does not exist. Let us now fix k_0 at an exceptional value and, for real λ, consider the equation

$$\varphi^{(\lambda)} = \varphi_0 + \lambda K\varphi^{(\lambda)}, \tag{4.3}$$

which goes over into (2.7) as $\lambda \to 1$. The function φ_0 depends on $\hat{k}$, and we fix it at a value for which $\varphi = \varphi^{(1)}$ does not exist. (Since k_0 is an exceptional point of the second kind, there exists a set of positive measure of such $\hat{k}$.) We have

$$\varphi^{(\lambda)} = (1-\lambda K)^{-1}\varphi_0. \tag{4.4}$$

Because K is compact, the resolvent $(1-\lambda K)^{-1}$ is a meromorphic function of λ with a pole of finite order at $\lambda=1$. Let the pole of $(1-\lambda K)^{-1}\varphi_0$ be of order N. Then

$$\lim_{\lambda\to 1}(\lambda-1)^N\varphi^{(\lambda)}=\varphi \tag{4.5}$$

exists, fails to vanish for almost all $\mathbf{x}$, and satisfies the homogeneous equation

$$K\varphi=\varphi.$$

There must exist a set of directions Ω of positive measure such that for all $\hat{n}\in\Omega$

$$\int (d\mathbf{x})\exp(-ik_0\hat{n}\cdot\mathbf{x})\,|V(\mathbf{x})|^{1/2}\varphi'(\mathbf{x})\neq 0, \tag{4.6}$$

where

$$\varphi'=\varphi^* v$$

satisfies

$$K^\dagger\varphi'=\varphi'.$$

Since φ is of the form $|V|^{1/2}\psi$, where ψ satisfies (3.1), (4.6) means that

$$\int (d\mathbf{x})\exp(ik_0\hat{n}\cdot\mathbf{x})\,V(\mathbf{x})\,\psi(\mathbf{x})\neq 0 \tag{4.7}$$

for all $\hat{n}\in\Omega$.

We now form the T matrix $T^{(\lambda)}$ out of $\varphi^{(\lambda)}=|V|^{1/2}\psi^{(\lambda)}$ as in (2.16). It satisfies the optical theorem (2.21) with $\hat{k}'=\hat{k}$ for all $\lambda\neq 1$. Multiplying (2.21) by $(\lambda-1)^{2N}$ and letting $\lambda\to 1$, we obtain

$$\int d\hat{k}''\,|\tilde{T}(k_0;\hat{k}'',\hat{k})|^2=0, \tag{4.8}$$

where

$$\begin{aligned}\tilde{T}(k_0;\hat{k}'',\hat{k})&=\lim_{\lambda\to 1}(\lambda-1)^N T^{(\lambda)}(k_0;\hat{k}'',\hat{k})\\&=-(1/4\pi)\int (d\mathbf{x})\exp(-ik_0\hat{k}''\cdot\mathbf{x})\,V(\mathbf{x})\,\psi(\mathbf{x}).\end{aligned} \tag{4.9}$$

Equation (4.8) contradicts (4.7). Consequently, our supposition that k_0 is an exceptional point of the second kind is false. We have thus proved the following:

Theorem[17a]: If the potential satisfies (2.1), then the set of real exceptional points of the second kind contains at most the point $k=0$.

Together with Kato's proof of the absence of positive-energy bound states, this leads to the

Corollary: For potentials that satisfy (2.4), $D(k)\neq 0$ for all real $k\neq 0$.

Let us make a connection between the exceptional points $k_0\neq 0$ on the real axis, and the structure of the spectrum of H. We expect the set of solutions $\psi(\mathbf{k},\mathbf{x})$ of (2.5) and of the bound states $\psi_n(\mathbf{x})$ to be complete in the sense that for every $f\in L^2$

$$\int d\mathbf{x}\,|f(x)|^2=\sum|\alpha_n|^2+\int (d\mathbf{k})\,|\tilde{f}(\mathbf{k})|^2, \tag{4.10}$$

where

$$\alpha_n=\int (d\mathbf{x})\,\psi_n^*(\mathbf{x})f(\mathbf{x}),\quad \tilde{f}(\mathbf{k})=\int (d\mathbf{x})\,\psi^*(\mathbf{k},\mathbf{x})f(\mathbf{x}).$$

However, if there are real exceptional points $k_0\neq 0$ of the second kind, then this cannot be true, as $\psi(\mathbf{k},\mathbf{x})$ is of the form

$$\psi(\mathbf{k},\mathbf{x})=\tilde{\psi}(\mathbf{k},\mathbf{x})/D(k)$$

and $D(k)$ tends to zero at least as $(k-k_0)$ for $k\to k_0$.[17b] Thus the integral on the right-hand side of (4.10) will diverge at $k=k_0$, at least for some functions $f\in L^2$. The set of real exceptional points $k_0\neq 0$ of the second kind therefore presumably constitutes the singular continuous spectrum of H. This presumption is strengthened by Simon's proof[18] that the singular spectrum is part of the exceptional set, and by its incompatibility with the unitarity of the S matrix.[19] If the exceptional set is countable, then the singular spectrum is known to be empty.[20] Analytic properties of the potential also have been proved to be sufficient to rule out a singular spectrum.[21]

In view of Ref. 2, p. 117, the above corollary implies that for potentials in the class (2.4), the singular continuous spectrum is empty. We also note that it implies strong asymptotic completeness in Simon's terminology.[22] These conclusions for potentials in the class (2.4) are not new.[23]

5. THE POINT $k=0$

Let us write

$$G=G_0+G'=G_0+ikG_1+G''=G_0+ikG_1+k^2G_2+G''', \tag{5.1}$$

where

$$G_0(\mathbf{x},\mathbf{y})=-(1/4\pi)(1/|\mathbf{x}-\mathbf{y}|), \tag{5.2}$$

$$G_1(\mathbf{x},\mathbf{y})=-1/4\pi, \tag{5.3}$$

$$G_2(\mathbf{x},\mathbf{y})=|\mathbf{x}-\mathbf{y}|/8\pi, \tag{5.4}$$

and one readily finds that for $\mathrm{Im}k\geq 0$

$$|G'(k;\mathbf{x},\mathbf{y})|\leq C|k|/1+|k||\mathbf{x}-\mathbf{y}|, \tag{5.5}$$

$$|G''(k;\mathbf{x},\mathbf{y})|\leq C|k|^2|\mathbf{x}-\mathbf{y}|/(1+|k||\mathbf{x}-\mathbf{y}|), \tag{5.5'}$$

$$|G'''(k;\mathbf{x},\mathbf{y})|\leq C|k|^3|\mathbf{x}-\mathbf{y}|^2/(1+|k||\mathbf{x}-\mathbf{y}|) \tag{5.5''}$$

for some constant C. Because $-G_0$ is the kernel of a positive semidefinite operator, we may define

$$A=(-G_0)^{1/2} \tag{5.6}$$

as positive semidefinite, and

$$B=A^{-1}GVA. \tag{5.7}$$

Corresponding to (5.1), we then have

$$B=B_0+B_R, \tag{5.8}$$

where

$$B_0=-AVA=B_0^\dagger \tag{5.9}$$

and

$$B_R=ikB_1+R_1=ikB_1+k^2B_2+R_2. \tag{5.10}$$

If V satisfies (2.3), then B_0 is Hilbert–Schmidt.

Note that since the power-series expansion of $\det_2(1-K)$ contains K only in the form of $\mathrm{tr}K^n$, $n=2,\cdots$, we have

$$D(0)=\det_2(1-K_0)=\det_2(1-B_0). \tag{5.11}$$

Assume now that $k=0$ is an exceptional point,

$$D(0)=0. \tag{5.12}$$

Then there exist functions $\chi\in L^2$ and $\varphi\in L^2$ such that

$$K_0\varphi=\varphi, \tag{5.13}$$

$$B_0\chi=\chi. \tag{5.13'}$$

To every solution $\chi\in L^2$ of (5.13′) there corresponds a solution

$$\varphi=|V|^{1/2}(-G_0)^{1/2}\chi \tag{5.14}$$

of (5.13), and

$$(\varphi,\varphi)=(\chi,(-G_0)^{1/2}|V|(-G_0)^{1/2}\chi)<\infty$$

because $\mathrm{tr}G_0|V|G_0|V|<\infty$. Conversely, to every solution $\varphi\in L^2$ of (5.13) there corresponds a solution

$$\chi=(-G_0)^{1/2}V^{1/2}\varphi \tag{5.14'}$$

of (5.13′), and

$$(\chi,\chi)=-(\varphi,V^{1/2}G_0V^{1/2}\varphi)<\infty.$$

We define $\psi=G_0V^{1/2}\varphi=-(-G_0)^{1/2}\chi$, or, explicitly,

$$\psi(\mathbf{x})=\int(d\mathbf{y})\,G_0(\mathbf{x},\mathbf{y})V^{1/2}(\mathbf{y})\,\varphi(\mathbf{y}), \tag{5.15}$$

which means that

$$\psi(\mathbf{x})=\int(d\mathbf{y})\,G_0(\mathbf{x},\mathbf{y})\,V(\mathbf{y})\,\psi(\mathbf{y}). \tag{5.16}$$

As before, it follows from the fact that $\varphi\in L^2$ and the assumptions (2.2) and $V\in L^1$ that $\psi(\mathbf{x})$ is well-defined pointwise for every $\mathbf{x}$ and satisfies the differential equation

$$\Delta\psi=V\psi \tag{5.17}$$

at least in a generalized sense. Furthermore, it has the asymptotic form

$$\psi(\mathbf{x})=-(1/4\pi|\mathbf{x}|)\int(d\mathbf{y})\,V(\mathbf{y})\,\psi(\mathbf{y})+h(\mathbf{x}), \tag{5.18}$$

where $h\in L^2(R^3)$, because of assumption (2.1). (See Appendix A.)

Consequently, $\psi(\mathbf{x})$ is in L^2 if and only if

$$\int(d\mathbf{x})\,V(\mathbf{x})\,\psi(\mathbf{x})=0. \tag{5.19}$$

We shall refer to a function ψ that satisfies (5.16) but not (5.19) as a *half-bound state*. If such a ψ exists, $k=0$ is an exceptional point of the second kind.

Suppose there are two linearly independent half-bound states at $k=0$. Then one can always form a linear combination that satisfies (5.19) and hence is a bound state. If there are n linearly independent solutions of (5.16), we can therefore assume without loss of generality that they have been so arranged that at most one of them violates (5.19). They can also be assumed to be real functions.

Thus, if 1 is an eigenvalue of K_0 and B_0 then there are the following three possibilities:

(1) (5.13′) has $n\geq 1$ linearly independent solutions χ_m, $m=1,\ldots,n$, and all of the corresponding functions $\psi_m=A\chi_m$ obey (5.19). Then 1 is an n-fold degenerate eigenvalue of G_0V and $k=0$ is an exceptional point of the first kind. Because B_0 is compact, n is necessarily finite.

(2) (5.13′) has exactly one nontrivial solution χ and for the corresponding function $\psi=A\chi$, (5.19) does not hold. Then 1 is not an eigenvalue of G_0V, ψ is a half-bound state, and $k=0$ is an exceptional point of the second kind.

(3) (5.13′) has $n+1>1$ linearly independent solutions χ_m, $m=1,\ldots,n+1$, such that the functions ψ_m, $m=1,\ldots,n$, that correspond to the first n χ's obey (5.19), whereas ψ_{n+1}, corresponding to χ_{n+1}, does not. Then 1 is an n-fold degenerate eigenvalue of G_0V and, in addition, there is a half-bound state. Now $k=0$ is an exceptional point of the first and second kind. We may assume that the n solutions χ_m, $m=1,\ldots,n$, of (5.13′) are mutually orthogonal and normalized to unity. But we cannot necessarily expect them to be orthogonal to χ_{n+1}.

Let P be the orthogonal projection on the eigenspace of B_0 at the eigenvalue 1, and $Q=1-P$, so that $B_0P=P$ and $\det_2(1-B_0Q)\neq 0$. Because B_0 is self-adjoint, it commutes with P, $[P,B_0]=0$. Now

$$\begin{aligned}(1-B)&=1-B_0P-B_0Q-B_R\\&=(1-B_0Q)\{1-[1-(1-B_0Q)^{-1}B_R]^{-1}P\}\\&\quad\times[1-(1-B_0Q)^{-1}B_R],\end{aligned}$$

and

$$\begin{aligned}D(k)&=\det{}_2(1-B_0Q)\exp(\mathrm{tr}(P+B_R))\det[1-(1-B_0Q)^{-1}B_R]\\&\quad\times\det\{1-[1-(1-B_0Q)^{-1}B_R]^{-1}P\}.\end{aligned} \tag{5.20}$$

The first factor on the right is real and different from zero. The second approaches $\exp(\mathrm{tr}P)$ as $k\to 0$ because (5.5) shows that

$$\mathrm{tr}B_R=O(k) \tag{5.21}$$

if $V\in L^1$. Because the same hypothesis and (5.5) leads to

$$\mathrm{tr}B_RB_R^{\dagger}=O(k^2), \tag{5.22}$$

the third factor tends to unity. The last factor equals

$$\begin{aligned}\det\{\ \}&=\det{}_P\{1-[1-(1-B_0Q)^{-1}B_R]^{-1}\}\\&=\det{}_P\{B_R[1-(1-B_0Q)^{-1}B_R]^{-1}\},\end{aligned}$$

where[24] $\det_P$ denotes the (finite-dimensional) determinant on the range of P. Because of (5.22), the behavior of $D(k)$ near $k=0$ is therefore determined by $\det_PB_R$. We must now examine the three possibilities denumerated above. Let us first examine

Case 2: In this case the range of P is one dimensional, and (5.19) does not hold. Then

$$\begin{aligned}\det{}_PB_R&=ik(\chi,B_1\chi)+O(k^2)\\&=-ik(V\psi,G_1V\psi)+O(k^2)\end{aligned} \tag{5.23}$$

because (5.13′) implies that $\psi=-A\chi$ and $\chi=AV\psi$. That the remainder is $O(k^2)$ is shown in Appendix B. Equation (5.3) and the violation of (5.19) therefore show that

$$D(k)=ikc+o(k), \tag{5.24}$$

where c is real and $c\neq 0$.

By the same arguments given in Sec. 4 for real exceptional points $k_0\neq 0$ of the second kind, it follows from the violation of (5.19) that (2.5) for $k=0$ has no solution and that the scattering cross section is infinite at $k=0$.

Case 1: If (5.19) holds, then B_1 contributes nothing to $\det_P B_R$ and the leading term comes from B_2:

$$\det_P B_R = \det_P(k^2 B_2) + \cdots = k^{2n} \det_P B_2 + o(k^{2n}). \quad (5.25)$$

That the remainder is $o(k^{2n})$ follows from (2.1), (5.18), and (5.5″) as is shown in Appendix B. We now make use of the fact that $A^{-2} = \nabla^2$ and therefore

$$A^{-2} G_2 = -G_0.$$

Consequently

$$(\chi_m, B_2 \chi_{m'}) = (\psi_m, A^{-2} G_2 V \psi_{m'}) = -(\psi_m, G_0 V \psi_{m'})$$
$$= -(\psi_m, \psi_{m'}),.$$

and hence

$$\det_P B_R = (-1)^n \det(\psi_m, \psi_{m'}) \neq 0.$$

As a result

$$D(k) = k^{2n} c + o(k^{2n}), \quad (5.26)$$

where c is real and $c \neq 0$.

By the same arguments given in Sec. 4 for real exceptional points $k_0 \neq 0$ of the first kind, (2.5) for $k = 0$ now has a well-defined solution and the scattering cross section is finite.

Case 3: If ψ_m, $m = 1, \ldots, n$, satisfy (5.19) but ψ_{n+1} does not and if χ_m, $m = 1, \ldots, n+1$, are the corresponding solutions of (5.13′), then

$$(\chi_m, B_1 \chi_{m'}) = -(\int V \psi_{n+1})^2 / 4\pi, \quad m = m' = n+1,$$
$$= 0, \quad \text{otherwise},$$

and

$$(\chi_m, B_2 \chi_{m'}) = -(\psi_m, \psi_{m'}).$$

We use these vectors to evaluate the determinant, even though χ_{n+1} is not necessarily orthogonal to χ_m, $m = 1, \ldots, n$.[25] As a result[26]

$$D(k) = \text{const } \det_P(ikB_1 + k^2 B_2 + R_2)$$
$$= ik^{2n+1} c + o(k^{2n+1}) \quad (5.27)$$

where c is real and $c \neq 0$. Again it follows that the scattering cross section at $k = 0$ is infinite.

We may also examine the behavior of the solution $\psi(\mathbf{k}, \mathbf{x})$ of (2.5) near $k = 0$. As we have already seen, in case 1 the function remains well defined and finite at $k = 0$ because of (5.19). It is shown in Appendix C that in Cases 2 and 3, that is, whenever there is a half-bound state at $k = 0$,

$$\lim_{k \to 0} D(k) \psi(\mathbf{k}, \mathbf{x}) = \tilde{\psi}(0, \mathbf{x}) \quad (5.28)$$

exists in the sense that $|V(\mathbf{x})|^{1/2} \tilde{\psi}(0, \mathbf{x})$ exists almost everywhere as a function in L^2 with positive norm.

The spectral expansion of the resolvent $(k^2 - H)^{-1}$ is given by

$$\mathcal{G}(k; \mathbf{x}, \mathbf{y}) = (2\pi)^{-3} \int (d\mathbf{k}') \frac{\psi(\mathbf{k}', \mathbf{x})\, \psi^*(\mathbf{k}', \mathbf{y})}{k^2 - k'^2}$$
$$+ \sum_n \frac{\psi_n(\mathbf{x})\, \psi_n^*(\mathbf{y})}{k^2 - k_n^2} \quad (5.29)$$

for $\text{Im} k > 0$. If $k = 0$ is an exceptional point of the second kind, i.e., if there is a half-bound state (as well as possibly being exceptional of the first kind, so that there are also $k = 0$ bound states), then it follows from (5.28) that the Green's function (5.29) diverges as k^{-1} near $k = 0$.

We note that these results are exactly analogous to the well-known situation for central potentials. In that case (5.19) is trivially satisfied whenever ψ is a wavefunction of angular momentum $l > 0$. It is less easy to see from the three-dimensional point of view why (5.19) is never true when ψ has angular momentum $l = 0$.

There is one important property of central potentials that has not been generalized: If ψ is the *first* "bound state," then (5.19) does not hold. In other words, the following proposition is valid for central potentials with finite first and second absolute moments: *If zero is an exceptional point and there are no negative eigenvalues, then* (5.19) *does not hold.* I have not succeeded in proving this proposition for noncentral potentials and do not know if it is generally valid.

6. THE DETERMINANT OF THE S MATRIX

At a fixed value of k, the S matrix may be regarded as an integral operator on the unit sphere, and its kernel is given in (2.19). It is unitary. Since T was shown in Secs. 2 and 4 to be well defined and finite for all $\hat{k}$ and $\hat{k}'$ for each $k > 0$, it is the kernel of a trace-class operator for each $k > 0$. The Fredholm determinant of the S matrix is therefore well defined:

$$\det S = \det[1 + (ik/2\pi)\, T]. \quad (6.1)$$

We now calculate[27,28]

$$D^*(k) = \det_2(1 - G^\dagger V)$$
$$= \det_2\{(1 - GV)[1 - (1 - GV)^{-1}(G^\dagger - G)\, V]\}$$
$$= \det_2(1 - GV)\, \{\exp[2\pi i \,\text{tr}\, V \delta(k^2 - H_0)]\}$$
$$\times \det\{1 - (1 - GV)^{-1} 2\pi i \delta(k^2 - H_0)\, V\}, \quad (6.2)$$

where $\delta(k^2 - H_0)$ is the operator whose kernel is

$$\Delta(\mathbf{x}, \mathbf{y}) = \tfrac{1}{2} k (2\pi)^{-3} \int d\hat{k}' \exp(ik\hat{k}' \cdot (\mathbf{x} - \mathbf{y})).$$

Then $(1 - GV)^{-1} \delta(k^2 - H_0)\, V$ has the kernel $-kM(\mathbf{x}, \mathbf{y})/(2\pi)^2$ where

$$M(\mathbf{x}, \mathbf{y}) = -(1/4\pi) \int d\hat{k}' \psi(k\hat{k}', \mathbf{x}) \exp(-ik\hat{k}' \cdot \mathbf{y})\, V(\mathbf{y}).$$

Because one readily sees that

$$\text{tr} M^n = \text{tr} T^n,$$

where the trace on the right is over the unit sphere, it follows that

$$\det\{1 - 2\pi i (1 - GV)^{-1} \delta(k^2 - H_0)\, V\} = \det[1 + (ik/2\pi) T] = \det S. \quad (6.3)$$

Furthermore,

$$\text{tr} V \delta(k^2 - H_0) = k (2\pi)^{-2} \int (d\mathbf{x})\, V(\mathbf{x})$$

and (6.2) and (6.3) yield[29]

$$\det S(k) = \exp[-(ik/2\pi) \int (d\mathbf{x}) V]\, D^*(k)/D(k). \quad (6.4)$$

For large k the dominant term in T is

$$T(k;\hat{k}',\hat{k}) = -(1/4\pi)\int (d\mathbf{x})\, V(\mathbf{x}) \exp(ik(\hat{k}-\hat{k}')\cdot\mathbf{x} + o(1),$$

and therefore the behavior of $\det S$ is as

$$\ln \det S = -(ik/2\pi)\int (d\mathbf{x})\, V(\mathbf{x}) + O(1).$$

In view of (2.10) the exponential factor in (6.4) is thus needed to give the right-hand side the correct behavior for large k.

If $k=0$ is exceptional, then the results of Sec. 5, together with (6.4), show directly that if there is a half-bound state, then

$$\det S(0) = -1, \tag{6.5}$$

whereas otherwise

$$\det S(0) = 1. \tag{6.5'}$$

Note that it follows from the fact that $D(0)$ is real [because of (2.11)] that (6.5') holds also whenever $k=0$ is not exceptional.

Because S is unitary, we may define, $\mathrm{mod}\pi$, a real number δ by

$$\det S = \exp(2i\delta). \tag{6.6}$$

T being compact, its spectrum consists of a denumerable set of point eigenvalues only, and the eigenvalues of S may be written in the form $\exp(2i\delta_n)$, $n=1,\ldots,\infty$. They accumulate at 1; hence the δ_n may be defined so that they accumulate at zero. These are the *eigenphase shifts*. We then have

$$\delta = \sum_{n=1}^{\infty} \delta_n (\mathrm{mod}\pi). \tag{6.7}$$

The result (6.4) implies that we may define

$$\delta(k) = -(k/4\pi)\int (d\mathbf{x})\, V(\mathbf{x}) - \eta(k), \tag{6.8}$$

where

$$\eta(k) = \arg D(k).$$

It follows from (6.5') that, unless there is a zero-energy half-bound state, $\delta(0) = m\pi$, where m is an integer. If there is a half-bound state, then (6.5) shows that $\delta(0) = (m+\frac{1}{2})\pi$. The significance of these integers is the subject of the generalized Levinson theorem.

We note that if V is central, the eigenphase shifts become the ordinary phase shifts δ_l, and an eigenvalue of angular momentum l has a $(2l+1)$-fold degeneracy. Hence

$$\delta = \sum_{0}^{\infty} (2l+1)\,\delta_l. \tag{6.9}$$

The relation between $D(k)$ and the Jost function is such that[30]

$$\eta = -\sum_l (2l+1)\{\delta_l + k^{-1}\int_0^\infty dr\, V(r)[u_l(kr)]^2\}$$

$$= -\delta - (k/4\pi)\int (d\mathbf{x})\, V(|\mathbf{x}|),$$

which is identical with (6.8).

7. THE GENERALIZED LEVINSON THEOREM

The generalized Levinson theorem is now proved by the same method as in the central case.[31] The function $D(k)$ is analytic in the upper half-plane, continuous on the real axis, and it satisfies (2.10) as well as (2.11). It has no zeros on the real axis, except possibly at $k=0$. The zeros of $D(k)$ in $\mathrm{Im}k>0$ are the bound states, and their multiplicity equals the degeneracy. We may therefore proceed in the standard method by evaluating the contour integral

$$(1/2\pi i)\int_C d\ln D(k) = n$$

over a contour C along the real axis from $-R$ to $+R$, avoiding the origin by a small semicircle of radius ϵ in the upper half-plane, and closing it in the upper half-plane by a large semicircle of radius R. Then n is the number of bound states, counted m times if m-fold degenerate. Evaluation of the integral in the limit as $\epsilon \to 0$ and $R \to \infty$ gives

$$\eta(0) - \eta(\infty) = -\pi(n + \tfrac{1}{2}q)$$

if we define $\eta(-k) = -\eta(k)$, as we may because of (2.11). Here n is the number of bound states, including those at $k=0$, and $q=1$ if there is a half-bound state at $k=0$; otherwise $q=0$. Again, each bound state is counted as many times as it is degenerate.

The generalized Levinson theorem thus reads as follows[32]:

$$\delta(0) - \lim_{k\to\infty}[\delta(k) + (k/4\pi)\int (d\mathbf{x})\, V] = \pi(n + \tfrac{1}{2}q). \tag{7.1}$$

At the end of Sec. 6 we observed that there exists an integer m such that

$$\delta(0) = \pi(m + \tfrac{1}{2}q). \tag{7.2}$$

Comparison with (7.1) now allows us to conclude that there must exist an integer p such that

$$\lim_{k\to\infty}[\delta(k) + (k/4\pi)\int (d\mathbf{x})\, V] = \pi p. \tag{7.3}$$

We may choose $p=0$ and thereby specify $\delta(k)$ uniquely. In that case the generalized Levinson theorem[33] is simply (7.2), where m is the total number of bound states, including those at $k=0$ and counting each as often as its degeneracy, and where $q=1$ if there is a half-bound state at $k=0$, and $q=0$ otherwise. Equation (7.3) appears to be new.

After this paper was written, I learned of an article by Wollenberg[34] with similar results. It is, however, less explicit and based on much more restrictive assumptions.

ACKNOWLEDGMENTS

I am happy to acknowledge useful contributions by Mr. David Moore, as well as instructive correspondence with Barry Simon and Tommy Dreyfus.

APPENDIX A

We want to demonstrate (4.1). It follows from (2.2) that ψ is bounded. Hence by (3.1) and (2.1)

$$|\psi(\mathbf{x})| \le \frac{C}{b+|\mathbf{x}|}\int (d\mathbf{y})\,\frac{|V(\mathbf{y})|(b+|\mathbf{x}|)|\mathbf{x}-\mathbf{y}|}{|\mathbf{x}-\mathbf{y}|^2} \le \frac{C'}{b+|\mathbf{x}|} \tag{A1}$$

for some $b>0$, and therefore by (4.1)

$$\psi(\mathbf{x}) = \frac{\exp(ik_0|\mathbf{x}|)}{4\pi|\mathbf{x}|}\int(d\mathbf{y})\exp(-ik_0\hat{x}\circ\mathbf{y})\,V(\mathbf{y})\,\psi(\mathbf{y}) + h(\mathbf{x}), \tag{A2}$$

where $h = h_1 + h_2$,

$$h_1(\mathbf{x}) = -\int\frac{(d\mathbf{y})}{4\pi}\,V(\mathbf{y})\,\psi(\mathbf{y})\exp(ik_0|\mathbf{x}-\mathbf{y}|)\left(\frac{1}{|\mathbf{x}-\mathbf{y}|} - \frac{1}{|\mathbf{x}|}\right), \tag{A3}$$

$$h_2(\mathbf{x}) = \frac{-1}{4\pi|\mathbf{x}|}\int(d\mathbf{y})\,V(\mathbf{y})\,\psi(\mathbf{y})\,[\exp(ik_0|\mathbf{x}-\mathbf{y}|) - \exp(ik_0(|\mathbf{x}| - \hat{x}\cdot\mathbf{y}))]. \tag{A4}$$

From (A3)

$$\int(d\mathbf{x})\,|h_1|^2 \le \frac{1}{16\pi^2}\int(d\mathbf{x})\left[\int(d\mathbf{y})\left|V(\mathbf{y})\,\psi(\mathbf{y})\left(\frac{1}{|\mathbf{x}-\mathbf{y}|} - \frac{1}{|\mathbf{x}|}\right)\right|\right]^2. \tag{A5}$$

Inverting the order of integrations (which is allowed because the integrand is positive) we first consider

$$\int(d\mathbf{x})\left(\frac{1}{|\mathbf{x}-\mathbf{y}|} - \frac{1}{|\mathbf{x}|}\right)^2 = |\mathbf{y}|\int(d\mathbf{z})\left(\frac{1}{|\hat{n}-\mathbf{z}|} - \frac{1}{|\mathbf{z}|}\right)^2,$$

where $\mathbf{x} = |\mathbf{y}|\mathbf{z}$ and $\mathbf{y} = |y|\hat{n}$. But

$$\left(\frac{1}{|\hat{n}-\mathbf{z}|} - \frac{1}{|\mathbf{z}|}\right)^2 = \frac{(1-2\hat{n}\circ\mathbf{z})^2}{|\hat{n}-\mathbf{z}|^2|\mathbf{z}|^2(|\mathbf{z}|+|\hat{n}-\mathbf{z}|)^2}$$

and the integral clearly converges. Thus

$$\int(d\mathbf{x})\left(\frac{1}{|\mathbf{x}-\mathbf{y}|} - \frac{1}{|\mathbf{x}|}\right)^2 \le C|\mathbf{y}|. \tag{A6}$$

Consequently by Schwarz's inequality

$$\int(d\mathbf{x})\left|\left(\frac{1}{|\mathbf{x}-\mathbf{y}|} - \frac{1}{|\mathbf{x}|}\right)\left(\frac{1}{|\mathbf{x}-\mathbf{y}'|} - \frac{1}{|\mathbf{x}|}\right)\right| \le C|\mathbf{y}|^{1/2}|\mathbf{y}'|^{1/2}$$

and

$$\int(d\mathbf{x})\,|h_1|^2 \le C\left(\int(d\mathbf{y})\,|V(\mathbf{y})\,\psi(\mathbf{y})|\,|\mathbf{y}|^{1/2}\right) \le C\left(\int(d\mathbf{y})\,|V|\right)^2 \le C' \tag{A7}$$

because of (A1).

For h_2 we write

$$g = |\mathbf{x}-\mathbf{y}| - |\mathbf{x}| + \hat{x}\circ\mathbf{y} \equiv |\mathbf{x}|f,$$
$$f = (1 + |\mathbf{z}|^2 - 2\hat{x}\cdot\mathbf{z})^{1/2} - 1 + 2\hat{x}\cdot\mathbf{z}, \quad \mathbf{y} = |\mathbf{x}|\mathbf{z}.$$

Since f is continuous and

$$f \sim |\mathbf{z}|(1 + 2\hat{x}\cdot\hat{z}) \text{ as } |\mathbf{z}| \to \infty,$$
$$f \sim C|\mathbf{z}|^2 \text{ as } |\mathbf{z}| \to 0,$$

it follows that

$$|f| \le C|\mathbf{z}|^2/(a + |\mathbf{z}|)$$

and hence

$$|g| \le C|\mathbf{y}|^2/(a|\mathbf{x}| + |\mathbf{y}|).$$

But

$$|\exp(ik_0 g) - 1| = |\tfrac{1}{2}\sin\tfrac{1}{2}k_0 g| \le C\frac{|g|}{b + |g|} \le C'\frac{|\mathbf{y}|^2}{c|\mathbf{x}| + d|\mathbf{y}| + |\mathbf{y}|^2}.$$

Now

$$\int(d\mathbf{x})\,|h_2|^2 \le \frac{1}{16\pi^2}\int\frac{(d\mathbf{x})}{|\mathbf{x}|^2}\left[\int(d\mathbf{y})\,|V(\mathbf{y})\,\psi(\mathbf{y})| \times|\exp(ik_0 g) - 1|\right]^2. \tag{A8}$$

Again we look at the x-integral first:

$$\int\frac{(d\mathbf{x})}{|\mathbf{x}|^2}\,|\exp(ik_0 g) - 1|^2 \le C\int\frac{(d\mathbf{x})\,|\mathbf{y}|^4}{|\mathbf{x}|^2(c|\mathbf{x}| + d|\mathbf{y}| + |\mathbf{y}|^2)^2} \le C|\mathbf{y}|^2\int\frac{(d\mathbf{z})}{|\mathbf{z}|^2(c|\mathbf{z}|+1)^2} \le C|\mathbf{y}|^2. \tag{A9}$$

That $h_2 \in L^2(R^3)$ now follows from (A8) and (A9) as (A7) did from (A6).

APPENDIX B

We show here that the remainder in (5.23) is $O(k^2)$, and those in (5.26) and (5.27) are $o(k^{2n})$.

According to (5.10) the remainder in $\det_P B_R$ is

$$(\chi, R_1\chi) = (\chi, A^{-1}G''VA\chi) = (-AV\chi, A^{-1}G''VA\chi) = (\psi, VG''V\psi) \tag{B1}$$

and hence by (5.5′), (A1), and (2.1)

$$|(\chi, R_1\chi)| \le C|k|\int(d\mathbf{x})(d\mathbf{y})\,\frac{|V(\mathbf{x})\,V(\mathbf{y})\,\psi(\mathbf{x})\,\psi(\mathbf{y})|}{1 + |k||\mathbf{x}-\mathbf{y}|}\,|k||\mathbf{x}-\mathbf{y}| \le C|k|^2\int(d\mathbf{x})(d\mathbf{y})\,\frac{|V(\mathbf{x})\,V(\mathbf{y})|\,(|\mathbf{x}|+|\mathbf{y}|)}{(b+|\mathbf{x}|)(b+|\mathbf{y}|)} \le C|k|^2. \tag{B2}$$

This justifies (5.23).

We similarly obtain

$$(\chi, R_2\chi) = (\psi, VG'''V\psi), \tag{B3}$$

and hence by (5.5″)

$$\begin{aligned}
&|(\chi, R_2\chi)| \\
&\le C|k|^2\int(d\mathbf{x})(d\mathbf{y})\,\frac{|V(\mathbf{x})V(\mathbf{y})\,\psi(\mathbf{x})\,\psi(\mathbf{y})|}{1+|k||\mathbf{x}-\mathbf{y}|}\,|k||\mathbf{x}-\mathbf{y}|^2 \\
&\le C|k|^2\int(d\mathbf{x})(d\mathbf{y})\,\frac{|V(\mathbf{x})V(\mathbf{y})||k||\mathbf{x}-\mathbf{y}|}{1+|k||\mathbf{x}-\mathbf{y}|}\cdot\frac{|\mathbf{x}-\mathbf{y}|}{(b+|\mathbf{x}|)(b+|\mathbf{y}|)} \\
&\le C'|k|^2\int(d\mathbf{x})(d\mathbf{y})\,|V(\mathbf{x})\,V(\mathbf{y})|\,\frac{|k|\,(|\mathbf{x}|+|\mathbf{y}|)}{1+|k|\,(|\mathbf{x}|+|\mathbf{y}|)} \\
&\le C|k|^2\left[|k|^{1/2}\int_{|\mathbf{x}|+|\mathbf{y}|<|k|^{-1/2}}(d\mathbf{x})(d\mathbf{y})\,|V(\mathbf{x})\,V(\mathbf{y})|\right. \\
&\quad\left.+\int_{|\mathbf{x}|+|\mathbf{y}|\ge|k|^{-1/2}}(d\mathbf{x})(d\mathbf{y})\,|V(\mathbf{x})\,V(\mathbf{y})|\right] = o(k^2) \text{ as } k\to 0.
\end{aligned} \tag{B4}$$

This justifies (5.25) and (5.27).

APPENDIX C

We want to solve the equation

$$(1-B)\chi=\chi_0 \tag{C1}$$

in the vicinity of $k=0$, assuming that $k=0$ is an exceptional point and we are in case 3 of Sec. 5. (Case 1 is a special instance with $n=0$.)

The equation reads

$$(1-B_0-ikB_1-R_1)\chi=\chi_0 \tag{C2}$$

and we write $\chi=P\chi+Q\chi$, in the notation of Sec. 5. Then

$$\lim_{k\to 0} Q\chi=(1-B_0Q)^{-1}Q\chi_0,$$

which is finite.

Using a biorthogonal set of χ_m, $m=1,\ldots,n+1$, $B_0\chi_m=\chi_m$, and η_m such that $(\eta_m,\chi_p)=\delta_{mp}$ we may write in dyadic notation

$$P=\sum_{m=1}^{n+1}\eta_m\chi_m^\dagger=\sum_{m=1}^{n+1}\chi_m\eta_m^\dagger. \tag{C3}$$

Then

$$\chi=Q\chi+\sum_{m=1}^{n}\eta_m(\chi_m,\chi)+\eta_{n+1}(\chi_{n+1},\chi). \tag{C4}$$

Since for $m=1,\ldots,n$

$$0=(\chi_m,(1-B_0)\chi)=\lim_{k\to 0}(\chi_m,\chi_0)$$

the coefficients (χ_m,χ) can be calculated with no divergences. The only coefficient that leads to a divergence at $k=0$ is that for $m=n+1$. We find, by (5.7) and the fact that $\psi_{n+1}=-A\chi_{n+1}$ violates (5.19),

$$\int(d\mathbf{x})\,V(\mathbf{x})\,\psi(\mathbf{x})=4\pi ik^{-1}+O(1). \tag{C5}$$

But insertion of (C4) gives

$$\int(d\mathbf{x})\,V(\mathbf{x})\,\psi(\mathbf{x})=-\left(\int(d\mathbf{x})\,VA\eta_{n+1}\right)(\chi_{n+1},\chi)+O(1).$$

Expanding

$$\eta_m=\sum_p a_{mp}\chi_p \tag{C6}$$

we have

$$\int(dx)\,VA\eta_{n+1}=a_{n+1,\,n+1}$$

because of (5.19), and hence

$$(\chi_{n+1},\chi)=ck^{-1}+O(1)$$

with $c\neq 0$. Thus also

$$\psi=ck^{-1}+O(1), \tag{C7}$$

where the vector $c\neq 0$.

APPENDIX D

We want to prove here first (2.9′).

By Schwarz's inequality and (2.2)

$$\int(d\mathbf{z})\,\frac{|V(\mathbf{z})|}{|\mathbf{x}-\mathbf{z}||\mathbf{y}-\mathbf{z}|}$$

$$\leq\left[\int(d\mathbf{z})\,\frac{|V(\mathbf{z})|}{|\mathbf{x}-\mathbf{z}|^2}\right]^{1/2}\left[\int(d\mathbf{z}')\,\frac{|V(\mathbf{z}')|}{|\mathbf{y}-\mathbf{z}|^2}\right]^{1/2}\leq M. \tag{D1}$$

Therefore by Schwarz's inequality, (2.3), and the fact that $V\in L^1$,

$$\int(d\mathbf{x})(d\mathbf{y})(d\mathbf{z})\,\frac{|V(\mathbf{x})||V(\mathbf{y})||V(\mathbf{z})|}{|\mathbf{x}-\mathbf{y}||\mathbf{y}-\mathbf{z}||\mathbf{z}-\mathbf{x}|}$$

$$\leq M\int(d\mathbf{x})(d\mathbf{y})\,\frac{|V(\mathbf{x})||V(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|}$$

$$\leq M\int(d\mathbf{x})\,|V(\mathbf{x})|\left[\int(d\mathbf{y})(d\mathbf{z})\,\frac{|V(\mathbf{y})||V(\mathbf{z})|}{|\mathbf{y}-\mathbf{z}|^2}\right]^{1/2}<C. \tag{D2}$$

Hence the function

$$f(\mathbf{x},\mathbf{y},\mathbf{z})=V(\mathbf{x})\,V(\mathbf{y})\,V(\mathbf{z})/|\mathbf{x}-\mathbf{y}||\mathbf{y}-\mathbf{z}||\mathbf{x}-\mathbf{z}|$$

is in $L^1(R^9)$. It follows, as in the argument on p. 24 of Simon,[2] that as $\mathrm{Re}k\to\infty$, in $\mathrm{Im}k\geq 0$, for K defined by (2.8),

$$\lim \mathrm{tr}K^3=\lim\frac{1}{(-4\pi)^3}\int(d\mathbf{x})(d\mathbf{y})(d\mathbf{z})\,f(\mathbf{x},\mathbf{y},\mathbf{z})$$

$$\times\exp(ik(|\mathbf{x}-\mathbf{y}|+|\mathbf{y}-\mathbf{z}|+|\mathbf{z}-\mathbf{x}|)=0$$

by Lebesgue's lemma. As $\mathrm{Im}k\to\infty$, it follows by the dominated convergence theorem.

In order to establish (2.10), we argue by Schwarz's inequality, for $n\geq 4$,

$$|\mathrm{tr}K^n|=|\mathrm{tr}K^2K^{n-2}|\leq\|K^2\|_2\,\|K^{n-2}\|_2$$

$$=\|K^2\|_2\|K^2K^{n-4}\|_2\leq\|K^2\|_2^2\|K^{n-4}\|\leq\|K^2\|_2^2\|K\|^{n-4},$$

where $\|\cdot\|$ is the operator norm and $\|\cdot\|_2$ is the Hilbert–Schmidt norm. Furthermore,

$$\|K^2\|_2^2=\mathrm{tr}K^2K^{2\dagger}$$

$$=\mathrm{tr}KK^\dagger K^\dagger K\leq[\mathrm{tr}(KK^\dagger)^2]^{1/2}[\mathrm{tr}(K^\dagger K)^2]^{1/2}=\mathrm{tr}(KK^\dagger)^2$$

and therefore

$$|\mathrm{tr}K^n|\leq\mathrm{tr}(KK^\dagger)^2\|K\|^{n-4}. \tag{D3}$$

Now the second part of (2.9) implies that

$$\lim\|K\|=0.$$

Consequently, for $|k|$ large enough the expansion

$$D(k)=\exp\left(-\sum_2^\infty\frac{1}{n}\,\mathrm{tr}K^n\right)$$

converges absolutely, and (2.10) follows from (2.9), (2.9′), and (D3).

*Supported in part by the National Science Foundation.

[1]R. Jost and A. Pais, Phys. Rev. 82, 840–51 (1951).
[2]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, N.J., 1971).
[3]A generalization of Levinson's theorem to the one-particle Schrödinger equation with an interaction that is much more general than a local (multiplicative) central potential was given recently by T. Dreyfus, unpublished Ph. D. thesis (University of Geneva, 1976) and J. Math. Anal. Appl. (to be

published). His methods, however, are quite different and he does not obtain as much detail in the local potential case as does the present paper.

[4]T. A. Osborn and D. Bollé, J. Math. Phys. **18**, 432 (1977). The connection between their formulation and the one in the present paper is easily made by means of their Eq. (I. 11) which implies that $\mathrm{tr}q(E) = 2d\delta/dE$, where δ is defined by (6. 6). Their class of potentials is $L^1 \cap L^2$, and thus larger than the present, but they do not treat the case in which $k = 0$ is an exceptional point. Their result generalizes to a larger class of potentials than first obtained by V. S. Buslaev, Dokl. Akad. Nauk SSSR **143**, 1067 (1962) [Sov. Phys. Dokl. **7**, 295 (1962)].

[5]The constant C used throughout does not necessarily always have the same value.

[6]It is understood that $k = |\mathbf{k}|$.

[7]See C. Zemach and A. Klein, Nuovo Cimento **10**, 1078 (1958) and B. Simon, Ref. 2, p. 23.

[8]We use † to denote the adjoint, and *, the complex conjugate.

[9]"Modified Fredholm theory" refers to the theory as applicable to L^2-kernels.

[10]See, for example, R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), p. 259.

[11]We write $\hat{\mathbf{k}} = \mathbf{k}/|\mathbf{k}|$.

[12]$\delta(\hat{k}, \hat{k}')$ is the solid-angle delta function.

[13]B. Simon, Ref. 2, p. 106. See also discussion there and references to the earlier proofs, particularly by Kato and Kuroda, under somewhat different assumptions.

[14]This was first shown in R. G. Newton, Czech. J. of Phys. B **24**, 1195—1204 (1974), Appendix. The brief proof is repeated here for completeness.

[15]This definition allows a point to be exceptional of both kinds. There may be a solution for which (4. 2) holds, and another for which it fails.

[16]T. Kato, Comm. Pure Appl. Math. **12**, 403 (1959); see also R. Lavine, J. Funct. Anal. **12**, 30—54 (1973).

[17a]A similar theorem was proved under somewhat different assumptions by P. Alsholm and G. Schmidt, Arch. Ratl. Mech. Anal. **40**, 281 (1971), Appendix 3.

[17b]This can be shown by arguments similar to those of Sec. 5.

[18]Ref. 2, p. 117.

[19]B. Simon, Ref. 2, p. 101, proves that asymptotic completeness holds if and only if the singular continuous spectrum is empty.

[20]B. Simon, Ref. 2, p. 130.

[21]J. Aguilar and J. M. Combes, Comm. Math. Phys. **22**, 269 (1971); L. E. Thomas, Helv. Phys. Acta **45**, 1057 (1972).

[22]B. Simon, Ref. 2, pp. 99 and 130.

[23]T. Ikebe, Arch. Ratl. Mech. Anal. **5**, 1—34 (1960); S. Agmon, Ann. Sculoa Norm. Sup. Pisa, Ser. IV, **2**, 151—218 (1975); Alsholm and Schmidt, Ref. 17a, Theorem 1(b); S. T. Kuroda, J. Math. Soc. Japan **25**, 75 (1973), Theorem 5. 21.

[24]That $\det(1 - MP) = \det_P(1 - M)$ follows directly from the expression of both in terms of traces of powers of MP.

[25]If the functions η_m, $m = 1, \ldots, n+1$, form an orthonormal basis then $\det_p B = \det_P(\chi_m, B\chi_{m'})/|\det a|^2$, where $a_{mm'} = (\eta_{m'}, \chi_m)$ and $\det a \neq 0$ because the χ_m, $m = 1, \ldots, n+1$, are linearly independent.

[26]See Appendix B for an examination of the remainder.

[27]Note that $\det_2[(1 - A)(1 - B)] = \det_2(1 - A)\det(1 - B)| \times \exp[\mathrm{tr}(1 - A)B]$.

[28]This technique is originally due to R. Sugar and R. Blankenbecler, Phys. Rev. **136**, B472 (1965).

[29]This result was implied by V. S. Buslaev (Ref. 4), who credits L. D. Faddeev for it, and independently stated by R. G. Newton, in *Scattering Theory in Mathematical Physics*, edited by J. A. Lavita and J. -P. Marchand (Reidel, Dordrecht, Holland, 1974), p. 228.

[30]This follows from (11) of R. G. Newton, J. Math. Phys. **13**, 880—83 (1972).

[31]N. Levinson, Kgl. Danske Videnskab. Selskab, Mat. -Fys. Medd. **25** (9) (1949). See also Ref. 10. p. 356.

[32]This result was first derived for a very much more restricted class of potentials and under the assumption that $k = 0$ is not exceptional, by Buslaev, Ref. 4.

[33]This agrees with the result of Dreyfus, Ref. 3, Theorem 3, except that he considers only the case in which $k = 0$ is not exceptional. His specification of S also does not make use of (7. 3).

[34]M. Wollenberg, Math. Nachr. (in press).

Group contraction in a fiber bundle with Cartan connection

W. Drechsler

Max-Planck-Institut für Physik und Astrophysik, Munich, Germany
(Received 23 November 1976)

A contraction of the structural group with respect to the stability subgroup is performed in a fiber bundle with Cartan connection. The relation of the connections in the original and in the contracted bundle is examined. As an example interesting for physics the contraction of the SO(4,1) de Sitter bundle over space–time to the affine tangent bundle over space–time is discussed with the latter bundle possessing the Poincaré group as structural group.

I. INTRODUCTION

It has recently been proposed[1-3] that a fiber bundle with Cartan connection might be of considerable interest in physics as a basis for a differential geometric description of strong interactions. A fiber bundle of Cartan type constructed over space—time can be used as the geometrical substratum on which a gauge theory for the interaction of extended hadrons can be formulated. The structural group G of the bundle plays here the role of an internal dynamical group determining the internal motions of extended hadronic states (being associated with the fibers of the bundle) in a similar way as the Poincaré group determines the kinematics of pointlike objects in flat Minkowski space—time. The particular bundle suggested for a gauge formulation of strong interaction dynamics was a bundle with Cartan connection constructed over a curved (in the presence of gravitation) space—time manifold V_4 possessing the de Sitter group SO(4,1) as structural group. In this paper we study from a more general point of view the properties of a fiber bundle $E(M,F,G)$ with Cartan connection constructed over a base manifold M of dimension n possessing the fiber and structural group F and G, respectively, and determine its relation to the affine tangent bundle $T_A(M)$ in contracting the structural group G of the bundle with respect to a subgroup G'. The bundle $T_A(M)$ obtained in the contraction limit is here regarded as a bundle associated with the bundle of affine frames $A(M)$ over M and possesses, like $A(M)$, the affine group $A(n,R)$ as structural group. The affine tangent bundle is a simple example of a bundle soldered to M (see Sec. III below). This soldering of base space and fiber is a characteristic feature of a bundle with Cartan connection. In a bundle $E(M,F,G)$ with Cartan connection the fiber is isomorphic to the homogeneous space G/G', where G' is the stability subgroup of G in F, i.e., the subgroup of transformations of the space F leaving a particular point fixed. The contraction of the group G with respect to G' determines how the Cartan connection in $E(M,F,G)$, being given by a differential 1-form with values in the Lie algebra of the group G, decomposes in the contraction limit into a component with values in the Lie algebra of G' *and* a component related to translations. Both parts together constitute the Cartan connection on the affine tangent bundle $T_A(M)$ to which $E(M,F,G)$ reduces in the contraction limit. By this process a Cartan bundle with simple or semisimple structural group G goes over into a Cartan bundle possessing as structural group the affine group $A(n,R)$ having an n-dimensional Abelian subgroup.

The plan of the paper is as follows. After recalling in Sec. II the notion of a principal and an associated fiber bundle over a differentiable manifold M and defining the concept of a connection in a principal fiber bundle in the general case, we discuss in Sec. III the properties of Cartan connections, in particular following the treatment given by Ehresmann[4] and Kobayashi.[5] In Secs. II and III essentially known results are stated without proofs in order to make the paper self-contained and easier readable for physicists. In Sec. IV we treat the group contraction process and study the resulting connection on the bundle of affine frames. In Sec. V we, finally, discuss as an example the contraction of the SO(4,1) de Sitter bundle $T^R(V_4)$ with base space V_4, used in Refs. 1—3, to the affine tangent bundle $T_A(V_4)$ over space—time in the limit $R \to \infty$, where R is the radius of curvature of the standard fiber of the bundle $T^R(V_4)$. Correspondingly, the de Sitter frame bundle $L^R(V_4)$, being a principal fiber bundle over space—time with structural group SO(4,1), contracts to the bundle of affine frames $A(V_4)$ over space—time. $A(V_4)$ is a principal fiber bundle over V_4 possessing the Poincaré group P as structural group with $P = \mathrm{ISO}(3,1)$ being the group of motion in each local Minkowski tangent space $T_x(V_4)$ to V_4. The bundle $A(V_4)$ and the associated bundle $T_A(V_4)$ allow the discussion of a Poincaré gauge group as studied in the framework of a Lagrangian formalism by Kibble[6] and by Hayashi and Nakano.[7] The familiar Lorentz frame bundle $L(V_4)$ with structural group SO(3,1) is in a natural way a subbundle of $A(V_4)$.

II. CONNECTIONS IN A PRINCIPAL FIBER BUNDLE

In this section we recall some of the notions and definitions well known from the differential geometric literature.[8]

A fiber bundle $E(M,F,\pi,G,\Phi)$ is a differentiable manifold defined by the following collection of objects:

(a) the bundle space E,

(b) the base space M, covered by a family of coordinate neighborhoods $\{U_i\}$ with the index i in a set J,

(c) a space F called the fiber (or the standard fiber),

(d) a mapping π of E onto M called the projection,

(e) the structural group G of the bundle, being a Lie group acting effectively and differentiably on F,

(f) a family Φ of homeomorphisms $\{\phi_i\}$ corresponding to the open covering $\{U_i\}$ of M mapping in each coordinate neighborhood $U_i \times F$ onto $\pi^{-1}(U_i)$. The property f states that the bundle is locally trivial, i.e., is locally the topological product of an open subset of M and F. $\pi^{-1}(x) = F_x$ is called *the fiber over* x. Furthermore, one has the following essential property: If $x \in U_i \cap U_j$ and $\phi_i(x)$ is the differentiable mapping

$$\phi_i(x) : F \to \pi^{-1}(x) = F_x, \tag{1}$$

being the representative of ϕ_i at $x \in U_i$ and, correspondingly, if $\phi_j(x)$ is the mapping of the standard fiber into F_x as determined by the representative of ϕ_j at $x \in U_j$, then the mapping

$$\psi_{ij}(x) = \phi_i(x)\, \phi_j^{-1}(x) \tag{2}$$

coincides with an element g of the structural group G. The family of mappings $\psi_{ij}(x)$ are called the *transition functions* of the bundle $E(M, F, \pi, G, \Phi)$ corresponding to the covering $\{U_i\}$ of M.

If F and G are identical manifolds and G acts on itself by left (or right) translation, one speaks of a *principal* fiber bundle over M which will be denoted by $P(M, G, \pi_P, \Phi)$ or simply by $P(M, G)$.

To give some familiar examples of fiber bundles, we remark that the *bundle of linear frames* of a manifold M of dimension n, i.e., the space of all frames λ_x with origin x at all points of M,

$$L(M) = \bigcup_{x \in M} \lambda_x, \tag{3}$$

is a principal fiber bundle with structural group $\mathrm{Gl}(n, R)$. Furthermore, the *tangent bundle*

$$T(M) = \bigcup_{x \in M} T_x(M), \tag{4}$$

being the union of all tangent spaces to the manifold M at all points of M, is a fiber bundle with structural group $\mathrm{Gl}(n, R)$ and standard fiber R^n. If one regards the tangent spaces $T_x(M)$ as *affine* spaces on which the affine group $A(n, R)$ acts, one obtains the affine tangent bundle, denoted by $T_A(M)$ in the Introduction, possessing the standard fiber R^n and the structural group $A(n, R)$ being the semidirect product of $\mathrm{Gl}(n, R)$ and R^n.

A fiber bundle E with standard fiber F is frequently referred to as a bundle $E(M, F, \pi_E, G, P, \Phi')$ *associated* with the principal fiber bundle $P(M, G, \pi_P, \Phi)$ in the following way[5]: Let F be a manifold on which G acts effectively as a transformation group. If E is identified with the coset space $B = (P \times F)/G$ and π_E is the mapping of B onto M induced by the mapping π_P of P onto M, then one can construct[9] a family of homomorphisms $\{\phi_i'\}$ of $U_i \times F$ onto $\pi_E^{-1}(U_i)$. In this terminology the tangent bundle $T(M)$ over M is the bundle with standard fiber R^n associated with $L(M)$, and the affine tangent bundle $T_A(M)$ is a bundle associated with the bundle of affine frames $A(M)$. In Sec. V below we shall encounter another example of an associated bundle. In this case the standard fiber is the coset space $F = G/G'$ with G' being a closed subgroup of G. The group G acts effectively on G/G'. Choosing $G = \mathrm{SO}(4, 1)$ and $G' = \mathrm{SO}(3, 1)$ the de Sitter bundle $T^R(V_4)$ constructed over the space–time manifold V_4 possesses as standard fiber a Riemannian space of constant curvature,[10] $F = V_4'$, isomorphic to the coset space $\mathrm{SO}(4, 1)/\mathrm{SO}(3, 1)$ and has as the structural group the SO(4, 1) de Sitter group. The bundle $T^R(V_4)$ is associated with the de Sitter frame bundle $L^R(V_4)$, being a principal fiber bundle over space–time with structural group SO(4, 1).

As a final example of a fiber bundle we mention the bundle of linear differential forms of degree q over M. Let $\wedge\, T_x^*(M)$ denote the exterior algebra over $T_x^*(M)$ with $T_x^*(M)$ being the dual space of $T_x(M)$. A q-form on M is an assignment of an element of degree q in $\wedge\, T_x^*(M)$ to every $x \in M$. The fiber bundle $D^q(M)$ of differential q-forms is the bundle over M with standard fiber $F = R^{n*} \wedge R^{n*} \wedge \cdots \wedge R^{n*}$ (q times) and structural group $\otimes\, \mathrm{Gl}(n, R)$ associated with the bundle of frames $L(M)$ over M. A q-form on M is thus a cross section of $D^q(M)$ which can be in a local coordinate system $x^1, x^2, \ldots, x^n$ on $U \subset M$ be expressed as

$$\omega = \sum_{i_1 < i_2 < \cdots < i_q} \omega_{i_1 i_2 \cdots i_q}(x)\, dx^{i_1} \wedge dx^{i_2} \cdots \wedge dx^{i_q}, \tag{5}$$

with $\omega_{i_1 i_2 \cdots i_q}(x)$ denoting the differentiable coefficients of ω. We remark in passing that one speaks of a vector valued q-form if the $\omega_{i_1 i_2 \cdots i_q}(x)$ in Eq. (5) assume values in an r-dimensional vector space V having the basis E_s, $s = 1, 2, \ldots, r$. Accordingly one writes

$$\omega_V = \sum_{i_1 < i_2 < \cdots < i_q} \omega^s_{i_1 i_2 \cdots i_q}(x)\, E_s \otimes dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_q}. \tag{6}$$

In the case that the vector space V admits the structure of a Lie algebra $\mathfrak{g}$, with E_s defining a basis for $\mathfrak{g}$, the form ω_V is called a differential q-form with values in $\mathfrak{g}$.

A connection[11] in a principal fiber bundle $P(M, G)$ over M is given by a linear mapping σ_u of the tangent space $T_x(M)$ to M at x into the tangent space $T_u(P(M, G))$ to $P(M, G)$ at $u \in \pi^{-1}(x)$ with the properties that

$$\text{(i)}\ \sigma_u(T_x(M)) = H_u, \tag{7}$$

H_u being the horizontal subspace of $T_u(P(M, G))$, $u \in F_x$,

(ii) $d\pi \circ \sigma_u$ being the identity mapping, and

(iii) σ_u depending differentially on u.

The subspace H_u of horizontal vectors at u', with $u' = ug = R_g u$ obtained from u by right translation[12] in the fiber with an element g of the structural group G, is given by

$$H_{u'} = H_{ug} = R_g H_u. \tag{8}$$

Calling the tangent space to the *fiber* at u the vertical subspace of $T_u(P(M, G))$, i.e.,

$$T_u(F_x(M)) = V_u, \tag{9}$$

one obtains the following unique decomposition of any vector $X \in T_u(P(M, G))$ into horizontal and vertical components:

$$X = X_v + X_h \tag{10}$$

with $X_v \in V_u$ and $X_h \in H_u$. (Compare the schematic drawing shown in Fig. 1.)

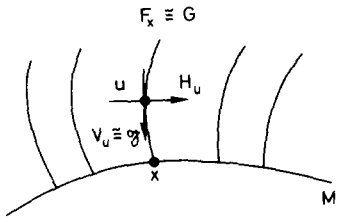


FIG. 1. Horizontal and vertical subspaces in a principal fiber bundle $P(M,G)$.

According to Eq. (10) a connection in $P(M,G)$ can be defined by a differentiable mapping of $T_u(P(M,G))$ onto $T_u(F_x(M))$, the space tangent to the fiber at u, i.e., onto the vertical subspace of $T_u(P(M,G))$, which is isomorphic to the Lie algebra $\mathfrak{g}$ of the structural group. This gives rise to a linear differential 1-form—called the form ω of the connection—with values in the Lie algebra of the structural group of the bundle. The 1-form ω possess the following properties:

$$\omega(udg) = g^{-1}dg, \tag{11a}$$

$$\omega(Xg) = g^{-1}\omega(X)g, \tag{11b}$$

with $u \in P(M,G)$, $\pi u = x$, $g \in G$, and $X \in T(P(M,G))$. In Eq. (11a) dg is the element of $T_g(G)$ tangent to the group G at g and $g^{-1}dg$ is the element of $T_e(G)$ tangent to the group at the unit element e; i.e., $g^{-1}dg \in \mathfrak{g}$.[13] Equation (11a) states that udg, being vertical, is mapped into the element of the Lie algebra corresponding to dg. Furthermore, Eq. (11a) implies that $\omega(X) = 0$ for X being horizontal. Denoting the right translation of the vector field X by $R_g X = Xg$, we see that Eq. (11b) states that ω transforms according to the (inverse) adjoint representation of the group G in $\mathfrak{g}$.

III. SOLDERING AND CARTAN CONNECTIONS

A fiber bundle $E(M,F,G,P)$ [14] associated with a principal fiber bundle $P(M,G)$ is called *soldered to* M if the following conditions are satisfied[4,5]:

(A) The group G acts transitively on F, i.e., F is the homogenous space G/G' where G' is the stability subgroup of G leaving the point O of F fixed.

(B) $\dim F = \dim M = n$.

(C) The bundle $E(M,F,G,P)$ admits a cross section which will be identified with M.

(D) If $T'(M)$ is the space of all tangent vectors to F_x at $x \in M$ for all x and $T(M)$ is the tangent bundle over M, then one can identify $T'(M)$ and $T(M)$ by an isomorphism.

The property D states that the fiber over x is tangent to the base space at x for every $x \in M$. $T'(M)$ is a fiber bundle over M with fiber R^n and structural group $\mathrm{Gl}(n,R)'$, which can be identified with the space $(P' \times R^n)/G'$, being a fiber bundle associated with $P'(M,G')$.[15] Here the principal bundle $P'(M,G')$ is a subspace of $P(M,G)$ obtained by restricting the homeomorphisms $F \to F_x$ of $P(M,G)$ in such a way that $O \in F$ is always mapped into x, i.e., into the point of contact of fiber and base space at $x \in M$. With the principal fiber bundle $P'(M,G')$ with structural group G' is associated in the usual way a bundle $E(M,F,G',P')$, the existence of which was first shown by Ehresmann to define the soldering in $E(M,F,G,P)$.

Now let ω denote the form of a Cartan connection in $P(M,G)$ [and thereby in $E(M,F,G,P)$], and let $\bar{\omega}$ be the restriction of ω to the bundle $P'(M,G')$ [and thereby to $E(M,F,G',P')$]. Furthermore, let $\mathfrak{g}'$ denote the Lie algebra of G'. Then $\bar{\omega}$ is a $\mathfrak{g}$-valued linear differential 1-form satisfying the following conditions:

$$\bar{\omega}(u'dg') = g'^{-1}dg', \tag{12a}$$

$$\bar{\omega}(X'g') = g'^{-1}\bar{\omega}(X')g', \tag{12b}$$

$$\text{if } \bar{\omega}(X') = 0, \text{ then } X' \text{ is the zero vector.} \tag{12c}$$

Here $u' \in P'(M,G')$, $\pi' u' = x$ [with π' denoting the projection in $P'(M,G')$], $g' \in G'$, $dg' \in T_{g'}(G')$, and $X' \in T(P'(M,G'))$. Equations (12a) and (12b) state that $\bar{\omega}$ defines a connection in $P'(M,G')$, and Eq. (12c) implies that any horizontal vector in $P'(M,G')$ is the zero vector. This last mentioned property is a consequence of the soldering of $E(M,G',P')$ to M.

Starting from Eqs. (12a)—(12c), it is now simple to specify uniquely the form ω of a Cartan connection in $E(M,F,G,P)$ by the following $\mathfrak{g}$-valued differential form:

$$\omega(X) = g^{-1}\bar{\omega}(X')g + g^{-1}dg, \tag{13}$$

where $g \in G$, $dg \in T_g(G)$ and $X = R_g X' = X'g$ with $X' \in T(P'(M,G'))$ and $X \in T(P(M,G))$. It is seen from Eq. (13) that the restriction of ω to $P'(M,G')$ is $\bar{\omega}$.

Let us now decompose the Lie algebra $\mathfrak{g}$ of G into the subalgebra $\mathfrak{g}'$ and a vector subspace $\mathfrak{t}$:

$$\mathfrak{g} = \mathfrak{g}' \dotplus \mathfrak{t}. \tag{14}$$

Then the tangent space $T_0(F)$ at $O \in F$ is isomorphic with $\mathfrak{t}$. If, furthermore,

$$[\mathfrak{g}', \mathfrak{t}] \subseteq \mathfrak{t}. \tag{15}$$

$T_0(F)$ can be identified with $\mathcal{A}$ and there exists a linear $\mathfrak{t}$-valued differential 1-form θ on $P'(M,G')$, called the *form of soldering*,[16] with the properties[17]:

$$\theta(X') = 0 \text{ for } X' \in T(P'(M,G'))$$

$$\text{if and only if } d\pi'(X') = 0, \tag{16a}$$

$$\theta(X'g') = g'^{-1}\theta(X')g' \tag{16b}$$

for all $X' \in T(P'(M,G'))$ and $g' \in G'$.

Equation (16a) states that θ vanishes for all vertical vectors on $P'(M,G')$.

It can be shown[5] that if the bundle $E(M,F=G/G',G,P)$ satisfies the conditions A, B, and C, and if the Lie algebra $\mathfrak{g}$ of G and the Lie algebra $\mathfrak{g}'$ of the stability subgroup G' of G decomposes according to Eq. (14) and obeys Eq. (15), then the bundle $E(M,F=G/G',G,P)$ is *soldered* to M (according to property D) if and only if there exists a $\mathfrak{t}$-valued linear differential form θ satisfying Eqs. (16a) and (16b).

The interest in the form θ lies in the fact that if ω' is a $\mathfrak{g}'$-valued 1-form defining a connection in $P'(M, G')$, the sum

$$\bar{\omega} = \omega' + \theta \tag{17}$$

is a $\mathfrak{g}$-valued 1-form defining in a one-to-one way the restriction to $P'(M, G')$ of a Cartan connection ω in $P(M, G)$ [and correspondingly in the associated bundle $E(M, F = G/G', GP)$ soldered to M]. The proof of this statement, i.e., that $\bar{\omega}$ as defined by Eq. (17) satisfies Eqs. (12a)–(12c), follows directly from the properties (16a) and (16b) and the definition of the connection ω' in $P'(M, G')$.

To conclude our review of Cartan bundles and their soldering property, we finally write down the structural equation of Cartan for the space $E(M, F = G/G', G, P)$, i.e., (d denotes the exterior derivative),

$$d\bar{\omega} + \tfrac{1}{2}[\bar{\omega}, \bar{\omega}] = \bar{\Omega} \tag{18}$$

with $\bar{\Omega}$ being the curvature 2-form of the connection $\bar{\omega}$ and $[\ ,\ ]$ denoting the exterior product of forms with values in a Lie algebra. Using in Eq. (18) the decomposition (17) and separating both sides of the equation into a $\mathfrak{g}'$-valued and a $\mathfrak{t}$-valued part, remembering Eq. (15), and putting

$$\bar{\Omega} = \bar{\Omega}_{\mathfrak{g}'} + \bar{\Omega}_{\mathfrak{t}} \tag{19}$$

and similarly

$$[\theta, \theta] = [\theta, \theta]_{\mathfrak{g}'} + [\theta, \theta]_{\mathfrak{t}}, \tag{20}$$

one obtains the following relations, moreover, by making use of the structural equation

$$d\omega' + \tfrac{1}{2}[\omega', \omega'] = \Omega' \tag{21}$$

as defined by the connection ω' in $P'(M, G')$:

$$\Omega' = \bar{\Omega}_{\mathfrak{g}'} - \tfrac{1}{2}[\theta, \theta]_{\mathfrak{g}'}, \tag{22}$$

$$d\theta + \tfrac{1}{2}[\omega', \theta] + \tfrac{1}{2}[\theta, \omega'] = D'\theta = \tau. \tag{23}$$

Here

$$\tau = \bar{\Omega}_{\mathfrak{t}} - \tfrac{1}{2}[\theta, \theta]_{\mathfrak{t}} \tag{24}$$

denotes the torsion form of the connection and $D'\theta = d\theta + [\omega', \theta]$ is the exterior covariant derivative of the form θ with respect to the connection ω'.

If now the space $F = G/G'$ satisfies the condition

$$[\mathfrak{t}, \mathfrak{t}] \subseteq \mathfrak{g}' \tag{25}$$

as is the case for the physically interesting example of a Cartan bundle treated in Sec. V as well as for the discussion of the group contraction presented in the next section, where we shall start from the assumption that the fiber $F = G/G'$ is a Riemannian globally symmetric space, Eqs. (22) and (24) reduce to the form

$$\bar{\Omega}_{\mathfrak{g}'} = \Omega' + \tfrac{1}{2}[\theta, \theta], \tag{26a}$$

$$\bar{\Omega}_{\mathfrak{t}} = \tau. \tag{26b}$$

If, on the other hand,

$$[\mathfrak{t}, \mathfrak{t}] = 0 \tag{27}$$

which is a condition valid for the affine tangent bundle $T_A(M)$ with fiber $F = A(n, R)/\mathrm{Gl}(n, R) = R^n$, obtainable from Eq. (26a) in the limit of group contraction (see the next section), one finds

$$\bar{\Omega}_{\mathfrak{g}'} = \Omega', \tag{28}$$

$$\bar{\Omega}_{\mathfrak{t}} = \tau. \tag{29}$$

IV. CONTRACTION OF THE STRUCTURAL GROUP

Let us choose the fiber of the Cartan bundle $E(M, F, G, P)$ to be a Riemannian globally symmetric space[18] $F = G/G'$ of dimension n. Such a space is characterized by the decomposition $\mathfrak{g} = \mathfrak{g}' \oplus \mathfrak{t}$ [see Eq. (14)] of the Lie algebra $\mathfrak{g}$ obeying

$$[\mathfrak{g}', \mathfrak{g}'] \subseteq \mathfrak{g}', \tag{30a}$$

$$[\mathfrak{g}', \mathfrak{t}] \subseteq \mathfrak{t}, \tag{30b}$$

$$[\mathfrak{t}, \mathfrak{t}] \subseteq \mathfrak{g}', \tag{30c}$$

where $\mathfrak{t}$ is here required to be a vector subspace of $\mathfrak{g}$ with dimension $n = \dim G - \dim G' = \dim M$. The coset space $F = G/G'$ has certain compact or noncompact forms depending on the particular quadratic form with signature (p, q) left invariant by the group G and the subgroup G'. This quadratic form specifies on the one hand G and G' as particular metric-preserving subgroups of the general linear group, and on the other hand determines the special form of a hypersurface F will represent if the space F is embedded into an $(n+1)$-dimensional flat space of definite signature (p', q') with $p' + q' = n + 1$. In Sec. IV we shall discuss as an example of particular interest in physics the case of a Cartan bundle possessing the ten-parameter semisimple group $G = \mathrm{SO}(4, 1)$ as structural group leaving a quadratic form with signature $(p, q) = (- - - -, +) = (4, 1)$ invariant, corresponding, when represented in a space $R^n_{p_0} = R^5_{4,1}$, to a one-shell (hyper)-hyperboloid, V'_4, called the de Sitter hyperboloid. This hyperbolic space is isomorphic to the noncompact coset space $F = \mathrm{SO}(4, 1)/\mathrm{SO}(3, 1)$ being a space of constant negative curvature regarded as the fiber of the Cartan bundle $E(V_4, \mathrm{SO}(4, 1)/\mathrm{SO}(3, 1), \mathrm{SO}(4, 1), P)$ constructed over a (Riemann–Cartan) space–time manifold V_4. In this section, however, we do not specify the group G in detail except for demanding that G be the identity component of a particular metric-preserving subgroup of $\mathrm{Gl}(n+1, R)$ admitting G' [being similarly a metric-preserving subgroup of $\mathrm{Gl}(n, R)$] as a subgroup in such a way that the space $F = G/G'$ is a globally symmetric Riemannian space of dimension $n = \dim M$ in accordance with the statement B of the previous section. These requirements restrict the groups G and G' to the class of the special orthogonal groups, $\mathrm{SO}(p, q)$, having dimension $\frac{1}{2}(p+q)(p+q-1)$. Calling $G' = \mathrm{SO}(p, q)$ from now on, with $p + q = n$, then G is either $\mathrm{SO}(p+1, q)$ or $\mathrm{SO}(p, q+1)$, and the space $F = G/G'$ is a Riemannian space V_n of constant curvature admitting a metric with signature (p, q).

Let us now contract the structural group of the Cartan bundle with respect to its stability subgroup G' of the point O in F. The Inönü–Wigner contraction of the algebra $\mathfrak{g} = \mathfrak{g}' \oplus \mathfrak{t}$ can be defined as the following transformation $U(\epsilon)$ of $\mathfrak{g}$ being singular in the limit $\epsilon \to 0$,[19]

$$U(0)\,\mathfrak{g}' = \mathfrak{g}', \tag{31a}$$

$$U(0)\,\mathfrak{t} = 0, \tag{31b}$$

resulting in an algebra $\tilde{\mathfrak{g}} = \mathfrak{g}' \oplus \tilde{\mathfrak{t}}$ obeying

$$[\mathfrak{g}', \mathfrak{g}'] \subseteq \mathfrak{g}', \tag{32a}$$

$$[\mathfrak{g}', \tilde{\mathfrak{t}}] \subseteq \tilde{\mathfrak{t}}, \tag{32b}$$

$$[\tilde{\mathfrak{t}}, \tilde{\mathfrak{t}}] = 0. \tag{32c}$$

The algebra $\mathfrak{g}$ is characterized by the fact that it has the same dimension as $\mathfrak{g}$ with $\mathfrak{g}'$ being a subalgebra in $\mathfrak{g}'$ as well as in $\mathfrak{g}$, and with $\tilde{\mathfrak{t}}$ being an *Abelian* invariant subalgebra of $\tilde{\mathfrak{g}}$. Thus $\tilde{\mathfrak{g}}$ is nonsemisimple.

Correspondingly the Riemannian space $F = G/G'$ on which G acts as a group of motion degenerates in the contraction limit $\epsilon \to 0$ into a flat space. The limit $\epsilon \to 0$ corresponds to the limit $1/K \to 0$, where K is the constant of curvature characterizing F. In this discussion, where we did not specify explicitly the particular special orthogonal groups G and G', we can say that F degenerates in the limit $\epsilon \to 0$ into the space $R^n_{p,q}$[20] isomorphic to $E(p,q)/\mathrm{SO}(p,q)$ with $G = E(p,q) = \mathrm{ISO}(p,q)$ denoting the subgroup of the affine group $A(n,R)$ in $n = p+q$ dimensions leaving a quadratic form with signature (p,q) invariant. The group $E(p,q) = \mathrm{ISO}(p,q)$ is the semidirect product of $\mathrm{SO}(p,q)$ and the space $R^n_{p,q}$ possessing the Lie algebra characterized by Eqs. (32a–c). The Abelian subgroup is generated by the n elements of $\tilde{\mathfrak{t}}$ corresponding to translations. The group $E(p,q) = \mathrm{ISO}(p,q)$, being the group of motions in $R^n_{p,q}$, can in the familiar way be parametrized by the $(n+1)\times(n+1)$ matrices $\begin{pmatrix} \Lambda & v \\ 0 & 1 \end{pmatrix}$ with $\Lambda = (\Lambda^i_j) \in \mathrm{SO}(p,q)$ and $v = \{v^i\} \in R^n_{p,q}$.

The group contraction is thus seen to relate the bundle with Cartan connection $E(M, F = G/G', G, P)$, associated with $P(M,G)$, having the fiber $F = \mathrm{SO}(p+1,q)/\mathrm{SO}(p,q)$ or $F = \mathrm{SO}(p,q+1)/\mathrm{SO}(p,q)$ being *curved* Riemannian spaces, to the special Cartan bundle $E(M, F = R^n_{p,q}, \mathrm{ISO}(p,q), A_{p,q}(M))$, associated with the bundle of affine frames $A_{p,q}(M) = P(M, \mathrm{ISO}(p,q))$, and possessing a *flat* space as fiber; i.e., $E(M, R^n_{p,q}, \mathrm{ISO}(p,q), A_{p,q}(M))$ is the subbundle characterized by the signature (p,q) of the affine tangent bundle called $T_A(M)$ in Sec. II above, being obtained by restricting the general affine group $A(n,R)$ operating in $T_A(M)$ to the group $\mathrm{ISO}(p,q)$.

The expression (17) of a Cartan connection in $P(M,G)$ [or, more exactly, its reduction to $P'(M,G')$] reduces in the contraction limit in a natural way to the form

$$\bar{\tilde{\omega}} = \tilde{\omega}' + \tilde{\theta} \tag{33}$$

of a Cartan connection [usually called a *generalized affine connection*[21]] in $T_A(M)$ with $\tilde{\omega}'$ being an $\mathfrak{so}(p,q)$-valued 1-form and $\tilde{\theta}$ being a $R^n_{p,q}$-valued 1-form. Here $\bar{\tilde{\omega}}$ is the restriction to $L_{p,q}(M)$ of a 1-form $\tilde{\omega}$ defining an affine connection in $A_{p,q}(M)$, with $L_{p,q}(M)$ and $A_{p,q}(M)$ being subbundles corresponding to the signature (p,q) of $L(M)$ and $A(M)$, respectively, having structural groups $\mathrm{SO}(p,q)$ and $\mathrm{ISO}(p,q)$, respectively. The form $\tilde{\omega}'$ in Eq. (33) is the form of a *linear connection* in M; more specifically it is the form of a connection in the bundle of orthogonal frames $L_{p,q}(M)$ over M, the latter regarded as manifold with signature (p,q) of its metric. The form $\tilde{\omega}'$ is unrelated to the form ω' in Eq. (17). ω' corresponds to a connection in $P'(M,G')$ with G' viewed as the stability subgroup of the point O in the *fiber* of $E(M, G/G', G, P)$, whereas $\tilde{\omega}'$ represents the connection form on $L_{p,q}(M)$ related to the *base space* (compare Ref. 15 in this context). If the base space M of the bundle is flat (i.e., $M = R^n_{p,q}$) $\tilde{\omega}'$ is integrable, i.e, $\tilde{\Omega}' = 0$, where

$$d\tilde{\omega}' + \tfrac{1}{2}[\tilde{\omega}', \tilde{\omega}'] = \tilde{\Omega}'. \tag{34}$$

However, Ω' defined by Eq. (21) does not vanish even if M is a flat space.

The form $\tilde{\theta}$ in Eq. (33) is *not* the fundamental 1-form on the manifold M. Let us for later reference, introduce the fundamental 1-form φ on M as the following $R^n_{p,q}$-valued form

$$\varphi = \omega^i(x)\, e_i(x), \tag{35}$$

where $\{e_i(x)\}$, $i = 1, 2, \ldots, n$, represents a basis in $T_x(M)$. The exterior covariant derivative of $\tilde{\theta}$ with respect to $\tilde{\omega}'$ is the torsion form [compare Eq. (23)], i.e.,

$$d\tilde{\theta} + [\tilde{\omega}', \tilde{\theta}] = \tilde{\tau}. \tag{36}$$

Equations (34) and (36) are the structural equations of Cartan for the affine tangent space over M. The latter viewed here as a manifold admitting a metric with signature (p,q). Expressed in terms of the generalized affine (or special Cartan) connection (33) the structural equations (18) read

$$\bar{\tilde{\Omega}} = \tilde{\Omega}' + \tilde{D}'\tilde{\theta} = \tilde{\Omega}' + \tilde{\tau} \tag{37}$$

with $\tilde{\Omega}'$ and $\tilde{\tau}$ as defined by Eqs. (34) and (35), respectively. Equation (37) for the affine case is in complete analogy to Eqs. (19), (22), and (23) taking into account the Abelian nature of the subalgebra $\tilde{\mathfrak{t}}$. The connection (33) on $T_A(M)$ is called torsion free if $\tilde{D}'\tilde{\theta} = 0$; it is called flat if $\tilde{\Omega}' = 0$, $\tilde{\tau} = 0$.

V. THE DE SITTER CASE

In Ref. 1 the de Sitter fiber bundle $T^R(V_4)$ constructed over a four-dimensional space–time manifold $M = V_4$ of Riemann–Cartan type with Lorentzian signature $(-\,-\,-\,+)$ was introduced and proposed as a generalized space providing the geometric framework for a gauge formulation of strong interaction physics. Let us discuss in this section the differential geometric properties of this eight-dimensional Cartan bundle space and the associated de Sitter frame bundle denoted by $L^R(V_4)$ in Ref. 1. Using the notation of the previous section, we see that $T^R(V_4)$ is isomorphic to the fiber bundle [compare Fig. 2 for a schematic drawing, where also the tangent spaces $T_0(F_x)$ and $T_x(V_4)$, to be identified by an isomorphism in the soldering of the bundle, are shown and drawn separately for clarity]

$$T^R(V_4) \simeq E(V_4, F = \mathrm{SO}(4,1)/\mathrm{SO}(3,1), \mathrm{SO}(4,1), L^R(V_4)) \tag{38}$$

associated with

$$L^R(V_4) = P(V_4, \mathrm{SO}(4,1)). \tag{39}$$

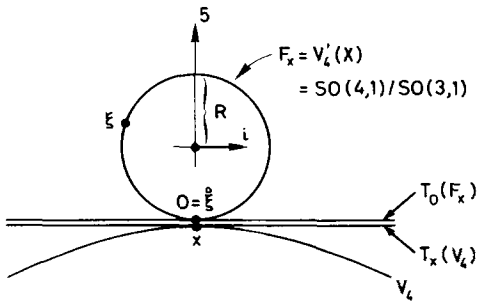


FIG. 2. Schematic drawing of the fiber bundle $T^R(V_4)$.

The fiber of $T^R(V_4)$ is a four-dimensional Riemannian space of constant curvature $K = 1/R^2$ (R being the curvature radius), i.e., a (4,1) de Sitter space, V_4', isomorphic to the coset space SO(4,1)/SO(3,1). With regard to strong interactions the curvature radius of the fiber V_4' was taken in Refs. 1–3 to have a fixed value of $R \approx 10^{-13}$ cm characterizing the bundle $T^R(V_4)$ as a whole in a way similar to the way the base manifold V_4 [or rather the tangent bundle $T(V_4)$ over V_4] is characterized by the Lorentz structure associated with a fixed limiting velocity $c \approx 3 \cdot 10^{10}$ cm/sec. The structural group of $T^R(V_4)$ as well as of the principal fiber bundle $L^R(V_4)$ is the de Sitter group SO(4,1) acting as a group of motion in the fiber of $T^R(V_4)$ and by left translation in the group itself, i.e., in the fiber of $L^R(V_4)$.

A basis for the Lie algebra of SO(4,1) is given by the ten elements

$$M_{ab} = -M_{ba}, \quad a,b = 0,1,2,3,5, \tag{40}$$

satisfying the commutation relations

$$i[M_{ab}, M_{cd}] = \eta_{ac} M_{bd} + \eta_{bd} M_{ac} - \eta_{ad} M_{bc} - \eta_{bc} M_{ad} \tag{41}$$

with

$$\eta_{ab} = \mathrm{diag}(1, -1, -1, -1, -1). \tag{42}$$

In order to exhibit the subgroup structure of SO(4,1) more clearly and to exemplify the contraction with respect to the stability subgroup SO(3,1) of the point $O \in F = V_4'$ (see Fig. 2), we introduce the elements

$$\Pi_i = (1/R) M_{5i}. \tag{43}$$

With (43) and $\eta_{ik} = \mathrm{diag}(1, -1, -1, -1)$ the commutation relations (41) can be rewritten as

$$i[M_{ij}, M_{kl}] = \eta_{ik} M_{jl} + \eta_{jl} M_{ik} - \eta_{il} M_{jk} - \eta_{jk} M_{il}, \tag{44a}$$

$$i[\Pi_i, M_{jk}] = \eta_{ik} \Pi_j - \eta_{ij} \Pi_k, \tag{44b}$$

$$i[\Pi_i, \Pi_j] = -(1/R^2) M_{ij}, \tag{44c}$$

which are of the form (30) with subalgebra $\mathfrak{g}'$ spanned by the six elements M_{ij}, $i,j = 0,1,2,3$, generating the SO(3,1) subgroup, and the four-dimensional vector subspace $\mathfrak{t}$ spanned by the elements Π_i, $i = 0,1,2,3$, corresponding to a four-parameter family of special de Sitter transformations, the so-called de Sitter boosts.

The de Sitter space V_4' on which SO(4,1) acts as a group of motion can be represented by a hypersurface[22] in an $R^5_{4,1}$, i.e.,

$$\xi^a \xi_a = \xi^a \xi^b \eta_{ab} = -R^2 \tag{45}$$

with $\xi^a = (\xi^0, \xi^1, \xi^2, \xi^3, \xi^5) \in R^5_{4,1}$. SO(4,1) is then the group of hyperbolic rotations in $R^5_{4,1}$ leaving the quadratic form (45) invariant. The space V_4' represented by (45) is compact in its "spatial" directions $\xi^1, \xi^2, \xi^3, \xi^5$ and noncompact in "time" ξ^0. A set of differential operators acting in $R^5_{4,1}$ and satisfying Eqs. (41) or (43) and (45) is provided by

$$L_{ab}(\xi) = i(\xi_a \partial_b - \xi_b \partial_a), \quad a,b = 0,1,2,3,5, \tag{46}$$

with $\partial_a = \partial/\partial \xi^a$. When applied to a function defined on V_4' the operators $L_{ab}(\xi)$ produce a result lying again in the hypersurface V_4'.

Contracting now the algebra (44) according to the prescription (31) of Sec. IV corresponding to the limit $R \to \infty$ and calling the elements of the Abelian subalgebra $\tilde{\mathfrak{t}}$ appearing in the limit P_i, one sees that the algebra (44) contracts to the algebra of the Poincaré group $P = \mathrm{ISO}(3,1)$ spanned by the elements M_{ij} and P_i satisfying the familiar commutation rules

$$i[M_{ij}, M_{kl}] = \eta_{ik} M_{jl} + \eta_{jl} M_{ik} - \eta_{il} M_{jk} - \eta_{jk} M_{il}, \tag{47a}$$

$$i[P_i, M_{jk}] = \eta_{ik} P_j - \eta_{ij} P_k, \tag{47b}$$

$$[P_i, P_j] = 0. \tag{47c}$$

In view of the replacement (43) corresponding for the $L_{ab}(\xi)$ to

$$\Pi_i = ((1/R) L_{5i}(\xi))_{\xi = \overset{o}{\xi}} \overset{R \to \infty}{\longrightarrow} P_i = i\tilde{\partial}_i, \tag{48}$$

where $\overset{o}{\xi}{}^a = (0,0,0,0,-R)$ denotes the coordinates of the point $O \in F_x$ and $\tilde{\partial}_i$, $i = 0,1,2,3$ being the differential operators in a flat four-dimensional space (see below); the quadratic form (45) can be represented as[22]

$$\xi^i \xi^j \eta_{ij} - \tilde{\xi}^5 \tilde{\xi}^5 R^2 = -R^2 \tag{49}$$

with

$$\tilde{\xi}^5 = \xi^5/R \tag{50}$$

approaching a constant in the limit $R \to \infty$. Dividing by R^2, we see that the hypersurface (49) reduces in the contraction limit to the hyperplanes $\tilde{\xi}^5 = \pm 1$, ξ^i arbitrary, being the equations for two spaces $R^4_{3,1}$. Only the one of these two hyperplanes with $\tilde{\xi}^5 < 0$ (corresponding to $O \in F_x$) is soldered to the base space, i.e., we have to take the space $R^4_{3,1}$ with $\tilde{\xi}^5 = -1$ as a solution of Eq. (49) in the limit $R \to \infty$.

It is thus seen that the de Sitter bundle $T^R(V_4)$ contracts in the limit $R \to \infty$ to the affine tangent bundle $T_A(V_4)$ over space–time, and the associated principal fiber bundle $L^R(V_4)$ contracts to the bundle of affine frames $A(V_4)$ over space–time containing the Lorentz frame bundle as a subbundle. $A(V_4)$ is a principal fiber bundle with structural group ISO(3,1) being the semidirect product of the Lorentz group SO(3,1) and Minkowski space $R^4_{3,1}$ (called M_4 for brevity) appearing here in the limit $R \to \infty$ as a gauge group acting in the fibers of $A(V_4)$ and $T_A(V_4)$. In the bundle $T^R(V_4)$ with structural (or gauge) group SO(4,1) the translational degrees of freedom are contained in the vector subspace

spanned by the Π_i being elements of the *semisimple* Lie algebra (44).

Let us now discuss Cartan connections on $T^R(V_4)$ and $T_A(V_4)$, respectively, where $T^R(V_4)$ and $T_A(V_4)$ are associated with the frame bundles $L^R(V_4)$ and $A(V_4)$, respectively. For reference to earlier work[1-3] let us discuss *spinor* connections on $T^R(V_4)$ [or $T_A(V_4)$], which are required in the definition of a de Sitter gauge invariant (or Poincaré gauge invariant) differentiation process for a spinor field defined on $T^R(V_4)$ [or on $T_A(V_4)$]. A spinor connection in $T^R(V_4)$ is a connection in the so-called spinor frame bundle $S^R(V_4)$ over V_4 possessing the covering group $\bar{G}=\overline{SO(4,1)}$ of $SO(4,1)$ as structural group. The homomorphism $S^R(V_4)\to L^R(V_4)$ is a bundle mapping corresponding to the group homomorphism $\overline{SO(4,1)}\to SO(4,1)$ [see Eq. (61) below]. The covering group of $SO(4,1)$ is the group $USp(2,2)=U(2,2)\cap Sp(4,C)$ being isomorphic to a subgroup of $Gl(2,Q)$, the general linear group of 2×2 matrices over the field of quaternions. On the other hand, the covering group of the Poincaré group is the semidirect product of $Sl(2,C)$ and the translations in Minkowski space.

We now first discuss the de Sitter case. A connection Γ^R on $T^R(V_4)$, associated with $S^R(V_4)$, is given by the following matrix-valued 1-form on V_4 ($x\in V_4$):

$$\Gamma^R(x)=\tfrac{1}{2}\omega^R_{ab}(x)M^{ab} \tag{51}$$

with $\omega^R_{ab}(x)=-\omega^R_{ba}(x)$ being the 1-form coefficients of the Cartan connection ω in $L^R(V_4)$, and M^{ab} being a spinor representation of the algebra of $SO(4,1)$. The lowest-dimensional representation of the algebra of $USp(2,2)$ is provided by the 4×4 matrices [we keep the same symbol as in Eq. (41)]

$$M^{ab}=(i/4)[\gamma^a,\gamma^b] \tag{52}$$

with $\gamma^a=(\gamma^i,\gamma^5)$, $i=0,1,2,3$, $\gamma^5=\gamma^0\gamma^1\gamma^2\gamma^3$ being the five anticommuting Dirac matrices obeying

$$\gamma^{a\dagger}=\gamma^0\gamma^a\gamma^0,\quad \{\gamma^a,\gamma^b\}=2\eta^{ab}\,1 \tag{53}$$

with η^{ab} as defined in Eq. (42).

The $\omega^R_{ab}(x)$ appearing in Eq. (51) can be expanded in terms of the 1-form components $\omega^i(x)$ of the fundamental form φ defined in Eq. (35) representing a complete set of 1-forms in $T^*_x(V_4)$. In terms of the coordinate differentials dx^μ of a system of global space–time coordinates on V_4 (written with a Greek index) the $\omega^i(x)$ can be expanded as

$$\omega^i(x)=\lambda^i_\mu(x)\,dx^\mu, \tag{54}$$

where the sixteen fields $\lambda^i_\mu(x)$ are referred to as the vierbein fields. For completeness we add that the covariant components of the metric tensor in V_4 are given by

$$g_{\mu\nu}(x)=\lambda^i_\mu(x)\lambda^k_\nu(x)\eta_{ik}, \tag{55}$$

being a symmetric tensor field on V_4 possessing ten independent components. The additional six degrees of freedom at each space–time point x contained in the $\lambda^i_\mu(x)$ refer to the freedom to Lorentz rotate the local Lorentz frame $e_k(x)$, $k=0,1,2,3$, in $T_x(V_4)$. After these remarks we write Eq. (51) finally as

$$\Gamma^R(x)=\tfrac{1}{2}\omega^i(x)\,\Gamma_i{}^R_{ab}(x)M^{ab} \tag{56}$$

with $\Gamma_i{}^R_{ab}(x)=-\Gamma_i{}^R_{ba}(x)$ denoting the 40 coefficients of a Cartan connection in $T^R(V_4)$. Equation (56) could, furthermore, be brought into the form (6) by using Eq. (54).

The form $\Gamma^R(x)$ defines a gauge invariant absolute differentiation process for a de Sitter spinor quantity $\psi(x,\xi)=\{\psi^{A'}(x,\xi);A'=1,2,3,4\}$, $x\in V_4$, $\xi\in F_x=V_4'(x)$ defined on the bundle space $T^R(V_4)$ and referred to a particular local frame in $S^R(V_4)$ called a gauge on $T^R(V_4)$ [being a system of moving de Sitter frames on $T^R(V_4)$ corresponding to a cross section of $S^R(V_4)$ [23]]. It is given by

$$D\psi(x,\xi)=d\psi(x,\xi)+i\Gamma^R(x)\,\psi(x,\xi) \tag{57}$$

with

$$d=dx^\mu\partial_\mu=\omega^k(x)\,\partial_k \tag{58}$$

where $\partial_\mu=\partial/\partial x^\mu$ and

$$\partial_k=\lambda^\mu_k(x)\,\partial_\mu, \tag{59}$$

the latter denoting the Pfaffian derivatives. Here $\lambda^\mu_k(x)$ are the inverse vierbein fields to $\lambda^k_\mu(x)$ obeying $\lambda^\mu_k(x)\lambda^i_\mu(x)=\delta^i_k$.

A de Sitter gauge transformation of $\psi(x,\xi)$ is defined by

$$\hat{\psi}(x,\xi')=\hat{S}(x)\,\psi(x,\xi) \tag{60}$$

with $\xi'^a=[A(x)]^a_b\,\xi^b$, $A(x)\in SO(4,1)$, and $\hat{S}(x)\in USp(2,2)$ [both with x-dependent parameters]. $A(x)$ and $\hat{S}(x)$ are related by the equation

$$\gamma^a[A^{-1}(x)]^b_a=\hat{S}(x)\gamma^b\hat{S}^{-1}(x), \tag{61}$$

defining the homomorphism $\pm\hat{S}(x)\to A(x)$ of $USp(2,2)$ onto $SO(4,1)$. The gauge invariance of the differentiation D in Eq. (57) entails the gauge transformation property of the connection form $\Gamma^R(x)$ according to the formula

$$\Gamma^R(x)=\hat{S}^{-1}(x)\,\hat{\Gamma}^R(x)\,\hat{S}(x)-i\hat{S}^{-1}(x)\,d\hat{S}(x). \tag{62}$$

This equation is analogous to Eq. (13) above.

We now investigate the decomposition of the connection form $\hat{\Gamma}^R(x)=dx^\mu\Gamma^R_\mu(x)$ according to Eq. (17) using Eq. (54), i.e.,

$$\lambda^i_\mu(x)\,\hat{\Gamma}_i{}^R(x)=\hat{\Gamma}_\mu{}^R(x)=\hat{\Gamma}_\mu{}^{R(\mathfrak{g}')}(x)+\hat{\Gamma}_\mu{}^{R(\mathfrak{t})}(x) \tag{63}$$

with

$$\hat{\Gamma}_\mu{}^{R(\mathfrak{g}')}(x)=\tfrac{1}{2}\hat{\Gamma}_\mu{}^R_{ij}(x)M^{ij} \tag{64}$$

and

$$\hat{\Gamma}_\mu{}^{R(\mathfrak{t})}(x)=\hat{\Gamma}_\mu{}^R_{5i}(x)M^{5i} \tag{65}$$

denoting the Lorentz valued and de Sitter boost valued components of $\hat{\Gamma}_\mu{}^R(x)$, respectively. Clearly, the decomposition (63) is not de Sitter gauge invariant. Nevertheless, this decomposition is interesting in the discussion of certain gauge fixing conditions which may be enforced on a physical theory based on a Cartan bundle formalism. To see this, we write down the analog of the curvature form (18) in the present case. In terms of the matrix coefficient $\hat{\Gamma}_\mu{}^R(x)$ the curvature

expression reads

$$\hat{R}^R_{\mu\nu}(x) = \partial_\mu \hat{\Gamma}_\nu{}^R(x) - \partial_\nu \hat{\Gamma}_\mu{}^R(x) + i[\hat{\Gamma}_\mu{}^R(x), \hat{\Gamma}_\nu{}^R(x)] \quad (66)$$

where the second term on the rhs now refers, in contradistinction to Eq. (18), only to the commutation in the Lie algebra $\mathfrak{g}$ of the bundle. By using Eq. (63) the rhs of Eq. (66) decomposes according to Eqs. (19), (26a), and (26b) into

$$\hat{R}^R_{\mu\nu}(x) = \hat{R}^{R'}_{\mu\nu}(x) + i[\hat{\Gamma}_\mu{}^{R(\mathfrak{t})}(x), \hat{\Gamma}_\nu{}^{R(\mathfrak{t})}(x)] + D'_\mu \hat{\Gamma}_\nu{}^{R(\mathfrak{t})}(x) - D'_\nu \hat{\Gamma}_\mu{}^{R(\mathfrak{t})}(x). \quad (67)$$

The first term on the rhs of Eq. (67) is the curvature constructed from the $\hat{\Gamma}_\mu{}^{R(\mathfrak{g}')}(x)$ due to the Lorentz subgroup, the second term is a $\mathfrak{g}'$-valued contribution to the curvature originating from $\hat{\Gamma}_\mu{}^{R(\mathfrak{t})}(x)$, and the last two terms correspond to the torsion [see Eq. (23)] with

$$D'_\mu \hat{\Gamma}_\nu{}^{R(\mathfrak{t})}(x) = \partial_\mu \hat{\Gamma}_\nu{}^{R(\mathfrak{t})}(x) + i[\hat{\Gamma}_\mu{}^{R(\mathfrak{g}')}(x), \hat{\Gamma}_\nu{}^{R(\mathfrak{t})}(x)]. \quad (68)$$

A gauge fixing condition would consist of the assumption that a certain term on the rhs of Eq. (67) is absent. For example, gravitation gives rise to a similar term as $\hat{R}^{R'}_{\mu\nu}(x)$ corresponding in this case to the $\mathfrak{g}'$-valued connection form $\tilde{\omega}'$ on the Lorentz frame bundle $L(V_4)$. It is thus an interesting question to ask whether the strong interactions—provided they can indeed be characterized by a curvature expression $\hat{R}^R_{\mu\nu}(x)$ derived from a Cartan connection—may possess an *integrable* Lorentz component $\hat{\Gamma}_\mu{}^{R(\mathfrak{g}')}(x)$ in a certain gauge, i.e., have a *vanishing* $\hat{R}^{R'}_{\mu\nu}(x)$. We shall not investigate gauge fixing conditions and their effects on the solution of certain equations relating geometrical and matter quantities further in this paper, and we turn, finally, to a brief discussion of connections on the affine tangent bundle $T_A(V_4)$. Before we do so let us, however, remark, in concluding our discussion of the de Sitter bundle, that even if the base space of the bundle $T^R(V_4)$ is flat Minkowski space—time, characterized by $\lambda^i_\mu(x) = \delta^i_\mu$ and the rhs of Eqs. (34) and (36) being zero, that despite these facts all contributions on the rhs of Eq. (67) may in principle be present. As mentioned, it requires the additional postulate of a gauge fixing condition, motivated by physics, to exclude a certain contribution on the rhs of Eq. (67) having itself a well-defined geometrical significance.

Returning to the generalized affine connection (33) and expanding the $R^n_{p,q}$-valued 1-form $\tilde{\theta}$ in analogy to Eqs. (35) and (54) for $M = V_4$ as

$$\tilde{\theta} = dx^\mu v^i_\mu(x) e_i(\tilde{\xi}) \quad (69)$$

with $(\tilde{\xi}; e_0(\tilde{\xi}), e_1(\tilde{\xi}), e_2(\tilde{\xi}), e_3(\tilde{\xi}))$ denoting an *affine* base in $T_x(V_4)$ with origins $M(x) = (\tilde{\xi}^0, \tilde{\xi}^1, \tilde{\xi}^2, \tilde{\xi}^3)$, the $v^i_\mu(x)$ are seen to determine a field of gauge translations in $T_A(V_4)$. A four-component Dirac spinor field $\phi(x, \tilde{\xi}) = \{\phi^A(x, \tilde{\xi}); A = 1, 2, 3, 4\}$ defined on $T_A(V_4)$, where the first argument, x, refers to the point on V_4 and the second argument, $\tilde{\xi}$, refers to the point in the affine tangent space at x, can now be differentiated in a Poincare gauge invariant way according to the formula [compare Eqs. (51) and (57)]

$$\tilde{D}\phi(x, \tilde{\xi}) = d\phi(x, \tilde{\xi}) + i\Gamma(x)\phi(x, \tilde{\xi}) = dx^\mu [\partial_\mu + (i/2)\Gamma_{\mu ij}(x) M^{ij} + v^i_\mu(x) \tilde{\partial}_i] \phi(x, \tilde{\xi}). \quad (70)$$

Here we have represented the basis vectors $e_i(\tilde{\xi})$ by the operators $\tilde{\partial}_i = \partial/\partial\tilde{\xi}^i$, $i = 0, 1, 2, 3$, affecting the second argument of $\phi(x, \tilde{\xi})$. In Eq. (70), $M^{ij} = (i/2)[\gamma^i, \gamma^j]$ are the generators of SO(3, 1) in the 4×4 spinor representation (here with unprimed matrix indices to distinguish them from the indices in the de Sitter case treated above), and $\Gamma_{\mu ij}(x) = -\Gamma_{\mu ji}(x) = \lambda^k_\mu(x)\Gamma_{kij}(x)$ are the coefficients of the connection on the Lorentz frame bundle with the first index turned into a global Greek index. Finally, the sixteen fields $v^i_\mu(x)$ are the contributions related to gauge translations, i.e., to the shift of the origin, $M(x)$, of the affine frame in the local affine tangent space at $x \in V_4$. Poincaré gauge transformations for a spinor field on $T_A(V_4)$ can now be defined by

$$\phi'(x, \tilde{\xi}') = \exp(-i\bar{v}^i(x) P_i) S(x) \phi(x, \tilde{\xi}). \quad (71)$$

Here the 4×4 matrix $S(x)$ describes an x-dependent *rotation* of the local Lorentz frame $e_i(\tilde{\xi})$ (generated by the M^{ij} introduced before), and the exponential factor represents an x-dependent *translation* [with parameter $\bar{v}^i(x)$, and P_i as defined in Eq. (48)] of the origin of the local frame in the local affine tangent space at $x \in V_4$.

Restricting the gauge group ISO(3, 1) to the homogenous Lorentz subgroup, i.e., disregarding gauge translation, Eqs. (70) and (71) assume the following form where we have, furthermore, made use of Eqs. (54), (58), and (59), and kept the same symbol for the spinor ϕ but dropped the now superfluous argument $\tilde{\xi}$:

$$\tilde{D}'\phi(x) = dx^\mu [\partial_\mu \phi(x) + (i/2)\Gamma_{\mu ij}(x) M^{ij} \phi(x)] = \omega^k(x)[\partial_k + (i/2)\Gamma_{kij}(x) M^{ij}] \phi(x), \quad (72)$$

$$\phi'(x) = S(x)\phi(x). \quad (73)$$

Equation (72) represents the usual formula for the covariant differentiation of a spinor quantity, $\phi(x)$, defined on a curved space—time manifold V_4 with Eq. (73) describing the effect on $\phi(x)$ of a change of gauge, i.e., a transition to another moving Lorentz frame on V_4 or, synonymously, a transition to another cross section on the Lorentz frame bundle $L(V_4)$ over V_4. Furthermore, the $\tilde{\omega}'_{ij}(x) = \omega^k(x)\Gamma_{kij}(x)$ appearing in Eq. (72) define a connection in $L(V_4)$. Equation (72) can be viewed as resulting from the Cartan connection (33) with $\tilde{\theta}$ being identified with the fundamental 1-form φ defined in Eqs. (35) and (54), i.e.,

$$\tilde{\omega}' + \varphi \xrightarrow{\text{in spinor form}} (i/2)\omega^k(x)\Gamma_{kij}(x) M^{ij} + \omega^k(x) e_k(x), \quad (74)$$

the latter being identical with the operator $\tilde{D}'$ of Eq. (72) when $e_k(x) = \partial_k$, where now the ∂_k can be identified with the $\tilde{\partial}_k$ introduced before. Thus to each *generalized affine connection* (or special Cartan connection) (33) corresponds in a unique way a linear connection $\tilde{\omega}'$

which is because of historical reasons often also called an affine connection.

[1]W. Drechsler, Fortschr. Physik 23, 607 (1975).

[2]W. Drechsler, "Wave equations for extended hadrons," in *Group Theoretical Methods in Physics, Fourth Int. Colloquium, Nijmegen* 1975, edited by A. Janner, T. Janssen, and M. Boon, Lecture Notes in Physics, Vol. 50 (Springer-Verlag, Berlin, 1976), p. 37.

[3]W. Drechsler, "Currents in a theory of strong interactions based on a fiber bundle geometry," Preprint MPI-PAE/PTh 8/76 (Max-Planck-Institut, 1976, Found. Phys., to be published); see also Proceedings of the Int. Symposium on Mathematical Physics, Mexico City, January 1976, p. 495.

[4]Ch. Ehresmann, "Les connexions infinitesimales dans un espace fibré differentiable," Colloque de Topologie, Bruxelles, 1950, p. 29.

[5]S. Kobayashi, Ann. Math. 43, 119 (1957).

[6]T.W.B. Kibble, J. Math. Phys. 2, 212 (1961).

[7]K. Hayashi and T. Nakano, Progr. Theor. Phys. 38, 491 (1967).

[8]See, for example, S. Kobayashi and K. Nomizu, *Foundation of Differential Geometry* (Wiley, New York, 1963), Vol. 1, A. Lichnerowicz, *Théorie globale des connexions et des groupes d'holomie* (Edizioni Cremonese, Rome, 1962).

[9]For details see also S. Kobayashi and K. Nomizu, Ref. 8.

[10]R denotes the radius of curvature of V_4'.

[11]For a more detailed discussion and proofs see the literature quoted in Ref. 8 and also the book of Y. Choquet-Bruhat, *Géométrie différentielle et systèmes extérieurs* (Dunod, Paris, 1968).

[12]Analogously, left translation by an element g is defined by $L_g u = gu$.

[13]$g^{-1}(g+dg) = e + g^{-1}dg$.

[14]For brevity we leave out the projection and the family of homeomorphisms in the argument of E.

[15]There exists a mapping of G' [viewed as transformation group in F leaving $O \in F$ fixed] into $Gl(n,R)'$ [viewed as a transformation group in $T_0(F)$] which can also be used to associate $E(M,F,G',P')$ (see below) with $T'(M)$. In our later discussion in Sec. IV we shall assume that this mapping establishes an isomorphism, i.e., that G' can be identified with $Gl(n,R)'$. The group $Gl(n,R)'$ operating in $T_0(F)$ and the group $Gl(n,R)$ operating in $T(M)$ are conceptually two different gauge groups. The former belonging, as subgroup of G, to the fiber of the Cartan bundle, the latter belonging to the fiber of the frame bundle over M.

[16]Compare Ref. 5 where the case of a so-called weakly reductive bundle E is treated for which $[\mathfrak{g}, \mathfrak{t}] \subseteq \mathfrak{t}$.

[17]If Eq. (15) does not hold true, one can only say that $g' \in G'$ induces a linear transformation $L_{g'}$ of $T_0(F)$ which does not correspond to the adjoint representation of G' in $T_0(F)$. In this case Eq. (16b) reads $\theta(X'g') = L_{g'}^{-1}\theta(X')$ with θ being a $T_0(F)$-valued form.

[18]A comprehensive discussion of these coset spaces can be found in R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974). However, for a more advanced treatise see also S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).

[19]See R. Gilmore, Ref. 18.

[20]$R^n_{p,q}$ is an n-dimensional pseudo-Euclidean space with signature given by $(p,q) = (---\cdots p \text{ times}, +++\cdots q \text{ times})$ and $p + q = n$.

[21]Compare S. Kobayashi and K. Nomizu, Ref. 8, Chap. III. Since the structural group of $T_A(M)$, as discussed here, is related to the special *orthogonal* group $SO(p,q)$, one also refers to $\bar{\omega}'$ as to the form of a *metric* connection.

[22]Identical covariant and contravariant indices are automatically summed in the following way: a, b, c, $\cdots$ over 0, 1, 2, 3, 5; i, j, k, $\cdots$ over 0, 1, 2, 3, and Greek indices, appearing below, over 0, 1, 2, 3.

[23]For a more detailed discussion in particular with regard to the sixteen possible orientations one can choose in the soldering of the bundle $T^R(V_4)$ see Appendix A of Ref. 3.

Covariant inverse problem of the calculus of variations

Henry F. Ahner

Adelphi University, Garden City, New York 11530

Allan E. Moose

159 Dolphin Road, East Quogue, New York 11942
(Received 4 March 1977)

The solution of the covariant inverse problem of the calculus of variations (that of finding a Lagrangian for a given set of dynamical equations) is presented as a generalization of the flat space formalism of Atherton and Homsy. Known Lagrangians such as those for the complex scalar field, vector gauge fields, and Einstein's equations are used as examples. Additional insight into the formalism is provided by new examples that include a two-tensor theory, a method for obtaining conservation laws directly from dynamical equations, and a Hamiltonian formulation for higher order, nonlinear, differential equations.

INTRODUCTION

The idea of using the calculus of variations to obtain, from a variational principle, a dynamical system's equations of motion is well known. Briefly, the determination of the stationary points of a suitably chosen functional, the Lagrangian, yields the equations of motion. Until recent work by Atherton and Homsy,[1] (AH), solution of the inverse problem, finding the Lagrangian whose stationary points are described by the equations of motion, depended largely on the skill and intuition of the investigator. Atherton and Homsy's procedure allows one to construct a variational principle for any given set of equations. Their paper cites applications to nonlinear equations in flat space.

The major purpose of this paper is to present the generalization of Atherton and Homsy's methods to tensor equations in curved space–time (Sec. I). The paper's secondary purposes, illustrating the potential power of the new formalism, include discussions of Noether's theorem (Sec. III), the construction of a generalized Hamiltonian mechanics directly from the equations of motion (Sec. IV), and the presentation of a two-tensor theory of gravitation (Sec. II D).

I. FORMALISM

A. Inverse problem of calculus of variations

The solution of the inverse problem of the calculus of variations is best approached within the context of functional analysis. Consider an expression of the form

$$F^{|}(y)=\int N(y)\,\phi dV=0, \tag{1}$$

where the slash, |, denotes the canonical variational procedure, $N(y)$ is a differential operator acting on a variable y, and ϕ is the variation of y. The fundamental lemma of the calculus of variations provides the result that

$$N(y)=0. \tag{2}$$

This is the Euler–Lagrange equation. From the more general view of functional analysis, one could read the | in Eq. (1) as "an appropriate derivative operator," while ϕ is regarded as an arbitrary element of function space.

Atherton and Homsy[1] show that the differential operator needed on the left side of Eq. (1) is the Frechet derivative. It is defined on the function spcae and is a higher order abstract analog of the ordinary derivative. We give an operational definition of the Frechet derivative as follows. Consider a general tensor operator, $N^A(y_B; y_{B,\mu}; y_{B,\mu\nu})$. (Capital letters A, B, $\cdots$ indicate any number of tensor indices. A, B, $\cdots$ will be raised and lowered by g^{AB} and g_{AB}. Greek letters μ, ν, $\cdots$ range over space–time. A repeated index of any type implies the summation convention. The comma denotes a partial derivative with respect to the indices to its right.) The operator N^A may be of any order.

With these conventions the Fréchet derivative of the operator N^A is defined as

$$N^{A|B}\phi_B=\lim_{\epsilon\to 0}\left(\frac{1}{\epsilon}N^A[(y_B+\epsilon\phi_B)_{,\Pi}]-\frac{1}{\epsilon}N^A(y_{B,\Pi})\right). \tag{3}$$

Here the capital Greek indices Π, Λ, $\cdots$ signify the set of all combinations of tensor indices, $\Pi=\{\text{None},\ \mu,\ \mu\nu,\ \mu\nu\rho,\ \cdots\}$. For example, if N^A is a second-order differential operator, Eq. (3) is shorthand for

$$N^{A|B}\phi_B=\lim_{\epsilon\to 0}\left(\frac{1}{\epsilon}N^A[(y_B+\epsilon\phi_B);\ (y_B+\epsilon\phi_B)_{,\mu};\ (y_B+\epsilon\phi_B)_{,\mu\nu}]-\frac{1}{\epsilon}N^A(y_B;\ y_{B,\mu};\ y_{B,\mu\nu})\right).$$

As with other index types, the summation convention will be used for pairs of capital Greek letters. Equation (3) is

$$N^{A|B}\phi_B=\frac{\partial}{\partial\epsilon}\{N^A[(y_B+\epsilon\phi_B)_{,\Pi}]\}\Big|_{\epsilon=0}. \tag{4}$$

Equations (3) and (4) define the Frechet derivative of the operator N^A in the direction ϕ_B.

Now consider the functional F defined by

$$F\equiv\int y_A\int_0^1 N^A(\lambda y_{B,\Pi})(-g)^{1/2}d\lambda\,d^4x, \tag{5}$$

where λ is a parameter that is homogeneous with $y_{B,\Pi}$. We shall show that the requirement that the Frechet derivative of F be zero, $F^{|B}\phi_B=0$, yields the Euler–Lagrange equations

$$N^A(y_B)=0. \tag{6}$$

In the process, we will obtain a set of conditions that the

operator N^A must satisfy in order to be obtainable from this type of variational principle. Following AH, operators satisfying these conditions will be called potential. It turns out that operators whose highest order is odd cannot be potential. Furthermore, the conditions for an operator to be potential are quite restrictive. Fortunately there is a technique for formulating a variational principle in these excluded situations. AH called this second principle the composite variational principle.

We begin by taking the Frechet derivative of Eq. (5),

$$F^{|B}\phi_B = \frac{\partial}{\partial\epsilon}\left(\int(y_A+\epsilon\phi_A)\int_0^1 N^A[\lambda(y_B+\epsilon\phi_B)_{,\Pi}](-g)^{1/2}d\lambda\,d^4x\right), \tag{7}$$

or

$$F^{|B}\phi_B = \int\left(\phi_A\int_0^1 N^A(\lambda y_B)d\lambda + y_A\int_0^1 N^{A|\lambda B}\phi_B d\lambda\right)\times(-g)^{1/2}d^4x. \tag{8}$$

Here the λ in the superscript λB is a reminder that N^A is a functional of λy_B. Now define the adjoint operator $\tilde{N}^{B|A}$ such that

$$\int\psi_A N^{A|B}\phi_B(-g)^{1/2}d^4x = \int\phi_B\tilde{N}^{B|A}\psi_A(-g)^{1/2}d^4x, \tag{9}$$

for all functions ψ_A and ϕ_B in function space. Using the adjoint in Eq. (8) we obtain

$$F^{|B}\phi_B = \int\left(\phi_A\int_0^1 N^A(\lambda y_B)d\lambda + \phi_B\int_0^1\tilde{N}^{B|\lambda A}y_A d\lambda\right)\times(-g)^{1/2}d^4x. \tag{10}$$

We require that the operator $N^{A|B}$ satisfy the condition

$$\int\psi_A N^{A|B}\phi_B(-g)^{1/2}d^4x = \int\phi_A N^{A|B}\psi_B(-g)^{1/2}d^4x. \tag{11}$$

Referring to Eq. (9), we see that this implies that

$$\int\phi_A N^{A|B}\psi_B(-g)^{1/2}d^4x = \int\phi_B\tilde{N}^{B|A}\psi_A(-g)^{1/2}d^4x. \tag{12}$$

Substituting Eq. (12) into Eq. (10) gives us

$$F^{|B}\phi_B = \int\left(\phi_A\int_0^1 N^A(\lambda y_B)d\lambda + \phi_A\int_0^1 N^{A|\lambda B}y_B d\lambda\right)\times(-g)^{1/2}d^4x. \tag{13}$$

Note that the Frechet derivative in Eq. (13) is

$$N^{A|\lambda B}y_B = \frac{\partial N^A(\lambda y_B)}{\partial(\lambda y_B)_{,\Pi}}(\lambda y_B)_{,\Pi}. \tag{14}$$

That is, for a second order operator

$$N^{A|\lambda B}y_B = \frac{\partial N^A}{\partial(\lambda y_B)}(\lambda y_B) + \frac{\partial N^A}{\partial(\lambda y_B)_{,\mu}}(\lambda y_B)_{,\mu} + \frac{\partial N^A}{\partial(\lambda y_B)_{,\mu\nu}}(\lambda y_B)_{,\mu\nu}.$$

Substituting the identity

$$N^{A|\lambda B}y_B = \lambda\frac{d}{d\lambda}N^A(\lambda y_B) = \frac{d}{d\lambda}[\lambda N^A(\lambda y_B)] - N^A(\lambda y_B) \tag{15}$$

into Eq. (13), we obtain

$$F^{|B}\phi_B = \int\phi_A\int_0^1\frac{d}{d\lambda}[\lambda N^A(\lambda y_B)]d\lambda(-g)^{1/2}d^4x. \tag{16}$$

This becomes

$$F^{|B}\phi_B = \int\phi_A N^A(y_B)(-g)^{1/2}d^4x. \tag{17}$$

Setting $F^{|B}\phi_B = 0$ implies $N^A(y_B) = 0$, since ϕ_A is arbitrary. Therefore, the conditions that N^A must satisfy to be called potential are embodied in the symmetry condition

$$\int\phi_A N^{A|B}\psi_B(-g)^{1/2}d^4x = \int\phi_A\tilde{N}^{A|B}\psi_B(-g)^{1/2}d^4x. \tag{18}$$

Now let us derive an explicit expression for the symmetry condition. Using Eq. (14) we can write

$$\int\psi_A N^{A|B}\phi_B(-g)^{1/2}d^4x = \int\psi_A\frac{\partial N^A}{\partial y_{B,\Pi}}\phi_{B,\Pi}(-g)^{1/2}d^4x. \tag{19}$$

The individual terms of the right-hand side of Eq. (19) may be integrated by parts. If we require that the surface terms vanish on the boundary of the domain of integration, Eq. (19) becomes

$$\int\psi_A N^{A|B}\phi_B(-g)^{1/2}d^4x = \sum_\Pi(-1)^{|\Pi|}\int\left((-g)^{1/2}\psi_A\frac{\partial N^A}{\partial y_{B,\Pi}}\right)_{,\Pi}\phi_B d^4x. \tag{20}$$

Here $|\Pi| = 0$ if there is an even number of indices and $|\Pi| = 1$ if there is an odd number of indices. The expression in the parenthesis on the right side of Eq. (20) is the adjoint operator. From Eq. (18) we see that the symmetry condition requires that

$$N^{A|B}\psi_B = \sum_\Pi(-1)^{|\Pi|}(-g)^{-1/2}\left((-g)^{1/2}\psi_B\frac{\partial N^B}{\partial y_{A,\Pi}}\right)_{,\Pi}$$

or

$$\frac{\partial N^A}{\partial y_{B,\Pi}}\psi_{B,\Pi} = \sum_\Pi(-1)^{|\Pi|}(-g)^{-1/2}\left((-g)^{1/2}\psi_B\frac{\partial N^B}{\partial y_{A,\Pi}}\right)_{,\Pi}. \tag{21}$$

On comparison of coefficients of ψ_B to arbitrary power of derivative Π, we obtain

$$\frac{\partial N^A}{\partial y_{B,\Pi}} = \sum_\Lambda(-1)^{|\Pi+\Lambda|}\binom{\Pi+\Lambda}{\Lambda}\times(-g)^{-1/2}\left((-g)^{1/2}\frac{\partial N^B}{\partial y_{A,\Pi\Lambda}}\right)_{,\Lambda}. \tag{22}$$

Here $\binom{\Pi+\Lambda}{\Lambda}$ is the Bernoulli symbol involving factorials of the number of each type of index.

As an example of the conditions expressed in Eq. (22), we list the symmetry conditions for a second order operator:

$$\frac{\partial N^A}{\partial y_B} = \frac{\partial N^B}{\partial y_A} - (-g)^{-1/2}\left[\left((-g)^{1/2}\frac{\partial N_B}{\partial y_{A,\mu}}\right)_{,\mu} - \left((-g)^{1/2}\frac{\partial N^B}{\partial y_{A,\mu\nu}}\right)_{,\mu\nu}\right]$$

$$\frac{\partial N^A}{\partial y_{B,\mu}} = -\frac{\partial N^B}{\partial y_{A,\mu}} + 2(-g)^{-1/2}\left((-g)^{1/2}\frac{\partial N^B}{\partial y_{A,\mu\nu}}\right)_{,\nu}, \tag{23}$$

$$\frac{\partial N^A}{\partial y_{B,\mu\nu}} = \frac{\partial N^B}{\partial y_{A,\mu\nu}}.$$

The requirement that the surface terms obtained in integrating Eq. (19) by parts vanish leads to a number of

boundary terms of the form

$$\frac{\partial N^A}{\partial y_{B,\Pi}} = 0, \tag{24}$$

for all Π on the boundary of the region.

From Eq. (21) or (22), we see why there is no potential type variational principle for odd order equations. The presence of the $(-1)^{|\Pi|}$ on the right-hand side rules out the possibility of odd order equations satisfying the symmetry conditions. Furthermore, as illustrated by Eq. (23), the symmetry conditions are quite restrictive. Thus it is convenient to have a procedure for constructing a variational principle that avoids the restrictions of the symmetry requirement.

Such a procedure is the formulation of the composite variational principle. Suppose we have an operator $N^A(z_B)$ of any differential order. Obtain an associated set of variables v_A and form the functional

$$F = \int v_A N^A(z_B)(-g)^{1/2} d^4x. \tag{25}$$

Now take the Frechet derivative of Eq. (25), treating v_A and z_A independently,

$$F' = \frac{\partial}{\partial \epsilon} \int (v_A + \epsilon \psi_A) N^A(z_B + \epsilon \phi_B)(-g)^{1/2} d^4x \Big|_{\epsilon=0}, \tag{26}$$

$$F' = \int \psi_A N^A(z_B)(-g)^{1/2} d^4x + \int v_A N^{A|B} \phi_B (-g)^{1/2} d^4x. \tag{27}$$

Introduce the adjoint operator, $\tilde{N}^{B|A}$, into the second integral in Eq. (27) to obtain

$$F' = \int \psi_A N^A(z_B)(-g)^{1/2} d^4x + \int \phi_B \tilde{N}^{B|A} v_A (-g)^{1/2} d^4x. \tag{28}$$

Since ψ_A and ϕ_B are arbitrary functions, Eq. (28) implies

$$N^A(z_B) = 0 \tag{29}$$

and

$$\tilde{N}^{B|A} v_A = 0, \tag{30}$$

when $F' = 0$. The price paid to obtain Eq. (29) from a variational principle, when the symmetry conditions need not be satisfied, is the introduction of new variables v_A satisfying equations of motion given by Eq. (30). Only if Eq. (30) and v_A are physically meaningful entities can this procedure be allowed in physics. Examples will be given to illustrate this point later.

B. Frechet derivative—covariant form

We know that for an integral expression in curved space—time to have meaning, the integrand must be a scalar density. It is a straightforward matter to show that an integrand involving the Frechet derivative, for example, in the expression

$$\int y_A \frac{\partial N^A}{\partial y_{B,\Pi}} \phi_{B,\Pi} (-g)^{1/2} d^4x,$$

is a scalar density. However, the individual terms in this integrand do not possess tensor character because partial differentiation is not a covariant operation. In certain applications it is convenient to have all quantities expressed covariantly. Re-expression of the Frechet derivative in terms of covariant derivatives may be illustrated as follows. For a second order operator $N^A(y_B; y_{B,\mu}; y_{B,\mu\nu})$ whose Frechet derivative is

$$N^{A|B}\phi_B = \frac{\partial N^A}{\partial y_B}\phi_B + \frac{\partial N^A}{\partial y_{B,\mu}}\phi_{B,\mu} + \frac{\partial N^A}{\partial y_{B,\mu\nu}}\phi_{B,\mu\nu},$$

write down the definitions of $y_{B;\mu}$ and $y_{B;\mu\nu}$. Now using these definitions and the chain rule, express $\partial N^A/\partial y_B$, $\partial N^A/\partial y_{B,\mu}$, and $\partial N^A/\partial y_{B,\mu\nu}$ in terms of covariant derivatives and Christoffel symbols. The result may be regrouped into

$$N^{A|B}\phi_B = \frac{\partial N^A}{\partial y_B}\phi_B + \frac{\partial N^A}{\partial y_{B;\mu}}\phi_{B;\mu} + \frac{\partial N^A}{\partial y_{B;\mu\nu}}\phi_{B;\mu\nu}. \tag{31}$$

The general expression for an operator of any order is

$$N^{A|B}\phi_B = \frac{\partial N^A}{\partial y_{B;\Pi}}\phi_{B;\Pi}. \tag{32}$$

Investigation of the transformation of the individual terms in Eqs. (31) and (32) under general coordinate transformations $y_B \to y'_B = y'_B(x^\mu)$ shows that the quantity $\partial N^A/\partial y_{B;\Pi}(-g)^{1/2}$ transforms as a tensor density of the type $T^{AB\Pi}$. One may show that the presence of the covariant derivative in Eq. (32) does not affect the integration by parts procedure used in obtaining the symmetry conditions in the previous section. For example,

$$\begin{aligned} 0 &= \int [(-g)^{1/2} X^{A\mu} Y_A]_{,\mu} d^4x \\ &= \int (X^{A\mu} Y_A)_{;\mu} (-g)^{1/2} d^4x \\ &= \int X^{A\mu}{}_{;\mu} Y_A (-g)^{1/2} d^4x + \int X^{A\mu} Y_{A;\mu} (-g)^{1/2} d^4x. \end{aligned}$$

Therefore, one may rewrite the symmetry condition, Eq. (22), as

$$\frac{\partial N^A}{\partial y_{B;\Pi}} = \sum_\Lambda (-1)^{|\Pi - \Lambda|} \binom{\Pi\Lambda}{\Lambda} \left(\frac{\partial N^B}{\partial y_{A;\Pi\Lambda}}\right)_{;\Lambda}. \tag{33}$$

II. EXAMPLES

A. Complex scalar field

As an example of the use of the composite principle, consider a complex scalar field ψ. As the auxiliary set of variables choose the complex conjugate ψ^*. Then the variational principle, Eq. (25), becomes

$$F(\psi, \psi^*) = \int \psi^* N(\psi)(-g)^{1/2} d^4x. \tag{34}$$

Therefore, from Eqs. (29) and (30), the Euler—Lagrange equations are

$$N(\psi) = 0 \text{ and } \tilde{N}^1(\psi^*) = 0.$$

Thus if the scalar field satisfies the Schrödinger equation

$$N(\psi) = +(2m)^{-1}\nabla^2\psi - V(x, y, z)\psi + i\psi_{,t} = 0$$

with $\hbar = 1$, m the particle's mass, and V a potential function, then Eq. (34) is in flat space,

$$F(\psi, \psi^*) = \int [\, \psi^*(2m)^{-1}\nabla^2\psi - \psi^*\nabla\psi + i\psi^*\psi_{,t}]\, d^4x. \tag{35}$$

Computation of the adjoint operator $\tilde{N}^1\psi^* = 0$, yields

$$(2m)^{-1}\nabla^2\psi^* - V\psi^* - i\psi^*_{,t} = 0.$$

We note that the adjoint equation represents a parti-

cle whose energy is the negative of the original particle. Since this equation is not potential, only the action given by Eq. (35) and the composite principle may be considered.

B. Gauge fields

Consider the case of a space–time vector field that also possesses an internal transformation group. Let ψ_i be a space–time scalar field[2] that transforms under an internal group as

$$\bar{\delta}\psi_i = i\epsilon^A L_A{}^j{}_i \psi_j,$$

where the $L_A{}^j{}_i$ are a particular matrix representation[3] of the group and ϵ^A are transformation parameters. It is well known that fields $B^A{}_\mu$ must be introduced to construct the covariant derivatives of ψ_i. From $B^A{}_\mu$, which does not transform like a representation of the internal group, one constructs field tensors, $G^A{}_{\mu\nu}$, that do;

$$G^A{}_{\mu\nu} \equiv B^A{}_{\mu,\nu} - B^A{}_{\nu,\mu} + igC^A{}_{BC} B^B{}_\mu B^C{}_\nu.$$

Note that $C^A{}_{BC}$ are structure constants for the internal group and recall[2] that

$$\psi_{i;\mu} = \psi_{i,\mu} - g\, i L_A{}^j{}_i B^A{}_\mu \psi_j,$$
$$(L_A, L_B)^j{}_i = C^C{}_{AB} L^j{}_i,$$
$$L^A{}_{BC} = -C^A{}_{BC},$$
$$G^A{}_{\mu\nu} = B^A{}_{\mu;\nu} - B^A{}_{\nu;\mu} - ig C^A{}_{BC} B^B{}_\mu B^C{}_\nu.$$

Now allowing g^{AB} to be a group metric that may be used to raise and lower internal indices, we consider the sourceless field equations for the field $B^A{}_\mu$,

$$N^{A\mu} = G^{A\mu\nu}{}_{;\nu}$$
$$= G^{A\mu\nu}{}_{;\nu} + igC^A{}_{BC} B^C{}_\nu G^{B\mu\nu} = 0. \tag{36}$$

In this section, the index pair $A\mu$ plays the role of the index A in the previous sections. One verifies by substitution that Eq. (36) satisfies conditions (23). Accordingly, we have the result that the gauge field equations (36) are potential. Equation (5) becomes

$$F = \int B_{A\mu} \int_0^1 N^{A\mu}(\lambda B^B{}_\nu)(-g)^{1/2} d\lambda\, d^4x, \tag{37}$$

and accordingly the Lagrangian $\mathcal{L}$ is

$$\mathcal{L} = (-g)^{1/2} \int_0^1 B_{A\mu} N^{A\mu}(\lambda B^B{}_\nu) d\lambda. \tag{38}$$

To verify that this is equivalent to the standard gauge field Lagrangian given by

$$\tfrac{1}{4}(-g)^{1/2} G_{A\mu\nu} G^{A\mu\nu}, \tag{39}$$

proceed by substituting Eq. (36) into Eq. (37). After integration of the first term of the resulting expression by parts over space–time, perform the indicated λ integration. The terms regroup immediately to (39).

This example is typical of most of the Lagrangians obtained from field equations. They are the expected ones, modulo an integration by parts over space–time. As a general rule, the standard Lagrangian may be obtained by integrating by parts so that the differential order of $\mathcal{L}$ is minimized.

C. Einstein's equations

The verification that Einstein's equations are potential is a straightforward but tedious exercise. Defining $I_{\mu\nu}$ as the Frechet derivative of $R_{\mu\nu}$,

$$I_{\mu\nu} \equiv \frac{\partial R_{\mu\nu}}{\partial g_{\alpha\beta}} f_{\alpha\beta} + \frac{\partial R_{\mu\nu}}{\partial g_{\alpha\beta,\gamma}} f_{\alpha\beta,\gamma} + \frac{\partial R_{\mu\nu}}{\partial g_{\alpha\beta,\gamma\delta}} f_{\alpha\beta,\gamma\delta},$$

we find that,[4]

$$I_{\mu\nu} = \tfrac{1}{2} g^{\alpha\beta}[f_{\alpha\beta;(\mu\nu)} - f_{\alpha\mu;\nu\beta} - f_{\alpha\nu;\mu\beta} + f_{\mu\nu;\alpha\beta}]. \tag{40}$$

Using Eq. (40), the Frechet derivative of Einstein's tensor, $G_{\mu\nu}$, may be calculated to be

$$G_{\mu\nu|\rho\sigma} f^{\rho\sigma} = I_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} g^{\rho\sigma} I_{\rho\sigma} - \tfrac{1}{2} f_{\mu\nu} R + \tfrac{1}{2} g_{\mu\nu} R^{\alpha\beta} f_{\alpha\beta}. \tag{41}$$

Noting that the terms in $G_{\mu\nu}$ are all of the same order in λ, λ^1, (i.e., the order of $g^{\mu\nu}$ is λ^{-1}) we find the Lagrangian to be

$$\mathcal{L} = g^{\mu\nu} \int_0^1 \lambda(R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R)(-g)^{1/2} d\lambda.$$

This simplifies to

$$\mathcal{L} = -\tfrac{1}{2}(-g)^{1/2} R,$$

immediately.

D. Two-tensor gravity

So far all the Lagrangians discussed were well known examples. As an example of the possibilities of generating new theories from the formalism developed, consider the case of Einstein's equations but use the composite potential method. Taking $f^{\mu\nu}$ as a symmetric second rank tensor, the associated field, we write a composite Lagrangian

$$\mathcal{L} = (-g)^{1/2} f^{\mu\nu}(G_{\mu\nu} + \tfrac{1}{2}\Lambda g_{\mu\nu} - \kappa T_{\mu\nu}). \tag{42}$$

Here $G_{\mu\nu}$, Λ, κ, and $T_{\mu\nu}$ are the Einstein tensor, the cosmological constant, the gravitational coupling constant, and the matter stress-energy tensor respectively. All indices are raised and lowered with $g^{\mu\nu}$ or $g_{\mu\nu}$.

From Eqs. (29), (30), and (41), the Euler–Lagrange equations are

$$G_{\mu\nu} + \tfrac{1}{2}\Lambda g_{\mu\nu} = \kappa T_{\mu\nu},$$

as desired, and

$$0 = I_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} I - \tfrac{1}{2} f_{\mu\nu} R + \tfrac{1}{2} g_{\mu\nu} f_{\rho\sigma} R^{\rho\sigma}$$
$$- \frac{\Lambda}{2} f_{\mu\nu} - f^{\rho\sigma} \kappa \frac{\partial T_{\rho\sigma}}{\partial g^{\mu\nu}}. \tag{43}$$

Note $I \equiv I_\mu{}^\mu$. It has been assumed here that no derivatives of the metric appear in $T_{\mu\nu}$. Observe that while the equation for $g_{\mu\nu}$ is highly nonlinear, the equation for $f_{\mu\nu}$ is linear. Letting

$$\gamma_{\mu\nu} \equiv f_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} f,$$

and considering the special case where

$$\gamma_\mu{}^\rho{}_{;\rho} \equiv \tau_\mu = 0,$$

we find that Eq. (43) reduces to

$$\Box^2 \gamma_{\mu\nu} - \Lambda(\gamma_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}\gamma) = J_{\mu\nu}$$
$$\equiv 2 f^{\rho\sigma} \kappa \frac{\partial T_{\rho\sigma}}{\partial g^{\mu\nu}} \tag{44}$$

in the limit of flat space. Writing Eq. (44) in terms of

its irreducible tensor components $\Gamma_{\mu\nu}$ and γ,

$$\gamma=\gamma_{\mu\nu}\eta^{\mu\nu},$$

$$\Gamma_{\mu\nu}=\gamma_{\mu\nu}-\tfrac{1}{4}\eta_{\mu\nu}\gamma,$$

we find

$$\Box^2\gamma+\Lambda\gamma=J=J_{\mu\nu}\eta^{\mu\nu}, \tag{45}$$

$$\Box^2\Gamma_{\mu\nu}-\Lambda\Gamma_{\mu\nu}=J_{\mu\nu}-\tfrac{1}{4}\eta_{\mu\nu}J. \tag{46}$$

If we consider the particle fields in $T_{\mu\nu}$ in Eq. (42) as part of the Fréchet derivative, then the additional Euler–Lagrange equations must be the particles' equations of motion.[5]

III. CONSERVATION LAWS

The formalism developed offers the possibility of obtaining conservation laws directly from equations of motion. To explore this possibility, we will briefly review Noether's theorem[6] and then extend it to cover the present situation.

Let $\mathcal{L}$ be a Lagrangian density that is a function of some fields y_A and the derivatives of these fields $y_{A,\Pi}$. Then under a coordinate transformation, $x^\mu \to x'^\mu = x^\mu + \xi^\mu$, the change in $\mathcal{L}$, $\bar{\delta}\mathcal{L}$, is given by

$$\bar{\delta}\mathcal{L}=\frac{\partial\mathcal{L}}{\partial y_{A,\Pi}}\bar{\delta}y_{A,\Pi}, \tag{47}$$

where $\bar{\delta}y_A$ is the change in y_A under the transformation. The variational derivative of $\mathcal{L}$, $\delta\mathcal{L}/\delta y_A$ is given by

$$\frac{\delta\mathcal{L}}{\delta y_A}=\sum_\Pi(-1)^{|\Pi|}\left(\frac{\partial\mathcal{L}}{\partial y_{A,\Pi}}\right)_{,\Pi}. \tag{48}$$

Since $\mathcal{L}$ is a scalar density of weight $+1$, we have

$$\bar{\delta}\mathcal{L}=-(\mathcal{L}\xi_\rho)_{,\rho}. \tag{49}$$

Combining Eqs. (47), (48), and (49), we obtain

$$\frac{\delta\mathcal{L}}{\delta y_A}\bar{\delta}y_A=\left[-\mathcal{L}\xi^\mu-\sum_{\Pi,\Theta}(-1)^{|\Pi|}\bar{\delta}y_{B,\Theta}\left(\frac{\partial\mathcal{L}}{\partial y_{B,\mu\Pi\Theta}}\right)_{,\Pi}\right]_{,\mu}\equiv t^\mu{}_{,\mu}. \tag{50}$$

Whenever y_A is an absolute object and, $\bar{\delta}y_A=0$, a conservation law, $t^\mu{}_{,\mu}=0$, follows with

$$t^\mu=-\mathcal{L}\xi^\mu-\sum_{\Pi,\Theta}(-1)^{|\Pi|}\delta y_{B,\Theta}\left(\frac{\partial\mathcal{L}}{\partial y_{B,\mu\Pi\Theta}}\right)_{,\Pi}, \tag{51}$$

being the conserved quantities.

Equation (51) may be expressed in terms of the equations of motion themselves. Using Eq. (5), we find

$$t^\mu=-\int_0^1 d\lambda\left[(-g)^{1/2}y^AN_A(\lambda y_B)\xi^\mu+\sum_{\Pi,\Theta}(-1)^{|\Pi|}\bar{\delta}y_{B,\Theta}\left((-g)^{1/2}y^A\frac{\partial N_A(\lambda y_B)}{\partial y_{B,\mu\Pi\Theta}}\right)_{,\Pi}\right].$$

Defining M_A as

$$M_A=\int_0^1 d\lambda N_A(\lambda y_B),$$

we have

$$t^\mu=-\left[(-g)^{1/2}y^AM_A\xi^\mu+\sum_{\Pi,\Theta}(-1)^{|\Pi|}\bar{\delta}y_{B,\Theta}\left((-g)^{1/2}y^A\frac{\partial M_A}{\partial y_{B,\mu\Pi\Theta}}\right)_{,\Pi}\right]. \tag{52}$$

Note that the conserved quantities have been expressed as directly calculatable relations involving the field equations and the transformations. Thus, if a set of field equations is invariant under either the Lorentz or Galilean groups, Eq. (52) will immediately give the conserved quantities, the momentum, the energy, and the angular momentum of the field.

As an example of the usefulness of this section, consider the case of a second order equation in one variable that we may write as a power series,

$$N=N(\Phi_{xx},\Phi_x,\Phi)=a_{n,m,l}\Phi^n{}_{xx}\Phi^m{}_x\Phi^l=0. \tag{53}$$

The condition that N be potential requires that $n=0$, or $n=1$, and that $la_{1,m,l}=(m+2)a_{0,l-1,m+2}$. Substitution of these conditions into Eq. (51) using the symmetry group given by translations of the origin ($\bar{\delta}\Phi=-\Phi_{,x}\xi^x$; $\xi^x=\epsilon$; $t^x=\epsilon t$, where ϵ and t are constants) yields

$$t=-(l+1)^{-1}a_{0,m,l}\Phi^m{}_x\Phi^{l+1}-(m+2)^{-1}a_{1,m,0}\Phi^{m+2}{}_x.$$

Treating this equation as a power series in Φ_x and solving for Φ_x, one has obtained a solution of Eq. (53) via a quadrature.

IV. HAMILTONIANS

The formalism we have presented, which enables one to write Lagrangians for a given set of equations, may be extended to give a treatment that is analogous to the customary Hamiltonian approach. This treatment may be developed using the potential and composite principles. We shall present the results for both variational principles here.

Consider a potential operator $N^A(y_B;y_{B;\Pi})$ that is now expressed in terms of the field variables and their covariant derivatives. As we have seen, the Lagrangian for this operator is

$$\mathcal{L}=(-g)^{1/2}y_A\int_0^1 N^A(\lambda y_B;y_{B;\Pi})d\lambda.$$

In this section the range of Π is restricted to exclude the case of no indices. We define the quantities $p^{B\Pi}$ by

$$\int p^{B\Pi}(-g)^{1/2}d^4x=\frac{\partial}{\partial y_{B;\Pi}}\int\mathcal{L}(-g)^{1/2}d^4x. \tag{54}$$

Since Lagrangians are only specified up to a complete divergence, there may be times when one may not be able to express the $y_{B;\Pi}$ in terms of the $p^{B\Pi}$. This never occurs when complete divergences are added to $\mathcal{L}$ so as to minimize its differential order. The use of an integral expression to define the $p^{B\Pi}$ carries with it the assumption that $\mathcal{L}$ has been expressed in its lowest differential order and we indicate this by writing

$$p^{B\Pi}=*\frac{\partial\mathcal{L}}{\partial y_{B;\Pi}}*,$$
$$p^{B\Pi}=*(-g)^{1/2}y_A\int_0^1\frac{\partial N^A}{\partial y_{B;\Pi}}d\lambda*. \tag{55}$$

Thus for a potential operator $N^\mu(y_\nu;y_{\nu;\sigma};y_{\nu;\sigma\rho})$ we have two "momenta"

$$p^{\nu\sigma}=*(-g)^{1/2}y_\mu\int_0^1\frac{\partial N^\mu}{\partial y_{\nu;\sigma}}d\lambda*$$

and

$$p^{\nu\sigma\rho} = *(-g)^{1/2} y_\mu \int_0^1 \frac{\partial N^\mu}{\partial y_{\nu;\sigma\rho}} d\lambda *.$$

We now define the quantity $H(y_B, p^{B\Pi})$,

$$H(y_B, p^{B\Pi}) = p^{B\Pi} y_{B;\Pi} - \mathcal{L}(y_B, y_{B;\Pi}). \tag{56}$$

From this, we form the variational principle

$$F = \int [p^{B\Pi} y_{B;\Pi} - H(y_B, p^{B\Pi})](-g)^{1/2} d^4x,$$

and integrate the first term by parts to put the derivatives on to the $p^{B\Pi}$,

$$F = \int \left\{ \sum_\Pi [(-1)^{|\Pi|} p^{B\Pi}{}_{;\Pi} y_B - H(y_B, p^{B\Pi})] \right\} (-g)^{1/2} d^4x. \tag{57}$$

Now treat the y_B and each $p^{B\Pi}$ as independent entities and calculate the Frechet derivative,

$$F^1 = \int \left\{ \sum_\Pi \left[(-1)^{|\Pi|} (\Phi^{B\Pi}{}_{;\Pi} y_B + p^{B\Pi}{}_{;\Pi} \psi_B) - \left(\frac{\partial H}{\partial y_B} \psi_B + \frac{\partial H}{\partial p^{B\Pi}} \Phi^{B\Pi} \right) \right] \right\} (-g)^{1/2} d^4x. \tag{58}$$

Integrate the first term in Eq. (58) by parts and regroup to obtain

$$F^1 = \int \left[\Phi^{B\Pi} \left(y_{B;\Pi} - \frac{\partial H}{\partial p^{B\Pi}} \right) + \psi_B \left(\sum_\Pi (-1)^{|\Pi|} p^{B\Pi}{}_{;\Pi} - \frac{\partial H}{\partial y_B} \right) \right] \times (-g)^{1/2} d^4x. \tag{59}$$

Since each of the $\Phi^{B\Pi}$ and ψ_B are independent, arbitrary functions, we have on setting $F^1 = 0$,

$$\frac{\partial H}{\partial p^{B\Pi}} = y_{B;\Pi} \tag{60}$$

and

$$\frac{\partial H}{\partial y_B} = \sum_\Pi (-1)^{|\Pi|} p^{B\Pi}{}_{;\Pi}. \tag{61}$$

Thus, for a second order operator, $N^\mu(y_\nu, y_{\nu;\sigma})$, we have

$$\frac{\partial H}{\partial p^{\nu\sigma}} = y_{\nu;\sigma},$$

$$\frac{\partial H}{\partial p^{\nu\sigma\rho}} = y_{\nu;\sigma\rho},$$

$$\frac{\partial H}{\partial y_B} = -p^{\nu\sigma}{}_{;\sigma} + p^{\nu\sigma\rho}{}_{;\sigma\rho}.$$

The formalism just presented may be easily shown to be consistent with the Lagrangian formalism. We proceed as follows: Into Eq. (61) substitute

$$\frac{\partial H}{\partial y_B} = -(-g)^{1/2} \delta^B{}_A \int_0^1 N^A(\lambda y_B, \lambda y_{B;\Pi}) d\lambda - y_A \int_0^1 \frac{\partial N^A}{\partial y_B} d\lambda \tag{62}$$

and

$$p^{B\Pi}{}_{;\Pi} = *(-g)^{1/2} \left(y_A \int_0^1 \frac{\partial N^A}{\partial y_{B;\Pi}} d\lambda \right)_{;\Pi} *, \tag{63}$$

obtained from Eqs. (56) and (55) respectively. The result is

$$-\delta^B{}_A \int_0^1 N^A(\lambda y_B, \lambda y_{B;\Pi}) d\lambda - y_A \int_0^1 \frac{\partial N^A}{\partial y_B} d\lambda = (-1)^{|\Pi|} \left(y_A \int_0^1 \frac{\partial N^A}{\partial y_{B;\Pi}} \right)_{;\Pi}. \tag{64}$$

The $* *$ notation is omitted for convenience. Transpose the second term on the left to the right and recall the symmetry condition, Eq. (22). We now have

$$-\delta^B{}_A \int_0^1 N^A(\lambda y_B, \lambda y_{B;\Pi}) d\lambda = \int N^{B|A}(\lambda y_A) d\lambda. \tag{65}$$

We may now use Eq. (15) to reduce Eq. (65) to

$$\int_0^1 \frac{d}{d\lambda} [\lambda N^B(\lambda y_A, \lambda y_{A;\Pi})] d\lambda = 0$$

or

$$N^B(y_A, y_{A;\Pi}) = 0.$$

Construction of the Hamiltonian type formalism in the composite case is similar to the procedure just described. From the Lagrangian,

$$\mathcal{L} = (-g)^{1/2} v_A N^A(y_B, y_{B;\Pi}),$$

define the quantities

$$p^{B\Pi} = *(-g)^{1/2} v_A \frac{\partial N^A}{\partial y_{B;\Pi}} *, \tag{66}$$

where as before, the $* *$ notation indicates that the Lagrangian has been expressed in its lowest differential order. Define an $H(v_A, y_B, p^{B\Pi})$,

$$H(v_A, y_B, p^{B\Pi}) = (-g)^{1/2} [p^{B\Pi} y_{B;\Pi} - v_A N^A(y_B; y_{B;\Pi})].$$

Form the variational principle,

$$F = \int (-g)^{1/2} [p^{B\Pi} y_{B;\Pi} - v_A N^A(y_B, y_{B;\Pi})] d^4x,$$

and continue exactly as in the case of a potential operator to obtain the following relations:

$$\frac{\partial H}{\partial y_B} = \sum_\Pi (-1)^{|\Pi|} p^{B\Pi}{}_{;\Pi} \tag{67}$$

$$\frac{\partial H}{\partial p^{B\Pi}} = y_{B;\Pi}, \tag{68}$$

$$\frac{\partial H}{\partial v_A} = 0. \tag{69}$$

As in the potential case, consistency with the Lagrangian formalism is easily demonstrated. Substitution of the appropriate definitions in Eq. (67) yields

$$\tilde{N}^{B|A} v_A = 0,$$

while from Eq. (69) it follows directly that

$$N(y_B, y_{B;\Pi}) = 0.$$

As an illustration of the formalism discussed for the potential case, consider the Korteweg–de Vries (KdV) equation. A Lagrangian for this equation[7] is

$$\mathcal{L} = \tfrac{1}{2}\Phi_{,x}\Phi_{,t} + \tfrac{1}{6}\Phi_{,x}{}^3 + \Phi_{,x}\Phi_{,xxx} + \tfrac{1}{2}\Phi_{,xx}{}^2.$$

As noted earlier, the Lagrangian must be expressed in its lowest differential order. Therefore, integrate the

third term by parts and use as the Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\Phi_{,x}\Phi_{,t} + \tfrac{1}{6}\Phi_{,x}{}^{3} - \tfrac{1}{2}\Phi_{,xx}{}^{2}. \tag{70}$$

From the definition, Eq. (55), we find

$$p^{t} = \tfrac{1}{2}\Phi_{,x},$$

$$p^{x} = \tfrac{1}{2}\Phi_{,t} + \tfrac{1}{2}\Phi_{,x}{}^{2},$$

$$p^{xx} = -\Phi_{,xx}.$$

These expressions are easily inverted and one obtains for H,

$$H = 2p^{x}p^{t} - \tfrac{4}{3}(p^{t})^{3} - \tfrac{1}{2}(p^{xx})^{2}.$$

For

$$\frac{\partial H}{\partial p^{B\Pi}} = y_{B;\Pi},$$

and

$$\frac{\partial H}{\partial y_{B}} = \sum_{\Pi}(-1)^{|\Pi|}(p^{B\Pi})_{;\Pi},$$

one obtains

$$2p^{x} - 4(p^{t})^{2} = \Phi_{,t},$$

$$2p^{t} = \Phi_{,x},$$

$$-p^{xx} = \Phi_{,xx},$$

and

$$-p^{t}{}_{,t} - p^{x}{}_{,x} + p^{xx}{}_{,xx} = 0,$$

respectively. The first three of these expressions repeat the definitions of $p^{B\Pi}$, while if one substitutes for the $p^{B\Pi}$ in the last, one obtains the KdV equation in a form given by AH.

V. CONCLUSION

In the article we have shown how the valuable methods of Atherton and Homsy may be extended to find Lagrangians for any nonlinear equations involving fields[8] in curved spaces. Additionally, we have shown how conservation laws and a Hamiltonian formalism may be obtained from these techniques.

We believe that the potential application of these methods, which enable Lagrangian techniques to be applied to a wide range of physical problems, is enormous. The examples used to illustrate the methods, including the new method of solution of a large number of nonlinear equations in one variable, only sample a few of the opportunities. Nothing concerning approximation schemes has been shown here. Similarly, while the two-tensor theory discovered herein may not be useful in describing nature, the increase in flexibility in one's perspective on how Lagrangian theories may be constructed cannot help but be useful. Finally, the potential benefits of a generalized Hamiltonian formalism for nonlinear equations have not yet been realized.

[1] R.W. Atherton and G.M. Homsy, Stud. Appl. Math. **54**, 31 (1975).

[2] The notation follows J.L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967).

[3] In this section only lower case Latin indices $i, j, k, \cdots$ and capital Latin indices $A, B, C, \cdots$ range over the internal group. Semicolons denote space—time covariant derivatives while colons denote complete covariant derivatives.

[4] The symmetrizer $(\mu\nu)$ stands for $\frac{1}{2}\mu\nu + \frac{1}{2}\nu\mu$.

[5] This theory has many similarities to the two-tensor theory of C.J. Isham, A. Salam, and J. Strathdee, Phys. Rev. D **3**, 867 (1971). One difference is that in their theory, Einstein tensors for both f and g are utilized in the Lagrangian. We hope to attempt to discuss the physics of this theory in a later paper.

[6] Follows presentation in J.L. Anderson, Ref. 2, p. 92, and the pages following.

[7] Ref. 1 (AH) obtains this Lagrangian in illustrating how a non-potential operator may be transformed to a potential operator by a change of variable.

[8] In an effort to keep the formalism at a minimum, we have not generalized our treatment to cover cases of Lagrangians where the group representation space is also varied [i.e., H.F. Ahner, Phys. Rev. D **13**, 250 (1976)].

Static stars: Some mathematical curiosities

C. B. Collins

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
(Received 18 August 1976)

The equations of structure of static Newtonian and general relativistic stars are investigated. By using Lie group theory, it is shown that, in each case, the condition that there should exist a simple "homologous" family of similar solutions necessitates precisely those equations of state for the stellar matter that are usually invoked by means of extraneous physical arguments. In the relativistic case, a diagram which depicts these families is drawn, using the qualitative theory of differential equations. This vividly exhibits the nature of the general soutions, and the exceptional character of the Misner-Zapolsky solution. This diagram is contrasted with similar ones obtained by Chandrasekhar in the Newtonian case.

1. INTRODUCTION

In the conventional treatment of both Newtonian and general relativistic stellar systems, it is customary to examine those special cases of interest that result upon imposing a definite equation of state. This equation of state is obtained by appeal to purely physical arguments that are to a considerable extent divorced from the gravitational aspect of the problem. For example,[1,2] by considering the limiting case in which the constituent particles have relativistic energies that render their rest masses negligible, one can obtain the equation of state of an ideal Fermi gas at zero temperature, in the form $p = \frac{1}{3}\rho$, where p is the isotropic pressure and ρ is the mass-energy density, and hence determine the structure of an ideal static neutron star.[3]

For Newtonian stars in equilibrium, it seems physically plausible to demand that a simple family of solutions should exist, the individual members being related to each other by transformations of the form $r \to \tilde{r} = ar$, $\rho \to \tilde{\rho} = b\rho$, and $M \to \tilde{M} = cM$, where M is the total mass enclosed within a radius r, and a, b, and c are constants. Chandrasekhar[4] refers to this change of scale as a "homologous transformation," and he shows that for both polytropic and isothermal gas spheres homologous families of solutions exist. In Sec. 2, we examine the consequences of the requirement that a homologous family should exist, and discover that it *necessitates* either a polytropic or an isothermal equation of state, i.e., an equation of state that is usually postulated by consideration of extraneous physical aspects of the problem.

Because the r coordinate is less well-defined in general relativity, the above scaling procedure is not necessarily justified as a means of defining homologous families of solutions. However, it seems reasonable to require that a simple family of solutions should exist, whose individual members are related by more general transformations of the form $r \to \tilde{r}(r)$, $\rho \to \tilde{\rho}(\rho)$ and $M \to \tilde{M}(M)$. In fact, the only allowable transformations do involve a simple rescaling, as we show in Sec. 2. It is deduced that in this case the equation of state is necessarily of the "γ-law" type, viz. $p = (\gamma - 1)\rho$, where γ is a constant; this is the equation of state most relevant to relativistic stellar structure.

To some extent, we could assert that in each case the mathematics "knows" in advance the physics of the situation. Whether or not this is significant remains a purely speculative matter, but it is tempting to suggest that it is a reflection of the "subliminal" role that mathematics can play in physical systems.[5,6]

In Sec. 3, we exhibit a qualitative diagram that depicts the homologous family of solutions in the relativistic case. This is achieved by observing that, when $p = (\gamma - 1)\rho$, the equations of structure form an autonomous system of two ordinary differential equations, to which the qualitative theory of ordinary differential equations can be applied. This diagram is compared with similar diagrams obtained for the homologous families of solutions in the Newtonian case, of which Chandrasekhar[4] has given an exhaustive treatment. Some comments are made on the special features of the solution of Misner and Zapolsky.[7]

2. HOMOLOGOUS FAMILIES OF SOLUTIONS

A. Lie group theory and ordinary differential equations

In differential equations, Lie group theory is usually employed to investigate systems of partial, rather than ordinary, differential equations. The reason for this appears to be that a generic system of partial differential equations always admits a nontrivial group, whereas although the same is true of ordinary differential equations, the problem of the discovery of the group is equivalent to the problem of integrating the original system.[8]

It can be shown[8] that the system of ordinary differential equations

$$\frac{du^k}{dx} = f^k(x, \mathbf{u}) \quad (k = 1, 2, \ldots, m)$$

where $\mathbf{u} = (u^1, u^2, \ldots, u^m)$, is invariant under the action of the infinitesimal generator $\mathbf{X} = \xi(x, \mathbf{u})(\partial/\partial x) + \eta^k(x, \mathbf{u})(\partial/\partial u^k)$, if and only if the following equations are satisfied:

$$\frac{\partial \eta^k}{\partial x} + f^j \frac{\partial \eta^k}{\partial u^j} - f^k \frac{\partial \xi}{\partial x} - f^k f^j \frac{\partial \xi}{\partial u^j} = \xi \frac{\partial f^k}{\partial x} + \eta^j \frac{\partial f^k}{\partial u^j} \quad (k = 1, 2, \ldots, m). \tag{2.1}$$

In the particular case where $\mathbf{X}$ generates "quasihomologous" transformations (i.e., transformations of the form $x \to \tilde{x}(x)$, $u^j \to \tilde{u}^j(u^j)$, where no sum is taken over j), this requires $\xi = \xi(x)$ and $\eta^j = \eta^j(u^j)$, with no sum over j.

As a result, Eq. (2.1) becomes

$$\frac{d\eta^k}{du^k}(u^k) - \frac{d\xi}{dx}(x) = \mathbf{X}(\ln f^k) \tag{2.2}$$

with no sum over k. The form of the left side of this equation places certain restrictions on the set of possible quasihomologous transformations.

B. The Newtonian static star

The structure equations for a Newtonian star in equilibrium are

$$\frac{dp}{dr} = -\frac{GM}{r^2}\rho \tag{2.3}$$

and

$$\frac{dM}{dr} = 4\pi r^2 \rho, \tag{2.4}$$

with the supplementary conditions

$$p = 0 \quad \text{when } \rho = 0 \tag{2.5a}$$

and

$$M = 0 \quad \text{when } r = 0. \tag{2.5b}$$

We shall assume further that $p \geq 0$. If p is independent of r, then from (2.3) either $M \equiv 0$ or $\rho \equiv 0$. In the former case it follows from (2.4) that $\rho \equiv 0$, so in either case $\rho \equiv 0$ and both p and M are constant. With the supplementary conditions (2.5), we obtain $p \equiv \rho \equiv M \equiv 0$. This corresponds to the trivial solution where there is no star. We shall henceforth assume that $\rho p' \not\equiv 0$, where a prime ($'$) denotes differentiation with respect to r.

In the general case we rewrite (2.3) and (2.4) in the form

$$\frac{dp}{dr} = -\frac{GM\rho}{r^2 p'} \tag{2.3$'$}$$

and

$$\frac{dM}{dr} = 4\pi r^2 \rho, \tag{2.4}$$

and determine the homologous transformations by applying expression (2.2) to Eqs. (2.3$'$) and (2.4). Thus, if

$$\mathbf{X} = \xi(r)\frac{\partial}{\partial r} + \eta^1(\rho)\frac{\partial}{\partial \rho} + \eta^2(M)\frac{\partial}{\partial M},$$

we find

$$\frac{d\eta^1}{d\rho}(\rho) - \frac{d\xi}{dr}(r) = \frac{-2\xi(r)}{r} + \eta^1(\rho)\frac{d}{d\rho}\left[\ln\left(\frac{\rho}{p'}\right)\right] + \frac{\eta^2(M)}{M} \tag{2.6}$$

and

$$\frac{d\eta^2}{dM}(M) - \frac{d\xi}{dr}(r) = \frac{2\xi(r)}{r} + \frac{\eta^1(\rho)}{\rho}. \tag{2.7}$$

From the functional dependence on M in Eq. (2.6), it follows that $\eta^2(M) = aM$, where a is a constant. By adding Eqs. (2.6) and (2.7) we now see that in view of the functional dependence on ρ and r, $\xi(r) = br + c$ for constants b and c. Substituting into (2.6) and (2.7), we obtain

$$c = 0,$$

$$\eta^1(\rho) = (a - 3b)\rho, \tag{2.8}$$

and

$$(a - 3b)\rho\frac{\partial}{\partial \rho}\left[\ln\left(\frac{\rho}{p'}\right)\right] = -2b.$$

If $a = 3b$, then from (2.8) $b = 0$ and hence $a = 0$. Consequently $\xi \equiv \eta^1 \equiv \eta^2 \equiv 0$, and we obtain the trivial identity transformation ($\mathbf{X} = 0$).

If $a \neq 3b$, then we can integrate (2.8) to obtain

$$p = A\rho^{2(a-2b)/(a-3b)} + B \quad (a \neq 2b, a \neq 3b) \tag{2.9}$$

and

$$p = A \ln \rho + B, \quad (a = 2b \neq 3b), \tag{2.10}$$

where A and B are constants. If $a \neq b$ we may rewrite (2.9) in the form

$$p = A\rho^{1+1/n} + B, \tag{2.11}$$

where the "polytropic index," $n = (a - 3b)/(a - b) \neq -1$. To comply with condition (2.5a) we must have $B = 0$ and $1/n > -1$ (i.e., $n > 0$ or $n < -1$). If $a = b$, Eq. (2.9) becomes

$$p = A\rho + B \tag{2.12}$$

and again $B = 0$ for the validity of condition (2.5a). Equation (2.11) with $B = 0$ is the usual form for a polytropic gas, whereas (2.12) is the equation of state for an isothermal gas.[4] It is clear that Eq. (2.12) can be considered as being obtained formally from (2.11) by letting n tend to infinity.

In the general case where $a \neq b$, $a \neq 2b$, and $a \neq 3b$, or in the special case where $a = 2b \neq 3b$ (i.e., $n = -1$), then $\mathbf{X}$ is proportional to

$$(n - 1)r\frac{\partial}{\partial r} - 2n\rho\frac{\partial}{\partial \rho} + (n - 3)M\frac{\partial}{\partial M}.$$

If $a = b \neq 3b$ (i.e., formally, $n = \infty$), then $\mathbf{X}$ is proportional to

$$r\frac{\partial}{\partial r} - 2\rho\frac{\partial}{\partial \rho} + M\frac{\partial}{\partial M}.$$

The operator $\mathbf{X}$ has two independent invariants. These are readily seen to be $Mr^{-(n-3)/(n-1)}$ and $\rho r^{2n/(n-1)}$ if $n \neq 1$, and r and M/ρ if $n = 1$ (or any two independent functions of these invariants).

In each case, the finite transformation generated by $\mathbf{X}$ is a simple rescaling: $r \to fr, \rho \to g\rho, M \to hM$, where f, g, and h are constants.

If n is finite, we put $\rho \propto \theta^2$ and $r \propto \xi$. The homology theorem of Chandrasekhar[4] states that if $\theta(\xi)$ is a solution, then so also is $C^{2/(n-1)}\theta(C\xi)$, where C is an arbitrary constant. In our notation, this reduces to the observation that $\rho^{1/n} r^{2/(n-1)}$ is invariant, or that the equations are invariant under a transformation $r \to Ar$, $\rho \to A^{-2n/(n-1)}\rho$, and $M \to A^{(n-3)/(n-1)}M$.

If n is infinite, we put $\rho \propto e^{-\psi}$ and $r \propto \xi$; the homology theorem states that if $\psi(\xi)$ is a solution, then so also is $\psi(C\xi) - 2\ln C$, where C is an arbitrary constant. It reduces to the observation that ρr^2 is invariant, or that

the equations are invariant under a transformation $r \to Ar$, $\rho \to A^{-2}\rho$, and $M \to AM$.

Some aspects of the group properties discussed above are treated by Kurth,[9] but from a different viewpoint.

C. The relativistic static star

The structure equations (we use geometrical units, in which $G = c = 1$) for a static star in general relativity are

$$\frac{dp}{dr} = -\frac{(\rho + p)(M + 4\pi r^3 p)}{r^2(1 - 2M/r)} \tag{2.13}$$

and

$$\frac{dM}{dr} = 4\pi r^2 \rho, \tag{2.14}$$

with the supplementary conditions (2.5).

If p is independent of r, then from (2.13) either $M + 4\pi r^3 p = 0$ or $p = -\rho$. In the fomer case it follows from (2.14) that $p = -\frac{1}{3}\rho$. Thus in either case the equation of state is unrealistic. If $\rho = 0$, then by (2.5) $p = 0$ and this corresponds to the trivial solution where there is no star. We shall henceforth assume that $\rho p' \neq 0$.

In the general case, we proceed as in the Newtonian case above. However the calculations involved are considerably more complicated, and we shall omit algebraic details. The expression (2.2) becomes

$$\begin{aligned}\frac{d\eta^1}{d\rho}(\rho) - \frac{d\xi}{dr}(r) &= \xi(r)\frac{\partial}{\partial r}\left[\ln\frac{1}{r^2}\frac{(M + 4\pi r^3 p)}{(1 - 2M/r)}\right] + \eta^1(\rho)\frac{\partial}{\partial \rho} \\ &\times \left[\ln\frac{(\rho + p)(M + 4\pi r^3 p)}{p'(\rho)}\right] + \eta^2(M)\frac{\partial}{\partial M}\left[\ln\frac{(M + 4\pi r^3 p)}{(1 - 2M/r)}\right]\end{aligned} \tag{2.15}$$

and

$$\frac{d\eta^2}{dM}(M) - \frac{d\xi}{dr}(r) = \frac{2\xi(r)}{r} + \frac{\eta^1(\rho)}{\rho}. \tag{2.16}$$

From the functional dependence on M in (2.16), we have $\eta^2(M) = aM + \alpha$, where a and α are constants. This implies that $\eta^1(\rho) = (a - 3b)\rho$ and $\xi(r) = br + c/r^2$, where b and c are constants. The conditions for compatibility of these relationships with Eq. (2.15) is found to be (i) $a = b = c = \alpha = 0$ or (ii) $a = b \neq 3b$, $c = \alpha = 0$. The first case corresponds to the identity transformation ($\mathbf{X} = 0$). The second case corresponds to the transformation generated by $\mathbf{X} = r(\partial/\partial r) - 2\rho(\partial/\partial\rho) + M(\partial/\partial M)$. The quantities M/r and ρr^2 (or any two independent functions of them) are homologous invariants. The equation of state is necessarily $p = (\gamma - 1)\rho$, where γ is constant. Because of the close analogy between this case and that of the isothermal gas considered in the Newtonian case in Sec. 2B above, we obtain a homology theorem for the relativistic case:

Homology Theorem: For an isothermal gas in equilibrium in general relativity, if $\rho(r)$ is a solution, then so also is $C^2\rho(Cr)$, where C is an arbitrary constant.

This is again equivalent to the observation that ρr^2 is invariant, or that the equations are invariant under a transformation $r \to Ar$, $\rho \to A^{-2}\rho$, and $M \to AM$.

It is interesting to note that Michalski[10] has discovered that the most general continuous transformation which leaves the equations of stellar structure invariant in the *nonstatic* case (i.e., explicit time dependence) is a homology transformation in which $r \to Ar$, $p \to A^{-2}p$, and $\rho \to A^{-2}\rho$, thus forcing the equation of state $p = (\gamma - 1)\rho$.

3. QUALITATIVE DESCRIPTION OF RELATIVISTIC STARS IN EQUILIBRIUM

In this section, we discuss the structure equations (2.13) and (2.14) for a static star in general relativity, in the case where a homologous family of solutions exists. As we have seen, this requires matter with an equation of state $p = (\gamma - 1)\rho$, where γ is constant.

It is convenient to express Eqs. (2.13) and (2.14) in terms of the homologous invariants $m \equiv M/r$ and $\mu \equiv 4\pi\rho r^2$.

They become

$$\frac{d\mu}{dt} = \frac{\mu}{1 - 2\mu}\left[2 - \frac{(5\gamma - 4)m}{(\gamma - 1)} - \gamma\mu\right] \tag{3.1}$$

and

$$\frac{dm}{dt} = \mu - m, \tag{3.2}$$

where $t = \ln r$. Since we are assuming throughout that $\rho p' \neq 0$, it follows that $\gamma \neq 1$ and hence that division by $(\gamma - 1)$ in (3.1) is justified.

Equations (3.1) and (3.2) form a plane autonomous system of differential equations and one can depict the nature of the solutions in a diagram in the $m - \mu$ plane (see Fig. 1; cf. Refs. 11–14). Only the physically relevant region $m > 0$, $\mu > 0$ is depicted. Those solutions for which $m > 1/2$ are not admissible, since they correspond to stars lying completely inside their Schwarzschild radii, i.e., to black holes, and condition (2.5b) cannot be satisfied. Of those solutions for which $0 < m < 1/2$, the majority are unrealistic since they cross the μ axis at finite t. This has the interpretation that $M = 0$ for some finite nonzero r_1; for realistic matter content, $\rho \geq 0$, and so equation (2.14) shows that in this case condition (2.5b) is satisfied only if $\rho \equiv 0$ for $0 \leq r \leq r_1$, which is absurd.

There are two other solutions. One is represented in Fig. 1 by the center of the spiral, at the point P where $\mu = m = 2(\gamma - 1)/[(\gamma + 2)^2 - 8]$. Thus

$$\rho = \frac{1}{2\pi r^2}\cdot\frac{\gamma - 1}{(\gamma + 2)^2 - 8} \quad \text{and} \quad M = \frac{2(\gamma - 1)r}{(\gamma + 2)^2 - 8}.$$

This is the special solution of Misner and Zapolsky[7]; it is the relativistic analogue of a special singular solution described by Chandrasekhar[4] for Newtonian polytropes with index n satisfying $3 < n \leq \infty$. The other "realistic" solution in Fig. 1 starts out at the origin (where $r = 0$) and spirals into the point P. It possesses the property that ρ is finite and nonzero at $r = 0$, and it extends out to infinite values of r, the density decreasing approximately as $1/r^2$, as in the Misner–Zapolsky solution.

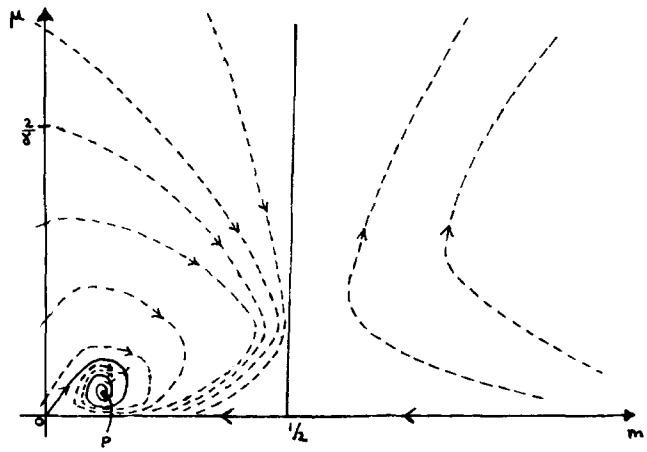


FIG. 1. Integral curves of the system (3.1) and (3.2) in the region $\mu > 0, m > 0$. The physically relevant solutions are that of Misner and Zapolsky (represented by the point P) and the solution which starts at the origin, 0, and spirals into P.

In each case the space–time metric is obtained from calculating the coefficients $A(r)$ and $B(r)$ in

$$ds^2 = -B(r)dt^2 + A(r)dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2).$$

In order to satisfy Einstein's equations with $A(0)$ finite, we must have $A(r) = [1 - 2m]^{-1} = [1 - 2M/r]^{-1}$; the function $B(r)$ must satisfy

$$B'/B = -2p'/(p + \rho).$$

Weinberg's solution[2] of this, in the case where the model is asymptotically flat [i.e., $B(\infty) = 1$], is

$$B(r) = \exp \int_r^\infty \frac{2p'}{p + \rho} dr \tag{3.3}$$

but this is valid only if the integral converges. This *will* be the case for a realistic model, where there is a region outside the star, but it is interesting to note that there are cases where (3.3) is not valid. For instance, in the solution of Misner and Zapolsky described above, $B \propto \rho^{(-2/\gamma)(\gamma-1)} \propto r^{(4/\gamma)(\gamma-1)}$; however this solution is not asymptotically flat.

ACKNOWLEDGMENT

I am grateful to the National Research Council of Canada for financial support.

[1]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[2]S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).
[3]J. R. Oppenheimer and G. M. Volkoff, Phys. Rev. **55**, 374 (1939).
[4]S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (Dover, New York, 1939).
[5]C. B. Collins, Commun. Math. Phys. **39**, 131 (1974).
[6]C. B. Collins, "The role of mathematics in gravitational physics: from the sublime to the subliminal?," Essay on Gravitation selected for Honorable mention by the Gravity Research Foundation (1976).
[7]C. W. Misner and H. S. Zapolsky, Phys. Rev. Lett. **12**, 635 (1964).
[8]L. V. Ovsjannikov, "Group Properties of Differential Equations," translated from the Russian "Gruppovye Svoystva Differentsialny Uravnenii" by G. W. Bluman. Original Russian edition published in 1962 at Novosibirsk by Siberian Section of the Academy of Science of USSR.
[9]R. Kurth, *Dimensional Analysis and Group Theory in Astrophysics* (Pergamon, Oxford, 1972).
[10]H. Michalski (private communication).
[11]C. B. Collins, Commun. Math. Phys. **23**, 137 (1971).
[12]C. B. Collins, Commun. Math. Phys. **27**, 37 (1972).
[13]I. S. Shikin, Doklady Akad. Nauk SSSR **176**, 1048 (1967) [Sov. Phys.—Dokl. **12**, 950 (1968)].
[14]I. S. Shikin, Zh. Eksp. Teor. Fiz. **63**, 1529 (1972) [Sov. Phys.—JETP **36**, 811 (1973)].

Complete integrability conditions of the Einstein–Petrov equations, type I

Carl H. Brans

Loyola University, New Orleans, Louisiana 70118
(Received 30 August 1976)

The post-Bianchi equations, defined as the integrability conditions of the Bianchi equations, are explicitly stated for the algebraically general (type I) Einstein–Petrov vacuum equations. A computer analysis of these equations has shown that they constitute a completely integrable set. Hence all conditions imposed by the Einstein equations of this type on the derivatives of the dependent variables are now known.

This paper presents the results of one phase of a computer assisted study of the problem of obtaining the exact solutions of the vacuum Einstein equations. At this point, we have developed the full set of integrability conditions for the Einstein–Petrov equations in the algebraically general, type I, case. The result is that the post-Bianchi equations, which are the integrability conditions of the Bianchi identities, are themselves completely integrable modulo the other equations. These post-Bianchi equations are given explicitly in Newman–Penrose notation in Eqs. (22) below.

Thus, all derivative conditions imposed by the Einstein–Petrov equations are now known. Some applications of this result will be explored in another paper. Several points may be mentioned here, however. First, the explicit statement of all differential conditions may be helpful in obtaining new solutions directly. At least the nature of the arbitrary functions generating the most general local analytic solutions can now be described in full. In connection with this, the explicit expression of the full integrability set provides a mechanism for a straightforward proof of the local existence theorem using a Cauchy–Kowalewski type of theorem.[1] Finally, it should be noted that these results have invariant content, since if the Petrov scalars are functionally independent, a unique canonical local coordinate system is obtained, as discussed in the Appendix. Different functional forms for metrics expressed in this standard coordinate system thus describe physically different metrics.

The general approach used in this work was described in a previous paper,[2] with a very early version of the computer techniques presented in another.[3] The equations of interest here are essentially the structure equations, which define the connection and curvature forms, together with the condition that the curvature components be of the Petrov type I. Since the Lorentz orthonormal frame in which the curvature components have this form is not integrable in general it is necessary to use a differential geometric language adapted to such nonholonomic bases (sometimes also referred to as "tetrads" or "moving frames"). The language of differential forms, developed by Cartan, is most convenient for this purpose. The book by Misner, Thorne, and Wheeler[4] contains an extensive introduction to this formalism.

Let ω^α be a Lorentz-orthonormal basis (tetrad, vierbein) for the space of 1-forms. Thus, the metric can be expressed

$$ds^2 = \eta_{\alpha\beta}\omega^\alpha\omega^\beta, \quad (\alpha, \beta, \cdots = 0, 1, 2, 3), \tag{1}$$

where $\eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$. Since each ω^α is itself a form, it can be expressed in some local coordinate system in terms of components, ω^α_i, $i = 0, 1, 2, 3$, which transform as covariant vectors with respect to the coordinate index i. In general, we will use greek indices to denote components of objects relative to a form basis such as ω^α, while latin indices will describe components relative to a coordinate basis. Thus if ρ is a 1-form,

$$\rho = \rho_\alpha\omega^\alpha = \rho_i dx^i, \tag{2}$$

and, in particular,

$$\omega^\alpha = \omega^\alpha{}_i\, dx^i. \tag{3}$$

Differentiation relative to a coordinate will be denoted by the usual comma notation while form derivatives are indicated by a vertical bar. Thus, if f is a function,

$$df = f_{,i}\, dx^i = f_{|\alpha}\omega^\alpha, \tag{4}$$

so that

$$f_{|\alpha}\omega^\alpha{}_i \equiv f_{,i}. \tag{5}$$

Since the results obtained in this paper are expressed in terms of form derivatives with respect to a basis such as ω^α, or the Newman–Penrose complex null basis, it is necessary to use Eq. (5) to translate into ordinary, coordinate derivatives. However, these coordinates must be chosen in some standard, canonical way in order to maintain the invariance of the results. Two approaches to this problem are discussed in the Appendix.

The basic geometric equations in differential form notation are the Cartan structure equations. The first of these essentially defines the connection forms, $\omega^\alpha{}_\beta$,

$$d\omega^\alpha = \omega^\beta \wedge \omega^\alpha{}_\beta, \quad \omega_{\alpha\beta} + \omega_{\beta\alpha} = 0. \tag{6}$$

The components of the connection forms, $\gamma^\alpha{}_{\beta\mu}$, are sometimes referred to as the Ricci rotation coefficients of the frame ω^α. The curvature forms, $\Omega^\alpha{}_\beta$, are given by the second structure equation,

$$d\omega^\alpha{}_\beta + \omega^\alpha{}_\gamma \wedge \omega^\gamma{}_\beta = \Omega^\alpha{}_\beta. \tag{7}$$

The usual Riemann curvature tensor components are then the components of $\Omega^\alpha{}_\beta$,

$$\Omega^\alpha{}_\beta = (\tfrac{1}{2})R^\alpha{}_{\beta\mu\nu}\omega^\mu \wedge \omega^\nu. \tag{8}$$

The condition that the curvature components be of the type I form is most easily expressed in terms of the

complex notation for this problem introduced by Petrov in his original paper.[5] Petrov set out to solve the classification problem for tensors having the algebraic properties of a Ricci-free Riemann tensor. Such tensors can be most conveniently analyzed in the SO(3, C) representation of the Lorentz group which makes use of the one-to-one correspondence between antisymmetric two-tensors and complex three-dimensional vectors. This correspondence can be established by associating the "electric" and "magnetic" parts of such a tensor with the real and imaginary parts of the complex vector. Thus, if $f_{\alpha\beta}$ is the tensor, define

$$E_a = f_{0a}, \tag{9}$$

and

$$B_a = \epsilon_{abc} f_{bc}/2, \tag{10}$$

where $a, b, \cdots = 1, 2, 3$, and ϵ_{abc} is the alternating symbol. Then $f_{\alpha\beta}$ is associated with ϕ_a where,

$$\phi_a = E_a + iB_a. \tag{11}$$

Lorentz transformations on $f_{\alpha\beta}$ are then associated with complex rotations on ϕ_a. In this notation, Ricci-free curvature tensors become complex three by three symmetric traceless matrices. The Petrov problem is then one of finding a canonical form for such a matrix. The general type I situation is the one for which this matrix has three distinct eigenvalues and can be diagonalized. Of course, because of the condition of being traceless, the sum of these eigenvalues is zero.

In order to make use of this complex notation in the structure equations from the start, let us replace the 1-form basis, ω^α, by the corresponding complex 2-form basis, σ^a, defined by

$$\sigma^a = \omega^0 \wedge \omega^a + (i/2)\epsilon^a{}_{bc}\,\omega^b \wedge \omega^c. \tag{12}$$

Further, define the complexified connection forms, X^a, by

$$X^a = (\tfrac{1}{2})\epsilon^a{}_{bc}\,\omega^{bc} - i\omega^a{}_0. \tag{13}$$

In this notation, the Einstein–Petrov type I structure equations become

$$d\sigma^a = \epsilon^a{}_{bc} X^b \wedge \sigma^c, \tag{14}$$

$$dX^a - (\tfrac{1}{2})\epsilon^a{}_{bc} X^b \wedge X^c = \alpha^a \sigma^a \quad \text{(no sum on } a\text{)}. \tag{15}$$

Here α^a are the three complex Petrov scalars, whose sum is zero. From now on, ω^α will represent this preferred basis.

These, then, are the equations whose integrability structure must be investigated. Thus if two of these equations determine different derivatives of the same function, the differential consistency condition must be checked. Since we are using form derivatives, rather than ordinary commutative differentiation, such conditions have the form

$$f_{|\alpha|\beta} - f_{|\beta|\alpha} - f_{|\mu}(\gamma^\mu{}_{\alpha\beta} - \gamma^\mu{}_{\beta\alpha}) = 0. \tag{16}$$

The integrability conditions for the set (14) and (15) can be obtained simply by applying the exterior derivative operator, d, and using the fact that $d^2 = 0$. This then leads to the Bianchi conditions,

$$d\alpha^a \wedge \sigma^a + (\alpha^a - \alpha^c)X^b \wedge \sigma^c - (\alpha^a - \alpha^b)X^c \wedge \sigma^b = 0. \tag{17}$$

Here, and throughout the remainder of the paper, a, b, c, = cyclic(1, 2, 3) and the summation convention is suspended. The use of ϵ_{abc} only complicates the notation. Using (12) these can be explicitly expanded to give

$$\alpha^a{}_{|a} = (\alpha^a - \alpha^b)X^c{}_b + (\alpha^c - \alpha^a)X^b{}_c, \tag{18}$$

$$\alpha^a{}_{|b} = i(\alpha^c - \alpha^a)X^b{}_0 + (\alpha^b - \alpha^a)X^c{}_a, \tag{19}$$

$$\alpha^a{}_{|c} = i(\alpha^b - \alpha^a)X^c{}_0 + (\alpha^a - \alpha^c)X^b{}_a, \tag{20}$$

$$\alpha^a{}_{|0} = i(\alpha^c - \alpha^a)X^b{}_b + i(\alpha^b - \alpha^a)X^c{}_c. \tag{21}$$

It is straightforward to check that these equations are consistent with the condition $\sum \alpha^a = 0$, so that there are only eight complex independent equations among them.

The differential consistency of the set (18)–(21) must now be checked, making use of (16). The computer was used to generate these conditions, the post-Bianchi equations, substituting where necessary for derivatives previously determined from the structure equations (14), (15), and the Bianchi equations themselves. Twelve equations were obtained of which only nine are independent. In full detail these equations are rather lengthy in this notation. However, they take a simpler form when expressed in the Newman–Penrose formalism,

$$\begin{aligned}3\Psi_2(&2\tau\lambda - 2\rho\nu - 3\nu\epsilon + \nu\bar{\rho} - \nu\bar{\epsilon} + 3\lambda\beta + \lambda\bar{\pi} - \lambda\bar{\alpha} + \delta\lambda - D\nu)\\ &+\Psi_0(-\tau\epsilon - \tau\bar{\rho} + \tau\bar{\epsilon} - 3\sigma\pi - 8\sigma\alpha + \rho\beta - \rho\bar{\pi} + \rho\bar{\alpha}\\ &+3\mu\kappa + 8\kappa\gamma - \delta\rho + D\tau) = 0,\end{aligned} \tag{22a}$$

$$\begin{aligned}3\Psi_2(&-\Delta\lambda + \tau\kappa - 2\sigma\rho - 5\sigma\epsilon - \sigma\bar{\epsilon} - \pi\nu - \nu\bar{\tau} + 2\mu\lambda + 5\lambda\gamma + \lambda\bar{\gamma}\\ &+\kappa\bar{\pi} - D\sigma) + \Psi_0(\Delta\rho + 4\tau\pi + 8\tau\alpha + \tau\bar{\tau} - \rho\gamma - \rho\bar{\gamma} - \pi\bar{\pi} - 4\nu\kappa\\ &+\mu\epsilon + \mu\bar{\epsilon} - 8\gamma\epsilon - 8\gamma\bar{\epsilon} + 8\alpha\bar{\pi} + D\mu - 8D\gamma) + 4\Psi_2\Psi_0 = 0,\end{aligned} \tag{22b}$$

$$\begin{aligned}3\Psi_2(&-\sigma\bar{\kappa} - \rho\kappa + \pi\lambda - \nu\bar{\sigma} + 5\lambda\alpha - 5\kappa\epsilon + \kappa\bar{\epsilon} + \lambda\bar{\beta} - \delta\lambda - D\kappa)\\ &+\Psi_0(\tau\bar{\sigma} + 4\rho\pi + 7\rho\bar{\alpha} - \rho\bar{\beta} + \pi\epsilon - \pi\bar{\epsilon}\\ &+\mu\bar{\kappa} - 4\lambda\kappa - 8\gamma\bar{\kappa} - 8\epsilon\alpha + 8\alpha\bar{\epsilon} + \delta\bar{\rho} + D\pi - 8D\alpha) = 0,\end{aligned} \tag{22c}$$

$$\begin{aligned}3\Psi_2(&-\Delta\nu - \tau\sigma - 5\sigma\beta - \sigma\bar{\alpha} + \nu\mu + 5\nu\gamma - \nu\bar{\gamma} + \lambda\bar{\nu} + \kappa\bar{\lambda} - \delta\sigma)\\ &+\Psi_0(\Delta\tau - 8\Delta\beta - 4\tau\mu - \tau\gamma + \tau\bar{\gamma} + 4\sigma\nu - \rho\bar{\nu} - \pi\bar{\lambda} - 7\mu\beta\\ &+\mu\bar{\alpha} + 8\gamma\beta + 8\epsilon\bar{\nu} + \delta\mu - 8\beta\bar{\gamma}) = 0,\end{aligned} \tag{22d}$$

$$\begin{aligned}3\Psi_2(&\Delta\lambda + 3\tau\kappa - 2\sigma\rho - 3\sigma\epsilon + \sigma\bar{\epsilon} - 3\pi\nu - 8\nu\alpha + \nu\bar{\tau} + 2\mu\lambda\\ &+3\lambda\gamma - \lambda\gamma + 8\kappa\beta - \kappa\bar{\pi} + D\sigma) + \Psi_0(-\Delta\rho - \tau\bar{\tau} + 4\sigma\lambda - 4\rho\mu\\ &-7\rho\gamma + \rho\bar{\gamma} + \pi\bar{\pi} - \mu\epsilon - \mu\bar{\epsilon} + 8\gamma\bar{\rho} + 8\beta\alpha - 8\alpha\bar{\alpha} - 8\delta\alpha - D\mu)\\ &+4\Psi_2\Psi_0 = 0,\end{aligned} \tag{22e}$$

$$\begin{aligned}3\Psi_2(&\Delta\rho + D\mu + \tau\bar{\tau} - \rho\gamma - \rho\bar{\gamma} - \pi\bar{\pi} + \mu\epsilon + \mu\bar{\epsilon})\\ &+\Psi_0(-\Delta\lambda - D\sigma + \tau\kappa + 2\sigma\rho + 3\sigma\epsilon - \sigma\bar{\epsilon} - \pi\nu - \nu\bar{\tau}\\ &-2\mu\lambda - 3\lambda\gamma + \lambda\bar{\gamma} + \kappa\bar{\pi}) = 0,\end{aligned} \tag{22f}$$

$$\begin{aligned}3\Psi_2(&\tau\epsilon + \tau\bar{\rho} - \tau\bar{\epsilon} - \sigma\pi - \rho\beta + \rho\bar{\pi} - \rho\bar{\alpha} + \mu\kappa + \delta\rho - D\nu)\\ &+\Psi_0(2\tau\lambda - 2\rho\nu - 5\nu\epsilon - \nu\bar{\rho} + \nu\bar{\epsilon} + 5\lambda\beta\\ &-\lambda\bar{\pi} + \lambda\bar{\alpha} - \delta\lambda + D\nu) = 0,\end{aligned} \tag{22g}$$

$$3\Psi_2(\tau\bar{\sigma}-\rho\alpha-\rho\bar{\beta}+\pi\epsilon-\pi\bar{\epsilon}+\mu\bar{\kappa}+\delta\rho+D\pi)+\Psi_0(-\sigma\bar{\kappa}$$
$$+3\rho\kappa-3\pi\lambda-\nu\bar{\sigma}-3\lambda\alpha+\lambda\bar{\beta}+3\kappa\epsilon+\kappa\bar{\epsilon}-\bar{\delta}\lambda-D\kappa)=0, \tag{22h}$$

$$3\Psi_2(\Delta\tau-\tau\gamma+\tau\bar{\gamma}-\rho\bar{\nu}-\pi\bar{\lambda}+\mu\beta+\mu\bar{\alpha}+\delta\mu)$$
$$+\Psi_0(-\Delta\nu+3\tau\sigma+3\sigma\beta-\sigma\bar{\alpha}-3\nu\mu-3\mu\gamma$$
$$+\lambda\nu+\kappa\bar{\lambda}-\mu\bar{\gamma}-\delta\sigma)=0. \tag{22i}$$

In order to translate from the SO(3, C) notation to Newman–Penrose (N–P), rewrite the metric (1) in terms of null form ρ^A, $A=1,\ldots,4$,

$$ds^2=2\rho^1\rho^2+2\rho^3\rho^4, \tag{23}$$

where

$$\rho^1=(\omega^3-\omega^0)/\sqrt{2}, \tag{24}$$

$$\rho^2=(\omega^3+\omega^0)/\sqrt{2}, \tag{25}$$

$$\rho^3=(\omega^2+i\omega^1)/\sqrt{2}, \tag{26}$$

$$\rho^4=\rho^{-3}. \tag{27}$$

Next, define Z^a, the complex connection forms adapted to the ρ^A basis,

$$Z^1=(X^2-iX^1)/2, \tag{28}$$

$$Z^2=(X^2+iX^1)/2, \tag{29}$$

$$Z^3=iX^3/2. \tag{30}$$

It is important to note that the components $Z^a{}_A$ are relative to the basis ρ^A while the $X^a{}_\alpha$ are relative to the ω^α. Thus, for example (28) implies that

$$Z^3{}_1=i(X^3{}_3-X^3{}_0)/2\sqrt{2}\,. \tag{31}$$

The N–P formalism uses a basis closely related to the ρ^A. The N–P derivative operators $(D,\Delta,\delta,\bar{\delta})$ are simply form derivatives with respect to $(-\rho^1,\rho^2,\rho^3,\rho^4)$. The N–P spin coefficients are related to the $Z^a{}_A$ by

$$Z^1{}_A=(\pi,-\nu,-\mu,-\lambda), \tag{32}$$

$$Z^2{}_A=(\kappa,-\tau,-\sigma,-\rho), \tag{33}$$

$$Z^3{}_A=(\epsilon,-\gamma,-\beta,-\alpha). \tag{34}$$

Finally, the independent Petrov scalars, α^1 and α^2 are related to Ψ_0 and Ψ_2 by

$$\Psi_0=(\alpha^1-\alpha^2)/2, \tag{35}$$

$$\Psi_2=(\alpha^1+\alpha^2)/2. \tag{36}$$

These were used to derive Eqs. (22) from the corresponding SO(3, C) equations.

The next step is to determine the integrability conditions generated by the post-Bianchi conditions. For example, in the X^a notation, three different derivatives of $X^1{}_0$ are determined. The consistency of these determinations with (14) and (15) must be guaranteed, etc. The computation of these equations in detail proved to be enormously complicated and was accomplished entirely by computer techniques. The surprising result was that, *in spite of the intermediate complexity, when all substitutions for previously determined derivatives were made, the integrability conditions of Eqs. (22) turned out to be trivial, i.e., the structure equations, plus Bianchi conditions, plus the post-Bianchi equations (22) are a completely integrable set.* The Einstein–Petrov equations impose no further restrictions on the derivatives of dependent variables.

That such a simple result should emerge from such enormously complicated intermediate stages probably indicates that it could have been anticipated without explicit computation. However, it is not obvious how this can be done. For example, it is true that the post-Bianchi equations are themselves the integrability conditions developed from a preceding set. This however, does not imply that they should be completely integrable since the Bianchi conditions, themselves the integrability conditions of the structure equations, are not completely integrable. This is so because supplementary conditions were placed on the components of the curvature forms. Otherwise, of course, the Bianchi conditions for an arbitrary curvature tensor are completely integrable. It might be expected that the restriction on the form of the curvature components would carry over to place nontrivial integrability conditions on the post-Bianchi equations. Nevertheless, this did not turn out to be the case. Perhaps an understanding of this problem might provide further insight into the structure of the type I solutions.

APPENDIX

The Einstein–Petrov differential equations discussed in this paper have been described in terms of form, or N–P, derivatives, rather than the usual partial derivatives of field functions. However, actual solutions must be described in terms of the functional form of the field variables relative to some local coordinate system. Thus, in practice, it is necessary to introduce a coordinate system, and to define the $\omega^\alpha{}_i$ used in (5). Here we will discuss two methods of defining such coordinate systems in a canonical way so as not to disturb the invariance of the results, showing in each case how the $\omega^\alpha{}_i$ are determined by the equations. First, if the four real Petrov scalars are functionally independent, they can obviously serve as coordinates. In terms of the complex α^a these coordinates, x^i, $i=0,1,2,3$, could be defined by

$$\alpha^1=x^0+ix^1, \tag{A1}$$

$$\alpha^2=x^2+ix^3, \tag{A2}$$

assuming of course that the real part of α^1 is a timelike function. (Recall that $\alpha^3=-\alpha^1-\alpha^2$.) If not, obvious changes could be made. Taking the exterior derivatives of (A1) and (A2) and using (4) we find that we can obtain all of the form derivatives, $x^i{}_{|\beta}$, of these canonical coordinates in terms of the $\alpha^a{}_{|\beta}$. However, these latter quantities are determined by the Bianchi conditions, (18)–(21). Finally, the $\omega^\alpha{}_i$ can be computed as the inverse of $x^i{}_{|\beta}$ since

$$\omega^\alpha=\omega^\alpha{}_i\,dx^i=\omega^\alpha{}_i\,x^i{}_{|\beta}\omega^\beta. \tag{A3}$$

On the other hand, if the Petrov functions are not fully independent a canonical coordinate system can still be defined in terms of the uniquely defined $\omega^\alpha{}_i$, up to an arbitrary choice of origin as shown by the following the-

orem.

Theorem: Given four independent forms ω^α defined over a neighborhood U of a point P, there exists a unique coordinate system, x^i, defined over a neighborhood V of P with $V \subset U$ and $x^i(P) = 0$ and for which

$$\omega^\alpha{}_0 = \delta^\alpha{}_0,$$
$$\omega^\alpha{}_1 = \delta^\alpha{}_1, \quad \text{for} \quad x^0 = 0,$$
$$\omega^\alpha{}_2 = \delta^\alpha{}_2, \quad \text{for} \quad x^0 = x^1 = 0,$$
$$\omega^\alpha{}_3 = \delta^\alpha{}_3, \quad \text{for} \quad x^0 = x^1 = x^2 = 0.$$

The proof of this theorem is given in detail in another paper.[6] Intuitively, the idea is as follows. Let e_α be the tangent vector basis dual to ω^α. Starting from P, the x^3 line is defined as the unique curve through P tangent to e^3 with proper distance providing the coordinate scale. This gives the $x^0 = x^1 = x^2 = 0$ line. Next, starting from points with so-defined x^3 values along this line, proceed out along the unique curves tangent to e^2, with proper distance providing the x^2 coordinate scale. This defines the x^1, x^2 coordinates over the surface $x^0 = x^1 = 0$. Continuing in this manner, the coordinate patch having the required properties can be constructed.

In this case the values of $\omega^\alpha{}_i$ are obtained by integrating the structure equations (6) using the theorem's results as initial values. Thus, $\omega^\alpha{}_0$ are fully determined, while $\omega^\alpha{}_1$ can be found by integrating (6) with respect to x^0, using the fact that $\omega^\alpha{}_1 = \delta^\alpha{}_1$ at $x^0 = 0$. Similar results hold for the other components.

*Work supported in part by a grant from the Research Corporation.

[1]I am grateful to the referee for pointing this out.

[2]Carl H. Brans, J. Math. Phys. **6**, 94 (1965).
[3]Carl H. Brans, J. Assoc. Comput. Mach. **14**, 45 (1967).
[4]C. W. Misner, K. S. Thorne, and J. S. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[5]A. Petrov, Sci. Not. Kazan State Univ. **114**, 55 (1954); A. Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).
[6]Carl H. Brans, J. Math. Phys. **11**, 1210 (1970).

Orbits of the rotation group on the density matrices of spin-1 particles*

M. Daumens† and M. Perroud

Centre de Recherches Mathématiques, Université de Montréal, Montréal H3C 3J7, Canada
(Received 30 December 1976)

The SO(3) orbits of the spin-one mixed states, contained in each SU(3) orbit are shown to be characterized by the *squares* (defined by means of a symmetric bilinear form) of the eigenvectors of the density matrices. The orbits of matrices diagonalizable by a rotation in a spherical basis are deduced, and quadrupole matrices are considered as a special case.

Recently, Bacry[1] has given a graphical representation of the pure states of spin j by putting $2j$ points on the sphere S^2, and from it, he has deduced the stability groups of the states. In particular he has determined the orbits and the strata of pure states of spin one. Our purpose is to find the orbits and the strata of the mixed states of spin one which are described by a density matrix belonging to the polarization domain D_1.[2]

For this we first characterize the orbits of pure states by the *squares* of the states defined by means of a symmetric bilinear form and we give the connection with Bacry's representation. Then we consider the action of the SU(3) group on D_1 and we label the SU(3) orbits of D_1 by the eigenvalues of the density matrices. Finally we show that the SO(3) orbits of D_1, contained in a given SU(3) orbit of D_1 can be characterized by the squares of the eigenvectors, and that, for simple (nondegenerate) eigenvalues, the domain of these squares is a tetrahedron.

As an application, we give the conditions which a density matrix must satisfy to be diagonalizable by a rotation in a given spherical basis. We deduce thus, the SO(3) orbits of such matrices and pay special attention to the particular case of quadrupole matrices. We also show that each SU(3) orbit of D_1 contains only one SO(3) orbit of quadrupole matrices.

1. SO(3) ORBITS OF PURE STATES

Let us consider the antiunitary operator A induced by the time reversal operation,[3] on the pure spin-one states, described by rays in the Hilbert space of dimension three, H_3. Its action on the spherical basis $|m\rangle$ $(m=+1,0,-1)$ of H_3 is defined by

$$A|m\rangle=(-1)^{1+m}|-m\rangle\,, \tag{1}$$

and, on an arbitrary state $|\phi\rangle=\sum_m \phi^m|m\rangle$ by

$$A\phi=\sum_m \phi^{m*}A|m\rangle=\sum_m (-1)^{1+m}\phi^{m*}|-m\rangle\,. \tag{2}$$

This operator A is an involution and Hermitian $(A=A^{-1}=A^{\dagger})$. Note that in a given spherical basis, it can be expressed as the product of the complex conjugation K times the Wigner matrix

$$A=K\Gamma=\Gamma K \quad \text{with} \quad \Gamma=D^1(0,\pi,0)\,. \tag{3}$$

The Wigner matrices $D^1(R)$ are unitary in a spherical basis but orthogonal (i.e., real) in a Cartesian one.

This property can be expressed in terms of the operator A in the following way.

All the matrices $D^1(R)$ commute with A and conversely each SU(3) matrix commuting with A is a matrix $D^1(R)$

$$\forall\, U\in \mathrm{SU}(3),\quad [A,U]=0 \Leftrightarrow \exists\, R\in \mathrm{SO}(3),\quad U=D^1(R)\,. \tag{4}$$

The Hermitian product of H_3 and the antiunitary operator A allow us to define the symmetric bilinear form

$$\varphi,\psi\in H,\quad (\varphi,\psi)=(\psi,\varphi)=\langle A\varphi,\psi\rangle\,. \tag{5}$$

From the preceding property of operator A, it follows that this form is also rotation invariant. Conversely, each SU(3) matrix keeping this form invariant is a Wigner matrix $D^1(R)$

$$\forall\, U\in \mathrm{SU}(3),\quad (U\varphi,U\psi)=(\varphi,\psi)\Leftrightarrow \exists\, R\in \mathrm{SO}(3),\quad U=D^1(R)\,. \tag{6}$$

This last property, essential for the following, is only valid in the spin-one case where the representation $D^1(R)$ has the same dimension as the element R of the rotation group.

Let us define the *square* of a vector of H_3 as the modulus of the bilinear form

$$\text{square of } \varphi \equiv |\varphi| = |\langle A\varphi,\varphi\rangle|\,, \tag{7}$$

with the property

$$\langle\varphi|\varphi\rangle=1 \Rightarrow 0\leq|(\varphi,\varphi)|\leq 1\,. \tag{8}$$

For example, the squares of the elements of the spherical basis $|m\rangle$ are

$$\text{square of } |m\rangle \equiv |m|,\quad |+|=|-|=0,\ \text{and}\ |0|=1\,. \tag{9}$$

In fact, the states are rays in H_3, i.e., unit elements up to an arbitrary phase. This arbitrariness could be eliminated by requiring that the elements of H_3 describing pure states have a real, positive bilinear form (φ,φ). This additional condition fixes the arbitrary phase.

Rotational invariance of the bilinear form and the arbitrariness of the phase imply that two rays in H_3 are on the same SO(3) orbit O iff they have the same square

$$\varphi,\psi\in O \Leftrightarrow |\varphi|=|\psi|\,. \tag{10}$$

In other words, the squares of the rays characterize the SO(3) orbits of rays in H_3,

$$\varphi\in H_3,\quad O(\varphi)=\{\psi\in H_3 : |\varphi|=|\psi|\}\,. \tag{11}$$

In the spherical basis $|m\rangle$ of H_3, we can choose as a representative of the orbit corresponding to the square $x \in [0, 1]$ the state $|x\rangle$ defined by

$$|x\rangle = \sqrt{1-x}\,|+1\rangle + \sqrt{x}\,|0\rangle . \tag{12}$$

The stabilizers of these states are easily determined. They are, up to an SO(3) conjugacy,

(i) $x = 0$: $\mathrm{SO}(2) = \{D^1(\alpha, 0, 0)\}$ with $\alpha \in [0, 2\pi]$,

(ii) $0 < x < 1$: $Z_2 = \{I, D^1(\pi, \theta_x, 0)\}$

$$\text{with } \cos\theta_x = (1-3x)/(1+x), \tag{13}$$

(iii) $x = 1$: $\mathrm{S}[\mathrm{O}(2)\times Z_2] = \{D^1(\alpha, 0, 0), D^1(0, \pi, 0)\}$

$$\text{with } \alpha \in [0, 2\pi] .$$

$$\simeq \mathrm{O}(2) .$$

In the Bacry representation, each state is represented by two points, not necessarily distinct, on the unit sphere S^2 embedded in $\mathbf{R}^3$. If we choose the same frame in $\mathbf{R}^3$ as is used to define the spherical basis, then the representative states $|x\rangle$ are represented by the north pole of the sphere and the point with spherical coordinates $(\theta_x, 0)$ shown in Fig. 1. To characterize the orbits, Bacry uses θ_x, readily related to x by the above formula (13).

The space of pure states then splits into three strata (unions of orbits with the same stabilizer up to conjugacy): $\{\mathrm{SO}(2)\}$, $\{Z_2\}$, and $\{\mathrm{O}(2)\}$ corresponding to the three previous stabilizers. The strata $\{\mathrm{SO}(2)\}$ and $\{\mathrm{O}(2)\}$ each consist of just one two-dimensional orbit with $x = 0$, $\theta_0 = 0$, and $x = 1$, $\theta_1 = \pi$. The stratum $\{Z_2\}$ is a continuous set of three-dimensional orbits characterized by $0 < x < 1$ or $0 < \theta_x < \pi$.

Now let us consider some properties of the bilinear form which will be used below. First, if two vectors of H_3 are orthonormal, then they satisfy the inequality

$$|(\varphi, \psi)|^2 \geq (1 - |\varphi|)(1 - |\psi|) . \tag{14}$$

Actually let φ and ψ be two orthonormal vectors. If $|\psi| = x$, there exists a rotation R such that

$$D^1(R)\psi = \sqrt{x}\,|0\rangle + \sqrt{1-x}\,|+\rangle .$$

In writing $D^1(R)\varphi$ under the form

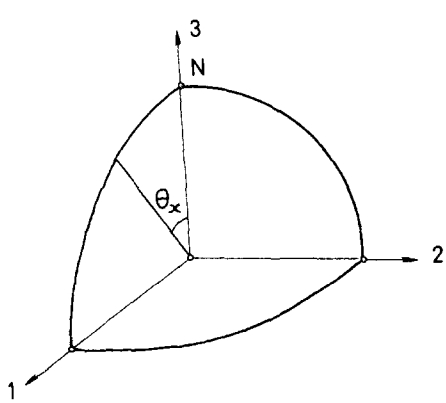


FIG. 1. Bacry representation of the representative states $|x\rangle = \sqrt{1-x}\,|+\rangle + \sqrt{x}\,|0\rangle$ with $\cos\theta_x = (1-3x)/(1+x)$.

$$D^1(R)\varphi = e^{-i\lambda}(y_-|-\rangle + y_0|0\rangle + y_+|+\rangle), \quad y_0 \in \mathbf{R},$$

we obtain

$$(\varphi, \psi) = e^{-2i\lambda}(e^{i\lambda}D^1(R)\varphi, e^{i\lambda}D^1(R)\psi)$$

$$= e^{-i\lambda}(-\sqrt{x}\,y_0 + \sqrt{1-x}\,y_-) .$$

But

$$\langle\varphi, \psi\rangle = \langle D^1(R)\varphi, D^1(R)\psi\rangle = e^{i\lambda}(\sqrt{x}\,y_0 + \sqrt{1-x}\,y_+^*) = 0 ,$$

then

$$|(\varphi, \psi)|^2 = (1-x)|y_+^* + y_-|^2 = (1-x)[1 - y_0^2 + 2\,\mathrm{Re}(y_+ y_-)] .$$

Now

$$e^{2i\lambda}(\varphi, \varphi) = (e^{i\lambda}D^1(R)\varphi, e^{i\lambda}D^1(R)\varphi) = y_0^2 + 2y_+y_- ,$$

hence

$$|(\varphi, \psi)|^2 = (1-x)[1 + \mathrm{Re}(e^{2i\lambda}(\varphi, \varphi))]$$

$$\geq (1-x)[1 - |e^{2i\lambda}(\varphi, \varphi)|] ,$$

from which follows the inequality (14). If three vectors φ_i $(i = 1, 2, 3)$ form an orthonormal set, the bilinear forms satisfy

$$\sum_{i=1}^{3} |(\varphi_i, \varphi_j)|^2 = 1, \quad \forall\, j = 1, 2, 3 , \tag{15}$$

which implies

$$2|(\varphi_i, \varphi_j)|^2 = 1 + |\varphi_k|^2 - |\varphi_i|^2 - |\varphi_j|^2 ,$$

$$(i, j, k) = \mathrm{perm}(1, 2, 3) . \tag{16}$$

Combining relations (14) and (16) we obtain four inequalities for the squares of the vectors of an orthonormal basis

$$\sum_{i=1}^{3} |\varphi_i| \geq 1, \quad \sum_{i=1}^{3} |\varphi_i| \leq 1 + 2|\varphi_j|, \qquad \forall j = 1, 2, 3 . \tag{17}$$

If we consider two orthonormal sets $\{\varphi_i\}$, $\{\psi_j\}$, Eq. (16) implies the equivalence

$$|\psi_i| = |\varphi_i|, \quad i = 1, 2, 3$$

$$\Leftrightarrow |(\psi_i, \psi_j)| = |(\varphi_i, \varphi_j)| \quad \forall\, i, j = 1, 2, 3 , \tag{18}$$

which allows us to show that, if the squares of the elements of two orthogonal bases are respectively equal, these bases are related by a rotation, up to multiplications by phase factors. Thus

$$|\psi_i| = |\varphi_i|, \quad i = 1, 2, 3$$

$$\Leftrightarrow \exists\, R \in \mathrm{SO}(3): \ \psi_i = e^{i\alpha_i}D^1(R)\varphi_i, \qquad i = 1, 2, 3 , \tag{19}$$

where the phases α_i are defined by $(\psi_i, \psi_j) = \exp[i(\alpha_i + \alpha_j)]\,(\varphi_i, \varphi_j)$.

2. SU(3) ORBITS OF MIXED STATES

Mixed states are described by 3×3 density matrices belonging to the polarization domain D_1 of the 3×3 Hermitian, positive, and unit trace matrices. The density matrices can be diagonalized by an SU(3) transformation. The SU(3) orbits of D_1 only depend on the eigenvalues of the matrices and can be parametrized by two eigenvalues λ_1 and λ_2 such that

$$0 \leq \lambda_1 \leq \lambda_2, \quad \lambda_1 + 2\lambda_2 \leq 1 . \tag{20}$$

The corresponding domain is the triangle $[NPF]$ represented in Fig. 2. These orbits can be gathered into three strata corresponding to different multiplicities of the eigenvalues[4]:

(i) *A triple eigenvalue* $\lambda = \frac{1}{3}$

The nonpolarized state $\rho_0 = \frac{1}{3} I$ is the fixed point with the entire SU(3) group as its stabilizer. In itself it forms an orbit and a stratum.

(ii) *A double eigenvalue*

This stratum with stabilizer $S[U(2) \times U(1)] \simeq U(2)$ is the union of two disjoint continuous subsets defined by

$$0 \leq \lambda_1 = \lambda_2 < \tfrac{1}{3} \quad \text{and} \quad \tfrac{1}{3} < \lambda_2 = \frac{1-\lambda_1}{2} \leq \frac{1}{2} \tag{21}$$

represented by the segments $(NP]$ and $(NF]$ respectively in Fig. 2. All the orbits of this stratum have dimension four and are in a 1-1 correspondence with the space of pure states. Note also that each subset contains an exceptional orbit (i.e., an orbit the elements of which have rank less than three) corresponding to the points P and F in Fig. 2,

P: $\lambda_1 = \lambda_2 = 0$, $\lambda_3 = 1$,

orbit of the pure states (rank = 1),

F: $\lambda_1 = 0$, $\lambda_2 = \lambda_3 = \frac{1}{2}$,

orbit of rank-2 matrices belonging to the boundary of D_1.

(iii) *Three single eigenvalues*

The stabilizer $U(1) \times U(1)$ of this stratum is given by the Cartan algebra of SU(3). In Fig. 2 this stratum is represented by the triangle $[NPF]$ without the sides NP and NF. Each orbit has dimension six. This two-dimensional stratum contains a one-dimensional subset of exceptional orbits of rank-two matrices belonging to the boundary of D_1 and represented by the open segment (PF) in Fig. 2.

3. SO(3) ORBITS OF MIXED STATES

Since the rotation group SO(3) is a subgroup of SU(3), the SO(3) orbits are contained in the SU(3) ones. We consider a given SU(3) orbit and we wish to determine the SO(3) orbits contained in it and their stabilizers. All the matrices in an SU(3) orbit have the same eigenvalues characterizing this orbit. The contained SO(3) orbits hence only depend on the eigenvectors and more specifically on their squares.

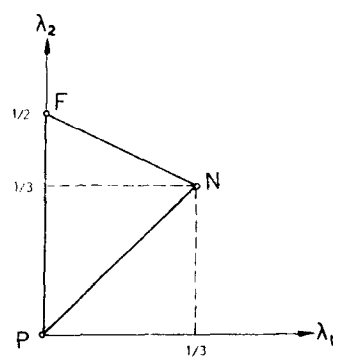


FIG. 2. SU(3) orbits and strata of the polarization domain D_1. Each orbit is a point (λ_1, λ_2) belonging to the triangle (NPF) defined by $0 \leq \lambda_1 \leq \lambda_2$ and $\lambda_1 + 2\lambda_2 \leq 1$. The three strata are $\{N\}$, $(NP] \cup (NF]$, and $\Delta[NFP] \setminus [NP] \cup [NF]$.

Let us study the different types of orbits belonging to the three SU(3) strata of D_1 defined in the previous section:

(i) *A triple eigenvalue*

The nonpolarized state is obviously also the fixed point for the rotation group. It is in itself an orbit and a stratum with stabilizer SO(3).

(ii) *A double eigenvalue*

In this case, the density matrix can be expressed as

$$\rho = \lambda_d I + (\lambda_s - \lambda_d) |s\rangle \langle s| , \tag{22}$$

where λ_d and λ_s are respectively the double and the simple eigenvalues, and $|s\rangle$ the eigenvector associated with λ_s, is the only unknown. We are thus in the same situation as for pure states, we can characterize the SO(3) orbits by the square $|s|$ of $|s\rangle$, and the stabilizer of ρ is the stabilizer of $|s\rangle$.

Thus SU(3) stratum can thus be split into SO(3) orbits of dimension two or three, characterized by the simple eigenvalue λ_s and the square $|s|$ of the associated eigenvector,

$$0 \leq \lambda_s < \tfrac{1}{3} \ (\lambda_2 = \lambda_3) \quad \text{or} \quad \tfrac{1}{3} < \lambda_s \leq 1 \ (\lambda_1 = \lambda_2) \quad \text{and} \quad 0 \leq |s| \leq 1 . \tag{23}$$

The stabilizer of each pure state $|s\rangle$ has already been given in Sec. 1.

(iii) *Three simple eigenvalues*

In this most general case the eigenvalues are labeled unambiguously by $\lambda_1 < \lambda_2 < \lambda_3$ and each density matrix can be written as

$$\rho = \sum_{i=1}^{3} \lambda_i |\varphi_i\rangle \langle \varphi_i| \quad \text{with} \sum_{i=1}^{3} \lambda_i = 1 , \tag{24}$$

where the eigenvectors form a complete orthonormal set.

Making use of the uniqueness (up to arbitrary phases) of expansion (24) and property (19) of orthonormal sets, one easily shows that two matrices with the same eigenvalues (which must be simple) are on the same SO(3) oribt iff the associated eigenvectors have the same squares, respectively. The four inequalities (17) hold for the orthonormal set of eigenvectors. In the three-dimensional space $|\varphi_i|$, $i = 1, 2, 3$, they define a regular tetrahedron with vertices $V_1(1, 0, 0)$, $V_2(0, 1, 0)$, $V_3(0, 0, 1)$, and $V(1, 1, 1)$ plotted in Fig. 3.

Hence any SU(3) orbit of D_1 characterized by three simple eigenvalues splits into a three-dimensional set of SO(3) orbits represented by this tetrahedron. The stabilizer of a matrix is the intersection of the stabilizers of these eigenvectors but in fact, because of the completeness of the eigenvectors, it is sufficient to take the intersection of only two of them. For the different points of the tetrahedron, the stabilizers are $[(i, j, k) = \text{perm } (1, 2, 3)]$:

(i) $S[O(2) \times Z_2] \simeq O(2)$ for $|\varphi_1| = |\varphi_2| = |\varphi_3| = 1$ (vertex V in Fig. 3),

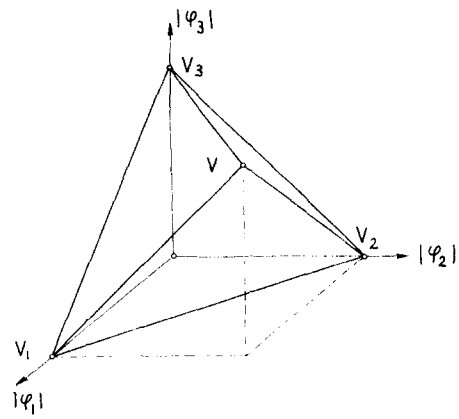


FIG. 3. SO(3) orbits of $\mathcal{D}_1$ included in a SU(3) orbit with given $\lambda_1<\lambda_2<\lambda_3$. Each orbit is a point $(|\varphi_1|,|\varphi_2|,|\varphi_3|)$ in the tetrahedron $(VV_1V_2V_3)$ defined by $\sum_{i=1}^{3}|\varphi_i|\geq 1$, $\sum_{i=1}^{3}|\varphi_i|\leq 1+2|\varphi_j|$, $j=1,2,3$. The vertices $V_1V_2V_3$ correspond to matrices diagonalizable by rotation and the vertex V to quadrupole matrices.

(ii) SO(2) for $|\varphi_i|=1$, $|\varphi_j|=|\varphi_k|=0$ (vertices V_i in Fig. 3), $i=1,2,3$,

(iii) Z_2 for $|\varphi_i|=1$, $0<|\varphi_j|=|\varphi_k|<1$ (open segments VV_i in Fig. 3), $i=1,2,3$,

(iv) I for $\sum_i|\varphi_i|\geq 1$, $\sum_i|\varphi_i|<1+2|\varphi_j|$, $j=1,2,3$ (the tetrahedron without the segments VV_i in Fig. 3).

The polarization domain $\mathcal{D}_1$ can thus be split into five strata according to the different stabilizers. These strata are listed in Table I. For each stratum, we give the dimension of the orbits, the subsets of the stratum and their dimension.

This characterization of the SO(3) orbits of $\mathcal{D}_1$ gives us a simple answer to the following question: Which are the density matrices diagonalizable by a rotation in a spherical basis? Such matrices belong to the same orbit as a diagonal matrix (in a given spherical basis). The orthonormal set of eigenvectors must satisfy

$$|\varphi_i|=1,\quad |\varphi_j|=|\varphi_k|=0,\quad (i,j,k)=\text{perm}\,(1,2,3)\,. \tag{25}$$

If the three eigenvalues are different, condition (ii) defines three two-dimensional sets of orbits represented by the triangle (NPF) without NP, NF in Fig. 2 and the vertices V_1, V_2, V_3 of the tetrahedron in Fig. 3. When one eigenvalue is double, the matrix is diagonalizable iff the square of the eigenvector associated to the simple eigenvalue is either 0 or 1 (of course, ρ_0 is diagonal in any basis).

Note that the action of the rotation group on $\mathcal{D}_1$ is not irreducible. The tensor product $D^1(R)\otimes D^1(R)$ can be reduced to the sum $D^0(R)+D^1(R)+D^2(R)$. Correspondingly, the domain $\mathcal{D}_1$ splits into subspaces globally invariant under rotations and each matrix of D_1 is a sum of elements of these subspaces:

$$\begin{aligned}&\mathcal{D}_1=\{\rho_0\}\oplus\mathcal{D}_1(1)\oplus\mathcal{D}_1(2)\,,\\&\rho=\rho_0+\rho_1+\rho_2,\quad \rho_L\in\mathcal{D}_1(L),\quad L=1,2\,.\end{aligned} \tag{26}$$

Using the antiunitary operator, we define a new matrix $\tilde{\rho}$ having the following property:

$$\tilde{\rho}=A\rho A=\Gamma\rho^*\Gamma=\rho_0-\rho_1+\rho_2\,. \tag{27}$$

A particular case, very important for practical applications,[5] occurs when $\rho_1=0$, i.e., $\rho=\tilde{\rho}$. The corresponding matrices are quadrupole ones. We can ask which are the orbits of quadrupole matrices (besides ρ_0). The condition $\rho=\tilde{\rho}$ implies

(i) $|\varphi_1|=|\varphi_2|=|\varphi_3|=1$ if $\lambda_1<\lambda_2<\lambda_3$ (vertex V in Fig. 3),

(ii) $|\varphi_s|=1$ if λ_s is simple and λ_d double.

Two eigenvalues are then sufficient to characterize the

TABLE I. SO(3) strata of the polarization domain $\mathcal{D}_1$, the orbits characterized by the eigenvalues λ_i of the matrices and the squares $|\varphi_i|$ of the associated eigenvectors.

Stabilizers	Dim. of orbits	Subsets of the strata $(i,j,k)=\text{perm}(1,2,3)$	Dim. of subsets
SO(3)	0	$\{\lambda_1=\lambda_2=\lambda_3=\frac{1}{3}\}$	0
O(2)	2	$\{\lambda_1<\lambda_2=\lambda_3,\ \|\varphi_1\|=1\}$	1
		$\{\lambda_1=\lambda_2<\lambda_3,\ \|\varphi_3\|=1\}$	1
		$\{\lambda_1<\lambda_2<\lambda_3,\ \|\varphi_1\|=\|\varphi_2\|=\|\varphi_3\|=1\}$	2
SO(2)	2	$\{\lambda_1<\lambda_2=\lambda_3,\ \|\varphi_1\|=0\}$	1
		$\{\lambda_1=\lambda_2<\lambda_3,\ \|\varphi_3\|=0\}$	1
		$\{\lambda_1<\lambda_2<\lambda_3,\ \|\varphi_i\|=1,\ \|\varphi_j\|=\|\varphi_k\|=0,\ i=1,2,3\}$	2
Z_2	3	$\{\lambda_1<\lambda_2=\lambda_3,\ 0<\|\varphi_1\|<1\}$	2
		$\{\lambda_1=\lambda_2<\lambda_3,\ 0<\|\varphi_3\|<1\}$	2
		$\{\lambda_1<\lambda_2<\lambda_3,\ \|\varphi_i\|=1,\ 0<\|\varphi_j\|=\|\varphi_k\|<1,\ i=1,2,3\}$	3
I	3	$\left\{\lambda_1<\lambda_2<\lambda_3,\ \sum_{i=1}^{3}\|\varphi_i\|\geq 1,\ \sum_{i=1}^{3}\|\varphi_i\|<1+2\|\varphi_j\|,\ j=1,2,3\right\}$	5

SO(3) orbits of the quadrupole matrices of $\mathcal{D}_1$. This means that each SU(3) orbit of $\mathcal{D}_1$ contains one and only one SO(3) orbit of quadrupole matrices. Doncel, Michel, and Minnaert[6] have already characterized the quadrupole orbits for arbitrary spins by means of two invariants which are functions of the eigenvalues.

The matrix $\tilde{\rho}$ allows us to build rotational invariants which can be used to characterize the SO(3) orbits of $\mathcal{D}_1$. Similarly, instead of the eigenvalues, the degree of polarization and the determinant of the density matrices can be used to specify the SU(3) orbits, of $\mathcal{D}_1$. These quantities and their relations to the eigenvalues and the squares of the eigenvectors are given in the Appendix.

Fano in Ref. 7 characterizes the states of a spin one-particle in a weak external electromagnetic field. This problem is actually equivalent to characterize the orbits of the rotation group. Fano chooses a frame in which

$$\langle J_i J_j + J_j J_i \rangle = 0, \quad i \neq j .$$

The density matrix here takes the form

$$\rho = \tfrac{1}{3} I + \tfrac{1}{2} \sum_{i=1}^{3} \langle J_i \rangle J_i + \tfrac{1}{6} \langle Q_0 \rangle Q_0 - \tfrac{1}{2} \langle Q_2 \rangle Q_2$$

where $Q_0 = 3J_3^2 - 2$, $Q_2 = J_1^2 - J_2^2$. The five parameters which label the orbits are given by the three components of $\langle \mathbf{J} \rangle$, $\langle Q_0 \rangle$, and $\langle Q_2 \rangle$. It is to be noticed that this labeling is not invariant by rotation. Moreover the range of these five parameters is of a rather complicated nature.

In conclusion, this method of characterizing the SO(3) orbits contained in the orbits of the special unitary group could be generalized for higher spin density matrices, but unfortunately contrary to the spin-one case, we do not have a simple parametrization of the orbits of pure states for arbitrary spins.

ACKNOWLEDGMENT

The authors thank Professor Louis Michel for several enlightening discussions on this subject and related topics.

APPENDIX

To characterize the SU(3) orbits of $\mathcal{D}_1$, instead of two eigenvalues, we can use, the degree of polarization and the determinant up to a numerical factor. They are related to the eigenvalues by

$$d = [\tfrac{1}{2}(3\,\mathrm{tr}\rho^2 - 1)]^{1/2} = [1 - 3(\lambda_1 + \lambda_2 - \lambda_1\lambda_2 - \lambda_1^2 - \lambda_2^2)]^{1/2}, \tag{A1}$$

$$\Delta = 27 \det \rho = 27\lambda_1\lambda_2(1 - \lambda_1 - \lambda_2) . \tag{A2}$$

Then each orbit is a point (d, Δ) in the domain defined by the inequalities

$$0 \leq d \leq 1, \quad \Delta \geq 0, \quad \Delta_- \leq \Delta \leq \Delta_+ , \tag{A3}$$

with

$$\Delta_\pm = 1 - 3d^2 \pm 2d^3 . \tag{A4}$$

The matrix $\tilde{\rho}$ associated to ρ, see Eq. (A2), can be written as

$$\rho = \sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|, \quad \tilde{\rho} = \sum_i \lambda_i |A\varphi_i\rangle\langle A\varphi_i| , \tag{A5}$$

and because of the commutation rule (4) the quantities

$$\mathrm{tr}\rho^n \tilde{\rho}^m = \sum_{i,j} (\lambda_i)^n (\lambda_j)^m |(\varphi_i, \varphi_j)|^2 \tag{A6}$$

are rotational invariants [but not invariants under the action of SU(3)]. By using Eq. (20), they can be expressed as quadratic functions of the squares of the eigenvectors. To characterize the SO(3) orbits we can choose in addition to the two SU(3) invariants the three independent quantities $\mathrm{tr}\rho\tilde{\rho}$, $\mathrm{tr}\rho^2\tilde{\rho}$, $\mathrm{tr}\rho^2\tilde{\rho}^2$. However these quantities are quadratic on the squares and the bounds (21) cannot be easily applied to them.

*Work supported by the Echanges France–Québec and by the Ministère de l'Education du Gouvernement du Québec.

†Permanent address: Laboratoire de Physique Théorique Université de Bordeaux I, Chemin du Solarium 33170 Gradignan, France.

[1] H. Bacry, J. Math Phys. **15**, 1686 (1974).

[2] P. Minnaert, Phys. Rev. Lett. **16**, 672 (1966) (for spin one) and in *Particle Physics, Les Houches 1971,* edited by C. DeWitt and C. Itsykson (Gordon and Breach, New York, 1972).

[3] E. P. Wigner, in *Group Theoretical Concepts and Methods in Elementary Particle Physics, Istanbul 1964,* edited by F. Gürsey (Gordon and Breach, New York, 1964).

[4] L. Michel and L. Radicati have studied the orbits and strata of SU(3) in the octet (which is an equivalent problem) in Ann. Inst. Henri Poincaré **XVIII**, 185 (1973).

[5] M. G. Doncel, L. Michel, and P. Minnaert, "Matrices Densité de Polarisation," lectures at the Summer School of Gif-sur-Yvette, 1970, edited by R. Salmeron (Laboratoire de l'Ecole Polytechnique, Paris, 1970).

[6] M. G. Doncel, L. Michel, and P. Minnaert, Nucl. Phys. B 38, 477 (1972).

[7] U. Fano, in *Spectroscopy and Group Theoretical Methods in Physics* (North-Holland, Amsterdam, 1968).

Generating functions for the eigenvalues of the Casimir operators of the orthogonal and symplectic groups

C. O. Nwachuku

Department of Mathematics, University of Benin, Benin City, Nigeria

M. A. Rashid*

Department of Physics, University of Ibadan, Ibadan, Nigeria
(Received 28 July 1976)

By constructing the appropriate generating functions, the eigenvalues of the Casimir operators for the orthogonal and the symplectic groups are expressed in terms of *power sums* which are formally the same for the $O(2n)$, $Sp(2n)$, $O(2n+1)$ groups as for the $U(n)$ groups. The results for the $O(2n)$, $Sp(2n)$, and the $O(2n+1)$ groups are written as the corresponding results for the $U(n)$ groups plus very simple *correction terms*. This approach unifies the treatment of the problem for the semisimple Lie groups. Explicit evaluation of the eigenvalues of the Casimir operators becomes very simple.

1. INTRODUCTION

The computation of the eigenvalues of the Casimir operators (the invariant operators) of the semisimple Lie groups has received considerable attention in the literature.[1-8] In a recent paper[9] (hereafter referred to as I), the authors, by diagonalizing directly the matrices introduced by Perelomov and Popov, obtained the eigenvalues C_p of the Casimir operators of order p of the orthogonal and the symplectic groups in a closed and simple form which is convenient for studying their structures [Eq. (2.1) below]. In particular, this form manifests the asymptotic behavior of the C_p for large p. However, its main shortcoming is that each term in the sum is a *fractional* function of the variable λ_i whereas the quantity C_p, as a whole, is indeed a polynomial. Clearly, there must be cancellations. Also since each term contains products of $2n[(2n+1)]$ factors for the cases of $O(2n)$ and $Sp(2n)$ $[O(2n+1)]$, the calculations using Eq. (2.1) become tedious except for very small values of n.

In what follows, we derive the appropriate generating functions $G(z)$ for the orthogonal and the symplectic groups analogous to the ones of Perelomov and Popov for the unitary groups. Then we are able to express the eigenvalues C_p for the orthogonal and the symplectic groups as a finite series of power sums, which are formally the same for the $U(n)$, $O(2n)$, $Sp(2n)$, and the $O(2n+1)$ groups. Our answers are extremely simple, and manifest the *correction effects* obtained for the $O(2n)$, $Sp(2n)$, and $O(2n+1)$ groups from the corresponding results for the $U(n)$ groups [Eqs. (4.9) and (4.15)]. Also it turns out that the answers for the $O(2n+1)$ groups are formally obtainable from those for the $O(2n)$ groups by simply replacing n by $n+\frac{1}{2}$. Our work, which follows closely the work of Ref. 3 for the unitary groups, thus unifies the treatment of this problem for the semisimple Lie groups.

This paper is arranged as follows: Section 2 contains an outline of our notations together with some of the results derived in I which are relevant for our present discussion. In Sec. 3, the generating functions $G(z)$ for the eigenvalues of the Casimir operators are derived for the orthogonal and the symplectic groups. The $G(z)$'s as well as the C_p's are then expressed in terms of the power sums S_k in Sec. 4 which contains the key results of this paper. Section 5 deals with the more explicit evaluation of these eigenvalues for particular values of p and for special types of representations.

2. NOTATION

We restrict our considerations to *tensor* representations only. The irreducible (tensor) representations of the orthogonal and the symplectic groups may be characterized[9] by n integers $f_n, f_{n-1}, \ldots, f_1$ ordered such that

$$f_n \geq f_{n-1} \geq \cdots \geq f_1.$$

These n integers are the eigenvalues of the n diagonal generators of these groups in the highest state of the representation. In the following we require f_i $(-n \leq i \leq n)$ where the f_{-i} are related to f_i by

$$f_{-i} = -f_i.$$

[Consistently this implies that $f_0 = 0$ for the case of the $O(2n+1)$ groups.]

The eigenvalue $C_p(f_1, f_2, \ldots, f_n)$ of the Casimir operator of order p corresponding to this representation characterized by $f_n, f_{n-1}, \ldots, f_1$ has been shown [Eq. (3.6) in I] to be

$$C_p = \sum_{i=-n}^{n} \lambda_i^p \left\{ \begin{matrix} \lambda_{-i} - \lambda_i \\ 2 + \lambda_{-i} - \lambda_i \\ -(1+\lambda_0-\lambda_i)(1+\lambda_{-i}-\lambda_i) - (\lambda_0 - \lambda_i) \end{matrix} \right\} \times \prod_{\substack{j=-n \\ \neq 0, \pm i}}^{n} (1+\lambda_j-\lambda_i) \Big/ \prod_{\substack{j=-n \\ \neq i}}^{n} (\lambda_j - \lambda_i), \tag{2.1}$$

where the terms in $\{\}$ apply to the $O(2n)$, $Sp(2n)$, and $O(2n+1)$ respectively, and the summation as well as the products in the denominators include zero only for the case of the $O(2n+1)$. In the above equations [as is displayed in Eq. (2.14) in I]

for the $O(2n)$:

$$\lambda_i = f_i + n + i - (1+\epsilon_i) \tag{2.2}$$

and

$$\lambda_{-i} = -\lambda_i + 2n - 2, \tag{2.3}$$

for the Sp(2n):

$$\lambda_i = f_i + n + i \tag{2.4}$$

and

$$\lambda_{-i} = -\lambda_i + 2n, \tag{2.5}$$

for the O(2n+1):

$$\lambda_i = f_i + n + i - \theta_{0i}, \tag{2.6}$$

$$\lambda_{-i} = -\lambda_i + 2n - 1 \quad (i \neq 0), \tag{2.7}$$

and

$$\lambda_0 = n \tag{2.8}$$

where

$$\epsilon_i = 1, -1, 0 \text{ for } i > 0,\ i < 0, \text{ and } i = 0,$$

respectively, and

$$\theta_{ji} = 1\ (0) \text{ for } j < i\ (j \geq i).$$

We may also introduce[10] the power sums S_k defined by

$$S_k = \sum_{i=-n}^{n} (\lambda_i^k - \rho_i^k), \tag{2.9}$$

where

$$\rho_i = \lambda_i - f_i = n + i - (1 + \epsilon_i), n + i, n + i - \theta_{0i}, \tag{2.10}$$

for O(2n), Sp(2n), and O(2n+1) groups, respectively. Evidently

$$S_0 = S_1 = 0. \tag{2.11}$$

Using Eqs. (2.3), (2.5), (2.7), and (2.8), we can rewrite Eq. (2.1) in the form

$$C_p = \sum_{i=-n}^{n} \lambda_i^p \left\{ \begin{array}{c} \dfrac{\lambda_i - n + 1}{\lambda_i - n + \frac{1}{2}} \\ \dfrac{\lambda_i - n - 1}{\lambda_i - n - \frac{1}{2}} \\ \dfrac{\lambda_i - n - \frac{1}{2}}{\lambda_i - n - 1} \end{array} \right\} \prod_{\substack{j=-n \\ \neq i}}^{n} \left(1 - \frac{1}{\lambda_i - \lambda_j}\right), \tag{2.12}$$

where again the summation and the products include zero only for the O(2n+1) groups.

3. THE GENERATING FUNCTIONS

(a) *The case of the* O(2n) *and the* Sp(2n) *groups*: We now transform Eq. (2.12) into a "contour integral" in the λ plane,[11] obtaining

$$C_p = -\frac{1}{2\pi i} \oint d\lambda\, \lambda^p \frac{\lambda - n \pm 1}{\lambda - n \pm \frac{1}{2}} \prod_{i=-n}^{n} \left(1 - \frac{1}{\lambda - \lambda_i}\right) \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^p \prod_{i=-n}^{n} \left(1 - \frac{1}{n \mp \frac{1}{2} - \lambda_i}\right), \tag{3.1}$$

where the upper (lower) sign applies to the O(2n) [Sp(2n)] case. This convention is to be understood throughout. The path of the integration may be taken (in a positive sense) along any large circle with the origin as center and containing all the poles of the integrand in λ. Note that the additional term in Eq. (3.1) above takes into account the pole at $\lambda = n \mp \frac{1}{2}$ present in the integrand which has no counterpart term in the expression in Eq. (2.12).

Using Eqs. (2.3) and (2.5), it is easy to show that

$$\prod_{i=-n}^{n} \left(1 - \frac{1}{n \mp \frac{1}{2} - \lambda_i}\right) = 1,$$

so that, making the change of variable $\lambda = 1/z$, Eq. (3.1) becomes

$$C_p = -\frac{1}{2\pi i} \oint \frac{f(z)\,dz}{z^{p+2}} \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^p, \tag{3.2}$$

where

$$f(z) = \frac{1 - (n \mp 1)z}{1 - (n \mp \frac{1}{2})z} \prod_{i=-n}^{n} \left(1 - \frac{z}{1 - \lambda_i z}\right). \tag{3.3}$$

In Eq. (3.2), the integration is taken (in the positive sense) along any closed path containing the origin but excluding all the poles in the function $f(z)$.

From Eq. (3.2), it is clear that the quantity $\pm \frac{1}{2}(n - \frac{1}{2})^p - C_p$ is the coefficient of z^{p+1} in the Taylor series expansion of $f(z)$ about the origin, i.e.,

$$f(z) = 1 - \sum_{p=0}^{\infty} (C_p \mp \tfrac{1}{2}(n \mp \tfrac{1}{2})^p)\, z^{p+1}. \tag{3.4}$$

In terms of the generating function $G(z)$ defined by

$$G(z) = \sum_{p=0}^{\infty} C_p z^p, \tag{3.5}$$

Eq. (3.4) can be rewritten as[12]

$$zG(z) = 1 - f(z) \pm \tfrac{1}{2} \sum_{p=0}^{\infty} (n \mp \tfrac{1}{2})^p z^{p+1}, \tag{3.6}$$

where $f(z)$ is given in Eq. (3.3).

(b) *The case of the* O(2n+1) *groups*: The "contour integral representation" for the C_p in this case is[11]

$$C_p = -\frac{1}{2\pi i} \oint d\lambda\, \lambda^p \frac{\lambda - n - \frac{1}{2}}{\lambda - n - 1} \prod_{i=-n}^{n} \left(1 - \frac{1}{\lambda - \lambda_i}\right) + \tfrac{1}{2} n^p \prod_{\substack{i=-n \\ \neq 0}}^{n} \left(1 - \frac{1}{n - \lambda_i}\right). \tag{3.7}$$

We remark that in the derivation of the above equation, the following points were especially noted:

(i) The factor in the product in the integrand corresponding to $i = 0$ eliminates the apparent pole at $\lambda = n + 1$.

(ii) The term corresponding to $i = 0$ in the summation in Eq. (2.11) is not correctly reproduced by the residue at the pole at $\lambda = \lambda_0 = n$ in Eq. (3.7). The additional term in this equation is to make up for this deficiency.

As in case (a) above, using Eq. (2.7), it can be easily shown that

$$\prod_{\substack{i=-n \\ \neq 0}}^{n} \left(1 - \frac{1}{n - \lambda_i}\right) = 1,$$

so that on changing the variable of integration to $z = 1/\lambda$, Eq. (3.7) becomes

$$C_p = -\frac{1}{2\pi i} \oint \frac{f(z)\,dz}{z^{p+2}} + \tfrac{1}{2} n^p, \tag{3.8}$$

where

$$f(z) = \frac{1 - (n + \frac{1}{2})z}{1 - (n + 1)z} \prod_{i=-n}^{n} \left(1 - \frac{z}{1 - \lambda_i z}\right) \tag{3.9}$$

in terms of which, the generating function for the $O(2n+1)$ groups takes the form[12]

$$zG(z) = 1 - f(z) + \tfrac{1}{2}\sum_{p=0}^{\infty} n^p z^{p+1}. \tag{3.10}$$

4. EXPRESSION OF THE GENERATING FUNCTIONS AND THE EIGENVALUES OF THE CASIMIR OPERATORS IN TERMS OF THE POWER SUMS

(a) *The case of the* $O(2n)$ *and the* $Sp(2n)$ *groups*: Equation (3.3) leads to

$$\ln f(z) = -\sum_{k=1}^{\infty} \frac{z^k}{k}\left[(n \mp 1)^k - (n \mp \tfrac{1}{2})^k + \sum_{i=-n}^{n}((\lambda_i + 1)^k - \lambda_i^k)\right]. \tag{4.1}$$

Next we try to express $\sum_{i=-n}^{n}[(\lambda_i+1)^k - \lambda_i^k]$ in terms of power sums which were defined in Eq. (2.9). Indeed

$$\begin{aligned}\sum_{i=-n}^{n}((\lambda_i+1)^k - \lambda_i^k) &= \sum_{l=0}^{k-1}(k_l)\left[\sum_{i=-n}^{n}(\lambda_i^l - \rho_i^l) + \sum_{i=-n}^{n}\rho_i^l\right] \\ &= \sum_{l=0}^{k-1}(k_l)S_l + \sum_{i=-n}^{n}((\rho_i+1)^k - \rho_i^k) \\ &= \sum_{l=0}^{k-1}(k_l)S_l + (2n \mp 1)^k + n^k - (n \mp 1)^k. \end{aligned} \tag{4.2}$$

In the last step above, Eq. (2.10) has been used.

The above two equations lead to

$$\ln f(z) = -\sum_{k=1}^{\infty}\frac{z^k}{k}\quad (2n \mp 1)^k + n^k - (n \mp \tfrac{1}{2})^k + \sum_{l=0}^{k-1}(k_l)S_l \tag{4.3}$$

or

$$f(z) = -\frac{[1-(2n \mp 1)z](1-nz)}{[1-(n \mp \frac{1}{2})z]}\exp[-\phi(z)] \tag{4.4}$$

where, since $S_0 = S_1 = 0$ [Eq. (2.11)],

$$\phi(z) = \sum_{k=3}^{\infty} a_k z^k, \quad a_k = \frac{1}{k}\sum_{l=2}^{k-1}(k_l)S_l. \tag{4.5}$$

Combining Eqs. (3.6) and (4.4) results in

$$G(z) = \frac{1}{z[1-(n \mp \frac{1}{2})z]}\{1-(n \mp 1)z - [1-(2n \mp 1)z](1-nz) \times \exp(-\phi(z)\}. \tag{4.6}$$

Further simplication is achieved by introducing the quantities B_p defined by[3]

$$\exp[-\phi(z)] = 1 - \sum_{p=2}^{\infty} B_p z^{p+1} \quad (\text{or } B_0 = B_1 = 0) \tag{4.7}$$

in terms of which

$$G(z) = 2n + \sum_{p=1}^{\infty}(B_p - n\,B_{p-1})z^p - \sum_{p=2}^{\infty}\sum_{q=1}^{p-1} \times (B_q - nB_{q-1})(n \mp \tfrac{1}{2})^{p-q}z^p. \tag{4.8}$$

Comparing Eq. (4.8) with the definition of $G(z)$ in Eq. (3.5), we finally obtain

$$C_p = 2n\delta_{p0} + (B_p - nB_{p-1}) - \sum_{q=1}^{p-1}(B_q - nB_{q-1})(n \mp \tfrac{1}{2})^{p-q}. \tag{4.9}$$

(b) *The case of the* $O(2n+1)$ *groups*: For this case, Eq. (2.10) gives

$$\sum_{i=-n}^{n}[(\lambda_i+1)^k - \lambda_i^k] = \sum_{l=0}^{k-1}(k_l)S_l + (2n)^k + (n+1)^k - n^k \tag{4.10}$$

which when substituted into $\ln f(z)$ computed from Eq. (3.9) results in

$$\ln f(z) = -\sum_{k=1}^{\infty}\frac{z^k}{k}\left[(2n)^k + (n+\tfrac{1}{2})^k - n^k + \sum_{l=0}^{k-1}(k_l)S_l\right] \tag{4.11}$$

or

$$f(z) = \frac{(1-2nz)[1-(n+\frac{1}{2})z]}{1-nz}\exp[-\phi(z), \tag{4.12}$$

where $\phi(z)$ is formally the same as in Eq. (4.5) for case (a). Combining the above equation with Eq. (3.10) results in

$$G(z) = \frac{1}{z(1-nz)}\{1-(n-\tfrac{1}{2})z - (1-2nz)[1-(n+\tfrac{1}{2})z]\exp[-\phi(z)\}. \tag{4.13}$$

In terms of the B_p defined in Eq. (4.7), we have

$$G(z) = (2n+1) + \sum_{p=1}^{\infty}[B_p - (n+\tfrac{1}{2})B_{p-1}]z^p - \sum_{p=2}^{\infty}\sum_{q=1}^{p-1}[B_q - (n+\tfrac{1}{2})B_{q-1}]n^{p-q}z^p. \tag{4.14}$$

Finally

$$C_p = (2n+1)\delta_{p0} + B_p - (n+\tfrac{1}{2})B_{p-1} - \sum_{q=1}^{p-1}[B_q - (n+\tfrac{1}{2})B_{q-1}]n^{p-q}. \tag{4.15}$$

A short discussion of our main results contained in Eqs. (4.8), (4.9), (4.14), and (4.15) is in order.

(i) Noting that the eigenvalues of the Casimir operators of the $U(n)$ groups can be written in the form [Eq. (17) of Ref. 3]

$$C_p = n\delta_{p0} + B_p - nB_{p-1}, \tag{4.16}$$

we see that our Eqs. (4.9) and (4.15) manifest the *correction effects* for the $O(2n)$, $Sp(2n)$, and the $O(2n+1)$ groups over the corresponding results for the $U(n)$ groups.

(ii) The results for the $O(2n+1)$ groups are formally the same as for the $O(2n)$ groups with the replacement $n \to n+\frac{1}{2}$. This results in a *synthesis* of the two types of results. This synthesis is fully exploited in Section 5, where detailed evaluation of the C_p's for the special cases is given.

(iii) Equations (4.9) and (4.15) are extremely simple in form and easy to handle. To compute C_p, all we need are the B_q's for $1 \le q \le p$ in terms of the power sums S_k.

5. SPECIAL CASES

In this section, we work out the complete answers for some particular values of p and in some special cases for the $O(2n)$ and the $Sp(2n)$ groups. The corresponding answers for the $O(2n+1)$ groups are automatically obtained using ansatz (ii) in Sec. 4.

Eq. (4.5) gives

$$\phi(z) = S_2 z^3 + (S_3 + \tfrac{3}{2}S_2)z^4 + (S_4 + 2S_3 + 2S_2)z^5 + (S_5 + \tfrac{5}{2}S_4 + \tfrac{10}{3}S_3 + \tfrac{5}{2}S_2)z^6 + (S_6 + 3S_5 + 5S_4 + 5S_3 + 3S_2)\, z^7 + \cdots. \tag{5.1}$$

Computing $\exp[-\phi(z)]$ and using the definition of B_p in Eq. (4.7) we find

$$\begin{aligned} &B_0 = B_1 = 0, \\ &B_2 = S_2, \\ &B_3 = S_3 + \tfrac{3}{2}S_2, \\ &B_4 = S_4 + 2S_3 + 2S_2, \\ &B_5 = S_5 + \tfrac{5}{2}S_4 + \tfrac{10}{3}S_3 + \tfrac{5}{2}S_2 - \tfrac{1}{2}S_2^2, \\ &B_6 = S_6 + 3S_5 + 5S_4 + 5S_3 + 3S_2 - \tfrac{3}{2}S_2^2 - S_2S_3, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.2}$$

Using $B_0 = B_1 = 0$, Eq. (4.9) can be rewritten in a somewhat more convenient form (from the point of view of actual computation) as

$$C_p = 2n\delta_{p0} + B_p - (2n \mp \tfrac{1}{2})\, B_{p-1} \pm \tfrac{1}{2} \sum_{q=1}^{p-3} (n \mp \tfrac{1}{2})^q\, B_{p-q-1} \tag{5.3}$$

from which we obtain

$$\begin{aligned} &C_0 = 2n, \\ &C_1 = 0, \\ &C_2 = B_2, \\ &C_3 = B_3 - (2n \mp \tfrac{1}{2})\, B_2, \\ &C_4 = B_4 - (2n \mp \tfrac{1}{2})\, B_3 \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})\, B_2, \\ &C_5 = B_5 - (2n \mp \tfrac{1}{2})\, B_4 \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})\, B_3 \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^2 B_2, \\ &C_6 = B_6 - (2n \mp \tfrac{1}{2})\, B_5 \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})\, B_4 \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^2 B_3 \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^3 B_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.4}$$

For the $O(2n+1)$ groups, to compute the corresponding C_p's, just replace n by $n + \frac{1}{2}$ and use the upper sign for $O(2n)$.

(a) *The completely symmetric representation*: For the completely symmetric representation $(f, 0, 0, \cdots)$ for which $f_n = -f_{-n} = f$, $f_i = -f_{-i} = 0$ $(1 \leq i \leq n-1)$, we have:

for the $_0(2n)$:

$$\begin{aligned} &S_2 = 2f(f + 2n - 2), \\ &S_3 = 3(n-1)S_2, \\ &S_4 = [f^2 + 2(n-1)f + 8(n-1)^2]S_2, \\ &S_5 = 5(n-1)S_4 - 20(n-1)^3 S_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.5}$$

and

$$\begin{aligned} &C_2 = S_2, \\ &C_3 = (n-1)C_2, \\ &C_4 = [f^2 + 2(n-1)f + 2(n-1)^2 - (n-1)]C_2, \\ &C_5 = (3n-2)C_4 - (n+1)^2(2n-1)C_2 - \tfrac{1}{2}C_2^2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.6}$$

for the $\mathrm{Sp}(2n)$:

$$\begin{aligned} &S_2 = 2f(f + 2n), \\ &S_3 = 3nS_2, \\ &S_4 = (f^2 + 2nf + 8n^2)S_2, \\ &S_5 = 5nS_4 - 20n^3 S_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.7}$$

and

$$\begin{aligned} &C_2 = S_2, \\ &C_3 = (n+1)C_2, \\ &C_4 = (f^2 + 2nf + 2n^2 + n + 1)C_2, \\ &C_5 = (3n+2)C_4 - (n-1)^2(2n+1)\, C_2 - \tfrac{1}{2}C_2^2. \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.8}$$

We note that the Casimir operators of odd order p are not independent and can be expressed in terms of those of even order $2q$ with $2q < p$. Thus the results we get for C_3 and C_5 in Eqs. (5.4) and (5.7) (also for S_3 and S_5) hold for any representation, and will not be repeated in the cases considered below.

(b) *The block representation*: Here we consider the block representation $(f, f, \ldots, f, 0, 0, \ldots, 0)$ with f boxes in the first k rows, defined by

$$\begin{aligned} &f_n = f_{n-1} = \cdots = f_{n-k+1} = f, \\ &f_{n-k} = f_{n-k-1} = \cdots f_1 = 0, \end{aligned}$$

for the $O(2n)$:

$$\begin{aligned} &S_2 = 2fk(f + 2n - k - 1), \\ &S_4 = 2fk[(f-k)(f^2 - fk + k^2 + 12n^2 - 18n + 7) \\ &\qquad + (2n-1)[2(f-k)^2 + fk + 2(n-1)(4n-3)]], \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.9}$$

and

$$\begin{aligned} &C_2 = S_2, \\ &C_4 = 2fk[(f-k)(f^2 - fk + k^2) + (2n-1) \\ &\qquad \times [kf + (f-k)(3n - 2 + 2f - 2k) + 2(n-1)(n+2)]], \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.10}$$

for the $\mathrm{Sp}(2n)$:

$$\begin{aligned} &S_2 = 2fk(f + 2n - k + 1), \\ &S_4 = 2fk[(f-k)(f^2 - fk + k^2 + 12n^2 \\ &\qquad + 6n + 1) + (2n+1)[fk + 2(f-k)^2 + 2n(4n+1)]], \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.11}$$

and

$$\begin{aligned} &C_2 = S_2, \\ &C_4 = 2fk[(f-k)(f^2 - fk + k^2) + (2n+1) \\ &\qquad \times [kf + (f-k)(3n + 2 + 2f - 2k) + (n+1)(2n+1)]]. \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5.12}$$

(c) *The completely antisymmetric representation*: The results for the completely antisymmetric repre-

sentation $\{1^k\} = (1, 1, \ldots, 1, 0, 0, \ldots, 0)$ with one box in the first k rows are deduced from Eqs. (5.10) and (5.12) by setting $f = 1$:

for the O(2n):

$$C_2 = S_2 = 2k(2n - k),$$
$$C_4 = 2k\{(1 - k)(1 - k + k^2) + (2n - 1)[k + (1 - k)(3n - 2k) + 2(n - 1)(n + 2)]\},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (5.13)$$

for the Sp(2n):

$$C_2 = S_2 = 2k(2n + 2 - k),$$
$$C_4 = 2k\{(1 - k)(1 - k + k^2) + (2n + 1) \times [k + (1 - k)(3n + 4 - 2k) + (n + 1)(2n + 1)]\},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (5.14)$$

We note that the representations $\{1^k\}$ for $k = 1, 2, \ldots, n$ are the simplest representations the products of which on reduction will give all the irreducible tensor representations of the orthogonal and the symplectic groups.

In all cases, the results for the O(2n + 1) are obtained from the corresponding results for the O(2n) by replacing n by $n + \frac{1}{2}$. Our answers agree with those of Perelomov and Popov, and Yeh and Wong,[7] for those particular cases treated here which were considered by them.

*On leave of absence from the Department of Mathematics, Quaid-e-Azam University, Islamabad (Pakistan).

[1]A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. 3, 676 (1966).
[2]A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. 3, 819 (1966).
[3]A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. 5, 489 (1967).
[4]J. D. Louck, Am. J. Phys. 38, 3 (1970).
[5]J. D. Louck and L. C. Biedenharn, J. Math. Phys. 11, 2368 (1970)
[6]S. Okubo, J. Math. Phys. 16, 528 (1975).
[7]M. X. F. Wong and H. Y. Yeh, J. Math. Phys. 16, 1239 (1975).
[8]Our list of references is not exhaustive. A better list appears in Ref. 7 and the Introduction in this reference contains information on the contents of these references.
[9]C. O. Nwachuku and M. A. Rashid, J. Math. Phys. 17, 1611 (1976).
[10]Note that our S_k is formally the same as that defined for the $U(n)$ in Eq. (7) of Ref. 3 and is different from that defined in Eq. (19) of Ref. 2 in which the r_i obeys a reflection symmetry $r_i = -r_{-i}$; whereas our ρ_i does not.
[11]The inverted commas on the contour integral signify the presence of an additional term in Eqs. (3.1) and (3.7).
[12]We could have replaced the last term in Eq. (3.6) by $\frac{1}{2}z/[1 - (n_+^-1/2)z]$ but it is not necessary. A similar remark applies to Eq. (3.10).

Hydrostatic density distribution of fluids in steep field gradients and near critical points

D. J. Gates

C.S.I.R.O. Division of Mathematics and Statistics, Canberra, A.C.T., Australia
(Received 20 September 1976)

A new perturbation expansion for the hydrostatic density distribution of a fluid in a steep field gradient is presented. It is a power series in the scale γ of an external potential $\psi(\gamma x)$, and is asymptotically valid as $\gamma \to 0$, although convergence is not proved. Significant corrections to the conventional hydrostatic equations are found near the critical point of a fluid, even in the gravitational field.

1. INTRODUCTION

The hydrostatic equilibrium of a fluid is described by the equation

$$\frac{1}{\bar{n}}\frac{\partial}{\partial \mathbf{y}} p(\bar{n}) = -\frac{\partial}{\partial \mathbf{y}}\psi, \tag{1.1}$$

where $\bar{n}(\mathbf{y})$ is the number density at a point $\mathbf{y}$, $\psi(\mathbf{y})$ is an external potential, and $p(\rho)$ is the pressure of a uniform fluid of density ρ. The equation is derived from a macroscopic model of the fluid and does not describe small scale density variations near surfaces. It, in effect, assumes that the change in potential energy is small over distances of the order of a molecular diameter. We consider here the response of fluids to steeper field gradients and the resulting corrections to (1.1). For a fluid close to its gas–liquid critical point, we find that these corrections are significant even in the gravitational field.

For states well away from the critical point, the corrections are significant only for very steep field gradients such as those imposed by container walls or by a porous medium saturated with the fluid.

If the system is subject to an external field $\psi(\gamma\mathbf{x})$ and we define

$$\bar{n}(\mathbf{y}) = \lim_{\gamma \to 0} \langle n(\mathbf{y}/\gamma, \psi)\rangle, \tag{1.2}$$

where $n(\mathbf{x}, \psi)$ is the exact number density for a system of interacting molecules in this field, and $\langle\ \rangle$ denotes a suitable space average on the molecular scale, then one can prove that (1.1) holds exactly.[1,2] The limit implies a vanishing field gradient on the molecular scale. We examine the case where the field gradient is small but nonvanishing on this scale by expanding $n(\mathbf{y}/\gamma, \psi)$ as a power series in γ.

In Secs. 2 and 3 we derive a prescription for the general coefficient of this expansion. In Secs. 4 and 5 the deviations from (1.1) are estimated.

A different approach was adopted by Lebowitz and Percus[3] (see also Ref. 4). Their method is based on a functional Taylor expansion of the chemical potential in terms of density gradients. They obtain only the first term of this expansion and do not give a prescription for obtaining further correction terms. Our method is more systematic and is known to be at least asymptotically correct[2] in the limit $\gamma \to 0$.

2. EXPANSION IN TENSOR FORM

The modified Ursell correlation functions $\hat{u}(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mu)$ of a continuous classical system may be expressed as functional derivatives of the local density, or one-particle distribution function,[5,6]

$$\hat{u}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mu) = \left(\frac{-1}{\beta}\right)^{k-1} \frac{\delta^{k-1}}{\delta\Psi(\mathbf{x}_2)\cdots\delta\Psi(\mathbf{x}_k)} n(\mathbf{x}_1, \mu, \Psi)\big|_{\Psi=0}, \tag{2.1}$$

where μ is the chemical potential, β the reciprocal temperature, $\Psi(\mathbf{x})$ is an external potential, and $n(\mathbf{x}, \mu, \Psi)$ denotes the one-particle distribution function for a system in this potential. The notation indicates that the derivative is evaluated in zero field to give $\hat{u}_k$ in zero field.

If we take

$$\Psi(\mathbf{x}) = \psi(\gamma\mathbf{x}) - \psi(\gamma\mathbf{x}_1) \tag{2.2}$$

where $\mathbf{x}_1$ is regarded as fixed (since it does not enter the differentiation), then $\psi(\gamma\mathbf{x}_1)$ simply subtracts from the chemical potential. Thus (2.1) takes the form

$$\hat{u}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mu - \psi(\gamma\mathbf{x}_1)) = \left(\frac{-1}{\beta}\right)^{k-1} \frac{\delta^{k-1}}{\delta\psi(\gamma\mathbf{x}_2)\cdots\delta\psi(\gamma\mathbf{x}_k)} n(\mathbf{x}_1, \mu, \psi)\big|_{\psi(\gamma\mathbf{x})=\psi(\gamma\mathbf{x}_1)}. \tag{2.3}$$

Consequently, the functional Taylor expansion of $n(\mathbf{x}_1, \mu, \psi)$ about the value $\psi(\gamma\mathbf{x}_1)$ of the field, at the point $\mathbf{x}_1$ where the density is represented, takes the form

$$\begin{aligned} n(\mathbf{x}_1, \mu, \psi) = {} & \rho[\mu - \psi(\gamma\mathbf{x}_1)] \\ & + \sum_{n=2}^{\infty} \frac{(-\beta)^{n-1}}{(n-1)!} \int d\mathbf{x}_2 \cdots \\ & \times \int d\mathbf{x}_n \hat{u}_n[\mathbf{x}_1, \ldots, \mathbf{x}_n, \mu - \psi(\gamma\mathbf{x}_1)] \\ & \times \prod_{i=2}^{n} [\psi(\gamma\mathbf{x}_i) - \psi(\gamma\mathbf{x}_1)], \end{aligned} \tag{2.4}$$

where the integrations are over all of the coordinate space, and $\rho(\mu)$ is the one-particle density in the absence of a field and is constant for the fluid states considered here.

Now we put $\mathbf{x}_1 = \mathbf{y}/\gamma$ and change to the integration variables $\mathbf{r}_i = \mathbf{x}_{i+1} - \mathbf{y}/\gamma$, which yields

$$\begin{aligned} n(\mathbf{y}/\gamma, \mu, \psi) = {} & \rho[\mu - \psi(\mathbf{y})] + \sum_{v=1}^{\infty} [(-\beta)^v / v!] \\ & \times \int d\mathbf{r}_1 \cdots \int d\mathbf{r}_v \hat{F}_{v+1}[\mathbf{r}_1, \ldots, \mathbf{r}_v, \mu - \psi(\mathbf{y})] \\ & \times \prod_{j=1}^{v} [\psi(\mathbf{y} + \gamma\mathbf{r}_j) - \psi(\mathbf{y})], \end{aligned} \tag{2.5}$$

where we have put $v = n - 1$, $j = i - 1$, and

$$\hat{F}_n(\mathbf{r}_1,\ldots,\mathbf{r}_{n-1},\mu)=\hat{u}_n(\mathbf{x},\mathbf{x}+\mathbf{r}_1,\ldots,\mathbf{x}+\mathbf{r}_{n-1},\mu), \quad (2.6)$$

which is independent of $\mathbf{x}$ because $\hat{u}_n$ is translationally invariant for fluid states. To obtain an expansion in powers of γ, we take ψ to be an analytic function and expand $\psi(\mathbf{y}+\gamma\mathbf{r})$ as a Taylor series:

$$\psi(\mathbf{y}+\gamma\mathbf{r})=\sum_{n=0}^{\infty}\frac{1}{n!}\gamma^n\left(\mathbf{r}\cdot\frac{\partial}{\partial\mathbf{y}}\right)^n\psi(\mathbf{y}), \quad (2.7)$$

where $\partial/\partial\mathbf{y}$ denotes the gradient operator. We can write

$$\psi(\mathbf{y}+\gamma\mathbf{r})=\sum_{n=0}^{\infty}\gamma^n[\mathbf{r}]^n\cdot\mathbf{S}^n(\mathbf{y}), \quad (2.8)$$

where

$$\mathbf{S}^n(\mathbf{y})\equiv\frac{1}{n!}\left[\frac{\partial}{\partial\mathbf{y}}\right]^n\psi(\mathbf{y}). \quad (2.9)$$

Here $[\mathbf{v}]^n$ denotes the tensor product of $\mathbf{v}$ with itself n times (i.e., the tensor with components $v_{i_1}v_{i_2}\cdots v_{i_n}$).

The product $\mathbf{A}\circ\mathbf{B}$ for two tensors of the same rank r means the sum

$$\sum_{i_1,\ldots,i_r}A_{i_1\cdots i_r}B_{i_1\cdots i_r}. \quad (2.10)$$

Using (2.8), we obtain

$$\begin{aligned}(-\beta)^v\prod_{j=1}^{v}&[\psi(\mathbf{y}+\gamma\mathbf{r}_j)-\psi(\mathbf{y})]\\ &=\sum_{n_1,n_2,\ldots,n_v}[\gamma\mathbf{r}_1]^{n_1}[\gamma\mathbf{r}_2]^{n_2}\cdots[\gamma\mathbf{r}_v]^{n_v}\cdot\mathbf{S}^{n_1}\mathbf{S}^{n_2}\circ\cdots\mathbf{S}^{n_v}(-\beta)^v\\ &=\sum_{\mathbf{n}\in N_v}\gamma^{\Sigma n_i}\mathbf{R}^{\mathbf{n}}\cdot\mathbf{S}^{\mathbf{n}}, \quad (2.11)\end{aligned}$$

where

$$\mathbf{R}^{\mathbf{n}}=[\mathbf{r}_1]^{n_1}\cdots[\mathbf{r}_v]^{n_v}, \quad (2.12)$$

$$\begin{aligned}\mathbf{S}^{\mathbf{n}}&=(-\beta)^v\mathbf{S}^{n_1}\cdots\mathbf{S}^{n_v}\\ &=\prod_{j=1}^{v}\left(\frac{1}{n_j!}\left[\frac{\partial}{\partial\mathbf{y}}\right]^{n_j}\{-\beta\psi(\mathbf{y})\}\right), \quad (2.13)\end{aligned}$$

both being tensors of rank $n_1+n_2+\cdots+n_v$ and N_v is the space of v-tuples $(n_1,\ldots,n_v)$ where $n_i\geq 1$ for all i.

Substituting (2.10) in (2.5) yields

$$\begin{aligned}n(\mathbf{y}/\gamma,\mu,\psi)&=\rho^0[\mu-\psi(\mathbf{y})]\\ &+\sum_{v=1}^{\infty}\sum_{\mathbf{n}\in N_v}\gamma^{\Sigma n_i}\frac{1}{v!}\int d\mathbf{r}_1\cdots\int d\mathbf{r}_v\\ &\times\hat{F}_{v+1}[\mathbf{r}_1,\ldots,\mathbf{r}_v,\mu-\psi(\mathbf{y})]\mathbf{R}^{\mathbf{n}}\cdot\mathbf{S}^{\mathbf{n}}\\ &=\rho[\mu-\psi(\mathbf{y})]+\sum_{\tau=1}^{\infty}\gamma^{\tau}\sum_{\mathbf{n}\in N_\tau}\mathbf{M}^{\mathbf{n}}[\mu-\psi(\mathbf{y})]\cdot\mathbf{S}^{\mathbf{n}},\end{aligned} \quad (2.14)$$

where

$$\mathbf{M}^{\mathbf{n}}(\mu)=\frac{1}{v!}\int d\mathbf{r}_1\cdots\int d\mathbf{r}_v\,\hat{F}_{v+1}(\mathbf{r}_1,\ldots,\mathbf{r}_v,\mu)\mathbf{R}^{\mathbf{n}} \quad (2.15)$$

and N_τ is the set of vectors $\mathbf{n}$ ($n_i\geq 1$), of any dimension, such that $\sum n_j=\tau$. Both $\mathbf{M}^{\mathbf{n}}$ and $\mathbf{S}^{\mathbf{n}}$ have rank τ.

Equation (2.14) is the desired expansion of $n(\mathbf{y}/\gamma,\mu,\psi)$ as a power series in γ. The first term is precisely the result derived rigorously[2] in the limit $\gamma\to 0$, and gives the result obtained from macroscopic hydrostatics.

The study of the convergence of the expansion appears to be a formidable task in view of our limited knowledge of the Ursell functions. Some progress may be possible using the "strong cluster properties" of Duneau *et al.*[8]

Only in the case of the ideal gas is the convergence of our expansion guaranteed *à priori*. In this case

$$n(\mathbf{y}/\gamma,\psi)=\Lambda^{-3}\exp[\beta\mu-\psi(\mathbf{y})] \text{ exactly}$$

so that only the first term of the expansion appears.

Lemma 1: The $\mathbf{M}^{\mathbf{n}}\circ\mathbf{S}^{\mathbf{n}}$ are invariant under any permutation of the components $(n_1,\ldots,n_v)$ of $\mathbf{n}$.

This follows from the definitions (2.10), (2.13), and (2.15), and enables us to write (2.14) in the form

$$n(\mathbf{y}/\gamma,\mu,\psi)=p[\mu-\psi(\mathbf{y})]+\sum_{\tau=1}^{\infty}\gamma^{\tau}\sum_{\mathbf{n}\in N'_\tau}K(\mathbf{n})\mathbf{M}^{\mathbf{n}}\cdot\mathbf{S}^{\mathbf{n}}, \quad (2.16)$$

where N'_τ is the subset of vectors $\mathbf{n}\in N_\tau$ such that $n_1\leq n_2\leq n_3\cdots$, and $K(\mathbf{n})$ is the number of distinct vectors which can be obtained by permuting the components of $\mathbf{n}$. It is given by

$$K(\mathbf{n})=\left(\sum_{i=1}^{\tau}k_i(\mathbf{n})\right)!\Big/\prod_{i=1}^{\tau}k_i(\mathbf{n})!, \quad (2.17)$$

where $k_i(\mathbf{n})$ is the number of components of $\mathbf{n}$ with value i. Clearly $\sum_i k_i(\mathbf{n})=v$ if $\mathbf{n}$ has v components.

For one-dimensional models the tensors are all scalars, so that (2.14) is the final form. We are however primarily interested in three-dimensional systems.

3. REDUCTION TO SCALAR INVARIANT FORM

Using properties of the tensors $\mathbf{M}^{\mathbf{n}}$ and $\mathbf{S}^{\mathbf{n}}$ enables one to reduce (2.14) to a form more suitable for numerical computations.

Lemma 2: The $\mathbf{M}^{\mathbf{n}}$ are isotropic tensors.

Proof: It is required to prove that the components $M^{\mathbf{n}}_{\alpha_1\cdots\alpha_\tau}$ w.r.t. any Cartesian coordinate system are invariant under rotations, i.e.,

$$\sum_{\alpha_1\cdots\alpha_\tau}A_{\alpha_1\beta_1}A_{\alpha_2\beta_2}\cdots A_{\alpha_\tau\beta_\tau}M^{\mathbf{n}}_{\alpha_1\cdots\alpha_\tau}=M^{\mathbf{n}}_{\beta_1\cdots\beta_\tau}, \quad (3.1)$$

where $A_{\alpha\beta}$ is an arbitrary unitary matrix. To prove this, we note that the left side is the $(\beta_1,\ldots,\beta_\tau)$ component of the expression (omitting μ from the notation)

$$\int d\mathbf{r}_1\cdots d\mathbf{r}_v\hat{F}_{v+1}(\mathbf{r}_1,\ldots,\mathbf{r}_v)[\mathbf{r}_1\cdot\mathbf{A}]^{n_1}\circ\cdots\circ[\mathbf{r}_v\cdot\mathbf{A}]^{n_v}. \quad (3.2)$$

Putting

$$\mathbf{r}'_i=\mathbf{r}_i\cdot\mathbf{A}\quad(i=1,2,\ldots,v) \quad (3.3)$$

yields

$$d\mathbf{r}_i=d\mathbf{r}'_i\quad(i=1,2,\ldots,v) \quad (3.4)$$

by the invariance of volume elements under rotation, and

$$\hat{F}_{v+1}(\mathbf{r}_1,\ldots,\mathbf{r}_v)=\hat{F}_{v+1}(\mathbf{r}'_1,\ldots,\mathbf{r}'_v) \quad (3.5)$$

by the invariance of the fluid state under rotations. Hence (3.2) reduces to $\mathbf{M}^{\mathbf{n}}$, as required.

Let $\mathbf{e}_1,\ldots,\mathbf{e}_\nu$ be an orthornormal basis of R^ν, and define the *elementary tensors*

$$\mathbf{E}(P)=\sum_{\alpha_1\cdots\alpha_\tau}\prod_{1\leq i<j\leq\tau}(\delta_{\alpha_i\alpha_j})^{P_{ij}}\mathbf{e}_{\alpha_1}\circ\cdots\circ\mathbf{e}_{\alpha_\tau}, \quad (3.6)$$

where P is a $\tau\times\tau$ permutation matrix of τ-tuples, so that $P_{ii}=0$, $P_{ij}=P_{ji}$ and every row and column contains a single 1 and $\tau-1$ zeros. Equivalently P may be represented by a graph with every vertex bonded to one other vertex, where P_{ij} denotes the number of bonds connecting vertex i to vertex j. Thus P defines a pairing of the integers $1,2,\ldots,\tau$. For example,

$$\mathbf{E}\begin{pmatrix}0&1\\1&0\end{pmatrix}=\mathbf{E}\left(\begin{array}{c}\bullet\!\!-\!\!-\!\!\bullet\\ 1\quad 2\end{array}\right)=\sum_{\alpha_1,\alpha_2}\delta_{\alpha_1\alpha_2}\mathbf{e}_{\alpha_1}\mathbf{e}_{\alpha_2}$$
$$=\sum_{\alpha}\mathbf{e}_{\alpha}\mathbf{e}_{\alpha}, \tag{3.7}$$

which is the rank-2 unit tensor. Similarly

$$\mathbf{E}\begin{pmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}=\mathbf{E}\left(\begin{array}{cc}1&2\\ \bullet\!\!-\!\!-\!\!\bullet&\\ \bullet\!\!-\!\!-\!\!\bullet&\\ 4&3\end{array}\right)=\sum_{\alpha,\beta}\mathbf{e}_{\alpha}\mathbf{e}_{\alpha}\mathbf{e}_{\beta}\mathbf{e}_{\beta}. \tag{3.8}$$

One can readily show that the $\mathbf{E}(P)$ are isotropic. The most general isotropic tensor of rank τ is expressible in the form of a sum over graphs (or matrices)

$$\mathbf{I}=\sum_{P\in\mathcal{P}_\tau}A(P)\mathbf{E}(P), \tag{3.9}$$

where the $A(P)$ are arbitrary constants and $\mathcal{P}_\tau$ is the set of graphs P with τ vertices. As always, τ is even.

Each P also defines a contraction or scalar invariant of $\mathbf{I}$:

$$I(P)=\mathbf{I}\cdot\mathbf{E}(P)$$
$$=\sum_{P'}A(P')\mathbf{E}(P')\cdot\mathbf{E}(P)$$
$$=\sum_{P'}A(P')E(P',P), \tag{3.10}$$

where

$$E(P,P')=\mathbf{E}(P)\cdot\mathbf{E}(P')=E(P',P)$$
$$=\sum_{\alpha_1\cdots\alpha_\tau}\prod_{ij}(\delta_{\alpha_i\alpha_j})^{P_{ij}P'_{ij}}. \tag{3.11}$$

Suppose Γ is an arbitrary graph (matrix) with Γ_{ij} bonds joining vertex i to vertex j ($\Gamma_{ij}=0,1,2,3,\cdots$). We define the *value* of the graph by

$$v(\Gamma)=\sum_{\alpha_1\cdots\alpha_\tau}\prod_{ij}(\delta_{\alpha_i\alpha_j})^{\Gamma_{ij}}. \tag{3.12}$$

Lemma 3: If Γ consists of n disconnected graphs $\Gamma^1,\ldots,\Gamma^n$, where each Γ^i is itself connected, then

$$v(\Gamma)=\nu^n \tag{3.13}$$

for a ν-dimensional system.

Proof: We can write

$$v(\Gamma)=\sum_{\alpha_1\cdots\alpha_\tau}\prod_{k=1}^{n}\prod_{i,j\in S^k}(\delta_{\alpha_i\alpha_j})^{\Gamma^k_{ij}}, \tag{3.14}$$

where S^k is the set of vertices contained in Γ^k. This reduces to

$$v(\Gamma)=\prod_{k=1}^{n}\sum_{\alpha_m:m\in S^k}\prod_{i,j\in S^k}(\delta_{\alpha_i\alpha_j})^{\Gamma^k_{ij}}$$
$$=\prod_{k=1}^{n}v(\Gamma^k). \tag{3.15}$$

Since each Γ^k is connected, there is a connected permutation $i_1,i_2,\ldots,i_r$ say, of the vertices of Γ^k connected in sequence (i_1 to i_2, i_2 to i_3, and so on). Any additional connections have no effect on $v(\Gamma^k)$ because of the properties of the Krönecker delta. We can therefore write

$$v(\Gamma^k)=\sum_{\alpha_{i_1}\cdots\alpha_{i_r}}\delta_{\alpha_{i_1}\alpha_{i_2}}\delta_{\alpha_{i_2}\alpha_{i_3}}\cdots\delta_{\alpha_{i_{r-1}}\alpha_{i_r}}$$
$$=\sum_{\alpha_{i_1}}\delta_{\alpha_{i_1}\alpha_{i_1}}=\nu. \tag{3.16}$$

Combining (3.15) and (3.16) proves the statement (3.13) of the lemma. In particular we have

$$E(P,P')=v(P+P')$$
$$=\nu^{n(P+P')} \tag{3.17}$$

where $n(P+P')$ is the number of separate parts of $P+P'$ each of which is a connected graph. Since every vertex of $P+P'$ has exactly two bonds, it consists entirely of distinct closed circuits. For example, if $\tau=2$ the only P is $\bullet\!\!-\!\!-\!\!\bullet$ (1, 2) and

$$P+P=\ 1\!\!\bigcirc\!\!2 \tag{3.18}$$

so that $n(P+P)=1$ and $E(P,P)=\nu$. For $\tau=4$ we have three graphs

$$P_1=\ \begin{array}{cc}1&2\\ \bullet\!\!-\!\!-\!\!\bullet\\ \bullet\!\!-\!\!-\!\!\bullet\\ 4&3\end{array}\ ,\quad P_2=\ \begin{array}{cc}1&2\\ \bullet\quad\bullet\\ \bullet\quad\bullet\\ 4&3\end{array}\ ,\quad P_3=\ \begin{array}{cc}1&2\\ \bullet\ \bullet\\ \bullet\ \bullet\\ 4&3\end{array} \tag{3.19}$$

so that

$$P_1+P_1=2P_1=\ \begin{array}{c}1\!\!\bigcirc\!\!2\\ 4\!\!\bigcirc\!\!3\end{array}\ , \tag{3.20}$$

$$P_1+P_2=\ \begin{array}{cc}1&2\\ \bullet\quad\bullet\\ \bullet\quad\bullet\\ 4&3\end{array}\ ,$$

and so on. We can represent the $E(P,P')$ in matrix form. For $\tau=4$ the matrix is

$$E(P_i,P_j)=\begin{pmatrix}\nu^2&\nu&\nu\\ \nu&\nu^2&\nu\\ \nu&\nu&\nu^2\end{pmatrix}. \tag{3.21}$$

In general $E(P_i,P_j)$ is a symmetric matrix of dimension

$$d=\frac{\tau!}{(\tau/2)!\,2^{\tau/2}} \tag{3.22}$$

and it is clear that

$$E(P,P)=\nu^{\tau/2}.$$

Lemma 4: The matrix $E(P,P')$ is positive definite.

To show this, let $\varphi(P_i)$ ($i=1,\ldots,d$) be any nonzero vector, and put

$$\Phi=\sum_P\varphi(P)\mathbf{E}(P), \tag{3.23}$$

which, because of the independence of the $\mathbf{E}(P)$, has at least one nonzero component $\Phi_{\alpha_1\cdots\alpha_\tau}$ with respect to the Cartesian basis $\mathbf{e}_i$. Now we have

$$\sum_{P,P'}\varphi(P)E(P,P')\varphi(P')=\Phi\cdot\Phi$$
$$=\sum_{\alpha_1\cdots\alpha_\tau}(\Phi_{\alpha_1\cdots\alpha_\tau})^2>0, \tag{3.24}$$

as required.

It follows that $E(P,P')$ is a nonsingular matrix.

Denoting its inverse by $E^{-1}(P, P')$, we deduce from (3.10) that

$$A(P) = \sum_{P'} I(P')E^{-1}(P, P'). \tag{3.25}$$

Thus (3.9) becomes

$$\mathbf{I} = \sum_{P} I(P)\mathbf{E}'(P), \tag{3.26}$$

where

$$\mathbf{E}'(P) = \sum_{P'} E^{-1}(P, P')\mathbf{E}(P'). \tag{3.27}$$

We have therefore expressed an arbitrary isotropic tensor $\mathbf{I}$ in terms of its contractions $I(P)$.

Now writing $\mathbf{M}^{\mathfrak{n}}$ in terms of its contractions $M^{\mathfrak{n}}(P)$:

$$\mathbf{M}^{\mathfrak{n}} = \sum_{P} M^{\mathfrak{n}}(P)\mathbf{E}'(P), \tag{3.28}$$

we deduce that

$$\mathbf{M}^{\mathfrak{n}} \cdot \mathbf{S}^{\mathfrak{n}} = \sum_{P, P'} M^{\mathfrak{n}}(P)S^{\mathfrak{n}}(P')E^{-1}(P, P') \tag{3.29}$$

where

$$S^{\mathfrak{n}}(P) = \mathbf{E}(P) \cdot \mathbf{S}^{\mathfrak{n}}. \tag{3.30}$$

One readily shows that

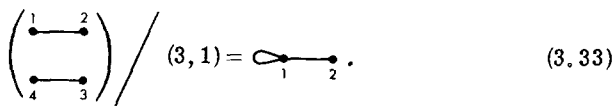
$$M^{\mathfrak{n}}(P) = \int d\mathbf{r}_1 \cdots d\mathbf{r}_v \hat{F}_{v+1}(\mathbf{r}_1, \ldots, \mathbf{r}_v) \prod_{1 \le i < j \le v} (\mathbf{r}_i \cdot \mathbf{r}_j)^{\Gamma_{ij}}$$

$$= m(\Gamma), \quad \text{say,} \tag{3.31}$$

where Γ is a graph obtained from P and $\mathfrak{n} = (n_1, n_2 \cdots n_v)$ by coalescing the first n_1 vertices of P, the next n_2 vertices of P and so on. We formally write

$$\Gamma = P/\mathfrak{n}, \tag{3.32}$$

which is a graph with v vertices, and $b = \frac{1}{2}\sum n_i$ bonds. For example,



(3.33)

With the notation of (2.14) we now have

$$\sum_{\mathfrak{n} \in N_\tau} \mathbf{M}^{\mathfrak{n}} \cdot \mathbf{S}^{\mathfrak{n}} = \sum_{\mathfrak{n} \in N_\tau} \sum_{P, P'} M^{\mathfrak{n}}(P)S^{\mathfrak{n}}(P')E^{-1}(P, P')$$

$$= \sum_{\mathfrak{n}} \sum_{\Gamma, \Gamma'} \sum_{\substack{P: P/\mathfrak{n} = \Gamma \\ P': P'/\mathfrak{n} = \Gamma'}} m(\Gamma)s(\Gamma')E^{-1}(P, P')$$

$$= \sum_{\Gamma, \Gamma'} \alpha(\Gamma, \Gamma')m(\Gamma)s(\Gamma'), \tag{3.34}$$

where

$$s(\Gamma) = S^{\mathfrak{n}}(P) \quad \text{for } \Gamma = P/n$$

$$= \left(\prod_{i=1}^{v} \frac{1}{n_i!}\right) \left[\prod_{1 \le i < j \le v} \left(\frac{\partial}{\partial \mathbf{y}_i} \cdot \frac{\partial}{\partial \mathbf{y}_j}\right)^{\Gamma_{ij}}\right]$$

$$\times \prod_{j=1}^{v} [-\beta\psi(\mathbf{y}_j)]\big|_{\mathbf{y}_i = \mathbf{y}} \tag{3.35}$$

and

$$\alpha(\Gamma, \Gamma') = \sum_{\mathfrak{n}} \sum_{\substack{P: P/\mathfrak{n} = \Gamma \\ P': P'/\mathfrak{n} = \Gamma'}} E^{-1}(P, P'). \tag{3.36}$$

Now (2.14) itself reduces to

$$n(\mathbf{y}/\gamma, \mu, \psi)$$

$$= \rho[\mu - \psi(\mathbf{y})] + \sum_{b=1}^{\infty} \gamma^{2b} \sum_{\substack{\Gamma, \Gamma' \\ \text{with } b \text{ bonds}}} \alpha(\Gamma, \Gamma')m(\Gamma)s(\Gamma')$$

$$= \rho[\mu - \psi(\mathbf{y})] + \sum_{\Gamma, \Gamma'} \alpha(\Gamma, \Gamma')\gamma^{b(\Gamma)}m(\Gamma)\gamma^{b(\Gamma')}s(\Gamma'), \tag{3.37}$$

where $b(\Gamma)$ is the number of bonds in Γ. We note that $\alpha(\Gamma, \Gamma') = 0$ unless Γ and Γ' have the same number of bonds and vertices. As before $m(\Gamma)$ is a function of $\mu - \psi(\mathbf{y})$. It is, of course, a simple matter to include $\rho[\mu - \psi(\mathbf{y})]$ in the summation using suitable definitions.

The graphs in (3.7) are labelled. As an alternative we can use (3.4) and obtain

$$n(\mathbf{y}/\gamma, \mu, \psi) = \rho[\mu - \psi(\mathbf{y})]$$

$$+ \sum_{\substack{\Gamma, \Gamma' \\ \text{unlabelled}}} \gamma^{2b(\Gamma)}K(\Gamma)m(\Gamma)\alpha(\Gamma, \Gamma')K(\Gamma')s(\Gamma'), \tag{3.38}$$

where K is given by (3.5), or equivalently by

$$K(\Gamma) = v!/\prod_i k_i(\Gamma)$$

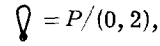
$$= \text{number of inequivalent labellings of } \Gamma, \tag{3.39}$$

where $k_i(\Gamma)$ is the number of vertices of Γ which have i bonds attached. We note that $K(\Gamma) = K(\Gamma')$ if $\alpha(\Gamma, \Gamma') \neq 0$.

Equation (3.37), or (3.38), gives the desired expansion of the density on powers of γ. It involves only scalar moments $m(\Gamma)$ of the modified Ursell correlation functions.

4. INITIAL TERMS OF THE SERIES

The coefficient of γ^2 in the series (3.38) involves graphs Γ and Γ' which have one bond, namely

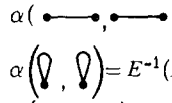


(4.1)

There is only one permutation graph P with 2 vertices, viz.,

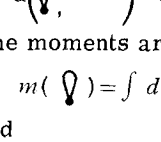
(4.2)

Hence the graphs (4.1) have the unique decompositions



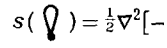
(4.3)

since $\mathfrak{n} = (n_1, n_2) \in N_2'$. Then (3.36) gives



(4.4)

The moments are



(4.5)

and

(4.6)

and the field derivatives, using (2.13) and (3.35), are

(4.7)

and

$$s(\bullet\!\!-\!\!\bullet) = \left|\frac{\partial}{\partial y}\beta\psi(\mathbf{y})\right|^2, \tag{4.8}$$

while, from (3.39)

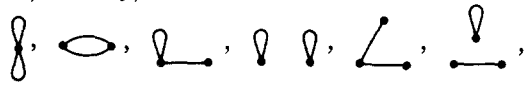
$$K(\text{loop}) = K(\bullet\!\!-\!\!\bullet) = 1. \tag{4.9}$$

The complete coefficient of γ^2 is therefore

$$-(\beta/2\nu)\left(\int d\mathbf{r}\,\hat{F}_2(\mathbf{r})|\mathbf{r}|^2\right)\nabla^2\psi$$

$$+(\beta^2/\nu)\left(\int d\mathbf{r}_1\int d\mathbf{r}_2\,\hat{F}_3(\mathbf{r}_1,\mathbf{r}_2)\mathbf{r}_1\mathbf{r}_2\right)|\partial\psi/\partial\mathbf{y}|^2. \tag{4.10}$$

The coefficient of γ^4 involves all graphs with two bonds, namely,



and

(4.11)

With the definitions (3.19), these have the representations

 $= P_i/(4), \quad i=1,2,3,$

 $= P_i/(2,2), \quad i=2,3,$

 $= P_i/(1,3), \quad i=1,2,3,$

 $= P_1/(2,2),$

 $= P_i/(1,1,2), \quad i=2,3,$

 $= P_1/(1,1,2),$

 $= P_i/(1,1,1,1), \quad i=1,2,3.$ (4.12)

To calculate the α's, we need the inverse of the matrix (3.21),

$$E^{-1}(P_i,P_j) = \frac{1}{\nu(\nu-2)(\nu+2)}\begin{pmatrix} \nu+1 & -1 & -1 \\ -1 & \nu+1 & -1 \\ -1 & -1 & \nu+1 \end{pmatrix}. \tag{4.13}$$

Then we obtain

$$\alpha\left(\text{loop}\;\text{loop}\right) = \sum_{i=1}^{3}\sum_{j=1}^{3} E^{-1}(P_i,P_j) = \frac{3}{\nu(\nu+1)}$$

while

$$\alpha(\text{loop}, \text{loop}) = \sum_{i=2,3}\sum_{j=2,3} E^{-1}(P_i,P_j) = 2/(\nu-1)(\nu+2).$$

In a similar manner we find

$$\alpha(\text{loop}, \text{loop}) = \alpha(\text{loop}) = 3/\nu(\nu+2),$$

$$\alpha(\text{loop}\;\text{loop}, \text{loop}\;\text{loop}) = \alpha(\text{loop}, \text{loop}) = (\nu+1)/\nu(\nu-1)(\nu+2),$$

and

$$\alpha(\text{loop}, \text{loop}) = 2(\nu+1)/\nu(\nu-1)(\nu+2).$$

Since two different graphs make a contribution only if they can be represented via the same $\mathbf{n}$, we have

$$\alpha(\text{loop}, \text{loop}\;\text{loop}) = \sum_{i=2,3} E^{-1}(P_i,P_1)$$

$$= -2/\nu(\nu-1)(\nu+2)$$

$$= \alpha(\text{loop}, \text{loop}),$$

while all other α's are zero. All that remains is to find the $K(\Gamma)$, write down the $m(\Gamma)$ and $s(\Gamma)$, and substitute in (3.38).

5. THE DENSITY PROFILE USING CLASSICAL CORRELATION FUNCTIONS

From Eq. (4.10) we deduce that for a uniform gravitational potential

$$\psi(\mathbf{y}) = \mathbf{k}\cdot\mathbf{y}, \tag{5.1}$$

the density has the expansion

$$n(\mathbf{y}/\gamma,\mu) = \rho(\mu - \mathbf{k}\cdot\mathbf{y}) + [(\beta\gamma|\mathbf{k}|)^2/2\nu]\int d\mathbf{r}_1\int d\mathbf{r}_2$$

$$\times\hat{F}_3(\mathbf{r}_1,\mathbf{r}_2,\mu-\mathbf{k}\cdot\mathbf{y})\mathbf{r}_1\cdot\mathbf{r}_2 + O(\gamma^4). \tag{5.2}$$

For $\hat{F}_3$ we take the generalization of the Ornstein–Zernike correlation function,[7] derived from a mean-field formulation. It has the Fourier transform

$$\mathcal{F}_3(\mathbf{p}_1,\mathbf{p}_2) = \int d\mathbf{r}_1\int d\mathbf{r}_2\exp[2\pi i(\mathbf{p}_1\cdot\mathbf{r}_1+\mathbf{p}_2\cdot\mathbf{r}_2)]\hat{F}_3(\mathbf{r}_1,\mathbf{r}_2)$$

$$= -(a_3^0/\beta^2)[\{a_2^0+\tilde{K}(\mathbf{p}_1)\}\{a_2^0+\tilde{K}(\mathbf{p}_2)\}$$

$$\times\{a_2^0+\tilde{K}(\mathbf{p}_1+\mathbf{p}_2)\}]^{-1} \tag{5.3}$$

where

$$\tilde{K}(\mathbf{p}) = \int d\mathbf{r}\,K(\mathbf{r})\exp(2\pi i\mathbf{p}\cdot\mathbf{r}). \tag{5.4}$$

Here $K(\mathbf{r})$ is the attractive part of the interaction potential and

$$a_n^0 = \partial^n a^0/\partial\rho^n, \tag{5.5}$$

where $a^0(\rho)$ is the free energy per unit volume in the absence of the attractive interaction K, and ρ is the density, both evaluated for the chemical potential $\mu - \mathbf{k}\cdot\mathbf{y}$.

We deduce from (5.3) that

$$\int d\mathbf{r}_1\int d\mathbf{r}_2\,\hat{F}_3(\mathbf{r}_1,\mathbf{r}_2)\mathbf{r}_1\cdot\mathbf{r}_2$$

$$= \frac{-1}{(2\pi)^2}\frac{\partial}{\partial\mathbf{p}_1}\cdot\frac{\partial}{\partial\mathbf{p}_2}\mathcal{F}_3(\mathbf{p}_1,\mathbf{p}_2)\big|_{\mathbf{p}_1=\mathbf{p}_2=0}$$

$$= +m_2a_3^0(a_2^0+\alpha)^{-4}\beta^{-2}, \tag{5.6}$$

where

$$\alpha = \int d\mathbf{r}\,K(\mathbf{r}), \tag{5.7}$$

$$m_2 = \int d\mathbf{r}\,|\mathbf{r}|^2K(\mathbf{r}). \tag{5.8}$$

The second term in (5.2) therefore reduces to

$$+(\gamma|k|)^2/2\nu m_2a_3^0(a_2^0+\alpha)^{-4}. \tag{5.9}$$

When only a single phase is present in the system, it is known[7] that the free energy per unit volume is given by

$$a(\rho) = a^0(\rho) + \tfrac{1}{2}\alpha\rho^2. \tag{5.10}$$

Consequently,

$$a_2^0 + \alpha = \frac{\partial\mu}{\partial\rho} = \left(\frac{\partial\rho}{\partial\mu}\right)^{-1} \tag{5.11}$$

and

$$\begin{aligned} a_3^0 &= \frac{\partial}{\partial\rho}\frac{\partial\mu}{\partial\rho} \\ &= \frac{\partial\mu}{\partial\rho}\cdot\frac{\partial}{\partial\mu}\frac{\partial\mu}{\partial\rho} \\ &= (\rho')^{-1}\frac{\partial}{\partial\mu}(\rho')^{-1} \\ &= -\frac{\partial^2\rho}{\partial\mu^2}\Big/\left(\frac{\partial\rho}{\partial\mu}\right)^3. \end{aligned} \tag{5.12}$$

Consequently, (5.2) reduces to

$$n(\mathbf{y}/\gamma, \mu) = \rho(\mu - \mathbf{k}\cdot\mathbf{y}) - [(\gamma|\mathbf{k}|)^2/2\nu]m_2 \times\rho'(\mu - \mathbf{k}\cdot\mathbf{y})\rho''(\mu - \mathbf{k}\cdot\mathbf{y}) + O(\gamma^4). \tag{5.13}$$

To estimate the magnitude of the correction term, we assume that the van der Waals equation

$$p = \rho kT/(1 - \rho\delta) + \tfrac{1}{2}\alpha\rho^2 \tag{5.14}$$

holds for the fluid. Here δ is the molecular volume and $\alpha < 0$ is given by (5.7). The resulting chemical potential is given by

$$\beta\mu = \log\Lambda^3/\delta + \log[\eta/(1-\eta)] + \eta/(1-\eta) + C\eta \tag{5.15}$$

where

$$C = \alpha\beta/\delta, \tag{5.16}$$

$$\eta = \rho\delta \tag{5.17}$$

and Λ is the thermal wavelength. Differentiating (5.15) gives

$$\rho' = \frac{\eta'}{\delta} = \frac{\beta^2}{\delta}\left(\frac{1}{\eta(1-\eta)^2} + C\right)^{-1} \tag{5.18}$$

and

$$\begin{aligned} \rho'' &= \frac{\eta''}{\delta} \\ &= -\frac{\beta^2}{\delta}\left(\frac{1}{\eta(1-\eta)^2} + C\right)^{-3}\frac{(3\eta-1)}{\eta^2(1-\eta)^3}. \end{aligned} \tag{5.19}$$

For the interaction potential we take a Lennard–Jones (12, 6) potential

$$K(s) = \begin{cases} 0 & \text{for } s < \sigma \\ 4\epsilon[(\sigma/s)^{12} - (\sigma/s)^6] & \text{for } s > \sigma, \end{cases} \tag{5.20}$$

which has a minimum $-\epsilon$ at $s = 2^{1/6}\sigma$. We deduce that in $\nu = 3$ dimensions

$$\alpha = -(32\pi/9)\epsilon\sigma^3 \tag{5.21}$$

and

$$m_2 = -(96\pi/7)\epsilon\sigma^5. \tag{5.22}$$

Substituting (5.18 to 22) in (5.13) and setting $\gamma = 1$ gives

$$n(\mathbf{y}, \mu)\delta = \eta - (16\pi/7)(k^2\beta^3\epsilon\sigma^5/\delta)\phi(\eta) \tag{5.23}$$

where

$$\phi(\eta) = \frac{(3\eta-1)}{\eta^2(1-\eta)^3}\left(\frac{1}{\eta(1-\eta)^2} + C\right)^{-4} \tag{5.24}$$

and η has the argument $\mu - \mathbf{k}\cdot\mathbf{y}$. The van der Waals equation of state has a critical temperature given by

$$\beta_c = 27\delta/4|\alpha| = (3^5/2^7\pi)(\delta/\epsilon\sigma^3). \tag{5.25}$$

If we introduce the dimensionless temperature

$$\theta = \beta_c/\beta, \tag{5.26}$$

then (5.23) reduces to

$$n(\mathbf{y}, \mu)\delta \approx \eta - \frac{3^{15}}{7\cdot 2^{17}\pi^2}\frac{k^2\delta^2}{\epsilon^2\sigma^4}\frac{\phi(\eta)}{\theta^3} \tag{5.27}$$

Taking σ to be the effective molecular diameter gives

$$\delta = \pi\sigma^3/6, \tag{5.28}$$

and (5.27) reduces further to

$$n(\mathbf{y}, \mu)\delta \approx \eta - \frac{3^{14}}{7\cdot 2^{18}}\left(\frac{|\mathbf{k}|\sigma}{\epsilon}\right)^2\frac{\phi(\eta)}{\theta^3}, \tag{5.29}$$

where the numerical coefficient reduces to $2\cdot 61$. In the case of argon we have the values

$$\epsilon = 1\cdot 65\times 10^{-21} \text{ Joule},$$

$$\sigma = 3\cdot 421\times 10^{-10} \text{ meter}.$$

We are primarily interested in the gravitational field where $k = mg$. Here

$$m = 6.7\times 10^{-26} \text{ kgm}$$

is the mass of the argon atom and $g = 9.8$ meter/sec^2 is the acceleration due to gravity. Then we have

$$|\mathbf{k}|\sigma/\epsilon = 1.36\times 10^{-13}, \tag{5.30}$$

which measures the ratio of the gravitation potential energy to the interaction potential energy over the same distance σ.

Consequently, the second term in (5.29) is small except near the critical point where $\phi(\eta)$ becomes arbitrarily large. Suppose θ has its critical value 1 and η is close to its critical value $\frac{1}{3}$,

$$\eta = \tfrac{1}{3} + \lambda,$$

where λ is small. Then to leading order in λ we find that

$$\phi(\eta) \approx (2^{13}/3^{18})\lambda^{-7} \tag{5.31}$$

so that (5.29) becomes

$$n(\mathbf{y}, \mu) \approx \tfrac{1}{3} + \lambda - (1\cdot 36\times 10^{-13}/7\cdot 2^5\cdot 3^4)\lambda^{-7}. \tag{5.32}$$

Consequently, for λ of order 1/100 or smaller, the correction term becomes significant.

For states far removed from the critical condition the correction term is negligible. This indicates that the hydrostatic law (1.1) holds with a high degree of accuracy.

It should be emphasized that the analysis gives only an indication of the departure from the hydrostatic laws since we have not estimated the later terms. We note

however that later terms contain higher powers of the factor $|\mathbf{k}|\sigma/\epsilon$, so that it is reasonable to expect that their contribution will be relatively small.

The sign of the correction term can best be seen from (5.13). We note that $\rho' > 0$ and $m_2 < 0$, while

$$\rho'' \begin{cases} > 0 & \text{for } \rho\delta < \tfrac{1}{3}, \\ < 0 & \text{for } \rho\delta > \tfrac{1}{3}. \end{cases}$$

We conclude that the correction term is positive for $\rho\delta < \frac{1}{3}$ and negative for $\rho\delta > \frac{1}{3}$. Furthermore, both ρ' and ρ'' tend to zero as $|\mu - \mathbf{k}\cdot\mathbf{y}|$ tends to infinity. Consequently, (5.13) indicates a correction to the shape of the interface $\rho(\mu - \mathbf{k}\cdot\mathbf{y})$ which is negligible for distances y from the interface which are much greater than $|\mu_c/mg|$, where μ_c is the chemical potential at the critical point.

[1]M. Kac and C.J. Thompson, k. Nor. Vidensk. Selsk. Forh. 42, 63 (1969).
[2]D.J. Gates, Physica 57, 267—80 (1972).
[3]J.L. Lebowitz and J.K. Percus, J. Math. Phys. 4, 116 (1963).
[4]F.H. Stillinger and F.P. Buff, J. Chem. Phys. 37, 1 (1962).
[5]J.K. Percus, "The Pair Distribution Function," in *The Equilibrium Theory of Classical Fluids*, edited by H.L. Frisch and J.L. Lebowitz (Benjamin, New York, 1964).
[6]G. Stell, "Cluster Expansions for Classical Systems in Equilibrium," in Ref. 5.
[7]D.J. Gates, Physica A 81, 47—71 (1975).
[8]M. Duneau, D. Iagolintzer, and B. Souillard, Commun. Math. Phys. 35, 307—20 (1974).

The class of continuous timelike curves determines the topology of spacetime*

David B. Malament

Department of Philosophy, University of Chicago, Chicago, Illinois 60637
(Received 23 August 1976)

The title assertion is proven, and two corollaries are established. First, the topology of every past and future distinguishing spacetime is determined by its causal structure. Second, in *every* spacetime the path topology of Hawking, King, and McCarthy codes topological, differential, and conformal structure.

1. SUMMARY

Suppose one has two spacetimes (M,g) and (M',g') together with a bijection $f: M \to M'$, where both f and f^{-1} preserve continuous timelike curves; i.e., if $\gamma: I \to M$ is a continuous timelike curve in (M,g), then $f\circ\gamma: I \to M'$ is a continuous timelike curve in (M',g'); and symmetrically for f^{-1}. We show that f must be a homeomorphism. In this sense the class of continuous timelike curves in spacetime determines its topology.

The result is of interest because, at least in some sense, we directly experience whether events on our worldlines are "close" or not. That experience alone, it appears, allows a complete determination of topological structure. The result also has two consequences which are of independent interest.

It is well known that in all strongly causal spacetimes the Alexandroff topology is equal to the manifold topology.[1] Hence, at least in strongly causal spacetimes, if one knows of all points p and q whether it is possible that a particle travel from p to q, then one can recover the topology of spacetime. The question naturally arises whether the condition of strong causality is necessary for this recovery. We show that it is not. The weaker condition of past and future distinguishability suffices. One has the following result: If (M,g) and (M',g') are past and future distinguishing spacetimes and if $f: M \to M'$ is a causal isomorphism (i.e., a bijection where both f and f^{-1} preserve the causal connectibility relation $\ll$), then f must be a homeomorphism. But we also show that the assertion becomes false if the hypothesis of past *and* future distinguishability is relaxed to that of future distinguishability (or past distinguishability) alone.

A second consequence of our theorem is an improvement of a result of Hawking, King, and McCarthy.[2] They define a *path topology* on spacetimes and prove that, in the presence of strong causality, the path topology "codes" (standard) topological, differential, and conformal structure. We show that their hypothesis of strong causality is unnecessary. Indeed their result is true of *all* spacetimes.

2. STANDARD DEFINITIONS AND RESULTS

In what follows a spacetime (M,g) is taken to be a connected, four-dimensional smooth manifold without boundary M, together with a smooth pseudo-Riemannian metric of Lorentz signature g. Spacetimes are assumed to be temporally orientable and endowed with a particular temporal orientation.

Given subsets A and O of M with O open, $I^+(A,O)$ is the set of points q in O such that there exists a future directed smooth timelike curve $\gamma: I \to O$ (where $I \subseteq R$ is connected) and points $t_1, t_2 \in I$ such that $t_1 < t_2$, $\gamma(t_1) \in A$, and $\gamma(t_2) = q$. $I^+(A,O)$ is called the *chronological future of A relative to O*. *The causal future of A relative to O*, $J^+(A,O)$, is the union of $A \cap O$ with the set of points q in O such that there exists a future directed smooth causal curve (i.e., a smooth curve whose tangent vectors are everywhere nonvanishing, nonspacelike, and future directed) $\gamma: I \to O$ and points $t_1, t_2 \in I$ such that $t_1 < t_2$, $\gamma(t_1) \in A$, and $\gamma(t_2) = q$. Finally, the *horismos future of A relative to O*, $E^+(A,O)$, is the set $J^+(A,O) - I^+(A,O)$. These sets have duals $I^-(A,O)$, $J^-(A,O)$, and $E^-(A,O)$ which are defined analogously (substitute past directed curves for future directed curves). $I(A,O)$ is the union $I^+(A,O) \cup I^-(A,O)$. The sets $J(A,O)$ and $E(A,O)$ are defined similarly.

If $A = \{p\}$, we write $I^+(p,O)$ instead of $I^+(A,O)$ and $I^+(p)$ instead of $I^+(p,M)$. Similarly for the other I, J, E sets. The relations $q \in I^+(p,O)$, $q \in J^+(p,O)$, and $q \in E^+(p,O)$ will sometimes be written as $p \ll q(O)$, $p < q(O)$, $p \to q(O)$. Furthermore, $p \ll q(M)$, $p < q(M)$, and $p \to q(M)$ will sometimes be written as $p \ll q$, $p < q$, and $p \to q$.

The I, J, E sets have the following basic properties.[3] If $q \in I^+(p,O)$, then $p \in I^-(q,O)$ and conversely (similarly for the J and E sets). Both $I^+(p,O)$ and $I^-(p,O)$ are open. If $p \ll q(O)$ and $q < r(O)$, then $p \ll r(O)$. Similarly, if $p < q(O)$ and $q \ll r(O)$, then $p \ll r(O)$. If $p \to q(O)$, then, if $\gamma: [0,1] \to 0$ is a future directed smooth causal curve with $\gamma(0) = p$ and $\gamma(1) = q$, γ must be a null geodesic.

An open set O is *convex* iff given any two points p and q in O there is a geodesic $\gamma: [0,1] \to 0$ with $\gamma(0) = p$, $\gamma(1) = q$ and γ is unique (up to reparametrization). If O is an open convex set, then, for all points p in O, $J^+(p,O) = \mathrm{Cl}[I^+(p,O)]$ = the closure in O of $I^+(p,O)$; and $E^+(p,O) = \mathrm{Bnd}[I^+(p,O)]$ = the boundary of $I^+(p,O)$ in O. {These assertions are false in general if O is not convex. But $J^+(p,O) \subseteq \mathrm{Cl}[I^+(p,O)]$ and $E^+(p,O) \subseteq \mathrm{Bnd}[I^+(p,O)]$ are always true.} Dual assertions hold for J^- and E^-. The open convex sets form a basis for the manifold topology; i.e., given any point p and any open set U containing p, there is an open convex set O with $p \in O \subseteq U$.

A set A is *achronal* in O iff for all points p and q in $A \cap O$, it is not the case that $p \ll q(O)$.

A spacetime (M,g) is *chronological* iff it admits no closed, future-directed smooth timelike curves. (M,g) is *causal* iff it admits no closed, future-directed smooth causal curves.

A spacetime (M,g) is *future* (resp. *past*) *distinguishing* iff for all p and q: $I^+(p)=I^+(q) \Rightarrow p=q$ (resp. $I^-(p)=I^-(q) \Rightarrow p=q$). Equivalently, (M,g) is future (resp. past) distinguishing iff for all p in M and all open sets O containing p, there exists an open set O_1 with $p \in O_1 \subseteq O$ such that no future (resp. past) directed smooth timelike curve through p which leaves O_1 ever returns to it.

Finally, a spacetime is *strongly causal* iff, for all points p and all open sets O containing p, there exists an open set O_1 with $p \in O_1 \subseteq O$ such that no future directed smooth timelike curve which leaves O_1 (whether or not if passes through p) ever returns to O_1.

If (M,g) is a spacetime and $O \subseteq M$ is a connected open set, then we may think of $(O, g_{|O})$ as a spacetime in its own right. If O is convex, $(O, g_{|O})$ is necessarily strongly causal.

These "causality conditions" can be ordered in terms of (strictly) increasing strength:

strong causality
$\Downarrow$
future *and* past distinguishability
$\Downarrow$
future (or past) distinguishability
$\Downarrow$
causality
$\Downarrow$
chronology

The respective converse implications are all false.

If (M,g) is a spacetime, the *Alexandroff topology on* M, $\mathcal{T}_A$, is the coarsest topology on M in which all sets $I^+(p)$ and $I^-(q)$ are open. The collection of all sets of form $I^+(p) \cap I^-(q)$ form a basis for $\mathcal{T}_A$. If $\mathcal{T}$ is the (standard) manifold topology on M, then it is always true that $\mathcal{T}_A \subseteq \mathcal{T}$. But the condition $\mathcal{T}_A = \mathcal{T}$ is equivalent to strong causality. Suppose (M,g) is strongly causal. Then the condition that a set $A \subseteq M$ be open (in $\mathcal{T}$) is explicitly definable in terms of the relation $\ll$: A is open iff, for all points p in A, there exist points r and s in A such that $p \in I^+(r) \cap I^-(s) \subseteq A$.

Given two spacetimes (M,g) and (M',g'), a bijection $f: M \to M'$ is a *smooth isometry* iff f and f^{-1} are smooth, and $f_*(g)=g'$. f is a *smooth conformal isometry* iff f and f^{-1} are smooth, and there is a smooth nonvanishing map $\Omega: M' \to R$ such that $f_*(g)=\Omega^2 g'$.

So far "causal structure" has been developed entirely in terms of smooth curves. For our purposes it is essential to work with the larger class of continuous curves. Suppose $\gamma: I \to M$ is a continuous curve. We say that γ is *future directed* and *timelike* iff, for all $t_0 \in I$ and all open convex sets O containing $f(t_0)$, there exists an open (i.e., open in the relative topology on I) subinterval $\bar{I} \subseteq I$ containing t_0 such that

$$\begin{aligned} &t \in \bar{I} \text{ and } t < t_0 \Rightarrow \gamma(t) \ll \gamma(t_0) \quad (O), \\ &t \in \bar{I} \text{ and } t_0 < t \Rightarrow \gamma(t_0) \ll \gamma(t) \quad (O). \end{aligned} \tag{*}$$

We say that γ is *future directed* and *causal* iff the above condition obtains but with $\ll$ replaced by $<$ in (*). Finally, we say that γ is a *future directed null geodesic* iff the above condition obtains but with (*) replaced by

$$t_1, t_2 \in \bar{I} \text{ and } t_1 < t_2 \Rightarrow \gamma(t_1) \to \gamma(t_2) \quad (O).$$

Note that every future directed continuous null geodesic can be reparametrized so as to become a (smooth) future directed null geodesic. (The corresponding assertions for continuous timelike and causal curves are false.) Dual definitions can be given for past directed continuous timelike (causal, null geodesic) curves.

The sets $I^+(A,O)$, $J^+(A,O)$, $E^+(A,O)$ could be redefined in terms of continuous curves, but doing so would not affect the resultant point sets. For example, $p \ll q(O)$ (according to our definition involving smooth timelike curves) iff there is a future directed continuous timelike curve $\gamma: I \to C$ and points $t_1, t_2 \in I$ with $t_1 < t_2$, $\gamma(t_1)=p$, and $\gamma(t_2)=q$.

When there is no chance of confusion we shall not distinguish between curves $\gamma: I \to M$ and their point set images $\gamma[I]$. Also, we shall sometimes refer, simply, to continuous (causal, null geodesic) curves and it should be understood that the curves are *either* future or past directed.

3. FROM TOPOLOGICAL STRUCTURE TO DIFFERENTIAL AND CONFORMAL STRUCTURE

We shall prove that the class of future directed continuous timelike curves determines the topology of spacetime. Having done so, it will follow automatically that this class of curves also determines the differential and conformal structure of spacetime. This is all that one can hope for since all conformally equivalent Lorentz metrics on a manifold induce the same continuous timelike curves.

That differential and conformal structure will follow on the heels of topological structure is a consequence of:

Hawking's theorem[4]: Suppose (M,g) and (M',g') are spacetimes and $f: M \to M'$ is a homeomorphism where both f and f^{-1} preserve future directed continuous null geodesics. Then f is a smooth conformal isometry.

To avail ourselves of this result, we need a simple lemma.

Lemma 1: Suppose (M,g) and (M',g') are spacetimes and $f: M \to M'$ is a homeomorphism where both f and f^{-1} preserve future directed continuous timelike curves. Then both f and f^{-1} preserve future directed continuous null geodesics.

Proof: It suffices to observe that the future directed continuous null geodesics of a spacetime (M,g) can be characterized in terms of its future directed continuous timelike curves and its topology.

First, given any open set U and points p,q in U, we have that $q \in \mathrm{Bnd}[I^+(p,U)]$ iff for all future directed continuous timelike curves $\sigma: (0,1) \to U$, if $\sigma(t_0)=q$ for some t_0 where $0<t_0<1$, then there exist t_1, t_2 where $0<t_1<t_0<t_2<1$ such that $\sigma(t_1) \notin I^+(p,U)$, but $\sigma(t_2) \in I^+(p,U)$.

Next, note that if $\gamma: I \to M$ is a continuous curve, then γ is a future directed null geodesic iff for all $t_0 \in I$ and all open sets O containing $\gamma(t_0)$, there exists an open set $U \subseteq O$ containing $\gamma(t_0)$ such that for all $t_1, t_2 \in I$ with $t_1 < t_2$, if $\gamma(t_1)$, $\gamma(t_2) \in U$ then $\gamma(t_2) \in \mathrm{Bnd}[I^+(\gamma(t_1), U)]$./

4. THE PRINCIPAL RESULT AND ITS CONSEQUENCES

Theorem 1: Suppose (M,g) and (M',g') are spacetimes and $f: M \to M'$ is a bijection where both f and f^{-1} preserve future directed continuous timelike curves. Then f is a homeomorphism. (By Hawking's theorem f must also be a smooth conformal isometry.)

A proof of the theorem is given in the next section.

As it is stated, the hypothesis of the theorem is slightly stronger than necessary. It suffices that f and f^{-1} take (past or future directed) continuous timelike curves to (past or future directed) continuous timelike curves.[5] This follows immediately from the following lemma.

Lemma 2: Suppose (M,g) and (M',g') are spacetimes and $f: M \to M'$ is a bijection. Suppose further that both f and f^{-1} preserve continuous timelike curves. Then either: (a) Both f and f^{-1} preserve future directed continuous timelike curves, or (b) both f and f^{-1} take future directed continuous timelike curves to past directed continuous timelike curves.

Proof: Let p be any point in M. Suppose there are future directed continuous timelike curves γ and σ through p such that $f \circ \gamma$, but not $f \circ \sigma$, is future directed in (M',g'). Let γ^- be the "lower segment" of γ with future end point p. Let σ^+ be the "upper segment" of σ with past end point p. Then the continuous timelike curve which results from "linking" γ^- with σ^+ is one whose image under f is not a continuous timelike curve at all. This is impossible. So at least as restricted to continuous timelike curves through some particular point in M, f either systematically preserves or systematically reverses orientation.

Let A (resp. B) be the set of points in M at which f preserves (resp. reverses) orientation. We show A is open. Suppose p is in A and $p \ll q$ for some point q. Then there is an open set O with $p \in O \subseteq I^-(q)$. Let γ be a future directed continuous timelike curve with initial point p and terminal point q. Suppose now there is a point $r \in O \cap B$. Let σ be any future directed continuous timelike curve with initial point r and terminal point q. Then the result of linking γ with σ is not a continuous timelike curve, but its image under f *is* a continuous timelike curve. This is impossible since f^{-1} preserves continuous timelike curves. Therefore, $O \subseteq A$ and so A is open as claimed. A symmetric argument establishes that B is open.

It thus follows that f either systematically preserves or systematically reverses the orientation of continuous timelike curves. The same argument applies to f^{-1} and, of course, f preserves orientation iff f^{-1} does too./

We consider now the question whether the topological structure of spacetime can be recovered from its causal structure. Rather than thinking of the topological, differential, and conformal structure of spacetime as given and abstracting a causal connectibility relation $\ll$, we ask if the construction can be turned "on its head" with the relation $\ll$ construed as primitive. It turns out that it can be if the spacetime in question is sufficiently well behaved in its causal structure. "Sufficiently well behaved" means "at least past and future distinguishing."

If (M,g) and (M',g') are spacetimes, a map $f: M \to M'$ is a *causal isomorphism* iff f is a bijection and for all points p and q in M: $p \ll q \Leftrightarrow f(p) \ll f(q)$. Our result follows from the following lemma.

Lemma 3: Suppose (M,g) and (M',g') are past and future distinguishing spacetimes and that $f: M \to M'$ is a causal isomorphism. Then f and f^{-1} preserve future directed continuous timelike curves.

Proof: Suppose $\gamma: I \to M$ is an arbitrary future directed continuous timelike curve in (M,g). Suppose $p = \gamma(t_0)$ with $t_0 \in I$, and suppose O' is an arbitrary open convex set containing $f(p)$. We must show that there exists an open subinterval $\bar{I} \subseteq I$ with $t_0 \in \bar{I}$ such that

$$t \in \bar{I} \text{ and } t < t_0 \Rightarrow (f \circ \gamma)(t) \ll f(p) \quad (O'),$$

$$t \in \bar{I} \text{ and } t_0 < t \Rightarrow f(p) \ll (f \circ \gamma)(t) \quad (O'). \qquad (*)$$

Since (M',g') is future distinguishing, there is an open set U' with $f(p) \in U' \subseteq O'$ such that no future directed timelike curve from $f(p)$ which leaves U' ever reenters. Let $f(q)$ be any point in $I^+(f(p), U')$. Since $f(p) \ll f(q)$, we must have $p \ll q$. So there must exist an open convex set O with $p \in O \subseteq I^-(q)$. Since γ is a future directed continuous timelike curve, there must exist an open subinterval $\bar{I}_1 \subseteq I$ with $t_0 \in \bar{I}_1$ such that

$$t \in \bar{I}_1 \text{ and } t_0 < t \Rightarrow p \ll \gamma(t) \quad (O).$$

We claim now that

$$p \ll \gamma(t) \quad (O) \Rightarrow f(p) \ll (f \circ \gamma)(t) \quad (O').$$

For, if $p \ll \gamma(t)(O)$, we have $p \ll \gamma(t) \ll q$. Hence $f(p) \ll (f \circ \gamma)(t) \ll f(q)$. So there exists a future directed smooth timelike curve through $f(p)$, $(f \circ \gamma)(t)$, and $f(q)$ in sequence. We know that this curve cannot leave U' between $f(p)$ and $f(q)$. So we must have $(f \circ \gamma)(t) \in I^+(f(p), U') \subseteq I^+(f(p), O')$.

A parallel argument using past distinguishability of (M',g') establishes that there is an open subinterval $\bar{I}_2 \subseteq I$ with $t_0 \in \bar{I}_2$ such that:

$$t \in \bar{I}_2 \text{ and } t < t_0 \Rightarrow (f \circ \gamma)(t) \ll f(p) \quad (O').$$

Hence the set $\bar{I} = \{t \in \bar{I}_1 / t \geq t_0\} \cup \{t \in \bar{I}_2 / t \leq t_0\}$ is an open subinterval of I with $t_0 \in \bar{I}$ which satisfies (*)./

Thus we have

Theorem 2: Suppose (M,g) and (M',g') are past and future distinguishing spacetimes and $f: M \to M'$ is a causal isomorphism. Then f is a homeomorphism. (By Hawking's theorem f must also be a smooth conformal isometry.)

As was the case with Theorem 1, Theorem 2 can be recast so as to be completely "time symmetric" in formulation.[5] Let τ be the symmetric causal connect-

ibility relation on spacetime points defined by $p\tau q \Leftrightarrow p \ll q$ or $q \ll p$. Given two spacetimes (M,g) and (M',g') a map $f: M \to M'$ is a *symmetric causal isomorphism* iff f is a bijection and for all points p and q in M: $p\tau q \Leftrightarrow f(p)\tau f(q)$. To recast Theorem 2 in symmetric form, it suffices to prove the following lemma and invoke Lemma 2.

Lemma 4: Suppose (M,g) and (M',g') are past and future distinguishing spacetimes and that $f: M \to M'$ is a symmetric causal isomorphism. Then f and f^{-1} preserve continuous timelike curves.

One proves the lemma by compounding the constructions of Lemmas 2 and 3. We skip the argument as it is somewhat tedious and involves no new ideas.

The following example shows that the hypothesis of past and future distinguishability in Theorem 2 (and hence Lemma 3) cannot be relaxed to either future distinguishability or past distinguishability alone. We give the example in a two-dimensional version to simplify matters.

Start with the two-dimensional plane carrying a metric:

$$ds^2 = (\cosh t - 1)^2(dt^2 - dx^2) + dtdx$$

with respect to global Cartesian coordinates t,x. Next form a vertical cylinder by identifying the point $(t,0)$ with all points $(t,2n)$ for all n. Finally excise two closed half-lines: $\{(t,x): x=0 \text{ and } t \geq 0\}$ and $\{(t,x): x=1 \text{ and } t \geq 0\}$ (see Fig. 1.) Along the "equator" $t=0$ the metric reduces to the form $ds^2 = dtdx$ and its associated null cones are horizontal, pointing in the direction of increasing x. But as $|t| \to \infty$, the cones "tip to the left" and asymptotically approach the upright position they have in Minkowski spacetime. Because of the excisions the spacetime is future distinguishing. But it is not past distinguishing. Every point on the $t=0$ equator has for its chronological past the entire region of the spacetime falling below the equator.

Now let f be a bijection of the spacetime onto itself defined by

$$f: (t,x) \to \begin{cases} (t,x) & \text{if } t < 0, \\ (t,x+1) & \text{if } t \geq 0. \end{cases}$$

f leaves the "lower open half" of the spacetime fixed, but reverses the position of the two upper slabs. f is surely discontinuous along the $t=0$ equator; it "cuts" continuous timelike curves which cross the equator.

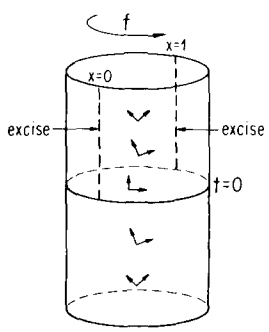


FIG. 1.

But f is a causal isomorphism. The important thing to notice here is that every point below the $t=0$ equator has all points in both upper slabs in its chronological future.

This establishes that the condition in the hypothesis of Theorem 2 cannot be relaxed to future distinguishability. A symmetric example (with excisions below the $t=0$ equator) shows that it cannot be relaxed to past distinguishability either.

Finally, we use Theorem 1 to generalize a result of Hawking, King, and McCarthy.[2] They define the *path topology* on a spacetime to be the finest topology which induces on all continuous timelike curves the same topology induced on them by the standard manifold topology. Equivalently, if (M,g) is a spacetime with $A \subseteq M$, A is open in the path topology on M iff given any continuous timelike curve $\gamma: I \to M$ there exists a (standard) open set O such that $\gamma[I] \cap A = \gamma[I] \cap O$. Their interest in the new topology is motivated in part by the belief that, in some sense, we "experience" continuity along future directed continuous timelike curves. The standard topology, they claim, has no comparable physical significance.

Hawking, King, and McCarthy prove that given any strongly causal spacetime (M,g), if $f: M \to M$ is a homeomorphism with respect to the path topology, then f must be a smooth conformal isometry. But along the way they prove the following:

Lemma 5[2]: If (M,g) is a spacetime and $f: M \to M$ is a homeomorphism with respect to the path topology, then both f and f^{-1} preserve continuous timelike curves.

Thus it follows immediately that we have

Theorem 3: If (M,g) is an arbitrary spacetime and $f: M \to M$ is a homeomorphism with respect to the path topology, then f is a smooth conformal isometry.

One can easily reformulate the theorem so as to be parallel in form to Theorems 1 and 2. One simply takes $f: M \to M'$ to be a path topology homeomorphism between arbitrary spacetimes (M,g) and (M',g'). The conclusion is affected not at all.

5. PROOF OF THEOREM 1

If it were assumed that f preserves *all* continuous curves, it would follow immediately that f is continuous. Given any sequence $\{p_i\}$ converging to p, one could find a continuous curve "threading" all the p_i in sequence and then p. Its image would have to be a continuous curve threading all the $f(p_i)$ in sequence and then $f(p)$. Hence $\{f(p)\}$ would have to converge to $f(p)$. Under our hypotheses, however, this construction can only cope with sequences $\{p_i\}$ which converge chronologically to p. The problem is with those sequences $\{p_i\}$ which converge to p but are locally spacelike related to p.

Our proof is rather long and so is divided into a sequence of lemmas. The crucial idea is this: To show that f is continuous at p, one proves that one may as well assume that f is continuous over a nice-looking

region near p (Lemma E). Then one uses continuous null geodesic segments in that "safe region" to characterize the convergence of points to p. This does the trick because (by Lemma 1 above) continuous null geodesics in the safe region are necessarily preserved by f.

In what follows $\mathcal{D}$ (resp. $\mathcal{D}'$) is taken to be the set of points at which f (resp. f^{-1}) is discontinuous.

Lemma A: If O is an open set in M, O' is an open convex set in M', and $f[O] \subseteq O'$, then $O \subseteq M - \mathcal{D}$.

Proof: Let p be any point in O. To show f is continuous at p, it suffices to show that given any open set U' containing $f(p)$, $f^{-1}[O' \cap U']$ is open in M. Since O' is convex, the spacetime $(O', g_{lO'})$ is strongly causal. So the Alexandroff topology on O' is equal to the relative manifold topology induced in O' Thus $U' \cap O'$ is open in the Alexandroff topology on O'. But $f_{lO} : O \to O'$ is certainly continuous with respect to the Alexandroff topologies on O and O'. So $f^{-1}[U' \cap O']$ must be open in the Alexandroff topology on O. *A fortiori* $f^{-1}[U' \cap O']$ is open in (the manifold topology on) M./

Lemma B: Given p in M, there is an open set O in M containing p such that $I(p, O) \subseteq M - \mathcal{D}$. (So f is at least continuous over "local futures and pasts.")

Proof: Let O' be an open convex set containing $f(p)$. We show first that there is an open set O containing p such that $f[I^+(p, O)] \subseteq O'$.

Suppose there is no such O. Then given any open O_1 containing p there must be a point p_1 in O_1 such that $p_1 \in I^+(p, O_1)$ but $f(p_1) \notin O'$. Since $I^-(p_1, O_1)$ is open, we can find an open set $O_2 \subseteq O_1$ containing p such that $O_2 \subseteq I^-(p_1, O_1)$. There must exist a p_2 in O_2 such that $p_2 \in I^+(p, O_2) \subseteq I^+(p, O_1)$ but $f(p_2) \notin O'$. Clearly $p_2 \ll p_1 (O_1)$. Continuing in this way, we can generate a nested sequence of open sets $O_1 \supseteq O_2 \supseteq O_3 \cdots$ all containing p, and a sequence of points $\{p_i\}$ where, for all i, $p_i \in O_i$, $p_{i+1} \ll p_i(O_i)$, $p \ll p_i(O_i)$, but $f(p_i) \notin O'$ (see Fig. 2). Furthermore, we may choose the $\{O_i\}$ so that they converge to p (i.e., so that their intersection is $\{p\}$). Now we can certainly join p_{i+1} to p_i with a continuous future directed timelike curve segment γ_i contained in O_i. Linking these segments together and adjoining the point p, we obtain a future directed continuous timelike curve γ through p which "threads" all the p_i. By our construction no initial segment of $f \circ \gamma$ can intersect O'. But this is impossible since $f \circ \gamma$ is a continuous timelike curve through $f(p)$.

Therefore, as claimed, there is an open set containing p—call it O_1—such that $f[I^+(p, O_1)] \subseteq O'$. Similarly, there is an open set O_2 such that $f[I^-(p, O_2)] \subseteq O'$. Let $O = O_1 \cap O_2$. Then clearly, $f[I(p, O)] \subseteq O'$. It now follows by Lemma A that $I(p, O) \subseteq M - \mathcal{D}$./

Lemma C: f and f^{-1} preserve continuous causal curves.

Proof: Let $\gamma : I \to M$ be a future directed continuous causal curve in M with $\gamma(t_0) = p$ for some $t_0 \in I$. Let O' be any open convex set containing $f(p)$. We must show that there exists an open subinterval $\bar{I} \subseteq I$ containing t_0 such that:

$$t \in \bar{I} \text{ and } t < t_0 \Rightarrow (f \circ \gamma)(t) < f(p) \quad (O'),$$

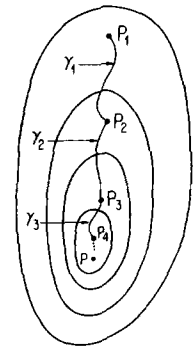


FIG. 2.

$$t \in \bar{I} \text{ and } t_0 < t \Rightarrow f(p) < (f \circ \gamma)(t) \quad (O'). \qquad (*)$$

Just as in the proof of Lemma B we can show that there must exist an open set O in M containing p such that $f[I(p, O)] \subseteq O'$. By moving to a subset we may take O to be convex. We choose $\bar{I} \subseteq I$ containing t_0 so that:

$$t \in \bar{I} \text{ and } t < t_0 \Rightarrow \gamma(t) < p \quad (O),$$

$$t \in \bar{I} \text{ and } t_0 < t \Rightarrow p < \gamma(t) \quad (O).$$

Now if $\gamma(t) < p$, then every continuous timelike curve segment through $\gamma(t)$ intersects $I^-(p, O)$. Hence every continuous timelike curve segment through $(f \circ \gamma)(t)$ intersects $I^-(f(p), O')$. Thus $(f \circ \gamma)(t) \in \mathrm{Cl}[I^-(f(p), O')]$ and therefore, since O' is convex, $(f \circ \gamma)(t) \in J^-(f(p), O')$. Thus the first half of (*) is established. The second half is symmetric. Hence $f \circ \gamma$ is a future directed continuous causal curve. (The argument for f^{-1} is, of course, symmetric.)/

Lemma D: (i) $\mathcal{D}$ is closed in M; $\mathcal{D}'$ is closed in M'.

(ii) For all $p \in M$, $p \in \mathcal{D}$ iff $f(p) \in \mathcal{D}'$.

(iii) If $p \in \mathcal{D}$, then there is an inextendible future directed continuous causal curve through p fully contained in $\mathcal{D}$.

Proof: Suppose f is continuous at p. Let O' be any open convex set containing $f(p)$. Let O be an open set with $p \in O \subseteq f^{-1}[O']$. Then, applying Lemma A, we have that $O \subseteq M - \mathcal{D}$. Thus $M - \mathcal{D}$ is open. Similarly $M' - \mathcal{D}'$ is open. So (i).

Suppose p is in $\mathcal{D}$. Then there exists a sequence $\{p_i\}$ which converges to p and an open convex set O' in M' which contains $f(p)$ but none of the $f(p_i)$. We can find sequences $\{r_i\}$ and $\{s_i\}$ converging chronologically to p from below and above respectively such that for each i there is a *local* future directed continuous timelike curve γ_i through p_i with initial point r_i and terminal point s_i. The only accumulation point of the γ_i is p.

Now $\{f(r_i)\}$ and $\{f(s_i)\}$ must converge to $f(p)$. So (passing to a subsequence if necessary) we may assume that all $f \circ \gamma_i$ begin and end in O'. But since $f(p_i) \notin O'$, each of these curves $f \circ \gamma_i$ must leave O' as well. There will be a future directed inextendible continuous causal curve Δ through $f(p)$ every point of which is an accumulation point of the $f \circ \gamma_i$.[6] Since the only accumulation point of the γ_i is p, it must be the case that $\Delta - \{f(p)\} \subseteq \mathcal{D}'$. Since $\mathcal{D}'$ is closed, it follows that $\Delta \subseteq \mathcal{D}'$. Thus $p \in \mathcal{D} \Rightarrow f(p) \in \mathcal{D}'$. The converse is symmetric. So we have (ii). For (iii) we need only repeat this past argument with respect to $f(p)$ and f^{-1}./

Lemma E: If $\mathcal{D} \neq \emptyset$, then there exists an open convex set O with $\mathcal{D} \cap O \neq \emptyset$ such that:

(i) $\mathcal{D}$ is achronal in O.

(ii) Through each point p in $\mathcal{D} \cap O$ there passes a unique continuous null geodesic Γ_p such that $\Gamma_p \cap O \subseteq \mathcal{D}$.

(iii) Given any continuous null geodesic Γ which intersects $\mathcal{D} \cap O$, either $\Gamma \cap O \subseteq \mathcal{D}$ or $\Gamma \cap O \cap \mathcal{D}$ is a singleton.

Proof: First note that (ii) and (iii) follows from (i) in view of Lemma D. For (i) suppose $\mathcal{D} \neq \emptyset$ but no O exists satisfying the required conditions. Let O_1 be any open convex set meeting $\mathcal{D}$ with compact closure. By our assumption we can find points r_1 and s_2 in $O_1 \cap \mathcal{D}$ such that $r_1 \ll s_1(O_1)$. Now let O_2 be any open convex set where $r_1 \in O_2 \subseteq I^-(s_1, O_1)$. Repeating the argument with respect to O_2, we can find points r_2 and s_2 in $O_2 \cap \mathcal{D}$ such that $r_2 \ll s_2(O_2)$. Certainly $s_2 \ll s_1(O_1)$. Continuing in this fashion, we generate a sequence $\{s_i\}$ in $O_1 \cap \mathcal{D}$ with $s_{i+1} \ll s_i(O_1)$ for all i. This sequence must have an accumulation point s. But now if we apply Lemma B to s, we find that there must exist an open set O containing s such that $I^+(s, O) \subseteq M - \mathcal{D}$. This leads to a contradiction since eventually all the s_i must enter $I^+(s, O_1)$./

Proof of the Theorem: Suppose $\mathcal{D} \neq \emptyset$ and O is as in Lemma E. Let p be any point in $\mathcal{D} \cap O$ with corresponding Γ_p. Clearly $I(\Gamma_p \cap O, O) \subseteq M - \mathcal{D}$. There must exist a sequence $\{p_i\}$ converging to p and an open convex set O' containing $f(p)$ but none of the $f(p_i)$.

Let Ω be *any* future directed continuous null geodesic segment through p distinct from Γ_p which is sufficiently "short" that $f \circ \Omega$ is fully contained in O'. There exist continuous null geodesic segments Ω_i within O, passing through p_i respectively, which converge to Ω in the sense that every open set which intersects Ω intersects eventually all Ω_i. We may choose $\{\Omega_i\}$ so that it has no convergence points off Ω. Eventually all Ω_i enter $I(\Gamma_p \cap O, O)$ and hence $M - \mathcal{D}$. It follows from Lemma E (iii) that, for eventually all i, $\Omega_i \cap \mathcal{D}$ is either empty or a singleton. The intersection point of Ω_i with $\mathcal{D}$ (if there is one) comes either "before p_i," at p_i itself, or "after p_i." Without loss of generality we may assume that there is an infinite subset of $\{\Omega_i\}$ in each member of which the intersection point with $\mathcal{D}$ (if there is one) does not come before p_i. Now let Ω_i^- be the "lower-half" of Ω_i with future end point p_i included. By moving to a subsequence we can thus find a sequence of continuous null geodesic segments $\{\Omega_i^-\}$ in O with the following properties (see Fig. 3):

(i) $\{\Omega_i^-\}$ converges to the lower half Ω^- of Ω, but has no convergence points off Ω^-.

(ii) For each i, $\Omega_i^- \cap \mathcal{D} \subseteq \{p_i\}$.

From (ii), Lemma C, and Lemma 1, it follows that each image curve $f \circ \Omega_i^-$ is a continuous null geodesic segment in M'. From (i) and the fact that $\Omega^- - \{p\} \subseteq M - \mathcal{D}$,

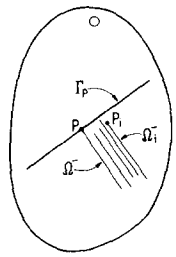


FIG. 3.

it follows that these segments converge to $f \circ \Omega^-$.

Now recall that no point $f(p_i)$ lies within O'. So, though the $f \circ \Omega_i^-$ converge to $f \circ \Omega^-$, they must all leave O' before reaching their respective $f(p_i)$. Let $f(q)$ be any point of the null geodesic extension of $f \circ \Omega^-$. We claim $f(q) \in \mathcal{D}'$. For suppose to the contrary that $f(q) \in M - \mathcal{D}'$. Then, since $f(q)$ is a convergence point of $\{f \circ \Omega_i^-\}$, q must be a convergence point of $\{\Omega_i^-\}$. This is impossible since $q \notin \Omega^-$.

In our construction we assumed that Ω satisfied the "not before p_i" clause for an infinite subset of Ω_i. Dropping that assumption, we have the following conclusion. If Ω^- and Ω^+ are the respective lower and upper segments of Ω, then *either* the future null geodesic extension of $f \circ \Omega^-$ or the past null geodesic extension of $f \circ \Omega^+$ is a future directed continuous causal curve segment through $f(p)$ lying within $\mathcal{D}'$. But this is true of *all* future directed continuous null geodesic segments; Ω was chosen arbitrarily. Thus, since f is a bijection, it follows that there exist *distinct* future directed continuous causal curves through $f(p)$ lying within $\mathcal{D}'$. Their preimages under f^{-1} must be distinct future directed continuous causal curves through p lying within $\mathcal{D}$. But this contradicts our assumption that $\mathcal{D}$ is achronal in O.

Thus, $\mathcal{D}$ is empty, and, hence, $\mathcal{D}'$ is empty as well./

ACKNOWLEDGMENTS

I wish to thank Professor Robert Geroch for several suggestive discussions, and Professor Howard Stein for catching a number of errors in an earlier version of this paper.

[1] E.H. Kronheimer and R. Penrose, Proc. Camb. Phil. Soc. 63, 481 (1967).

[2] S.W. Hawking, A.R. King, and P.J. McCarthy, J. Math. Phys. 17, 174 (1976).

[3] Proofs of these and subsequent claims can be found in S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, 1973); R. Penrose, *Techniques of Differential Topology in Relativity* (SIAM, Philadelphia, 1972).

[4] A proof is given in Hawking, King, and McCarthy (Ref. 2). The theorem is not formulated in exactly this from, but the argument carries over intact.

[5] This version of the theorem is applicable to all temporally orientable spacetimes whether or not a particular temporal orientation is distinguished.

[6] Hawking and Ellis (Ref. 3) prove this in detail in their Lemma 6.2.1.

Equivalence transformations for nonlinear evolution equations*

Franco Magri

Istituto di Matematica del Politecnico, Piazza Leonardo da Vinci, 32-20133 Milano, Italy
(Received 9 August 1976)

A systematic approach to the study of nonlinear evolution equations based on the theory of the equivalence transformations is suggested. In this paper it is applied to the Burgers and to the Korteweg–de Vries equations. The main result is that the Hopf–Cole transformation for the Burgers equation and the Miura, Bäcklund, and Hirota transformations for the Korteweg–de Vries equation (together with the linear equations of the inverse scattering theory) are all deduced from a single general equivalence condition.

INTRODUCTION

The aim of this paper is to suggest a *systematic approach to the study of nonlinear evolution equations based on the theory of the equivalence transformations*. This method, well known in the theory of classical dynamical systems,[1] seems not only conceptually simple but also effective in the applications. In this paper it is applied, in particular, to the Burgers and to the Korteweg–de Vries equation (hereafter abbreviated as KdV equation).

In Sec. 1, we present an *operator formulation* of the theory of the equivalence transformations for evolution equations. The use of the methods of the nonlinear functional analysis allows one to unify all the equivalence conditions for particular equations into the *single operator equivalence condition* (1.6). This condition holds both for linear and nonlinear equivalence transformations and is the starting point of a systematic analysis of these transformations. In particular, we show how the equivalence transformations may be used in the study of the symmetry transformations, of the conservation laws, and of the initial value problem for a given evolution equation. For the convenience of the reader, the few notions of the nonlinear operator theory which are needed in this section are summarized in the Appendix.

Section 2 deals with the study of the equivalence transformations of the Burgers equation (2.1). As the simplest example, we find the well-known Hopf–Cole transformation which reduces the Burgers equation to the linear diffusion equation. This result gives a new interpretation of this transformation and displays the constructive character of the procedure suggested in Sec. 1.

In the following two sections we search for equivalence transformations of the KdV equation (3.1). This approach, which differs from that followed in the original papers, leads in a way which seems more systematic to the other equations (such as the modified KdV equation and the linear equations of the inverse scattering theory) usually associated with the KdV equation. As a particular example of equivalence transformation we obtain the so-called generalized Miura transformation (3.13).

In the last section, we show how the Bäcklund transformation for the KdV equation can be viewed from the standpoint of the equivalence transformations previously found.

The study of the conservation laws of the evolution equations and of their link with symmetry transformations shall be dealt with in a subsequent paper.

1. THE METHOD OF THE EQUIVALENCE TRANSFORMATIONS

In this section we formulate the theory of the equivalence transformations for a system of nonlinear evolution equations like $(A, B = 1, 2, \ldots, n)$

$$\partial_t u^A(x,t) = \kappa^A(u^B, u^B_j, u^B_{je}, \cdots), \tag{1.1}$$

where the field functions $u^A(x,t)$ are supposed to be defined, at any instant of time, in a fixed region Ω of $\mathbb{R}^3$ and the subscripts denote the partial derivatives of these functions with respect to the space coordinates x^j.

Dealing with evolution equations it is useful to consider the time t as a parameter and to emphasize the dependence of the field functions on the time. Consequently, we shall simply denote by $u(t_0) = (u^1(x,t_0), \ldots, u^n(x,t_0))$ the n-tuple of the field functions evaluated at the time instant t_0, regarded as functions of the space coordinates only.

Now, in order to develop the theory of the equivalence transformations from a general and unified standpoint, we introduce the following *operator formulation*. We consider the linear function space U of the field functions regarded as functions of the space coordinates only. Any n-tuple $u(t_0)$ will then be referred to as a point of this space, and the function $u = u(t)$ will be regarded as defining a line in U (Fig. 1). The functions $\kappa^A(u^B, u^B_j, u^B_{ji}, \cdots)$ define a nonlinear formal differential operator with domain and range in the space U, which we shall denote by $\mathbf{K}$. The given evolution equations may then be synthesized into the single operator equation

$$\partial_t u = \mathbf{K}(u). \tag{1.2}$$

This equation is called an *abstract evolution equation*. Its solutions, if any, may be regarded as the field lines of the abstract vector field defined by the operator $\mathbf{K}$.[2] This geometric standpoint allows to introduce in a conceptually simple way the notion of equivalence transformations, as follows.

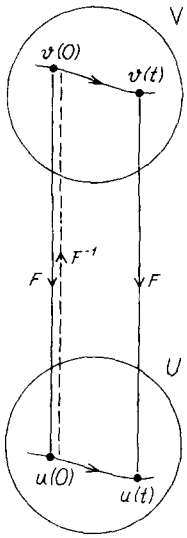


FIG. 1.

A. Equivalent equations

Consider a second evolution equation

$$\partial_t v = \mathbf{H}(v) \tag{1.3}$$

defined on a second linear function space V and let $\mathbf{F}$ be a generally nonlinear transformation which relates the spaces U and V. We write

$$u = \mathbf{F}(v) \tag{1.4}$$

and we suppose that this transformation admits an inverse mapping and is Gateaux differentiable (see the Appendix).

Equation (1.3) is said to be *equivalent* to the given Eq. (1.2) under the transformation (1.4), if *every solution $v = v(t)$ of Eq. (1.3) is mapped into a solution $u = u(t)$ of Eq. (1.2)* (Fig. 1).[3]

B. Condition of equivalence

If the operator $\mathbf{F}$ does not depend explicitly on the time t (as we suppose henceforth, for simplicity), we get

$$\begin{aligned} \partial_t u - \mathbf{K}(u) &\overset{(1.4)}{=} \partial_t \mathbf{F}(v) - \mathbf{K}(\mathbf{F}(v)), \\ &\overset{(A4)}{=} \mathbf{F}'_v \partial_t v - \mathbf{K}(\mathbf{F}(v)), \\ &\overset{(1.3)}{=} \mathbf{F}'_v \mathbf{H}(v) - \mathbf{K}(\mathbf{F}(v)), \end{aligned} \tag{1.5}$$

where $\mathbf{F}'_v$ is the linear Gateaux derivative of the operator $\mathbf{F}$. Hence in order that $u = F(v)$ be a solution of the evolution equation (1.2) it must be

$$\mathbf{F}'_v \mathbf{H}(v) - \mathbf{K}(\mathbf{F}(v)) \overset{\circ}{=} \theta_U \tag{1.6}$$

for every solution v of Eq. (1.3). Here θ_U denotes the null element of the space U, and we have used the symbol $\overset{\circ}{=}$ to mean that the equality must hold only for the solutions of Eq. (1.3). Obviously this condition is sufficient as well. It is the *operator equivalence condition* we were looking for.

If we know both the evolution equations (1.2) and (1.3), we can use this condition to verify if they are equivalent under a mapping (1.4). On the other hand, if we know only the evolution equation (1.2), we can use this condition to produce both the equivalent equation (1.3) and the equivalence transformation (1.4). It suffices to look for a pair of operators $\mathbf{H}: V \to V$ and $\mathbf{F}: V \to U$ such that condition (1.6) be identically verified (the condition being in this way *a fortiori* verified on the manifold of the solutions). In the next three sections we shall give examples of this procedure, considering in particular the Burgers and the KdV equations.

C. Symmetry properties

Let us assume a symmetry transformation of Eq. (1.3), i.e., a mapping $\mathbf{S}: V \to V$ which maps solutions of this equation once more into solutions.[4] Then, any equivalence transformation $\mathbf{F}: V \to U$ induces in a natural way a symmetry mapping $\mathbf{P}: U \to U$ of the given equation (1.2), according to the scheme shown in Fig. 2. Namely, a *symmetry mapping of the given equation corresponds to every symmetry mapping of the equivalent equation.* In Sec. 5 we obtain in this way the Bäcklund transformation for the KdV equation.

D. Local conservation laws

We shall deal with the relation between the conservation laws of two equivalent equations in a subsequent paper.[5] For completeness, here we limit ourselves to stating the following result. Let us denote in operator form,

$$\mathrm{R}(u) = \rho(u^B, u^B_j, \cdots), \tag{1.7}$$

a conserved density[6] for the given evolution equation (1.2). Then $\mathbf{R}(\mathbf{F}(v))$ is a conserved density for Eq. (1.3). Namely, a *conservation law of the equivalent equation is associated with every conservation law of the given evolution equation.*

E. Initial value problem

The problem consists in finding the solution $u = u(t)$ corresponding to an initial condition $u(0) = u_0$. If we know an equivalent evolution equation which is simpler

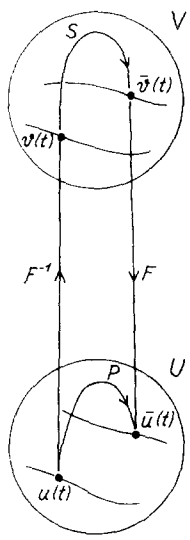


FIG. 2.

to solve than the given equation, we can think of solving this problem according to the following scheme (see Fig. 1):

1. to obtain the initial condition v_0 for the equivalent equation, by means of the *inverse* equivalence transformation;

2. to solve the initial value problem for the equivalent equation;

3. to obtain the solution $u(t)$ of the given initial value problem, by means of the *direct* equivalence transformation.

In this way we replace the study of the given equation by the study of the pair of equations

$$\mathbf{F}(v)=u, \quad \partial_t v=\mathbf{H}(v). \tag{1.8}$$

In Sec. 4, looking for the simplest equivalent equations of the KdV equation, we are lead in this way to the pair of linear equations of the inverse scattering problem associated with this equation.[7,8]

2. THE BURGERS EQUATION

As an introduction to the study of the KdV equation, in this section we apply the method of the equivalence transformations to the Burgers equation (Ref. 7, p. 96),

$$u_t+uu_x-au_{xx}=0. \tag{2.1}$$

The main result is that this method leads, in a very natural way, to the well-known Hopf–Cole transformation (Ref. 7, p. 97), which allows us to linearize the Burgers equation.

We look for operators H and F of the form

$$\mathbf{H}(v)=h(v,v_x,v_{xx}), \tag{2.2}$$

$$\mathbf{F}(v)=f(v,v_x,v_{xx},\cdots), \tag{2.3}$$

where h and f are two functions of $v(x,t)$ and its space derivatives which are to be determined so as to verify the equivalence condition (1.6). The form of the operator K,

$$\mathbf{K}(u)=\partial_x(au_x-\tfrac{1}{2}u^2) \tag{2.4}$$

suggests that we choose in particular a function f of the kind

$$f(v,v_x,\cdots)=\partial_x g(v,v_x,\cdots). \tag{2.5}$$

As a first attempt, we try a function g which depends only on v,

$$\mathbf{F}(v)=\partial_x g(v). \tag{2.6}$$

In this case, the condition of equivalence (1.6) becomes

$$\begin{aligned}&\mathbf{K}(\mathbf{F}(v))-\mathbf{F}'_v\mathbf{H}(v)\\&=\partial_x[ag''(v)v_x^2+ag'(v)v_{xx}-\tfrac{1}{2}g'^2(v)v_x^2-g'(v)h(v,v_x,v_{xx})]\\&=0\end{aligned} \tag{2.7}$$

[where $g'(v)$ is dg/dv]. The simplest way of verifying this condition is to choose

$$h(v,v_x,v_{xx})=av_{xx} \tag{2.8}$$

and $g(v)$ such that

$$ag''(v)-\tfrac{1}{2}g'^2(v)=0. \tag{2.9}$$

This condition yields

$$g(v)=-2a\ln(v+A). \tag{2.10}$$

In this way we obtain the equivalence transformation

$$u=-2a\partial_x\ln(v+A). \tag{2.11}$$

It is the Hopf–Cole transformation. It reduces the Burgers equation to the linear diffusion equation

$$v_t=av_{xx}, \tag{2.12}$$

a well-known result.

Then the *Hopf–Cole transformation appears as a particular example of equivalence transformation* of the Burgers equation.

3. THE KORTEWEG-DE VRIES EQUATION

In the following two sections we study the equivalence transformations of the KdV equation (Ref. 7, p. 577),

$$u_t+auu_x+u_{xxx}=0. \tag{3.1}$$

The main result is that in this way we are lead quite directly to the other evolution equations which have been related, in the original papers, to the KdV equation.[9] In particular, in this section we find a three-parameter family of equivalence transformations including as a particular case the so-called *generalized Miura transformation* (Ref. 7, p. 600). They relate the KdV equation to a three-parameter family of equivalent equations, which includes as a particular case the modified KdV equation. In the next section we shall obtain the *linear equations* of the inverse scattering theory.

According to the method of the equivalence transformations we look for operators H and F of the form

$$\mathbf{H}(v)=h(v,v_x,v_{xx},v_{xxx}), \tag{3.2}$$

$$\mathbf{F}(v)=f(v,v_x), \tag{3.3}$$

which verify the equivalence condition

$$\begin{aligned}&\mathbf{F}'_v\mathbf{H}(v)-\mathbf{K}(\mathbf{F}(v))\\&=\frac{\partial f}{\partial v}h(v\cdots v_{xxx})+\frac{\partial f}{\partial v_x}\left(\frac{\partial h}{\partial v}v_x+\frac{\partial h}{\partial v_x}v_{xx}+\frac{\partial h}{\partial v_{xx}}v_{xxx}\right.\\&\left.+\frac{\partial h}{\partial v_{xxx}}v_{xxxx}\right)+af(vv_x)\left(\frac{\partial f}{\partial v}v_x+\frac{\partial f}{\partial v_x}v_{xx}\right)+\left(\frac{\partial^3 f}{\partial v^3}v_x^3\right.\\&+3\frac{\partial^2 f}{\partial v^2\,\partial v_x}v_x^2v_{xx}+3\frac{\partial^3 f}{\partial v\,\partial v_x^2}v_xv_{xx}^2+\frac{\partial^3 f}{\partial v_x^3}v_{xx}^3+3\frac{\partial^2 f}{\partial v^2}v_xv_{xx}\\&\left.+3\frac{\partial^2 f}{\partial v\,\partial v_x}v_{xx}^2\right)+v_{xxx}\left(3\frac{\partial^2 f}{\partial v\,\partial v_x}v_x+3\frac{\partial^2 f}{\partial v_x^2}v_{xx}+\frac{\partial f}{\partial v}\right)\\&+v_{xxxx}\frac{\partial f}{\partial v_x}\\&=0.\end{aligned} \tag{3.4}$$

From it we can obtain a set of explicit conditions on the unknown functions h and f as follows.

Firstly, we must impose that the coefficient of v_{xxxx} vanishes. This condition yields

$$h(v,v_x,v_{xx},v_{xxx})=-v_{xxx}+l(v,v_x,v_{xx}). \tag{3.5}$$

Now, we insert this expression into (3.4) and we impose the condition that the coefficient of v_{xxx} vanishes. In this way we obtain the condition

$$\left(-\frac{\partial f}{\partial v}+\frac{\partial f}{\partial v_x}\cdot\frac{\partial l}{\partial v_{xx}}\right)+\left(3\frac{\partial^2 f}{\partial v\,\partial v_x}v_x+3\frac{\partial^2 f}{\partial v_x^2}v_{xx}+\frac{\partial f}{\partial v}\right)=0. \tag{3.6}$$

In order to keep the expression of $h(vv_xv_{xx}v_{xxx})$ as simple as possible, we attempt a solution in the form

$$l=l(v,v_x). \tag{3.7}$$

Then the condition (3.6) implies

$$f(v,v_x)=Av_x+B(v). \tag{3.8}$$

Again, we insert this expression in the equivalence condition (3.4) and we impose the condition that the coefficient of v_{xx} vanishes. We obtain the condition

$$A\frac{\partial l}{\partial v_x}+aA(Av_x+B(v))+3\frac{d^2B(v)}{dv^2}v_x=0 \tag{3.9}$$

which yields

$$l(v,v_x)=-\frac{1}{2}\left(aA+\frac{3}{A}\frac{d^2B}{dv^2}\right)v_x^2-aB(v)v_x. \tag{3.10}$$

Lastly, we insert this expression in (3.4) and we impose the conditions that the coefficients of v_{xx}^3, v_{xx}^2, and v_x separately vanish. These conditions yield

$$B(v)=-\tfrac{1}{6}aA^2v^2+Bv+C, \tag{3.11}$$

or

$$B(v)=D, \tag{3.12}$$

where A, B, C, and D are arbitrary constants.

In this way, we have obtained the following two classes of equivalence transformations of the KdV equation:

$$u=Av_x-\tfrac{1}{6}aA^2v^2+Bv+C \tag{3.13}$$

and

$$u=Aw_x+D \tag{3.14}$$

corresponding to the choice (3.11) and (3.12), respectively.

In (3.14) we used w instead of v in order to keep the two classes quite distinct. The associated equivalent equations are given by

$$v_t+v_{xxx}-\tfrac{1}{6}a^2A^2v^2v_x+aBvv_x+aCv_x=0 \tag{3.15}$$

and by

$$w_t+w_{xxx}+\tfrac{1}{2}aAw_x^2+aDw_x=0. \tag{3.16}$$

The transformation (3.13), with $A=1$ and $B=C=0$ is the *Miura transformation* (Ref. 7, p. 599); with $A=i\epsilon/a$ and $B=1$, $C=0$ is the *generalized Miura transformation* (Ref. 7, p. 600). Then these transformations appear as *particular examples of equivalence transformations* of the KdV equation. The transformation (3.13), with $B=0$, relates the KdV equation to the so-called *modified KdV equation*,

$$v_t+v_{xxx}-\tfrac{1}{6}a^2A^2v^2v_x+aCv_x=0. \tag{3.17}$$

This equation is then a particular example of an equivalent equation to the KdV equation. It is of interest to note that Miura found his transformation by noticing a correspondence between the conservation laws of the KdV and of the modified KdV equations, rather than a correspondence between their solutions (Ref. 6, p. 1207). The general relation between the conservation laws of every pair of equivalent equations, illustrated in Sec. 1, gives the reason for this fact.

4. THE MODIFIED KORTEWEG-DE VRIES EQUATION

A way of pursuing the study of the equivalence transformations is to search for *higher-order transformations*, i.e., for transformations which depend on higher-order derivatives than the first order, to which we have restricted ourselves in the previous section. An alternative (and more convenient) procedure is to search for first-order transformations of the equivalent equations already obtained. In fact, combining in sequence two first-order transformations we obtain a second-order transformation, and so forth.

According to this point of view, in this section we study the first-order transformations of the modified KdV equation (3.17). From them we shall deduce second-order equivalence transformations of the KdV equation (see Fig. 3).

The modified KdV equation is characterized by the nonlinear operator

$$\mathbf{H}(v)=\partial_x\left(\frac{a^2A^2}{18}v^3-aCv-v_{xx}\right). \tag{4.1}$$

In analogy with the Burgers equation, we search for an equivalence transformation of the kind

$$v=\partial_x g(\psi)=\mathbf{G}(\psi). \tag{4.2}$$

Let

$$\partial_t\psi=n(\psi,\psi_x,\psi_{xx},\psi_{xxx})=\mathbf{N}(\psi) \tag{4.3}$$

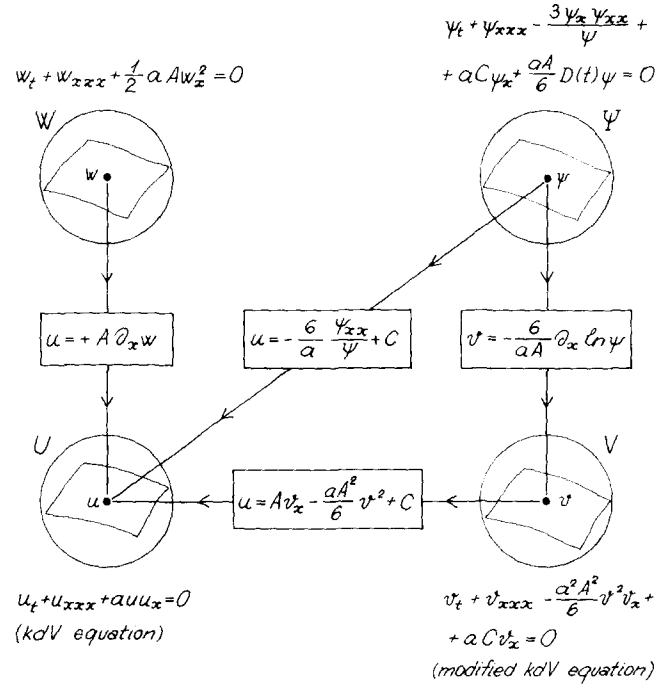


FIG. 3.

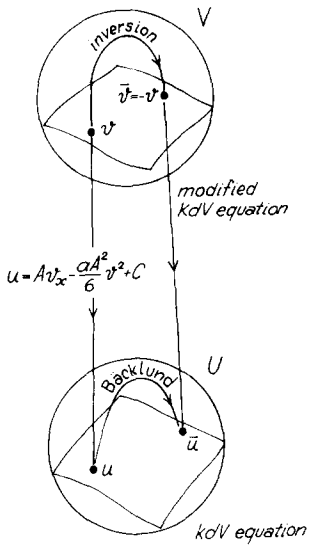


FIG. 4.

be the corresponding equivalent equation. The equivalence condition (1.6) becomes

$$G'_{\psi}N(\psi)-H(G(\psi))$$
$$=\partial_x\Big[g'(\psi)n(\psi\cdots\psi_{xxx})+\Big(g'''-\frac{a^2A^2}{18}g'^3\Big)\psi_x^3$$
$$+aCg'\psi_x+3g''\psi_x\psi_{xx}+g'\psi_{xxx}\Big]=0. \tag{4.4}$$

The simplest way of verifying this condition is to choose $g(\psi)$ and $n(\psi\cdots\psi_{xxx})$ such that

$$g'''-\frac{a^2A^2}{18}g'^3=0 \tag{4.5}$$

and

$$g'(\psi)[n(\psi\cdots\psi_{xxx})+\psi_{xxx}+aC\psi_x]+3g''(\psi)\psi_x\psi_{xx}=D(t) \tag{4.6}$$

where $D(t)$ is an arbitrary function of the time. The first condition yields

$$g(\psi)=-\frac{6}{aA}\ln\psi \tag{4.7}$$

and then, from (4.6) we obtain

$$n(\psi\cdots\psi_{xxx})=3\frac{\psi_x\psi_{xx}}{\psi}-\psi_{xxx}-aC\psi_x-\frac{aA}{6}D(t)\psi. \tag{4.8}$$

The transformation (4.7) is once more *the Hopf–Cole transformation.* It *is an equivalence transformation also for the modified KdV equation.* The associated equivalent equation is

$$\psi_t+\psi_{xxx}+aC\psi_x-3\frac{\psi_x\psi_{xx}}{\psi}+\frac{aA}{6}D(t)\psi=0. \tag{4.9}$$

Combining in sequence the equivalence transformations $u=\mathbf{F}(v)$ and $v=\mathbf{G}(\psi)$ given by (3.13) with $B=0$ and by (4.2), we obtain

$$u=-\frac{6}{a}\frac{\psi_{xx}}{\psi}+C. \tag{4.10}$$

This is the second-order equivalence transformation which directly relates the KdV equation to the equivalent equation (4.9).

For the convenience of the reader, we summed up the main classes of equivalence transformations of the KdV equation previously found in Fig. 3. In it we also summarized the main equivalent equations which have been related to this equation.

In Fig. 3 we have symbolically represented the manifold of the solutions of the various equations as curved surfaces in the corresponding linear function spaces.

Now, as we have seen in Sec. 1, we can replace the study of the initial value problem for the KdV equation by the study of the pair of equations formed by any equivalence transformation and by the associated equivalent equation. A comparison of the equivalence transformations in Fig. 3 shows that it is convenient to choose the pair of equations (4.9) and (4.10). In fact, they also may be written in the form

$$\psi_{xx}+\frac{a}{6}(u-C)\psi=0, \tag{4.11}$$

$$\psi_t+\psi_{xxx}+\frac{a}{2}(u+C)\psi_x+\frac{aA}{6}D(t)\psi=0. \tag{4.12}$$

In this way, we have replaced the study of the given nonlinear evolution equation by the study of a pair of linear equations in the new unknown function ψ. They are the *linear* equations of the inverse scattering theory (Ref. 7, p. 587; Ref. 8)

Thus, the theory of the equivalence transformations seems to supply a *systematic procedure in the study of nonlinear evolution equations.*

5. THE BÄCKLUND TRANSFORMATION FOR THE KdV EQUATION

In this section, from the study of the symmetries of the equivalent equations previously found, we deduce the Bäcklund transformation for the KdV equation.[10] This example points out how Bäcklund transformations for evolution equations can be viewed from the standpoint of the theory of the equivalence transformations.

Following Chen,[11] let us observe that the modified KdV equation (3.17) manifestly admits the discrete symmetry

$$\bar{v}=-v. \tag{5.1}$$

Then by means of the equivalence transformation (3.13) (with $B=0$), which relates the modified to the KdV equation we at once obtain the following symmetry mapping $u\to\bar{u}$ of the KdV equation:

$$u=Av_x-\frac{aA^2}{6}v^2+C,$$
$$\bar{u}=-Av_x-\frac{aA^2}{6}v^2+C, \tag{5.2}$$

where v is any solution of the modified equation. Equations (5.2) may be regarded as the parametric equations of the symmetry mapping, where the function v plays the role of the parameter. *This symmetry mapping is the Bäcklund transformation for the KdV equation.* (See Fig. 4.)

Now, by means of the other equivalence transformations previously found we can recast this Bäcklund transformation into different forms.

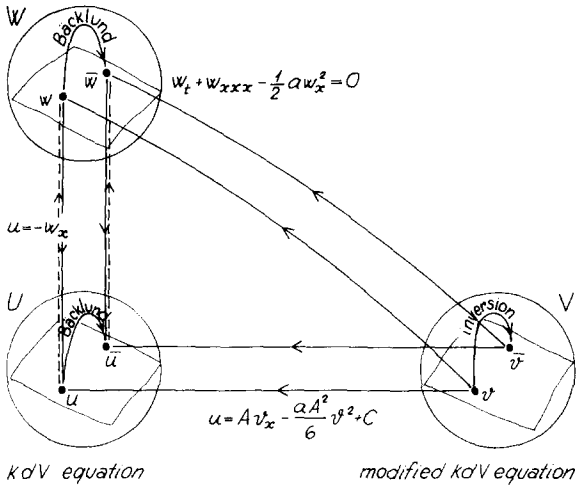


FIG. 5.

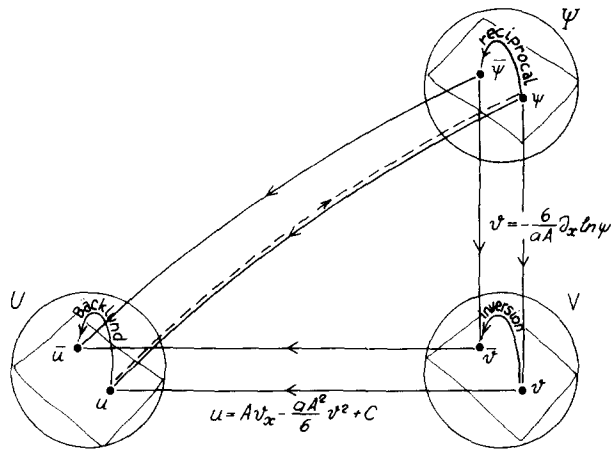


FIG. 6.

Firstly, we can work it out in the space W linked to the original space U by the equivalence transformation

$$u = -w_x. \tag{5.3}$$

(See Fig. 5.) In this space the parametric equations become

$$\begin{aligned} -w_x &= Av_x - \frac{aA^2}{6}v^2 + C, \\ -\bar{w}_x &= -Av_x - \frac{aA^2}{6}v^2 + C. \end{aligned} \tag{5.4}$$

The usual form of the Bäcklund transformation is then obtained by *eliminating the parameter* v between these parametric equations [observing that from (5.4) we get $2Av = \bar{w} - w$].[11]

Another form of the Bäcklund transformation can be obtained passing from the parameter v to the parameter ψ according to the scheme shown in Fig. 6. In this new parameter, the parametric equations become

$$\begin{aligned} u &= -\frac{6}{a}\frac{\psi_{xx}}{\psi} + C, \\ \bar{u} &= +\frac{6}{a}\frac{\psi_{xx}}{\psi} - \frac{12}{a}\frac{\psi_x^2}{\psi^2} + C. \end{aligned} \tag{5.5}$$

The first parametric equation is exactly the linear equation (4.11) of the inverse scattering theory. This fact points out clearly the strict connection between the Bäcklund transformation and the inverse scattering theory.

Lastly, observe that from (5.5) we get

$$\bar{u} - u = \frac{12}{a}\partial_{xx}\ln\psi. \tag{5.6}$$

Let us take, in particular, $u = 0$ as the starting solution of the KdV equation. Then the new solution $\bar{u}$ is given by

$$\bar{u} = \frac{12}{a}\partial_{xx}\ln\psi. \tag{5.7}$$

This relation is just the transformation suggested by Hirota.[12] This fact shows that the *Hirota tranformation is a particular case of the Bäcklund transformation.*

Thus, the theory of the equivalence transformations seems to supply a *systematic procedure to study the symmetries of the nonlinear evolution equations.*

CONCLUSION

The aim of this paper was to point out how the study of the equivalence transformations may supply a unified view in the theory of nonlinear evolution equations. The study of the Burgers and of the KdV equations emphasizes the constructive character of this method. The main result is that the transformations associated with these equations can be all deduced from the single operator equivalence condition (1.6). The same procedure can be followed also in the study of other evolution equations (such as the nonlinear Schrödinger equation), as we shall see in a subsequent paper.

APPENDIX

The main concept of nonlinear operator theory which we need is the notion of the Gateaux derivative of an operator $\mathbf{F}: V \to U$. It may be denoted by the symbol $\mathbf{F}'_v$[13] and is defined by

$$\mathbf{F}'_v\varphi = \frac{d}{d\epsilon}\mathbf{F}(v + \epsilon\varphi)\big|_{\epsilon=0}. \tag{A1}$$

If the operator $\mathbf{F}$ is of the form

$$\mathbf{F}(v) = f(v, v_x, v_{xx}, \cdots) \tag{A2}$$

we have

$$\begin{aligned} \mathbf{F}'_v\varphi &= \frac{d}{d\epsilon}f(v + \epsilon\varphi, v_x + \epsilon\varphi_x, \cdots)\big|_{\epsilon=0} \\ &= \frac{\partial f}{\partial v}\varphi + \frac{\partial f}{\partial v_x}\varphi_x + \frac{\partial f}{\partial v_{xx}}\varphi_{xx} + \cdots. \end{aligned} \tag{A3}$$

Then, we at once verify the relation

$$\begin{aligned} \partial_t\mathbf{F}(v) &= \frac{\partial f}{\partial v}\partial_t v + \frac{\partial f}{\partial v_x}\partial_t v_x + \cdots \\ &= \frac{\partial f}{\partial v}v_t + \frac{\partial f}{\partial v_x}\partial_x v_t + \cdots \\ &= \mathbf{F}'_v\partial_t v. \end{aligned} \tag{A4}$$

It also holds for operators $\mathbf{F}$ which are more general than (A2).

*This work has been sponsored by the Consiglio Nazionale delle Ricerche, Gruppo per la Fisica-Matematica.

[1]E.T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Reprinted, Cambridge U.P., London, 1970), Chap. 11.

[2]This geometric standpoint in the study of nonlinear functional analysis seems originally due to V. Volterra, *Leçons sur les fonctions de lignes* (Gauthier-Villars, Paris, 1913), pp. 43—8. It is particularly useful in the problem of giving a variational formulation to a given system of equations. See M.N. Vainberg, *Variational Methods for the Study of Nonlinear Operators* (Holden-Day, San Francisco, 1964); E. Tonti, Acad. R. Belg. Bull. Cl. Sci. 55, 137 (1969); F. Magri, Ann. Phys. 99, 196 (1976); Int. J. Eng. Sci. 12, 537 (1974).

[3]V. Arnold, *Equations differentielles ordinaires* (Editions MIR, Moscow, 1974), p. 142.

[4]F. Magri, Nuovo Cimento B **34**, 334 (1976).

[5]F. Magri, "An operator approach to conservation laws" (to be published).

[6]See, for example, R.M. Miura, C.S. Gardner, and M.D. Kruskal, J. Math. Phys. 9, 1204 (1968).

[7]See, for example, G.B. Whitman, *Linear and Nonlinear Waves* (Wiley, New York, 1974), p. 587.

[8]C.S. Gardner, J.M. Greene, M.D. Kruskal, and R.M. Miura, Phys. Rev. Lett. 19, 1095 (1967); P.D. Lax, Commun. Pure Appl. Math. 21, 467 (1968).

[9]For an alternative approach see H.D. Wahlquist and F.B. Estabrook, J. Math. Phys. 16, 1 (1975).

[10]H.D. Wahlquist and F.B. Estabrook, Phys. Rev. Lett. 31, 1386 (1973); G.L. Lamb, Jr., J. Math. Phys. 15, 2157 (1974).

[11]H.H. Chen, in *Bäcklund Transformations*, edited by R. Miura (Springer, Lecture Notes 515, New York, 1976), p. 242.

[12]R. Hirota, Phys. Rev. Lett. 27, 1192 (1971).

[13]S.G. Mikhlin, *The Numerical Performance of Variational Methods* (Wolters-Noordhoff, Groningen, 1971), p. 284.

Inclusions of arbitrary shape in an elastic medium

F. C. Chen and K. Young

Department of Physics, The Chinese University of Hong Kong, Hong Kong
(Received 9 August 1976)

The problem of elasticity where a homogeneous linear solid of arbitrary shape and elastic constants is embedded in an infinite homogeneous isotropic medium is discussed in a manner which is a straightforward generalization of Eyges' work on the Laplace equation. An exact integral equation is recast into an infinite system of algebraic equations and a systematic scheme of approximation is obtained by truncation. The lowest order approximation is discussed in detail, and its accuracy is shown to be quite good even for large differences in moduli between the two phases. Application to the effective moduli of composite materials is briefly discussed.

1. INTRODUCTION

In this paper we consider the problem of a homogeneous linear solid of arbitrary shape and arbitrary elastic constants (phase 2, or the inclusion) embedded in an infinite, homogeneous, isotropic medium (phase 1). In such a problem, one typically wishes to find the strain in the inclusion in terms of the strain imposed at infinity. An exact solution cannot be expected in general, so the utility of any method must be measured by the possibility of systematic approximations. The usual approach of first finding the general solution in each phase and then matching solutions at the boundary results in a set of linear algebraic equations for the coefficients in the general solutions. But where the shape is complicated, the imposition of boundary conditions becomes very difficult and it is not easy to formulate approximate boundary conditions either. Thus this method does not naturally give rise to a practical algorithm for obtaining approximate solutions.

This problem is treated here by an integral equation,[1,2] which will be derived in Sec. 2; here we only wish to point out two important features. (1) The equation incorporates both the differential equation of elasticity and the boundary conditions, so that the necessity for matching solutions is obviated. (2) The integral equation connects the strain in the inclusion directly to the strain at infinity and does not involve the strain in finite parts of phase 1. It is therefore tailored to the problem at hand.

The advantage of the integral equation lies in the possibility of systematic approximations, and it is the main task of this paper to discuss these in detail, in a manner which simply generalizes the work of Eyges on the Laplace equation[3] and the Schroedinger equation[4] and that of Waterman on the Helmholtz equation.[5] The main idea is quite simple: We expand the strain in the inclusion in a complete set, for example, a Taylor series, so that the integral equation becomes an infinite set of linear algebraic equations for the expansion coefficients. These equations are analogous to a similar set for the coefficients in the general solutions in the traditional approach, but is considerably simpler because only the strain inside the inclusion is involved. Finite subsets of these equations obtained by truncation can be solved, and taking larger subsets yields higher orders of approximation.

The lowest nontrivial order of approximation consists of retaining only one term and one equation and physically corresponds to the approximation where the strain inside the inclusion is regarded as uniform. This approximation is discussed in detail in Sec. 3. It is interesting to note that this approximation becomes exact when the inclusion is an ellipsoid. In this case, our result, of course, agrees with the classic work of Eshelby[6] on ellipsoidal inclusion, but the method used here is perhaps more accessible to those who are not expert in elastic theory. More importantly, we provide an approximate solution to nonellipsoidal inclusions. As an illustration, the result for a cubical inclusion is presented in Sec. 4. An important application of elastic inclusion theory is the calculation of the effective moduli of a composite material, and we present the result for a dilute dispersion of randomly oriented cubical inclusions.

It is, of course, important to know the form and size of the correction to the lowest order approximation. To this end, in Sec. 5 we present a simple higher-order calculation, which indicates that higher-order approximations yield a continued fraction expansion and that the error in the lowest order is small even for large differences in moduli between the two phases.

2. THE INTEGRAL EQUATION

A number of methods have been used to derive integral equations for various physical systems (dielectric,[3] quantum mechanical,[4] acoustic,[5] and elastic[1,2]) involving an inclusion embedded in a medium and governed by second order differential equations. Primarily to establish notations, we present a brief derivation for the case of elasticity based on the Green's function technique, which should underline the unity in the mathematical structure of all such systems.

Consider an infinite solid with a space-dependent stiffness tensor $c_{ij,kl}(\mathbf{x})$. Its energy density is

$$\mathcal{E}(\mathbf{x}) = \tfrac{1}{2} c_{ij,kl}(\mathbf{x})\, \partial_i u_j(\mathbf{x})\, \partial_k u_l(\mathbf{x}), \tag{1}$$

where $u_i(\mathbf{x})$ is the displacement field. Taking $\int \mathcal{E}(\mathbf{x})\, d^3x$ to be stationary gives the equation of static equilibrium

$$\partial_k [c_{ij,kl}(\mathbf{x})\, \partial_i u_j(\mathbf{x})] = 0. \tag{2}$$

Now suppose the solid to consist of an inclusion (phase 2) embedded in a medium (phase 1), both being homoge-

neous. Then

$$c_{ij,kl}(\mathbf{x}) = c^{(\alpha)}_{ij,kl}, \quad \mathbf{x} \in V_\alpha, \quad \alpha = 1,2, \tag{3}$$

and Eq. (2) can be written as

$$c^{(1)}_{ij,kl}\partial_k\partial_l u_j = \partial_k\{[c^{(1)}_{ij,kl} - c_{ij,kl}(\mathbf{x})]\partial_l u_j\} \tag{4}$$

or, more compactly,

$$D_{ij}u_j(\mathbf{x}) = \partial_k P_{ik}(\mathbf{x}), \tag{5}$$

where P_{ik} is the expression inside the curly brackets in (4) and vanishes in phase 1. The operator D_{ij} is

$$D_{ij} = c^{(1)}_{ij,kl}\partial_k\partial_l. \tag{6}$$

If we define the Green's function G_{jm} for the medium by

$$D_{ij}G_{jm}(\mathbf{x}) = \delta_{im}\delta^3(\mathbf{x}), \tag{7}$$

then Eq. (5) has the formal solution

$$u_j(\mathbf{x}) = u^0_j(\mathbf{x}) + \int d^3y\, G_{jm}(\mathbf{x}-\mathbf{y})[\partial_k P_{mk}(\mathbf{y})], \tag{8}$$

where u^0_j is a homogeneous solution. Integrating by parts and making use of the explicit form for P_{mk}, we convert (8) to[7]

$$u_j(\mathbf{x}) = u^0_j(\mathbf{x}) + \Delta c_{il,km}\int_2 d^3y\, \partial_k G_{jm}(\mathbf{x}-\mathbf{y})\,\partial_l u_i(\mathbf{y}), \tag{9}$$

where $\Delta c_{il,km} = c^{(2)}_{il,km} - c^{(1)}_{il,km}$ and the integral runs over phase 2 only.

This is the starting point of our approximation scheme to be discussed in the next sections. However, we have yet to know the Green's function G_{jm}. To keep the problem tractable, we specialize to cases in which phase 1 is isotropic, characterized by the bulk and shear moduli K_1 and G_1:

$$c^{(1)}_{il,km} = K_1\delta_{il}\delta_{km} + G_1(\delta_{ik}\delta_{lm} + \delta_{im}\delta_{lk} - \tfrac{2}{3}\delta_{il}\delta_{km}). \tag{10}$$

The Green's function can then be found from (7) by standard Fourier transform technique, the result being simply expressed as follows:

$$G_{ij}(\mathbf{x}) = -(1/8\pi)[G_1^{-1}(\delta_{ij}\nabla^2 - \partial_i\partial_j) + M_1^{-1}\partial_i\partial_j]|\mathbf{x}|, \tag{11}$$

where $M_1 = K_1 + \frac{4}{3}G_1$ is the longitudinal modulus.

We are now ready to identify the homogeneous term $u^0_j(\mathbf{x})$. If we let $|\mathbf{x}| \to \infty$, then, from (11), $G_{jm}(\mathbf{x}-\mathbf{y}) \to 0$, so that $u_j(\mathbf{x}) \to u^0_j(\mathbf{x})$, showing that $u^0_j(\mathbf{x})$ is the displacement at infinity.

3. SYSTEMATIC APPROXIMATION

The greatest advantage of (9) is the possibility of making systematic approximations. To proceed, we expand the strain in the inclusion in a Taylor series about the origin, which is chosen to be inside the inclusion:

$$u_j(\mathbf{x}) = \sum_n \frac{1}{n!}u_{j,p_1\ldots p_n}x_{p_1}\cdots x_{p_n}, \tag{12}$$

and similarly for $u^0_j(\mathbf{x})$.[8] We now derive a system of linear algebraic equations for the coefficients $u_{j,p_1\ldots p_n}$. Differentiate (9) n times and evaluate the result at the origin:

$$u_{j,p_1\ldots p_n} - u^0_{j,p_1\ldots p_n} = (-1)^n\Delta c_{il,km}\int_2 d^3y[\partial_{p_1}\cdots\partial_{p_n}\partial_k G_{jm}(\mathbf{y})\,\partial_l u_i(\mathbf{y})]. \tag{13}$$

On the right-hand side of (13), $\partial_l u_i(\mathbf{y})$ can likewise be expanded, so we get

$$u_{j,p_1\ldots p_n} - u^0_{j,p_1\ldots p_n} = (-1)^n\Delta c_{il,km}\sum_s (1/s!)T_{jm,kp_1\ldots p_n,q_1\ldots q_s}u_{l,iq_1\ldots q_s}, \tag{14}$$

where

$$T_{jm,kp_1\ldots p_n,q_1\ldots q_s} = \int_2 d^3y[\partial_k\partial_{p_1}\cdots\partial_{p_n}G_{jm}(\mathbf{y})]y_{q_1}\cdots y_{q_s},^9 \tag{15}$$

which only depends on the elastic constants of the medium (phase 1) and the geometry of the inclusion, but not on any strain in the problem.

If the system of equations (14) is truncated to a finite number of equations involving an equal number of unknowns, then the coefficients $u_{j,p_1\ldots p_n}$ can be solved in terms of the strain at infinity. In fact, the lowest order truncation of (14) contains enough nontrivial and useful features to warrant a detailed discussion. We take the $n=1$ equation and on the right-hand side keep terms only up to $s=0$. Then the following closed system of equations is obtained:

$$u_{j,p} - u^0_{j,p} = -\Delta c_{il,km}T_{jm,kp}u_{l,i}, \tag{16}$$

where from (15) [10]

$$T_{jm,kp} = \int dV\,\partial_k\partial_p G_{jm}. \tag{17}$$

Using (11) for the Green's function, we have

$$T_{jm,kp} = -\tfrac{1}{2}[G_1^{-1}\delta_{jm}t_{pkqq} + M_1^{-1}t_{jmkp}], \tag{18}$$

where the shape factors t_{jmkp} depend only on the shape of the inclusion and is totally symmetric and dimensionless[11]:

$$t_{jmkp} = (1/4\pi)\int dV\,\partial_j\partial_m\partial_k\partial_p|\mathbf{x}|. \tag{19}$$

In Appendix A we list the shape factors for spheroids, cylinders, and oblongs.

If we first decompose $u_{j,p}$ into the strain u_{jp} and the antisymmetric part a_{jp},

$$u_{jp} = \tfrac{1}{2}(u_{j,p} + u_{p,j}), \tag{20a}$$

$$a_{jp} = \tfrac{1}{2}(u_{j,p} - u_{p,j}), \tag{20b}$$

then (16) decouples into two sets. The six components of strain can be solved from the algebraic equations

$$u_{jp} - u^0_{jp} = -\Delta c_{il,km}T^+_{jm,kp}u_{li}, \tag{21}$$

while a_{jp} can be obtained from

$$a_{jp} - a^0_{jp} = -\Delta c_{il,km}T^-_{jm,kp}u_{li}. \tag{22}$$

Here we have used the abbreviations

$$T^\pm_{jm,kp} = \tfrac{1}{2}(T_{jm,kp} \pm T_{pm,kj}). \tag{23}$$

For inclusions with inversion symmetry the right-hand side of (22) vanishes so that $a_{jp} = a^0_{jp}$.

In general, Eq. (20) for the strain is only an approximation, based on regarding the strain in the inclusion as uniform. It is known[6] that for ellipsoidal inclusions the exact solution indeed yields a uniform strain. We shall prove this from the integral equation in Appendix

C. From the proof it will be seen that the possibility of an exact solution for ellipsoidal inclusions in the case of elasticity has the same geometric origin as the possibility of an exact solution in the case of the Laplace equation.[3]

4. AN EXAMPLE: A CUBE

We illustrate this method by treating a cubical inclusion—probably the simplest geometry for which no exact solution exists. For simplicity we take the cube to be elastically isotropic with bulk and shear moduli K_2, G_2, and choose the coordinate axes along the principal directions of the cube. With shape factors known (Appendix A), the solution of (21) for the strain $u_{j\rho}$ is straightforward and results in

$$u_{j\rho} = \Gamma_{j\rho,rs} u^0_{rs} \tag{24}$$

where the Γ's are

$$\begin{aligned}
&\Gamma_{12,12} = \tfrac{1}{2}[1 + c_1 \Delta G/G_1]^{-1}, \\
&\Gamma_{11,11} = \tfrac{2}{3}[1 + c_2 \Delta G/G_1]^{-1} + \tfrac{1}{3}[1 + c_3 \Delta K/K_1]^{-1}, \\
&\Gamma_{11,22} = \tfrac{1}{3}[1 + c_2 \Delta G/G_1]^{-1} - \tfrac{1}{3}[1 + c_3 \Delta K/K_1]^{-1},
\end{aligned} \tag{25}$$

and the c's depend only on the ratio of moduli in phase 1:

$$c_i = (3\alpha_i K_1 + 4\beta_i G_1)/(3K_1 + 4G_1), \tag{26}$$

where

$$\begin{aligned}
&\alpha_1 = \tfrac{2}{3} - 2/\sqrt{3}\pi = 0.30, \quad && \beta_1 = \tfrac{2}{3} - 1/2\sqrt{3}\pi = 0.57, \\
&\alpha_2 = \sqrt{3}/\pi = 0.55, && \beta_2 = \sqrt{3}/4\pi + \tfrac{1}{2} = 0.64, \\
&\alpha_3 = 1.00, && \beta_3 = 0.00.
\end{aligned}$$

Since K_1 and G_1 are positive, the c's lie between α_i and β_i. For a typical ratio $K_1/G_1 \sim 2$, all three c's are close to 0.5.[12]

Once we know the strain of a single inclusion, it is straightforward to calculate the effective moduli K and G of a dilute mixture containing a volume fraction f ($f \ll 1$) of such inclusions with random orientation. For inclusions of any shape the result can be written in the form

$$\begin{aligned}
K &= K_1 + f\Delta K S_K, \\
G &= G_1 + f\Delta G S_G,
\end{aligned} \tag{27}$$

where the dimensionless parameters S_K, S_G depend on shape. For randomly oriented cubical inclusions we find

$$S_K = [1 + c_3 \Delta K/K_1]^{-1} \quad \text{(cube)}, \tag{28a}$$

$$S_G = \tfrac{2}{5}[1 + c_2 \Delta G/G_1]^{-1} + \tfrac{3}{5}[1 + c_1 \Delta G/G_1]^{-1} \quad \text{(cube)}. \tag{28b}$$

It is interesting to compare with the well-known result for spherical inclusions [which can be readily obtained from (21) by using the values in Appendix A]. For spheres, S_K is exactly the same as (28a), while S_G is given by

$$S_G = [1 + c_4 \Delta G/G_1]^{-1} \quad \text{(sphere)}, \tag{29}$$

where c_4 is again of the form (26), with

$$\alpha_4 = \tfrac{2}{5}, \quad \beta_4 = \tfrac{3}{5}.$$

Since c_4 is again close to 0.5, there is actually very little difference between (28b) and (29), i.e., between cubical and spherical inclusions.

In this section, the cube is chosen merely for its simplicity. The main point is that the strain in an inclusion and the resultant effective moduli of dilute mixtures can now be extended (approximately) to nonellipsoidal inclusions. Previous work on effective moduli, except for certain bounds[13–15] have been limited chiefly to spheres and spheroids.[16,17] The present method opens the way to a more general discussion. One situation involving nonellipsoidal inclusions in the case of semicrystalline polymers,[18] in which the crystalline regions (phase 2) form lamellae which are probably best described as thin discs.

It is, of course, essential to know the accuracy of the lowest order approximation used in this section. To this end we discuss in the next section an example to next nontrivial order. The result of the calculation will then give a good idea of the size of the error expected in the lowest order.

5. HIGHER ORDER CALCULATION: AN EXAMPLE

In this section we show how the calculation can be carried to higher order. Such a calculation illustrates the completeness of the method—how increasing accuracy can in principle be obtained, and more importantly gives the form of the higher order corrections and hence an estimate of the accuracy of the lowest order approximation. Since higher order calculations are bound to proliferate indices, we continue with the example of a unit cube and for simplicity consider a purely hydrostatic strain at infinity

$$u^0_{j,\rho} = U\delta_{j\rho} \tag{30}$$

and calculate only to second order in the strain, i.e., up to $u_{j,\rho_1\rho_2\rho_3}$. (First order is trivial on account of reflection symmetry.) By symmetry, there are only three independent nonzero coefficients to this order.

$$A = u_{x,x}, \quad B = u_{x,xxx}, \quad C = u_{x,xyy}. \tag{31}$$

We take only the $n=1$ and $n=3$ equations in (14), and keep terms only up to $s=2$ on the right, thus yielding three equations for A, B and C. The $n=1$ equation gives, after some simplification,

$$\begin{aligned}
&[1 + \Delta K T^*_{jk,jk}]A + (\Delta G/3)T^*_{jn,jn,nn} B + (\Delta G/3) \\
&\quad \times [2T^*_{jl,ji,li} - 3T^*_{jn,jn,nn}]C = U,
\end{aligned} \tag{32}$$

The T's are defined by (15), repeated indices are summed as usual and we have adopted the convention that an index appearing four times is to be successively set equal to 1, 2, 3 (corresponding to the principal directions) and then summed, i.e.,

$$T^*_{jn,jn,nn} = T^*_{j1,j1,11} + T^*_{j2,j2,22} + T^*_{j3,j3,33}.$$

The $n=3$ case gives two independent equations

$$B + 2C = 0, \tag{33}$$

$$\begin{aligned}
&\{6 + \Delta G[T^*_{ni,lnnn,jj} + 2T^*_{ni,jnnn,ij} - 5T^*_{nm,mnnn,mm}]\}C \\
&\quad = 3\Delta T^*_{kn,knnn} A,
\end{aligned} \tag{34}$$

where in (34) we have already used (33) to eliminate B. Equation (33) is independent of the shape, and is in fact a consequence of the differential equation alone, as can be directly verified. All the T's can be evaluated analytically (Appendix B); putting these into (32)—(34), we finally obtain

$$A = U\left\{1 + \frac{\Delta K}{M_1} - \frac{0.11\Delta K\,\Delta G/M_1^2}{1 + \Delta G(0.72/M_1 + 0.25/G_1)}\right\}^{-1}, \quad (35a)$$

$$C = -\tfrac{1}{2}B = 0.74\,\frac{\Delta K}{M_1}\left\{1 + \Delta G\left(\frac{0.72}{M_1} + \frac{0.25}{M_1}\right)\right\}^{-1} A. \quad (35b)$$

In contrast, the lowest order approximation gives

$$A = U\{1 + \Delta K/M_1\}^{-1}, \quad (36)$$

with no information on B and C.

This suggests that the present method gives a continued fraction expansion. Take, for example, (in arbitrary units)

$$G_1 = 1, \quad K_1 = 2, \quad G_2 = 5, \quad K_2 = 10;$$

then the lowest order approximation (36) gives $A = 0.294\ U$ while the next nontrivial order (35a) gives $A = 0.304U$, only showing a 3% difference. Thus even for one phase much stiffer than the other ($K_2/K_1 = 5$ in this example), the lowest order approximation is quite accurate. Indeed it can be seen that (35a) and (36) never differ by more than 10% for all values of ΔK and ΔG. In contrast, low order "Born approximation," i.e., the first few terms of the Neumann series solution to (9), is unlikely to be accurate except for small $\Delta K/K_1$, $\Delta G/G_1$.

6. CONCLUSION

We conclude by recalling the limitations of this method. First, the discussion is confined to *linear* solids, so that large deformations cannot be treated. Secondly the medium (phase 1) has to be isotropic, for otherwise the Green's function G_{ij} is unknown. We emphasize, however, that an anisotropic inclusion (phase 2) presents no problem except that of greater complexity in the algebra.

Higher-order calculations are possible in principle but are likely to be cumbersome, on account of the intrinsic complexity of elasticity. The lowest-order approximation is, however, sufficiently accurate to be practically useful, for example, in studying the effective moduli of composite materials.

It is also to be stressed that the only mathematical tools used in this study are "standard" ones, i.e., the same ones that are encountered in treating the more familiar equations of mathematical physics.[3-5] The tools, and indeed the language, specific to elastic theory (e.g., biharmonic functions, "cuts" and "traction") which are perhaps not familiar to all physicists have been completely avoided.

APPENDIX A

The shape factors t_{ijkl} are defined by

$$t_{ijkl} = (1/4\pi)\int dV\,\partial_i\partial_j\partial_k\partial_l|\mathbf{x}|. \quad (A1)$$

Spheroid

For a spheroid with principal axes a, a, b, and the b axis along $\hat{n}$, the shape factor is

$$\begin{aligned} t_{ijkl} = {} & \alpha(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \\ & + \beta(\delta_{ij}n_kn_l + \delta_{ik}n_jn_l + \delta_{il}n_jn_k \\ & + \delta_{kl}n_in_j + \delta_{jl}n_in_k + \delta_{jk}n_in_l) \\ & + \gamma n_in_jn_kn_l, \end{aligned} \quad (A2)$$

where

$$\alpha = -\tfrac{1}{4} + \tfrac{1}{24}(1-k^2) + \tfrac{1}{8}(3+k^2)L,$$

$$\beta = \tfrac{1}{4} + \tfrac{1}{8}(1-k^2) - \tfrac{3}{8}(5-k^2)L,$$

$$\gamma = -\tfrac{3}{4} - \tfrac{15}{8}(1-k^2) + \tfrac{15}{8}(7-3k^2)L,$$

$$k^2 = -\kappa^2 = 1 - a^2/b^2,$$

and

$$L = \begin{cases} \dfrac{1-k^2}{k^4}\left(\dfrac{1}{2k}\ln\dfrac{1+k}{1-k} - 1 - \tfrac{1}{3}k^2\right), & b \geq a \ \text{(prolate)}, \\ \dfrac{1+\kappa^2}{\kappa^4}\left(\dfrac{1}{\kappa}\tan^{-1}\kappa - 1 + \tfrac{1}{3}\kappa^2\right), & b \leq a \ \text{(oblate)}. \end{cases}$$

For a sphere ($k = 0$) it reduces to

$$t_{ijkl} = -\tfrac{2}{15}(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$$

while for an extremely oblate spheroid ($k \to +\infty$)

$$t_{ijkl} = -2n_in_jn_kn_l.$$

Cylinder

For a right circular cylinder of radius R and length $2l$, with cylinder axis along $\hat{n}$, the shape factor is again given by (A2), where now

$$\alpha = -\tfrac{1}{8}(\xi + \xi^3), \quad \beta = -\tfrac{1}{8}(3\xi - 5\xi^3),$$

$$\gamma = -2 + \tfrac{5}{8}(9\xi - 7\xi^3), \quad \xi = l/(l^2 + R^2)^{1/2}.$$

Oblong

For a oblong with edges $a\hat{i}$, $b\hat{j}$, $c\hat{k}$, there are only two independent shape factors:

$$t_{1111} = -\frac{4}{\pi}\tan^{-1}\frac{cb}{a\Delta} + \frac{2}{\pi}\frac{abc}{\Delta}\frac{a^2+\Delta^2}{a^2\Delta^2 + b^2c^2},$$

$$t_{1122} = -\frac{2}{\pi}\frac{abc}{(a^2+b^2)\Delta},$$

where $\Delta = (a^2 + b^2 + c^2)^{1/2}$. All other components can be obtained by permutation. For a cube, these reduce to

$$t_{1111} = -\tfrac{2}{3} + \frac{2}{\sqrt{3}\pi}, \quad t_{1122} = \frac{-1}{\sqrt{3}\pi}.$$

A useful check on these results is given by the identity $\nabla^2\nabla^2|\mathbf{x}| = -8\pi\delta^3(\mathbf{x})$, which implies $t_{iikk} = -2$. Since long rods/flat disks can be regarded as the limit of spheroids, cylinders, or oblongs, consistency in these limits provides another check.

APPENDIX B

Here we list the T's for a cube ($|x|$, $|y|$, $|z| \leq 1$)

that occur in Eqs. (32)—(34):

$$T^{+}_{jk,jk} = 1/M_1,$$

$$T^{+}_{jn,jn,nn} = -(1/M_1)[(6/\pi)\ln(2+\sqrt{3})-2] = -0.515/M_1,$$

$$T^{+}_{jl,jj,li} = -(1/M_1)[(6/\pi)\ln(2+\sqrt{3})-1] = -1.515/M_1,$$

$$T^{+}_{ni,innn,jj} = (1/M_1)[14\sqrt{3}/\pi - 2] = 5.72/M_1,$$

$$T^{+}_{ni,jnnn,ij} = (12-12\sqrt{3}/\pi)/M_1 + (20\sqrt{3}/\pi - 8)/G_1 = 5.38/M_1 + 3.03/G_1,$$

$$T^{+}_{nm,mnnn,mm} = (-10\sqrt{3}/3\pi + 6)/M_1 + (10\sqrt{3}/\pi - 4)/G_1 = 4.16/M_1 + 1.51/G_1,$$

$$T^{+}_{kn,knnn} = (8\sqrt{3}/3\pi)/M_1 = 1.47/M_1.$$

APPENDIX C

Here we prove that (16) is exact for an ellipsoidal inclusion. The approximation of uniform strain would be exact if, upon substituting a constant for $\partial_i u_l(\mathbf{y})$ into the right-hand side of (9), the resultant integral is linear in x, for after solving (16) we would then have satisfied (9) exactly. So we require $\int d^3y\, \partial_k G_{jm}(\mathbf{x}-\mathbf{y})$ to be linear in x, that is,

$$\int d^3y\, \partial_{\alpha_1}\cdots\partial_{\alpha_n}\partial_k G_{jm}(\mathbf{y}) = 0, \quad n \geq 2.$$

Since $G_{jm}(\mathbf{y})$ is linearly related to second derivatives of $|\mathbf{y}|$ (r) by (11), it is sufficient if

$$\int dV\, \partial_{\alpha_1}\cdots\partial_{\alpha_N} r = 0, \quad N = n+3 \geq 5, \tag{C1}$$

For an ellipsoid centered at the origin, it is only necessary to consider even N, say $N = 2L$, $L \geq 3$.

First consider a sphere. Symmetry considerations dictate that (C1) must have the form

$$\int dV\, \partial_{\alpha_1}\cdots\partial_{\alpha_n} r = A_N[\delta_{\alpha_1\alpha_2}\delta_{\alpha_3\alpha_4}\cdots + \cdots], \tag{C2}$$

where [ ] contains all possible contraction of $\alpha_1\cdots\alpha_N$ in pairs. Summing over each pair of indices $\alpha_1\alpha_2$, $\alpha_3\alpha_4$, etc., gives

$$\int dV(\nabla^2)^L r = A_N \times \text{positive constant}. \tag{C3}$$

Since $\nabla^2\nabla^2 r = -8\pi\delta^3(\mathbf{r})$ and here $L \geq 3$, the integrand in (C3) is the derivative of a delta function, which, upon conversion to a surface integral, proves $A_N = 0$ and hence (C1).

Next consider a spheroid symmetric about the z axis. A general ellipsoid can be treated in exactly the same way. Since (C1) has already been proved for any sphere, it suffices to integrate over the spheroid minus a small sphere of radius ϵ. So (C1) is

$$\int_{r>\epsilon} dV\, \partial_{\alpha_1}\cdots\partial_{\alpha_N} r = \int d\Omega \int_{\epsilon}^{R(\Omega)} dr\, r^2\, \partial_{\alpha_1}\cdots\partial_{\alpha_N} r, \tag{C4}$$

where $R(\Omega)$ is the polar representation of the surface, in the case of a spheroid given by

$$R(\Omega)^{-2} = \frac{1}{a^2} + \left(\frac{1}{b^2} - \frac{1}{a^2}\right)\cos^2\Theta. \tag{C5}$$

The integrand $(\partial_{\alpha_1}\cdots\partial_{\alpha_N} r)$ is dimensionally r^{1-N} and is totally symmetric among its indices. Again from $\nabla^2\nabla^2 r = -8\pi\delta^3(\mathbf{r})$ we see that the tensor vanishes upon summation over any two or more pairs of indices, since the volume of integration excludes the origin. Hence the tensor is a linear combination of terms with angular momentum $l = N$ and $l = N-2$, i.e.,[19]

$$(\partial_{\alpha_1}\cdots\partial_{\alpha_N} r) = r^{1-N}\left\{\sum_m a_m Y_{N,m}(\Omega) + \sum_m b_m Y_{N-2,m}(\Omega)\right\}, \tag{C6}$$

where the dependence of a_m, b_m on $\alpha_1\cdots\alpha_N$ is suppressed. Inserting (C6) into (C4) and doing the radial integral gives

$$\int_{r>\epsilon} dV(\partial_{\alpha_1}\cdots\partial_{\alpha_N} r) = \int d\Omega\, \frac{1}{4-N}\left[R(\Omega)^{4-N} - \epsilon^{4-N}\right] \times\left\{\sum_m a_m Y_{N,m}(\Omega) + \sum_m b_m Y_{N-2,m}(\Omega)\right\}.$$

The term involving ϵ clearly vanishes, while from (C5) we see that $R(\Omega)^{4-N}$ is a polynomial of degree $N-4$ in $\cos\Theta$ and hence has zero projection on $Y_{N,m}$ and $Y_{N-2,m}$. This completes the proof of (C1) for spheroids. The proof relies on the fact that for ellipsoids $R(\Omega)^{-2}$ contains only up to $l = 2$ terms, and it is therefore clear that uniform strain cannot be expected to hold for shapes other than ellipsoids.

A similar result for the Laplace equation has been given by Eyges.[3]

[1] R.J. Knops, Proc. Edin. Math. Soc. **14**, 61 (1964).
[2] V.D. Kupradze, Prog. Solid Mech. **3**, 1 (1963).
[3] L. Eyges, Ann. Phys. (N.Y.) **90**, 266 (1975).
[4] L. Eyges, Ann. Phys. (N.Y.) **81**, 567 (1973).
[5] P.C. Waterman, J. Acoust. Soc. Am. **45**, 1417 (1969).
[6] J.D. Eshelby, Proc. Roy. Soc. A **241**, 376 (1957); Prog. Solid Mech. **2**, 88 (1961).
[7] ∂_k operating on a function of $x-y$ shall always mean $\partial/\partial y_k$.
[8] u^0_j will only contain terms up to $n=1$ if, as is usually the case, the strain at infinity is constant.
[9] Repeated integration by parts show that there is no problem at the origin.
[10] From now on, all integrals will be understood to be over phase 2 only.
[11] Power counting may suggest a logarithmic divergence at the origin, but actually no such divergence occurs because the volume integral can be converted to a surface integral.
[12] A value of $c_1 = 0$ would mean $u_{12} = u^0_{12}$, corresponding to a "parallel" arrangement of the two phases, while $c_1 = 1$ means $G_2 u_{12} = G_1 u^0_{12}$, i.e., equality of stresses, corresponding to a "series" arrangement. The fact that $c_i \sim \frac{1}{2}$ can be heuristically interpreted as saying that the actual situation is midway between the two.
[13] Z. Hashin, J. Appl. Mech. **29**, 143 (1962).
[14] Z. Hashin and J. Shtrikman, J. Mech. Phys. Solids **11**, 127 (1963).
[15] R. Hills, J. Mech. Phys. Solids **11**, 357 (1963).
[16] L.J. Walpole, J. Mech. Phys. Solids **14**, 151 (1966).
[17] R. Roscoe, Rheol. Acta **12**, 404 (1973).
[18] R. Gray and N. McCrum, J. Polym. Sci. A-2 **7**, 1329 (1969).
[19] Equation (C7) holds only when put under the integral in (C4).

Geometrization of configuration space

B. E. Eichinger

Department of Chemistry, University of Washington, Seattle, Washington 98195
(Received 2 August 1976)

Projective geometry provides the means to geometrize configuration space, granted three elementary postulates: (1) It is possible to describe physical systems as collections of a countable number n of particles; (2) the mass-inertial tensor of such a system is positive definite; (3) the Galilei group acting on elastic bodies can be rationally extended to the whole of $GL(4,R)$. From these it is possible to deduce that any of the three symmetric spaces $O(4,n)/O(4)\times O(n)$, $U(2,n)/U(2)\times U(n)$, and $Sp(1,n)/Sp(1)\times Sp(n)$ represents the configuration space. These three spaces are Einstein manifolds of constant sectional curvature. Several geometrical theorems are given for all three spaces; the metric on each is given in both Cartesian and polar forms; infinitesimal generators of the Lie algebras $so(4,n)$, $su(2,n)$, and $sp(1)\times sp(n)$ are also obtained. Physical quantities are presumed to be obtained from eigenfunctions of the Laplace–Beltrami operator on the three spaces, and partial solutions for such functions are obtained on $Sp(1,n)/Sp(1)\times Sp(n)$. These solutions fall into two classes: Those with integral values of a quantum number $l\geq 0$ form the sequence $2(l+1)^2$; those with half-integral $l\geq 1/2$ form a series with $(2l+1)(l+3/2)$ members, and possibly represent leptons, quarks, etc.

I. INTRODUCTION

That physical space is non-Euclidean is generally accepted. Locally, the geometry of space has long been taken to be that of the Euclidean space E^3; but over large distances the Riemannian curvature is nonzero. It is here suggested that it might be of benefit to presume that space is locally projectively flat (P^3). By investigating the group of automorphisms acting on the configuration space of n particles imbedded in P^3, it is possible to discover that the global nature of such a space is hyperbolic.

There are several interesting analogies between projective geometry and the physical world. For example, the plane-point duality[1] of P^3 is analogous to wave-particle duality. Furthermore, P^3 is not simply connected, and in consequence half-integral spin is comprehensible. In essense, the step from E^3 to P^3 admits of transformations to and from infinity. In classical terms, such transformations are capable of creating and annihilating charge and mass in exchange for the potential at infinity. This paper does not purport to calculate charge or mass; it contains only the basic formulation of the projective geometry of n particle systems.

The geometry of the configuration space of an n-particle system may be deduced from consideration of the group of automorphisms acting on the space in such a way as to preserve the positive definite character of the mass-inertial tensor of the system. Let particle i, $1\leq i\leq n$, considered to be a point with mass m_i, be located in flat Euclidian space at y_i^α, $1\leq\alpha\leq 3$. The matrices $M^{1/2}=\mathrm{diag}(m_1^{1/2}, m_2^{1/2}, \ldots, m_n^{1/2})$ and

$$Y=\begin{bmatrix} y_1^1 & y_2^1 & \cdots & y_n^1 \\ y_1^2 & y_2^2 & \cdots & y_n^2 \\ y_1^3 & y_2^3 & \cdots & y_n^3 \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

yield the mass-inertial tensor T on construction of $T=XX'$, where $X=YM^{1/2}$ and where X' is the transpose of X. The tensor T has components

$$\begin{bmatrix} \sum m_i y_i^1 y_i^1 & \sum m_i y_i^1 y_i^2 & \sum m_i y_i^1 y_i^3 & \sum m_i y_i^1 \\ & \sum m_i y_i^2 y_i^2 & \sum m_i y_i^2 y_i^3 & \sum m_i y_i^2 \\ \text{sym} & & \sum m_i y_i^3 y_i^3 & \sum m_i y_i^3 \\ & & & \sum m_i \end{bmatrix}$$

and is symmetric. The matrix X may be interpreted as an indefinite Grassmann manifold[2] (as will be seen later); the coordinates $(m_i^{1/2}y_i^1, m_i^{1/2}y_i^2, m_i^{1/2}y_i^3, m_i^{1/2}) = (x_i^1, x_i^2, x_i^3, x_i^4)$ are homogeneous coordinates[1] of the projective space P^3. The later observation stems from the fact that $y_i^\alpha = x_i^\alpha/x_i^4$. The points y_i of P^3 may be considered to be the rays of V^4, the vector space x_i of four dimensions. The homogeneous coordinates x_i are defined by the y_i^α only to within a constant of proportionality, at least insofar as the intrinsic geometry of the manifold is concerned.

Suppose that T is brought to diagonal form by the Galilean transformation (supposing time to be an implicit variable)

$$\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}\begin{bmatrix} I & y_0 \\ y_0' & m \end{bmatrix}\begin{bmatrix} R' & 0 \\ t' & 1 \end{bmatrix}=\begin{bmatrix} I & 0 \\ 0 & m \end{bmatrix},$$

where $R\in SO(3)$,[3] t is a translation vector, $I^{\alpha\beta}=\sum m_i y_i^\alpha y_i^\beta$, $m=\sum m_i$, and $y_0^\alpha=\sum m_i y_i^\alpha$. (The choice $t=-m^{-1}Ry_0$ accomplishes the task.) The fundamental assertion that the matrix T is positive definite and hence bounded from below is equivalent to saying that systems cannot vanish from the world and that masses are nonnegative. Let the lower bound be $a^2\mathbf{1}_4$, where a is any number less than the minimum eigenvalue of XX', so that

$$XX'>a^2\mathbf{1}_4,$$
$$a^{-1}XX'a^{-1}-\mathbf{1}_4>0. \tag{I. 1a}$$

(The matrix equation $A>0$ will mean that A is positive definite; $A<0$ signifies that A is negative definite.) Since the coordinates y_i^α are unaffected by $a^{-1}X\to X$, Eq. (I. 1a) is geometrically equivalent to

$$XX'-1>0 \tag{I. 1b}$$

Where there is no chance for confusion, the order of the identity will henceforth be suppressed.

Those transformations, considered as a subgroup of GL(4, R), which leave infinity fixed, i.e., do not move the hyperplane $x_i^4 = 0$, include the Galilei translations and rotations as well as the discrete crystallographic groups. In addition, arbitrary homogeneous states of strain may be induced in the system if we allow all 12 parameters of

$$\begin{bmatrix} R_2SR_1 & t \\ 0 & 1 \end{bmatrix}$$

to be arbitrary. Here the R_i are rotation matrices (3 + 3 parameters), S is a diagonal deformation gradient matrix (3 parameters) and t is a translation vector (3 parameters). (This group will be called the extended Galilei group.) To get all of GL(4, R), which is the group of automorphisms of P^3, we need to include the subgroup of matrices of the form

$$\begin{bmatrix} 1 & 0 \\ r' & a \end{bmatrix},$$

which are not traditionally considered in the mechanics of either rigid or elastic bodies.

Consider first the action of the extended Galilei group on T, as given by

$$\begin{bmatrix} R_2SR_1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & y_0 \\ y_0' & m \end{bmatrix} \begin{bmatrix} R_1'S'R_2' & 0 \\ t' & 1 \end{bmatrix} = \begin{bmatrix} \bar{I} & \bar{y}_0 \\ \bar{y}_0' & \bar{m} \end{bmatrix}.$$

The action is such as to leave the mass invariant, $\bar{m} = m$, while the distribution of points is compressed or dilated, rotated, and translated. In terms of the individual $y_i^\alpha = x_i^\alpha / x_i^4$, the transformation gives $\bar{y}_i^\alpha = (R_2SR_1)_\beta^\alpha y_i^\beta + t^\alpha$.

The remainder of GL(4, R) consists of scalar dilations of the mass, which maps $y_i^\alpha \to a^{-1} y_i^\alpha$ corresponding to a change of scale, as well as of those with matrices

$$\begin{bmatrix} 1 & 0 \\ r' & 1 \end{bmatrix},$$

whose action is to be investigated. Suppose that T has been brought to the diagonal form

$$\begin{bmatrix} I & 0 \\ 0 & m \end{bmatrix}$$

by the action of the Galilei group. Now apply the transformation

$$\begin{bmatrix} 1 & 0 \\ r' & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} I & Ir \\ r'I & m + r'Ir \end{bmatrix},$$

which increases the mass of the system, and translates it as well. The translation could now be undone by a Galilei translation which does not affect the mass. Hence, the action of this part of GL(4, R) corresponds to an increase of energy (on use of relativistic mass–energy equivalence). Suppose now that T is not diagonal, and choose $r = -I^{-1}y_0$. Then

$$\begin{bmatrix} 1 & 0 \\ -y_0'I^{-1} & 1 \end{bmatrix} \begin{bmatrix} I & y_0 \\ y_0' & m \end{bmatrix} \begin{bmatrix} 1 & -I^{-1}y_0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & m - y_0'I^{-1}y_0 \end{bmatrix}.$$

The scalar of energy may thus be diminished by such a transformation.

Altogether, a matrix $g^{-1} \in$ GL(4, R) acting on T according to $g^{-1}T(g^{-1})'$ translates, rotates, deforms isotropically and/or anisotropically, and either increases or decreases the energy of the system of particles. This group of automorphisms thus extends the classical mechanics of rigid and elastic bodies to encompass the effect of heat exchange in a geometrical way. This is indeed peculiar to projective geometry. The physical consequences of these transformations may be understood on closer inspection.

The action of the part of GL(4, R) considered above is such as to (possibly) map some particles to infinity, while other particles are (possibly) mapped from infinity to the finite domain. Consider

$$\begin{bmatrix} 1 & 0 \\ r' & 1 \end{bmatrix} \begin{bmatrix} x_i^4 y_i \\ x_i^4 \end{bmatrix},$$

where $x_i^4 y_i$ is the three-dimensional part of x_i. The coordinates y_i are mapped to

$$\bar{y}_i = y_i / (1 + r'y_i)$$

by such a transformation. It is apparent from this equation that points on the plane $r'y = 0$ are invariant points, points on the plane $r'y = -1$ are sent to infinity, and points at infinity are sent to $r/(r'r)$ (as can be seen from the inverse transformation).

Particles with $x_i^4 = 0$ are massless, and so the points at infinity may well represent photons. Hence, the matrix X is constituted of the homogeneous coordinates of both matter and the electromagnetic field. Projective duality[1] on P^3 asserts that points and planes are equivalent, which suggests that wave-particle duality is nothing other than the projective duality of P^3. Matter can be created or annihilated by mappings with points at infinity, alias photons.

In summary, the import of these considerations is very simple: The "world" composed of n particles (photons included) is taken to be locally projectively flat rather than locally Euclidean flat. The difference between projective and affine geometry is that mappings to and from infinity are allowed in the former. Such mappings may correspond to emission and absorption of radiation.

II. REAL PROJECTIVITIES OF CONFIGURATION SPACE

Let $g^{-1} = (C_\alpha^\beta) \in$ GL(4, R). The action of g^{-1} on y_i is given by linear fractional transformations

$$\bar{y}_i^\beta = (C_\alpha^\beta y_i^\alpha + C_4^\beta) / (C_\gamma^4 y_i^\gamma + C_4^4),$$

which constitute all projectivities of P^3. The matrix $XX' - 1$ is taken to

$$\bar{X}\bar{X}' - 1 = g^{-1}(XX' - 1)g'^{-1} > 0$$

since $(g'g)^{-1}$ is positive definite. We now wish to obtain the projectivity $X \to \bar{X}$. First note that

$$P(XX' - 1)P' = PXQ'QX'P' - 1 = X_1X_1' - 1,$$

where $X_1 = PXQ'$, $P \in O(4)$, and $Q \in O(n)$, gives an equivalence class. That is, all matrices PXQ', with P and Q orthogonal and of rank 4 and n respectively, are equivalent to X that they are characterized as belonging to the same equivalence class of mass-inertial tensors T. This suggests that the space X will be a coset space $G/O(4) \times O(n)$,[4] and indeed this is what will now be shown.

The transformation

$$\bar{X}\bar{X}' - 1 = g^{-1}(XX' - 1)g'^{-1} \tag{II. 1}$$

induces

$$\bar{X} = G(X),$$

and the nature of $G(X)$ is to be determined. From the form of Eq. (II. 1) it is clear that one may take

$$\bar{X} = g^{-1}h,$$

where $g = g(X)$ and $h = h(X)$. Then

$$hh' - gg' = XX' - 1. \tag{II. 2}$$

Let

$$h = h_0 + h_1Xh_2 + h_3Xh_4X'h_5 + \cdots$$

and

$$g = g_0 + g_1Xg_2 + g_3Xg_4X'g_5 + \cdots;$$

on multiplying the two series in the manner indicated in Eq. (II. 2) one finds that

$$\begin{aligned} &h_0h_0' - g_0g_0' = -1_4, \quad h_1 = g_1 = 1_4, \\ &h_2h_0' = g_2g_0', \quad h_0h_2' = g_0g_2', \\ &h_2h_2' - g_2g_2' = 1_n \end{aligned} \tag{II. 3}$$

and that all other h_i and g_i must vanish. Hence, the linear fractional transformation is

$$\bar{X} = (XB' + A')^{-1}(XD' + C'), \tag{II. 4a}$$

where $A' = g_0$, $B' = g_2$, $D' = h_2$, and $C' = h_0$. These matrices constitute the group $O(4, n)$, as can be seen from Eq. (II. 3). That is,

$$\begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix} \begin{bmatrix} 1_4 & 0 \\ 0 & -1_n \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1_4 & 0 \\ 0 & -1_n \end{bmatrix},$$

giving

$$\begin{aligned} &A'A - C'C = g_0g_0' - h_0h_0' = 1_4, \\ &A'B - C'D = g_0g_2' - h_0h_2' = 0, \\ &B'A - D'C = g_2g_0' - h_2h_0' = 0, \\ &B'B - D'D = g_2g_2' - h_2h_2' = -1_n \end{aligned}$$

on comparison with Eqs. (II. 3). Let

$$\Gamma_R = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad H_R = \begin{bmatrix} 1_4 & 0 \\ 0 & -1_n \end{bmatrix}$$

so that

$$\Gamma_R'H_R\Gamma_R = H_R. \tag{II. 5a}$$

We also have

$$\Gamma_RH_R\Gamma_R' = H_R. \tag{II. 5b}$$

Furthermore,

$$\bar{X} = (AX + B)(CX + D)^{-1} = (XB' + A')^{-1}(XD' + C'), \tag{II. 4b}$$

from which all invariants may be constructed.

We may now return to the transformation equation (II. 1) to find

$$\begin{aligned} \bar{X}\bar{X}' - 1 &= (XB' + A')^{-1}[(XD' + C')(DX' + C) \\ &\quad - (XB' + A')(BX' + A)](BX' + A)^{-1} \\ &= (XB' + A')^{-1}(XX' - 1)(BX' + A')^{-1} \end{aligned}$$

on use of Eqs. (II. 5).

The whole of the space X splits into separable domains under the action of these transformations. Points X in the domain $XX' - 1 > 0$ remain in this domain, points on the boundary $XX' - 1 = 0$ stay on the boundary, and points in the Cartan domain $XX' - 1 < 0$ map into the same domain. (The word point is here understood to mean a point in configuration space.)

The space $XX' - 1 > 0$ is a symmetric space, as will now be shown. A symmetric space is defined as one for which every point is an isolated fixed point of an involution.[5] An involution is one for which

$$H_R\Gamma_R'H_R = \Gamma_R, \tag{II. 6}$$

or $A' = A$, $D' = D$, $-C' = B$. The solution of

$$X = (AX + B)(-B'X + D)^{-1} = (XB' + A)^{-1}(XD - B)$$

gives the fixed point $X = -(A - 1)^{-1}B$. This point lies in the domain $XX' - 1 = (A - 1)^{-1}[BB' - (A - 1)^2](A - 1)^{-1} = 2(A - 1)^{-1} > 0$ provided $A > 0$. The positive definite character of $(A - 1)^{-1}$ results from $A^2 - BB' = 1$; $A - 1$ is positive definite if A is, since BB' is positive definite (except on subspaces of lower dimension). Owing to the equivalence relation $X \sim P'XQ$, the symmetric space may be identified with one of the connected components of the homogeneous and symmetric spaces $O(4, n)/O(4) \times O(n)$. However, we will not make explicit use of the theory[2,3,6] of symmetric spaces here.

III. COMPLEX AND QUATERNION SPACES

That the homogeneous coordinates of a point are even in number permits of a representation of the configuration space by complex matrices. That they are four in number further allows a representation in terms of quaternions. Since much of the development of the geometries over the real, complex, and quaternion fields is formally the same, the basis for these alternative representations will now be presented, beginning with the complex.

Define the $2 \times n$ complex matrix Z, with elements

$$\begin{bmatrix} z_j^1 \\ z_j^2 \end{bmatrix} = \begin{bmatrix} x_j^1 + ix_j^2 \\ x_j^3 + ix_j^4 \end{bmatrix}, \quad 1 \le j \le n.$$

Then ZZ^*, where Z^* is the complex conjugate of the transpose of $Z (Z^* = \bar{Z}')$, is bounded from below if XX' is. Hence, we may set

$$ZZ^* - 1_2 > 0 \tag{III. 1}$$

to define the geometry of physically accessible space. The group Γ_c consisting of matrices

$$\Gamma_c = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{III. 2a}$$

satisfying

$$\Gamma_c H_c \Gamma_c^* = H_c \tag{III. 2b}$$

and

$$\Gamma_c^* H_c \Gamma_c = H_c \tag{III. 2c}$$

where

$$H_c = \begin{bmatrix} 1_2 & 0 \\ 0 & -1_n \end{bmatrix} \tag{III. 2d}$$

is the group $U(2, n)$ and acts upon Z according to

$$Z_1 = (AZ + B)(CZ + D)^{-1} = (ZB^* + A^*)^{-1}(ZD^* + C^*), \tag{III. 3}$$

as can be shown with the procedure used to obtain Eq. (II. 4b). The space $ZZ^* - 1 > 0$ may be identified with the coset space $U(2, n)/U(2) \times U(n)$.

The space composed of n quaternions is represented by $2 \times 2n$ matrices Q. Let the elements of Q be the 2×2 quaternions q_j, with

$$q_j = \begin{bmatrix} z_j^1 & z_j^2 \\ -\bar{z}_j^2 & \bar{z}_j^1 \end{bmatrix}$$

or

$$q_j = \sum_{\alpha=1}^{4} \sigma_\alpha x_j^\alpha .$$

The σ_α are defined[7] by

$$\sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix},$$

$$\sigma_3 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \sigma_4 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix},$$

the latter three of which are equivalent to Pauli spin matrices. The quaternion conjugate to

$$q_j = \sigma_1 x_j^1 + \sum_{\alpha=2}^{4} \sigma_\alpha x_j^\alpha$$

is

$$\hat{q}_j = \sigma_1 x_j^1 - \sum_{\alpha=2}^{4} \sigma_\alpha x_j^\alpha ,$$

and it is easily seen that $\hat{q}_j = q_j^\dagger$. Thus

$$QQ^* = \sum_{j=1}^{n} q_j q_j^\dagger = \left[\sum_{\alpha=1}^{4} \sum_{j=1}^{n} (x_j^\alpha)^2 \right] 1_2 > a^2 1_2$$

or, since the projective geometry is invariant to $q_j \to a q_j$,

$$QQ^* - 1_2 > 0 \tag{III. 4}$$

is the space of interest. Apply now a permutation to bring Q to the form

$$\begin{bmatrix} z^1 & z^2 \\ -\bar{z}^2 & \bar{z}^1 \end{bmatrix},$$

where z^α is the αth row of Z and is a $1 \times n$ vector with complex entries. Define the matrix

$$J_n = \begin{bmatrix} 0 & 1_n \\ -1_n & 0 \end{bmatrix} \tag{III. 5}$$

and note that

$$Q^* = J_n Q' J_1', \quad Q = J_1 \bar{Q} J_n', \tag{III. 6}$$

which will be used shortly.

Use of the Eq. (III. 4) alone gives the group of automorphisms

$$\Gamma_q H_q \Gamma_q^* = H_q, \qquad \Gamma_q^* H_q \Gamma_q = H_q, \tag{III. 7a}$$

with

$$H_q = \begin{bmatrix} 1_2 & 0 \\ 0 & -1_{2n} \end{bmatrix} \tag{III. 7b}$$

and where

$$\Gamma_q = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

which acts upon Q by means of linear fractional transformations

$$Q_1 = (AQ + B)(CQ + D)^{-1} = (QB^* + A^*)^{-1}(QD^* + C^*). \tag{III. 8}$$

However, the structure of Γ_q is not as simple as Eqs. (III. 7) suggest owing to Eq. (III. 6). Use of Eq. (III. 6) in conjunction with Eq. (III. 8) gives $J_1 \bar{A} J_1' = A$, $J_1 \bar{B} J_n' = B$, $J_n \bar{C} J_1' = C$, and $J_n \bar{D} J_n' = D$, so that Γ_q is the symplectic group $Sp(1, n)$ satisfying

$$\Gamma_q K_{1,n} \Gamma_q^* = \Gamma_q^* K_{1,n} \Gamma_q = K_{1,n} \tag{III. 9a}$$

where

$$K_{1,n} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1_n & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1_n \end{bmatrix} \tag{III. 9b}$$

and where Γ has the structure

$$\Gamma = \begin{bmatrix} a_1 & B_1 & a_2 & B_2 \\ C_1 & D_1 & C_2 & D_2 \\ -\bar{a}_2 & -\bar{B}_2 & a_1 & \bar{B}_1 \\ -\bar{C}_2 & -\bar{D}_2 & \bar{C}_1 & \bar{D}_1 \end{bmatrix}. \tag{III. 9c}$$

Here a_i is a complex number, B_i is a $1 \times n$ complex vector, and D_i is an $n \times n$ complex matrix. The space $QQ^* - 1 > 0$ may be identified[2,3] with $Sp(1, n)/Sp(1) \times Sp(n)$; it is called the quaternion hyperbolic space.

The three different spaces are homeomorphic to one another, since there are obviously unique mappings with unique inverses that take one space into another. Hence, the spaces are topologically equivalent. However, the groups of automorphisms for large n are quite different, there being more parameters in $Sp(1, n)$ than in $U(2, n)$ and more in $U(2, n)$ than in $O(4, n)$ [$(n+1)(2n+3)$, $(n+2)^2$, and $(n+4)(n+3)/2$ real parameters respectively[3]].

For small n there are some special isomorphisms,[3] e.g., $su(2,2) \approx so(2,4)$ between Lie algebras, that may be especially interesting. Some of these are given in Appendix A. However, in these cases the literal interpretation of homogeneous coordinates demands that the dimensionality of the projective space be one less than the number of particles. Thus, a system consisting of but one particle has no dimensions; a space consisting of two particles has one dimension (the distance between them), and such a space could be represented with homogeneous coordinates as $O(2,2)/O(2)\times O(2)$, as $U(1,2)/U(1)\times U(2)$, or as $Sp(1,1)/Sp(1)\times Sp(1)$; a space of three particles defines a plane, and the only permissible representation is $O(3,3)/O(3)\times O(3)$. It is only when there are four or more particles that three-dimensional space takes form. [Of course, there is nothing to prevent academic consideration of spaces such as $O(2,n)/O(2)\times O(n)$ and $O(3,n)/O(3)\times O(n)$ where all particles are constrained to lie on a line or a plane respectively.] Hence, the groups that have been of recent importance in particle physics,[8-10] e.g., SO(4, 2) and SU(6), could be considered to be approximate symmetries and/or subgroups of $O(4,n)$, $U(2,n)$, or $Sp(1,n)$ as appropriate. These dimensional approximations notwithstanding, it will later be shown useful to consider P^3 spaces containing fewer than four particles so as to interpret the theory.

Since the three spaces considered are substantially different from one another, it seems that one should be able to make a definite statement as to which is the best representation for a system. However, each space has properties which recommend it over others depending on the point of view. For example, crystalline solids may well be best represented by $O(4,n)/O(4)\times O(n)$ since the structure of the crystallographic groups are apparent in this case. On the other hand, the success of SU(2), $SU(2)\times U(1)$, SU(3), and SU(6) in particle physics suggests the importance of $U(2,n)/U(2)\times U(n)$. Finally, the particles comprising $Sp(1,n)/Sp(1)\times Sp(n)$ may be moved independently of one another while the origin of Q is fixed (as can be seen from the first version of Q, upon which the subgroup $Sp(1)\times Sp(1)\times\cdots\times Sp(1)$, n factors, acts on the right), and so this space seems to permit of a greater variety of configurations than do the real and complex spaces. Perhaps different systems can be metrically different from one another even though they are topologically identical. However, this problem remains to be resolved.

In the following the differentials dX, dZ, and dQ will be encountered, and these matrices will consist of $4n$ infinitesimals. There are two ways to interpret this operation. We might first assert that the fourth row of X contains hidden variables, so that $dm_i^{1/2}$ is in fact $d(m_i^{1/2}y_i^4)$. Thus, the operator d corresponds to an infinitesimal change of configuration as observed from a fixed frame. Alternatively, one may consider the configuration to be fixed, and the frame to be moving, in which case hidden variables are unnecessary, and $dm_i^{1/2}$ is nonzero owing to relativistic effects. In such a case the group of automorphisms can be understood as a transformation from one moving frame to another.

IV. METRIZATION OF THE SPACES[11]

Since the group of automorphisms of X, Z, and Q act in the same manner, it is possible to develop the metric for the generic case. For this purpose, let M be the $\rho\times n$ matrix X, Z, or Q, which is transformed to

$$M_1 = (AM+B)(CM+D)^{-1} = (M\tilde{B}+\tilde{A})^{-1}(M\tilde{D}+\tilde{C}), \quad \text{(IV. 1)}$$

where $\tilde{A}=A'$, etc., or $\tilde{A}=A^*$, etc., as required by the nature of the groups $O(4,n)$, $U(2,n)$, or $Sp(1,n)$. Now,

$$M_1\tilde{M}_1 - \mathbf{1} = (M\tilde{B}+\tilde{A})^{-1}(M\tilde{M}-\mathbf{1})(B\tilde{M}+A)^{-1} \quad \text{(IV. 2a)}$$

and

$$\tilde{M}_1 M_1 - \mathbf{1} = (\tilde{M}\tilde{C}+\tilde{D})^{-1}(\tilde{M}M-\mathbf{1})(CM+D)^{-1}. \quad \text{(IV. 2b)}$$

We also have

$$dM_1 = AdM(CM+D)^{-1} - (AM+B)(CM+D)^{-1}CdM(CM+D)^{-1}$$
$$= (M\tilde{B}+\tilde{A})^{-1}dM(CM+D)^{-1} \quad \text{(IV. 3)}$$

on use of Eqs. (IV. 1), (III. 7), (III. 2), and (II. 5). The differential operator

$$\partial_M = \begin{bmatrix} \partial/\partial m_1^1 & \partial/\partial m_2^1 & \cdots & \partial/\partial m_n^1 \\ \partial/\partial m_1^2 & \partial/\partial m_2^2 & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \\ \partial/\partial m_1^\rho & \partial/\partial m_2^\rho & \cdots & \partial/\partial m_n^\rho \end{bmatrix} \quad \text{(IV. 4)}$$

transforms according to

$$\partial'_{M_1} = (CM+D)\partial'_M(M\tilde{B}+\tilde{A}) \quad \text{(IV. 5)}$$

since $\mathrm{Tr}(dM\partial'_M)$ must be the invariant operator d. In Eq. (IV. 5) and others to follow, ∂'_M does not act on coordinates to the right; only formal matrix multiplication is intended.

The distance function

$$ds^2 = \mathrm{Tr}[(M\tilde{M}-1)^{-1}dM(\tilde{M}M-1)^{-1}d\tilde{M}] \quad \text{(IV. 6)}$$

is the only function which can be constructed on the spaces which is both quadratic in the dm_i^α and is invariant to the action of $O(4,n)$, $U(2,n)$, or $Sp(1,n)$. The corresponding Laplace–Beltrami operator is

$$\Delta = \mathrm{Tr}[(M\tilde{M}-1)\bar{\partial}_M(\tilde{M}M-1)\partial'_M], \quad \text{(IV. 7)}$$

where $\bar{\partial}_M$ is the complex conjugate of ∂_M in case M is Z or Q, and is ∂_X for $M=X$. It will later prove convenient to multiply the right-hand sides of Eqs. (IV. 6) and (IV. 7) by $\frac{1}{2}$ for the quaternion spaces.

There is a multitude of other invariants that can be constructed. For example, the eigenvalues of the cross ratio[12]

$$R(a,b,c,d) = (M_a\tilde{M}_b - 1)(M_c\tilde{M}_b - 1)^{-1}$$
$$\times(M_c\tilde{M}_d - 1)(M_a\tilde{M}_d - 1)^{-1}$$

are invariants, where M_a, M_b, etc., are different points in the configuration space.

The invariant volume elements of the three spaces are likewise obtained by construction. From Eq. (IV. 3), the Jacobian of the map of $m\times n$ matrices X to X_1 is[11]

$$|XB'+A'|^{-n}|CX+D|^{-m},$$

which is easily seen if one writes dX as the row $(dx_1^1, dx_2^1, \ldots, dx_n^1, dx_1^2, \ldots, dx_n^m) = d\hat{X}$. Then $P'dXQ = d\hat{X}(P\otimes Q)$, and $|P\otimes Q| = |P|^n|Q|^m$. Hence the invariant volume element dV_r of the real space is

$$dV_r = |XX'-1|^{-n/2}|X'X-1|^{-m/2}\dot{X} = |XX'-1|^{-(n+m)/2}\dot{X} \quad \text{(IV. 8a)}$$

where $\dot{X} = dx_1^1 dx_1^2 \cdots dx_n^m$ and where use has been made of Eqs. (IV. 2). For $p\times n$ complex matrices Z, the Jacobian of the map (IV. 1) is

$$\|ZB^* + A^*\|^{-2p}\|CZ+D\|^{-2n},$$

where $\|S\|$ is the absolute value (modulus) of the determinant $|S|$. Hence, the invariant volume element dV_c of the complex space is

$$dV_c = |ZZ^*-1|^{-(n+p)}\dot{Z}, \quad \text{(IV. 9a)}$$

where $\dot{Z} = \dot{X}$. The space of n quaternions has as invariant volume element dV_q the quantity

$$dV_q = |QQ^*-1|^{-(n+1)}\dot{Q} \quad \text{(IV. 10a)}$$

with $\dot{Q} = \dot{X}$. In all of the above, use is made of $\|\tilde{M}M - 1\| = \|M\tilde{M} - 1\|$.

We shall now obtain the polar forms of Eq. (IV. 6). There are considerable differences between the several spaces that become apparent when the metrics are expressed in this alternative fashion.

Any $m\times n$ $(n \geq m)$ X may be written as[11,13]

$$X = P'(\Lambda, 0)Q = P'\Lambda U, \quad \text{(IV. 11)}$$

where $PP' = P'P = \mathbf{1}_m$, $QQ' = Q'Q = \mathbf{1}_n$, Λ is an $m\times m$ diagonal matrix, and U is the first m rows of Q. Then

$$(XX'-1)^{-1} = P'(\Lambda^2-1)^{-1}P,$$
$$(X'X-1)^{-1} = Q\begin{bmatrix}(\Lambda^2-1)^{-1} & 0\\ 0 & -1\end{bmatrix}Q'. \quad \text{(IV. 12)}$$

The orthogonal matrix P depends upon $m(m-1)/2$ variables, Λ contains m variables, and U contains $mn - m(m+1)/2$ variables since $UU' = \mathbf{1}_m$; altogether there are mn variables as required. We have

$$dX = dP'(\Lambda, 0)Q + P'(d\Lambda, 0)Q + P'(\Lambda, 0)\,dQ,$$
$$PdXQ' = (d\Lambda, 0) + (\Lambda, 0)\delta Q - \delta P(\Lambda, 0), \quad \text{(IV. 13a)}$$

where $\delta Q = dQQ' = -QdQ'$ and $\delta P = -P\,dP' = dPP'$ are antisymmetric since P and Q are orthogonal matrices. Let

$$\delta Q = \begin{bmatrix}\delta A & \delta B\\ -\delta B' & \delta C\end{bmatrix},$$

where $\delta A = -\delta A'$ is $m\times m$, δB is $m\times(n-m)$, and $\delta C = -\delta C'$ is $(n-m)\times(n-m)$, so that

$$PdXQ' = (d\Lambda + \Lambda\delta A - \delta P\Lambda, \Lambda\delta B) = (d\Lambda + \delta\Omega, \Lambda\delta B). \quad \text{(IV. 13b)}$$

The metric equation (IV. 6) is then given by

$$ds^2 = \mathrm{Tr}[(\Lambda^2-1)^{-1}(d\Lambda+\delta\Omega)(\Lambda^2-1)^{-1}(d\Lambda+\delta\Omega') - (\Lambda^2-1)^{-1}\Lambda^2\delta B\delta B'],$$

in which the term in $\delta B\delta B'$ is absent if $m = n$. The matrix $\delta\Omega$ may be expressed uniquely as the sum of symmetric $\Lambda d\theta - d\theta\Lambda$, $d\theta = -d\theta' = (\delta P + \delta A)/2$, and antisymmetric $\Lambda d\phi + d\phi\Lambda$, $d\phi = -d\phi' = (\delta P - \delta A)/2$, parts so that

$$ds^2 = \mathrm{Tr}[(\Lambda^2-1)^{-2}d\Lambda^2 + (\Lambda^2-1)^{-1}(\Lambda d\theta - d\theta\Lambda)(\Lambda^2-1)^{-1} \times(\Lambda d\theta - d\theta\Lambda) - (\Lambda^2-1)^{-1}(\Lambda\,d\phi + d\phi\Lambda)(\Lambda^2-1)^{-1} \times(\Lambda d\phi + d\phi\Lambda) - (\Lambda^2-1)^{-1}\Lambda^2\delta B\delta B']. \quad \text{(IV. 14a)}$$

Use of the substitution $\lambda_\alpha = \coth\xi_\alpha$ finally gives

$$ds^2 = \sum_{\alpha=1}^{m} d\xi_\alpha^2 + 2\sum_{\beta>\alpha}\sinh^2(\xi_\alpha - \xi_\beta)\,d\theta_{\alpha\beta}^2 + \sinh^2(\xi_\alpha+\xi_\beta)\,d\phi_{\alpha\beta}^2 - \sum_{\alpha=1}^{m}\cosh^2\xi_\alpha\sum_{j=1}^{n-m}d\omega_{\alpha j}^2, \quad \text{(IV. 14b)}$$

where the $d\omega_{\alpha j}$ are the elements of δB.

The invariant volume element in polar coordinates is

$$dV_r = \left[\prod_{\beta>\alpha}|\cosh 2\xi_\alpha - \cosh 2\xi_\beta|\,d\theta_{\alpha\beta}\,d\phi_{\alpha\beta}\right] \times\left[\prod_{\alpha=1}^{m}\cosh^{n-m}\xi_\alpha\,d\xi_\alpha\prod_{j=1}^{n-m}d\omega_{\alpha j}\right]. \quad \text{(IV. 8b)}$$

The Laplace—Beltrami operator is calculated from the classical formula $\Delta = g^{-1/2}\partial_\mu(g^{1/2}g^{\mu\nu}\partial_\nu)$; it is not given here owing to the fact that it is a lengthy formula that does not appear to be computationally useful.

Consider now the complex space. Let P^*ZZ^*P be the diagonal matrix $\mathrm{diag}(\lambda_1^2, \lambda_2^2)$. Further, let P_1 be $\mathrm{diag}(\exp i\phi_1, \exp i\phi_2)$. Then $P_1P^*ZZ^*PP_1^{-1} = P^*ZZ^*P$, so that P is the coset $\mathrm{U}(2)/\mathrm{U}(1)\times\mathrm{U}(1)$. Thus, the polar form

$$Z = P\Lambda U,$$

where $UU^* = \mathbf{1}_2$, depends upon $\dim(P) + \dim(\Lambda) + \dim(U) = 2 + 2 + (4n-4) = 4n$ real variables as it should.[11] Calculations similar to those above give for the general complex space the metric

$$ds^2 = \mathrm{Tr}[(\Lambda^2-1)^{-2}d\Lambda^2 + (\Lambda^2-1)^{-1}(\Lambda d\theta - d\theta\Lambda)(\Lambda^2-1)^{-1} \times(\Lambda d\theta - d\theta\Lambda) - (\Lambda^2-1)^{-1}(\Lambda d\phi + d\phi\Lambda)(\Lambda^2-1)^{-1} \times(\Lambda\,d\phi + d\phi\Lambda) - \Lambda^2(\Lambda^2-1)^{-1}\delta B\delta B^*], \quad \text{(IV. 15a)}$$

where $d\theta = -d\theta^*$ and $d\phi = -d\phi^*$ are $p\times p$ skew Hermitian matrices, and δB is a $p\times(n-p)$ complex matrix. With the change of variables $\lambda_\alpha = \coth\xi_\alpha$ this equation gives

$$ds^2 = \sum_{\alpha=1}^{p}d\xi_\alpha^2 + \sinh^2 2\xi_\alpha\,d\phi_\alpha^2 + 2\sum_{\beta>\alpha}\sinh^2(\xi_\alpha-\xi_\beta)\,d\theta_{\alpha\beta}\,d\bar{\theta}_{\alpha\beta} + \sinh^2(\xi_\alpha+\xi_\beta)\,d\phi_{\alpha\beta}\,d\bar{\phi}_{\alpha\beta} - \sum_{\alpha=1}^{p}\cosh^2\xi_\alpha\sum_{j=1}^{n-p}d\omega_{\alpha j}\,d\bar{\omega}_{\alpha j}, \quad \text{(IV. 15b)}$$

where the $d\phi_\alpha$ are the moduli of the diagonal elements of $d\phi$.

The invariant volume element is

$$dV_c = \prod_{\beta>\alpha} (\cosh 2\xi_\alpha - \cosh 2\xi_\beta)^2 \, d\theta_{\alpha\hat{\beta}} \, d\phi_{\alpha\hat{\beta}} \times \prod_{\alpha=1}^{p} \sinh 2\xi_\alpha \cosh^{2(n-p)}\xi_\alpha \, d\xi_\alpha \, d\phi_\alpha \prod_{j=1}^{n-p} d\omega_{\alpha\hat{j}}, \tag{IV.9b}$$

where $d\psi_{a\hat{b}}$ is the product of the real and imaginary parts of $d\psi_{ab}$. The Laplace–Beltrami operator is again a complicated object, and is not given here. Suffice it to say that the angular parts of both the real and complex differential operators yield simply solvable differential equations, but the radial parts are extremely intricate. This is not the case for the quaternion space, to which we now turn.

Since $QQ^* = N^2 \mathbf{1}_2$, where N is the norm of the n-dimensional quaternion Q, one may write

$$Q = NKU,$$

where $UU^* = U^*U = \mathbf{1}_{2n}$, $UJ_nU' = J_n$, $KK' = 1_2$, and $KJ_nK' = J_1$ on account of Eq. (III.6). The $2\times 2n$ matrix K has but two nonzero elements, the locations of which will be later chosen for convenience. We require the polar form of

$$ds^2 = (1/2)(N^2-1)^{-1}\mathrm{Tr}[dQ(Q^*Q-1)^{-1}dQ^*].$$

Now $(dQ)U^* = dNK + NK\delta U$, where $\delta U = dUU^*$ has the structure

$$\delta U = \begin{bmatrix} ida_1 & db_1 & da_2 & db_2 \\ -db_1^* & dD_1 & db_2' & dD_2 \\ -d\bar{a}_2 & -d\bar{b}_2 & -ida_1 & d\bar{b}_1 \\ -db_2^* & -d\bar{D}_2 & -db_1' & -dD_1' \end{bmatrix},$$

in which a_1 is a real parameter, b_1 is a $1\times(n-1)$ complex vector, and a_2 is a complex parameter. It is convenient to choose K such that $U(Q^*Q-1)^{-1}U^* = (N^2K'K-1)^{-1}$ conforms; thus

$$(N^2K'K-1)^{-1} = \begin{bmatrix} (N^2-1)^{-1} & 0 & 0 & 0 \\ 0 & -1_{n-1} & 0 & 0 \\ 0 & 0 & (N^2-1)^{-1} & 0 \\ 0 & 0 & 0 & -1_{n-1} \end{bmatrix}$$

and

$$K = \begin{bmatrix} 1 & 0_{n-1} & 0 & 0_{n-1} \\ 0 & 0_{n-1} & 1 & 0_{n-1} \end{bmatrix}.$$

The matrix $K\delta U$ has the form

$$\begin{bmatrix} ida_1 & db_1 & da_2 & db_2 \\ -d\bar{a}_2 & -d\bar{b}_2 & -ida_1 & d\bar{b}_1 \end{bmatrix}.$$

These quantities give

$$ds^2 = (N^2-1)^{-2}[dN^2 + N^2(da_1^2 + da_2 d\bar{a}_2)] - (N^2-1)^{-1}N^2(db_1 d\bar{b}_1 + db_2 d\bar{b}_2) \tag{IV.16a}$$

or

$$ds^2 = d\xi^2 + \sinh^2\xi\cosh^2\xi \sum_{\alpha=1}^{3} d\sigma_\alpha^2 - \cosh^2\xi \sum_{l=1}^{4(n-1)} d\eta_l^2, \tag{IV.16b}$$

where $d\sigma \to da$ and $d\eta \to db$, and where $N = \coth\xi$.

The volume element is

$$dV_q = \sinh^3\xi\cosh^{4n-1}\xi \, d\xi \prod_{\alpha=1}^{3} d\sigma_\alpha \prod_{l=1}^{4(n-1)} d\eta_l \tag{IV.10b}$$

and the Laplace–Beltrami operator is

$$\Delta = \frac{\partial^2}{\partial\xi^2} + [3\coth\xi + (4n-1)\tanh\xi]\frac{\partial}{\partial\xi} + 4\,\mathrm{csch}^2 2\xi \times \sum_{\alpha=1}^{3}\frac{\partial^2}{\partial\sigma_\alpha^2} - \mathrm{sech}^2\xi \sum_{l=1}^{4(n-1)}\frac{\partial^2}{\partial\eta_l^2}. \tag{IV.17}$$

This naive rendition of the metric misses the important fact that $\sum d\sigma_\alpha^2$ and $\sum d\eta_l^2$ are not simply the metrics on R^3 and $R^{4(n-1)}$. We have instead $\sum d\sigma_\alpha^2$ isomorphic to the metric on Sp(1), and $\sum d\eta_l^2$ isomorphic to the metric on $\mathrm{Sp}(n)/\mathrm{Sp}(1)\times\mathrm{Sp}(n-1)$. The same considerations apply to the real and complex spaces.

Let $A \in \mathrm{Sp}(1)$; then

$$\sum d\sigma_\alpha^2 = \mathrm{Tr}(dA\, dA^*).$$

Represent[14] A by

$$\begin{bmatrix} \cos(\omega/2)\exp[-i(\theta+\phi)/2] & \sin(\omega/2)\exp[-i(\theta-\phi)/2] \\ -\sin(\omega/2)\exp[i(\theta-\phi)/2] & \cos(\omega/2)\exp[i(\theta+\phi)/2] \end{bmatrix},$$

so that

$$\sum d\sigma_\alpha^2 = \tfrac{1}{4}(d\theta^2 + d\phi^2 + d\omega^2 + 2\cos\omega\, d\theta\, d\phi)$$

or

$$d\sigma_1 = \tfrac{1}{2}(d\theta + \cos\omega\, d\phi),$$
$$d\sigma_2 = \tfrac{1}{2}\sin\omega\, d\phi,$$
$$d\sigma_3 = \tfrac{1}{2}\, d\omega.$$

Furthermore,

$$\sum\frac{\partial^2}{\partial\sigma_\alpha^2} = 4\left[\frac{\partial^2}{\partial\omega^2} + \cot\omega\frac{\partial}{\partial\omega} + \csc^2\omega\left(\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial\phi^2} - 2\cos\omega\frac{\partial^2}{\partial\theta\partial\phi}\right)\right]$$

is a well-known operator encountered in the rotational Hamiltonian. The metric on the compact space $\mathrm{Sp}(n)/\mathrm{Sp}(1)\times\mathrm{Sp}(n-1)$ is left as an exercise.

V. GEODESICS AND CURVATURE

The connection forms and curvature tensors for the three spaces follow from equations for parallel displacement of a vector. Let $F = (f^{\alpha l})$ be a contravariant vector with components in the real, complex, or quaternion space. The change of basis, Eqs. (IV.1) and (IV.3), gives

$$F_1 = (M\tilde{B} + \tilde{A})^{-1}F(CM + D)^{-1}, \tag{V.1}$$

so that

$$F\cdot F = \mathrm{Tr}[(M\tilde{M}-1)^{-1}F(\tilde{M}M-1)^{-1}\tilde{F}] \tag{V.2}$$

is an invariant. The scalar $F\cdot F$ is not altered on transport of F along a geodesic, so that

$$d(F\cdot F)/ds = 0.$$

This equation is satisfied identically if

$$\dot{F} = F(\tilde{M}M - 1)^{-1}\tilde{M}\dot{M} + \dot{M}\tilde{M}(M\tilde{M} - 1)^{-1}F$$
$$= F\tilde{M}(M\tilde{M} - 1)^{-1}\dot{M} + \dot{M}\tilde{M}(M\tilde{M} - 1)^{-1}F, \quad (V.3)$$

in which $\dot{F} = dF/ds$ and where the last equality in (V.3) follows from $\tilde{M}(M\tilde{M} - 1) = (\tilde{M}M - 1)\tilde{M}$. The Christoffel symbol is thus obtained as[12,15]

$$\Gamma^{\alpha i}_{\beta j \gamma k} = -(t_{j\gamma}\delta^{\alpha}_{\beta}\delta^{i}_{k} + t_{k\beta}\delta^{\alpha}_{\gamma}\delta^{i}_{j}), \quad (V.4a)$$

where

$$T = (t_{j\gamma}) = \tilde{M}(M\tilde{M} - 1)^{-1}. \quad (V.4b)$$

The equations for geodesics[15] are

$$\ddot{M} - 2\dot{M}T\dot{M} = 0, \quad (V.5)$$

which can be obtained by a somewhat tedious calculation based on the variational principle $\delta s = 0$. Alternatively, Eq. (V.5) follows from the classical formula $\ddot{X}^{\mu} + \Gamma^{\mu}_{\alpha\beta}\dot{X}^{\alpha}\dot{X}^{\beta} = 0$ on replacing α by αi, etc.

Since the point $(\Lambda, 0)$ in the real or complex space or the point NK in the quaternion space can be moved to any other respective point by a linear fractional transformation, it is sufficient to solve Eq. (V.5) at these points. Hence, all geodesics are images of

$$[\Lambda, 0] = \begin{bmatrix} \coth\eta_1 s & 0 & 0 & 0 & \\ 0 & \coth\eta_2 s & 0 & 0 & \\ 0 & 0 & \coth\eta_3 s & 0 & , \quad 0 \\ 0 & 0 & 0 & \coth\eta_4 s & \end{bmatrix} \quad (V.6a)$$

for the real space, where $\sum \eta_{\alpha}^2 = 1$, or of

$$[\Lambda, 0] = \begin{bmatrix} \coth\zeta_1 s & 0 & \\ & & , \quad 0 \\ 0 & \coth\zeta_2 s & \end{bmatrix} \quad (V.6b)$$

for the complex space, where $\sum \xi_{\alpha}^2 = 1$, or of

$$NK = \coth s K \quad (V.6c)$$

for the quaternion space. In this form the geodesics have a singular origin which corresponds to formation of the system. It will later be demonstrated that the arc length s is essentially identified with physical time t.

The curvature tensor $R_{\alpha i \beta j} = R^{\gamma k}_{\alpha i \gamma k \beta j}$ on the real space is obtained from the classical formula as

$$R_{\alpha i \beta j} = -(n+2)[(XX' - 1)^{-1}]_{\alpha\beta}[(X'X - 1)^{-1}]_{ij}; \quad (V.7a)$$

that on the complex space follows from results on Kähler manifolds[6] and is given by

$$R_{\alpha i \beta \bar{j}} = -(n+2)[(ZZ^* - 1)^{-1}]_{\alpha\beta}[(Z^*Z - 1)^{-1}]_{ij}; \quad (V.7b)$$

for the quaternion hyperbolic space

$$R_{\alpha i \beta j} = -(n+4)[(QJ_nQ' - J_1)^{-1}]_{\alpha\beta}[(Q'J_1'Q - J_n')^{-1}]_{ij}. \quad (V.7c)$$

The spaces are Einstein manifolds of constant negative sectional curvature. It should not be considered strange that the constants in Eq. (V.7a) and (V.7b) are different from that of the quaternion space, Eq. (V.7c); topological equivalence is not sufficient to guarantee equivalent Ricci tensors.

VI. LIE ALGEBRAS

The infinitesimal generators of the Lie algebra so(4, n) are

$$M_{\alpha\beta} = x_{\alpha l}\partial_{\beta l} - x_{\beta l}\partial_{\alpha l} \sim X\partial' - (X\partial')',$$
$$M_{ij} = x_{\mu i}\partial_{\mu j} - x_{\mu j}\partial_{\mu i} \sim X'\partial - (X'\partial)',$$
$$M_{\alpha i} = x_{\alpha l}x_{\mu i}\partial_{\mu l} - x_{\alpha i} - \partial_{\alpha i} \sim X(X'\partial)' - X - \partial,$$

in which the summation convention on repeated indices is used in the first equations and where matrix notation is also given. These operators commute with the Laplace—Beltrami operator, and so are conserved. The total (spin plus isospin) angular momentum operators are the $M_{\alpha\beta}$; the M_{ij} are internal symmetry operators; the $M_{\alpha i}$ are related to dipole transitions and to the energy, if extrapolation from the work of Barut and Kleinert[16] is warranted.

For su(2, n) the operators commuting with the Laplace—Beltrami operator are

$$L_{\alpha\beta} = z_{\alpha l}\partial_{\beta l} - \bar{z}_{\beta l}\bar{\partial}_{\alpha l} \sim Z\partial' - (Z\partial')^*,$$
$$L_{ij} = \bar{z}_{\mu j}\bar{\partial}_{\mu i} - z_{\mu i}\partial_{\mu j} \sim (Z'\partial)^* - Z'\partial,$$
$$L_{\alpha i} = z_{\alpha l}z_{\mu i}\partial_{\mu l} - \bar{\partial}_{\alpha i} \sim Z(Z'\partial)' - \bar{\partial},$$

where $\bar{\partial}_{\alpha i} = \partial/\partial\bar{z}_{\alpha i}$. The generators of sp(1) and sp(n) are given as

$$J_1Q\partial' - (Q\partial')'J_1 \quad \text{and} \quad J_nQ'\partial - (Q'\partial)'J_n,$$

but the noncompact part of sp(1, n) has so far eluded the author.

VII. INTERPRETATION

The introduction of projective coordinates for representation of the states of a system has been accomplished without reference to the temporal coordinate. It is clearly necessary to uncover the nature of arc length, to see if there is perhaps some connection between this parameter and duration. For this purpose, it is necessary to consider a system composed of a single particle in P^3, the projective space of three dimensions. As noted previously, this is inconsistent with geometry; however, it is consistent with the usual physical representation of the world line of a single particle by a four-dimensional manifold. That is, the symmetric space O(4, 1)/O(4)×O(1) should correspond in some way to an established precedent.

Let

$$X = (x_1, x_2, x_3, x_4) = \sigma u, \quad \sigma^2 = XX', \quad uu' = 1.$$

Then, since

$$(X'X - 1)^{-1} = X'(XX' - 1)^{-1}X - 1 = (\sigma^2 - 1)^{-1}X'X - 1,$$

we have

$$ds^2 = (\sigma^2 - 1)^{-1}\,dX[(\sigma^2 - 1)^{-1}X'X - 1]\,dX'$$
$$= (\sigma^2 - 1)^{-2}(XdX')\,(XdX') - (\sigma^2 - 1)^{-1}\,dX\,dX' \quad (VII.1)$$

from Eq. (IV.6). But $dX = d\sigma u + \sigma du$, $udu' = duu' = 0$, so

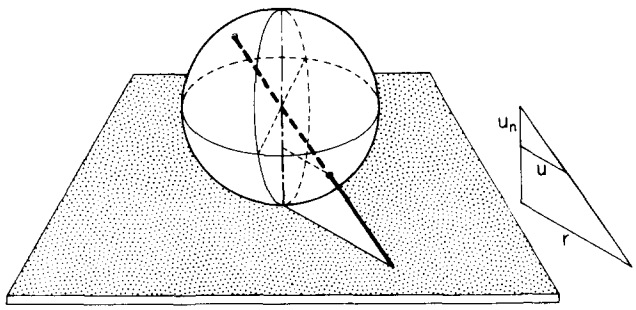


FIG. 1. Antipodal projection of the unit sphere S^{n-1} onto P^{n-1}. Opposing points on the diameter are identified, so that P^{n-1} is isomorphic to $S^{n-1}/\{\pm 1\}$. The auxiliary figure shows that the $(n-1)$-dimensional vectors r and u satisfy the proportion $r/1 = u/u_n$; the u_i are homogeneous coordinates of the projective space P^{n-1}.

that

$$ds^2 = (\sigma^2 - 1)^{-2}\, d\sigma^2 - \sigma^2(\sigma^2 - 1)^{-1}\, du\, du' \qquad \text{(VII. 2a)}$$

or

$$ds^2 = d\xi^2 - \cosh^2\xi\, du\, du' \qquad \text{(VII. 2b)}$$

where $\sigma = \coth\xi$. The points u such that $uu' = 1$ constitute the space S^3; hence $du\, du' = d\omega^2 + \sin^2\omega(d\theta^2 + \sin^2\theta\, d\phi^2)$. Use is now made of the well-known equivalence $P^3 \sim S^3/\{\pm 1\}$; that is, the three-dimensional projective space is topologically equivalent to the three-dimensional sphere with diametrically opposed points identified. As is shown in Fig. 1, the antipodal projection of the sphere onto the plane is one for which

$$y_i/1 = u_i/u_n,$$

and the equivalence is demonstrated. The parametric representation $[\sin\omega\sin\theta\cos\phi, \sin\omega\sin\theta\sin\phi, \sin\omega\cos\theta, \cos\omega]$ of u gives $y \sim \tan\omega[\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]$; in the neighborhood of the poles $\omega = 0$, i.e., for small distances, and for large ξ, the metric Eq. (VII. 2b) becomes

$$ds^2 \approx d\xi^2 - e^{2\xi}[(d\omega/2)^2 + (\omega/2)^2(d\theta^2 + \sin^2\theta\, d\phi^2)]. \qquad \text{(VII. 2c)}$$

The substitutions $\xi = \kappa ct$ and $\kappa r = \omega/2$, where c is the velocity of light, t is time, and κ is a constant having the dimensions of reciprocal length, give

$$(ds/\kappa)^2 \approx c^2\, dt^2 - \exp(2\kappa ct)[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)] \qquad \text{(VII. 3)}$$

which is equivalent[17] to the de Sitter metric. For small t this approximate metric reduces to that of special relativity.

What has been shown is that regions of the space $O(4, 1)/O(4)\times O(1)$ reduce to the matter free space of general relativity with cosmological term $\lambda = 3\kappa^2$. It is eminently reasonable that general relativity correponds to times long after the origin of the system, and to points near the gravitating body. General relativity was devised to explain Newtonian gravitation, which is known with certainty only for local phenomena. Since the transformation $X = \coth(\kappa ct)u$ accomplishes the reduction, and since also all geodesics are images of $X_0 = (\coth s, 0, 0, 0)$ [it may be noted from Eq. (V. 5) that s is defined only to within an equivalence class $s \sim as + b$], it is apparent that s is essentially physical time t.

VIII. SOME SOLUTIONS

Geodesics on the polar form of $Sp(1, n)/Sp(1)\times Sp(n)$ are readily obtained, and prove interesting. The function

$$f^2 = \dot\xi^2 + \sinh^2\xi\cosh^2\xi \sum_1^3 \dot\sigma_\alpha^2 - \cosh^2\xi \sum_1^{4(n-1)} \dot\eta_l^2 = 1$$

together with the Euler equations

$$\frac{d}{ds}\frac{\partial f}{\partial \dot x_a} - \frac{\partial f}{\partial x_a} = 0$$

give

$$\begin{aligned} &\dot\sigma_\alpha = a_\alpha/\sinh^2\xi\cosh^2\xi, \\ &\dot\eta_l = b_l/\cosh^2\xi, \qquad \text{(VIII. 1)} \\ &\dot\xi^2 + a^2/\sinh^2\xi\cosh^2\xi - b^2/\cosh^2\xi = 1, \end{aligned}$$

where $a^2 = \sum a_\alpha^2$ and $b^2 = \sum b_l^2$, all of which are integration constants. The solutions of these equations are

$$\begin{aligned} &\cosh 2\,\xi(s) = r\cosh 2s - b^2, \\ &\eta_l(s) = b_l(a^2 + b^2)^{-1/2}\cot^{-1}(x\coths) + \eta_l(0), \qquad \text{(VIII. 2)} \\ &\sigma_\alpha(s) = a_\alpha a^{-1}\cot^{-1}(y\coths) \\ &\qquad - a_\alpha(a^2 + b^2)^{-1/2}\cot^{-1}(x\coths) + \sigma_\alpha(0), \end{aligned}$$

with

$$\begin{aligned} &r^2 = 4a^2 + (b^2 + 1)^2, \\ &x^2 = (r - b^2 + 1)/(r + b^2 - 1), \\ &y^2 = (r - b^2 - 1)/(r + b^2 + 1). \end{aligned}$$

If $a = b = 0$, $\xi = s$; otherwise ξ possesses a minimum $\xi_{\min} = (1/2)\cosh^{-1}(r - b^2)$ at $s = 0$. In terms of the norm N of the quaternion Q, the maximum $N_{\max} = [(r - b^2 + 1)/(r - b^2 - 1)]^{1/2}$. The norm approaches coths asymptotically.

The evolution of the system can be pictured by means of the projective sphere. Particles comprising the system may be represented as points on a sphere of radius N, which points are projected onto the P^3 plane. Initially, the sphere has a very large radius and the points are clustered around the poles. With increasing s, the sphere shrinks, and the particles migrate towards the equator. At infinite time the sphere becomes the unit sphere, and the points become distributed more or less uniformly upon the surface as determined by the initial conditions.

Physical quantities on each of the spaces are assumed to be derivable from functions that at the very least are eigenfunctions of the LB operator with eigenvalue zero. Such harmonic functions can be obtained from the Poisson kernel.[11,18] The kernels are not given here; instead parts of the harmonic functions on $Sp(1, n)/Sp(1)\times Sp(n)$ will be found by more traditional methods. The Laplace–Beltrami operator as given by Eq. (IV. 17) is separable; the radial part gives

$$F'' + [3\coth\xi + (4n - 1)\tanh\xi]F' - 16l(l + 1)\,\mathrm{csch}^2 2\xi F - k\,\mathrm{sech}^2\xi F = 0$$

where $F' = dF(\xi)/d\xi$. Here we have set $\sum \partial^2 F/\partial\sigma_\alpha^2 = -4l(l+1)F$ and $\sum \partial^2 F/\partial\eta_i^2 = kF$. The identification of ξ with t leads one to look for solutions which are finite at finite ξ. The first step to a solution is to eliminate the term in $\operatorname{csch}^2(2\xi)F$ by means of the substitution $F = \tanh^\beta(\xi) f(\xi)$, whereupon one finds $\beta(\beta+2) = 4l(l+1)$, i.e., $\beta = 2l$, $-2(l+1)$, and

$$f'' + \frac{(2\beta+3) + 2(2n+1)\sinh^2\xi}{\sinh\xi\cosh\xi} f' + \frac{4\beta n - k}{\cosh^2\xi} f = 0. \tag{VIII. 3}$$

The term in $\tanh(\xi) f'$ is eliminated by means of $f = g/\cosh^{2n+1}\xi$, and Eq. (VIII. 3) reduces to

$$g'' + \frac{2\beta+3}{\sinh\xi\cosh\xi} g' + \frac{4n(n-1) - k - (2\beta+3)}{\cosh^2\xi} g - (2n+1)^2 g = 0.$$

One may now look for solutions of the form

$$g = \sum_j a_j \cosh^j\xi,$$

for which the indicial and recurrence relations are

$$[4n(n-1) - k - (2\beta+3)]a_0 = 0, \tag{VIII. 4a}$$

$$[4n(n-1) - k]a_1 = 0, \tag{VIII. 4b}$$

$$a_{j+2} = \frac{(2n+1)^2 - j^2}{4n(n-1) - k - (j+1)(j-1-2\beta)} a_j. \tag{VIII. 4c}$$

The odd series terminates at $j = 2n+1$, provided the denominator in Eq. (VIII. 4c) does not vanish. This possibility is eliminated if $\beta = -2(l+1)$, whereupon

$$a_0 = 0,$$

$$k = 4n(n-1),$$

$$a_{j+2} = \frac{j^2 - (2n+1)^2}{(j+1)(j+4l+3)} a_j,$$

or

$$a_{2j+3} = -\frac{(n-j)(n+j+1)}{(j+1)(j+2l+2)} a_{2j+1}, \quad 0 \le j \le n.$$

Thus

$$g(\xi) = \cosh\xi \sum_{j=0}^{n} \frac{(-1)^j (n+j)!\,(2l+1)!}{j!\,(n-j)!\,(2l+j+1)!} \cosh^{2j}\xi,$$

and one complete solution of the ξ equation is

$$F_n^l(\xi) = \coth^{2(l+1)}\xi \sum_{j=0}^{n} \frac{(-1)^j (n+j)!}{j!\,(n-j)!\,(2l+1+j)!} \operatorname{sech}^{2(n-j)}\xi. \tag{VIII. 5}$$

The other solution is obtained upon eliminating the term in $\operatorname{sech}^2(\xi)F(\xi)$ by means of the substitution $F(\xi) = \cosh^\alpha(\xi) f(\xi)$. Standard methods give

$$G_n^l(\xi) = (\cosh\xi)^{-2(n-l-1)} (\sinh\xi)^{2l} \times \sum_{j=0} j! \frac{\Gamma(n/2 + l + 1 + j)(\sinh 2\xi)^{2j}}{(2l+1+j)!\,\Gamma((n+1)/2 - l - j)}$$

which is linearly independent of $F_n^l(\xi)$ in Eq. (VIII. 5). This solution is a finite polynomial if the maximum value for l is $(n-1)/2$ and if l is half-integral for even n or integral for odd n.

The solutions on Sp(1) of the equation[19]

$$\frac{\partial^2\psi}{\partial\omega^2} + \cot\omega\frac{\partial\psi}{\partial\omega} + \csc^2\omega\left(\frac{\partial^2\psi}{\partial\theta^2} + \frac{\partial^2\psi}{\partial\phi^2} - 2\cos\omega\frac{\partial^2\psi}{\partial\theta\partial\phi}\right) = -l(l+1)\psi$$

are of the form

$$\psi = \exp\{i[\tfrac{1}{2}(j+k)\theta + \tfrac{1}{2}(j-k)\phi]\}\,\Omega(\omega),$$

where

$$\Omega'' + \cot\omega\Omega' - \left(\frac{j^2/4}{\cos^2\omega/2} + \frac{k^2/4}{\sin^2\omega/2}\right)\Omega + l(l+1)\Omega = 0.$$

The choice $j + k \ge 0$ is tantamount to orienting the coordinate frame on P^3; for initial inspection choose both j and k positive. It is clear that the solutions of the ξ equation allow l to be either integral or half-integral.

The substitution

$$\Omega = \cos^j\omega/2 \sin^k\omega/2 f(\omega)$$

gives

$$f'' + [(k+1/2)\cot\omega/2 - (j+1/2)\tan\omega/2]f' + (l-\mu)(l+\mu+1)f = 0,$$

where $\mu = (j+k)/2$. The two solutions of this equation are

$$f(\omega) = \sum_n (-1)^n \frac{j!\,(l-\mu)!\,(n+l+\mu)!}{n!\,(n+j)!\,(l+\mu)!\,(l-\mu-n)!} \cos^{2n}\omega/2$$

and

$$f(\omega) = \sum_n (-1)^n \frac{k!\,(l-\mu)!\,(n+l+\mu)!}{n!\,(n+k)!\,(l+\mu)!\,(l-\mu-n)!} \sin^{2n}\omega/2.$$

Hence, $l-\mu$ is integral and ≥ 0 if the polynomials are finite, and j and k are integral. The solutions on the angles θ, ϕ, and ω can have period $\ne 2\pi$ owing to the fact that the projective sphere is doubly connected. Solutions for $l = 0$ are infinite series, one of which (cosine series) diverges at the origin of P^3 and the other of which (sine series) diverges at the P^3 infinity. States of any value of μ appear to be allowed, and may correspond to excited states.

The most interesting aspect of these solutions is that integral l and half-integral l series fall into separate categories. Since there are two allowed solutions for each l, $\mu = (j+k)/2$ and $\mu' = (j-k)/2$, there are $2(l+1)^2$ solutions for each integral l. These are elementary solutions. For half-integral l the solutions number $4(l = 1/2)$, $12(l = 3/2)$, $24(l = 5/2)$, etc. Might not these be leptons (4), a quartet of tricolored quarks (12), etc. ? Table I contains a list of what might be called ground state solutions for j, $k > 0$ and for the first few values of l.

These solutions form the basis for further analysis, which is not pursued here. Suffice it to say at this point that particles with half-integral spin are comprehensible in terms of projective geometry. Wave-particle duality and spin find a natural home in projective geometry. The projective space $P^3 \sim S^3/\{\pm 1\} \approx SO(3)$ is not simply connected,[7] and is somewhat more difficult to imagine than is the Euclidean space E^3. However, this small extension from E^3 to P^3 holds promise of expanding our understanding of the physical world. There are many

TABLE I. Allowed quantum numbers from the Sp(1) equation.

			Integral series		
l	$1/2(j+k)$	j	k	$1/2(j-k)$	Number of solutions
0	0	0	0	0	2[a]
1	0	0	0	0	
	1	2	0	1	
		1	1	0	
		0	2	−1	8
2	0	0	0	0	
	1	2	0	1	
		1	1	0	
		0	2	0	
	2	4	0	2	
		3	1	1	
		2	2	0	
		1	3	−1	
		0	4	−2	18
L	$0 \le \mu \le L$	$0 \le j \le 2L$	$2L-j$	$-\mu \le \mu' \le \mu$	$2(L+1)^2$

			Half-integral series		
l	$1/2(j+k)$	j	k	$1/2(j-k)$	Number of solutions
1/2	1/2	1	0	1/2	
		0	1	−1/2	4
3/2	1/2	1	0	1/2	
		0	1	−1/2	
	3/2	3	0	3/2	
		2	1	1/2	
		1	2	−1/2	
		0	3	−3/2	12
5/2	1/2	1	0	1/2	
		0	1	−1/2	
	3/2	3	0	3/2	
		2	1	1/2	
		1	2	−1/2	
		0	3	−3/2	
	5/2	5	0	5/2	
		4	1	3/2	
		3	2	1/2	
		2	3	−1/2	
		1	4	−3/2	
		0	5	−5/2	24
L	$1/2 \le \mu \le L$	$0 \le j \le 2L$	$2L-j$	$-\mu \le \mu' \le \mu$	$(2L+1)(L+3/2)$

[a] $l=0$ is a special case, for which more solutions are allowed than is indicated.

difficult problems to be resolved; these include the formulation of suitable norms over infinite volumes, investigation of the topology of the spaces, and classification of solutions in terms of the eigenvalues of the infinitesmal generators of the Lie algebras. However, the fact that one arrives at the mathematical structure found here from two very different viewpoints[20] lends support to the theory.

Note added in proof: The Laplace–Beltrami operator Eq. (IV. 7) is correct for the complex space, but not for the real and quaternion spaces. The correct operators are

$$\Delta = \mathrm{Tr}[(XX' - 1)\partial(X'X - 1)\partial'] + 2\mathrm{Tr}[(XX' - 1)X\partial']$$

and

$$\Delta = \mathrm{Tr}[(QQ^* - 1)\bar{\partial}(Q^*Q - 1)\partial'] - 2\mathrm{Tr}[(QQ^* - 1)Q\partial'],$$

respectively. The generators of the noncompact parts of the groups are in all cases $M(M'\partial)' - \bar{\partial}$, thus correcting the operators for $so(4,n)$ and adding to those of $sp(1,n)$. In Eq. (V. 7c) the coefficient $(n+4)$ should be changed to $2(n+2)$, and the last sentence of the paragraph in which this equation appears can be deleted.

ACKNOWLEDGMENTS

I am grateful to Professor Ramesh Gangolli for his helpful answers to several questions concerning symmetric spaces and to Professor Verner Schomaker who increased my understanding of group theory.

APPENDIX

Some Lie group and algebra isomorphisms and homeomorphisms are listed here for convenience. All but one are taken from Helgason.[3] The symbol ≈ denotes topological isomorphsism, and ~ represents a homeomorphism. Groups are indicated by capital letters; their corresponding algebras are given small letters:

$$U(n) \sim SU(n) \times U(1),$$

$$S[U(m) \times U(n)] \sim SU(m) \times U(1) \times SU(n),$$

$$Sp(m,n) \cap U(2m+2n) \approx Sp(m) \times Sp(n),$$

$$SO(n) \approx Spin(n)/F,$$

$$su(2) \approx so(3) \approx sp(1),$$

$so(5) \approx sp(2)$,

$so(4) \approx so(3) \times so(3) \approx su(2) \times su(2) \approx sp(1) \times sp(1)$,

$so(3, 1) \approx sl(2, C)$,

$so(6) \approx su(4)$,

$so(4, 2) \approx su(2, 2)$,

$so(3, 3) \approx sl(4, R)$.

The group Spin(n) and the discrete subgroup F are defined by Chevalley, Ref. 7.

[1](a) O. Schrier and E. Sperner, *Projective Geometry of n Dimensions* (Chelsea, New York, 1961); (b) O. Veblen and J.W. Young, *Projective Geometry* (Ginn, New York, 1938).
[2]J.A. Wolf, *Spaces of Constant Curvature* (Publish or Perish, Boston, 1974), pp. 377—78.
[3]The notation for Lie groups and their algebras here coincides with that of S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[4]W.M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry* (Academic, New York, 1975), p. 60.
[5]Ref. 3, p. 311.
[6]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Vols. I and II* (Interscience, New York, 1963).
[7]C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, N.J., 1946), pp. 16—24.
[8]A.O. Barut and R.B. Haugen, Ann. Phys. (N.Y.) **71**, 519 (1972).
[9]A. DeRújula, H. Georgi, and S.L. Glashow, Phys. Rev. D **12**, 147 (1975).
[10]R. Penrose and M.A.H. MacCallum, Phys. Reports **6** (4), 242 (1972).
[11]L.K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*, Trans. of Math. Monographs, Vol. VI (American Mathematical Society, Providence, R.I., 1963).
[12]C.L. Siegel, *Symplectic Geometry* (Academic, New York, 1964).
[13]A. Ben-Israel and T.N. E. Greville, *Generalized Inverses: Theory and Applications* (Wiley-Interscience, New York, 1974), p. 244.
[14]E.P. Wigner, *Group Theory* (Academic, New York, 1959), Chap. 15.
[15]L.K. Hua, Amer. J. Math. **66**, 470 (1944).
[16]A.O. Barut and H. Kleinert, Phys. Rev. **160**, 1149 (1967).
[17]H.P. Robertson, Phil. Mag. **5**, 835 (1925).
[18]S. Helgason, in *Battelle Rencontres*, edited by C.M. DeWitt and J.A. Wheeler (Benjamin, New York, 1968).
[19]I.M. Gel'fand, R.A. Minlos, and Z.Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Macmillan, New York, 1963).
[20]B.E. Eichinger, Found. Phys., in press.

Polarization theorem for diffractive excitation at large momentum transfer

Alfred Actor

Abteilung Physik, Universität Dortmund, W. Germany
(Received 17 January 1973; revised manuscript received 15 October 1976)

The kinematics of two-body collisions with large s and $|t|$ favors small helicity states (0 for bosons, $\pm 1/2$ for fermions) over all larger helicities (supposing one or more particles with spin ≥ 1 are involved in the collision). We state and prove this result in the form of a theorem: If there exist no constraints among invariant amplitudes in the kinematic region $s \gg |t| \gg$ (mass)^2 for a two-body process with arbitrary spins, then only the smallest helicity states are important. This is a purely kinematic result, and dynamics can cause physical scattering amplitudes to have different properties, by providing constraints among the invariant amplitudes so as to evade the theorem. Nevertheless, the result is useful in that it clarifies the nature of two-body spin structure when $|t|$ is large, and enables one to isolate the purely dynamical conditions among the invariant amplitudes which must be satisfied if there is no "kinematic polarization" of the nature described. In diffractive excitation of higher spin states, sufficiently large values of s and $|t|$ can be reached to make the theorem of practical interest.

I. INTRODUCTION

In the kinematic region where momentum transfer $|t|$ is much larger than any (mass)^2 and s is much larger than $|t|$ there is practically no difference between the s- and t- channel amplitudes describing a two-body reaction. Up to a phase they coincide,

$$|f^{(s)}_{cdab}| \approx \delta_{-a,A}\delta_{-d,D}|f^{(t)}_{cADb}|. \tag{1}$$

[The notation used here is $ab \to cd$ for the s channel and $Db \to cA$ for the t channel, with $D=\bar{d}$ and $A=\bar{a}$; thus $s=(p_c+p_d)^2$, $t=(p_c-p_a)^2$, m_c is the mass of particle c, and so on.] In Eq. (1) we see the helicities of particles c, b attached to the continued t-channel amplitude coincide with the same labels on the s-channel amplitude, while the helicities of antiparticles A, D are opposite to those of particles a, d. This is caused by the reversal of the 4-momentum associated with particles a, d under $s-t$ crossing.

Proving Eq. (1) is easy. The $s-t$ crossing relations are[1,2]

$$f^{(s)}_{cdab} = \sum_{c',A',D',b'} d^{J_c}_{c'c}(\chi_c) d^{J_d}_{D'd}(\chi_d) d^{J_a}_{A'a}(\chi_a) \times d^{J_b}_{b'b}(\chi_b) f^{(t)}_{c'A'D'b'}, \tag{2}$$

where for large s

$$\begin{aligned} \cos\chi_c &\approx (t + m_c^2 - m_a^2)/T_{ca}, \\ \cos\chi_a &\approx -(t + m_a^2 - m_c^2)/T_{ca}, \\ \cos\chi_d &\approx -(t + m_d^2 - m_b^2)/T_{db}, \\ \cos\chi_b &\approx (t + m_b^2 - m_d^2)/T_{db}, \end{aligned} \tag{3}$$

with

$$T_{ca}^2 = |t|^2 + 2|t|(m_c^2 + m_a^2) + (m_c^2 - m_a^2)^2.$$

For large $|t|$, choosing the continuation $T_{ca} \to -|t|$, $T_{db} \to -|t|$ it follows that $\chi_c \to 0$, $\chi_a \to \pi$, $\chi_d \to \pi$, $\chi_b \to 0$.

Crossing equations which are formally identical to Eq. (1) are found when all particle masses are set equal to zero. However, in this case only the helicity amplitudes with maximum helicities are of interest, the reason being that for a massless particle with spin-J only the helicity states $|\lambda| = J$ are physical. All the unphysical amplitudes with $|\lambda| < J$ are decoupled from the physical amplitudes (the unphysical s- and t-channel amplitudes cross into each other). Normally one would set these unphysical amplitudes equal to zero.

The crossing equations (1) have a rather different meaning. In these equations all the helicity labels are physical, and it is not obvious that any one of them is more important than the others. If not, then there is no polarization at large $|t|$. However, we shall show there may very well be polarization, as a result of kinematic effects, whenever one of the final particles has spin ≥ 1. If there are no constraints among invariant amplitudes at large $|t|$, then the helicity amplitudes with the smallest helicity labels (0 for bosons, $\pm\frac{1}{2}$ for fermions) will be the only important ones. However, if certain constraints among the invariant amplitudes are satisfied, this need not be the case; some other helicity amplitudes may be more important.

In Sec. II we shall prove that for large s and $|t|$ the smallest helicities are kinematically favored over all higher ones; a sufficient condition being that all invariant amplitudes are independent in this region. The condition necessary to prove this result is weaker; it is necessary that the invariant amplitudes do *not* satisfy certain constraints involving s and t. This general result can be established very easily; one only needs a certain property of the equations which connect helicity amplitudes with invariant amplitudes. This property, which is nothing more than the appearance of certain mass factors in the invariant amplitude coefficients, may not be generally known and therefore we must say something about it. To minimize the complications associated with arbitrary spin we shall first discuss collisions with spins $0+0 \to L+0$ (i.e., diffractive excitation of one of the initial particles into a state with spin-L). For this process we give the equations relating helicity amplitudes and invariant amplitudes. Then we can show these equations have the mass factors we need, and it is easy to prove the large-$|t|$ result for this special case. Later we sketch the general proof.

As already mentioned, it is possible that the invariant amplitudes satisfy certain conditions, such that the kinematic preference for small helicity is avoided,

and other helicity states are dominant. This is also discussed in Sec. II, and we use our spin-L example to investigate the nature of these constraints. It seems that conditions among the invariant amplitudes cannot be found which lead to all the helicity states being comparably important. (For example, a spin-1 particle should have predominantly either helicity $=\pm 1$ or helicity $=0$.) Therefore, one can expect some sort of polarization at large $|t|$, although without a specific model one cannot predict exactly what this polarization should be.

II. SPIN STRUCTURE FOR LARGE s AND $|t|$

Consider the diffractive reaction $\pi+\pi\to\pi^*+\pi$ with spins $0+0\to L+0$ where the diffractively-excited state π^* with spin-L has unnatural parity when the corresponding initial meson π has unnatural parity. Parity is unchanged in the elastic transition $b\to d$, and there are $L+1$ independent amplitudes. In Appendix A we introduce a complete set of invariant amplitudes $A_1,\ldots,A_{L+1}$ and calculate the s- and t-channel helicity amplitudes in terms of them,

$$(m_c\sqrt{2})^L G_c^L f_{c000}^{(s)} = \sum_{N=0}^{L-|c|} (-)^{L-N}(L-N)!\,[(L-N+c)!\,(L-N-c)!\,]^{-1/2} \times S_{cd}^N T_{ca}^{L-N} d_{0c}^{L-N}(\chi_c) A_{1+N}, \tag{4}$$

$$(m_c\sqrt{2})^L G_c^L f_{c000}^{(t)} = \sum_{N=|c|}^{N} (-)^N N!\,[(N+c)!\,(N-c)!\,]^{-1/2} \times S_{cd}^N T_{ca}^{L-N} d_{0c}^{N}(\chi_c) A_{1+N}. \tag{5}$$

The crossing angle χ_c is defined by Eq. (A10). T_{ca} is given above and

$$S_{cd}^2 = s^2 - 2s(m_c^2+m_d^2) + (m_c^2-m_d^2)^2.$$

The rotation matrix elements for $n\geq 0$ are

$$d_{0n}^N(\chi) = (\sin\chi)^n \frac{1}{2^n N!}[(N+n)!\,(N-n)!\,]^{-1/2} P_n^N(\cos\chi),$$

where the polynomial

$$P_n^N(z) = 2^{N-n}\sum_{\beta=0}^{N-n}\binom{N}{\beta}\binom{N}{N-n-\beta}(-)^{N-n-\beta} \times (1+z)^\beta (1-z)^{N-n-\beta}$$

has the value

$$P_n^N(z=1) = \binom{N}{N-n}$$

at $z=1$. Thus for small χ_c,

$$d_{0n}^N(\chi_c) \approx (\sin\chi_c)^n \frac{1}{n!\,2^n}\left[\frac{(N+n)!}{(N-n)!}\right]^{1/2}.$$

We have already seen that $\chi_c\to 0$ in the kinematic region considered; indeed for $s\gg|t|\gg(\text{mass})^2$

$$\sin\chi_c \approx -2m_c/\sqrt{|t|}, \tag{6}$$

where we have chosen $T_{ca}\approx -|t|$. Therefore, in this region, for $c\geq 0$,

$$(m_c\sqrt{2})^L G_c^L f_{c000}^{(s)} \approx \left(-\frac{m_c}{\sqrt{|t|}}\right)^c \sum_{N=0}^{L-c}\frac{(L-N)!}{c!\,(L-N-c)!}\,s^N|t|^{L-N}A_{1+N}, \tag{7}$$

$$(m_c\sqrt{2})^L G_c^L f_{c000}^{(t)} \approx (-)^L\left(-\frac{m_c}{\sqrt{|t|}}\right)^c \sum_{N=|c|}^{L}\frac{N!}{c!\,(N-c)!}\,s^N|t|^{L-N}A_{1+N}. \tag{8}$$

In Eqs. (7) and (8) we already find the proof of our assertion that for large s and $|t|$ the kinematics favors small helicities. This is apparent from the overall factors $(m_c/\sqrt{t})^c$ on the right in these equations. If all the invariant amplitudes A_n are independent, then, because $m_c\gg\sqrt{|t|}$, the $c=0$ helicity amplitudes $f_{0000}^{(s)}\approx(-)^L f_{0000}^{(t)}$ will be dominant.

To make this even more explicit let us consider the simple case $L=1$, where the s- and t-channel amplitudes are

$$m_c\sqrt{2}f_{1000}^{(s)} = -m_c\sqrt{\phi}A_1/S_{cd}, \tag{9}$$

$$m_c 2f_{0000}^{(s)} = -T_{ca}\cos\chi_c A_1 + S_{cd}A_2, \tag{10}$$

$$m_c\sqrt{2}f_{1000}^{(t)} = -m_c\sqrt{\phi}A_2/T_{ca}, \tag{11}$$

$$m_c 2f_{0000}^{(t)} = T_{ca}A_1 - S_{cd}\cos\chi_c A_2. \tag{12}$$

In the limit of large s and $|t|$ these formulas become

$$m_c\sqrt{2}f_{1000}^{(s)} \approx s\left[-\frac{m_c\sqrt{|t|}}{s}\left(1-\frac{|t|}{s}\right)^{1/2}A_1\right], \tag{13}$$

$$m_c 2f_{0000}^{(s)} \approx s\left(\frac{|t|}{s}A_1 + A_2\right), \tag{14}$$

$$m_c\sqrt{2}f_{1000}^{(t)} \approx s\left[-\frac{m_c}{\sqrt{|t|}}\left(1-\frac{|t|}{s}\right)^{1/2}A_2\right], \tag{15}$$

$$m_c 2f_{0000}^{(t)} \approx s\left(-\frac{|t|}{s}A_1 - A_2\right). \tag{16}$$

All of Eqs. (9)–(16) have the same dimension. However, the invariant amplitude coefficients for $c=1$ in Eqs. (9) and (11) contain a factor m_c while the $c=0$ coefficients on the right in Eqs. (10) and (12) do not. For $c=0$ the dimension of the invariant amplitude coefficients is provided entirely by powers of $\sqrt{s}$ and $\sqrt{t}$. Because $\sqrt{s}$ and $\sqrt{t}$ are much larger than m_c the $c=0$ coefficients are the largest ones. Therefore, if A_1, A_2 are independent the $c=0$ helicity amplitudes $f_{0000}^{(s)}\approx -f_{0000}^{(t)}$ are dominant.

The rule for obtaining correct invariant amplitude formulas for large s and $|t|$ is to set all masses—excepting only overall mass factors—equal to zero. Thus one arrives at the situation one expects; for large s and $|t|$ all masses become unimportant, and essentially the only quantities with dimension which are available are s and t. In a sense, the overall mass factors in the dominant helicity amplitudes do not count as quantities with dimension because the same overall mass factor appears in each of these amplitudes as a normalization factor. This is true no matter whether the amplitudes with the smallest helicities are dominant, or whether, due to dynamical conditions among invariant amplitudes, some other helicity amplitudes are dominant.

Dynamical constraints can certainly exist. For example, we are free to imagine the condition

$$\frac{|t|}{s} A_1 + A_2 \lesssim O(m_c^2/t) \tag{17}$$

is satisfied by the invariant amplitudes A_1, A_2 in the spin $L=1$ example above. This constraint causes the $c=0$ amplitudes (14) and (16) to be small, of $O(m_c/\sqrt{t})$, with respect to the $c=1$ amplitudes (13) and (15), the latter becoming the same up to a sign in accordance with the crossing equations (1). Thus by imposing the condition (17) we make the $c=1$ amplitudes dominant rather than the $c=0$ ones.

It is important to note that one cannot assume that the right-hand side in Eq. (17) is $O(m_c/\sqrt{t})$, because of the properties attributed to invariant amplitudes. It is usually assumed that these functions do not depend explicitly on variables like $\sqrt{s}$ or $\sqrt{t}$, and therefore it should not be possible to satisfy a condition such as

$$\frac{|t|}{s} A_1 + A_2 \lesssim O(m_c/\sqrt{t}).$$

If this is true, then one cannot arrange for the $c=1$ amplitude and the $c=0$ amplitude to be equally important. One or the other will dominate, and there will be polarization.

For any given process one can easily write down the conditions which the invariant amplitudes must satisfy if some helicity states other than the lowest ones are to be important. Suppose in the spin-L amplitudes (7) and (8) the invariant amplitudes A_n satisfy the condition

$$\sum_{N=0}^{L} s^N |t|^{L-N} A_{1+N} \lesssim O(m_c^2/t). \tag{18}$$

Then the $c=1$ helicity amplitudes will dominate those with $c=0$ and $c \geq 2$. However, by introducing a second constraint

$$\sum_{N=1}^{L} \frac{N!}{(N-1)!} s^N |t|^{L-N} A_{1+N} \lesssim O(m_c^2/t) \tag{19}$$

we can make the $c=1$ amplitudes unimportant, of $O(m_c/\sqrt{t})$, with respect to the dominant amplitudes which are f_{0000} and f_{2000}. If we also change the right-hand side of Eq. (18) to $O(m_c^4/t^2)$ then only the $c=2$ amplitudes are dominant.

If the spin-L particle could be treated as a massless particle then only the amplitude f_{L000} would be important. The conditions which provide this are

$$\sum_{N=|c|}^{L} s^N |t|^{L-N} \frac{N!}{(N-c)!} A_{1+N} \lesssim O((m_c^2/t)^{L-c}), \tag{20}$$

where $c=0,1,\ldots,L-1$. An approximate solution of the conditions (20) is

$$A_{1+N} \approx (-)^N s^{L-N} |t|^N \frac{L!}{N!\,(L-N)!} f(s,t), \tag{21}$$

as one can verify using the identity

$$\sum_N (-)^N \frac{1}{(N-\alpha)!\,(\beta-N)!} = (-)^\beta \delta_{\alpha\beta}. \tag{22}$$

From these examples it is clear that for diffractive production of higher-spin states a variety of final polarizations are possible. However, this polarization cannot be entirely arbitrary, e.g., one cannot arrange for all helicity states of the higher-spin particle to be equally important. This is because one cannot impose constraints like (18) with odd powers of $(m_c/\sqrt{t})$ on the right-hand side; such constraints are inconsistent with the properties of invariant amplitudes. Therefore, some sort of final polarization should be present. In the $L=1$ example, we saw that either the $c=1$ helicity amplitudes *or* the $c=0$ ones are dominant. For higher spin, one can arrange for more complicated final polarization.

Let us now go on to the general problem, with all spins arbitrary. The diffractive excitation reactions of interest are $N\pi \to N^*\pi$, $N\pi^*$ and $N^*\pi^*$, $NN \to N^*N$ and N^*N^*, where N^* and π^* are excited nucleon and pion states with any spin. Limitations of space prevent us from giving here the explicit invariant amplitude coefficients for these reactions; they can be found in an unpublished work by the author.[3] The essential features of the problem are the same for any two-body reaction. The equations which give s-channel helicity amplitudes in terms of *any* complete set of invariant amplitudes A_n can always be written in the form

$$m_c^{J_c-|c|} m_d^{J_d-|d|} m_a^{J_a-|a|} m_b^{J_b-|b|} f_{cdab}^{(s)} = \sum_n M_{cdab}^n A_n, \tag{23}$$

where for any choice of the helicities (c,d,a,b) the coefficients M_{cdab}^n *do not contain any mass factor common to all the invariant amplitudes.* In other words, *no overall mass factor appears on the right in any of Eqs.* (23), although individual invariant amplitude coefficients may contain mass factors. Alternatively we can write Eq. (23) in the form

$$m_c^{J_c} m_d^{J_d} m_a^{J_a} m_b^{J_b} f_{cdab}^{(s)} = \sum_n \tilde{M}_{cdab}^n A_n, \tag{24}$$

where

$$\tilde{M}_{cdab}^n \equiv m_c^{|c|} m_d^{|d|} m_a^{|a|} m_b^{|b|} M_{cdab}^n. \tag{25}$$

All helicity amplitudes are dimensionless, and the dimension of Eq. (24) is fixed by the particle spins independently of helicity. Thus for a given n the coefficients (25) all have the same dimension independent of helicity. The available quantities with dimension are the masses, s and t. For large s and $|t|$ it is clear that the largest coefficients $\tilde{M}_{cdab}^n$ are those with no mass factors. The minimum values of $|c|$, $|d|$, $|a|$, and $|b|$ are preferred. This argument holds for each invariant amplitude separately. No redefinition of invariant amplitudes can affect the conclusion.

The same conclusion can be reached using covariant language, although this is a less transparent way to attack the problem. In Appendix B we briefly sketch the proof of the large-$|t|$ theorem using covariant wavefunctions for any spin.

The dominance of minimum helicities in a hard (large-$|t|$) collision would correspond to polarization of the final particles' spins perpendicular to their momentum. Dynamical constraints can exist with the form

$$\sum_n p_n(s,t) A_n(s,t) \sim 0 \tag{26}$$

where $p_n(s,t)$ is a polynomial in s and t, such that some other polarization obtains. We have seen that conditions of this sort can cause various helicity amplitudes, or

combinations of them, to be dominant. But not all helicity amplitudes can be made comparable in magnitude at once, and so polarization of some sort is expected at large $|t|$ when final particles with higher spins are involved.

The kinematic effect we have discussed in this article does not exist if the final masses are very large (e.g., "fireball" production). Neither does it exist for inclusive reactions, or for any exclusive reaction with three or more final particles. A simple example serves to illustrate the latter remark. Consider the reaction $a+b\to c+H$ where H is some (unspecified) system of hadrons and particle c has spin-1, while a, b are spinless. The cross section corresponding to the measurement of $p_{c\mu}$, with other final momenta either summed or unobserved, can be written

$$E_c\frac{d\sigma}{d^3p_c}=\sum_n \epsilon^*_{n\mu}(p_c)\epsilon_{n\nu}(p_c)W^{\mu\nu},$$

where

$$W_{\mu\nu}=g_{\mu\nu}W_1+\cdots.$$

The $g_{\mu\nu}$ term in $W_{\mu\nu}$ is independent of the others. Using the property $\epsilon^*_{m\mu}\epsilon^{\mu}_n=-\delta_{mn}$ we see that

$$E_c\frac{d\sigma}{d^3p_c}=\sum_n[-W_1+\cdots].$$

The W_1 term is helicity independent; in particular this term has no helicity-dependent mass factor. The additional terms do contain helicity-dependent mass factors, and for them one can find extreme kinematic configurations which favor minimal helicities. But this does not work for the W_1 term. Only if W_1 is negligable for dynamical reasons can one obtain a polarization effect. For two-body reactions a $g_{\mu\nu}$ term analogous to the one above is never present in the cross section.

APPENDIX A

Here we calculate the invariant amplitude formulas (4) and (5). There are $L+1$ independent amplitudes (L are dependent because of parity invariance). For the M function we write

$$\begin{aligned}M_{\mu_1\cdots\mu_L}&=(p_a\cdots p_a)_{\mu_1\cdots\mu_L}A_1\\&\quad+p_{d\mu_1}(p_a\cdots p_a)_{\mu_2\cdots\mu_L}A_2\\&\quad+\cdots+(p_d\cdots p_d)_{\mu_1\cdots\mu_L}A_{L+1}\end{aligned}\tag{A1}$$

which defines $L+1$ independent invariant amplitudes $A_1,\ldots,A_{L+1}$. The M function (A1) is to be contracted with a spin-L helicity wavefunction $\Phi_{c\mu_1\ldots}(p_c)$ for particle c in either the s channel or the t channel. These functions satisfy the decomposition rule[4]

$$\begin{aligned}G^L_c\Phi_{c\mu_1\cdots\mu_L}(p_c)&=\sum_n G^N_nG^{L-N}_{c-n}\Phi_{n\mu_1\cdots\mu_N}(p_c)\\&\quad\times\Phi_{c-n,\mu_{N+1}\cdots\mu_L}(p_c),\end{aligned}\tag{A2}$$

where

$$G^L_c=[(2L)!/(L+c)!\,(L-c)!\,]^{1/2}.\tag{A3}$$

The contraction formulas needed for the s channel are

$$\begin{aligned}&G^N_n(p_d\cdots p_d)^{\mu_1\cdots\mu_N}\Phi^{(s)}_{n\mu_1\cdots\mu_N}(p_c)\\&\quad=\delta_{n0}(S_{cd}/m_c\sqrt{2})^N,\end{aligned}\tag{A4}$$

$$\begin{aligned}&G^M_m(p_a\cdots p_a)^{\mu_1\cdots\mu_M}\Phi^{(s)}_{m\mu_1\cdots\mu_M}(p_c)\\&\quad=(-)^Md^M_{0m}(\chi_c)(T_{ca}/m_c\sqrt{2})^MM!\,[(M+m)!\,(M-m)!\,]^{-1/2}.\end{aligned}\tag{A5}$$

These are particular cases of general formulas.[5] With their help one easily obtains the s-channel amplitudes from the M function (A1),

$$\begin{aligned}&(m_c\sqrt{2})^LG^L_cf^{(s)}_{c000}\\&\quad=\sum_N(-)^{L-N}(L-N)!\,[(L-N+c)!\,(L-N-c)!\,]^{-1/2}\\&\qquad\times S^N_{cd}T^{L-N}_{ca}d^{L-N}_{0c}(\chi_c)A_{1+N}.\end{aligned}\tag{A6}$$

Similarly in the t channel

$$\begin{aligned}&G^N_n(p_D\cdots p_D)^{\mu_1\cdots\mu_N}\Phi^{(t)}_{n\mu_1\cdots\mu_N}(p_c)\\&\quad=(-)^Nd^N_{0n}(\chi_c)(S_{cd}/m_c\sqrt{2})^NN!\,[(N+n)!\,(N-n)!\,]^{-1/2},\end{aligned}\tag{A7}$$

$$\begin{aligned}&G^M_m(p_A\cdots p_A)^{\mu_1\cdots\mu_M}\Phi^{(t)}_{m\mu_1\cdots\mu_M}(p_c)\\&\quad=\delta_{m0}(T_{ca}/m_c\sqrt{2})^M.\end{aligned}\tag{A8}$$

The t-channel helicity amplitudes are

$$\begin{aligned}&(m_c\sqrt{2})^LG^L_cf^{(t)}_{c000}\\&\quad=\sum_N(-)^NN!\,[(N+c)!\,(N-c)!\,]^{-1/2}\\&\qquad\times S^N_{cd}T^{L-N}_{ca}d^N_{0c}(\chi_c)A_{1+N}.\end{aligned}\tag{A9}$$

The crossing angle χ_c is defined by

$$\begin{aligned}&S_{cd}T_{ca}\sin\chi_c=2m_c\sqrt{\phi},\\&S_{cd}T_{ca}\cos\chi_c=(s+m_c^2-m_d^2)(t+m_c^2-m_a^2)\\&\qquad\qquad-2m_c^2(m_c^2-m_d^2-m_a^2+m_b^2),\end{aligned}\tag{A10}$$

where ϕ is the physical boundary function. The fact that this angle appears explicitly in the s- and t-channel invariant amplitude coefficients makes it easy to obtain the crossing formula

$$f^{(s)}_{c000}=\sum_{c'}d^L_{c'c}(\chi_c)f^{(t)}_{c'000}\tag{A11}$$

from Eqs. (A6) and (A9). The formula (A11) follows from the identity

$$\begin{aligned}&(L-N)!\,[(L-N+c)!\,(L-N-c)!\,]^{-1/2}[G^L_c]^{-1}d^{L-N}_{0c}(\chi)\\&\quad=\sum_{c'}N!\,[(N+c)!\,(N-c)!\,]^{-1/2}\\&\qquad\times[G^L_{c'}]^{-1}d^L_{c'c}(\chi)d^N_{0c'}(-\chi),\end{aligned}\tag{A12}$$

and the fact that invariant amplitudes cross like spinless amplitudes. Equation (A12) can also be written in a more familiar form using 3-J symbols,

$$\begin{aligned}&\begin{pmatrix}L & N & L-N\\ c & 0 & -c\end{pmatrix}d^{L-N}_{c0}(\chi)\\&\quad=\sum_{c'}\begin{pmatrix}L & L-N & N\\ c' & 0 & -c'\end{pmatrix}d^L_{c'c}(\chi)d^N_{0c'}(\chi),\end{aligned}\tag{A13}$$

which is a special case of the Clebsch–Gordon decomposition formula. A technical point concerning Eqs. (A11) and (A12) is that one must change the sign of χ_c in Eq. (A9) when continuing to the s channel because the normal to the scattering plane flips over during the continuation.

APPENDIX B

The existence of the theorem proved in Sec. II can be inferred from the following covariant argument. This argument is not a proof; the only real proof is to work out the invariant amplitude coefficients in the s and t channels. However it does show in covariant language how the mass factors in Eq. (23) come to be there, and it may help one to understand why this formula is valid for all two-body reactions.

One knows that s-channel helicity amplitudes for any spins can be written in the form

$$f^{(s)}_{cdab} = \Psi^{(s)}_{c\mu_c}\Psi^{(s)}_{d\mu_d} M^{\mu_c\mu_d\mu_a\mu_b}\Psi^{(s)}_{a\mu_a}\Psi^{(s)}_{b\mu_b}, \tag{B1}$$

where the M function $M_{\mu_c\mu_d\mu_a\mu_b}$ and the helicity wavefunctions $\Psi^{(s)}_{\lambda_i\mu_i}$ carry Lorentz indices μ_i, as well as spinor indices which we suppress. A similar formula with the same M function holds for t channel helicity amplitudes. The wavefunctions $\Psi^{(s)}_{\lambda_i\mu_i}(p_i)$ for arbitrary spin can be constructed from wavefunctions $\epsilon_{n\mu}(p_i)$ for spin-1, and if the particle is a fermion then a Dirac spinor must also be included. This construction proceeds according to well-known rules[4,6]: for spin $J_i = L_i + \nu_i$ ($\nu_i = 0$ or $\frac{1}{2}$ for boson or fermion) the helicity wavefunctions $\Psi^{(s)}_{\lambda_i\mu_i}$ consist of a sum of terms, each of which is a product of L_i spin-1 wavefunctions $\epsilon_{n\mu}(p_i)$ multiplied by a Dirac spinor when $\nu_i = \frac{1}{2}$. The total helicity λ_i is the sum of the helicity labels on all the ϵ's appearing in the product, plus the helicity of the spinor if $\nu_i = \frac{1}{2}$. Dirac spinors contain no helicity-dependent overall mass factor and we can simply ignore them in the following discussion. The important part is the product of polarization vectors $\epsilon_{n\mu}(p_i)$, L_i of them, because these vectors do contain helicity-dependent mass factors. In an arbitrary frame

$$m\sqrt{2}\exp(\mp i\phi)\epsilon_{\pm1\mu}(p)$$
$$= m(0, \pm\cos\theta\cos\phi - i\sin\phi,$$
$$\pm\cos\theta\sin\phi + i\cos\phi, \mp\sin\theta) \tag{B2}$$

$$m\epsilon_{0\mu}(p) = E(|\mathbf{p}|/E, -\sin\theta\cos\phi, -\sin\theta\sin\phi, -\cos\theta) \tag{B3}$$

where $p = (E, \mathbf{p})$, $p^2 = m^2$, and the polar angles (θ, ϕ) specify the direction of $\mathbf{p}$. Both of the four-vectors (B2) and (B3) have the same dimension ~ (mass). However, the right-hand side in Eq. (B2) is a constant times an angular factor, while the right-hand side in Eq. (B3) is proportional to E. Therefore, it would seem that when $E \gg m$ the helicity-zero function (B3) should be much more important than the helicity-one function (B2). If this were true (it is not) then it is easy to see that the only important value of $|\lambda_i|$, the helicity of particle i, would be 0 or $\frac{1}{2}$ because λ_i is the sum of the helicity labels on the ϵ_n's, and only $n = 0$ is important. Thus in a kinematic region where all the E_i are large the conclusion would be that only the smallest helicity labels are important.

This is incorrect; we have ignored the fact that the $\epsilon_{n\mu}$ are four-vectors, and important cancellations can and do occur. In the s channel this is completely obvious. There, large E_i means large s (because $E_i \approx \sqrt{s}/2$), with t arbitrary. If the preceding discussion were correct then all s-channel helicities would be minimal at high energy, and of course this is not true. When the wavefunctions in Eq. (B1) are contracted with the tensors in the M function, many cancellations and rearrangements occur and the end result is that all s-channel helicity amplitudes have the same energy dependence (this must be true). All of the extra factors of $\sqrt{s}$ coming from E_i in Eq. (B3), which *a priori* seem to make ϵ_0 more important than $\epsilon_{\pm1}$, disappear. Effectively what happens is that cancellations cause these helicity-dependent factors of $\sqrt{s}$ to be replaced by $\sqrt{|t|}$ or by one of the masses m_j. When $|t|$ is not large no s channel helicity label is favored over any other. However, when $|t| \gg m^2$, then we can simply ignore all the masses, and then cancellations can only replace $\sqrt{s}$ by $\sqrt{|t|}$. Now we can see why the large-$|t|$ theorem exists. Small helicities correspond to many factors of ϵ_0 and therefore to many factors of $E_i \approx \sqrt{s}/2$ which have to be cancelled. When $|t|$ is large these factors of E_i become factors of $\sqrt{|t|}$. On the other hand, large helicities correspond to many factors of $\epsilon_{\pm1}$ and therefore to many explicit factors of $m_i \ll \sqrt{|t|}$ which nothing can cancel. The conclusion is obvious; small helicities are favored in the s channel. This argument holds for each invariant amplitude separately.

Things are a little different in the t channel. There one can easily see that $|t|$ must also be large if any simplification in the spin structure is to occur. For t-channel wavefunctions, the angle θ in Eqs. (B2), (B3) is either constant (0 or π) or it is the t-channel scattering angle θ_t or $\pi - \theta_t$, depending on which particle is considered. The azimuth ϕ is either 0 or π. Both $\sin\theta_t$ and $\cos\theta_t$ are proportional to s for large s,

$$-i\sin\theta_t \approx \cos\theta_t \approx 2s/T_{ca}T_{db}.$$

Therefore, no helicity is preferred over another when $s \to \infty$. However, when $|t|$ is also large, then E and $|\mathbf{p}|$ in Eq. (B3) are $\approx\sqrt{|t|}/2$ and it follows that ϵ_0 is more important than $\epsilon_{\pm1}$. When the t-channel wavefunctions are contracted with the tensors in the M function an overall helicity-independent power of $\sqrt{s}$ gets cancelled, and the t-channel helicity amplitudes all end up with the same energy dependence (this has to be true). Because masses can be neglected the cancelled helicity-independent power of $\sqrt{s}$ is replaced by a helicity-independent power of $\sqrt{|t|}$, the same for all helicity amplitudes. But the helicity amplitudes with large helicities also contain explicit factors of m_i coming from the polarization vectors (B2), while in the amplitudes with minimal helicities there are factors of $\sqrt{|t|}$ instead. No cancellation or rearrangement can overcome the advantage held by the latter amplitudes for large $|t|$. Therefore the minimal t-channel helicities are favored.

ACKNOWLEDGMENT

I am indebted to a referee for emphasizing the importance of constraints among invariant amplitudes at large $|t|$.

[1] T. L. Trueman and G. C. Wick, Ann. Phys. 26, 322 (1964).
[2] I. Muzinich, J. Math. Phys. 5, 1481 (1964).
[3] A. Actor (unpublished).
[4] Y. Frishman, Phys. Rev. 143, 1135 (1966).
[5] A. Actor, Lett. Nuovo Cimento 5, 902 (1972).
[6] P. Auvil and J. J. Brehm, Phys. Rev. 145, 1152 (1966).

Matrix elements of vector and axialvector operators in an algebraic formulation of baryon decays*

A. Böhm and R. B. Teese

Center for Particle Theory, Physics Department, University of Texas at Austin, Austin, Texas 78712
(Received 11 June 1976)

Properties of the matrix elements of baryon transition operators are derived from a minimal set of assumptions about the algebra of observables. Poincaré invariance, SU(3) classification, and V–A transitions are assumed. SU(3) is treated as a spectrum generating group rather than a symmetry group, and the usual form factors are expressed in terms of form factors which are invariant with respect to the spectrum generating SU(3). As a particular case, it is shown that the condition that the form factors be first class with respect to the spectrum generating SU(3) does not lead to a vanishing pseudotensor contribution.

INTRODUCTION

In this paper we present a fairly self-contained derivation, from a more general viewpoint than usual, of properties of the matrix elements of operators which occur in one-hadron processes. Our framework is relativistic quantum mechanics, i.e., the hadrons are described by an algebra of observables, and the basic assumptions about this algebra of observables are that it contains: (1) the algebra of the Poincaré group extended by parity, time inversion, and charge conjugation, $\mathcal{P}^{\text{ext}}$; (2) the algebra of SU(3); and (3) a set of vector and axialvector operators that have definite transformation properties under $\mathcal{P}^{\text{ext}}$ and SU(3) and which describe transitions between different hadron states. Since these assumptions are generally accepted, we will derive results which are restatements of expressions that the textbooks on particle physics begin with. The value of our presentation is not so much the derivation of familiar expressions from fundamental assumptions which are not burdened by field theoretic apparatus, but the fact that these expressions can be given in a modified version as a consequence of modified theoretical assumptions which lead to predictions for experimentally measurable quantities that cannot be derived under the usual assumptions. This will be the content of Sec. IV, where we combine the spacetime algebra of a spin-$\frac{1}{2}$ particle with the algebra of an SU(3) which is assumed to be a spectrum generating group, denoted $\text{SU}(3)_E$, rather than an "approximate symmetry" group. Sections I, II, and III are more pedagogical in nature and constitute a preparation for the last sections. Section I reviews the framework of our discussion and gives together with the Appendix all the material that is necessary for the following derivation, requiring of the reader just the knowledge of the fundamental facts of the representation of the Poincaré group and Lorentz group as they are given, e.g., in Ref. 10. In Sec. II it is shown that the matrix elements of Lorentz vector and axialvector operators may be written in terms of certain functions (form factors). Definite C, P, T and Hermiticity properties are assumed for these operators in Sec. III, and conditions on the form factors are derived. Finally, in Sec. V we relate the $\text{SU}(3)_E$-invariant form factors which were obtained in the previous sections to the non-$\text{SU}(3)_E$-invariant form factors that are commonly used for data analysis.

I. FRAMEWORK

The basis of a relativistic quantum mechanics for particle physics will be the association of a hadron with an irreducible representation of the Poincaré group.[1] This introduces as observables the generators P_μ, $L_{\mu\nu}$, $\nu, \mu = 0, 1, 2, 3$, of the Poincaré group with their well-known physical interpretation. If we associate with this algebra a Lorentz vector operator Γ_μ which together with the Poincaré group generators fulfills the following commutation relations:

$$[P_\mu, P_\nu] = 0, \tag{1}$$

$$[L_{\mu\nu}, P_\rho] = i(g_{\nu\rho}P_\mu - g_{\mu\rho}P_\nu), \tag{2}$$

$$[L_{\mu\nu}, L_{\rho\sigma}] = -i\,(g_{\mu\rho}L_{\nu\sigma} + g_{\nu\sigma}L_{\mu\rho} - g_{\mu\sigma}L_{\nu\rho} - g_{\nu\rho}L_{\mu\sigma}), \tag{3}$$

$$[M_{\mu\nu}, S_{\rho\sigma}] = 0, \tag{4}$$

$$\tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}P^\nu M^{\rho\sigma} = 0, \tag{5}$$

$$[S_{\mu\nu}, S_{\rho\sigma}] = -i(g_{\mu\rho}S_{\nu\sigma} + g_{\nu\sigma}S_{\mu\rho} - g_{\mu\sigma}S_{\nu\rho} - g_{\nu\rho}S_{\mu\sigma}), \tag{6}$$

$$[L_{\rho\sigma}, \Gamma_\mu] = [S_{\rho\sigma}, \Gamma_\mu] = i(g_{\sigma\mu}\Gamma_\rho - g_{\rho\mu}\Gamma_\sigma), \tag{7}$$

$$[\Gamma_\rho, \Gamma_\sigma] = -iS_{\rho\sigma}, \tag{8}$$

$$[P_\mu, \Gamma_\nu] = 0,$$

where

$$L_{\mu\nu} = M_{\mu\nu} + S_{\mu\nu},$$

$\mu, \nu, \rho, \sigma = 0, 1, 2, 3$ and $g_{00} = 1$, $g_{11} = g_{22} = g_{33} = -1$, then we obtain a relativistic symmetry[2] $\mathfrak{S}$. This $\mathfrak{S}$, which is the semidirect product[3] $\mathcal{P}_{P_\mu, L_{\mu\nu}} \boxdot \text{SO}(3,2)_{\Gamma_\mu, S_{\mu\nu}}$, we believe to be of fundamental importance in the relativistic quantum mechanics of particle physics.[4,5] A well-known example of this relativistic symmetry is obtained if, in addition to relations (1)–(8), one requires

$$\{\Gamma_\rho, \Gamma_\sigma\} = \tfrac{1}{2}g_{\rho\sigma}. \tag{9}$$

The irreducible representation space $\mathcal{H}(m, D)$ of $\mathfrak{S}$, in which (9) is fulfilled, is called the Dirac representation and is identical with the space of solutions of the Dirac equation.[6,7] $\mathcal{H}(m, D)$ is the direct sum of two equivalent representations of the Poincaré group

$$\mathcal{H}(m, D) = \mathcal{H}^{n=1/2}(m, s=1/2) \oplus \mathcal{H}^{n=-1/2}(m, s=1/2), \tag{10}$$

where the additional label $n = \pm 1/2$ is the eigenvalue of the operator $P_\mu M^{-1}\Gamma^\mu$ (or of Γ_0 for states at rest, since

the energy p_0 is assumed to always be positive). $\mathcal{H}(m, D)$ is an irreducible representation of the extended Poincaré group; $\mathcal{H}^{n=1/2}(m, s=1/2)$ describes particle states and $\mathcal{H}^{n=-1/2}(m, s=1/2)$ antiparticle states with negative relative parity.

There are two convenient basis systems in the Dirac representation, which are inherited from the corresponding basis systems of the Poincaré group. The canonical basis $|ps\sigma n\rangle$ is labeled by the 4-momentum p, the spin $s(=1/2)$, the third component of spin σ, and the additional quantum number n. Under a Lorentz transformation Λ, the canonical basis transforms as[1,8]

$$U(\Lambda)|ps\sigma n\rangle=\sum_{\sigma'}|(\Lambda p)\, s\sigma' n\rangle\, D^{s=1/2}_{\sigma'\sigma}(R), \tag{11}$$

where R is the Wigner rotation

$$R=L(\Lambda p)\,\Lambda L^{-1}(p),$$

L^{-1} is the "boost," $L(p)p=(m,0,0,0)$ and $D^{s=1/2}_{\sigma'\sigma}(R)$ is the matrix which represents R in the $s=1/2$ irreducible representation of the rotation group. The generalized basis vectors may be "normalized" such that

$$\langle p'n'\sigma'|\, pn\sigma\rangle=2p_0\delta^3(\mathbf{p}'-\mathbf{p})\,\delta_{\sigma'\sigma}\delta_{n'n}, \tag{12}$$

where we have dropped the $s=1/2$ label.

The spinor basis[8,9] $f^{(k_0c)j}_{j_3}(p)$ is defined to transform as

$$U(\Lambda)f^{(k_0c)j}_{j_3}(p)=\sum_{j'j_3'} f^{(k_0c)j'}_{j_3'}(\Lambda p)\,D^{j'j(k_0c)}_{j_3'j_3}(\Lambda), \tag{13}$$

where $D^{j'j(k_0c)}_{j_3'j_3}(\Lambda)$ is the matrix which represents the Lorentz transformation Λ in the (k_0c) representation of the Lorentz group. For the Dirac representation, $k_0=1/2$, $c=\pm 3/2$, and $j=j'=1/2$.[10]

The spinor basis is not orthogonal,[6,8,11] and so $\langle f^c_{j_3}(p)|$ does not transform contragradiently to $|f^c_{j_3}(p)\rangle$. We may define a new vector [omitting the constant indices k_0, j, and using the γ matrices defined by (A5)],

$$\langle \mathring{f}^c_{j_3}(p)|=(1/m)(p_\mu\gamma^\mu\gamma_0)^{cc'}_{j_3j_3'}\langle f^{c'}_{j_3'}(p)|, \tag{14}$$

which does transform contragradiently to $|f^c_{j_3}(p)\rangle$. Then the normalization is given by

$$\langle \mathring{f}^{c'}_{j_3'}(p')|f^c_{j_3}(p)\rangle=2p_0\delta^3(\mathbf{p}'-\mathbf{p})\,\delta^{c'c}\,\delta_{j_3'j_3}, \tag{15}$$

and the connection between the two basis systems is

$$|pn\sigma\rangle=|f^c_{j_3}(p)\rangle\, D^{cn}_{j_3\sigma}(p) \tag{16a}$$

and

$$\langle pn\sigma|=\mathring{D}^{nc}_{\sigma j_3}(p)\langle \mathring{f}^c_{j_3}(p)| \tag{16b}$$

(repeated indices are summed). The transformation matrices D, $\mathring{D}$ play essentially the same role as the usual positive energy Dirac spinors u, v and $\bar{u}$, $\bar{v}$ respectively. The relationship between the D's and the Dirac spinors is discussed in the Appendix.

By (16), the matrix elements of any operator J in the canonical basis may be written

$$\langle \alpha' p'n'\sigma'|J|\, pn\sigma\alpha\rangle$$
$$=\mathring{D}^{n'c'}_{\sigma'j_3'}(p')\langle^{\alpha'}\mathring{f}^{c'}_{j_3'}(p')\,|J|^{\alpha} f^c_{j_3}(p)\rangle D^{cn}_{j_3\sigma}(p), \tag{17}$$

where α represents any additional quantum numbers [SU(3), etc.] which may be present. We ignore these additional quantum numbers for the time being and return to them in Sec. IV. The transformation property (13) follows from the fact that the spinor basis is defined as a direct product:

$$f^c_{j_3}(p)=f^c_{j_3}\otimes f(p). \tag{18}$$

The $f^c_{j_3}$ are the basis elements of the four-dimensional representation space $(1/2,3/2)\oplus(1/2,-3/2)$ of $SO(3,1)_{S_{\mu\nu}}$, which is an irreducible representation space of $SO(3,2)_{S_{\mu\nu}\Gamma_\mu}$, while $f(p)$ is a generalized basis vector of the $(m, s=0)$ representation of the Poincaré group.[6,11]

Every operator which acts on a direct product space may be written as a linear combination of direct products of operators; thus we write for J

$$J=\sum_i J^I_{(i)}\otimes J^X_{(i)},$$

where $J^I_{(i)}$ acts only on $f^c_{j_3}$ and $J^X_{(i)}$ acts only on $f(p)$. It will be seen that, for the cases we consider here, $i=1,2$. Thus (17) may be written

$$\langle p'n'\sigma'|J|\, pn\sigma\rangle$$
$$=\mathring{D}^{n'c'}_{\sigma'j_3'}(p')\sum_i\langle \mathring{f}^{c'}_{j_3'}|J^I_{(i)}|\, f^c_{j_3}\rangle\,\langle \mathring{f}(p')|J^X_{(i)}|f(p)\rangle\, D^{cn}_{j_3\sigma}(p). \tag{19}$$

Equation (19) will be illustrated in the Appendix by showing how certain simple operators are evaluated in both basis systems.

II. FORM FACTORS

For the case that $J=V_\mu$ is a Lorentz vector, we may write in general

$$V_\mu=\sum_i (V^I_{(i)}\otimes V^X_{(i)})_\mu \tag{20}$$

so that the matrix elements of V_μ in the spinor basis are

$$\langle \mathring{f}^{c'}_{j_3'}(p')|\, V_\mu\,|f^c_{j_3}(p)\rangle$$
$$=\sum_i[\langle \mathring{f}^{c'}_{j_3'}|\, V^I_{(i)}\,|f^c_{j_3}\rangle\langle \mathring{f}(p')|\, V^X_{(i)}\,|f(p)\rangle]_\mu. \tag{21}$$

Since the matrix elements $\langle \mathring{f}(p')|\, V^X_{(i)}\,|f(p)\rangle$ are just functions (distributions) of p and p', we may write

$$=\langle \mathring{f}^{c'}_{j_3'}|\sum_i[\langle \mathring{f}(p')|\, V^X_{(i)}\,|f(p)\rangle\, V^I_{(i)}]_\mu\,|f^c_{j_3}\rangle$$

which is simply, for a given p' and p, a 4×4 matrix in the indices c, c', j_3, j_3'. This matrix may therefore be expanded in terms of the 16 basis elements of the Dirac algebra:

$$=a_\mu(1)^{c'c}_{j_3'j_3}+a_\mu{}^\sigma(\gamma_\sigma)^{c'c}_{j_3'j_3}$$
$$+a_\mu{}^{\sigma\rho}(\sigma_{\sigma\rho})^{c'c}_{j_3'j_3}+a_\mu{}^5(\gamma_5)^{c'c}_{j_3'j_3}+a_\mu{}^{5\sigma}(\gamma_\sigma\gamma_5)^{c'c}_{j_3'j_3}, \tag{21a}$$

where the expansion coefficients a_μ, $a_\mu{}^\sigma$, etc., are functions of p, p'. In order that both sides of this equation have the same property under Lorentz and parity transformations, it must be that

a_μ is a Lorentz vector,

$a_\mu{}^\sigma$ is a Lorentz second rank tensor,

$a_\mu{}^{\sigma\rho}$ is a Lorentz third rank tensor, (21b)

$a_\mu{}^5$ is a Lorentz pseudovector,

$a_\mu{}^{5\sigma}$ is a Lorentz second rank pseudotensor.

Since these quantites are functions of the momenta, we need the most general functions of the two Lorentz vectors p', p which satisfy the above Lorentz transformation properties. These are

$$a_\mu = (p'_\mu - p_\mu) F_1^- + (p'_\mu + p_\mu) F_1^+, \tag{22a}$$

$$a_\mu{}^\sigma = g_\mu{}^\sigma F_2 + p_\mu p'^\sigma F_3 + p'_\mu p^\sigma F_4 + p_\mu p^\sigma F_5 + p'_\mu p'^\sigma F_6, \tag{22b}$$

$$\begin{aligned} a_\mu{}^{\sigma\rho} = {} & g_\mu{}^\sigma (p'^\rho + p^\rho) F_7^+ + g_\mu{}^\sigma (p'^\rho - p^\rho) F_7^- \\ & + p'^\rho p^\sigma (p'_\mu + p_\mu) F_8^+ + p'^\rho p^\sigma (p'_\mu - p_\mu) F_8^- \\ & + \text{terms symmetric in } \rho, \sigma, \end{aligned} \tag{22c}$$

$$a_\mu{}^5 = 0, \tag{22d}$$

$$a_\mu{}^{5\sigma} = \epsilon_\mu{}^{\sigma\rho\delta} p'_\rho p_\delta F_9, \tag{22e}$$

where the F's are Lorentz scalar functions of p and p'. Equation (22d) follows from the fact that no pseudovector can be formed from the two momenta. In (22c), terms symmetric in ρ, σ were omitted because $\sigma_{\rho\sigma}$ is antisymmetric. When the above quantities are used in (21a) and (19), we obtain

$$\begin{aligned} &\langle p'n'\sigma' \,|\, V_\mu \,|\, pn\sigma\rangle \\ &= \dot{D}^{n'c'}_{\sigma' j'_3}(p')\{k_\mu F_1^+ + q_\mu F_1^- + \gamma_\mu F_2 + p_\mu \gamma_\sigma p'^\sigma F_3 \\ &\quad + p'_\mu \gamma_\sigma p^\sigma F_4 + p_\mu \gamma_\sigma p^\sigma F_5 + p'_\mu \gamma_\sigma p'^\sigma F_6 \\ &\quad + \sigma_{\mu\sigma}(k^\sigma F_7^+ + q^\sigma F_7^-) + \sigma_{\sigma\rho} p^\sigma p'^\rho (q_\mu F_8^- + k_\mu F_8^+) \\ &\quad + \gamma_5 \gamma_\rho \epsilon_\mu{}^{\rho\sigma\delta} p'_\sigma p_\delta F_9\}^{c'c}_{j'_3 j_3} D^{cn}_{j_3\sigma}(p), \end{aligned} \tag{23}$$

where $q_\mu = p'_\mu - p_\mu$ and $k_\mu = p'_\mu + p_\mu$. The terms involving $\gamma_\sigma p^\sigma$ and $\gamma_\sigma p'^\sigma$ may be simplified by using the Dirac equation (A22),

$$\begin{aligned} p_\mu (\gamma^\mu)^{cc'}_{j_3 j'_3} D^{c'n}_{j'_3\sigma}(p) &= \pi m D^{cn}_{j_3\sigma}(p), \\ \dot{D}^{nc'}_{\sigma j'_3}(p)\, p_\mu (\gamma^\mu)^{c'c}_{j'_3 j_3} &= \pi m \dot{D}^{nc}_{\sigma j_3}(p), \end{aligned} \tag{24}$$

where $\pi = \mathrm{sgn}(n)$ and $m = (p_\mu p^\mu)^{1/2}$. The result is

$$\begin{aligned} &\langle p'n'\sigma' \,|\, V_\mu \,|\, pn\sigma\rangle \\ &= \dot{D}^{n'c'}_{\sigma' j'_3}(p')\{\gamma_\mu f_1 + q_\mu f_2 + \sigma_{\mu\nu} q^\nu f_3 + \pi\gamma_\mu f_4 + \pi'\gamma_\mu f_5 \\ &\quad + (\pi' p'_\mu - \pi p_\mu) f_6 + \sigma_{\mu\nu}(\pi' p'^\nu - \pi p^\nu) f_7 \\ &\quad + \pi'\pi\gamma_\mu f_8 + \pi'\pi q_\mu f_9 + \pi'\pi\sigma_{\mu\nu} q^\nu f_{10} \\ &\quad + (\pi p'_\mu - \pi' p_\mu) f_{11} + \sigma_{\mu\nu}(\pi p'^\nu - \pi' p^\nu) f_{12}\}^{c'c}_{j'_3 j_3} D^{cn}_{j_3\sigma}(p), \end{aligned} \tag{25}$$

where the f's are linear combinations of the F's and other Lorentz scalar functions of p, p'.[12]

Equation (25) covers the general case where V_μ causes transitions from p, n, σ states to p', n', σ' states. In practice, however, we rarely have this general situation.[13] Therefore, we restrict ourselves to the case $n' = n$. Then $\pi' = \pi$ and $\pi\pi' = 1$, so (25) now contains only six functions:

$$\begin{aligned} &\tilde{f}_1 = f_1 + f_8, \quad i\tilde{f}_2 = f_3 + f_{10}, \quad \tilde{f}_3 = f_2 + f_9, \\ &\tilde{\tilde{f}}_1 = f_4 + f_5, \quad i\tilde{\tilde{f}}_2 = f_7 + f_{12}, \quad \tilde{\tilde{f}}_3 = f_6 + f_{11}. \end{aligned} \tag{26}$$

If we define

$$F_i = \tilde{f}_i + \pi\tilde{\tilde{f}}_i, \quad i = 1, 2, 3, \tag{27}$$

then (25) takes the familiar form

$$\begin{aligned} &\langle p'n\sigma' \,|\, V_\mu \,|\, pn\sigma\rangle \\ &= \dot{D}^{nc'}_{\sigma j'_3}(p')\{\gamma_\mu F_1 + i\sigma_{\mu\nu} q^\nu F_2 + q_\mu F_3\}^{c'c}_{j'_3 j_3} D^{cn}_{j_3\sigma}(p) \end{aligned} \tag{28}$$

(no sum over n). However, we should emphasize that these F_i's are not quite the usual form factors, because by (27) they depend on n as well as p and p'. That is, they contain a part $\tilde{f}_i$ which is invariant under charge conjugation C, and a part $\pi\tilde{\tilde{f}}_i$ which changes sign under C. This is required by the fact that we have an *ab initio* positive energy theory, and it will be discussed further when we consider charge conjugation.

For the case that $J = A_\mu$ is a Lorentz pseudovector, we must replace everywhere in (21b) vector, tensor, etc., by pseudovector, pseudotensor, etc. No pseudovector can be formed from the two 4-momenta, so we have

$$\begin{aligned} &a_\mu = 0, \quad a_\mu{}^\sigma = \epsilon_\mu{}^{\sigma\rho\pi} p'_\rho p_\pi G_1, \\ &a_\mu{}^{\sigma\rho} = k_\pi \epsilon_\mu{}^{\sigma\rho\pi} G_2^+ + q_\pi \epsilon_\mu{}^{\sigma\rho\pi} G_2^-, \\ &a_\mu{}^5 = k_\mu G_3^+ + q_\mu G_3^-, \\ &a_\mu{}^{5\sigma} = g_\mu{}^\sigma G_4 + p_\mu p^\sigma G_5 + p'_\mu p^\sigma G_6 \\ &\qquad + p_\mu p'^\sigma G_7 + p'_\mu p'^\sigma G_8. \end{aligned} \tag{29}$$

By using

$$\gamma_5 = -(i/4!)\epsilon_{\mu\nu\sigma\gamma}\gamma^\mu\gamma^\nu\gamma^\sigma\gamma^\gamma \tag{30}$$

we find that (30) leads to a form analogous to (25). Under the restriction $n = n'$, we have

$$\begin{aligned} &\langle p'n\sigma' \,|\, A_\mu \,|\, pn\sigma\rangle \\ &= \dot{D}^{nc'}_{\sigma j'_3}(p')\{\gamma_\mu\gamma_5 \tilde{g}_1 + \gamma_5 q_\mu \tilde{g}_3 + i\sigma_{\mu\sigma}\gamma_5 q^\sigma \tilde{g}_2 \\ &\quad + \pi\gamma_\mu\gamma_5 \tilde{\tilde{g}}_1 + \pi\gamma_5 q_\mu \tilde{\tilde{g}}_3 + \pi i\sigma_{\mu\sigma}\gamma_5 q^\sigma \tilde{\tilde{g}}_2\}^{c'c}_{j'_3 j_3} D^{cn}_{j_3\sigma}(p), \end{aligned} \tag{31}$$

where the g's are Lorentz scalar functions.[14] If we define G's [by analogy with Eq. (27)] which depend not only upon p and p' but also n, $G_i = \tilde{g}_i + \pi\tilde{\tilde{g}}_i$, then we may write

$$\begin{aligned} &\langle p'n\sigma' \,|\, A_\mu \,|\, pn\sigma\rangle \\ &= \dot{D}^{nc'}_{\sigma j'_3}(p')\{\gamma_\mu\gamma_5 G_1 + i\sigma_{\mu\sigma} q^\sigma \gamma_5 G_2 + \gamma_5 q_\mu G_3\}^{c'c}_{j'_3 j_3} D^{cn}_{j_3\sigma}(p). \end{aligned} \tag{32}$$

In the foregoing, we have explicitly displayed the indices on the transformation matrices D and $\dot{D}$. This served to emphasize that they are in fact matrices rather than column vectors, as they are usually treated. In the remainder of the paper, however, we will omit those indices which are summed over and list the free indices as variables, in order to simplify the formulas. For example, instead of

$$\dot{D}^{n'c'}_{\sigma' j_3'}(p')(\gamma_\mu)^{c'c''}_{j_3' j_3''} D^{c'' n}_{j_3'' \sigma}(p)$$

we will simply write

$$\dot{D}(p', n', \sigma')\gamma_\mu D(p, n, \sigma).$$

III. CONDITIONS ON THE FORM FACTORS

The assumption of definite transformation properties of V_μ and A_μ under the discrete transformations C, P, and T puts restrictions on the functions F_i and G_i of Eqs. (28) and (32). The P-transformation property has already been used in deriving (28) and (32); we now assume that V_μ and A_μ are normal under time reversal,[15] i.e., that

$$A_T V_\mu A_T^{-1} = \epsilon(\mu) V_\mu, \quad A_T A_\mu A_T^{-1} = \epsilon(\mu) A_\mu, \tag{33}$$

where

$$\epsilon(\mu) = -(1 - 2\delta_{\mu 0}) = \begin{cases} +1 & \text{for } \mu = 0, \\ -1 & \text{for } \mu = 1, 2, 3. \end{cases}$$

Like the assumption that V and A are vector and axial vector operators, (33) is a restrictive assumption that can only be justified by its experimental consequences.

In order to exploit this transformation property, we need an extension of the Dirac representation of the relativistic symmetry by T. We take for this the extension of the Poincaré group representation $\mathcal{H}(m, 1/2)$ by P and T which is realized by baryons.[1,4,16] In that case A_T does not affect n, and its action on the states is[4]

$$A_T | p_0, \mathbf{p}, n, \sigma\rangle = \alpha(n)(-1)^{\sigma+1/2} | p_0, -\mathbf{p}, n, -\sigma\rangle, \tag{34}$$

where $\alpha(n)$ is a phase factor which may depend on n. Then

$$\begin{aligned} &\langle p' n\sigma' | A_T^{-1} V_\mu A_T | p n\sigma\rangle \\ &\quad = \overline{(A_T | p' n\sigma'\rangle, V_\mu A_T | p n\sigma\rangle)} \\ &\quad = -(-1)^{\sigma+\sigma'} \overline{\langle -\mathbf{p}'\ n\ -\sigma' | V_\mu | -\mathbf{p}\ n\ -\sigma\rangle}, \end{aligned} \tag{35}$$

where we have used (34) and the antiunitary nature of A_T. According to (33), the lhs of (35) must be

$$= \epsilon(\mu)\langle p' n\sigma' | V_\mu | p n\sigma\rangle$$

so

$$\begin{aligned} &\langle p' n\sigma' | V_\mu | p n\sigma\rangle \\ &\quad = -\epsilon(\mu)(-1)^{\sigma+\sigma'} \overline{\langle -\mathbf{p}'\ n\ -\sigma' | V_\mu | -\mathbf{p}\ n\ -\sigma\rangle}. \end{aligned} \tag{36}$$

Writing the rhs of (36) in the form (28), we have, for $n' = n$,

$$\begin{aligned} &-\epsilon(\mu)(-1)^{\sigma+\sigma'} \overline{\langle -\mathbf{p}'\ n\ -\sigma' | V_\mu | -\mathbf{p}\ n\ -\sigma\rangle} \\ &\quad = -\epsilon(\mu)(-1)^{\sigma+\sigma'} \overline{\mathring{D}(-\mathbf{p}', n, -\sigma')\{\gamma_\mu F_1 + i\sigma_{\mu\nu} q^\nu F_2 \epsilon(\nu)} \\ &\quad \overline{+ q_\mu F_3 \epsilon(\mu)\} D(-\mathbf{p}, n, -\sigma)}. \end{aligned} \tag{37}$$

From the explicit forms for the bilinears (A49)–(A53), we find the properties

$$\begin{aligned} &(-1)^{\sigma+\sigma'} \mathring{D}(-\mathbf{p}', n, -\sigma')(\gamma_\mu) D(-\mathbf{p}, n, -\sigma) \\ &\quad = -\epsilon(\mu) \overline{\mathring{D}(\mathbf{p}', n, \sigma')\gamma_\mu D(\mathbf{p}, n, \sigma)}, \end{aligned} \tag{38}$$

etc. Using these properties to compare (37) with (28), we find

$$\begin{aligned} &F_1(p', p) = \overline{F_1(p', p)}, \quad F_2(p', p) = \overline{F_2(p', p)}, \\ &F_3(p', p) = \overline{F_3(p', p)}. \end{aligned} \tag{39}$$

Similarly, for the axial transition operator A_μ we find

$$\begin{aligned} &G_1(p', p) = \overline{G_1(p', p)}, \quad G_2(p', p) = \overline{G_2(p', p)}, \\ &G_3(p', p) = \overline{G_3(p', p)}. \end{aligned} \tag{40}$$

From the assumption that the V_μ and A_μ are Hermitian,

$$V_\mu^\dagger = V_\mu, \quad A_\mu^\dagger = A_\mu, \tag{41}$$

we derive further conditions on the F_i. Because of (41) we have

$$\langle p' n\sigma' | V_\mu | p n\sigma\rangle = \overline{\langle p n\sigma | V_\mu | p' n\sigma'\rangle}. \tag{42}$$

Using (28), this may be written

$$\begin{aligned} &\mathring{D}(p' n\sigma')\{\gamma_\mu F_1(p', p) + i\sigma_{\mu\nu} q^\nu F_2(p', p) + q_\mu F_3(p', p)\} D(p n\sigma) \\ &\quad = \overline{\mathring{D}(p n\sigma)\{\gamma_\mu F_1(p, p') - i\sigma_{\mu\nu} q^\nu F_2(p, p')} \\ &\quad \overline{- q_\mu F_3(p, p')\} D(p n\sigma')}. \end{aligned} \tag{43}$$

The explicit forms of the bilinears (A49)–(A53) may be used to compare the two sides of this equation. The result is

$$\begin{aligned} &F_1(p', p) = +\overline{F_1(p, p')}, \quad F_2(p', p) = +\overline{F_2(p, p')}, \\ &F_3(p', p) = -\overline{F_3(p, p')}. \end{aligned} \tag{44}$$

Since the F's are Lorentz scalars, they must be functions of scalars formed with p' and p. Thus they are symmetric in p' and p, i.e., $F_i(p', p) = F_i(q^2)$, where $q^2 = (p' - p)^2$, so that (44) may be compared with (39). The result is

$$F_1 \text{ and } F_2 \text{ real}, \quad F_3 = 0. \tag{45}$$

The hermiticity of A_μ leads to similar consequences:

$$\begin{aligned} &G_1(p', p) = \overline{G_1(p, p')}, \quad G_2(p', p) = -\overline{G_2(p, p')}, \\ &G_3(p', p) = \overline{G_3(p, p')}. \end{aligned} \tag{46}$$

The G's are also symmetric in p', p, i.e., $G_i(p', p) = G_i(q^2)$, so that (46) may be compared with (40) to give

$$G_2 = 0, \quad G_1 \text{ and } G_3 \text{ real}. \tag{47}$$

For physical reasons, the charge conjugation operator U_c must have the property

$$U_c | p n\sigma\rangle = a | p\ -n\ \sigma\rangle, \tag{48}$$

where a is a phase factor, and the relationship between the U_c and the Poincaré group is

$$U_c^{-1} P_\mu U_c = P_\mu, \quad U_c^{-1} L_{\mu\nu} U_c = L_{\mu\nu}. \tag{49}$$

We may find the action of U_c on Γ_μ as follows:

$$\begin{aligned} &U_c^{-1}\Gamma_0 U_c | \mathbf{p} = 0, n\sigma\rangle \\ &\quad = a(n)\, a(-n)(-n) | \mathbf{p} = 0,\ n\sigma\rangle \\ &\quad = -n | \mathbf{p} = 0, n\sigma\rangle \\ &\quad = -\Gamma_0 | \mathbf{p} = 0, n\sigma\rangle, \end{aligned} \tag{50}$$

where we have used the unitarity of U_c, i.e., $a(-n)\, a(n) = 1$. Thus

$$U_c^{-1}\Gamma_0 U_c = -\Gamma_0. \tag{51}$$

Moreover,

$$U_c^{-1}\Gamma_i U_c|\mathbf{p}=0,n\sigma\rangle$$
$$=a(n)\,U_c^{-1}\Gamma_i|\mathbf{p}=0,\ -n\ \sigma\rangle$$
$$=a(n)\,U_c^{-1}\sum_{-n',\sigma'}|\mathbf{p}=0,\ -n'\ \sigma'\rangle\langle -n'\sigma'|\Gamma_i|-n\ \sigma\rangle$$
$$=a(n)\sum_{-n',\sigma'}a(-n')|\mathbf{p}=0,n'\sigma'\rangle\langle -n'\ \sigma'|\Gamma_i|-n\ \sigma\rangle. \tag{52}$$

The matrix element of $2\Gamma_i$ is $(\gamma_i)^{n'n}_{\sigma'\sigma}$, which is given in Eq. (A4) of the Appendix. Thus

$$=a(n)\sum_{-n',\sigma'}a(-n')|\mathbf{p}=0,n\sigma\rangle\,\delta_{-n',n}(-\tfrac{1}{2}i\sigma_i)_{\sigma'\sigma}$$
$$=a(n)\,a(n)\sum_{\sigma'}|\mathbf{p}=0,\ -n\sigma'\rangle(-\tfrac{1}{2}i\sigma_i)_{\sigma'\sigma}$$
$$=[a(n)]^2\Gamma_i|\mathbf{p}=0,n\sigma\rangle$$

so that

$$U_c^{-1}\Gamma_i U_c=[a(n)]^2\,\Gamma_i. \tag{53}$$

Using (49), (51), and (53), we calculate

$$U_c^{-1}P_\mu\Gamma^\mu U_c|pn\sigma\rangle$$
$$=-\{P_0\Gamma^0-[a(n)]^2P_i\Gamma^i\}|pn\sigma\rangle. \tag{54}$$

On the other hand,

$$U_c^{-1}P_\mu\Gamma^\mu U_c|pn\sigma\rangle=U_c^{-1}P_\mu\Gamma^\mu\,a(n)|p\ -n\ \sigma\rangle$$
$$=a(n)\,a(-n)(-n)\,m\,|\,pn\sigma\rangle$$
$$=-P_\mu\Gamma^\mu|pn\sigma\rangle. \tag{55}$$

In order for (54) and (55) to be equal, it must be that

$$[a(n)]^2=-1.$$

This and the unitary of U_c, $a(n)\,a(-n)=1$, imply that

$$a(n)=\pm i\,\mathrm{sgn}\,n. \tag{56}$$

We may choose the + sign for convenience.

The V_μ are assumed to have the following C-transformation property:

$$U_cV_\mu U_c^{-1}=+cV_\mu, \tag{57}$$

where c is a phase. Like (33) and unlike (49), (51), (53), (56) this is an additional assumption. From (28) and (57) it follows that for $n'=n$,

$$\langle p'n\sigma'|V_\mu|pn\sigma\rangle=+c\,\langle p'\ -n\ \sigma'|V_\mu|p\ -n\ \sigma\rangle$$
$$=+c\,\mathring{D}(p'\ -n\ \sigma')\{F_1(-n)\gamma_\mu+iF_2(-n)\sigma_{\mu\nu}q^\nu$$
$$+F_3(-n)q_\mu\}D(p\ -n\ \sigma). \tag{58}$$

From Eqs. (A49)—(A53) we see that under the change $n\to -n$, the bilinears $\mathring{D}(p')D(p)$ and $\mathring{D}(p')\sigma_{\mu\nu}D(p)$ do not change sign, but $\mathring{D}(p')\gamma_\mu D(p)$ does change sign. Thus by comparing (58) and (28) we find

$$F_1(n)=-cF_1(-n),\quad F_2(n)=+cF_2(-n),$$
$$F_3(n)=+cF_3(-n). \tag{59}$$

These relations serve to restrict the number of arbitrary functions in (26). If, e.g., we choose $c=+1$, then according to (27) and (59) we have

$$F_1(n)=\mathrm{sgn}(n)\tilde{\tilde{f}}_1=-F_1(-n),$$
$$F_2(n)=\tilde{f}_2=F_2(-n),$$
$$F_3(n)=\tilde{f}_3=F_3(-n). \tag{60}$$

We are thus left with only three arbitrary scalar functions of the momenta, $\tilde{\tilde{f}}_1$, $\tilde{f}_2$, and $\tilde{f}_3$, which characterize the matrix elements (23).

For the axial vectors A_μ we assume the C-transformation property

$$U_cA_\mu U_c^{-1}=dA_\mu, \tag{61}$$

where d is a phase. In the same manner that we obtained (58), we now have

$$\langle p'n\sigma'|A_\mu|pn\sigma\rangle$$
$$=d\mathring{D}(p'\ -n\ \sigma')\{G_1(-n)\gamma_\mu\gamma_5+G_2(-n)\sigma_{\mu\nu}q^\nu\gamma_5$$
$$+G_3(-n)q_\mu\gamma_5\}D(p\ -n\ \sigma). \tag{62}$$

From Eqs. (A54)—(A58), we see that under the change $n\to -n$, the bilinears $\mathring{D}(p')\sigma_{\mu\nu}\gamma_5D(p)$ and $\mathring{D}(p')\gamma_5D(p)$ do not change sign, but $\mathring{D}(p')\gamma_\mu\gamma_5D(p)$ does change sign. Thus

$$G_1(n)=-dG_1(-n),\quad G_2(n)=dG_2(-n),$$
$$G_3(n)=dG_3(-n). \tag{63}$$

As before, we may thus characterize the matrix elements (31) with only three arbitrary scalar functions. By choosing $d=\pm 1$, we may write these in terms of the functions in Footnote 14.

IV. SPECTRUM GENERATING SU(3)

So far we have investigated the matrix elements of vector and axial vector operators V_μ, A_μ in the Dirac representation space of the relativistic symmetry. The physical system that could be described by this representation space would be a particular fermion—antifermion system. We shall now extend our description to the case where the physical system consists of a multiplet of baryon—antibaryon systems (whose external properties are assumed to be described by the Dirac representation). Though this formulation can be given for any group which classifies the baryons, we will choose SU(3).

Thus we assume that the space of physical states is the direct product space of the representation space of an SU(3) multiplet (octet) with the Dirac representation space:

$$\mathcal{H}^{SU(3)}(1,1)\otimes\mathcal{H}^{\Xi}(m,D). \tag{64}$$

It is generally well accepted that the transition between different baryon states, as they occur, e.g., in the semileptonic decays of baryons,

$$\text{baryon}\,\alpha\to\text{baryon}\,\alpha'+\text{lepton pair}\,l\nu,$$

are described by SU(3) octet Lorentz vector and axialvector operators. We will, therefore, investigate the matrix elements of SU(3)-octet Lorentz vector operators V_μ^β and SU(3)-octet Lorentz axialvector operators A_μ^β.[17] A basis system in this space is given by the direct product of the usual basis in $\mathcal{H}^{SU(3)}$ and the canonical basis system in $\mathcal{H}^D=\mathcal{H}(m,D)$,

$$|pn\sigma,\alpha\rangle = |pn\sigma\rangle\otimes|\alpha\rangle \tag{65}$$

where α represents I, I_3, Y.

Remark: Instead of the canonical basis system one could also have chosen the spinor basis system

$$|f^c_{j_3}(p),\alpha\rangle = f^c_{j_3}(p)\otimes|\alpha\rangle. \tag{66}$$

As long as no further connection between SU(3) and the spin SU(2) or the $SU(2)_{s_{ij}}$ is assumed, there is no reason to prefer one or the other basis system.

If SU(3) is not a symmetry group, but the mass is a function of the internal quantum numbers α,

$$m = m(\alpha), \tag{67}$$

then the space of the physical system will not be the direct product space (64), with only one mass, and we can no longer assume that P_μ commutes with SU(3). Instead of the relativistic symmetry $\mathfrak{S}_{P_\mu L_{\mu\nu}\Gamma_\mu}$ generated by the momenta P_μ and $L_{\mu\nu}$, Γ_μ, one may consider a relativistic symmetry generated by operators that do commute with SU(3). As an example we will assume that the velocity

$$\hat{P}_\mu = P_\mu M^{-1} \tag{68}$$

and $L_{\mu\nu}$, Γ_μ commute with the SU(3) which classifies the particles, i.e.,

$$[\hat{\mathfrak{S}}_{\hat{P}_\mu, L_{\mu\nu}, \Gamma_\mu}, \mathrm{SU}(3)_E] = 0. \tag{69}$$

Here $\hat{\mathfrak{S}}_{\hat{P}_\mu, L_{\mu\nu}, \Gamma_\mu}$ denotes the relativistic symmetry generated by $\hat{P}_\mu, L_{\mu\nu}$, Γ_μ {these fulfill the same algebraic relations (1)$\cdots$(8) (9) as P_μ, $L_{\mu\nu}$, Γ_μ if one asummes that $[\Gamma_\mu, M] = 0$}, and $SU(3)_E$ denotes the SU(3) which classifies the particles and which is now considered to be not a symmetry group but rather a spectrum generating group. Although there is no direct evidence for assumption (69),[18] it is not in evident contradiction with reality as is the assumption of an SU(3) symmetry group, i.e., the assumption that $\mathfrak{S}_{P_\mu, L_{\mu\nu}, \Gamma_\mu}$ commutes with SU(3).

By assumption (69), the space of physical states can be taken as the direct product space

$$\mathcal{H} = \mathcal{H}^{\mathrm{SU}(3)_E}(1,1)\otimes\mathcal{H}^{\hat{\mathfrak{S}}}(\hat{m}=1, D) \tag{70}$$

where $\mathcal{H}^{\hat{\mathfrak{S}}}(\hat{m}=1, D)$ denotes the Dirac representation space of $\hat{\mathfrak{S}}$ (the eigenvalue of $\hat{P}_\mu\,\hat{P}^\mu = 1$). The properties of $\mathcal{H}^{\hat{\mathfrak{S}}}$ are analogous to those of $\mathcal{H}^{\mathfrak{S}}$. Instead of the canonical basis of generalized momentum eigenvectors, one has the canonical basis of generalized velocity eigenvectors $|\hat{p}n\sigma\rangle$:

$$P_\mu M^{-1}|\hat{p}n\sigma\rangle = \hat{p}_\mu|\hat{p}n\sigma\rangle, \tag{71}$$

where $\hat{p}_\mu = p_\mu/m$. Correspondingly, there is a spinor basis $f^c_{j_3}(\hat{p})$. A basis system in $\mathcal{H}$ is then given by the direct product

$$|pn\sigma,\alpha\rangle = |\hat{p}n\sigma\rangle\otimes|\alpha\rangle. \tag{72}$$

After these preparations, our problem can be stated as the investigation of the properties of the matrix elements of the $SU(3)_E$-octet Lorentz vector and axial-vector operators V^β_μ, A^β_μ between the basis vectors (72).

Since V^β_μ is an $SU(3)_E$-octet operator,[19]

$$\begin{aligned}\langle\alpha'\hat{p}'n'\sigma'|V^\beta_\mu|\hat{p}n\sigma\alpha\rangle &= C(11,11,11,\gamma=1;\alpha\beta\alpha')\langle\hat{p}'n'\sigma'|V^{(f)}_\mu|\hat{p}n\sigma\rangle \\ &\quad + C(11,11,11,\gamma=2;\alpha\beta\alpha')\langle\hat{p}'n'\sigma'|V^{(d)}_\mu|\hat{p}n\sigma\rangle.\end{aligned} \tag{73}$$

The two reduced matrix elements $\langle\hat{p}'n'\sigma'|V^{(f,d)}_\mu|\hat{p}n\sigma\rangle$ define two operators $V^{(f,d)}_\mu$ in the space $\mathcal{H}^D$.[20] We may apply the results of the previous sections to these operators with only the modification of replacing p by $\hat{p}$, since the $V^{(f,d)}_\mu$ are operators in $\mathcal{H}^{\hat{\mathfrak{S}}}$ rather than $\mathcal{H}^{\mathfrak{S}}$.

Thus we have

$$\begin{aligned}\langle\hat{p}'n\sigma'|V^{(f,d)}_\mu|\hat{p}n\sigma\rangle &= \dot{D}(\hat{p}'n\sigma')\{\gamma_\mu F^{(f,d)}_1(\hat{q}^2) \\ &\quad + i\sigma_{\mu\nu}\hat{q}^\nu F^{(f,d)}_2(\hat{q}^2) + \hat{q}_\mu F^{(f,d)}_3(\hat{q}^2)\}D(\hat{p}n\sigma),\end{aligned} \tag{74}$$

where $\hat{q}_\mu = p'_\mu/m' - p_\mu/m$ and $D(\hat{p}) = D(p)$, i.e., the $D(p)$ are functions of the velocity only, as seen from Eq. (A11).

We shall now derive properties of the $F^{(f,d)}_i(\hat{q}^2)$ which follow from assumptions made for the V^β_μ. First we will consider charge conjugation. The charge conjugation operator should have the property

$$U_c|\hat{p}n\sigma,\alpha\rangle = a(n)|\hat{p}-n\,\sigma,\ \bar{\alpha}\rangle.$$

The derivation of

$$a(n) = i\,\mathrm{sgn}n$$

follows almost exactly as before, except that in calculating Eq. (56) we need

$$\Gamma_m|pn\sigma,\alpha\rangle = \sum_{\sigma'}|p-n\,\sigma'\,\bar{\alpha}\rangle\langle\sigma'|\Gamma_m|\sigma\rangle.$$

That is, since n and α are coupled, Γ_m must change α to $\bar{\alpha} = -\alpha$ as well as n to $-n$.

We assume the following C-transformation property for V^β_μ:

$$U_c^{-1}V^\beta_\mu U_c = -cV^{-\beta}_\mu, \tag{75}$$

where c is a phase factor. The matrix elements of (75) in the canonical basis are

$$\begin{aligned}\langle\alpha'p'n'\sigma'|U_c^{-1}V^\beta_\mu U_c|pn\sigma\alpha\rangle &= \langle-\alpha'p'-n'\sigma'|V^\beta_\mu|p-n\,\sigma-\alpha\rangle \\ &= -c\langle\alpha'p'n'\sigma'|V^{-\beta}_\mu|pn\sigma\alpha\rangle.\end{aligned} \tag{76}$$

Since the V^β_μ are octet operators, we may use (73) to rewrite (76):

$$\begin{aligned}\sum_{\gamma=1,2} C(\gamma;-\alpha,\beta,-\alpha')\langle p'-n'\,\sigma'|V^{(\gamma)}_\mu|p-n\,\sigma\rangle &= -c\sum_{\gamma=1,2} C(\gamma;\alpha,-\beta,\alpha')\langle p'n'\sigma'|V^{(\gamma)}_\mu|pn\sigma\rangle.\end{aligned} \tag{77}$$

[Instead of $V^{(f)}_\mu$ and $V^{(d)}_\mu$ we use the notation $V^{(\gamma=1)}_\mu$, $V^{(\gamma=2)}_\mu$ respectively.] Using the property of the Clebsch–Gordan coefficients that

$$\begin{aligned}C(\gamma=1;\alpha,-\beta,\alpha') &= -C(\gamma=1;-\alpha,\beta,-\alpha'), \\ C(\gamma=2;\alpha,-\beta,\alpha') &= +C(\gamma=2;-\alpha,\beta,-\alpha'),\end{aligned} \tag{78}$$

we see from (77) that

$$\begin{aligned}\langle p'-n'\,\sigma'|V^{(\gamma=1)}_\mu|p-n\,\sigma\rangle &= c\langle p'n'\sigma'|V^{(\gamma=1)}_\mu|pn\sigma\rangle, \\ \langle p'-n'\,\sigma'|V^{(\gamma=2)}_\mu|p-n\,\sigma\rangle &= -c\langle p'n'\sigma'|V^{(\gamma=2)}_\mu|pn\sigma\rangle,\end{aligned} \tag{79}$$

i.e., that the $V_\mu^{(\gamma)}$ satisfy

$$U_c^{-1} V_\mu^{(\gamma=1)} U_c = c V_\mu^{(\gamma=1)}, \tag{80}$$

$$U_c^{-1} V_\mu^{(\gamma=2)} U_c = -c V_\mu^{(\gamma=2)}. \tag{81}$$

According to (59) we have from (80)

$$\begin{aligned} F_1^{(\gamma=1)}(n) &= -cF_1^{(\gamma=1)}(-n), \\ F_2^{(\gamma=1)}(n) &= cF_2^{(\gamma=1)}(-n), \\ F_3^{(\gamma=1)}(n) &= cF_3^{(\gamma=1)}(-n) \end{aligned} \tag{82}$$

and likewise from (81)

$$\begin{aligned} F_1^{(\gamma=2)}(n) &= cF_1^{(\gamma=2)}(-n), \\ F_2^{(\gamma=2)}(n) &= -cF_2^{(\gamma=2)}(-n), \\ F_3^{(\gamma=2)}(n) &= -cF_3^{(\gamma=2)}(-n). \end{aligned} \tag{83}$$

Since c does not depend on n, relations (82), (83) [just as relations (59) did] will restrict the number of arbitrary functions needed to characterize the matrix elements of V_μ^β. If, e.g., we choose $c=+1$, then by (27) we have

$$\begin{aligned} &F_1^{(1)}(n) = \operatorname{sgn}(n)\tilde{\tilde{f}}_1^{(1)}, \quad F_2^{(1)}(n) = \tilde{f}_2^{(1)}, \\ &F_3^{(1)}(n) = \tilde{f}_3^{(1)}, \quad F_1^{(2)}(n) = \tilde{f}_1^{(2)}, \\ &F_2^{(2)}(n) = \operatorname{sgn}(n)\tilde{\tilde{f}}_2^{(2)}, \quad F_3^{(2)}(n) = \operatorname{sgn}(n)\tilde{\tilde{f}}_3^{(2)}. \end{aligned} \tag{84}$$

Likewise, if we assume for the axialvector operators,

$$U_c^{-1} A_\mu^\beta U_c = -dA_\mu^\beta, \tag{85}$$

then we obtain

$$\begin{aligned} G_1^{(\gamma=1)}(n) &= -dG_1^{(\gamma=1)}(-n), \\ G_2^{(\gamma=1)}(n) &= dG_2^{(\gamma=1)}(-n), \\ G_3^{(\gamma=1)}(n) &= dG_3^{(\gamma=1)}(-n) \end{aligned}$$

and

$$\begin{aligned} G_1^{(\gamma=2)}(n) &= dG_1^{(\gamma=2)}(-n), \\ G_2^{(\gamma=2)}(n) &= -dG_2^{(\gamma=2)}(-n), \\ G_3^{(\gamma=2)}(n) &= -dG_3^{(\gamma=2)}(-n). \end{aligned} \tag{86}$$

Besides the above assumption for the V_μ^β, A_μ^β, we assume that they have a definite A_T-transformation property:

$$A_T V_\mu^\beta A_T = \epsilon(\mu) V_\mu^\beta, \quad A_T A_\mu^\beta A_T = \epsilon(\mu) A_\mu^\beta \tag{87}$$

and a definite Hermiticity property

$$V_\mu^{\beta\dagger} = V_\mu^{-\beta}, \quad A_\mu^{\beta\dagger} = A_\mu^{-\beta}. \tag{88}$$

We will call V_μ^β and A_μ^β "first class with respect to the spectrum generating $SU(3)_E$" if (87) and (88) are fulfilled. This is in analogy with the first class condition when SU(3) is a symmetry group.

Since A_T does not affect the SU(3) quantum numbers, condition (87) leads by the same arguments as given in Sec. III to [see Eq. (39)]

$$F_i^{(f,d)}(\hat{q}^2) = \overline{F_i^{(f,d)}(\hat{q}^2)}, \quad i=1,2,3. \tag{89}$$

Similarly, for the axialvector operators we obtain [see Eq. (40)]

$$G_i^{(f,d)}(\hat{q}^2) = \overline{G_i^{(f,d)}(\hat{q}^2)}, \quad i=1,2,3. \tag{90}$$

We now use condition (88) to calculate

$$\begin{aligned} &\langle \alpha' \hat{p}' n' \sigma' \,|\, V_\mu^\beta \,|\, \hat{p} n \sigma \alpha \rangle \\ &= \overline{\langle \alpha \hat{p} n \sigma \,|\, V_\mu^{-\beta} \,|\, \hat{p}' n' \sigma' \alpha' \rangle} \\ &= \overline{C(11,11,11,\gamma=1;\alpha' - \beta\alpha)}\ \overline{\langle \hat{p} n \sigma \,|\, V_\mu^{(f)} \,|\, \hat{p}' n' \sigma' \rangle} \\ &\quad + \overline{C(11,11,11,\gamma=2;\alpha' - \beta\alpha)}\ \overline{\langle \hat{p} n \sigma \,|\, V_\mu^{(d)} \,|\, \hat{p}' n' \sigma' \rangle}. \end{aligned} \tag{91}$$

Comparing (73) with (91) and using the property of the SU(3) Clebsch–Gordan coefficients,

$$\overline{C(11,11,11,\gamma;\alpha' - \beta\alpha)} = C(11,11,11,\gamma;\alpha\beta\alpha'), \tag{92}$$

we obtain

$$\langle \hat{p}' n' \sigma' \,|\, V_\mu^{(f,d)} \,|\, \hat{p} n \sigma \rangle = \overline{\langle \hat{p} n \sigma \,|\, V_\mu^{(f,d)} \,|\, \hat{p}' n' \sigma' \rangle}. \tag{93}$$

(93) is the same as relation (42) of Sec. III (stating that $V_\mu^{(f,d)}$ are Hermitian operators) and so, by the same arguments as given there,

$$\begin{aligned} F_{1,2}^{(f,d)}(\hat{p}', \hat{p}) &= \overline{F_{1,2}^{(f,d)}((-\hat{q})^2)}, \\ F_3^{(f,d)}(\hat{p}', \hat{p}) &= -\overline{F_3^{(f,d)}((-\hat{q})^2)}. \end{aligned} \tag{94}$$

Likewise, the same calculations done for the A_μ^β yield [see Eq. (46)]

$$\begin{aligned} G_{1,3}^{(f,d)}(\hat{p}', \hat{p}) &= \overline{G_{1,3}^{(f,d)}((-\hat{q})^2)}, \\ G_2^{(f,d)}(\hat{p}', \hat{p}) &= -\overline{G_2^{(f,d)}((-\hat{q})^2)}. \end{aligned} \tag{95}$$

Then (89) and (94) together with (90) and (95) lead to

$$F_{1,2}^{(f,d)} \text{ real}, \quad F_3^{(f,d)} = 0, \tag{96}$$

$$G_{1,3}^{(f,d)} \text{ real}, \quad G_2^{(f,d)} = 0. \tag{97}$$

It is essential to remark that the above argument only holds for the $F_i^{(f,d)}(\hat{q}^2)$ and $G_i^{(f,d)}(\hat{q}^2)$ in the matrix elements between velocity eigenstates, and not for the usual "form factors" which appear in the decomposition of the matrix elements between momentum eigenstates. The reason for this is that, under assumption (69), the reduced matrix elements $\langle \hat{p}' n' \sigma' \,|\, V_\mu^{(f,d)} \,|\, \hat{p} n \sigma \rangle$ do not depend upon SU(3) quantum numbers, whereas the corresponding reduced matrix elements $\langle p' n' \sigma' \,|\, \tilde{V}_\mu^{(f,d)} \,|\, p n \sigma \rangle$ that appear in the matrix elements between momentum eigenstates

$$\begin{aligned} &\langle \alpha' p' n' \sigma' \,|\, V_\mu^\beta \,|\, p n \sigma \alpha \rangle \\ &= \sum_\gamma C(11,11,11,\gamma;\alpha\beta\alpha') \langle p' n' \sigma' \,|\, \tilde{V}_\mu^{(\gamma)} \,|\, p n \sigma \rangle \end{aligned} \tag{98}$$

do implicitly depend upon the SU(3) quantum numbers, so that expression (98) cannot follow from the Wigner–Eckart theorem.

V. COMPARISON WITH NONINVARIANT FORM FACTORS

In order to compare this theory with the experimental data, it is most convenient to derive formulas which relate the SU(3) invariant form factors F_i, G_i to the usual form factors f_i, g_i. This is because the data has traditionally been analyzed in terms of f_i, g_i, so that analysis programs written in terms of those form factors are already available. To derive these formulas, we will calculate a physical quantity (the decay rate) and compare the result to the conventional expression. The assumptions that we shall make will be mentioned as they are needed.

We assume that the process is weak in the sense that the part T of the Hamiltonian which is responsible for the decay causes a negligible level shift.[21] The initial decay rate is then given by (lowest order perturbation theory):

$$\Gamma = 2\pi \sum_b \sum_{A,A'} \delta(E_{A'} - E_b)\, T_{bA} \langle A | W(0) | A' \rangle \overline{T}_{bA'}, \tag{99}$$

where $W(t)$ is the statistical operator that describes the ensemble of decaying hadrons,

$$T_{bA} = \langle b | T | A \rangle, \quad \overline{T}_{bA'} = \langle A' | T | b \rangle.$$

$|A\rangle$, $|B\rangle$, $\cdots$ denote the eigenvectors of a complete system of commuting observables (CSCO), $|a\rangle$, $|b\rangle$, $\cdots$ denote the corresponding free eigenvectors [i.e., $|a\rangle$ denotes the eigenvector of the CSCO for $T=0$ that has the same eigenvalues as $|A\rangle$. This is always possible when the level shift can be neglected], $\sum_{AA'}$ means summation over all the eigenvectors of the CSCO, and $\sum_b$ means summation over all those eigenvalues whose eigenvectors span the space of final states that are observed. The sums in (99) are discrete if one chooses as the CSCO a set of operators with discrete spectrum. It is often more convenient to use observables with a continuous spectrum for the description of the experimental situation. In particular the momentum or velocity is an observable which is easily accessible in an experiment (though a physical system cannot be in an eigenstate of this observable). If we choose the vectors $|A\rangle$ to be the generalized eigenvectors of the momentum operator,

$$|A\rangle = |\boldsymbol{p}(A), \eta_A\rangle \tag{100}$$

with the normalization

$$\langle \boldsymbol{p}'\eta_{A'} | \boldsymbol{p}\eta_A \rangle = 2E(\boldsymbol{p})\delta^3(\boldsymbol{p}-\boldsymbol{p}')\delta_{\eta'_A \eta_A},$$

$$E(\boldsymbol{p}) = (m^2 + \boldsymbol{p}^2)^{1/2}, \tag{100'}$$

where η_A denotes the eigenvalues of the other observables that together with the $P_\mu^{(A)}$ make up the CSCO, then

$$\sum_A = \sum_{\eta_A} \int \frac{d^3\boldsymbol{p}}{2E(\boldsymbol{p})}. \tag{101}$$

For reasons which have been given in Sec. IV and which will again be discussed below, we may want to choose generalized eigenvectors of the velocity operator P_μ/M rather than the momentum operator, i.e., choose the canonical basis vector (72) of Sec. IV:

$$|A\rangle = |\hat{\boldsymbol{p}}_A, \hat{\eta}_A\rangle = |\hat{p} n \sigma \alpha\rangle. \tag{102}$$

[It may, of course, happen that the $\hat{\eta}_A$ in (102) are the same as the η_A in (100). This will be the case if the set of commuting observables whose set of eigenvalues is η_A commutes with P_μ as well as $\hat{P}_\mu$.]

The normalization of the eigenvectors is conveniently chosen:

$$\langle \hat{\eta}'\hat{\boldsymbol{p}} | \hat{p}\hat{n} \rangle = 2\hat{E}(\boldsymbol{p})\delta^3(\hat{\boldsymbol{p}} - \hat{\boldsymbol{p}}')\,\delta_{\hat{\eta}\hat{\eta}'}, \tag{102'}$$

where

$$\hat{E}(\boldsymbol{p}) = (1 + \boldsymbol{p}^2/m^2)^{1/2}.$$

Then

$$\sum_A = \sum_{\hat{\eta}} \int d^3\hat{\boldsymbol{p}}/2\hat{E}. \tag{103}$$

We assume that the eigenvectors $|A\rangle$ are baryon states, i.e., that they are the canonical basis vectors

$$|A\rangle = |\hat{p}, n, \sigma, \alpha\rangle, \tag{104}$$

which were discussed in Sec. IV. The baryon has a definite mass which we assume to be a function of the SU(3) quantum numbers $\alpha (= I, I_3, Y$ and the Casimir operators), and the decaying baryon state $W(t)$ is prepared to have definite internal quantum numbers α and $n = \frac{1}{2}$. If the polarization has not been measured in the preparation, one will have to average over σ. The statistical operator of the decaying system is then

$$W(0) = \tfrac{1}{2}\sum_\sigma \int (d^3\hat{k}/2\hat{E}_k) \int (d^3\hat{k}'/2\hat{E}_{k'}) \, |\hat{k} n \sigma \alpha\rangle\langle \alpha \sigma n \hat{k}'|$$

$$\times \phi(\hat{k})\,\overline{\phi}(\hat{k}'). \tag{105}$$

Instead of the generalized eigenvectors $|\hat{p} n \sigma \alpha\rangle$, one usually uses the generalized eigenvectors $|p n \sigma \alpha\rangle$ of the momentum operator P_μ. In that case one has, instead of (105),

$$W = \tfrac{1}{2}\sum_\sigma \int (d^3k/2E_k) \int (d^3k'/2E_{k'}) \, |k n \sigma \alpha\rangle\langle \alpha \sigma n k' | \, \phi(k)\,\overline{\phi}(k'). \tag{106}$$

As long as the observables whose eigenvalues are $\alpha(I, I_3, Y, \cdots)$ commute with the momentum, the $|p n \sigma \alpha\rangle$ will exist and (106) will be equivalent to (105). If, however, one of the operators whose eigenvalues are α, e.g., the Casimir operators of $SU(3)_E$, does not commute with P_μ, then $|p n \sigma \alpha\rangle$ does not exist. Therefore, under assumption (69) we use the generalized eigenvectors (104) with the normalization (102'). This will be the case for the baryon octet.

If we use the normalization (102') and the measure (103), then the requirement that W be normalized,

$$\mathrm{Tr} W = 1 \tag{107}$$

can be written as

$$\mathrm{Tr} W = \sum_{\alpha,\sigma} \int (d^3\hat{k}/2\hat{E}_k) \langle \alpha \sigma n \hat{k} | \, W \, | \hat{k} n \sigma \alpha \rangle$$

$$= \sum_{\alpha,\sigma} \int (d^3\hat{k}/2\hat{E}_k)\, \tfrac{1}{2}\sum_{\sigma'} \int (d^3\hat{k}'/2\hat{E}_{k'}) \int (d^3\hat{k}''/2\hat{E}_{k''})$$

$$\times \langle \alpha \sigma n \hat{k} \, | \, \hat{k}' n \sigma \alpha \rangle \langle \alpha \sigma n \hat{k}'' \, | \, \hat{k} n \sigma \alpha \rangle \, \phi(\hat{k}')\,\overline{\phi}(\hat{k}'')$$

$$= \int (d^3\hat{k}/2\hat{E}_k) \, |\phi(\hat{k})|^2 = 1. \tag{108}$$

The physical meaning of the distribution function $\phi(\hat{p})$ is that $|\phi(\hat{p})|^2$ is the probability that the momentum value of the decaying particle is $p = m\hat{p}$. For the idealized case that the system of decaying particles is prepared to have the exact momentum p_I the probability $|\phi(\hat{p})|^2$ will be zero unless $p = p_I$. In the normalization (108) this is expressed by

$$|\phi(\hat{p})|^2 = 2\hat{E}_{p_I}\delta^3(\hat{p} - \hat{p}_I). \tag{109}$$

If we choose the generalized eigenvectors $|p n \sigma \alpha\rangle$ for $|A\rangle$, then we obtain the decay rate by substituting (105) in (99) and using (102'):

$$\Gamma = 2\pi \sum_b \tfrac{1}{2}\sum_\sigma \int (d^3\hat{k}/2\hat{E}_k)(d^3\hat{k}'/2\hat{E}_{k'})\, \delta(E_b - E_k)\, \phi(\hat{k})\,\overline{\phi}(\hat{k}')$$

$$\times \langle b | T | \hat{k} n \sigma \alpha \rangle \langle \alpha n \sigma \hat{k}' | T | b \rangle. \tag{110}$$

We have so far only specified that the decaying state is a one-hadron state with $s=\frac{1}{2}$ and definite n, α. We now consider in particular the process

$$\alpha \to \alpha' + l + \bar{\nu},$$

where α, α' are baryons and l is either an electron or muon. For the vectors b we then have to take

$$|b\rangle = |\hat{p}', n=+\tfrac{1}{2}, \sigma', \alpha'\rangle \otimes |p_l, n=+\tfrac{1}{2}, \sigma_l, \alpha_l\rangle \otimes |p_{\bar{\nu}}, n=-\tfrac{1}{2}, \sigma_{\bar{\nu}}, \alpha_{\bar{\nu}}\rangle,$$

which we abbreviate

$$|b\rangle = |p', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu}\rangle.$$

We normalize the lepton states in the conventional way, i.e., according to (100′). Thus $\sum_b$ becomes

$$\sum_b = \sum_{\sigma_{\bar{\nu}}, \sigma_l, \sigma'} \int (d^3p_{\bar{\nu}}/2E_{\bar{\nu}})(d^3p_l/2E_l)(d^3\hat{p}'/2\hat{E}'). \tag{111}$$

Using this, (110) becomes

$$\Gamma = \pi \sum_{\text{spins}} \int \frac{d^3p_{\bar{\nu}}}{2E_{\bar{\nu}}} \frac{d^3p_l}{2E_l} \frac{d^3\hat{p}'}{2\hat{E}'} \int \frac{d^3\hat{k}}{2\hat{E}_k} \frac{d^3\hat{k}'}{2\hat{E}_{k'}} \times \delta(E_{\bar{\nu}} + E_l + E' - E)\, \phi(\hat{k})\, \overline{\phi}(\hat{k}') \times \langle \hat{p}', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} | T | \hat{k}\alpha \rangle \langle \alpha \hat{k}' | T | \hat{p}', \alpha'; p_l, l; p_\nu, \bar{\nu} \rangle. \tag{112}$$

The calculations leading to (112) contain only well-accepted principles of quantum mechanics applied to the case where the $\hat{P}_\mu = P_\mu M^{-1}$ rather than the P_μ commute with the other observables. They have been given here in order to show that there is absolutely nothing mysterious about the velocity eigenvectors $|\hat{p}, \alpha\rangle$ and that under assumption (69) it is advantageous to use them instead of the momentum eigenvectors $|p, \alpha\rangle$.

We will now specify the operator T further. We assume that T conserves total momenta.[22] Its matrix element may then be written as the product of the momentum δ function and a reduced matrix element:

$$\langle \hat{p}, \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} | T | k, \alpha \rangle = \delta^3(p' + p_l + p_{\bar{\nu}} - k) \langle\langle \alpha' l \bar{\nu} | T | \alpha \rangle\rangle. \tag{113}$$

Inserting this into (112) and using $\delta^3(p) = m^{-3}\delta^3(p/m)$ gives

$$\Gamma = \pi \sum_{\text{spins}} \int \frac{d^3p_{\bar{\nu}}}{2E_{\bar{\nu}}} \frac{d^3p_l}{2E_l} \frac{d^3\hat{p}'}{2\hat{E}'} \delta(E - E' - E_{\bar{\nu}} - E_l) \times \left| \phi\left(\frac{p_l + p_{\bar{\nu}} + p'}{m}\right) \right|^2 \frac{1}{m^6} \left(\frac{1}{2\hat{E}}\right)^2 \left| \left\langle\left\langle \hat{p}', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} \right| \times T \left| \frac{p_l + p_{\bar{\nu}} + p'}{m}\ \alpha \right\rangle\right\rangle \right|^2,$$

where

$$|\langle\langle \alpha' l \bar{\nu} | T | \alpha \rangle\rangle|^2 = \langle\langle \alpha' l \bar{\nu} | T | \alpha \rangle\rangle \langle\langle \alpha | T | \alpha' l \bar{\nu} \rangle\rangle.$$

This can be brought into a familiar form if one makes the idealized assumption that the ensemble of decaying baryons has a definite momentum p. Then $|\phi|^2$ is given by (109), and we obtain

$$\Gamma = \pi \int \frac{d^3p_l}{2E_l} \frac{d^3p_{\bar{\nu}}}{2E_{\bar{\nu}}} \frac{d^3p'}{2E'} \sum_{\text{spins}} \delta^4(p - p' - p_l - p_{\bar{\nu}}) \times \frac{1}{m^2 m'^2} \frac{1}{2E} |\langle\langle \hat{p}', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} | T | \hat{p}, \alpha \rangle\rangle|^2. \tag{114}$$

The decay rate is conventionally calculated using the momentum eigenvectors (100) rather than the velocity eigenvectors (102). By carrying out the above calculation using (100) and (100′) rather than (102) and (102′), we obtain the conventional expression[23] (except for irrelevant normalization factors of 2π to some power or different normalizations for the lepton measure):

$$\Gamma = \frac{1}{2(2\pi)^5} \int \frac{d^3p_l}{2E_l} \frac{d^3p_{\bar{\nu}}}{2E_{\bar{\nu}}} \frac{d^3p'}{2E'} \times \sum_{\text{spins}} \delta^4(p - p' - p_l - p_{\bar{\nu}}) \times \frac{1}{2E} |\langle\langle p', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} | T | p, \alpha \rangle\rangle|^2. \tag{115}$$

By comparing (114) and (115) we see that we may make the identification[24]

$$\pi \sum_{\text{spins}} \left(\frac{1}{mm'}\right)^2 |\langle\langle \hat{p}', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} | T | \hat{p}, \alpha \rangle\rangle|^2 = \frac{1}{2(2\pi)^5} \sum_{\text{spins}} |\langle\langle p', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} | T | p, \alpha \rangle\rangle|^2. \tag{116}$$

The difference between (114) and the conventional expression (115) is that the conventional invariant matrix element is a function of the momenta of the baryons whereas in (114) it is a function of their velocities, and this results in a factor of $1/mm'$.

We now have to specify the operator T further. We do this in analogy to the usual $V-A$ product form of the weak interaction. We assume that the matrix element of T is the product of a leptonic part and a hadronic part. The leptonic part is given by the usual $V-A$ matrix element for leptons, i.e., the lepton pair is treated as noninteracting particles (lowest order perturbation):

$$\langle\langle \hat{p}', \alpha'; p_l, l; p_{\bar{\nu}}, \bar{\nu} | T | \hat{p}, \alpha \rangle\rangle = \bar{u}(p_l) \gamma^\lambda (1 - \gamma_5) v(p_{\bar{\nu}}) \langle\langle \hat{p}' \alpha' | H_\lambda | \hat{p}\alpha \rangle\rangle, \tag{117}$$

where we have used the conventional notation for the lepton spinors. The conventional expression is the same as (117) except that $\hat{p}$, $\hat{p}'$ are replaced by p, p'.

We assume that the transition operator in the hadron subspace H_λ is the sum of a vector and axialvector, and further that it is the sum of the V_μ^α, A_μ^α of Sec. IV with weights $C(\alpha)$:

$$H_\lambda = g \sum_\alpha C(\alpha)(V_\mu^\alpha + A_\mu^\alpha). \tag{118}$$

If we take V, A, $C(\alpha)$ to be dimensionless, then the constant g which expresses the strength of the interaction has the dimension of mass, because the reduced matrix element $\langle\langle |T| \rangle\rangle$ has the dimension of (mass)^2. For a Cabibbo-type model, one would choose $C(\pm 1) = \cos\theta_C$, $C(\pm 2) = \sin\theta_C$, $C(\pm 3) = 0$, where θ_C is the Cabibbo angle.

Finally, we use (73) and (74), and their analogs for A_μ^α, to write

$$\langle\langle \hat{p}' \alpha' | H_\lambda | \hat{p}\alpha \rangle\rangle = \langle \hat{p}', n' = \tfrac{1}{2}, \sigma', \alpha' | H_\lambda | \hat{p}, n = \tfrac{1}{2}, \sigma, \alpha \rangle = g\hat{D}(\hat{p}', \tfrac{1}{2}, \sigma') \{ \gamma_\lambda F_1^{\alpha'\alpha}(\hat{q}^2) + i\sigma_{\lambda\nu}\hat{q}^\nu F_2^{\alpha'\alpha}(\hat{q}^2) + \hat{q}_\lambda F_3^{\alpha'\alpha}(\hat{q}^2) + \gamma_\lambda\gamma_5 G_1^{\alpha'\alpha}(\hat{q}^2) + i\sigma_{\lambda\nu}\hat{q}^\nu\gamma_5 G_2^{\alpha'\alpha}(\hat{q}^2) + \hat{q}_\lambda\gamma_5 G_3^{\alpha'\alpha}(\hat{q}^2) \} D(\hat{p}, \tfrac{1}{2}, \sigma), \tag{119}$$

where we have used the abbreviations

$$F_i^{\alpha'\alpha}(\hat{q}^2)=\sum_\beta C(\beta)\,F_i^{\alpha'\beta\alpha}(\hat{q}^2)$$
$$=\sum_\beta C(\beta)\sum_{\gamma=1,2} C(11,11,11,\gamma;\alpha\beta\alpha')\,F_i^{(\gamma)}(\hat{q}^2),$$
$$G_i^{\alpha'\alpha}(\hat{q}^2)=\sum_\beta C(\beta)\,G_i^{\alpha'\beta\alpha}(\hat{q}^2)$$
$$=\sum_\beta C(\beta)\sum_{\gamma=1,2} C(11,11,11,\gamma;\alpha\beta\alpha')\,G_i^{(\gamma)}(\hat{q}^2). \tag{120}$$

On the other hand, the conventional expression for the hadronic reduced matrix element is of the form

$$\langle\langle p'\alpha'|H_\lambda|p\alpha\rangle\rangle=(G/\sqrt{2})\bar{u}(p')\{\gamma_\lambda f_1^{\alpha'\alpha}(q^2)+\sigma_{\lambda\nu}q^\nu f_2^{\alpha'\alpha}(q^2)/m+q_\lambda f_3^{\alpha'\alpha}(q^2)/m+\gamma_\lambda\gamma_5 g_1^{\alpha'\alpha}(q^2)+\sigma_{\lambda\nu}\gamma_5 q^\nu g_2^{\alpha'\alpha}(q^2)/m+q_\lambda\gamma_5 g_3^{\alpha'\alpha}(q^2)/m\}u(p), \tag{121}$$

where m is the mass of the decaying baryon, $f_i^{\alpha'\alpha}(q^2)=\sum_\beta C(\beta)f_i^{\alpha'\beta\alpha}(q^2)$, etc., and $C(\beta)$ as in (118), could be the Cabibbo factor.

The leptonic reduced matrix element is the same as that used in (117). Thus for both cases, the leptonic parts of (116) are

$$\sum_{\text{spins}}\{\bar{u}(p_l)\gamma^\lambda(1-\gamma_5)v(p_{\bar{\nu}})\bar{v}(p_{\bar{\nu}})\gamma^\mu(1-\gamma_5)u(p_l)\}$$
$$=-8(p_l^\mu p_{\bar{\nu}}^\lambda+p_l^\lambda p_{\bar{\nu}}^\mu-g^{\lambda\mu}\mathbf{p}_l\cdot\mathbf{p}_{\bar{\nu}}+i\epsilon^{\lambda\mu\kappa\sigma}(p_l)_\kappa(p_{\bar{\nu}})_\sigma). \tag{122}$$

Since the leptonic part (122) is the same on both sides of (116), we may use the abbreviation (lepton part)$^{\lambda\mu}$ to represent the term given by (122). By using (119) and (121) we may then write (116) as

$$\pi(1/m'm)^2(\text{lepton part})^{\lambda\mu}\sum_{\text{spins}} g^2[\mathring{D}(\hat{p}',\tfrac{1}{2},\sigma')\{\gamma_\lambda F_1^{\alpha'\alpha}+\cdots+\hat{q}_\lambda\gamma_5 G_3^{\alpha'\alpha}\}D(\hat{p},\tfrac{1}{2},\sigma)\mathring{D}(\hat{p},\tfrac{1}{2},\sigma)\{\gamma_\mu\bar{F}_1^{\alpha'\alpha}+\cdots+\hat{q}_\mu\gamma_5\bar{G}_3^{\alpha'\alpha}\}D(\hat{p}',\tfrac{1}{2},\sigma)]$$
$$=(1/2(2\pi)^5)(\text{lepton part})^{\lambda\mu}\sum_{\text{spins}}\tfrac{1}{2}G^2[\bar{u}(p')\{\gamma_\lambda f_1^{\alpha'\alpha}+\cdots+q_\lambda g_3^{\alpha'\alpha}/m\}u(p)\bar{u}(p)\{\gamma_\mu\bar{f}_1^{\alpha'\alpha}+\cdots+q_\mu\bar{g}_3^{\alpha'\alpha}/m\}u(p')]. \tag{123}$$

By using the projection operators (A59) and the conventional projection operator for the Dirac spinors,[25]

$$\sum_{\text{spin}} u_\alpha(p)\bar{u}_\beta(p)=(\not{p}+m)_{\alpha\beta},$$

we may write (123) in terms of traces:

$$\pi(\text{lepton part})^{\lambda\mu}(1/mm')^2g^2\,\text{Tr}[(\gamma_\lambda F_1^{\alpha'\alpha}+\cdots+\hat{q}_\lambda\gamma_5 G_3^{\alpha'\alpha})(\hat{\not{p}}+1)(\gamma_\mu\bar{F}_1^{\alpha'\alpha}+\cdots+\hat{q}_\mu\gamma_5\bar{G}_3^{\alpha'\alpha})(\hat{\not{p}}'+1)]$$
$$=[1/2(2\pi)^5](\text{lepton part})^{\lambda\mu}\tfrac{1}{2}G^2\,\text{Tr}[(\gamma_\lambda f_1^{\alpha'\alpha}+\cdots+q_\lambda\gamma_5 g_3^{\alpha'\alpha}/m)(\not{p}+m)\times(\gamma_\mu\bar{f}_1^{\alpha'\alpha}+\cdots+q_\mu\gamma_5\bar{g}_3^{\alpha'\alpha}/m)(\not{p}'+m')]. \tag{124}$$

By evaluating the traces in (124) and comparing the two sides of the equation, one could find relations between the two sets of form factors. Instead of doing that, however, it will be much easier to define new functions (we suppress the $\alpha'\alpha$ indices):

$$\mathcal{F}_1=F_1+2F_2,\quad \mathcal{G}_1=G_1,$$
$$\mathcal{F}_2=-F_2-F_3,\quad \mathcal{G}_2=-G_2-G_3, \tag{125}$$
$$\mathcal{F}_3=-F_2+F_3,\quad \mathcal{G}_3=-G_2+G_3$$

and

$$\mathcal{f}_1=f_1+(m+m')f_2/m,\quad \mathcal{g}_1=g_1+(m'-m)g_2/m,$$
$$\mathcal{f}_2=-f_2-f_3,\quad \mathcal{g}_2=-g_2-g_3, \tag{126}$$
$$\mathcal{f}_3=-f_2+f_3,\quad \mathcal{g}_3=-g_2+g_3.$$

Inserting (125) and (126) into (124) then gives

$$\pi(\text{lepton part})^{\lambda\mu}(1/m'm)^2g^2\,\text{Tr}[(\gamma_\lambda\mathcal{F}_1+\hat{p}_\lambda\mathcal{F}_2+\hat{p}'_\lambda\mathcal{F}_3+\gamma_\lambda\gamma_5\mathcal{G}_1+\hat{p}_\lambda\gamma_5\mathcal{G}_2+\hat{p}'_\lambda\gamma_5\mathcal{G}_3)(\hat{\not{p}}+1)(\gamma_\lambda\bar{\mathcal{F}}_1+\cdots+\hat{p}'_\lambda\gamma_5\bar{\mathcal{G}}_3)(\hat{\not{p}}'+1)]$$
$$=[1/2(2\pi)^5](\text{lepton part})^{\lambda\mu}(m'm)\tfrac{1}{2}G^2\,\text{Tr}\{[\gamma_\lambda\mathcal{f}_1+\hat{p}_\lambda(m\mathcal{f}_2/m)+\hat{p}'_\lambda(m'\mathcal{f}_3/m)+\gamma_\lambda\gamma_5\mathcal{g}_1+\hat{p}_\lambda(m\mathcal{g}_2/m)+\hat{p}'(m'\mathcal{g}_3/m)]\times(\hat{\not{p}}+1)[\gamma_\lambda\bar{\mathcal{f}}_1+\cdots+\hat{p}'_\lambda(m'\bar{\mathcal{g}}_3/m)]\}, \tag{127}$$

where we have used $\not{p}=m\hat{\not{p}}$ and $\not{p}'=m'\hat{\not{p}}'$ in the rhs. By inspecting both sides of (127) we find

$$\pi(\mathcal{g}/m'm)\mathcal{F}_1=[1/2(2\pi)^5]\sqrt{m'm}\,(G/\sqrt{2})(\mathcal{f}_1),$$
$$\pi(\mathcal{g}/m'm)\mathcal{F}_2=[1/2(2\pi)^5]\sqrt{m'm}\,(G/\sqrt{2})(m\mathcal{f}_2),$$
$$\pi(\mathcal{g}/m'm)\mathcal{F}_3=[1/2(2\pi)^5]\sqrt{m'm}\,(G/\sqrt{2})(m'\mathcal{f}_3),$$
$$\pi(\mathcal{g}/m'm)\mathcal{G}_1=[1/2(2\pi)^5]\sqrt{m'm}\,(G/\sqrt{2})(\mathcal{g}_1),$$
$$\pi(\mathcal{g}/m'm)\mathcal{G}_2=[1/2(2\pi)^5]\sqrt{m'm}\,(G/\sqrt{2})(m\mathcal{g}_2),$$
$$\pi(\mathcal{g}/m'm)\mathcal{G}_3=[1/2(2\pi)^5]\sqrt{m'm}\,(G/\sqrt{2})(m'\mathcal{g}_3). \tag{128}$$

From these we obtain, by using (125) and (126) again,

$$f_1^{\alpha'\alpha}=(2\pi)^4\frac{\mathcal{g}}{G}\left(\frac{2}{m'm}\right)^{3/2}\left\{F_1^{\alpha'\alpha}+F_2^{\alpha'\alpha}\left(2-\frac{(m+m')^2}{2mm'}\right)+F_3^{\alpha'\alpha}\,\frac{m^2-m'^2}{2mm'}\right\},$$

$$\frac{f_2^{\alpha'\alpha}}{m}=4\pi^4\frac{\mathcal{g}}{G}\left(\frac{2}{m'm}\right)^{5/2}\times\{(m'+m)F_2^{\alpha'\alpha}+(m'-m)F_3^{\alpha'\alpha}\},$$

$$\frac{f_3^{\alpha'\alpha}}{m}=4\pi^4\frac{\mathcal{g}}{G}\left(\frac{2}{m'm}\right)^{5/2}\times\{(m'-m)F_2^{\alpha'\alpha}+(m'+m)F_3^{\alpha'\alpha}\},$$

$$g_1^{\alpha'\alpha}=(2\pi)^4\frac{\mathcal{g}}{G}\left(\frac{2}{m'm}\right)^{3/2}\times\left\{G_1^{\alpha'\alpha}+G_2^{\alpha'\alpha}\,\frac{m^2-m'^2}{2mm'}-G_3\frac{(m-m')^2}{2mm'}\right\},$$

$$\frac{g_2^{\alpha'\alpha}}{m}=4\pi^4\frac{\mathcal{g}}{G}\left(\frac{2}{m'm}\right)^{5/2}\times\{(m'+m)G_2^{\alpha'\alpha}+(m'-m)G_3^{\alpha'\alpha}\},$$

$$\frac{g_3^{\alpha'\alpha}}{m}=4\pi^4\frac{\mathcal{g}}{G}\left(\frac{2}{m'm}\right)^{5/2}$$

$$\times\{(m'-m)\,G_2{}^{\alpha'\alpha}+(m'+m)\,G_3{}^{\alpha'\alpha}\}. \tag{129}$$

We thus see that condition (97) does not require g_2 to vanish except in the symmetry limit.

APPENDIX

In this section we give a brief derivation of the transformation matrices D and $\mathring{D}$. The Lorentz transformation properties (11) and (13) mean that we may write

$$|\,pn\sigma\rangle = U(L^{-1}(p))(\phi(\mathbf{p}=0)\otimes|n\sigma\rangle), \tag{A1}$$

$$f^c_{j_3}(p) = f(p)\otimes f^c_{j_3}, \tag{A2}$$

where $\phi(p)$ and $f(p)$ are generalized eigenvectors of P_μ in the $(m, s=0)$ irreducible representation of $\rho_{P_\mu, M_{\mu\nu}}$, and $|n,\sigma\rangle$, $f^c_{j_3}$ are basis vectors of the four-dimensional $(1/2,3/2)\oplus(1/2,-3/2)$ representation spaces of $SO(3,1)_{\Gamma_i, S_{ij}}$ and $SO(3,1)_{S_{\mu\nu}}$, respectively.[6]

For the special case $J=\Gamma_\mu=I\otimes\Gamma^I_\mu$ and $\mathbf{p}=0$, Eq. (19) becomes, by using (A1),

$$\langle n'\sigma'|\,\Gamma^I_\mu\,|n\sigma\rangle = \mathring{D}^{n'c'}_{\sigma' j'_3}(0)\,\langle \mathring{f}^{c'}_{j'_3}|\,\Gamma^I_\mu\,|f^c_{j_3}\rangle D^{cn}_{j_3\sigma}(0), \tag{A3}$$

i.e., we see that $D(0)$, $\mathring{D}(0)$ transform the γ matrices from the spinor to the canonical basis. By using the facts that (1) n is the eigenvalue of Γ_0 for $\mathbf{p}=0$ in the canonical basis, (2) Γ_0 transforms between the $c=+3/2$ and $c=-3/2$ states in the spinor basis, and (3) Γ_i is a vector operator in a known representation space of $SO(3,1)$, it is possible to explicitly determine these γ matrices (as done in Ref. 6). In the canonical basis, we find

$$\langle n'\sigma'|\,\Gamma_0\,|n\sigma\rangle = (\gamma_0)^{n'n}_{\sigma'\sigma\ \mathrm{CAN}}$$

$$= \begin{array}{c} \begin{matrix} n=1/2 & n=-1/2 \end{matrix} \\ \begin{pmatrix} \delta_{\sigma'\sigma} & 0 \\ 0 & -\delta_{\sigma'\sigma} \end{pmatrix} \end{array} \begin{matrix} n'=1/2 \\ n'=-1/2 \end{matrix}, \tag{A4}$$

$$(\gamma_i)^{n'n}_{\sigma'\sigma\ \mathrm{CAN}} = i \begin{array}{c} \begin{matrix} n=+1/2 & n=-1/2 \end{matrix} \\ \begin{pmatrix} 0 & (\sigma_i)_{\sigma'\sigma} \\ (\sigma_i)_{\sigma'\sigma} & 0 \end{pmatrix} \end{array} \begin{matrix} n'=+1/2 \\ n'=-1/2 \end{matrix}$$

while in the spinor basis

$$\langle \mathring{f}^{c'}_{j'_3}|\,2\Gamma_0\,|f^c_{j_3}\rangle = (\gamma_0)^{c'c}_{j'_3 j_3\ \mathrm{SPIN}}$$

$$= \begin{array}{c} \begin{matrix} c=+3/2 & c=-3/2 \end{matrix} \\ \begin{pmatrix} 0 & \delta_{j'_3 j_3} \\ \delta_{j'_3 j_3} & 0 \end{pmatrix} \end{array} \begin{matrix} c'=+3/2 \\ c'=-3/2 \end{matrix},$$

$$(\gamma_i)^{c'c}_{j'_3 j_3\ \mathrm{SPIN}} = \begin{array}{c} \begin{matrix} c=+3/2 & c=-3/2 \end{matrix} \\ \begin{pmatrix} 0 & -(\sigma_i)_{j'_3 j_3} \\ (\sigma_i)_{j'_3 j_3} & 0 \end{pmatrix} \end{array} \begin{matrix} c'=+3/2 \\ c'=-3/2 \end{matrix}, \tag{A5}$$

where the σ_i are the usual Pauli matrices:

$$\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}. \tag{A6}$$

The matrix which relates these two sets of γ matrices is easily found to be (within a phase)

$$D^{cn}_{j_3\sigma}(0) = \frac{1-i}{2}\,\delta_{j_3\sigma} \begin{array}{c} \begin{matrix} n=+1/2 & n=-1/2 \end{matrix} \\ \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \end{array} \begin{matrix} c=+3/2 \\ c=-3/2 \end{matrix} \tag{A7}$$

and

$$\mathring{D}(0) = D^\dagger(0), \tag{A8}$$

since $D(0)$ is unitary.

For general p, we have by (16) and (A1),

$$\begin{aligned} |\,f^c_{j_3}(p)\rangle D^{cn}_{j_3\sigma}(p) &= |\,pn\sigma\rangle \\ &= U(L^{-1}(p))\,|\,\mathbf{p}=0, n\sigma\rangle \\ &= U(L^{-1}(p))\,|\,f^c_{j_3}(0)\rangle D^{cn}_{j_3\sigma}(0), \text{ using (6), and} \\ &= |\,f^c_{j'_3}(p)\rangle D^{1/2\,1/2(k_0c)}_{j'_3 j_3}(L^{-1}(p))\,D^{cn}_{j_3\sigma}(0), \end{aligned} \tag{A9}$$

by using (13), so that

$$D^{cn}_{j_3\sigma}(p) = D^{1/2\,1/2(k_0c)}_{j'_3 j_3}(L^{-1}(p))\,D^{cn}_{j_3\sigma}(0). \tag{A10}$$

By using explicit expressions for the $k_0=1/2$, $c=\pm 3/2$ Lorentz transformations,[1,6] (A10) may be written in terms of the γ matrices (A5),

$$D^{cn}_{j_3\sigma}(p) = C(1+\hat{p}_\mu\gamma^\mu\gamma_0)^{cc'}_{j_3 j'_3}\,D^{c'n}_{j'_3\sigma}(0), \tag{A11}$$

where $\hat{p}_\mu = p_\mu/m$ and

$$C=[2(1+\hat{p}_0)]^{-1/2}. \tag{A12}$$

However, we do not obtain $\mathring{D}(p)$ by taking the adjoint of (A11). To see this, we first write, using (A1),

$$\begin{aligned} &\langle p'n\sigma\,|\,f^c_{j_3}(p)\rangle \\ &= \langle\phi(\mathbf{p}=0)|\otimes\langle n\sigma\,|\,U(L(p'))\,|\,f^c_{j_3}(p)\rangle \\ &= \langle\phi(\mathbf{p}=0)|\otimes\langle n\sigma\,|\,U^I(p')\otimes U^X(p')\,|\,f^c_{j_3}\rangle\otimes|\,f(p)\rangle. \end{aligned} \tag{A13}$$

Letting U^X act on $\phi(\mathbf{p}=0)$ and U^I on $f^c_{j_3}$, we have

$$= \phi'\langle\phi(p')\,|\,f(p)\rangle\langle n\sigma\,|\,f^c_{j'_3}\rangle D^{1/2\,1/2(k_0c)}_{j'_3 j_3}(L(p')),$$

where ϕ' is a phase factor which may arise from different normalizations for ϕ and f. But $\langle\phi(p')\,|\,f(p)\rangle = \phi''\,2p_0\delta^3(\mathbf{p}'-\mathbf{p})$ and $\langle n\sigma\,|\,f^c_{j_3}\rangle = \mathring{D}^{nc}_{\sigma j_3}(0)$, so that (defining $\phi=\phi'\phi''$)

$$\begin{aligned} &\langle p'n\sigma\,|f^c_{j_3}(p)\rangle \\ &= \phi\mathring{D}^{nc}_{\sigma j'_3}(0)\,D^{1/2\,1/2(k_0c)}_{j'_3 j_3}(L(p'))\,2p_0\delta^3(\mathbf{p}'-\mathbf{p}). \end{aligned} \tag{A14}$$

Using this we may expand $|\,f^c_{j_3}(p)\rangle$ in terms of $|\,pn\sigma\rangle$. Carrying out the trivial p integration, we have

$$|\,f^c_{j_3}(p)\rangle = \phi\,|\,pn\sigma\rangle\,\mathring{D}^{nc}_{\sigma j'_3}(0)\,D^{1/2\,1/2(k_0c)\dagger}_{j'_3 j_3}(L(p')). \tag{A15}$$

Thus the scalar product of two spinor basis vectors is

$$\begin{aligned} \langle f^{c'}_{j'_3}(p')\,|\,f^c_{j_3}(p)\rangle &= |\,\phi\,|^2\,D^{1/2\,1/2(k_0c)\dagger}_{j'_3 j''_3}(L(p'))\,D^{cn'}_{j''_3\sigma'}(0) \\ &\times\langle p'n'\sigma'|\,pn\sigma\rangle\,\mathring{D}^{nc}_{\sigma j'_3}(0)\,D^{1/2\,1/2(k_0c)}_{j'_3 j_3}(L(p)). \end{aligned} \tag{A16}$$

Using (3) and the facts that $D(0)$ is unitary and $D^{1/2\,1/2(k_0c)}_{j'_3 j_3}(L)$ is self-adjoint,

$$= D^{1/2\,1/2(k_0c)}_{j'_3 j_3}(L(p)\,L(p))\,2p_0\delta^3(\mathbf{p}'-\mathbf{p})\,\delta^{c'c},$$

and, using an explicit form for the boost,[6,8]

$$= (\hat{p}_\mu \gamma_0 \gamma^\mu)^{c'c}_{j_3' j_3} 2p_0 \delta^3(\mathbf{p}' - \mathbf{p})$$

where the γ matrices are given by (A5). If we define new vectors $\dot{f}$ according to Eq. (14), then Eq. (A16) becomes the same as Eq. (15):

$$\langle \dot{f}^{c'}_{j_3'}(p') \,|\, f^{c}_{j_3}(p) \rangle = 2p_0 \delta^3(\mathbf{p}' - \mathbf{p})\, \delta^{c'c}\, \delta_{j_3' j_3}. \tag{15}$$

Using definition (14), the transformation equation (16a) is unchanged,

$$|\, pn\sigma \rangle = |\, f^{c}_{j_3}(p) \rangle D^{cn}_{j_3\sigma}(p), \tag{A16a}$$

but (16b) becomes

$$\langle pn\sigma \,| = \dot{D}^{nc''}_{\sigma j_3''}(p)(\hat{p}_\mu \gamma^\mu \gamma_0)^{c''c'}_{j_3'' j_3'} \langle f^{c'}_{j_3'}(p) |. \tag{A16b}$$

Now

$$\langle pn\sigma \,| = (|\, pn\sigma \rangle)^\dagger \quad \text{and} \quad \langle f^{c}_{j_3}(p) \,| = (|\, f^{c}_{j_3}(p) \rangle)^\dagger \tag{A17}$$

so from (A16b) we conclude that

$$D^{cn}_{j_3\sigma}(p)^\dagger = \dot{D}^{nc'}_{\sigma j_3'}(p)(\hat{p}_\mu \gamma^\mu \gamma_0)^{c'c}_{j_3' j_3} \tag{A18}$$

or

$$\dot{D}^{nc}_{\sigma j_3}(p) = D^{cn}_{j_3'\sigma}(p)^\dagger (\hat{p}_\mu \gamma_0 \gamma^\mu)^{c'c}_{j_3' j_3}. \tag{A19}$$

Using (A4), (A5), and (A11), we may write this as

$$\dot{D}^{nc}_{\sigma j_3}(p) = \underset{\text{CAN}}{(\gamma_0)^{nn'}_{\sigma\sigma'}}\, D^{c'n}_{j_3'\sigma'}(p)^\dagger \underset{\text{SPIN}}{(\gamma_0)^{c'c}_{j_3' j_3}}$$

or, using (A11),

$$= C\, \dot{D}^{nc'}_{\sigma j_3'}(0)(\delta^{c'c}\delta_{j_3' j_3} + (\gamma_0 \gamma^\mu \hat{p}_\mu)^{c'c}_{j_3' j_3}). \tag{A20}$$

It is easy to see that D and $\dot{D}$ satisfy the Dirac equation. For example,

$$\begin{aligned} P_\mu \Gamma^\mu \,|\, pn\sigma \rangle &= P_\mu \Gamma^\mu \,|\, f^{c}_{j_3}(p) \rangle D^{cn}_{j_3\sigma}(p) \\ &= (P^X_\mu \otimes 1)(1 \otimes \Gamma^I_\mu) \,|\, f(p) \otimes f^{c}_{j_3} \rangle D^{cn}_{j_3\sigma}(p) \\ &= \tfrac{1}{2} |\, f^{c'}_{j_3'}(p) \rangle (p_\mu \gamma^\mu)^{c'c}_{j_3' j_3} D^{cn}_{j_3\sigma}(p). \end{aligned}$$

But, since $P_\mu \Gamma^\mu$ is Lorentz invariant,

$$P_\mu \Gamma^\mu \,|\, pn\sigma \rangle = \tfrac{1}{2} m\pi \,|\, pn\sigma \rangle = \tfrac{1}{2} m\pi \,|\, f^{c}_{j_3}(p) \rangle D^{cn}_{j_3\sigma}(p) \tag{A21}$$

so that

$$(p_\mu \gamma^\mu)^{c'c}_{j_3' j_3} D^{cn}_{j_3\sigma}(p) = \pi m\, D^{c'n}_{j_3'\sigma}(p). \tag{A22a}$$

Similarly, we find

$$\dot{D}^{nc'}_{\sigma j_3'}(p)(p_\mu \gamma^\mu)^{c'c}_{j_3' j_3} = \pi m\, \dot{D}^{nc}_{\sigma j_3}(p). \tag{A22b}$$

From the above, in particular (A11), (A20), and (A22), it is apparent that the $D^{n=1/2,c}_{\sigma j_3}(p)$ are essentially the Dirac spinors $u^{c}_{j_3}(p,\sigma)$ and $D^{n=-1/2,c}_{\sigma j_3}(p)$ the $v^{c}_{j_3}(p,\sigma)$, while the $\dot{D}^{cn}_{j_3\sigma}(p)$ are the $\bar{u}$ and $-\bar{v}$. We will display this connection between the D and $\dot{D}$ and the explicit form of u and v in the usual representation.

The Dirac spinors are usually written[9]

$$u^{(\nu)}(p) = C(1 + \gamma_\mu \hat{p}^\mu) u^{(\nu)} \tag{A23}$$

and

$$v^{(\nu)}(p) = C(1 - \gamma_\mu \hat{p}^\mu) v^{(\nu)}, \tag{A24}$$

where C is given by (A12) and

$$u^{(1/2)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad u^{(-1/2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

$$v^{(1/2)} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad v^{(-1/2)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{A25}$$

We may arrange these u's and v's side by side in matrix form:

$$U(0) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A26}$$

so that (A23) and (A24) may be written together as

$$U(p) = C(1 + \gamma_\mu \gamma_0 \hat{p}^\mu)\, U(0) \tag{A27}$$

and similarly the Dirac adjoint is

$$\bar{U}(p) = U^\dagger(p)\, \gamma_0 = C U^\dagger(0)(1 + \hat{p}^\mu \gamma_0 \gamma_\mu)\, \gamma_0. \tag{A28}$$

The γ_0 is inserted in (A27) because of the sign difference between (A23) and (A24).

The γ matrices in this representation are

$$\gamma_k = \begin{pmatrix} 0 & -\sigma_k \\ \sigma_k & 0 \end{pmatrix}, \quad \gamma_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A29}$$

They are connected to our γ's in the spinor basis by

$$U^\dagger (\gamma_\mu)_{\text{usual}}\, U = (\gamma_\mu)_{\text{ours}}, \tag{A30}$$

where

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \tag{A31}$$

The connection between the usual spinors (A23), (A24) and our D, $\dot{D}$ matrices (A11), (A20) is

$$D(p) = U^\dagger U(p)\, UD(0) \tag{A32}$$

and

$$\begin{aligned} \dot{D}(p) &= \gamma_0 D^\dagger(p)\, \gamma_0 \\ &= \gamma_0 D^\dagger(0)\, U^\dagger \bar{U}(p)\, U \end{aligned} \tag{A33}$$

since $D^\dagger(p)\gamma_0$ corresponds to the usual Dirac adjoint spinor. We will verify (A32) by an explicit calculation. From (A11) and (A7) we have

$$D(p) = \frac{C}{m} \frac{1-i}{2} \times \begin{pmatrix} m + p_0 + \mathbf{p}\cdot\sigma & i(m + p_0 + \mathbf{p}\cdot\sigma) \\ m + p_0 - \mathbf{p}\cdot\sigma & -i(m + p_0 - \mathbf{p}\cdot\sigma) \end{pmatrix}. \tag{A34}$$

On the other hand,

$$\begin{aligned} &U^\dagger U(p)\, UD(0) \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \frac{C}{m} \begin{pmatrix} m + p_0 & -\mathbf{p}\cdot\sigma \\ -\mathbf{p}\cdot\sigma & m + p_0 \end{pmatrix} \\ &\times \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \frac{1-i}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \end{aligned}$$

$$= \frac{C}{m}\,\frac{1-i}{2} \times \begin{pmatrix} m+p_0+\mathbf{p}\cdot\boldsymbol{\sigma} & i(m+p_0+\mathbf{p}\cdot\boldsymbol{\sigma}) \\ m+p_0-\mathbf{p}\cdot\boldsymbol{\sigma} & -i(m+p_0-\mathbf{p}\cdot\boldsymbol{\sigma}) \end{pmatrix}, \tag{A35}$$

which is the same as (A34). Equation (A33) may be verified in the same manner.

To illustrate Eq. (19), we will consider the two simple examples $J=P_\mu$ and $J=\Gamma_\mu$. For the case $J=P_\mu$ we note that the momentum operator does not affect the internal quantum numbers, so in the spinor basis

$$P_\mu = I \otimes P_\mu^{\mathrm{EXT}} \tag{A36}$$

and (19) becomes

$$\langle p'n'\sigma' | P_\mu | pn\sigma\rangle = \mathring{D}^{n'c''}_{\sigma' j_3''}(p')\langle \mathring{f}^{c''}_{j_3''} | f^{c'}_{j_3'}\rangle \times \langle \mathring{f}(p') | P_\mu^{\mathrm{EXT}} | f(p)\rangle D^{c'n}_{j_3'\sigma}(p) \tag{A37}$$

but

$$\langle \mathring{f}(p') | P_\mu^{\mathrm{EXT}} | f(p)\rangle = p_\mu \langle \mathring{f}(p') | f(p)\rangle = p_\mu 2E(p)\,\delta^3(\mathbf{p}'-\mathbf{p}) \tag{A38}$$

and[6]

$$\langle \mathring{f}^{c''}_{j_3''} | f^{c'}_{j_3'}\rangle = \delta^{c''c'}\,\delta_{j_3''j_3'} \tag{A39}$$

so that

$$\langle p'n'\sigma' | P_\mu | pn\sigma\rangle = 2p_\mu E(p)\,\delta^3(\mathbf{p}'-\mathbf{p})\,\mathring{D}^{n'c}_{\sigma' j_3}(p)\,D^{cn}_{j_3\sigma}(p) = 2p_\mu E(p)\,\delta^3(\mathbf{p}'-\mathbf{p})\,\delta^{nn'}\delta_{\sigma\sigma'}, \tag{A40}$$

which is just what we would expect.

The example $J=\Gamma_\mu$ illustrates the differences between calculating in the two basis systems. Then (19) becomes (dropping the discrete indices),

$$\langle p'n'\sigma' | 2\Gamma_\mu | pn\sigma\rangle = \mathring{D}(p')\langle \mathring{f} | 2\Gamma_\mu^{\mathrm{INT}} | f\rangle\langle \mathring{f}(p') | f(p)\rangle D(p), \tag{A41}$$

where we used the fact that Γ_μ does not affect the momenta. The matrix elements of $2\Gamma_\mu^{\mathrm{INT}}$ have been calculated,[6] and are simply the γ matrices in our spinor basis:

$$\langle p'n'\sigma' | 2\Gamma_\mu | pn\sigma\rangle = \mathring{D}(p')\gamma_\mu 2E(p)\,\delta^3(\mathbf{p}'-\mathbf{p})\,D(p). \tag{A42}$$

Using (A11) and (A4), we have (for $p'=p$)

$$\mathring{D}(p)\gamma_0 D(p) = \begin{array}{c} \begin{matrix} n=1/2 & n=-1/2 \end{matrix} \\ \begin{bmatrix} \frac{p_0}{m} & i\frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m} \\ i\frac{\mathbf{p}\cdot\boldsymbol{\sigma}}{m} & -\frac{p_0}{m} \end{bmatrix} \end{array} \begin{array}{l} c=3/2 \\ c=-3/2 \end{array} \tag{A43}$$

$$\mathring{D}(p)\gamma_i D(p) = \begin{array}{c} \begin{matrix} n=1/2 & n=-1/2 \end{matrix} \\ \begin{bmatrix} \frac{p_i}{m} & i\left(\sigma_i + \frac{p_i\mathbf{p}\cdot\boldsymbol{\sigma}}{m(m+p_0)}\right) \\ i\left(\sigma_i + \frac{p_i\mathbf{p}\cdot\boldsymbol{\sigma}}{m(m+p_0)}\right) & -\frac{p_i}{m} \end{bmatrix} \end{array} \begin{array}{l} c=3/2 \\ c=-3/2 \end{array}. \tag{A44}$$

The calculation is made in the canonical basis as follows:

$$\begin{aligned} \langle p'n'\sigma' | 2\Gamma_\mu | pn\sigma\rangle &= \langle 0n'\sigma' | U(L(p'))\,2\Gamma_\mu U^{-1}(L(p'))\,U(L(p')) | pn\sigma\rangle \\ &= L_\mu^{-1\nu}(p')\langle 0n'\sigma' | 2\Gamma_\nu U(L(p')) | pn\sigma\rangle = L_\mu^{-1\nu}(p')\sum_{\gamma,s}\int d\mu(k)\langle 0n'\sigma' | 2\Gamma_\nu | k\gamma s\rangle\langle k\gamma s | U(L(p')) | pn\sigma\rangle \\ &= L_\mu^{-1\nu}(p')\langle 0n'\sigma' | 2\Gamma_\mu | 0n\sigma\rangle\langle 0n\sigma | U(L(p')) | pn\sigma\rangle = L_\mu^{-1\nu}(p')\langle 0n'\sigma' | 2\Gamma_\mu | 0n\sigma\rangle\,2E(p)\,\delta^3(\mathbf{p}'-\mathbf{p}), \end{aligned} \tag{A45}$$

where $L^{-1}(p)$ is the boost, which is given by[9]

$$L_\mu^{-1\nu}(p) = \begin{array}{c} \begin{matrix} \nu=0 & \nu=n \end{matrix} \\ \begin{bmatrix} \frac{p_0}{m} & -\frac{p^n}{m} \\ \frac{p_m}{m} & g_m^n - \frac{p_m p^n}{m(m+p_0)} \end{bmatrix} \end{array} \begin{array}{l} \mu=0 \\ \mu=m \end{array}. \tag{A46}$$

The matrix elements $\langle 0n'\sigma' | 2\Gamma_\mu | 0n\sigma\rangle$ are given by Eq. (A4), and so we may calculate

$$L_\mu^{-1\nu}(p')\langle 0n'\sigma' | 2\Gamma_\mu | 0n\sigma\rangle$$

by simply multiplying the matrices in (A4) with (A46). The result agrees with (A43) and (A44) so the calculations in the two different basis systems agree.

It is useful to have explicit expressions for the bilinears $\mathring{D}\gamma_\mu D$, etc. These are listed below for $n=n'$, as these are the only ones used in the text. We use the conventions

$$\sigma_{\mu\nu} = \tfrac{1}{2}i[\gamma_\mu, \gamma_\nu] \tag{A47}$$

and

$$\gamma_5 = -(i/4!)\,\epsilon^{\mu\nu\rho\sigma}\gamma_\mu\gamma_\nu\gamma_\rho\gamma_\sigma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{A48}$$

Greek indices run 0 to 3, Latin indices 1 to 3. The unit and Pauli matrices are abbreviated $\delta = \delta_{\sigma'\sigma}$ and $\sigma_i = (\sigma_i)_{\sigma'\sigma}$. Also, $\pi = \text{sign } n$ and

$$N = [4m'm(m+p_0)(m'+p_0)]^{-1/2},$$

$$\hat{D}(p')\,D(p) = N\{[(m'+p_0')(m+p_0) - \mathbf{p}\cdot\mathbf{p}']\,\delta + i(\mathbf{p}'\times\mathbf{p})\cdot\sigma\}, \tag{A49}$$

$$\hat{D}(p')\,\gamma_0 D(p) = \pi N\{[(m'+p_0')(m+p_0) + \mathbf{p}'\cdot\mathbf{p}]\,\delta + i(\mathbf{p}'\times\mathbf{p})\cdot\sigma\}, \tag{A50}$$

$$\hat{D}(p')\,\gamma_m D(p) = \pi N\{[(m'+p_0')\,p_m + (m+p_0)\,p_m']\,\delta + i\epsilon_{mli}\sigma_l[(m+p_0)\,p_i' - (m'+p_0')\,p_i]\}, \tag{A51}$$

$$\hat{D}(p')\,\sigma_{0j} D(p) = iN\{[(m'+p_0')\,p_j - (m+p_0)\,p_j']\,\delta - i\epsilon_{jli}\sigma_l[(m+p_0)\,p_i' + (m'+p_0')\,p_i]\}, \tag{A52}$$

$$\hat{D}(p')\,\sigma_{ij} D(p) = N\epsilon_{ijk}\{(m'+p_0')(m+p_0)\,\sigma_k + \mathbf{p}'\cdot\mathbf{p}\sigma_k - p_k\mathbf{p}'\cdot\sigma - p_k'\mathbf{p}\cdot\sigma - i\epsilon_{kml}p_l'p_m\delta\}, \tag{A53}$$

$$\hat{D}(p')\,\gamma_5 D(p) = -N\{(m+p_0)\,\mathbf{p}'\cdot\sigma - (m'+p_0')\,\mathbf{p}\cdot\sigma\}, \tag{A54}$$

$$\hat{D}(p')\,\gamma_0\gamma_5 D(p) = \pi N\{(m'+p_0')\,\mathbf{p}\cdot\sigma + (m+p_0)\,\mathbf{p}'\cdot\sigma\}, \tag{A55}$$

$$\hat{D}(p')\,\gamma_m\gamma_5 D(p) = \pi N\{[(m'+p_0')(m+p_0) - \mathbf{p}'\cdot\mathbf{p}]\,\sigma_m + p_m\,\mathbf{p}'\cdot\sigma + p_m'\,\mathbf{p}\cdot\sigma + i\epsilon_{mkl}p_kp_l'\delta\}, \tag{A56}$$

$$\hat{D}(p')\,\sigma_{0j}\gamma_5 D(p) = iN\{[(m'+p_0')(m+p_0) + \mathbf{p}'\cdot\mathbf{p}]\,\sigma_j - p_j\,\mathbf{p}'\cdot\sigma - p_j'\,\mathbf{p}\cdot\sigma - i\epsilon_{jkm}p_kp_m'\delta\}, \tag{A57}$$

$$\hat{D}(p')\,\sigma_{ij}\gamma_5 D(p) = N\{\epsilon_{ijk}[(m'+p_0')\,p_k - (m+p_0)\,p_k']\,\delta + i[\sigma_j((m'+p_0')\,p_i + (m+p_0)\,p_i') - \sigma_i((m'+p_0')\,p_j + (m+p_0)\,p_j')]\}. \tag{A58}$$

Finally, we give the useful projection operators:

$$D^{c\,1/2}_{0j_3}(p)\,\hat{D}^{1/2\,c'}_{j_3\sigma'}(p) = (\gamma_\mu\,\hat{p}^\mu + 1)^{cc'}_{\sigma\sigma'}, \tag{A59}$$

$$D^{c\,-1/2}_{0j_3}(p)\,\hat{D}^{-1/2\,c'}_{j_3\sigma'}(p) = -(\gamma_\mu\,\hat{p}^\mu - 1)^{cc'}_{\sigma\sigma'}. \tag{A60}$$

*Work supported by the U.S. Atomic Energy Commission Contract No. AT(40-1)3992.

[1]E.P. Wigner, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964), p. 37.

[2]P. Budini and C. Fronsdal, Phys. Rev. Lett. 14, 968 (1965).

[3]The subscripts X_i on the symbol for the group G_{X_i} indicate that X_i are the generators of G. This notation is necessary to allow us to distinguish between mathematically isomorphic groups, which have different physical observables.

[4]A. Bohm, Phys. Rev. D 3, 367 (1971).

[5]A. Bohm, Phys. Rev. D 3, 377 (1971).

[6]A. Bohm and G.B. Mainland, Fortschr. Phys. 18, 285 (1970). In the present paper, we use the convention that S_{0i} rather than S_{i0} has the same action as Γ_i. This leads to the spinor basis matrix elements of Γ_i having a different sign in the present paper than in the reference.

[7]We shall confine our discussions here to the Dirac representation, although we believe that other representations of γ are more likely candidates for the description of hadron systems.

[8]H. Joos, Fortschr. Phys. 10, 65 (1962), Sec. 3.2.

[9]S. Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1967).

[10]M.A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, Oxford, 1964). We will be concerned here with the representations $(k_0 = 1/2,\ c = 3/2)$ and $(k_0 = 1/2,\ c = -3/2)$, which are denoted in Ref. 9 by $D(1/2, 0)$ and $D(0, 1/2)$, respectively.

[11]See the Appendix.

[12]$f_1 = F_2 + \tfrac{1}{2}(m'+m)(mF_5 + m'F_6) + ip_\mu' p^\mu F_9$, $f_2 = F_1^- + iF_7^+ + 2ip_\mu' p^\mu F_8^-$, $f_3 = -iF_1^+ + iF_7^- + 2p_\mu' p^\mu F_8^+$, $f_4 = mF_1^+ + imF_7^+ + im(2p_\mu' p^\mu - m'^2)F_8^+$, $f_5 = m'F_1^+ - im'F_7^+ + im'(2p_\mu' p^\mu - m^2)F_8^+$, $f_6 = \tfrac{1}{2}(m'F_6 - mF_5)$, $f_7 = -\tfrac{1}{2}(mF_5 + m'F_6)$, $f_8 = \tfrac{1}{2}(m'+m)(m'F_3 + mF_4) + [m'm - i/2(m'+m)^2]F_9$, $f_9 = -2im'mF_8^-$, $f_{10} = -m'mF_8^+$, $f_{11} = \tfrac{1}{2}(mF_4 - m'F_3 + i(m'-m)F_9)$, $f_{12} = -\tfrac{1}{2}(m'F_3 + mF_4 - i(m'+m)F_9)$.

[13]A leptonic reduced matrix element might have $n' = -n$ (e.g., for an electron and antineutrino). However, the form factors for such matrix elements have been well established. For the hadronic matrix elements we consider here, we always have $n' = n$.

[14]$\tilde{g}_1 = G_4 + (p_\mu' p^\mu - 1/2(m'-m)^2)G_1$, $\tilde{g}_2 = 2G_2^- + G_3^+$, $\tilde{g}_3 = G_3^- + 2G_2^+$, $\tilde{\tilde{g}}_1 = -2mG_2^+ - mG_3^+ - 2mG_2^+ + m'G_3^+$, $\tilde{\tilde{g}}_2 = -(i/2)\,((m'-m)G_1 + mG_5 + mG_6 - m'G_7 - mG_8)$, $\tilde{\tilde{g}}_3 = \tfrac{1}{2}((m'+m)G_1 + mG_5 - mG_6 - m'G_7 + mG_3)$.

[15]C.W. Kim and H. Primakoff, Phys. Rev. 180, 1502 (1969).

[16]Haim Goldberg, Nuovo Cimento A 60, 509 (1969).

[17]We use the Cartan Weyl basis for the SU(3) Lie algebra in which the commutation relations are given by

$$[H_i,\ H_j] = 0, \qquad [H_j, E_\alpha] = r_j(\alpha)E_\alpha,$$

$$[E_\alpha,\ E_\beta] = N_{\alpha\beta}E_\gamma, \quad [E_\alpha,\ E_{-\alpha}] = r^i(\alpha)H_i.$$

The root vectors are (in our normalization)

$$r_i(1) = (1/\sqrt{3})(1, 0),\quad r_i(2) = (1/2\sqrt{3})(1, \sqrt{3}),\quad r_i(-\alpha) = -r_i(\alpha),$$
$$r_i(3) = (1/2\sqrt{3})(-1, \sqrt{3}),$$

and $N_{\alpha\beta} = \pm\sqrt{1/6}$ if $r(\alpha) + r(\beta) = r(\gamma)$ is also a nonvanishing root vector, $N_{\alpha\beta} = 0$ otherwise. In particular,

$$N_{1,3} = -N_{3,1} = N_{-3,-1} = -N_{-1,-3} = N_{3,-2}$$
$$= -N_{-2,3} = N_{-2,1} = -N_{1,-2} = N_{2,-3}$$
$$= -N_{-3,2} = N_{-1,2} = -N_{2,-1} = \sqrt{1/6}.$$

In the normalization we have used here, the hypercharge is

$$Y = 2H_2$$

and the isospin is

$$I_3 = \sqrt{3}\,H_1, \quad (I_1 \pm iI_2) = (\sqrt{6})E_{\pm 1}.$$

We also call $H_1 = E_0$, $H_2 = E_8 = E_{-8}$. The component of the octet operator with the same transformation property as E_α is denoted by V_μ^α or A_μ^α. Our notation differs from the one conventionally used in physics literature. The connection is

$$V_\mu^{\pm 1} \text{ corresponds to } \mathcal{J}_{1\mu} \pm i\mathcal{J}_{2\mu},$$

$A_\mu^{\pm 1}$ corresponds to $\mathcal{J}_{1\mu}^5 \pm i\,\mathcal{J}_{2\mu}^5$,

$V_\mu^{\pm 2}$ corresponds to $\mathcal{J}_{4\mu} \pm i\,\mathcal{J}_{5\mu}$,

$A_\mu^{\pm 2}$ corresponds to $\mathcal{J}_{4\mu}^5 \pm i\,\mathcal{J}_{5\mu}^5$, etc.

[18]First suggested by J. Werle in 1965, this approximation has been used in many calculations, e.g., A. Bohm, Phys. Rev. D 7, 2101 (1973) and references therein.

[19]We use the Clebsch—Gordon coefficients of J. G. Kuriyan, D. Lurie; and A. J. Macfarlane, J. Math. Phys. 6, 722 (1965). See also Ref. 15.

[20]If one defines two operators $V_{f,d}^\beta$ in the octet representation space of $SU(3)_E$ by $\langle \alpha' \mid V_{(f)}^\beta \mid \alpha \rangle = C(11, 11, 11, \gamma=1; \alpha\beta\alpha')$, $\langle \alpha' \mid V_{(d)}^\beta \mid \alpha \rangle = C(11, 11, 11, \gamma=2; \alpha\beta\alpha')$, then V_μ^β may be written as $V_\mu^\beta = V_{(f)}^\beta \otimes V_\mu^{(f)} + V_{(d)}^\beta \otimes V_\mu^{(d)}$.

[21]M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964).

[22]Note that P_μ is not the total momentum. P_μ is the momentum in the hadron subspace only.

[23]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964), Appendix B.

[24]This identification is made at this point in order to simplify subsequent formulas. It is justified by the fact that the following manipulations will not involve any parts of (114) or (115) except those in (116).

[25]We use the normalization $\pi u = 2m$ because of the measure (101). See, e.g., Ref. 23, Appendix B.

Subalgebras of real three- and four-dimensional Lie algebras

J. Patera and P. Winternitz

Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada
(Received 28 December 1976)

The Lie subalgebras of all real Lie algebras of dimension $d \leq 4$ are classified into equivalence classes under their groups of inner automorphisms. Tables of representatives of each conjugacy class are given.

1. INTRODUCTION

The purpose of this article is to find all Lie subalgebras of all real Lie algebras of dimension $d \leq 4$. More precisely, we shall classify the subalgebras of each such Lie algebra into conjugacy classes and present a representative of each class. Conjugacy in each case is considered under the group of inner automorphisms, i.e., the Lie group obtained by the exponentiation of the adjoint representation of the considered Lie algebra.

The main reason for our interest in this problem is in connection with physical applications. We are presently engaged in a project which involves the classification of subgroups of the "fundamental" groups of physics, namely the Poincaré group (inhomogeneous Lorentz group), the similitude group (Poincaré group extended by dilations), the de Sitter groups O(4, 1) and O(3, 2), the conformal group O(4, 2), and others (see Refs. 1–3 and forthcoming papers). The knowledge of all subalgebras of low dimensional real Lie algebras (in abstract form) is of great help not only in the above program, but is also of independent interest.

As further motivation let us mention several of the many applications of the subgroup structure of a Lie group (or subalgebra structure of a Lie algebra).

1. Consider a physical system (or any other system) described by, e.g., a set of differential equations. Let this system have a symmetry described by a Lie group G. A classification of subgroups of G provides a classification of possible symmetry breaking influences (like additional terms in the equations, boundary conditions, etc).[4]

2. A knowledge of the subgroups of G is important in the representation theory of G, e.g., it allows for the induction of representations from different subgroups, helps in the classification of irreducible representations, etc.[5]

3. A classification of subgroups is related to a classification of possible bases for representations, in that different bases may correspond to the reduction of the representations of G to different subgroups.

4. The invariant operators of different subgroups of G (if they exist)[3,6] will provide different sets of quantum numbers for a quantum mechanical system described by G (or different integrals of motion for the corresponding classical system).

5. A knowledge of the subgroup structure of G is needed if we are interested in considering all possible contractions of G to other groups.[7]

Complex Lie algebras of dimension $d \leq 4$ have been classified by Lie,[8] real ones with $d \leq 5$ by Mubarakzyanov.[9] We make use of this last classification that was reproduced in a modified form in an earlier article.[3]

In Sec. 2 we discuss very briefly the method used to classify subalgebras and then present the results in tables. Section 3 contains some conclusions and comments.

2. CLASSIFICATION OF SUBALGEBRAS

All three-dimensional indecomposable real Lie algebras can be classified into nine types.[3] We denote them $A_{3,i}$ (with $i = 1, \ldots, 9$), as in a previous paper[3] and present their commutation relations in convenient bases $\{e_1, e_2, e_3\}$ in Table I. Two of these algebras are simple, namely the algebras $A_{3,8}$ and $A_{3,9}$ of the groups O(2, 1) and O(3). The rest are solvable ($A_{3,1}$ is actually nilpotent) and can all be written as semidirect sums of a one-dimensional subalgebra $\{e_3\}$ and an Abelian ideal $\{e_1, e_2\}$.

All four-dimensional indecomposable real Lie algebras can be classified into 12 types[3] $A_{4,i}$ (with $i = 1, \ldots, 12$). We present their commutation relations in a convenient basis $\{e_1, e_2, e_3, e_4\}$ in Table II. They are all solvable ($A_{4,1}$ is nilpotent) and can all be written as semidirect sums of a one-dimensional Lie algebra $\{e_4\}$ and a three-dimensional ideal $N = \{e_1, e_2, e_3\}$. For $A_{4,1}, \ldots, A_{4,6}$ N is Abelian, for $A_{4,7}, \ldots, A_{4,11}$ it is of type $A_{3,1}$ (nilpotent), and for $A_{4,12}$ it is of the type $A_{3,3}$.

Two types of decomposable three-dimensional Lie algebras exist: an Abelian one, $3A_1$ and non-Abelian one, $A_2 \oplus A_1$. Twelve classes of decomposable four-dimensional Lie algebras exist, namely $4A_1, 2A_2, A_2 \oplus 2A_1$, and $A_1 \oplus A_{3,i}$ ($i = 1, \ldots, 9$), where A_2 is the two-dimensional solvable Lie algebra satisfying $[e_1, e_2] = e_2$.

The subalgebras of the simple Lie algebras are known, those of the decomposable algebras can be found using the Goursat twist method,[10] explained in an earlier paper.[1] The subalgebras of the indecomposable and nonsimple Lie algebras $A_{3,i}$ ($i = 1, \ldots, 7$) and $A_{4,i}$ ($i = 1, \ldots, 12$) can be classified into conjugacy classes under the groups $\mathcal{G}A_{3,i}$ or $\mathcal{G}A_{4,i}$ generated by these algebras, using a previously developed classification method.[1] Let us briefly present the classification algorithm for the case of relevance here, namely of an n-dimensional Lie algebra L_n that can be written as a semidirect product

$$L_n = \{e_n\} \square L_{n-1}, \tag{1}$$

where e_n is a certain element of L_n and L_{n-1} is an $(n-1)$-dimensional ideal of L_n.

Step 1. Find all subalgebras of the factor algebra $L_n/L_{n-1} = \{e_n\}$. This step is trivial, since there are only two subalgebras: $\{0\}$ and $\{e_n\}$.

Step 2. Find all subalgebras of L_{n-1} and classify them into orbits under the action of the group $\mathcal{G}L_n$. This can be done either by making use of some low dimensional faithful representation of $\mathcal{G}L_n$, or more abstractly, by making use of the Baker–Campbell–Hausdorf formula,[11]

$$e^{\alpha X} Y e^{-\alpha X} = Y + \alpha[X, Y] + \frac{\alpha^2}{2!}[X,[X, Y]] + \frac{\alpha^3}{3!}[X,[X,[X, Y]]] + \cdots, \tag{2}$$

where X and Y are elements of L_n. By induction we can assume that the subalgebras of L_{n-1} are already classified under $\mathcal{G}L_{n-1}$. We must hence only eliminate subalgebras from the $\mathcal{G}L_{n-1}$ list that are conjugate under the one-parameter group $\exp\alpha e_n$ to other ones in the list (since we have $\mathcal{G}L_n = \exp \mathrm{R} e_n \square \mathcal{G}L_{n-1}$).

Step 3. Find all splitting extensions of the algebra $\{e_n\}$. To do this we must simply find all subalgebras N_a of L_{n-1} that are invariant under e_n,

$$[e_n, N_a] \subseteq N_a, \quad N_a \subseteq L_{n-1}. \tag{3}$$

The invariant subalgebras N_a should then be classified into orbits under $\mathrm{Nor}_{\mathcal{G}L_n} e_n$, i.e., the normalizer of $\{e_n\}$ in $\mathcal{G}L_n$. A representative of each orbit, together with the element e_n then provides the required sublist.

Step 4. Find all subalgebras of L_n not contained in L_{n-1} and not containing any $\mathcal{G}L_n$ conjugate of e_n. They will be of the form

$$\left\{e_n + \sum_i x_i e_i, N_a\right\}, \tag{4}$$

where N_a is a subalgebra of L_{n-1} with an L_n normalizer not contained in L_{n-1} and x_i are real numbers, not all equal to zero and such that $\tilde{e}_n = e_n + \sum_i x_i e_i$ is not conjugate to e_n under $\mathcal{G}L_n$. The algebras (4) should be classified into orbits under $\mathrm{Nor}_{\mathcal{G}L_n} \tilde{e}_n$.

As an example of this method consider the algebra $A_{4,8}$ of Ref. 3 (and Table II). The nonzero commutation relations are

$$[e_2, e_3] = e_1, \quad [e_2, e_4] = e_2, \quad [e_3, e_4] = -e_3. \tag{5}$$

Step 1. The algebra $\{e_1, e_2, e_3\}$ is a nilpotent ideal of the type $A_{3,1}$. The factor algebra $\{e_4\}$ has two subalgebras: $\{0\}$ and $\{e_4\}$.

Step 2. Classify subalgebras of $A_{3,1}$ under $\mathcal{G}A_{4,8}$. Putting

$$A = x_1 e_1 + x_2 e_2 + x_3 e_3 \tag{6}$$

and applying (2) we have

$$\exp(\alpha e_2 + \beta e_3) A \exp(-\alpha e_2 - \beta e_3) = (x_1 + \alpha x_3 - \beta x_2) e_1 + x_2 e_2 + x_3 e_3 \tag{7}$$

and

$$\exp(\gamma e_4) A \exp(-\gamma e_4) = x_1 e_1 + x_2 e^{-\gamma} e_2 + x_3 e^{\gamma} e_3. \tag{8}$$

Choosing α, β, and γ appropriately we obtain the one-dimensional subalgebras

$$\{e_1\}, \{e_2\}, \{e_3\}, \text{ and } \{e_2 + \epsilon e_3\}, \quad \epsilon = \pm 1. \tag{9}$$

All two-dimensional subalgebras of $A_{3,1}$ are Abelian. We let B run through all subalgebras in (9), leave A in its general form, and require $[A, B] = 0$. This imposes $B = e_1$ (or $A = e_1$). Since e_1 commutes with e_2, e_3, and e_4 we can use (7) and (8) to similify A. We obtain the subalgebras

$$\{e_1, e_2\}, \{e_1, e_3\}, \text{ and } \{e_1, e_2 + \epsilon e_3\}, \quad \epsilon = \pm 1. \tag{10}$$

The only three-dimensional subalgebra is $\{e_1, e_2, e_3\}$ itself.

Step 3. The equation

$$[e_4, A] = \lambda A \tag{11}$$

has the solutions $A = e_1$, $A = e_2$, and $A = e_3$. It is easy to see that the only two-dimensional invariant subalgebras are $\{e_1, e_2\}$ and $\{e_1, e_3\}$. Thus all subalgebras of $A_{4,8}$ containing e_4 are, up to $\mathcal{G}A_{4,8}$ conjugacy,

$$\{e_4\}, \{e_4, e_1\}, \{e_4, e_2\}, \{e_4, e_3\},$$
$$\{e_4, e_1, e_2\}, \{e_4, e_1, e_3\}, \text{ and } \{e_4, e_1, e_2, e_3\}.$$

Step 4. Putting $\tilde{e}_4 = e_4 + x e_1 + y e_2 + z e_3$ and using

$$\exp(\alpha e_2 + \beta e_3) \tilde{e}_4 \exp(-\alpha e_2 - \beta e_3) = e_4 + (x + \alpha z - \beta y) e_1 + (y + \alpha) e_2 + (z - \beta) e_3$$

with $\alpha = -y$, $\beta = z$, we find the subalgebras

$$\{e_4 + x e_1\}, \{e_4 + x e_1, e_2\}, \text{ and } \{e_4 + x e_1, e_3\}, \quad x \neq 0, \tag{12}$$

not contained in $A_{3,1}$ and not containing $\mathcal{G}A_{4,8}$ conjugates of e_4. Together (9)–(12) provide a list of representatives of all $\mathcal{G}A_{4,8}$ conjugacy classes of subalgebras of $A_{4,8}$.

We proceed quite analogously in all cases and only present the results in tables. The decomposable Lie algebras under consideration are Abelian ($3A_1$ and $4A_1$), of the type $A_2 \oplus A_2$ or of the type $A_1 \oplus B$, i.e., direct sums of a one-dimensional algebra with a two- or three-dimensional one. A classification of the subalgebras of an n-dimensional Abelian Lie algebra reduces to a simple problem in analytical geometry, namely a characterization or parametrization of all m-dimensional subspaces of an n-dimensional Euclidean space for $1 \leq m \leq n-1$. This problem has been solved in at least two manners, namely the Hermitian row reduced normal form[12] and the Grassmann coordination.[13] The algebra $A_2 \oplus A_2$ is treated by a straightforward application of the Goursat twist method. This method can be further simplified for algebras of the type $A_1 \oplus B$. Indeed, denote the basis element of A_1 as e_n, those of B as $\{e_1, \ldots, e_{n-1}\}$. Assume that all subalgebras of B are known and denoted by $B_{j,k}$. A list of representatives of subalgebra classes of $A_1 \oplus B$ will consist of three sublists: 1. all subalgebras $B_{j,k}$ (including the trivial

TABLE I. Proper subalgebras of real Lie algebras of dimension 2 and 3. The range of parameters: $0 \leq \phi < \pi$, $-\infty < x, y < \infty$, $\epsilon = \pm 1$.

Algebra	Nonzero commutation relations	Subalgebras		
		A_2	$2A_1$	A_1
$2A_1$				$(e_1 \cos\phi + e_2 \sin\phi)$
A_2	$[e_1, e_2] = e_2$			(e_1), (e_2)
$3A_1$			$(e_1 + xe_3, e_2 + ye_3)$, $(e_1 + xe_2, e_3)$, (e_2, e_3)	$\{e_1 + xe_2 + ye_3\}$, $\{e_2 + xe_3\}$, $\{e_3\}$
$A_1 \oplus A_2$	$[e_1, e_2] = e_2$	$(e_1 + xe_3; e_2)$	(e_1, e_3), (e_2, e_3)	$\{e_1 \cos\phi + e_3 \sin\phi\}$, $\{e_2 + \epsilon e_3\}$, $\{e_2\}$
$A_{3,1}$ (Weyl)	$[e_2, e_3] = e_1$		$(e_1, e_2 \cos\phi + e_3 \sin\phi)$	$\{e_1\}$, $\{e_2 \cos\phi + e_3 \sin\phi\}$
$A_{3,2}$	$[e_1, e_3] = e_1$, $[e_2, e_3] = e_1 + e_2$	$(e_3; e_1)$	(e_1, e_2)	$\{e_1\}$, $\{e_2\}$, $\{e_3\}$
$A_{3,3}$ $(D \square T_2)$	$[e_1, e_3] = e_1$, $[e_2, e_3] = e_2$	$(e_3; e_1 \cos\phi + e_2 \sin\phi)$	(e_1, e_2)	$\{e_1 \cos\phi + e_2 \sin\phi\}$, $\{e_3\}$
$A_{3,4}$ $(\mathrm{E}(1,1))$	$[e_1, e_3] = e_1$, $[e_2, e_3] = -e_2$	$(e_3; e_1)$, $(e_3; e_2)$	(e_1, e_2)	$\{e_1\}$, $\{e_2\}$, $\{e_3\}$, $\{e_1 + \epsilon e_2\}$
$A^a_{3,5}$ $(0 < \lvert a \rvert < 1)$	$[e_1, e_3] = e_1$, $[e_2, e_3] = ae_2$	$(e_3; e_1)$, $(e_3; e_2)$	(e_1, e_2)	$\{e_1\}$, $\{e_2\}$, $\{e_3\}$, $\{e_1 + \epsilon e_2\}$
$A_{3,6}$ $(\mathrm{E}(2))$	$[e_1, e_3] = -e_2$, $[e_2, e_3] = e_1$		(e_1, e_2)	$\{e_1\}$, (e_3)
$A^a_{3,7}$ $(a > 0)$	$[e_1, e_3] = ae_1 - e_2$, $[e_2 e_3] = e_1 + ae_2$		(e_1, e_2)	$\{e_1\}$, (e_3)
$A_{3,8}$ $(\mathrm{SU}(1,1))$	$[e_1, e_2] = e_1$, $[e_2, e_3] = e_3$, $[e_3, e_1] = 2e_2$	$(e_2; e_1)$		$\{e_1\}$, $\{e_2\}$, $(e_1 + e_3)$
$A_{3,9}$ $(\mathrm{SU}(2))$	$[e_1, e_2] = e_3$, $[e_3, e_1] = e_2$, $[e_2, e_3] = e_1$			(e_1)

ones), 2. the direct sums $e_n \oplus B_{j,k}$ for all $B_{j,k}$, 3. "twisted" subalgebras obtained from $B_{j,k}$ by adding an arbitrary multiple of e_n to each element of $B_{j,k}$ that is not contained in the derived algebra $B'_{j,k}$. The ranges of the coefficients multiplying e_n must then be determined using the normalizer $\mathrm{Nor}_{\mathcal{G}B} B_{j,k}$.

As an example consider the sum $A_2 \oplus A_1$ and choose the generators $\{e_3\}$ for A_1 and e_1, e_2, satisfying $[e_1, e_2] = e_2$ for A_2.

1. The subalgebras of A_2 are: $\{e_1, e_2\}, \{e_2\}, \{e_1\}, \{0\}$.
2. To these we add: $\{e_2, e_3, e_1\}, \{e_2, e_3\}, \{e_1, e_3\}, \{e_3\}$.
3. The twisted subalgebras are: $\{e_1 + ae_3, e_2\}$, $\{e_1 + ae_3\}, \{e_2 + be_3\}$. Since the normalizer of $\{e_2\}$ is $\{e_1, e_2\}$ we can use e_1 to "scale" $b \neq 0$ into $\epsilon = \pm 1$. Thus we have $a \neq 0$, $b \equiv \epsilon = \pm 1$ and the list of subalgebras is completed.

In Table I we present representatives of each conjugacy class of subalgebras of all two- and three-dimensional real Lie algebras. The algebras are listed in the first column using notation introduced previously. We sometimes indicate the corresponding Lie group in brackets [e.g., the Weyl group, the semidirect product of dilations and translations $D \square T_2$, the Euclidean group E(2) and pseudo-Euclidean group E(1, 1), etc.] All nonzero commutation relations are in the second column. The subalgebras are then listed by dimension and type in the last three columns.

Table II for $d = 4$ algebras is arranged similarly. The subalgebras are given in three columns for dimensions $d = 3$, 2, and 1 respectively. Within each column they are ordered by type.

In both tables we list the generators of subgroups in such a manner that the generators of the derived algebra are written to the right of a semicolon, e.g., $\{e_2, e_3; e_1\}$ for $A_{3,1}$, $\{; e_1, e_2, e_3\}$ for $A_{3,9}$ [the algebra of O(3)]. For Abelian subalgebras the semicolon which should be on the extreme right is omitted.

Algebras are generally indicated by braces, e.g., $\{e_1; e_2\}$. However maximal subalgebras of each considered Lie algebra are enclosed in ordinary brackets, e.g., $(e_1; e_2)$.

3. CONCLUDING REMARKS

The results of this paper are summarized in Tables I and II. Together with our earlier results on the invariants of low dimensional Lie algebras[3] these present a fairly complete picture of the structure and properties of real Lie algebras with $d \leq 4$. This knowledge should be of use in the study of any physical system, or any system of equations, having a symmetry described by one of the studied algebras. Furthermore, the presented classification is helpful in the studies of the Lie subgroup structure of larger Lie groups, e.g., like the conformal group SU(2, 2). It should be mentioned that the classification is in a sense "maximal." Thus, conjugacy in this article was considered with respect to the identity component of the Lie group corresponding to each studied Lie algebra (i.e., the group of inner auto-

TABLE II. Proper subalgebras of real Lie algebras of dimension 4. The range of parameters: $-\infty < x, y, z, v < \infty$, $0 \leq \phi < \pi$, $\epsilon = \pm 1$.

Algebra	Nonzero commutation relations	Subalgebras Dimension 3	Dimension 2	Dimension 1
$4A_1$		$3A_1$: $(e_1+xe_4, e_2+ye_4, e_3+ze_4)$, (e_1+xe_2, e_3, e_4), (e_2, e_3, e_4) $(e_1+xe_3, e_2+ye_3, e_4)$	$2A_1$: $\{e_1+xe_3+ye_4, e_2+ze_3+ve_4\}$, $\{e_1+xe_2+ye_4, e_3+ze_4\}$, $\{e_2+xe_4, e_3+ye_4\}$, $\{e_1+xe_2+ye_3, e_4\}$ $\{e_2+xe_3, e_4\}$, $\{e_3, e_4\}$	$\{e_1+xe_2+ye_3+ze_4\}$, $\{e_2+xe_3+ye_4\}$, $\{e_3+xe_4\}$, $\{e_4\}$
$A_2 \oplus 2A_1$	$[e_1, e_2] = e_2$	$3A_1$: (e_1, e_3, e_4), (e_2, e_3, e_4) $A_2 \oplus A_1$: $(e_1+x(e_3\cos\phi+e_4\sin\phi),$ $e_3\sin\phi - e_4\cos\phi; e_2)$	A_2: $\{e_1+x(e_3\cos\phi+e_4\sin\phi); e_2\}$ $2A_1$: $\{e_1+x(e_3\cos\phi+e_4\sin\phi), e_3\sin\phi - e_4\cos\phi\}$, $\{e_3, e_4\}$, $\{e_2+\epsilon(e_3\cos\phi+e_4\sin\phi), e_3\sin\phi - e_4\cos\phi\}$, $\{e_2, e_3\sin\phi - e_4\cos\phi\}$	$\{e_2\}$, $\{e_3\cos\phi+e_4\sin\phi\}$, $\{e_1+x(e_3\cos\phi+e_4\sin\phi)\}$ $\{e_2+\epsilon(e_3\cos\phi+e_4\sin\phi)\}$
$2A_2$	$[e_1, e_2] = e_2$, $[e_3, e_4] = e_4$	$A_1 \oplus A_2$: $(e_1, e_3; e_2)$, $(e_1, e_4; e_2)$, $(e_1, e_3; e_4)$, $(e_2, e_3; e_4)$ $A_{3,3}$: $(e_1+e_3; e_2, e_4)$ $A_{3,4}$: $(e_1-e_3; e_2, e_4)$ $A^a_{3,5}$: $(e_1+xe_3; e_2, e_4)$, $a = \begin{cases} x, & 0<\lvert x\rvert<1 \\ 1/x, & 1<\lvert x\rvert<\infty \end{cases}$	A_2: $\{e_1+xe_3; e_2\}$, $\{e_3+xe_1; e_4\}$, $\{e_1+\epsilon e_4; e_2\}$, $\{e_3+\epsilon e_2; e_4\}$, $\{e_1+e_3; e_2+\epsilon e_4\}$ $2A_1$: $\{e_1, e_3\}$, $\{e_1, e_4\}$, $\{e_2, e_3\}$, $\{e_2, e_4\}$	$\{e_2\}$, $\{e_3\}$, $\{e_4\}$, $\{e_1+xe_3\}$ $\{e_1+\epsilon e_4\}$, $\{e_2+\epsilon e_4\}$, $\{e_2+\epsilon e_3\}$
$A_{3,1} \oplus A_1$	$[e_2, e_3] = e_1$	$3A_1$: $(e_1, e_2\cos\phi+e_3\sin\phi, e_4)$ $A_{3,1}$: $(e_2+xe_4, e_3+ye_4; e_1)$	$2A_1$: $\{e_1, e_4+x(e_2\cos\phi+e_3\sin\phi)\}$, $\{e_4, e_2\cos\phi+e_3\sin\phi\}$ $\{e_1+xe_4, e_2\cos\phi+e_3\sin\phi\}$	$\{e_1+xe_4\}$, $\{e_4\}$, $\{e_2\cos\phi+e_3\sin\phi+xe_4\}$
$A_{3,2} \oplus A_1$	$[e_1, e_3] = e_1$, $[e_2, e_3] = e_1+e_2$	$3A_1$: (e_1, e_2, e_4) $A_2 \oplus A_1$: $(e_3, e_4; e_1)$ $A_{3,2}$: $(e_3+xe_4; e_1, e_2)$	$2A_1$: $\{e_1+xe_4, e_2\}$, $\{e_1, e_2+\epsilon e_4\}$, $\{e_1, e_4\}$, $\{e_2, e_4\}$, $\{e_3, e_4\}$ A_2: $\{e_3+xe_4; e_1\}$	$\{e_1\}$, $\{e_1+\epsilon e_4\}$, $\{e_2+xe_4\}$, $\{e_3+xe_4\}$, $\{e_4\}$
$A_{3,3} \oplus A_1$	$[e_1, e_3] = e_1$, $[e_2, e_3] = e_2$	$3A_1$: (e_1, e_2, e_4) $A_2 \oplus A_1$: $(e_3, e_4; e_1\cos\phi+e_2\sin\phi)$ $A_{3,3}$: $(e_3+xe_4; e_1, e_2)$	$2A_1$: $\{e_1, e_2\}$, $\{e_1\cos\phi+e_2\sin\phi, e_4\}$, $\{e_3, e_4\}$, $\{e_1, e_2+\epsilon e_4\}$, $\{e_1+\epsilon e_4, e_2+xe_1\}$ A_2: $\{e_3+xe_4; e_1\cos\phi+e_2\sin\phi\}$	$\{e_1\cos\phi+e_2\sin\phi\}$, $\{e_3+xe_4\}$, $\{e_4\}$, $\{e_1\cos\phi+e_2\sin\phi+\epsilon e_4\}$
$A_{3,4} \oplus A_1$	$[e_1, e_3] = e_1$, $[e_2, e_3] = -e_2$	$3A_1$: (e_1, e_2, e_4) $A_2 \oplus A_1$: $(e_3, e_4; e_1)$, $(e_3, e_4; e_2)$ $A_{3,4}$: $(e_3+xe_4; e_1, e_2)$	$2A_1$: $\{e_1, e_2\}$, $\{e_1, e_4\}$, $\{e_2, e_4\}$, $\{e_3, e_4\}$, $\{e_1+\epsilon e_2, e_4\}$, $\{e_1, e_2+\epsilon e_4\}$, $\{e_1+\epsilon e_4, e_2+xe_1\}$ A_2: $\{e_3+xe_4; e_1\}$, $\{e_3+xe_4; e_2\}$	$\{e_1\}$, $\{e_2\}$, $\{e_4\}$, $\{e_3+xe_4\}$, $\{e_1+\epsilon e_2+xe_4\}$, $\{e_1+\epsilon e_4\}$, $\{e_2+\epsilon e_4\}$
$A^a_{3,5} \oplus A_1$ $(0<\lvert a\rvert<1)$	$[e_1, e_3] = e_1$, $[e_2, e_3] = ae_2$	$3A_1$: (e_1, e_2, e_4) $A_2 \oplus A_1$: $(e_3, e_4; e_1)$, $(e_3, e_4; e_2)$ $A^a_{3,5}$: $(e_3+xe_4; e_1, e_2)$	$2A_1$: $\{e_1, e_2\}$, $\{e_1, e_4\}$, $\{e_2, e_4\}$, $\{e_3, e_4\}$, $\{e_1+\epsilon e_2, e_4\}$, $\{e_1, e_2+\epsilon e_4\}$, $\{e_1+\epsilon e_4, e_2+xe_4\}$ A_2: $\{e_3+xe_4; e_1\}$, $\{e_3+xe_4; e_2\}$	$\{e_1\}$, $\{e_2\}$, $\{e_4\}$, $\{e_1+\epsilon e_4\}$ $\{e_2+\epsilon e_4\}$, $\{e_3+xe_4\}$, $\{e_1+\epsilon e_2+xe_4\}$
$A_{3,6} \oplus A_1$	$[e_1, e_3] = -e_2$, $[e_2, e_3] = e_1$	$3A_1$: (e_1, e_2, e_4) $A_{3,6}$: $(e_3+xe_4; e_1, e_2)$	$2A_1$: $\{e_1+xe_4, e_2\}$, $\{e_1, e_4\}$, (e_3, e_4) $(x \geq 0)$	$\{e_4\}$, $\{e_1+xe_4\}$, $\{e_3+ye_4\}$ $(x \geq 0)$

TABLE II. (*Continued*)

Algebra	Nonzero commutation relations	Subalgebras Dimension 3	Dimension 2	Dimension 1
$A_{3,7}^{a}\oplus A_1$ ($a>0$)	$[e_1,e_3]=ae_1-e_2$, $[e_2,e_3]=e_1+ae_2$	$3A_1$: (e_1,e_2,e_4) $A_{3,7}^{a}$: $(e_3+xe_4;e_1,e_2)$	$2A_1$: $\{e_1,e_4\}$, (e_3,e_4), $\{e_1+xe_4,e_2\}$ $(x\geq 0)$	$\{e_4\}$, $\{e_1+xe_4\}$, $\{e_3+ye_4\}$, $(x\geq 0)$
$A_{3,8}\oplus A_1$	$[e_3,e_1]=2e_2$, $[e_1,e_2]=e_1$, $[e_2,e_3]=e_3$	$A_2\oplus A_1$: $(e_2,e_4;e_1)$ $A_{3,8}$: $(;e_1,e_2,e_3)$	$2A_1$: $\{e_1,e_4\}$, $\{e_2,e_4\}$, (e_1+e_3,e_4) A_2: $\{e_2+xe_4;e_1\}$	$\{e_1\}$, $\{e_4\}$, $\{e_2+xe_4\}$ $(x\geq 0)$ $\{e_1+e_3+ye_4\}$, $\{e_1+\epsilon e_4\}$
$A_{3,9}\oplus A_1$	$[e_1,e_2]=e_3$, $[e_2,e_3]=e_1$, $[e_3,e_1]=e_2$	$A_{3,8}$: $(;e_1,e_2,e_3)$	$2A_1$: (e_1,e_4)	$\{e_4\}$, $\{e_1+xe_4\}$ $(x\geq 0)$
$A_{4,1}$	$[e_2,e_4]=e_1$, $[e_3,e_4]=e_2$	$3A_1$: (e_1,e_2,e_3) $A_{3,1}$: $(e_4+xe_3,e_2;e_1)$	$2A_1$: $\{e_1,e_2\}$, $\{e_1,e_3\}$, $\{e_2,e_3+xe_1\}$, $\{e_1,e_4+xe_3\}$	$\{e_1\}$, $\{e_2\}$, $\{e_3+xe_1\}$, $\{e_4+xe_3\}$
$A_{4,2}^{a}$ ($a\neq 0,1$)	$[e_1,e_4]=ae_1$, $[e_2,e_4]=e_2$, $[e_3,e_4]=e_2+e_3$	$3A_1$: (e_1,e_2,e_3) $A_{3,2}$: $(e_4;e_2,e_3)$ $A_{3,4}$: $(e_4;e_1,e_2)$, $a=-1$ $A_{3,5}^{v}$: $(e_4;e_1,e_2)$, $v=\begin{cases}a, & \lvert a\rvert<1\\ 1/a, & \lvert a\rvert>1\end{cases}$	$2A_1$: $\{e_1,e_2\}$, $\{e_1+xe_2,e_3\}$, $\{e_2,e_3\}$, $\{e_1+\epsilon e_3,e_2\}$ A_2: $\{e_4;e_1\}$, $\{e_4;e_2\}$	$\{e_1\}$, $\{e_2\}$, $\{e_4\}$, $\{e_1+\epsilon e_2\}$, $\{e_3+xe_1\}$
$A_{4,2}^{1}$	$[e_1,e_4]=e_1$, $[e_2,e_4]=e_2$, $[e_3,e_4]=e_2+e_3$	$3A_1$: (e_1,e_2,e_3) $A_{3,2}$: $(e_4;e_2,e_3+xe_1)$ $A_{3,3}$: $(e_4;e_1,e_2)$	$2A_1$: $\{e_1,e_2\}$, $\{e_1+xe_2,e_3\}$, $\{e_2,e_3+xe_1\}$ A_2: $\{e_4;e_1\cos\phi+e_2\sin\phi\}$	$\{e_1\cos\phi+e_2\sin\phi\}$, $\{e_3+xe_1\}$, $\{e_4\}$
$A_{4,3}$	$[e_1,e_4]=e_1$, $[e_3,e_4]=e_2$	$3A_1$: (e_1,e_2,e_3) $A_2\oplus A_1$: $(e_4+xe_3,e_2;e_1)$ $A_{3,1}$: $(e_3,e_4;e_2)$	$2A_1$: $\{e_1,e_2\}$, $\{e_1+xe_2,e_3\}$, $\{e_2,e_3\}$, $\{e_2,e_3+\epsilon e_1\}$, $\{e_2,e_4+xe_3\}$ A_2: $\{e_4+xe_3;e_1\}$	$\{e_1\}$, $\{e_2\}$, $\{e_1+\epsilon e_2\}$, $\{e_3+xe_1\}$, $\{e_4+xe_3\}$
$A_{4,4}$	$[e_1,e_4]=e_1$, $[e_2,e_4]=e_1+e_2$, $[e_3,e_4]=e_2+e_3$	$3A_1$: (e_1,e_2,e_3) $A_{3,2}$: $(e_4;e_1,e_2)$	$2A_1$: $\{e_1+xe_3,e_2\}$, $\{e_1,e_3\}$, $\{e_2,e_3\}$ A_2: $\{e_4;e_1\}$	$\{e_1+xe_3\}$, $\{e_2\}$, $\{e_3\}$, $\{e_4\}$
$A_{4,5}^{a,b}$ ($-1\leq a<b<1$, $ab\neq 0$)	$[e_1,e_4]=e_1$, $[e_2,e_4]=ae_2$, $[e_3,e_4]=be_3$	$3A_1$: (e_1,e_2,e_3) $A_{3,5}^{a}$: $(e_4;e_1,e_2)$ $A_{3,5}^{b}$: $(e_4;e_1,e_3)$ $A_{3,5}^{v}$: $(e_4;e_2,e_3)$, $v=\begin{cases}a/b, & \lvert a/b\rvert<1\\ b/a, & \lvert a/b\rvert>1\end{cases}$	$2A_1$: $\{e_1,e_2\}$, $\{e_1,e_3\}$, $\{e_2,e_3\}$, $\{e_1,e_2+\epsilon e_3\}$, $\{e_2,e_1+\epsilon e_3\}$, $\{e_3,e_1+\epsilon e_2\}$, $\{e_1+\epsilon e_3,e_2+xe_3\}$ $(x\neq 0)$ A_2: $\{e_4;e_1\}$, $\{e_4;e_2\}$, $\{e_4;e_3\}$	$\{e_1\}$, $\{e_2\}$, $\{e_3\}$, $\{e_4\}$ $\{e_1+\epsilon e_3\}$, $\{e_2+\epsilon e_3\}$, $\{e_1+\epsilon e_2+xe_3\}$ $(x\neq 0)$
$A_{4,5}^{a,a}$ ($-1\leq a<1$, $a\neq 0$)	$[e_1,e_4]=e_1$, $[e_2,e_4]=ae_2$ $[e_3,e_4]=ae_3$	$3A_1$: (e_1,e_2,e_3) $A_{3,3}$: $(e_4;e_2,e_3)$ $A_{3,5}^{a}$: $(e_4;e_1,e_2\cos\phi+e_3\sin\phi)$	$2A_1$: $\{e_1,e_2\cos\phi+e_3\sin\phi\}$, $\{e_2,e_3\}$, $\{e_3,e_1+\epsilon e_2\}$, $\{e_1+\epsilon e_3,e_2+xe_3\}$ A_2: $\{e_4;e_1\}$, $\{e_4;e_2\cos\phi+e_3\sin\phi\}$	$\{e_1\}$, $\{e_2\cos\phi+e_3\sin\phi\}$, $\{e_4\}$, $\{e_1+\epsilon e_3\}$, $\{e_1+\epsilon e_2+xe_3\}$

TABLE II. (*Continued*)

Algebra	Nonzero commutation relations	Subalgebras Dimension 3	Dimension 2	Dimension 1
$A_{4,5}^{a,1}$ $(-1 \leq a < 1, a \neq 0)$	$[e_1, e_4] = e_1$, $[e_2, e_4] = ae_2$ $[e_3, e_4] = e_3$	$3A_1$: (e_1, e_2, e_3) $A_{3,3}$: $(e_4; e_1, e_3)$ $A_{3,5}^{a}$: $(e_4; e_1\cos\phi + e_3\sin\phi, e_2)$	$2A_1$: $\{e_1\cos\phi + e_3\sin\phi, e_2\}$, $\{e_1, e_3\}$, $\{e_3, e_1 + \epsilon e_2\}$, $\{e_1 + xe_3, e_2 + \epsilon e_3\}$ A_2: $\{e_4; e_2\}$, $\{e_4; e_1\cos\phi + e_3\sin\phi\}$	$\{e_2\}$, $\{e_1\cos\phi + e_3\sin\phi\}$, $\{e_4\}$, $\{e_2 + \epsilon e_3\}$, $\{e_1 + \epsilon e_2 + xe_3\}$
$A_{4,5}^{1,1}$	$[e_1, e_4] = e_1$, $[e_2, e_4] = e_2$, $[e_3, e_4] = e_3$	$3A_1$: (e_1, e_2, e_3) $A_{3,3}$: $(e_4; e_1 + xe_3, e_2 + ye_3)$, $(e_4; e_1 + xe_2, e_3)$, $(e_4; e_2, e_3)$	$2A_1$: $\{e_1 + xe_3; e_2 + ye_3\}$, $\{e_1 + xe_2, e_3\}$, $\{e_2, e_3\}$ A_2: $\{e_4; e_1 + xe_2 + ye_3\}$, $\{e_4; e_2 + xe_3\}$, $\{e_4; e_3\}$	$\{e_1 + xe_2 + ye_3\}$, $\{e_2 + xe_3\}$, $\{e_3\}$, $\{e_4\}$
$A_{4,6}^{a,b}$ $(a \neq 0, b \geq 0)$	$[e_1, e_4] = ae_1$, $[e_2, e_4] = be_2 - e_3$, $[e_3, e_4] = e_2 + be_3$	$3A_1$: (e_1, e_2, e_3) $A_{3,7}^{b}$: $(e_4; e_2, e_3)$	$2A_1$: $\{e_1 + xe_3, e_2\}$, $\{e_2, e_3\}$ $(x \geq 0)$ A_2: $(e_4; e_1)$	$\{e_1 + xe_3\}$, $\{e_3\}$, $\{e_4\}$ $(x \geq 0)$
$A_{4,7}$	$[e_1, e_4] = 2e_1$, $[e_2, e_4] = e_2$, $[e_3, e_4] = e_2 + e_3$, $[e_2, e_3] = e_1$	$A_{3,1}$: $(e_2, e_3; e_1)$ $A_{3,5}^{1/2}$: $(e_4; e_1, e_2)$	$2A_1$: $\{e_1, e_2\}$, $\{e_1, e_3\}$ A_2: $\{e_4; e_1\}$, $\{e_4; e_2\}$	$\{e_1\}$, $\{e_2\}$, $\{e_3\}$, $\{e_4\}$
$A_{4,8}$	$[e_2, e_3] = e_1$, $[e_2, e_4] = e_2$, $[e_3, e_4] = -e_3$	$A_{3,1}$: $(e_2, e_3; e_1)$ $A_2 \oplus A_1$: $(e_4, e_1; e_2)$, $(e_4, e_1; e_3)$	$2A_1$: $\{e_1, e_2\}$, $\{e_1, e_3\}$, $\{e_1, e_2 + \epsilon e_3\}$, $\{e_4, e_1\}$ A_2: $\{e_4 + xe_1, e_2\}$; $\{e_4 + xe_1; e_3\}$	$\{e_1\}$, $\{e_2\}$, $\{e_2 + \epsilon e_3\}$, $\{e_3\}$ $\{e_4 + xe_1\}$
$A_{4,9}^{b}$ $(0 < \lvert b\rvert < 1)$	$[e_2, e_3] = e_1$, $[e_1, e_4] = (1 + b)e_1$, $[e_2, e_4] = e_2$, $[e_3, e_4] = be_3$	$A_{3,1}$: $(e_2, e_3; e_1)$ $A_{3,5}^{v}$: $(e_4; e_1, e_2)$, $v = \begin{cases} 1+b, & \lvert 1+b\rvert < 1 \\ 1/(1+b), & \lvert 1+b\rvert > 1 \end{cases}$ $A_{3,4}$: $(e_4; e_1, e_3)$, $b = -\frac{1}{2}$ $A_{3,5}^{w}$: $(e_4; e_1, e_3)$, $w = \begin{cases} b/(1+b), & \lvert b/(1+b)\rvert < 1 \\ (1+b)/b, & \lvert b/(1+b)\rvert > 1 \end{cases}$	$2A_1$: $\{e_1, e_2\}$, $\{e_1, e_3\}$, $\{e_1, e_2 + \epsilon e_3\}$ A_2: $\{e_4; e_1\}$, $\{e_4; e_2\}$, $\{e_4; e_3\}$	$\{e_1\}$, $\{e_2\}$, $\{e_3\}$, $\{e_4\}$, $\{e_2 + \epsilon e_3\}$
$A_{4,9}^{1}$	$[e_2, e_3] = e_1$, $[e_1, e_4] = 2e_1$, $[e_2, e_4] = e_2$, $[e_3, e_4] = e_3$	$A_{3,1}$: $(e_2, e_3; e_1)$ $A_{3,5}^{1/2}$: $(e_4; e_1, e_2\cos\phi + e_3\sin\phi)$	$2A_1$: $\{e_1, e_2\cos\phi + e_3\sin\phi\}$ A_2: $\{e_4; e_1\}$, $\{e_4; e_2\cos\phi + e_3\sin\phi\}$	$\{e_1\}$, $\{e_2\cos\phi + e_3\sin\phi\}$, $\{e_4\}$
$A_{4,9}^{0}$	$[e_2, e_3] = e_1$, $[e_1, e_4] = e_1$, $[e_2, e_4] = e_2$	$A_{3,1}$: $(e_2, e_3; e_1)$ $A_2 \oplus A_1$: $(e_3, e_4; e_1)$ $A_{3,3}$: $(e_4; e_1, e_2)$ $A_{3,2}$: $(e_4 + xe_3; e_1, e_2)$ $(x \neq 0)$	$2A_1$: $\{e_1, e_2\}$, $\{e_1, e_3\}$, $\{e_1, e_2 + \epsilon e_3\}$ $\{e_3, e_4\}$ A_2: $\{e_4 + xe_3; e_1\}$, $\{e_4; e_2\}$	$\{e_1\}$, $\{e_2\}$, $\{e_3\}$, $\{e_2 + \epsilon e_3\}$, $\{e_4 + xe_3\}$
$A_{4,10}$	$[e_2, e_3] = e_1$, $[e_2, e_4] = -e_3$, $[e_3, e_4] = e_2$	$A_{3,1}$: $(e_2, e_3; e_1)$	$2A_1$: $\{e_1, e_2\}$, (e_1, e_4)	$\{e_1\}$, $\{e_2\}$, $\{e_4 + xe_1\}$
$A_{4,11}^{a}$ $(0 < a)$	$[e_2, e_3] = e_1$, $[e_1, e_4] = 2ae_1$, $[e_2, e_4] = ae_2 - e_3$, $[e_3, e_4] = e_2 + ae_3$	$A_{3,1}$: $(e_2, e_3; e_1)$	$2A_1$: $\{e_1, e_2\}$ A_2: $(e_4; e_1)$	$\{e_1\}$, $\{e_2\}$, $\{e_4\}$
$A_{4,12}$	$[e_1, e_3] = e_1$, $[e_2, e_3] = e_2$, $[e_1, e_4] = -e_2$, $[e_2, e_4] = e_1$	$A_{3,3}$: $(e_3; e_1, e_2)$ $A_{3,6}$: $(e_4; e_1, e_2)$ $A_{3,7}^{\lvert x\rvert}$: $(e_4 + xe_3; e_1, e_2)$ $(x \neq 0)$	$2A_1$: $\{e_1, e_2\}$, (e_3, e_4) A_2: $\{e_3; e_1\}$	$\{e_1\}$, $\{e_3\}$, $\{e_4 + xe_3\}$

morphisms). When the algebra itself is realized as a subalgebra of some larger algebra, further conjugacies will enter into the picture and different subalgebras, that in this article are isomorphic but not conjugate, may become conjugate under the larger group.

A problem of importance in physics and other applications is the representation theory of each of the above Lie algebras. For some of them the representation theory has been completely developed [e.g., O(3), O(2, 1), E(2) *et al.*[14]]. For other Lie algebras, specially the solvable ones, this theory is much less complete. While a large amount of mathematical work on the representations of nilpotent and solvable groups has been performed,[15-18] it has not been systematically applied to low dimensional Lie algebras.

An article on the subalgebras of five-dimensional Lie algebras is in preparation, as are applications of the present classification to the study of subgroups of O(3, 2), SU(2, 2), and other Lie groups. Work is also in progress on the representation theory of various low dimensional Lie groups and algebras, in particular considering reductions to their different subgroups.

ACKNOWLEDGMENTS

In conclusion we thank M. Perroud, R. T. Sharp, and H. Zassenhaus for helpful discussions.

*Supported in part by the National Research Council of Canada and the North Atlantic Treaty Organization.

[1]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597, 1615 (1975); **17**, 717 (1976).
[2]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, Can. J. Phys. **54**, 950 (1976); J. Math. Phys. **17**, 977 (1976).
[3]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[4]C. P. Boyer, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976); E. G. Kalnins, W. Miller, Jr., and P. Winternitz, SIAM J. Appl Math. **30**, 630 (1976); J. Beckers, J. Patera, M. Perroud, and P. Winternitz, J. Math. Phys. **18**, 72 (1977).
[5]G. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[6]L. Abellanos and L. M. Alonso, J. Math. Phys. **16**, 1580 (1975).
[7]E. Inonu and E. P. Wigner, Proc. Nat. Acad. Sci. **39**, 510 (1953); **40**, 119 (1954); E. J. Saletan, J. Math. Phys. **2**, 1 (1961); C. W. Conatser, J. Math. Phys. **13**, 196 (1972); I. E. Segal, Duke Math. J. **18**, 221 (1951).
[8]S. Lie, *Theorie der Transformationsgruppen* (Teubner, Leipzig, 1893).
[9]G. M. Mubarakzyanov, Izv. Vysshikh Uchebn. Zavedeniĭ Mat. No. 1 (32), 114; No. 3 (34), 99; No. 4 (35), 104 (1963).
[10]E. Goursat, Ann. Sci. Ec. Norm. Super. (Paris) **6** (3), 9 (1889).
[11]N. Jacobson, *Lie Algebras* (Interscience, New York, 1955).
[12]E. Bodewig, *Matrix Calculus* (North-Holland, Amsterdam, 1959); W. H. Greub, *Linear Algebra* (Springer, New York, 1967).
[13]H. Grassmann, *Gesammelte Mathematische und Physikalische Werke* (Teubner, Leipzig, 1894), Band 1, Teil 1.
[14]N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, R. I., 1968).
[15]P. Bernat, N. Conze, M. Duflo, M. Lévy-Nahas, M. Rais, P. Renouard, and M. Vergne, *Rêpresentations des Groupes de Lie Rêsolubles* (Dunod, Paris, 1972).
[16]L. Auslander and B. Kostant, Bull. Am. Math. Soc. **73**, 692 (1967).
[17]J. Dixmier, Bull. Soc. Math. France **85**, 113, 325 (1957); **94**, 181 (1966).
[18]A. A. Kirillov, Uspekhi Mat. Nauk. **17**, 57 (1962). A. A. Kirillov, *Elementy Teorii Predstavleniĭ* (Elements of Representation Theory) (Nauka, Moscow, 1972).

New algorithms for the Molien function*

Marko V. Jarić and Joseph L. Birman

Physics Department, City College—CUNY, New York, New York 10031
(Received 28 January 1977)

Two new forms are given for the Molien (generating) function for multiplicity c_{n1} of the identity representation in the symmetrized nth power of representation Γ of finite group G. These are

$$M(\Gamma,G;z) = \Sigma_n c_{n1} z^n = 1/|G| \Sigma_g \exp\left[\Sigma_l z^l \chi(g^l)/l\right] = 1/|G| \Sigma_g \Pi_j [1 - z\gamma_j(g)]^{-\delta_j},$$

where g is an element in G, $\gamma_j(g)$ an irreducible character in the Abelian subgroup A generated by g, δ_j the subduction coefficient of Γ of $G|\gamma_j$ of A; we usually take Γ irreducible. These forms have the merit of only requiring characters. In the following paper these algorithms are used to compute the Molien functions for space group irreducible representations.

1. INTRODUCTION

The Molien function[1] for a finite group G is the generating function for the multiplicity c_{n1} with which the trivial or identity representation Γ^1 is contained in the symmetrized nth power of a representation Γ of G. The Molien function is also the formal power series

$$M(\Gamma, G; z) = \sum_{n=0}^{\infty} c_{n1} z^n . \tag{1.1}$$

By definition $c_{01} \equiv 1$. The usual form for the generating function is

$$M(\Gamma, G; z) = \frac{1}{|G|} \sum_{g} \frac{1}{\det[1 - z\Gamma(g)]} , \tag{1.2}$$

where the sum is over all elements g in the group G, $|G|$ is the order of G.

It may be helpful in the interpretation of (1.1), (1.2) to recall that if, in the l-dimensional representation Γ of G, we diagonalize the one matrix $\Gamma(\bar{g})$ for element $\bar{g}$ and find its l eigenvalues

$$\mathrm{diag}\Gamma(\bar{g}) = (\bar{\omega}_1, \ldots, \bar{\omega}_l), \tag{1.3}$$

then we can construct the partial Molien function for element $\bar{g}$ in the representation Γ:

$$m(\Gamma, \bar{g}; z) \equiv 1 \Big/ \prod_{i=1}^{l} (1 - z\bar{\omega}_i) \tag{1.4}$$

and

$$M(\Gamma, G; z) = \frac{1}{|G|} \sum_{g} m(\Gamma, g; z) . \tag{1.5}$$

Also, since the partial Molien function is a class function,

$$M(\Gamma, G; z) = \frac{1}{|G|} \sum_{k} c_k m(\Gamma, C_k; z), \tag{1.6}$$

where the sum is over all classes C_k in the group G, and c_k is the order of class C_k.

In actual application, the forms (1.2), (1.4) have the disadvantage of requiring diagonalization of all the Γ.

The present short paper reports on two equivalent presentations for the Molien function, only requiring the characters

$$\chi(g) = \mathrm{Tr}\Gamma(g), \tag{1.7}$$

which are more readily available. This should facilitate applications, especially to crystal space groups. In the accompanying paper we report on such applications and give, for the first time we believe, the Molien function for any space group representations.

2. FIRST NEW ALGORITHM FOR THE MOLIEN FUNCTION

From the character system for the symmetrized nth power of representation Γ we can find the c_{n1}. Let $\chi_{(n)}(g)$ be the character of g in the symmetrized nth power of Γ; then, since all characters of Γ^1 are 1,

$$c_{n1} = \frac{1}{|G|} \sum_{g} \chi_{(n)}(g) . \tag{2.1}$$

A closed, but cumbersome expression exists[2] for $\chi_{(n)}(g)$:

$$\chi_{(n)}(g) = \sum_{\kappa_1, \ldots, \kappa_n} \frac{\chi^{\kappa_1}(g) \cdots \chi^{\kappa_n}(g^n)}{1^{\kappa_1}\kappa_1! 2^{\kappa_2}\kappa_2! \cdots n^{\kappa_n}\kappa_n!} , \tag{2.2}$$

with

$$\sum_{l=1}^{n} l\kappa_l = n \tag{2.3}$$

and κ_l zero or a positive integer. The objective of this section is to rewrite $\chi_{(n)}$ in terms of a simpler generating function and thereby also c_{n1}.

First we include the condition (2.3) by incorporating it as a Kronecker delta in (2.2), and extending multiple sums on $\kappa_1 \cdots \kappa_n$ to infinity:

$$\chi_{(n)}(g) = \sum_{\kappa_1, \ldots, \kappa_n} \delta_{(n, \Sigma l\kappa_l)} \frac{\chi^{\kappa_1}(g) \cdots \chi^{\kappa_n}(g^n)}{1^{\kappa_1}\kappa_1! \cdots n^{\kappa_n}\kappa_n!} . \tag{2.4}$$

Next we use an integral representation of the delta

$$\delta_{p,0} = \mathrm{Res} \frac{1}{z^{p+1}} = \frac{1}{2\pi i} \oint_C \frac{dz}{z^{p+1}} , \tag{2.5}$$

with p integer, where C is a closed contour around $z = 0$ and Res means "residue." Substituting (2.5) into (2.4), we obtain

$$\chi_{(n)}(g) = \frac{1}{2\pi i} \oint_C \frac{dz}{z^{n+1}} \prod_{l=1}^{n} \sum_{\kappa_l=0}^{\infty} \frac{z^{l\kappa_l} \chi^{\kappa_l}(g^l)}{(l)^{\kappa_l}(\kappa_l)!} . \tag{2.6}$$

We also used

$$\prod_{l=1}^{n} z^{l\kappa_l} = \frac{1}{z^{-\Sigma l\kappa_l}} . \tag{2.7}$$

Now Eq. (2.6) can be rearranged to give

$$\chi_{(n)}(g)=\frac{1}{2\pi i}\oint_C\frac{dz}{z^{n+1}}\prod_{l=1}^{n}\sum_{\kappa_l=0}^{\infty}\left(\frac{z^l\chi(g^l)}{l}\right)^{\kappa_l}\frac{1}{(\kappa_l)!}, \tag{2.8}$$

which can now be recognized as follows:

$$\chi_{(n)}(g)=\frac{1}{2\pi i}\oint_C\frac{dz}{z^{n+1}}\prod_{l=1}^{n}\exp\frac{z^l\chi(g^l)}{l}$$
$$=\frac{1}{n!}\frac{d^n}{dz^n}\exp\left(\sum_{l=1}^{n}\frac{z^l\chi(g^l)}{l}\right)\Bigg|_{z=0}. \tag{2.9}$$

We may now formally extend the sum upon l to ∞. This sum is convergent since we may take $|z|<1$; also for finite groups the characters of powers of g are periodic [e.g., if p is the smallest integer such that $g^p=e$, then $\chi(g^{l+p})=\chi(g^l)$]:

$$\chi_{(n)}(g)=\frac{1}{n!}\frac{d^n}{dz^n}\left\{\exp\left(\sum_{l=1}^{\infty}\frac{z^l\chi(g^l)}{l}\right)\right\}\Bigg|_{z=0}, \tag{2.10}$$

which is the nth term in the Taylor expansion about $z=0$ of the function inside the curly brackets. Hence

$$\exp\left(\sum_{l=1}^{\infty}\frac{z^l\chi(g^l)}{l}\right)=\sum_{n=0}^{\infty}\chi_{(n)}(g)z^n, \tag{2.11}$$

where $\chi_0(g)\equiv 1$ by definition.

Recalling (2.1), we have

$$M(\Gamma,G;z)=\frac{1}{|G|}\sum_{g}\exp\left(\sum_{l=1}^{\infty}\frac{z^l\chi(g^l)}{l}\right)=\sum_{n=0}^{\infty}c_{n1}z^n. \tag{2.12}$$

This is our first expression for the Molien function in terms of a generating function which depends only on the character system of the representation.

To verify the connection with the usual form of the Molien function, we observe

$$\chi(g^l)=\mathrm{Tr}[\Gamma(g^l)]=\mathrm{Tr}[\Gamma(g)^l]. \tag{2.13}$$

Then the exponential function becomes

$$\exp\left(\mathrm{Tr}\sum_{l=1}^{\infty}\frac{z^l\Gamma(g)^l}{l}\right)=\det\exp\left(\sum_{l=1}^{\infty}\frac{[z\Gamma(g)]^l}{l}\right). \tag{2.14}$$

But for $|z|<1$

$$-\ln[1-z\Gamma(g)]=\sum_{l=1}^{\infty}\frac{[z\Gamma(g)]^l}{l}, \tag{2.15}$$

and, substituting back into (2.14), we have

$$\det\exp\{-\ln[1-z\Gamma(g)]\}$$
$$=\det\{1/[1-z\Gamma(g)]\}=1/\det[1-z\Gamma(g)]. \tag{2.16}$$

All that now remains is to average this expression over the group by multiplying by $1/|G|$ and summing on g to recover (1.2).

3. SECOND NEW ALGORITHM FOR THE MOLIEN FUNCTION

Let us return to (1.4), (1.5). In this equation the eigenvalues of $\Gamma(g)$ appear. If G is a finite group, then, as previously remarked, every element has finite period p: $g^p=e$. Of course, $p\leq|G|$. Consequently, $\Gamma(g^p)=\Gamma(e)$ and

$$\mathrm{diag}\Gamma(g^p)=(1,1,1,\cdots). \tag{3.1}$$

It follows that

$$\omega_j=\exp(2\pi i n_j/p), \tag{3.2}$$

where n_j is selected from the set of integers $1,\ldots,p$, i.e., the eigenvalues ω_j of Γ are selected from among the pth roots of unity. It may be that in a given representation $\Gamma(g)^t=\Gamma(e)$, where t divides p, but the argument does not depend on this.

In using (1.3), (1.4) we must determine which ω_j appear, and the multiplicity of appearance. Let us call δ_j the multiplicity [frequency of appearance of a particular ω_j in (1.3)].

Then (1.4) can be rewritten

$$1\Big/\prod_{j=1}^{p}(1-z\omega_j)^{\delta_j}, \tag{3.3}$$

where $\sum_{j=1}^{p}\delta_j=l$, the dimension of Γ.

Consider the subgroup A, generated by g, consisting of g and all its powers: $g,g^2,\ldots,g^p=e$. The irreducible representations of A are labeled by the pth roots of unity:

$$g\to\exp(2\pi i n_j/p)\equiv\gamma_j(g),\quad j=1,\ldots,p. \tag{3.4}$$

A little consideration then shows that if we consider Γ of G as a representation of A (i.e., we subduce),

$$\Gamma\text{ of }G\downarrow\gamma\text{ of }A, \tag{3.5}$$

then

$$\delta_j=\frac{1}{|A|}\sum_{m=1}^{p}\gamma_j(g^m)^*\chi(g^m). \tag{3.6}$$

That is, the multiplicity δ_j is the reduction coefficient of Γ of G, upon γ of A. To find the δ_j one simply reads off the characters $\chi(g^m)$ of the cyclic subgroup and carries out the reduction (3.6) in the usual manner; prior to this one has constructed all the needed basic set of characters γ_j of A, as pth roots of unity. In (3.6) $|A|=p$.

Assembling this we have

$$M(\Gamma,G;z)=\frac{1}{|G|}\sum_{g}\left\{1\Big/\prod_{j=1}^{p}[1-z\gamma_j(g)]^{\delta_j}\right\}. \tag{3.7}$$

Using the periodicity of the characters for finite groups mentioned above we have also verified that Eq. (3.7) follows from Eq. (2.12).

4. CONCLUSION

We have found both of our algorithms useful for the calculation of the Molien function, and we believe that both are new. In the accompanying paper we illustrate the use of these algorithms by calculating the Molien function for irreducible representations of O_h^3-Pm3n, an important nonsymmorphic space group.

The merit of our method is that it facilitates the calculation of the Molien function by only using characters, which are generally more readily available than representations.

Note added in proof: We thank Professor R. Stanley for recently informing us that (2.11) could also be obtained from (2.2) using a known combinatorial result.[3]

*Supported in part by Grants NSF No. DMR76-20641 and CUNY–FRAP No. 11453.

[1]T. Molien, Sitzungsber, Konig. Preuss. Akad. Wiss., 1152 (1897); W. Burnside, *Theory of Groups of Finite Order* (Dover, New York, 1952), 2nd ed.; R. P. Stanley (to be published); L. Michel, Proc. Sixth Int. Colloq. on Group Theory in Physics, Montreal (1976) (to be published).
[2]G. Y. Lyubarski, *Applications of Group Theory in Physics* (Pergamon, New York, 1960), p. 75.
[3]J. Riordan, *An Introduction to Combinational Analysis* (Wiley, New York, 1958), p. 68.

Calculation of the Molien generating function for invariants of space groups*

Marko V. Jarić and Joseph L. Birman

Physics Department, City College—CUNY, New York, New York 10031
(Received 28 January 1977)

The new algorithms for computing the Molien generating function for a representation of a finite group obtained in the preceding paper are applied to obtain an expression which can be used for irreducible representation (*$\mathbf{k}n$) of any crystallographic space group G. It proves convenient to express $M(\Gamma,G;z)$ as a sum: $M(\Gamma,G;z) = 1/|P|\Sigma_k c_k \bar{m}(\Gamma,g_k;z)$, where the partial Molien function $\bar{m}(\Gamma;g_k;z)$ is labelled by a coset representative g_k carrying the class index k of the point group $P = G/T$, T being the translation group, the sum is over classes k, and c_k is the order of class $C_k(P)$. The resulting form was used to compute $M(\Gamma,G;z)$ for irreducible representations *Γn, *$\mathbf{X}n$, *$\mathbf{R}n$ of nonsymmorphic space group A-15 or O_h^3-$Pm3n$ in which many high temperature superconducting crystals occur. Certain of these representations (matrix groups) are identified as generalized Coxeter groups, i.e., unitary groups generated by reflections. The Molien function for these groups has the required form given by Shephard and Todd: $M(\Gamma,G;z) = [\Pi_i(1-z^{d_i})]^{-1}$. The d_i satisfy dimensionality theorems.

1. INTRODUCTION

There are various physical problems for which it is of importance to determine polynomial invariants of a group G. The first step in the explicit determination of such invariants is to obtain the number of invariants of given degree. This can be accomplished by use of the Molien function,[1] for representation Γ of group G. Actual construction of the invariants is a separate step which can be accomplished by use of projection operator or an equivalent algebraic technique. Knowledge of the structure of the Molien function provides a useful guide for determination of the integrity basis of the ring of invariants.

If one is concerned with a phase transition in a crystal, one may have focussed attention upon some multicomponent order parameter $(x_1,\ldots,x_l)$, where x_j are vectors in a Hilbert space, such that under action of elements g in G

$$g:\ (x_1,\ldots,x_l)\to\Gamma(g), \tag{1.1}$$

where Γ is an l-dimensional irreducible representation of G. Take G to be a space group; for the present we suppress all indices relating to the representation. Then, the first step in modern renormalization group calculations[2] requires construction of the Hamiltonian for the system based on these order parameters. The Hamiltonian is taken as a power series in the set $(x_1,\ldots,x_l) \equiv (x)$

$$H(x) = H^0 + H^{(1)}(x) + H^{(2)}(x) + \cdots + H^{(s)} + \cdots, \tag{1.2}$$

where the terms in $H^{(s)}(x)$ will be a sum of the independent polynomial invariants, each homogeneous of degree s in the set (x). This so-called Landau–Ginzburg–Devonshire–Wilson expression is usually truncated at $s = 4$ or 6, but future work may require considering terms of higher degree, or the entire function $H(x)$.

Work in the framework of phenomenological (Landau) theory of phase transitions in crystals[3] requires knowledge of the basic invariants (integrity basis), and in some forms of the theory, one needs expansions of the Gibbs free energy $F(x)$ in a power series like (1.2).

A third example is the calculation of selection rules for high order optical processes such as multiphonon infrared or Raman processes.[4]

In the accompanying paper[1] two new algorithms were presented for computation of the Molien function $M(\Gamma, G; z)$. These forms will be used here. They are

$$M(\Gamma,G;z) = \frac{1}{|G|}\sum_g \exp\left(\sum_{m=1}^{\infty}\frac{z^m\chi(g^m)}{m}\right) \tag{1.3}$$

and

$$M(\Gamma,G;z) = \frac{1}{|G|}\prod_{J=1}^{p}[1-z\gamma_J(g)]^{-\delta_J}, \tag{1.4}$$

where $\chi(g)$ is the character of g in Γ, $\gamma_J(g)$ is the irreducible character of g in the cyclic group A of order p generated by g, and δ_J is the subduction coefficient of Γ of G upon γ_J of A.

2. THE CRYSTAL SPACE GROUP G

We shall assemble some formulas for the elements of a crystal space group which will be used later.

Let the translation group T of the crystal be generated by the fundamental translation $\mathbf{a}_i$ $(i = 1, 2, 3)$. Using periodic boundary conditions, we have

$$(e\,|\,\mathbf{a}_i)^{N_i} = (e\,|\,0),\quad i = 1, 2, 3. \tag{2.1}$$

Then $|\,T\,| = N_1N_2N_3$. The set of all lattice translations

$$(e\,|\,\mathbf{R}_L) = (e\,|\,l_i\mathbf{a}_i), \tag{2.2}$$

$$-N_i/2 \le l_i < N_i/2,\quad l_i \text{ integer}, \tag{2.3}$$

forms T. Sometimes below we denote a lattice vector by $\mathbf{R}$. It is simplest to consider a cubic system with $N_1 = N_2 = N_3 = N$; there is no loss of generality.

Call the space group G. The factor group G/T is isomorphic to a crystal point group P. Coset representatives in the decomposition of G with respect to T are

$$(\phi\,|\,\tau) \text{ with } \begin{cases} \tau = 0 \\ \text{or} \\ \tau = \text{a fractional translation}. \end{cases} \tag{2.4}$$

A general element g of G is

$$g=(\phi\,|\,\mathbf{t}),\quad \mathbf{t}=\boldsymbol{\tau}+\mathbf{R}_L\,. \tag{2.5}$$

If required, we may affix an index to ϕ and τ such as $\phi_\sigma,\boldsymbol{\tau}_\sigma$; other symbols will be defined as needed. Assume the rotational element ϕ has period p, so $\phi^p=e$. Then

$$g^p=(\phi\,|\,\mathbf{t})^p=(e\,|\,\mathbf{R}_p)\,, \tag{2.6}$$

where

$$\mathbf{R}_p=(\mathbf{t}+\phi\mathbf{t}+\cdots+\phi^{p-1}\mathbf{t})\equiv\{\phi\}^p\circ\mathbf{t} \tag{2.7}$$

and

$$\{\phi\}^p\equiv\sum_{\lambda=0}^{p-1}(\phi)^\lambda\,. \tag{2.8}$$

Thus

$$\mathbf{R}_p=p_i\mathbf{a}_i\quad(p_i\ \text{integer}) \tag{2.9}$$

is some lattice vector. The period of $\mathbf{R}_p$ [in (2.9)] will be (N/q), $q\geq 1$ with q an integer. We find it simpler, and no less general, to choose $q=1$ so that

$$g^{pN}=(e\,|\,0)\,. \tag{2.10}$$

Thus the period of an element g depends on the period of its rotational part in a simple fashion.

In (1.3) the element g^m occurs. Take for g the expression (2.5). Then

$$g^m=(\phi\,|\,\mathbf{t})^m=(\phi^m\,|\,\{\phi\}^m\mathbf{t})\,. \tag{2.11}$$

In order to take advantage of the period of ϕ, we write for m:

$$m=\mu+lp\quad\text{with}\ \mu=1,\ldots,p;\ l=0,\ldots,(N-1)\,. \tag{2.12}$$

Then

$$\phi^m=\phi^\mu \tag{2.13}$$

and

$$\begin{aligned}\{\phi\}^m\circ\mathbf{t}&=\{\phi\}^{lp}\circ\mathbf{t}+\{\phi\}^\mu\circ\mathbf{t}\\&=l\mathbf{R}_p+\{\phi\}^\mu\circ\mathbf{t}\\&=l\mathbf{R}_p+\{\phi\}^\mu\circ\mathbf{t}-\boldsymbol{\tau}_\mu+\boldsymbol{\tau}_\mu\\&=l\mathbf{R}_p+\mathbf{R}_\mu+\boldsymbol{\tau}_\mu\,,\end{aligned} \tag{2.14}$$

where

$$\mathbf{R}_\mu\equiv\{\phi\}^\mu\circ\mathbf{t}-\boldsymbol{\tau}_\mu \tag{2.15}$$

is some lattice vector, and $\boldsymbol{\tau}_\mu$ is the fractional (or zero) in the coset representative whose rotational part is ϕ^μ. Then

$$g^m=g^{\mu+lp}=(e\,|\,l\mathbf{R}_p)(e\,|\,\mathbf{R}_\mu)(\phi^\mu\,|\,\boldsymbol{\tau}_\mu)\,. \tag{2.16}$$

The element $g_\sigma^{-1}g^mg_\sigma$ will also be needed later. The element g is as before, in (2.5), while

$$g_\sigma\equiv(\phi_\sigma\,|\,\boldsymbol{\tau}_\sigma)\,. \tag{2.17}$$

Hence

$$g_\sigma^{-1}g^mg_\sigma=(e\,|\,\phi_\sigma^{-1}\circ(\mathbf{R}_\mu+l\mathbf{R}_p))((\widetilde{\phi^\mu})^\sigma\,|\,\mathbf{t}_{\bar\sigma\mu_\sigma}) \tag{2.18}$$

where

$$(\widetilde{\phi^\mu})^\sigma\equiv\phi_\sigma^{-1}\phi^\mu\phi_\sigma \tag{2.19}$$

and

$$\mathbf{t}_{\bar\sigma\mu_\sigma}\equiv(\phi_\sigma^{-1}\phi^\mu\boldsymbol{\tau}_\sigma+\phi_\sigma^{-1}\boldsymbol{\tau}_\mu-\phi_\sigma^{-1}\boldsymbol{\tau}_\sigma)\,. \tag{2.20}$$

The translation $\mathbf{t}_{\bar\sigma\mu_\sigma}$ is in general a sum of fractional plus lattice translation, and might be written

$$\mathbf{t}_{\bar\sigma\mu_\sigma}=\mathbf{R}'_{\mu_\sigma}+\boldsymbol{\tau}_{\mu_\sigma}\,, \tag{2.21}$$

where $\boldsymbol{\tau}_{\mu_\sigma}$ is the fractional associated with rotational part $(\widetilde{\phi^\mu})^\sigma$ in the coset representative, while $\mathbf{R}'_{\mu_\sigma}$ is defined by (2.21) and is a lattice vector. For later convenience we define another lattice vector as

$$\mathbf{R}_{\mu_\sigma}\equiv\phi_\sigma\circ\mathbf{R}'_{\mu_\sigma}\,. \tag{2.22}$$

3. REPRESENTATIONS

We shall be concerned with the Molien function for irreducible representation Γ of space group G. The construction of Γ of G and of $\chi(g)$ is well known,[5] and we shall simply cite some relevant expressions.

Let the irreducible representation Γ of the space group G be labeled $^*\mathbf{k}n$ where $^*\mathbf{k}=(\mathbf{k},\mathbf{k}_2,\ldots,\mathbf{k}_\sigma=\phi_\sigma\mathbf{k},\ldots,\mathbf{k}_s)$ and n refers to the allowable little group irreducible representation of $G(\mathbf{k})$. The canonical wave vector is $\mathbf{k}$, and $G(\mathbf{k})$ is the space group of the wave vector $\mathbf{k}$. Coset representatives in the canonical decomposition of G with respect to $G(\mathbf{k})$ are g_σ:

$$G=G(\mathbf{k})+g_2G(\mathbf{k})+\cdots+g_\sigma G(\mathbf{k})+\cdots+g_sG(\mathbf{k})\,. \tag{3.1}$$

We always reserve the index σ for such representatives. Then

$$\chi^{^*\mathbf{k}n}(g)=\sum_{\sigma=1}^{s}\dot\chi^{(\mathbf{k})(n)}(g_\sigma^{-1}gg_\sigma) \tag{3.2}$$

and $\dot\chi^{(\mathbf{k})(n)}(h)=0$ if $h\notin G(\mathbf{k})$; otherwise it has the value of the character of h in the nth allowable irreducible representation $D^{(\mathbf{k})(n)}$ of $G(\mathbf{k})$. Later we require the character of g^m.

Consider the Abelian subgroup A generated by group element g. Because we take the period of the elements in the space group to be pN the group A has pN distinct irreducible representations given by

$$\begin{aligned}&g\to\gamma_J(g)=\exp(2\pi in_J/pN)\,,\\&n_J=1,\ldots,pN\quad\text{and}\quad J=1,\ldots,pN\,.\end{aligned} \tag{3.3}$$

Below it will prove convenient to take $n_J=J$ and

$$\begin{aligned}&n_J=J=j+\nu N\,,\\&j=1,\ldots,N;\quad\nu=0,\ldots,p-1\,.\end{aligned} \tag{3.4}$$

Of course, we still retain the same total number of roots this way, but merely achieve a simpler labeling. Equivalently, j is an integer (mod N), and ν is an integer (mod p).

Recalling from the previous paper I(3.1)–(3.7), we see that the $\gamma_J(g)$ are the possible eigenvalues ω_J which occur in the expression for the Molien function.

4. MOLIEN FUNCTION FOR A SPACE GROUP

We are now in position to assemble all the previous results to calculate the Molien function for representation Γ (labeled $^*\mathbf{k}n$) of G. Our strategy must be to compute the subduction coefficients δ_J and the corresponding

eigenvalues γ_J. There is no difficulty in principle since we are merely evaluating the sum I(3.6):

$$\delta_J = \frac{1}{|A|}\sum_m \gamma_J(g^m)^*\chi(g^m)\,. \tag{4.1}$$

The sum goes over all elements in the cyclic subgroup A generated by g. However, because Γ is an induced representation and the structure of χ reflects this as in Eq. (3.2), the calculation is formally rather untidy. Here we shall simplify the formulas as much as seems possible at present.

We require that

$$\begin{aligned}\chi^{*\mathbf{k}n}(g^m) &= \sum_{\sigma=1}^{s} \dot\chi^{(\mathbf{k})(n)}(g_\sigma^{-1} g^m g_\sigma)\\ &= \sum_{\sigma=1}^{s} \dot\chi^{(\mathbf{k})(n)}((e\mid \phi_\sigma^{-1}\circ(l\mathbf{R}_p+\mathbf{R}_\mu))((\widetilde{\phi^\mu})^\sigma \mid \mathbf{t}_{\bar\sigma\mu_\sigma}))\\ &= \sum_\sigma \exp\{-[i\mathbf{k}_\sigma\circ(l\mathbf{R}_p+\mathbf{R}_\mu)]\}\,\dot\chi^{(\mathbf{k})(n)}((\widetilde{\phi^\mu})^\sigma \mid \mathbf{t}_{\bar\sigma\mu_\sigma})\\ &= \sum_\sigma \exp\{-i[l\mathbf{k}_\sigma\circ\mathbf{R}_p+\mathbf{k}_\sigma\circ\mathbf{R}_\mu+\mathbf{k}_\sigma\cdot\mathbf{R}_{\mu_\sigma}]\}\\ &\quad\times\dot\chi^{(\mathbf{k})(n)}((\widetilde{\phi^\mu})^\sigma\mid\tau_{\mu_\sigma})\,,\end{aligned} \tag{4.2}$$

where $\mathbf{k}_\sigma = \mathbf{k}\circ\phi_\sigma^{-1} = \phi_\sigma\circ\mathbf{k}$ is one of the members of $^*\mathbf{k}$, not equivalent to $\mathbf{k}$, since $(\phi_\sigma\mid\tau_\sigma)$ is not in $G(\mathbf{k})$. Note that in (4.2) the label l has been separated out, and we have taken $m=\mu+pl$ [see (2.12)] and used (2.18)–(2.21).

Then, returning to (4.1) and replacing the sum on m by appropriate sums on μ and l, we have as an intermediate step

$$\delta_J = \sum_{\sigma=1}^{s}\frac{1}{p}\sum_{\mu=1}^{p}\frac{1}{N}\sum_{l=0}^{N-1}\gamma_J(g^{\mu+lp})^*\dot\chi^{(\mathbf{k})(n)}(g_\sigma^{-1}g^{\mu+lp}g_\sigma)\,. \tag{4.3}$$

Now write, following (3.3) and (3.4), $m=\mu+pl$, and $n_J=j+\nu N$, so that

$$\begin{aligned}\gamma_J(g^m)^* &= \exp[-2\pi i(\mu+pl)(j+\nu N)/pN]\\ &= \exp[-2\pi i(\mu j/pN+\mu\nu/p+lj/N)]\,.\end{aligned} \tag{4.4}$$

It is already clear here that some indices are redundant, and we shall eliminate j shortly.

The first step is now to calculate the sum on l, and we isolate the relevant terms in (4.3) to obtain

$$\frac{1}{N}\sum_{l=0}^{N-1}\exp\left[-il\left(\frac{2\pi j}{N}+\mathbf{k}_\sigma\circ\mathbf{R}_p\right)\right] = \Delta\left(\frac{j}{N}+\frac{\mathbf{k}_\sigma\cdot\mathbf{R}_p}{2\pi}\right). \tag{4.5}$$

This will immediately be recognized as a "lattice delta"

$$\Delta(y) = \begin{cases}1 & \text{if } y=\text{integer}\,,\\ 0 & \text{otherwise}\,.\end{cases} \tag{4.6}$$

Taking account of the restrictions on j (an integer mod N) and $\mathbf{k}_\sigma\cdot\mathbf{R}_p = 2\pi(\text{integer})/N$, we find

$$2\pi j/N+\mathbf{k}_\sigma\cdot\mathbf{R}_p = 0 \tag{4.7}$$

as the only permitted value of j, which now allows us to write

$$\delta_J = \sum_{\sigma=1}^{s}\delta_{\nu_\sigma}\Delta\left(\frac{j}{N}+\frac{\mathbf{k}_\sigma\cdot\mathbf{R}_p}{2\pi}\right), \tag{4.8}$$

where we define

$$\begin{aligned}\delta_{\nu_\sigma} \equiv \frac{1}{p}\sum_{\mu=1}^{p}&\exp\left\{-i\left[\frac{2\pi\mu\nu}{p}+\mathbf{k}_\sigma\circ\left(\mathbf{R}_\mu-\frac{\mu}{p}\mathbf{R}_p+\mathbf{R}_{\mu_\sigma}\right)\right]\right\}\\ &\times\dot\chi^{(\mathbf{k})(n)}((\widetilde{\phi^\mu})^\sigma\mid\tau_{\mu_\sigma})\,.\end{aligned} \tag{4.9}$$

(δ_{ν_σ} should not be confused with a Kronecker delta $\delta_{\nu,\sigma}$ which we denote with comma between indices.) An interesting note about (4.9) is that there is no dependence on N [recall that $\nu=0,\ldots,(p-1)$].

Although it may appear cumbersome, Eqs. (4.8), (4.9) may be readily used in computations. It is assumed that the space group irreducible representations are known, so that, for each $^*\mathbf{k}n$, the set of coset representatives in $G(\mathbf{k})$ is known, as are the dotted characters $\dot\chi^{(\mathbf{k})(n)}$, and the set of elements g_σ. Computation of the translations $\mathbf{R}_p, \mathbf{R}_\mu, \mathbf{R}_{\mu_\sigma}$ is straightforward and so is the evaluation of the sum (4.9).

In the same fashion that the calculation of δ_J simplifies, so does the calculation of each term in (1.4), which is

$$\prod_J[1-z\gamma_J(g)]^{-\delta_J}\,, \tag{4.10}$$

and, substituting $J=j+\nu N$ as before, we find this becomes

$$\prod_{\nu=0}^{p-1}\prod_{j=1}^{N}\left[1-z\exp\left(\frac{2\pi i(j+\nu N)}{pN}\right)\right]^{-\delta_{j+\nu N}}. \tag{4.11}$$

However, as shown above

$$\delta_{j+\nu N} = \delta_J = \sum_{\sigma=1}^{s}\delta_{\nu_\sigma}\Delta\left(\frac{j}{N}+\mathbf{k}_\sigma\circ\frac{\mathbf{R}_p}{2\pi}\right)$$

for the representation under consideration. Now we may use the Δ again, to eliminate index j in favor of σ as in (4.7), and thus we change the range of the product appropriately, so that (4.11) becomes

$$\prod_{\nu=0}^{p-1}\prod_{\sigma=1}^{s}\{1-z\exp[i(2\pi\nu-\mathbf{k}_\sigma\circ\mathbf{R}_p)/P]\}^{-\delta_{\nu_\sigma}}. \tag{4.12}$$

In order to exhibit these formulas in more compact form, let us define the quantities

$$\omega_p \equiv \exp(2\pi i/p)\,, \tag{4.13}$$

$$\gamma_{\sigma\mu} \equiv \exp[-i\mathbf{k}_\sigma\circ(\mathbf{R}_\mu+\mathbf{R}_{\mu_\sigma})/p]\,. \tag{4.14}$$

Of course, these all refer to a single space group element g. Then

$$\delta_{\nu_\sigma} = \frac{1}{p}\sum_{\mu=1}^{p}(\omega_p^*)^{\mu\nu}(\gamma_{\sigma\mu}^*)^\mu(\gamma_{\sigma\mu})^p\,\dot\chi^{(\mathbf{k})(n)}((\widetilde{\phi^\mu})^\sigma\mid\tau_{\mu_\sigma}) \tag{4.15}$$

and

$$M(^*\mathbf{k}n,G;z) = \frac{1}{|G|}\sum_g\prod_{\nu=0}^{p-1}\prod_{\sigma=1}^{s}[1-z\gamma_{\sigma p}(\omega_p)^\nu]^{-\delta_{\nu_\sigma}}. \tag{4.16}$$

We continue with Eq. (4.16) and reduce it to a form in which is appears as a sum of partial Molien functions, each labeled by an index derived from an index of class of the point group p of space group G.

Let $P=G/T$. We call the kth conjugacy class of P $C_k(P)$, with elements

$$C_k(P):\ \phi_{ki},\quad i=1,\ldots,c_k, \tag{4.17}$$

where ϕ_{ki} is a rotation reflection.

Call the $(k\alpha)$th conjugacy class of G $C_{k\alpha}(G)$, with elements

$$t_{k\alpha i}g_{ki}\equiv(e\mid \mathbf{R}_{k\alpha i})(\phi_{ki}\mid \boldsymbol{\tau}_{ki}),\quad i=1,\ldots,c_k. \tag{4.18}$$

The index α will refer to those lattice translations associated with ϕ_{ki}, and $\boldsymbol{\tau}_{ki}$ is the canonical fractional; $t_{k\alpha i}$ is an element in T. To understand the structure of the conjugacy classes of G, form the conjugate of (4.18) with respect to general element $g=(\phi\mid \mathbf{t})$, where $\mathbf{t}=\boldsymbol{\tau}+\mathbf{R}$ with $\boldsymbol{\tau}$ the canonical fractional for ϕ and R a lattice vector as in Eq. (2.5). Then

$$g^{-1}(t_{k\alpha i}g_{ki})g=t'_{k\alpha j}g_{kj} \tag{4.19}$$

with

$$g_{kj}=(\phi^{-1}\phi_{ki}\phi\mid \boldsymbol{\tau}_{kj})\equiv(\phi_{kj}\mid \boldsymbol{\tau}_{kj}), \tag{4.20}$$

$$t'_{k\alpha j}\equiv(e\mid \phi^{-1}(\mathbf{R}_{k\alpha i}+\mathbf{R}')), \tag{4.21}$$

and

$$\phi^{-1}\mathbf{R}'\equiv(\phi_{kj}-e)\phi^{-1}\mathbf{t}+\phi^{-1}\boldsymbol{\tau}_{ki}-\boldsymbol{\tau}_{kj}. \tag{4.22}$$

When g runs through all elements in G, one will obtain on the right side of (4.19) the set of coset representatives g_{ki}, $i=1,\ldots,c_k$ whose rotational parts are all the members of the conjugacy class $C_k(P)$ of Eq. (4.17). Each such fixed rotational part (fixed k and i) will be associated with a set of translations, e.g., $t_{k\alpha i}$ which is a subset of T. For example if ϕ commutes with ϕ_{ki}, then the set (4.21) will occur with

$$\phi^{-1}\circ\mathbf{R}'=(\phi_{ki}-e)\phi^{-1}\mathbf{t}-\phi^{-1}\boldsymbol{\tau}_{ki}-\boldsymbol{\tau}_{kj}. \tag{4.23}$$

To label a class in G, we require an index k referring to set ϕ_{ki} and an index α referring to the particular subset of translations accompanying a specified representative such as g_{ki}. For fixed ki a subset $T_{k\alpha i}$ of translations occurs in class $C_{k\alpha}(G)$ accompanying g_{ki}. Then for fixed ki, the decomposition of T with respect to the class label α is disjoint

$$T=\sum_{\alpha}T_{k\alpha i},\quad T_{k\alpha i}\equiv\{t_{k\alpha i}\}. \tag{4.24}$$

Thus a single class $C_k(P)$ in P, with c_k elements gives rise to several classes $C_{k\alpha}(G)$ in G.

Let $f(t)$ be a function on the translations t of T. Then for fixed ki

$$\sum_{t\in T}f(t)=\sum_{\alpha}\left\{\sum_{t\in Tk\alpha i}f(t)\right\},\quad \text{fixed } ki. \tag{4.25}$$

Let χ be a character in a representation of G; then

$$\chi(t_{k\alpha i}g_{ki})=\chi(t_{k\alpha j}g_{kj}). \tag{4.26}$$

Since χ is a class function, it has the same value for all elements in a class, i.e., is independent of i,j since these refer to the same class.

TABLE I. Partial Molien function $\overline{m}(\Gamma,g_k;z)$ [text (4.29)] for irreducible representation $^*\Gamma n$ of O_h^3-$Pm3n$.[a] In the tables: $P_n=1-z^n$; $Q_n=1+z^n$; $R_4(n)=1+nz^2+z^4$; $S_2=1-z+z^2$; $T_2=1+z+z^2$.

c_k	g_k \ $^*kn$	$^*\Gamma1+$		$^*\Gamma1-$		$^*\Gamma2+$		$^*\Gamma2-$		$^*\Gamma3+$		$^*\Gamma3-$		$^*\Gamma4+$		$^*\Gamma4-$		$^*\Gamma5+$		$^*\Gamma5-$	
		$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$	$\chi(g_k)$	$\overline{m}$
1	$(E\mid 0)$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	2	$\frac{1}{P_1^2}$	2	$\frac{1}{P_1^2}$	3	$\frac{1}{P_1^3}$	3	$\frac{1}{P_1^3}$	3	$\frac{1}{P_1^3}$	3	$\frac{1}{P_1^3}$
8	$(C_3\mid 0)$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	−1	$\frac{1}{T_2}$	−1	$\frac{1}{T_2}$	0	$\frac{1}{P_3}$	0	$\frac{1}{P_3}$	0	$\frac{1}{P_3}$	0	$\frac{1}{P_3}$
3	$(C_4^2\mid 0)$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	2	$\frac{1}{P_1^2}$	2	$\frac{1}{P_1^2}$	−1	$\frac{1}{P_2Q_1}$	−1	$\frac{1}{P_2Q_1}$	−1	$\frac{1}{P_2Q_1}$	−1	$\frac{1}{P_2Q_1}$
1	$(I\mid 0)$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	2	$\frac{1}{P_1^2}$	−2	$\frac{1}{Q_1^2}$	3	$\frac{1}{P_1^3}$	−3	$\frac{1}{Q_1^3}$	3	$\frac{1}{P_1^3}$	−3	$\frac{1}{Q_1^3}$
8	$(IC_3\mid 0)$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	−1	$\frac{1}{T_2}$	1	$\frac{1}{S_2}$	0	$\frac{1}{P_3}$	0	$\frac{1}{Q_3}$	0	$\frac{1}{P_3}$	0	$\frac{1}{Q_3}$
3	$(\sigma\mid 0)$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	2	$\frac{1}{P_1^2}$	−2	$\frac{1}{Q_1^2}$	−1	$\frac{1}{P_2Q_1}$	1	$\frac{1}{P_1P_2}$	−1	$\frac{1}{P_2Q_1}$	1	$\frac{1}{P_1P_2}$
6	$(C_2\mid \tau)$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	−1	$\frac{1}{Q_1}$	0	$\frac{1}{P_2}$	0	$\frac{1}{P_2}$	1	$\frac{1}{P_1P_2}$	1	$\frac{1}{P_1P_2}$	−1	$\frac{1}{P_2Q_1}$	−1	$\frac{1}{P_2Q_1}$
6	$(C_4\mid \tau)$	1	$\frac{1}{P_1}$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	−1	$\frac{1}{Q_1}$	0	$\frac{1}{P_2}$	0	$\frac{1}{P_2}$	−1	$\frac{1}{Q_1Q_2}$	−1	$\frac{1}{Q_1Q_2}$	1	$\frac{1}{P_1Q_2}$	1	$\frac{1}{P_1Q_2}$
6	$(\sigma_v\mid \tau)$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	−1	$\frac{1}{Q_1}$	1	$\frac{1}{P_1}$	0	$\frac{1}{P_2}$	0	$\frac{1}{P_2}$	1	$\frac{1}{P_1P_2}$	−1	$\frac{1}{P_2Q_1}$	−1	$\frac{1}{P_2Q_1}$	1	$\frac{1}{P_1P_2}$
6	$(S_4\mid \tau)$	1	$\frac{1}{P_1}$	−1	$\frac{1}{Q_1}$	−1	$\frac{1}{Q_1}$	1	$\frac{1}{P_1}$	0	$\frac{1}{P_2}$	0	$\frac{1}{P_2}$	−1	$\frac{1}{Q_1Q_2}$	1	$\frac{1}{P_1Q_2}$	1	$\frac{1}{P_1Q_2}$	−1	$\frac{1}{Q_1Q_2}$

[a] Notation follows Ref. 7.

TABLE II. Partial Molien function $\bar{m}(\Gamma, g_k; z)$ for irreducible representations *$\mathbf{X}n$, and *$\mathbf{R}n$ of O_h^3-$Pm3n$.[a] See Table I heading.

c_k	g_k	*kn	*X1 and *X2	*X3 and *X4	*R1	*R2 and *R3	*R2⊕*R3	*R4
1	$(E\|0)$		$\frac{Q_2^3}{P_2^6}$	$\frac{Q_2^3}{P_2^6}$	$\frac{Q_2}{P_2^2}$	$\frac{Q_2}{P_2^2}$	$\frac{R_4(6)}{P_2^4}$	$\frac{Q_2 R_4(14)}{P_2^6}$
8	$(C_3\|0)$		$\frac{Q_6}{P_6^2}$	$\frac{Q_6}{P_6^2}$	$\frac{Q_2}{P_2^2}$	$\frac{P_4}{P_6}$	$\frac{R_4(3)}{R_4(1)^2}$	$\frac{Q_6}{P_6^2}$
3	$(C_4^2\|0)$		$\frac{Q_2}{P_2^4}$	$\frac{Q_2^3}{P_2^6}$	$\frac{Q_2}{P_2^2}$	$\frac{Q_2}{P_2^2}$	$\frac{R_4(6)}{P_2^4}$	$\frac{Q_2}{P_2^4}$
1	$(I\|0)$		$\frac{1}{P_2^3}$	$\frac{1}{P_2^3}$	$\frac{1}{P_2}$	$\frac{1}{P_2}$	$\frac{1}{P_2^2}$	$\frac{1}{P_2^3}$
8	$(IC_3\|0)$		$\frac{1}{P_6}$	$\frac{1}{P_6}$	$\frac{1}{P_2}$	$\frac{P_2^2}{P_6}$	$\frac{P_2}{P_6}$	$\frac{1}{P_6}$
3	$(\sigma\|0)$		$\frac{Q_2^2}{P_2^3}$	$\frac{1}{P_2^3}$	$\frac{1}{P_2}$	$\frac{1}{P_2}$	$\frac{1}{P_2^2}$	$\frac{1}{P_2^3}$
6	$(C_2\|\tau)$		$\frac{Q_4}{P_2 P_4^2}$	$\frac{Q_4}{P_2 P_4^2}$	$\frac{1}{P_2}$	$\frac{1}{P_2}$	$\frac{1}{P_2^2}$	$\frac{1}{P_2^3}$
6	$(C_4\|\tau)$		$\frac{1}{P_4 Q_2}$	$\frac{Q_4}{P_2 P_4^2}$	$\frac{1}{P_2}$	$\frac{1}{P_2}$	$\frac{1}{P_2^2}$	$\frac{1}{P_2 Q_2^2}$
6	$(\sigma_v\|\tau)$		$\frac{Q_4}{P_4^2 Q_2}$	$\frac{Q_4}{P_4^2 Q_2}$	$\frac{1}{Q_2}$	$\frac{1}{Q_2}$	$\frac{1}{Q_2^2}$	$\frac{1}{Q_2^3}$
6	$(S_4\|\tau)$		$\frac{1}{P_2 P_4}$	$\frac{Q_4}{P_4^2 Q_2}$	$\frac{1}{Q_2}$	$\frac{1}{Q_2}$	$\frac{1}{Q_2^2}$	$\frac{1}{P_2^2 Q_2}$

[a] Notation follows Ref. 7. Representations *X1, *X2, *X3, *X4 are called *$\mathbf{X}_2$, *$\mathbf{X}_4$, *$\mathbf{X}_1$, *$\mathbf{X}_3$, respectively, in Ref. 8.

Now let $F(\chi)$ be a function on the characters of χ of G, and let t be an arbitrary translation in T. Then

$$\sum_{t\in T} F(\chi(tg_{ki})) = \sum_{\alpha} \left\{ \sum_{t\in Tk\alpha i} F(\chi(tg_{ki})) \right\} = F_k . \tag{4.27}$$

The last step follows from (4.25) and (4.26), and now (4.27) can be used to rewrite the expression for the Molien function.

Returning to the expression I(1.5) for Molien function, we may write, for g an element of the space group G,

$$M(\Gamma, G; z) = (1/|G|)\sum_{g} m(\Gamma, g; z) . \tag{4.28}$$

Letting $g = tg_{ki}$

$$M(\Gamma, G; z) = (1/|P|)\sum_{k}\sum_{i} \cdot (1/|T|)\sum_{t} m(\Gamma, tg_{ki}; z)$$

$$= (1/|P|)\sum_{k} c_k \bar{m}(\Gamma, g_k; z) , \tag{4.29}$$

where

$$\bar{m}(\Gamma, g_k; z) \equiv (1/|T|)\sum_{t} m(\Gamma, tg_k; z) . \tag{4.30}$$

The last step in (4.29) follows owing to the use of (4.27).

The symbol g_k is a typical element or coset representative corresponding to class $C_k(P)$. The partial Molien function $\bar{m}(\Gamma, g_k; z)$ will be tabulated below. Then, instead of Eq. (4.16), we write

$$\bar{m}(\Gamma, g_k; z) = (1/|T|)\sum_{t} \prod_{\nu=1}^{p} \prod_{\sigma=1}^{s} [1 - z\gamma_{\sigma p}(\omega_p)^{\nu}]^{-\delta_{\nu\sigma}} . \tag{4.31}$$

All quantities refer to elements tg_k in G.

5. APPLICATION TO IRREDUCIBLE REPRESENTATIONS OF SPACE GROUP O_h^3-$Pm3n$

The space group O_h^3-$Pm3n$ is a nonsymmorphic space group of considerable current interest. Many superconducting compounds such as V_3Si, Nb_3Sn, crystallize in this space group in their room temperature forms.

In addition since O_h^3 is nonsymmorphic, the calculations of the Molien function in this case may illustrate typical problems which arise in such calculations for space groups. The results are also of interest since they may reveal new features peculiar to space groups.

We believe these to be the first reported calculated Molien functions for space group irreducible representations: these are the representations associated with *Rn and *$\mathbf{X}n$. Before dealing with these representations we shall discuss the representations *Γn.

TABLE III. Molien generating function for irreducible representations $*\Gamma n, *\mathbf{X}n, *\mathbf{R}n$ of O_h^3-$Pm3n$.

$*\Gamma 1+$	$\frac{1}{(1-z)}$
$*\Gamma 1-$	
$*\Gamma 2+$	$\frac{1}{(1-z^2)}$
$*\Gamma 2-$	
$*\Gamma 3+$	$\frac{1}{(1-z^2)(1-z^3)}$
$*\Gamma 3-$	$\frac{1}{(1-z^2)(1-z^6)}$
$*\Gamma 4+$	$\frac{1}{(1-z^2)(1-z^3)(1-z^4)}$
$*\Gamma 4-$	$\frac{1}{(1-z^2)(1-z^4)(1-z^6)}$
$*\Gamma 5-$	
$*\Gamma 5+$	$\frac{1+z^9}{(1-z^2)(1-z^4)(1-z^6)}$
$*\mathbf{X}1$	$\frac{1+z^4+3z^6+5z^8+z^{10}+z^{12}}{(1-z^2)(1-z^4)^3(1-z^6)^2}$
$*\mathbf{X}2$	
$*\mathbf{X}3$	$\frac{1+3z^4+7z^6+15z^8+13z^{10}+15z^{12}+8z^{14}+4z^{16}}{(1-z^2)(1-z^4)^3(1-z^6)^2}$
$*\mathbf{X}4$	
$*\mathbf{R}1$	$\frac{1}{(1-z^2)(1-z^4)}$
$*\mathbf{R}2$	$\frac{1}{(1-z^4)(1-z^6)}$
$*\mathbf{R}3$	
$*\mathbf{R}2\oplus*\mathbf{R}3$	$\frac{1+z^2+2z^4+4z^6+8z^8+4z^{10}+2z^{12}+z^{14}+z^{16}}{(1-z^4)^2(1-z^6)^2}$
$*\mathbf{R}4$	$\frac{1+2z^4+5z^6+11z^8+9z^{10}+11z^{12}+6z^{14}+3z^{16}}{(1-z^2)(1-z^4)^3(1-z^6)^2}$

The irreducible representations of this space group have been given by Gorzkowski[6] and in addition are listed in the standard compilation of Miller and Love.[7] We follow those authors' notations and we shall also indicate connection to another notation used by Gor'kov[8] (see Table II).

A. Representations $*\Gamma n$

At the center of the Brillouin zone, the group $G(\Gamma)$ is the entire space group and the factor group $G(\Gamma)/T$ is a group isomorphic to point group O_h since all translations in T map into the identity. All irreducible representations of O_h are well known as are the irreducible character systems.

In Table I we give: class multiplicity, coset representatives in $G(\Gamma)/T \sim O_h$ (first column), then in succeeding columns for each irreducible representation $*\Gamma n$ the irreducible character and the partial Molien function which arises by summing over all elements in T. The weighted sum of all the partial Molien functions as in Eq. (4.29) gives $M(*\Gamma n, O_h^3; z)$, and these are listed in Table III where the first seven rows refer to $*\Gamma n$. The notation of Miller and Love[7] is used. The result for the fundamental representation $*\Gamma 4-$ of O_h^3 agrees with a result of Meyer,[9] who gave Molien functions for the fundamental representations of all point groups.

B. Representations $*\mathbf{X}n$

At the point $\mathbf{X}(0, 0, \pi/a)$ in the Brillouin zone the allowable irreducible representations of $G(\mathbf{X})$ may be considered as ray representation of $G(\mathbf{X})/T \sim D_{4h} - 4/mmm$ with nontrivial factor systems.[10] All allowable $(\mathbf{X})(n)$ $(n = 1, \ldots, 4)$ are two dimensional, and the $*\mathbf{X}n$ are six dimensional. Some of these representations have been used in recent theories of properties of these systems.[8]

In Table II we give the functions $\bar{m}(*\mathbf{X}n, g_k; z)$ for $n = 1, \ldots, 4$. In Table II, lines 8 and 9, the full Molien function $M(*\mathbf{X}n, O_h^3; z)$ is given.

C. Representations $*\mathbf{R}n$

At the point $\mathbf{R}(\pi/a, \pi/a, \pi/a)$ in the Brillouin zone allowable irreducible representations of $G(\mathbf{R})$ are ray representations of $G(\mathbf{R})/T \sim O_h - m3m$. The factor system[10] is nontrivial. Allowable irreducible representations $*\mathbf{R}n$ are two dimensional, $n = 1, 2, 3$, and six dimensional, $n = 4$. The six-dimensional $*\mathbf{R}4$ is unusual; it was used in a recent theory of the electronic properties of these A-15 systems.[11] Representations $*\mathbf{R}2$ and $*\mathbf{R}3$ are time reverse, and it proves interesting to consider the four-dimensional physically irreducible representation (properly a corepresentation): $*\mathbf{R}2 \oplus *\mathbf{R}3$.

In Table II, last four columns, we list the $\bar{m}(*\mathbf{R}n, g_k; z)$. In Table III, last four rows, we give the five Molien functions $M(*\mathbf{R}n, O_h^3; z)$, including for the physically irreducible $*\mathbf{R}2 \oplus *\mathbf{R}3$.

6. RELATION TO UNITARY GROUPS GENERATED BY REFLECTIONS (GENERALIZED COXETER GROUPS)

Apart from its ultimate utility as a guide to constructing actual polynomial invariants, the Molien function may reveal important features of the structure of the matrix group associated with representation $*\mathbf{k}n$.

In particular for matrix groups Γ which are unitary groups[12] generated by reflections (u.g.g.r.) there are important dimensionality theorems which apply.[12,13] In case of a u.g.g.r. the Molien function can alwasy be written

$$M(\Gamma, G; z) = 1/(1-z^{d_1})(1-z^{d_2})\cdots(1-z^{d_m}). \tag{6.1}$$

The dimensionality theorems are *inter alia*

$$\prod_{i=1}^{m} d_i = |G|, \tag{6.2}$$

$$\sum_i (d_i - 1) = r, \tag{6.3}$$

where r is the number of pseudoreflections in the matrix group G. (These formulas apply somewhat more generally as well.[13])

Observing Table III, we see that Molien functions for representations $*\Gamma n$ for all n except 5+ are of the form (6.1). We easily verify that the relevant matrix groups

TABLE IV. Identification[a] of the matrix group *kn as an irreducible unitary group generated by reflections (u.g.g.r.).

*kn	Dimensionality of irr. u.g.g.r.	Symbol for the u.g.g.r.
*Γ1−, *Γ2±	1	[ ][2]
*Γ3+	2	$G(3,3,2)$
*Γ3−	2	$G(6,6,2)$
*Γ4+	3	$G(2,2,3)$
*Γ4−, *Γ5−	3	$G(2,1,3)$
*R1	2	$G(4,4,2)$
*R2, *R3	2	$G(6,3,2)$

[a]Table VII, p. 301 in Ref. 12.

are u.g.g.r. It is, of course, well known that the fundamental (Γ4−) representation of O_h is a matrix u.g.g.r., and line 6 of Table III agrees with Meyer.[9]

In Table IV we identify the irreducible matrix u.g.g.r. for representations *Γn±, $n=1,\ldots,5$, with the exception of *Γ5+ which as seen in Table III is not such a group. The notation of Shephard and Todd[12] has been used.

Again returning to Table III, we see that the irreducible *R1, *R2, and *R3 may be u.g.g.r.; we verified that they are. The groups so identified are given in Table IV, again following Shephard and Todd.

In this paper we limited ourselves to presenting the Molien function for space groups, and discussing some implications. Elsewhere we shall present and apply the integrity basis for the ring of invariants. Other aspects of the theory of invariants for space groups are presently being investigated.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor L. Michel for a discussion of our work. One of us (J. L. Birman) is also happy to acknowledge the stimulating effect of Professor Louis Michel's beautiful lecture[13] in rekindling interest in this subject; an earlier discussion with Professor Yu. Gufan was useful. We thank Professor R. Stanley for kindly sending us his preprint.[13]

APPENDIX: VERIFICATION OF DIMENSIONALITY FORMULA

From Eq. I(1.4) it is clear that the degree of the polynomial in the denominator is equal to l (which is the dimensionality of the representation Γ). To verify this property in (4.16) we have to show that

$$\sum_{\sigma=1}^{s}\sum_{\nu=0}^{p-1}\delta_{\nu\sigma}=sl_n\,, \tag{A1}$$

where sl_n is the dimensionality of the irreducible representation *kn of G and l_n is the dimensionality of the irreducible representation $D^{(\mathbf{k})(n)}$ of $G(\mathbf{k})$. Since ν appears in (4.15) only in $(\omega_p^*)^{\mu\nu}$, the sum over ν will give

$$\frac{1}{p}\sum_{\nu=0}^{p-1}(\omega_p^*)^{\mu\nu}=\Delta\left(\frac{\mu}{p}\right)=\delta_{\mu,p}\,, \tag{A2}$$

where we took into account the fact that μ is restricted to the values $1,\ldots,p$. Thus by (A2) we may eliminate the summation over μ, leaving only terms $\mu=p$. Then from the definition of $\gamma_{\sigma\mu}$ we see that $(\gamma_{\sigma p}{}^*)^{\mu}\,(\gamma_{\sigma\mu})^{p}\big|_{\mu=p}=1$, so that the left-hand side of (A1) reduces to:

$$\sum_{\sigma=1}^{s}\sum_{\nu=1}^{p}\delta_{\nu\sigma}=\sum_{\sigma=1}^{s}\dot{\chi}^{(\mathbf{k})(n)}((\widetilde{\phi^p})^{\sigma}\,|\,\tau_{p\sigma})\,. \tag{A3}$$

However, $(\widetilde{\phi^p})^{\sigma}=e$, by (2.20), and thus $\tau_{p\sigma}=0$. Finally since $\dot{\chi}^{(\mathbf{k})(n)}(e\,|\,0)=l_n$ (independent of σ), (A3) reduces to (A1).

*Supported in part by Grants NSF DMR76-20641 and CUNY-FRAP 11453.

[1]M. Jarić and J. L. Birman, J. Math. Phys. **18**, 1456 (1977). (preceding paper). Equations in that paper will be denoted I(1.6), etc.; also see references therein.

[2]D. Mukamel and S. Krinsky, Phys. Rev. B **13**, 5065 (1976); see also *ibid.* B **13**, 5078, 5086 (1976).

[3]J. L. Birman, "Symmetry Change in Continuous Phase Transitions in Crystals," Second International Colloquium on Group Theory Methods in Physics (Nijmegen, 1972), edited by A. Janner; also Phys. Rev. Lett. **17**, 1216 (1966); Y. U. Gufan and V. P. Sakhnenko, Zh. Eksp. Teor. Fiz. **69**, 1428 (1975) [Sov. Phys. JETP **42**, 728 (1976)].

[4]J. L. Birman, "Theory of Crystal Space Groups and Infra-Red and Raman Lattice Processes in Crystals," in *Handbuch der Physik* 25/2b (Springer-Verlag, Berlin, 1974).

[5]See Ref. 4, especially Eqs. (37.3).

[6]W. Gorzkowski, Phys. Stat. Solidi **3**, 910 (1963).

[7]S. C. Miller and W. F. Love, *Irreducible Representations of Space Group* (Pruett Press, Boulder, Col., 1967).

[8]L. P. Gor'kov, Zh. Eksp. Teor. Fiz. **65**, 1658 (1973) [Sov. Phys. JETP **38**, 830 (1974)].

[9]B. Meyer, Can. J. Math. **6**, 135 (1953).

[10]G. Bir and G. Picus, *Symmetry and Strain Induced Effects in Semiconductors* (Halsted, New York, 1974).

[11]T. K. Lee, J. L. Birman, and S. J. Williamson (to be published).

[12]G. C. Shephard and J. A. Todd, Can. J. Math. **6**, 274 (1953) and references therein.

[13]L. Michel, in Proc. Sixth Int'l. Colloq. on Group Theory in Physics, Montreal (1976), to be published, and R. P. Stanley "Relative Invariants of Finite Groups Generated by Pseudo-Reflections," MIT preprint (1976), unpublished.

On approximating the resolvent of a rotated Hamiltonian in the scattering region*

S. R. Singh

Department of Physics, University of Western Ontario, London, Canada
(Received 7 June 1976)

We show that, with certain restrictions on the potentials, the Bubnov–Galerkin method enables one to approximate the resolvent of a two-body rotated Hamiltonian in the strong sense. In the general N-body case we find that one might encounter some spurious singularities. However, we suggest a slight modification in the method that enables one to construct a sequence that converges strongly to the exact resolvent. This provides a procedure to approximate the scattering amplitudes for all N-body collisions.

1. INTRODUCTION

Let $H^N = H_0^N + V^N$ be the total Hamiltonian of an N-particle system with center of mass part removed. H_0^N and V^N are the kinetic energy and potential terms respectively. The scattering amplitudes $\tilde{T}^{\pm}$ for a particular reaction are given by $\tilde{T}^{\pm} = T_0 - T^{\pm}$ where $T_0 = -(\psi_f, V_i^N \psi_i)$, $T^{\pm} = \lim_{\epsilon \to 0} (\psi_f, V_f (E \pm i\epsilon - H^N)^{-1} \times V_i \psi_i)$ in the usual notation, where ψ_i and ψ_f are the initial and final free states respectively and V_i and V_f are the corresponding potential terms. In the following we consider the case of T^+. The case of T^- can be treated similarly.

If $V_f \psi_f = f$, $V_i \psi_i = g \in H_0^N = L^2(R^{3N-3})$, then T^+ is given by

$$T^+ = \lim_{\epsilon \to 0} (f, (E + i\epsilon - H^N)^{-1} g)_0. \tag{1}$$

$(\cdot, \cdot)_0$ denotes the scalar product in H_0^N. Thus the problem of computing the scattering amplitude reduces to computing the matrix elements of the resolvent, $R(\cdot)$, of H^N in the limit of the real line.

For a self-adjoint H^N, which usually is the case, and $\epsilon > 0$, $R(E + i\epsilon)$ can be approximated by a sequence $R_n(E + i\epsilon)$ of degenerate operators.[1] Thus a straightforward procedure to compute $(f, R(E + i0)g)_0$ is to set $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_k \underset{k \to \infty}{\to} 0$ and compute $(f, R(E + i\epsilon_k)g)_0 = \lim_{n \to \infty} (f, R_n(E + i\epsilon_k)g)_0$; and then seek $\lim_{k \to \infty} (f, R(E + i\epsilon_k)g)_0 = \lim_{k \to \infty} \lim_{n \to \infty} (f, R_n(E + i\epsilon_k) g)_0 = (f, R(E + i0)g)_0$.[2] Thus, in this method, a double limit must be taken in the prescribed order, which is a rather serious computational disadvantage. On the other hand, setting $\epsilon = 0$ in advance, not only poses a serious convergence problem, it also complicates the computational procedure.[3]

In order to circumvent these problems, the use of the rotated Hamiltonians has been made by several authors.[3] The method requires that the potentials satisfy certain analyticity conditions, but the computation procedure is significantly simplified. Although it is anticipated that the resolvent encountered may be approximated by a sequence of degenerate operators, to the best of our knowledge, no rigorous result to that effect is available. In the present article we study the convergence properties of the sequence obtained by the Bubnov–Galerkin method (BG).[4] In Sec. 2, following Simon,[5] we introduce a class of two-body potentials for which the physical information can be extracted from that on the rotated Hamiltonians, and collect some preliminary results. In Sec. 3, we show that for $N = 2$, BG yields a sequence that converges strongly to the desired limit. For $N > 2$ we have been unable to rule out the possibility that the algebraic equations encountered may be singular or nearly singular no matter how large n is. However we show that a nonsingular set of equations can be easily constructed from the singular ones which yields a sequence that converges strongly to the desired limit. These results enable one to approximate the scattering amplitudes by BG.

2. PRELIMINARIES

Let H_+^N be the completion of the domain $\mathcal{D}(\cdot)$, of H_0^N with respect to the norm $\|u\|_+ = \|(1 + H_0^N)^{1/2} u\|_0$ and H_-^N be the completion of H_0^N with respect to the norm $\|u\|_- = \|(1 + H_0^N)^{-1/2} u\|_0$, where $\|\cdot\|_0$ denotes the norm in H_0^N. Let $U(\theta)$ be the one-parameter group of dilatations on H_0^2 defined by $(U(\theta) f)(r) = e^{3\theta/2} f(re^{\theta})$. Simon[5] has introduced the class $\mathcal{J}_\alpha$ of two-body potentials $V: H_+^2 \to H_-^2$, defined by

(i) V is H_0^2 symmetric, i.e., $(\psi, V\psi)_0$ is real for all $\psi \in H_+^2$;

(ii) V is compact as a map from H_+^2 to H_-^2;

(iii) The family of bounded operators $V(\theta) = U(\theta) V U^{-1}(\theta)$ from H_+^2 to H_-^2 has an analytic continuation to a family of operators from H_+^2 to H_-^2 into the strip $|\mathrm{Im}\theta| \leq \alpha$.

It should be remarked here that some concrete examples of the potentials that satisfy these conditions are given in Ref. (5). Also the Hamiltonian, H_f^N, that represents an N-body system interacting via potentials in $\mathcal{J}_\alpha$ is defined by the method of forms,[5,6] which for a narrower class of potentials is identical to the usual definition of H^N. In the following, V^N will be assumed to have only the two-body forces which belong to $\mathcal{J}_\alpha$. Also, since no confusion will arise, the superscript N from $H_{0\pm}^N$ will be dropped. Furthermore, $\mathcal{B}(X, Y)$ and $\mathcal{C}(X, Y)$ will denote the classes of bounded and compact operators, respectively, from X to Y; and $\mathcal{B}(X, X), \mathcal{C}(X, X) = \mathcal{B}(X), \mathcal{C}(X)$.

Let $z = (z_1, \ldots, z_{N-1})$ and $D_\alpha^N = D_\alpha^1 X \cdots X D_\alpha^{N-1}$ where $D_\alpha^i = \{z_i : 0 \leq |z_i| < \infty, -\alpha \leq \arg(z_i) \leq \alpha\}$ and for any function ϕ, $\phi^\alpha = \phi(re^{i\alpha})$. The basis of the method of Ref. 4 is the following proposition:

Proposition 1: Let $f(z)$, $g(z)$ be analytic and square integrable on D_α^N and $V_{ij} \in \mathcal{J}_\alpha$, then

$$(f, R(E+i0)g)_0 = (f, (E+i0-H_f^N)^{-1}g)_0$$
$$= \exp[3i(N-1)\alpha]$$
$$\times (f^{-\alpha}, [E-\exp(-2i\alpha)H_f^N(\alpha)]^{-1}g^\alpha)_0,$$

where $H_f^N(\alpha): \mathcal{H}_+ \to \mathcal{H}_0$, is the Hamiltonian associated with the form $H^N(\alpha) = H_0^N + V^N(\alpha)$ with $V^N(\alpha) = \exp(2i\alpha) \times \sum V_{ij}^\alpha$.

Proposition 1 follows by replacing the integrals along the positive real line by that along the ray at an angle α in each variable. It is now clear that for $V_{ij} \in \mathcal{J}_\alpha$ the physical information is contained in $H_f^N(\alpha)$. Also, contrary to the case with H_f^N, E is at a positive distance from the spectrum $\sigma(\cdot)$ of $\exp(-2i\alpha)H_f^N(\alpha)$.[5]

In order to approximate $[E-\exp(-2i\alpha)H_f^N(\alpha)]^{-1}g^\alpha$ by BG, one chooses a linearly independent (l. i.) basis $\{\phi_i\}$ in $\mathcal{H}_+$ and solves the following set of equations:

$$\sum_{j=1}^n \alpha_j(\phi_i, (E\exp(2i\alpha) - H^N(\alpha))\phi_j)_0 = (\phi_i, \tilde{g}^\alpha)_0, \quad (2)$$
$$i = 1, \ldots n;$$

where $\tilde{g}^\alpha = \exp(2i\alpha)g^\alpha$. $\sum_{j=1}^n \alpha_j\phi_j$ is taken to be an approximation to $[E-\exp(-2i\alpha)H_f^N(\alpha)]^{-1}g^\alpha$. Before concentrating on the convergence properties of the method, we establish some auxiliary results. Unless otherwise stated, the results are valid for the general N-body case.

Lemma 1: Let $\{\phi_i\} \subset \mathcal{H}_+$ and $\tilde{g}^\alpha \in \mathcal{H}_0$. Then (2) is equivalent to

$$\sum_{j=1}^n \alpha_j(\phi_j, (1-\zeta B^+ + K)\phi_j)_+ = -(\phi_i, B\tilde{g}^\alpha)_+, \quad (3)$$

where $\zeta = (1+E\exp(2i\alpha))$, $B \in \mathcal{B}(\mathcal{H}_0, \mathcal{H}_+)$, $B^+ \in \mathcal{B}(\mathcal{H}_+)$, $K = \bar{B}V^N(\alpha) \in \mathcal{B}(\mathcal{H}_+)$, $\bar{B} \in \mathcal{B}(\mathcal{H}_-, \mathcal{H}_+)$.

Proof: Let $\bar{B}$ be the closure of $(1+H_0^N)^{-1}$ in $\mathcal{H}_- \times \mathcal{H}_+$. It is straightforward to check that $\bar{B} \in \mathcal{B}(\mathcal{H}_-, \mathcal{H}_+)$. Let B and B^+ be the (closed) restrictions of $\bar{B}$ to $\mathcal{H}_0$ and $\mathcal{H}_+$ respectively. Then $(\phi_i, \tilde{g}^\alpha)_0 = (\phi_i, (1+H_0^N)^{-1}\tilde{g}^\alpha)_+ = (\phi_i, B\tilde{g}^\alpha)_+$ and $(\phi_i, \phi_j)_0 = (\phi_i(1+H_0^N)^{-1}\phi_j)_+ = (\phi_i, B^+\phi_j)_+$.

The stated properties of B, B^+ are easy to check. Also since $(1+H_0^2)^{-1}V_{ij}(\alpha) \in \mathcal{B}(\mathcal{H}_+)$ we have that $(1+H_0^N)^{-1} \times V_{ij}(\alpha) \in \mathcal{B}(\mathcal{H}_+)$ and hence that $(1+H_0^N)^{-1}V^N(\alpha) \in \mathcal{B}(\mathcal{H}_+)$. From closability, this result extends to K.

Corollary 1: For $N=2$, K of Lemma 1 is in $\mathcal{C}(\mathcal{H}_+)$.

Proof: Follows from the fact that $V^2(\alpha) \in \mathcal{C}(\mathcal{H}_+, \mathcal{H}_-)$ and $\bar{B} \in \mathcal{B}(\mathcal{H}_-, \mathcal{H}_+)$.

Lemma 2: Let B, B^+, and K be as in Lemma 1. Then $B^{-1}(1-B^+ + K) = H_f^N(\alpha)$.

Proof: B is obviously invertible. Also $\mathcal{D}(B^{-1}(1-B^+ + K)) \subset \mathcal{H}_+$. Now for $u \in \mathcal{H}_0$, $v \in \mathcal{H}_+$ we have that

$$(u, B^{-1}(1-B^+ + K)v)_0 = (u, (H_0^N + V^N(\alpha))v)_0$$

and the result follows from the definition and uniqueness of $H_f^N(\alpha)$.[5-8]

We omit the trivial proof of the following result:

Lemma 3: Let $\{\phi_i\}$ be an orthonormal basis in $\mathcal{H}_+$. Then Eq. (3) is equivalent to

$$(1-\zeta B_n^+ + K_n)f_n = -P_n B\tilde{g}^\alpha, \quad (4)$$

where $f_n = \sum_{j=1}^n \alpha_j\phi_j$, $A_n = P_nAP_n$, and P_n is the orthoprojection on the n-dimensional subspace of $\mathcal{H}_+$ determined by $\{\phi_i\}$, $i=1$ to n.

3. THE CONVERGENCE OF BG

In the following theorem, the compactness of K is crucial. Hence the result is valid only for the two-body case (Corollary 1).

Theorem 1 ($N=2$): Let $\{\phi_i\}$ be a l. i. basis in $\mathcal{H}_+$ and E, $\alpha \neq 0$. Then for sufficiently large n, Eq. (2) has a unique solution and

$$\lim_{n\to\infty} \left\| \sum_{j=1}^n \alpha_j\phi_j - (E\exp(2i\alpha) - H_f^2))^{-1}\tilde{g}^\alpha \right\|_0 = 0.$$

Proof: From Lemma 1, (2) is equivalent to (3). Also, without loss of generality, we can assume $\{\phi_i\}$ to be an orthonormal basis in $\mathcal{H}_+$. Thus, from Lemma 3, it suffices to prove the result for Eq. (4). We divide the proof into several steps.

Step 1: $(1-\zeta B_n^+)^{-1} \underset{n\to\infty}{\to} (1-\zeta B^+)^{-1}$ strongly in $\mathcal{H}_+$.

Proof: For any $h \in \mathcal{H}_+$ we have that $\lim_{n\to\infty} \|(B_n^+ - B^+)h\|_+ = 0$. Also for E, $\alpha \neq 0$, $\mathrm{Im}\zeta \neq 0$ and hence $(1-\zeta B^+)^{-1} \in \mathcal{B}(\mathcal{H}_+)$, $(1-\zeta B_n^+)^{-1} \in \mathcal{B}(\mathcal{H}_+)$ and for each n,

$$\|(1-\zeta B_n^+)^{-1}\|_+ \leq |\zeta|/|\mathrm{Im}\zeta|.$$

It now follows that

$$\|[(1-\zeta B_n^+)^{-1} - (1-\zeta B^+)^{-1}]h\|_+$$
$$= \|(1-\zeta B_n^+)^{-1}\zeta(B_n^+ - B^+)(1-\zeta B^+)^{-1}h\|_+$$
$$\leq \frac{|\zeta|^2}{|\mathrm{Im}\zeta|} \|(B_n^+ - B^+)(1-\zeta B^+)^{-1}h\|_+$$
$$\underset{n\to\infty}{\to} 0,$$

for $(1-\zeta B^+)^{-1}h \in \mathcal{H}_+$.

Step 2: $(1+(1-\zeta B_n^+)K_n)^{-1} \underset{n\to\infty}{\to} (1+(1-\zeta B^+)^{-1}K)^{-1}$ uniformly in $\mathcal{H}_+$.

Proof: From the compactness of K and Step 1, it follows that

$$\lim_{n\to\infty} \|(1-\zeta B_n^+)^{-1}K_n - (1-\zeta B^+)^{-1}K\|_+ = 0.$$

Further, since

$$\rho = (1+(1-\zeta B^+)^{-1}K)^{-1}$$
$$= (1-\zeta B^+ + K)^{-1}(1-\zeta B^+)$$
$$= [B^{-1}(1-\zeta B^+ + K)]^{-1}B^{-1}(1-\zeta B^+)$$
$$= -[E\exp(2i\alpha) - H_f^2(\alpha)]^{-1}(B^{-1} - \zeta) \quad \text{(Lemma 2)},$$

ρ exists for α, $E \neq 0$.[5] Also as $(1-\zeta B^+)^{-1}K$ is compact, ρ must be bounded. The existence of the left member, for sufficiently large n, and uniform convergence to the right member, now follows from a standard argument (see, e.g., Theorem 1, Sec. 77, Ref. 4).

Step 3: $\lim_{n\to\infty} \|f_n - (E\exp(2i\alpha) - H_f^2(\alpha))^{-1}\tilde{g}^\alpha\|_0 = 0$.

Proof: Existence of $(1-\zeta B_n^+ + K_n)^{-1}$ for sufficiently large n is obvious from Step 2. Also for any $h \in \mathcal{H}_+$,

$$\lim_{n\to\infty} \|[(1-\xi B_n^+ + K_n)^{-1} - (1-\xi B^+ + K)^{-1}]h\|_+$$
$$= \lim_{n\to\infty} \|[(1+(1-\xi B_n^+)^{-1}K_n)^{-1}(1-\xi B_n^+)^{-1}$$
$$- (1+(1-\xi B^+)^{-1}K)^{-1}(1-\xi B^+)^{-1}]h\|_+$$
$$= 0 \quad \text{(Steps 1 and 2).}$$

Further, since for $\tilde{g}^\alpha \in \mathcal{H}_0$, $\lim_{n\to\infty}\|(P_nB - B)\tilde{g}^\alpha\|_+ = 0$, we have that $\lim_{n\to\infty}\|f_n - f\|_+ = 0$ where

$$f = -(1-\xi B^+ + K)^{-1}B\tilde{g}^\alpha$$
$$= (E\exp(2i\alpha) - H_f^2(\alpha))^{-1}\tilde{g}^\alpha \quad \text{(Lemma 2).}$$

The result now follows from the additional fact that

$$\|\cdot\|_0 \leq \|\cdot\|_+.$$

For the proof of Step 2 of Theorem 1 to go through, it is essential that K be compact. For $N \geq 3$ all we have is the boundedness of K. Therefore, the proof of Theorem 1 does not carry on to the general N-body case. Thus we are led to consider the case of a densely defined general sectorial form $H^N(\alpha)$, which reduces exactly to the form being considered presently.[8]

It is pertinent to remark here that some convergence properties of BG to approximate $(\xi - T)^{-1}$, where T is the m-sectorial operator associated with a densely defined closable sectorial form, have recently been studied.[8] The main result was that if ξ is at a positive distance from the numerical range of the form then the sequence obtained by BG converges strongly to $(\xi - T)^{-1}$. However, all we know about ξ $[= E\exp(2i\alpha)]$ in the present case is that it is at a positive distance from $\sigma(T)$ $(T = H_f^N(\alpha))$. It is clear from Theorem 1 that there is a class of sectorial operators for which the results of Ref. 8 can be strengthened to make only this assumption on ξ. However, in general this assumption is too weak to obtain convergence. For example, in case of a self-adjoint T, it is known that some points of the spectrum of the approximating operator could accumulate on ξ and/or move around in an arbitrarily small neighborhood of ξ (see, e.g., Ref. 9) and therefore a solution of Eq. (2) may not even exist. The situation in the present case gets only worse.

A modification in BG has been suggested in Ref. 10, which enables one to construct a sequence that approximates $(\xi - T)^{-1}$ in the strong sense where T is self-adjoint and ξ could be real but at a positive distance from $\sigma(T)$. In the following we generalize this method further to obtain a sequence that converges strongly to $(\xi - T)^{-1}$ where T is m-sectorial and $\text{dist}(\xi, \sigma(T)) > 0$.

Let $A = \xi B^+ - K$ and A_n be as in Lemma 3. We have that $A \in \mathcal{B}(\mathcal{H}_+)$ and, for E, $\alpha \neq 0$, $\text{dist}(1, \sigma(A)) > 0$, and therefore $(1-A)^{-1}$, $(1-A^\dagger)^{-1} \in \mathcal{B}(\mathcal{H}_+)$, where $\dagger$ denotes the adjoint. Let $L = A + A^\dagger - A^\dagger A$, $L_n = A_n + A_n^\dagger - A_n^\dagger A_n$. $(1-L)^{-1}$ is given by $(1-L)^{-1} = (1-A)^{-1}(1-A^\dagger)^{-1}$ and thus is in $\mathcal{B}(\mathcal{H}_+)$. It is now clear that $\text{dist}(1, \sigma(L)) > 0$. We also have that $L_n \to L$ strongly. Now let U_d and $U_{d+\delta}$ be two neighborhoods of 1 in the complex plane such that $U_d \subset U_{d+\delta}$, the boundary $\partial U_{d+\delta}$ of $U_{d+\delta}$ is a continuous closed contour of finite length l, dimeter of $U_d = d > 0$, $\text{dist}(1, \partial U_d) \geq d' > 0$, $\text{dist}(\sigma(A), \partial U_{d+\delta}) \geq \mu > 0$ and $\min(\text{dist}(U_d, \partial U_{d+\delta}), \text{dist}(0, U_{d+\delta})) \geq \delta > 0$. It is obviously possible to find such neighborhoods. Also let $(1+\epsilon_n^j)$, $j = 1$ to $m(n)$, be the points of $\sigma(L_n)$ such that $\{(1+\epsilon_n^j)\}_{j=1}^{m(n)} \subset U_d$, let p_n^j be the projection on the eigenspace of L_n corresponding to the eigenvalue $(1+\epsilon_n^j)$ and set $W_n = \sum_{j=1}^{m(n)}(1+\epsilon_n^j)p_n^j$, $L_n^- = L_n - W_n$. Owing to the self-adjointness of L_n, ϵ_n^j are all real, p_n^j are orthoprojections and thus W_n is self-adjoint and $\|W_n\|_+ < 1 + d$. Also it is clear that $(1 - L_n^-)^{-1}$ is a uniformly bounded sequence of operators from $\mathcal{H}_+$ to $\mathcal{H}_+$ with bound being $\leq 1/d'$, and, for $\eta \in \partial U_{d+\delta}$, $\|(\eta - W_n)^{-1}\|_+ \leq 1/\delta$. The following theorem (in fact a stronger result) was proven in the Appendix of Ref. 10:

Theorem 2: $(1 - L_n^-)^{-1} \underset{n\to\infty}{\to} (1-L)^{-1}$ strongly in $\mathcal{H}_+$.

However, we give a proof here for the following reasons. The present proof is relatively straightforward and makes certain points more transparent. Also, in contradistinction to Ref. 10, we do not make use of the spectral theorem here and thus there is a possibility of generalization to the case of nonnormal operators. We shall need

Theorem 3: $W_n \underset{n\to\infty}{\to} 0$ strongly in $\mathcal{H}_+$.

Proof: We divide the proof into different steps.

Step 1: For $\eta \in \partial U_{d+\delta}$ $[(\eta - W_n)^{-1}W_n - (\eta - L)^{-1}W_n] \underset{n\to\infty}{\to} 0$ weakly in $\mathcal{H}_+$.

Proof: For $u, v \in \mathcal{H}_+$ and $\eta \in \partial U_{d+\delta}$ we have that

$$|(u, (W_n - L)(\eta - W_n)^{-1}W_n v)_+|$$
$$= |(u(L_n - L)(\eta - W_n)^{-1}W_n v)_+| \quad (\text{for } W_n^2 = L_n W_n)$$
$$\leq \|(L_n - L)u\|_+ \|(\eta - W_n)^{-1}W_n v\|_+.$$

The first term goes to zero for $L_n \to L$ strongly and the second term is uniformly bounded by $(1+d)/\delta$:

$$\Delta(\eta) = |(u, [(\eta - W_n)^{-1} - (\eta - L)^{-1}]W_n v)_+|$$
$$= |(u, (\eta - L)^{-1}(W_n - L)(\eta - W_n)^{-1}W_n v)_+| \underset{n\to\infty}{\to} 0.$$

Step 2: The convergence is Step 1 is uniform with respect to $\eta \in \partial U_{d+\delta}$.

Proof: Since $\partial U_{d+\delta}$ is of finite length l, the result will follow from the continuity of $\Delta(\eta)$ in η. But for $\eta, \eta' \in \partial U_{d+\delta}$

$$|\Delta(\eta) - \Delta(\eta')|$$
$$= |\,|(u, [\eta - W_n)^{-1} - (\eta - L)^{-1}]W_n v)_+|$$
$$- |(u, [(\eta' - W_n)^{-1} - (\eta' - L)^{-1}]W_n v)_+|\,|$$
$$\leq |(u, [(\eta - W_n)^{-1} - (\eta' - W_n)^{-1}]W_n v)_+$$
$$- (u, [(\eta - L)^{-1} - (\eta' - L)^{-1}]W_n v)_+|$$
$$\leq |\eta - \eta'|\,[\,|(u, (\eta - W_n)^{-1}(\eta' - W_n)^{-1}W_n v)_+|$$
$$+ |(u, (\eta - L)^{-1}(\eta' - L)^{-1}W_n v)_+|\,]$$
$$\leq |\eta - \eta'|\ \|u\|_+ \|v\|_+ [1/\delta^2 + 1/\mu^2]\|W_+\|,$$

and the result follows from the fact that δ, $\mu > 0$ and W_n is uniformly bounded.

Step 3: $W_n \underset{n\to\infty}{\to} 0$ weakly in $\mathcal{H}_+$.

Proof: For $u, v \in \mathcal{H}_+$ let

$$I = \frac{1}{2\pi i}\int_{\partial U_{d+\delta}} d\eta (u, [(\eta - W_n)^{-1} - (\eta - L)^{-1}]W_n v)_+.$$

Since $(\zeta - L)^{-1}$ is analytic for $\zeta \in U_{d+\delta}$, the second term in I is equal to zero. The first term is equal to $(u, W_n v)_+$ as can be seen by distorting the contour into a circle at infinity and noticing that the residue at the origin is zero. Thus $I = (u, W_n v)_+$. Now

$$|I| \leq l \sup_{\eta \in \partial U_{d+\delta}} |(u, [(\eta - W_n)^{-1} - (\eta - L)^{-1}] W_n v)_+| \xrightarrow[n \to \infty]{} 0$$

from Steps 1 and 2.

Step 4: $W_n \xrightarrow[n \to \infty]{} 0$ strongly in H_+.

Proof: For $u \in H_+$

$$\| W_n u \|_+^2 = (u, W_n^2 u)_+ = (u, L_n W_n u) \xrightarrow[n \to \infty]{} 0$$

from Step 3 and the fact that $L_n \to L$ strongly.

Proof of Theorem 2: For $u \in H_+$ we have that

$$\| [(1 - L_n^-)^{-1} - (1 - L)^{-1}] u \|_+$$
$$= \| (1 - L_n^-)^{-1} (L_n - W_n - L)(1 - L)^{-1} u \|_+$$
$$\leq \| (1 - L_n^-)^{-1} \|_+ \| (L_n - W_n - L)(1 - L)^{-1} u \|_+ \xrightarrow[n \to \infty]{} 0$$

for $\|(1 - L_n^-)^{-1}\|_+ \leq 1/d'$, $L_n \to L$ strongly, and $W_n \to 0$ strongly (Theorem 3).

In the following theorem we obtain the present method to approximate $(1 - A)^{-1}$.

Theorem 4: $(1 - L_n^-)^{-1} (1 - A_n^\dagger) \xrightarrow[n \to \infty]{} (1 - A)^{-1}$ strongly in H_+.

Proof: Strong limit of the uniformly bounded sequence $(1 - L_n^-)^{-1}$ is $(1 - L)^{-1}$ and that of $(1 - A_n^\dagger)$ is $(1 - A^\dagger)$. Thus

$$(1 - L_n^-)^{-1}(1 - A_n^\dagger) \xrightarrow[n \to \infty]{} (1 - L)^{-1}(1 - A^\dagger) = (1 - A)^{-1}$$

strongly in H_+.

It is clear that $(1 - A_n^\dagger)$ in Theorem 4 could be replaced by $(1 - A^\dagger)$ without affecting the result. However, this result provides a convenient computational procedure to approximate $(1 - A)^{-1}$, which we explain in the following. Equation (2) is a matrix equation

$$M_n \alpha_n = \beta_n , \tag{5}$$

where M_n is an nxn matrix and β_n is an n-dimensional column vector. A vector solution α_n of Eq. (5) is sought. The result of Theorem 4 requires that one multiply Eq. (5) from left by $M_n^\dagger$ and obtain

$$M_n^\dagger M_n \alpha_n = M_n^\dagger \beta_n . \tag{6}$$

If this resulting set (6) remains singular or near singular with increasing n, we drop the offending part x_n. The resulting equation

$$(M_n^\dagger M_n - x_n) \alpha_n' = M_n^\dagger \beta_n \tag{7}$$

then has a unique solution α_n' which enables us to construct $f_n = \sum_{j=1}^n \alpha_j' \phi_j$. It is clear, from Lemmas 1, 2, 3, Theorem 4 and the additional fact that $P_n \to 1$ strongly in H_+, that $f_n' \xrightarrow[n \to \infty]{} [E \exp(2i\alpha) - H_f^N(\alpha)]^{-1} g^{\sim \alpha}$ strongly in H_+. Convergence in H_0 follows from that in H_+. We state this result formally as

Theorem 5: $(N \geq 3)$: Let ϕ_i be an l. i. basis in H_+ and E, $\alpha \neq 0$. Also let f_n' be the solution of the "reduced" set of equations obtained from Eq. (2). Then for $g^{\sim \alpha} \in H_0$ we have that

$$\lim_{n \to \infty} \| f_n' - [E \exp(2i\alpha) - H_f^N(\alpha)]^{-1} g^{\sim \alpha} \|_0 = 0.$$

4. CONCLUDING REMARKS

For $N = 2$ we have that the sequence obtained by the original BG converges strongly to the desired limit. This result also enables one to investigate the spectrum of a two body rotated Hamiltonian. For $N \geq 3$, we have the convergence of the sequence obtained by the modified BG. Computationally the modification is a straightforward procedure; analytically the result is achieved in two steps. The first step is to reduce the problem involving a non-self-adjoint operator to the problem involving a self-adjoint operator. The second step is to drop out any possible singularities. The latter modification in BG has been studied in Ref. 10 for a more general class of self-adjoint operators in that no assumption of boundedness, above or below, was made. The technique used there relied on the spectral theorem for the self-adjoint operators. The present method is easily extendable to the general class of self-adjoint and normal operators, and possibly also for some of the operators that may not have a spectral decomposition.

ACKNOWLEDGMENT

The author is grateful to Professor J. Nuttall for several helpful discussions and his hospitality. Thanks are also due to Dr. T. Ghemirés and Dr. F. Chatelin for pointing out an error that occurred in the original manuscript.

*Supported, in part, by the National Research Council of Canada and the Center for Interdisciplinary Studies in Chemical Physics, U.W.O.

[1] M.H. Stone, *Linear Transformations in Hilbert Space* (Am. Math. Soc., New York, 1932), Theorem 5.3.

[2] F.A. McDonald and J. Nuttall, Phys. Rev. A 4, 1821 (1971), and references therein.

[3] J. Nuttall and H.L. Cohen, Phys. Rev. 188, 1542 (1969), F.A. McDonald and J. Nuttall, Phys. Rev. C 6, 121 (1972), and references therein.

[4] S.G. Mikhlin, *Variational Methods in Mathematical Physics* (Pergamon, New York, 1964), Chap. IX.

[5] B. Simon, Commun. Math. Phys. 27, 1 (1972).

[6] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U.P., Princeton, N.J., 1971), Chaps. 2, 7.

[7] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Theorem 2.1, Chap. 6.

[8] S.R. Singh, Pac. J. Math. 65, 218 (1976). The argument used to conclude uniform boundedness of $[1 - T_n(z)]^{-1}$ in Theorem 1 is incorrect. However, it is easy to check that, in this case, $W(T_n(z)) \subset \overline{W(T(z))}$ and thus $\|[1 - T_n(z)]^{-1}\|_+ \leq 1/d'$.

[9] Yu.V. Vorobyev, *Method of Moments in Applid Mathematics* (Gordon and Breach, New York, 1965), Chap. 3.

[10] S.R. Singh and A.D. Stauffer, Nuovo Cimento B 22, 139 (1974).

Error bounds on some approximate solutions of the Fredholm equations and applications to potential scattering and bound state problems*

S. R. Singh

Department of Physics, University of Western Ontario, London 72, Canada

G. Turchetti†

Centre for Interdisciplinary Studies in Chemical Physics, University of Western Ontario, London 72, Canada
(Received 29 November 1976; revised manuscript received 16 February 1977)

One may produce sequences that converge to the eigenvalues of a compact operator and the solution of an inhomogeneous equation involving this operator by the Bubnov–Galerkin method. We produce easily computable, converging and accurate error bounds on these approximations, provided that the operator belongs to a fairly wide subclass of the compact operators. These bounds in turn enable one to compute error bounds, having the same properties, on the approximations to the scattering parameters and the bound state energies obtained by the Schwinger variational method for some two-body interactions. This provides a method of obtaining upper and lower bounds to these quantities.

1. INTRODUCTION

Many problems in physics can be reduced to solving a Fredholm type equation involving a compact operator. Several approximation schemes are known to solve a Fredholm equation, for example, the collocation method,[1] the collocation-variational method,[2] and the Bubnov–Galerkin method[3] (BG). The latter is closely related to the variational methods that are frequently used in physics.[4-6] Also the method of moments,[7] which proves quite useful in certain situations, is a special case of BG. Although it is known that BG provides sequences that converge to the exact values, in both the homogeneous and the inhomogeneous case, they may not have any bound property unless the operator involved belongs to a class that is too restrictive for the practical purposes.[8] Thus it is desirable to have an error bound on an approximate solution.

For an error bound to be practically useful, it should be convergent to zero as the order is increased, computable with reasonable amount of labor and should be small at moderately large orders. Most of the error bounds available in literature lack some of these properties.[1] In the present note we derive bounds on the solutions obtained by BG which have these characteristics, provided that the operator is in $\mathcal{B}_l(H)$, where $\mathcal{B}_l(H)$ is the class of linear operators from H to H such that $\|K\|_l = [\mathrm{Tr}(K^\dagger K)^{l/2}]^{1/l} < \infty$ and H is a separable Hilbert space. This class is sufficiently large to include a fairly large number of the cases of interest: For example, the two-body problem for a reasonably large class of potentials reduces to studying an operator $K \in \mathcal{B}_2(H)$, $H = L^2(R^3)$.[9] We give explicit error formulas for the approximations to the eigenvalues of $K \in \mathcal{B}_l(H)$ and to the solution of an inhomogeneous equation involving K. These results with some results of Refs. 4–6, enable one to compute error bounds with the same properties, on the approximate values of the eigenvalues and the scattering parameters obtained by the Schwinger variational method, for the two-body interactions compatible with Ref. 9. This obviously provides a practically satisfactory method to compute converging and accurate upper and lower bounds on these quantities, which have been difficult to obtain, and/or lack these properties.[10] The accuracy of the bounds is then numerically tested for two explicit examples, one of them being the one arising in the S-wave bound state problem involving an exponential potential.

2. THE ERROR BOUNDS

Let P_N be the orthoprojection on $\mathcal{E}_N$, where $\mathcal{E}_N$ is the linear span of some linearly independent set $\{\phi_j\}$, $j = 1$ to N; $\mathcal{E}_N \xrightarrow[N\to\infty]{} H$ and $K_N = P_N K P_N$. Then it is well known[3] that BG approximation λ_j^N to an eigenvalue λ_j of K, $j \leq N$, is an eigenvalue of K_N. Also BG approximation, Y_N, to the solution Y of the equation

$$(1 - \lambda K)Y = g \tag{1}$$

is the solution of

$$(1 - \lambda K_N)Y_N = P_N g . \tag{2}$$

Furthermore, if p_j^N, p_j are the eigenprojections of K_N, K corresponding to the eigenvalues λ_j^N, λ_j and K is compact, then[3] $\lambda_j^N \xrightarrow[N\to\infty]{} \lambda_j$, $\|p_j^N - p_j\| \xrightarrow[N\to\infty]{} 0$, $\|Y_N - Y\| \xrightarrow[N\to\infty]{} 0$, where $\|\cdot\|$ denotes the norm. The scalar product in H will be denoted by $\langle \cdot \,|\, \cdot \rangle$.

Theorem 1: Let $K \in \mathcal{B}_l(H)$ for some $l \geq 1$, and $K_N, p_j^N, \lambda_j, \lambda_j^N$ be as above. Then

$$|\lambda_j - \lambda_j^N| \leq \tfrac{1}{2}\alpha_N[1 - (1 - 4\gamma_n/\alpha_N^2)^{1/2}] = \delta_j^N \xrightarrow[N\to\infty]{} 0 \tag{3}$$

for each j such that

$$\delta_j^N + \|(1 - p_j^N)(K - K_N)(1 - p_j^N)\|_L < d_N , \quad L \geq l, \tag{4}$$

where

$$\gamma_N = \|p_j^N K (1 - p_j^N)\|_L^2 \xrightarrow[N\to\infty]{} 0 ,$$

$$\alpha_N = d_N - \|(1 - p_j^N)(K - K_N)(1 - p_j^N)\|_L$$

and d_N^2 is the smallest eigenvalue of $A_N^\dagger A_N$ with $A_N = [\lambda_j^N - (1 - p_j^N)K_N(1 - p_j^N)]$. [Given j, condition (4) is always satisfied for large enough $N(j)$.]

Proof: For any j, the eigenvalue equation for K reads

$$(\lambda_j - K)p_j = 0 \tag{5}$$

implying that

$$0 = p_j^N(\lambda_j - K)[p_j^N + (1 - p_j^N)]p_j$$
$$= (\lambda_j - p_j^N K p_j^N)p_j^N p_j - p_j^N K(1 - p_j^N)p_j .$$

Since $p_j^N K p_j^N = \lambda_j^N p_j^N$, we have that

$$(\lambda_j - \lambda_j^N)p_j^N p_j = p_j^N K(1 - p_j^N)p_j . \tag{6}$$

As $\| p_j^N - p_j \| \xrightarrow[N\to\infty]{} 0$, for large enough N, $p_j^N p_j \neq 0$, and hence it follows from (6) that

$$|\lambda_j - \lambda_j^N| \leq \| p_j^N K(1 - p_j^N)\| \, \| (1 - p_j^N)p_j \| / \| p_j^N p_j \|$$
$$\leq \| p_j^N K(1 - p_j^N) \|_L \| (1 - p_j^N)p_j \| / \| p_j^N p_j \| . \tag{7}$$

Similarly, starting from (5), it is straightforward to derive that $(1 - p_j^N)p_j$ is the unique solution of

$$[\lambda_j - (1 - p_j^N)K(1 - p_j^N)](1 - p_j^N)p_j = (1 - p_j^N)K p_j^N p_j ,$$

which yields

$$(1 - p_j^N)p_j = [\lambda_j^N - (1 - p_j^N)K_N(1 - p_j^N)]^{-1}[1 + T_N]^{-1} \times (1 - p_j^N)K p_j^N p_j , \tag{8}$$

where

$$T_N = [(\lambda_j - \lambda_j^N) - (1 - p_j^N)(K - K_N)(1 - p_j^N)] \times [\lambda_j^N - (1 - p_j^N)K_N(1 - p_j^N)]^{-1} .$$

Existence of $(1 + T_N)^{-1}$ is used to derive (8) which is ensured by the fact that, owing to the compactness of K, $\| T_N \| \xrightarrow[N\to\infty]{} 0$ and N may be chosen so large that $\| T_N \| < 1$. Also, for large enough N,

$$\| (1 + T_N)^{-1} \| \leq (1 - \| T_N \|)^{-1}$$
$$\leq \{1 - d_N^{-1}[\, |\lambda_j - \lambda_j^N| + \| (1 - p_j^N)(K - K_N)(1 - p_j^N) \|_L \,]\}^{-1}$$
$$= D_N .$$

From this and (8) it follows that

$$\| (1 - p_j^N)p_j \| \leq \| (\lambda_j^N - (1 - p_j^N)K_N(1 - p_j^N))^{-1} \| \times D_n \| (1 - p_j^N)K p_j^N \|_L \| p_j^N p_j \|$$
$$= d_N^{-1} D_N \| (1 - p_j^N)K p_j^N \|_L \| p_j^N p_j \| . \tag{9}$$

Finally, from (7) and (9) we have that

$$|\lambda_j - \lambda_j^N| \leq \gamma_N / (\alpha_N - |\lambda_j - \lambda_j^N|) , \tag{10}$$

with α_N and γ_N being as defined in the theorem. The optimum solution of the inequality (10) is given by (3). Condition (4) ensures that $\| T_N \| < 1$, $p_j^N p_j \neq 0$, and is satisfied, for large enough N, for the reasons that $\delta_j^N \xrightarrow[N\to\infty]{} 0$ and $\| K - K_N \|_L \xrightarrow[N\to\infty]{} 0$. QED

It is obvious, from the proof, that the inequality of Theorem 1 is valid if $\| \cdot \|_L$ is replaced by $\| \cdot \|$ or any other upper bound to $\| \cdot \|$, and K is assumed only to be compact, but then the bounds may not be computable. Computability was the motivation that we majorized $\| \cdot \|$ by $\| \cdot \|_L$ whenever $\| \cdot \|$ may not be computable. Also, the condition (4) could be replaced by slightly milder restriction, for the above reason and also by replacing δ_j^N by $|\lambda_j - \lambda_j^N|$. But, in contradistinction to the latter, the present condition is easy to check at each stage of computation.

It is also worth pointing out that if K is symmetric and λ_j is positive (negative) then λ_j^N is $\leq$ ($\geq$) λ_j and hence the bounds are given by

$$\lambda_j^N \lessgtr \lambda_j \lessgtr \lambda_j^N \pm \delta_j^N .$$

In the case of a nonsymmetric K this type of inequality is not possible, but using the present result one has that

$$|\lambda_j^N - \delta_j^N| \leq |\lambda_j| \leq |\lambda_j^N + \delta_j^N| .$$

Also δ_j^N could be taken to be the maximum error in the real and the imaginary parts of λ_j^N.

In the following theorem we consider the case of an inhomogeneous equation.

Theorem 2: Let $K \in \mathcal{B}_l(H)$, $l \geq 1$, and let Y, Y_N be as in Eqs. (1) and (2). Then

$$\| Y - Y_N \| \leq (1 - |\lambda| \eta_N)^{-1} [\, \| (1 - P_N)g \| + |\lambda| \, \| (1 - P_N)K P_N Y_N \| \,]$$
$$= \epsilon_N \xrightarrow[N\to\infty]{} 0 , \tag{11}$$

where

$$\eta_N = \| (1 - \lambda K_N)^{-1} P_N K(1 - P_N) \|_L + \| (1 - P_N)K \|_L \xrightarrow[N\to\infty]{} 0, \quad L \geq l , \tag{12}$$

provided that N is large enough so that $|\lambda| \eta_N < 1$.

Proof: Letting $Y' = Y - Y_N$, from (1) and (2), and the fact that $Y_N = P_N Y_N$, it follows that

$$(1 - \lambda K)Y' = (1 - P_N)g + \lambda(1 - P_N)K P_N Y_N ,$$

yielding

$$Y' = (1 - \lambda K)^{-1}[(1 - P_N)g + \lambda(1 - P_N)K P_N Y_N]$$
$$= (1 - \lambda T_N)^{-1}[(1 - P_N)g + \lambda(1 - P_N)K P_N Y_N] , \tag{13}$$

where

$$T_N = (1 - \lambda K_N)^{-1} P_N K(1 - P_N) + (1 - P_N)K$$

and we have used the relation $(1 - \lambda K_N)^{-1} = (1 - \lambda K_N)^{-1} P_N + (1 - P_N)$. $\| Y' \|$ is clearly majorized by ϵ_N of the theorem if $|\lambda| \eta_N < 1$, which obviously is the case for large enough N. QED

On majoring $\| \cdot \|$, by $\| \cdot \|_L$, the same remark applies as following Theorem 1.

In the scattering problems one is usually interested in computing expressions like $\langle g | y \rangle$ or $\langle f | y \rangle$, and BG approximation to $\langle g | y \rangle$, $\langle f | y \rangle$ is $\langle g | Y_N \rangle$, $\langle f | Y_N \rangle$.[4,6] It is clear that in either of the cases an error bound is computable in terms of ϵ_N. If $\| \lambda K \| < 1$, λ real, and K is symmetric, then it is known that $\langle g | Y_N \rangle \leq \langle g | Y \rangle$.[4] It follows in that case that $\langle g | Y_N \rangle \leq \langle g | Y \rangle \leq \langle g | Y_N \rangle + \| g \| \epsilon_N$. Although for a symmetric K encountered in potential scattering, the Hellmann–Feynman theorem enables one to prove that $\langle g | Y_N \rangle \leq \langle g | Y \rangle$ for sufficiently large N even when $\| \lambda K \| > 1$,[11] for an arbitrary K the result may not be true. It is clear that the present result allows one to compute bounds on $|\langle f | y \rangle|$, for any f, involving an ar-

bitrary $K \in \mathcal{B}_1(H)$. When $\langle \cdot | Y_N \rangle$ itself is not a bound to $\langle \cdot | y \rangle$, then the following closer approximation to y may be used to improve the bound of Theorem 2.

Let $\bar{Y}_N$ be the solution of

$$(1 - \lambda K_N)\bar{Y}_N = g; \tag{14}$$

$\bar{Y}_N$ is clearly given by $\bar{Y}_N = Y_N + (1 - P_N)g$.

Corollary 1: Let $\bar{Y}_N$ be as in (14). Then

$$\begin{aligned}\| Y - \bar{Y}_N \| &\leq |\lambda| (1 - |\lambda| \eta_N)^{-1} [\| (1 - \lambda K_N)^{-1} P_N K (1 - P_N) g \| \\ &\quad + \| (1 - P_N) K (1 - P_N) g \| + \| (1 - P_N) K P_N Y_N \|] \\ &= \bar{\epsilon}_N \xrightarrow[N \to \infty]{} 0\end{aligned}$$

for N so large that $|\lambda| \eta_N < 1$ and η_N is as defined in Theorem 2.

Proof: Since $Y - \bar{Y}_N = Y - Y_N - (1 - P_N)g$, from (13) it follows that

$$\begin{aligned}Y - \bar{Y}_N &= \lambda (1 - \lambda T_N)^{-1} [(1 - \lambda K_N)^{-1} K (1 - P_N) g \\ &\quad + (1 - P_N) K P_N Y_N]\end{aligned}$$

and the result follows by the usual estimates. QED

If the set $\{\phi_j\}$, $j = 1$ to N, is taken to be $\{K^{m-1}\psi_j\}$, $j = 1$ to J, $m = 1$ to M, $N = JM$, then BG reduces to the method of matrix moments.[12] (With $J = 1$, it is the ordinary method of moments[7]). The rate of convergence with this type of basis sets is usually faster but at the expense of some extra labor in iterating the kernel more number of times. At a finite order, however, a reasonable compromise may be made between the labor and accuracy by a careful choice of J and M with a specific problem at hand. Also from computational viewpoint it is sometimes more convenient, with either of the choices of the basis sets, that N be kept fixed and the set $\{\phi_j\}$ be varied to minimize the error.[13]

The problem of the two-body scattering involving a potential from the class $\mathcal{R} \cap L^1$, where $\mathcal{R}$ is the Rollnick class and L^1 is the class of absolutely integrable functions, creates no additional problem and Theorem 2 and Corollary 1 may be directly used to compute the error bounds on the BG approximation to the scattering amplitude and any other scattering parameter. Since BG in this case is equivalent to the Schwinger variational method,[1,6] this statement is applicable to the variational method as well. Although the two-body bound state problem can be reduced to studying a $K \in \mathcal{B}_2(H)$ for a much larger class than $\mathcal{R} \cap L^1$,[9] in order to obtain the binding energy from the information on K, one must overcome some additional analytic and computational problems. In the following we show that the result of Theorem 1 may be used to obtain upper and lower bounds on the binding energy.

Consider the Schrödinger equation

$$(H_0 - E_j - V)\psi_j = 0 \tag{15}$$

with symbols having the usual meaning and $V \geq 0$. Let $K(E) = \sqrt{V}\,(H_0 - E)^{-1}\sqrt{V}$ be the operator valued function of E on I where $I \subset (-\infty, 0)$ is an interval that contains all the eigenvalues in question in its interior. For a large class of potentials, including the asymptotically Coulombic ones, $K(E)$, $dK(E)/dE \in \mathcal{B}_2(H)$ are positive operator valued continuous functions of E on I.[14] By $dK(E)/dE \in \mathcal{B}_1(H)$ we mean that $\lim_{\epsilon \to 0}(1/\epsilon) \| K(E + \epsilon) - K(E) \|_1 < \infty$. As a consequence, the eigenvalues $\lambda_j(E)$, $j = 1$ to ∞, of $K(E)$ are positive continuously differentiable and monotonically increasing functions of $E \in I$, and E_j for $j = 1$ to n is given by[5]

$$1 - \lambda_j(E_j) = 0, \tag{16}$$

$n \leq \infty$, is the number of eigenvalues of $H_0 - V$ in I and I may be taken to be finite. For each $\lambda_j(E)$ such that $\lambda_j(E) \geq 1$ for some $E \in I$, (16) has a unique solution. Also it is easy to prove that the eigenvalues $\lambda_j^N(E)$, $j = 1$ to N, have these properties. One has to observe only that P_N are independent of E and proceed as in the case of $\lambda_j(E)$. This is sufficient to show that if $E_j < 0$ then for sufficiently large $N(j)$, the equation

$$1 - \lambda_j^N(E_j^N) = 0 \tag{17}$$

has a unique solution $E_j^N \geq E_j$ and $E_j^N \xrightarrow[N \to \infty]{} E_j$ for fixed $j = 1$ to n.[5] If n is finite, then N can be chosen to be independent of j. In the following we show that the result of Theorem 1 may be used to compute a converging lower bound on E_j. For the sake of brevity of the exposition, we restrict ourselves to the present case. Some extensions are possible which are easy to notice. Also we omit some trivial details. Further, in case $\lambda_j(-0)$, $d\lambda_j(E)/dE|_{E \to -0}$ is infinite, e.g., the case of long range potentials,[5] I will be chosen so that $I \subset (-\infty, -\epsilon)$, $\epsilon > 0$. In that case $\{E_j\} \subset I$ is always finite, and $\sup_{E \in I} |\lambda_j(E)|$, $(d/dE)\lambda_j(E)|_{j=1}^n < \infty$.

Theorem 3: Let all the symbols be as in the preceding paragraph, and $\delta_j^N(E) = \delta_j^N$ of Theorem 1 with $K(E)$ replacing K. Then given $j = 1$ to n, for sufficiently large N, the equation

$$1 - \lambda_j^N(\bar{E}_j^N) - \delta_j^N(\bar{E}_j^N) = 0 \tag{18}$$

has a unique solution $\bar{E}_j^N \leq E_j$ and $\bar{E}_j^N \xrightarrow[N \to \infty]{} E_j$.

We divide the proof into several steps.

Step 1:

$$\frac{d}{dE}\gamma_N(E) \xrightarrow[N \to \infty]{} 0 \quad \text{uniformly for } E \in I.$$

Proof: $\Delta^N(E) = \| p_j^N(E) K(E) (1 - p_j^N(E)) \|_L \xrightarrow[N \to \infty]{} 0$ uniformly for $E \in I$, for $\Delta^N(E)$ is continuous and I is finite. Continuity of $p_j^N(E)$ follows from the uniform continuity of $K_N(E)$. Thus the result will follow if $d\Delta^N(E)/dE$ is bounded uniformly with respect to N and E. Now

$$\begin{aligned}\left| \frac{d\Delta^N(E)}{dE} \right| &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \Big| \, \| p_j^N(E + \epsilon) K(E + \epsilon)(1 - p_j^N(E + \epsilon)) \|_L \\ &\quad - \| p_j^N(E) K(E)(1 - p_j^N(E)) \|_L \Big| \\ &\leq \lim_{\epsilon \to 0} \frac{1}{\epsilon} \| p_j^N(E + \epsilon) K(E + \epsilon)[1 - p_j^N(E + \epsilon)] \\ &\quad - p_j^N(E) K(E)[1 - p_j^N(E)] \|_L \\ &= \Big\| \frac{dp_j^N(E)}{dE} K(E)[1 - p_j^N(E)] \\ &\quad + p_j^N(E) \frac{dK(E)}{dE}[1 - p_j^N(E)] - p_j^N(E) K(E) \frac{dp_j^N(E)}{dE} \Big\|_L .\end{aligned}$$

All the factors involved are uniformly bounded with respect to N. Uniform boundedness on I follows from the finiteness of I and continuity of $d\Delta^N(E)/dE$, which is straightforward to check.

Step 2: $d\alpha_N(E)/dE$ is bounded uniformly with respect to N and $E \in I$.

Proof: Since $(d/dE)\left|\lambda_i^N(E) - \lambda_j^N(E)\right|$ is uniformly bounded for $E \in I$, the result will follow if

$$(d/dE)\left\|[1 - p_j^N(E)][K(E) - K_N(E)][1 - p_j^N(E)]\right\|_L$$

is uniformly bounded. For this we need to proceed exactly as in Step 1.

Step 3: Result of the theorem.

Proof: By differentiating $\delta_j^N(E)$ and using the results of Steps 1 and 2, one readily obtains that

$$\frac{d}{dE}\delta_j^N(E) \xrightarrow[N\to\infty]{} 0 \quad \text{uniformly for } E \in I.$$

Consequently, $(d/dE)[\lambda_j^N(E) + \delta_j^N(E)] > 0$ for sufficiently large N independent of E, for $(d/dE)\lambda_j^N(E) \to (d/dE)\lambda_j(E) > 0$ and the other factor goes to zero.

This is sufficient to prove the existence of a unique solution $\bar{E}_j^N$ of Eq. (18) such that $\bar{E}_j^N \le E_j$, using the additional fact that $\lambda_j^N(E) + \delta_j^N(E) \ge \lambda_j(E)$. The convergence to E_j is proven exactly as in the case of E_j^N,[5] observing that $\lambda_j^N(E) + \delta_j^N(E) \xrightarrow[N\to\infty]{} \lambda_j(E)$, $E \in I$. QED

The result of Theorem 3 can be extended to include potentials that have nonzero positive and negative parts but are capable of supporting some bound states. The only thing to observe is that only such a $\lambda_j(E)$ is of consequence which is positive and ≥ 1, and such a $\lambda_j(E)$ has all the properties required.[9] An alternative way to handle this case is to include the negative part of V in the definition of H_0.[10]

3. NUMERICAL EXAMPLES

The error bounds of Sec. 2 converge to zero in the limit of large N. Also some features of the formulas give some estimate of the labor involved and of their accuracy at a finite order. In order to compute δ_j^N and ϵ_N, $\bar{\epsilon}_N$ one must iterate K one more time than is necessary to compute λ_j^N and Y_N. However, this is not surprising if, for example, we compare the present method with the method of moments. In the latter, with some additional restrictions on K, one needs at least the first two moments to obtain a lower bound and at least three to obtain an upper bound. In order to improve the accuracy at a finite order, we have majorized the expressions by the smallest computable upper bound available. Owing to these careful estimates, for example, if $p_j^N = p_j$ for some N (Theorem 1), then $\delta_j^N = 0$. Nevertheless, a concrete numerical example would give a more accurate estimate of the amount of labor involved and the accuracy of the bounds.

TABLE I. Relative error bounds: $\Delta_j^L = (\lambda_j - \lambda_j^N)/\lambda_j$, $\Delta_L^u = (\lambda_j^N + \delta_j^N - \lambda_j)/\lambda_j$ for the eigenvalues λ_j of K of Eq. (19).

N \ j	Δ_1^L	Δ_1^U	Δ_2^L	Δ_2^U	Δ_3^L	Δ_3^U
2	2×10^{-2}	8×10^{-3}	3×10^{-1}	x	x	x
3	3×10^{-4}	8×10^{-5}	7×10^{-2}	x	4×10^{-1}	x
4	1×10^{-5}	3×10^{-6}	5×10^{-3}	8×10^{-3}	1×10^{-1}	x
5	5×10^{-8}	1×10^{-8}	5×10^{-4}	6×10^{-4}	2×10^{-2}	x
6	9×10^{-10}	2×10^{-10}	2×10^{-5}	2×10^{-5}	4×10^{-3}	x
7	2×10^{-12}	7×10^{-13}	8×10^{-7}	7×10^{-7}	3×10^{-4}	1×10^{-3}
8	9×10^{-14}	2×10^{-13}	1×10^{-8}	1×10^{-8}	3×10^{-5}	8×10^{-5}

In the following we test the present method for the case of some Hilbert–Schmidt operators ($l = 2$), since this is the type of operators one faces in the two-body scattering and bound state problems. Explicit expressions in terms of the moments of K for this case are given in the Appendix.

The first example we consider[15] is of K defined by

$$(Ku)(\chi) = \int_0^1 k(\chi, \xi)u(\xi)\,d\xi, \tag{19}$$

where

$$k(\chi, \xi) = \begin{cases} \frac{1}{2}(2 - \chi)\xi, & \xi \le \chi, \\ \frac{1}{2}(2 - \xi)\chi, & \chi < \xi. \end{cases}$$

In this case $H = L^2[0, 1]$, $K \in \mathcal{B}_2(H)$, $\|K\|_2 = \sqrt{11/180}$. The eigenvalue problem is exactly solvable, yielding

$$\lambda_j^{-1/2} + \tan\lambda_j^{-1/2} = 0, \quad j = 1, 2, \ldots. \tag{20}$$

For the basis set we chose

$$\phi_j(\chi) = \chi^{j-1}, \quad j = 1, \ldots, N. \tag{21}$$

The relative error bounds to the first three eigenvalues $\lambda_1 = .242962685095\ldots$, $\lambda_2 = .04142261498403\ldots$, $\lambda_3 = .0157086716133\ldots$ are shown in Table I for various values of N. It is remarkable that the upper bounds are as accurate as the lower bounds, and are obtained with little extra labor.

TABLE II. Error bounds ϵ_N and $\|Y - Y_N\|$ for the solution of Eq. (1) with K and Y of Eqs. (19) and (22).

N \ λ	1.0		5.0		−1.0		−5.0	
	$\|Y - Y_N\|$	ϵ_N	$\|Y - Y_N\|$	ϵ_N	$\|Y - Y_N\|$	ϵ_N	$\|Y - Y_N\|$	ϵ_N
2	2.5×10^{-2}	2.8×10^{-2}	1.4×10^{-1}		1.5×10^{-2}	2.5×10^{-2}	7.5×10^{-3}	5.3×10^{-2}
3	3.9×10^{-3}	4.1×10^{-3}	1.0×10^{-2}	3.2×10^{-2}	2.6×10^{-3}	3.7×10^{-3}	1.7×10^{-3}	5.4×10^{-3}
4	1.4×10^{-4}	1.4×10^{-4}	2.8×10^{-3}	3.8×10^{-3}	7.7×10^{-5}	7.9×10^{-5}	1.9×10^{-4}	2.2×10^{-4}
5	1.3×10^{-5}	1.3×10^{-5}	1.6×10^{-4}	1.9×10^{-4}	8.4×10^{-6}	8.5×10^{-6}	2.6×10^{-5}	2.8×10^{-5}

TABLE III. Relative error bounds: $\Delta_j^L = (\lambda_j - \lambda_j^N)/\lambda_j$, $\Delta_j^U = (\lambda_j^N + \delta_j^N - \lambda_j)/\lambda_j$ for the eigenvalues λ_j of $K(0.1)$ of Eq. (23).

N \ j	Δ_1^L	Δ_1^U	Δ_2^L	Δ_2^U
2	2×10^{-2}	9×10^{-3}	6×10^{-1}	x
3	3×10^{-3}	1×10^{-3}	3×10^{-1}	x
4	3×10^{-4}	1×10^{-4}	1×10^{-1}	x
5	2×10^{-4}	7×10^{-5}	3×10^{-2}	x
6	2×10^{-4}	6×10^{-5}	9×10^{-3}	3×10^{-2}
7	1×10^{-4}	4×10^{-5}	4×10^{-3}	1×10^{-2}
8	7×10^{-5}	2×10^{-5}	3×10^{-3}	8×10^{-3}

For the case of an inhomogeneous equation, involving this K, we take $g(\chi) = \chi^3/6 - \chi/3$ in Eq. (1). Exact solution for this case is

$$Y = \tfrac{1}{2}[\chi - 2\sin(\sqrt{\lambda}\chi)/(\sin\sqrt{\lambda} + \sqrt{\lambda}\cos\sqrt{\lambda})]. \tag{22}$$

ϵ_N (Theorem 2) with the basis set given by (21) for several N values is compared in Table II with $\|Y - Y_N\|$ for $\lambda = 1.0, 5.0, -1.0, -5.0$. Our bounds do not differ much from the best possible bounds, viz., $\|Y - Y_N\|$.

The other case we consider is of $K(p)$ defined by

$$(K(p)u)(\rho) = \int_0^\infty k(p;\rho,\xi)u(\xi)d\xi, \tag{23}$$

where

$$k(p;\rho,\xi) = \begin{cases} (1/2p)e^{-(\rho+\xi)/2-p\rho}(e^{p\xi} - e^{-p\xi}), & \xi \le \rho, \\ (1/2p)e^{-(\rho+\xi)/2-p\xi}(e^{p\rho} - e^{-p\rho}), & \xi > \rho. \end{cases} \tag{24}$$

The underlying $H = L^2[0,\infty)$, $K(p) \in \mathcal{B}_2(H)$ with $\|K(p)\|_2 = [2(1+p)(1+2p)^2]^{-1/2}$. This case arises in the S-wave bound state problem involving an exponential potential, with a proper choice of units and the binding energy $E_j = -p_j^2$. The exact eigenvalues $\lambda_j(p)$ are given by[16]

$$J_{2p}(2\lambda_j^{-1/2}(p)) = 0, \quad j = 1, 2, \ldots, \tag{25}$$

where $J_\nu(\cdot)$ is the Bessel function. The basis set was chosen to be

$$\phi_j(\rho) = \rho^j e^{-\rho/2}, \quad j = 1, 2, \ldots. \tag{26}$$

The relative error bounds to the first two eigenvalues $\lambda_1(p) = 0.545833$, $\lambda_2(p) = 0.117637$ for $p = 0.1$ are given in Table III. The same conclusions are reached as in the previous example.

In the tables x stands for the case when the quantity in question could not be computed because N was not large enough to satisfy the condition of the corresponding theorem.

ACKNOWLEDGMENTS

One of us (S. R. Singh) is thankful to Professor J. Nuttall for helpful discussions and his hospitality. The other (G. Turchetti) is thankful to Professor J. W. McGowan for his hospitality at the Centre for Interdisciplinary Studies in Chemical Physics of UWO. Thanks are also due to Dr. T. Ghemires and Dr. F. Chatelin for pointing out an error that occurred in the original manuscript.

APPENDIX

Let $(\mu_n)_{ij} = \langle \phi_i | K^n \phi_j \rangle$, $i,j = 1, \ldots, N$, $n = 0, 1, \ldots$, and let μ_n be the matrix with elements $(\mu_n)_{ij}$, and

$$\beta_j = \langle \phi_j | g \rangle, \quad j = 1, 2, \ldots, N.$$

Then λ_j^N, $j = 1$ to N, are given by

$$(\mu_1 - \lambda_j^N \mu_0)u_j^N = 0,$$

where u_j^N is a column vector with N components. Let U be the matrix with columns u_j^N, $j = 1$ to N, and let V be the matrix with elements V_{ij}, where

$$V_{ij} = U_{ij}[U^t \mu_0 U]_{jj}^{-1/2}, \quad i,j = 1, \ldots, N.$$

Then the expressions needed to compute δ_j^N (Theorem 1, $l = L = 2$) are given by

$$\gamma_N = (V^t \mu_2 V)_{jj} - (\lambda_j^N)^2$$

and

$$\alpha_N = \min_{j \ne i = 1, N} |\lambda_j^N - \lambda_i^N| - \left[\|K\|_2^2 - \sum_{i=1}^N (\lambda_i^N)^2 + 2(\lambda_j^N)^2 - 2(V^t \mu_2 V)_{jj} \right]^{1/2}.$$

For the inhomogeneous problem, let M be defined by

$$M = \mu_0 - \lambda \mu_1.$$

Then the expressions needed to compute ϵ_N (Theorem 2, $l = L = 2$) are given by

$$\|(1 - P_N)g\| = [\|g\|^2 - \beta^t \mu_0^{-1} \beta]^{1/2},$$

$$\|(1 - P_N)K P_N Y_N\| = [\beta^t M^{-1}(\mu_2 - \mu_1 \mu_0^{-1} \mu_1) M^{-1} \beta]^{1/2},$$

$$\|(1 - P_N)K\|_2 = [\|K\|_2^2 - \mathrm{Tr}(\mu_2 \mu_0^{-1})]^{1/2},$$

$$\|(1 - \lambda K_N)^{-1} P_N K (1 - P_N)\|_2 = [\mathrm{Tr} M^{-1}(\mu_2 - \mu_1^2) M^{-1}]^{1/2}.$$

*Supported in part by the National Research Council of Canada and Centre for Interdisciplinary Studies in Chemical Physics, UWO.

†Permanent address: Istituto di Física, Università di Bologna and INFN Sezione di Bologna.

[1]See, e.g., B. Noble, *Topics in Numerical Analysis*, edited by J. H. Miller (Academic, New York, 1973), pp. 211–32.
[2]S. R. Singh and M. C. Zerner, "The Collocation-Variational Method to Solve the Fredholm Integral Equations and Some Applications," submitted for publication.
[3]See, e.g., S. G. Mikhlin, *Variational Methods in Mathematical Physics* (Pergamon, New York, 1964), Chap. IX.
[4]S. R. Singh and A. D. Stauffer, Nuovo Cimento B **22**, 1 (1974).
[5]S. R. Singh and A. D. Stauffer, Nuovo Cimento B **25**, 547 (1975).
[6]S. R. Singh and A. D. Stauffer, J. Phys. A 8, 1379 (1975).
[7]See, e.g., Yu. V. Vorobyev, *Method of Moments in Applied Mathematics* (Gordon and Breach, New York, 1965).
[8]See, e.g., Ref. 4 and the references cited therein.
[9]See, e.g., S. Weinberg, Phys. Rev. **131**, 440 (1963); M. Scadron, S. Weinberg, and J. Wright, Phys. Rev. **135**, B202 (1964).
[10]See, e.g., T. Kato, Prog. Theor. Phys. **6**, 394 (1951); R. Sugar and R. Blakenbeckler, Phys. Rev. **136**, B472 (1964); M. Cohen, R. P. McEachran, S. Cameron, and J. A. Stauffer, J. Phys. B **4**, 1109 (1971), and references cited therein.
[11]S. R. Singh and A. D. Stauffer, "A rigorous derivation of the Hellmann–Feynman Theorem for bound and scattering states,"

submitted for publication.
[12]G. Turchetti, Lett. Nuovo Cimento **15**, 123 (1976); "Variational matrix Padé approximants in two body interactions," submitted for publication.
[13]D. Bessis, P. Mery, and G. Turchetti, "Variational bounds for positive interactions," submitted for publication.
[14]See, e.g., Ref. 11 and references cited therein.
[15]Reference 7, p. 45.
[16]H. A. Bethe and R. Bacher, Rev. Mod. Phys. **8**, 111 (1936).

Broken linear transformations

A. Gervois and M. L. Mehta*

Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, B.P. n°2-91190 Gif-sur-Yvette, France
(Received 10 May 1976)

The properties of mixing and complete mixing are studied for the baker's transformation and some generalizations with the usual measure $d\mu(x) = dx$. When each linear portion maps on the whole segment, mixing is true. When some linear portions do not map on the whole segment, mixing can hold or not, depending on the transformation even with a more general measure.

1. THE BAKER'S TRANSFORMATION

One of the most studied broken linear transformations is the baker's transformation. To make bread, the baker first prepares a paste of flour, water, salt, etc. To make sure that the paste is homogeneous, he repeats certain movements which may be schematized for a one-dimensional paste as folding it over and then stretching it to its original length:

$$Tx = \begin{cases} 2x & \text{if } 0 \leq x \leq \frac{1}{2}, \\ 2(1-x) & \text{if } \frac{1}{2} \leq x \leq 1. \end{cases} \tag{1.1}$$

A function $P(x)$ defined for $0 \leq x \leq 1$ (for example, the distribution of salt in the paste) is changed by this transformation to

$$\mathbf{T}P(x) = \tfrac{1}{2}[P(x/2) + P(1 - x/2)]. \tag{1.2}$$

Transformations T and $\mathbf{T}$ are related through

$$\int_0^1 P(x)Q(Tx)\,dx = \int_0^1 [\mathbf{T}P(x)]Q(x)\,dx \tag{1.3}$$

for arbitrary functions $P(x)$ and $Q(x)$.

Several well-known properties of T or $\mathbf{T}$ may be noted.

1.1. The fixed points of T are 0 and $\frac{2}{3}$, i.e., $Tx = x$ for $x = 0$ and $x = \frac{2}{3}$. One may also say that 0 and $\frac{2}{3}$ form 1-periods of T.

1.2. Successive transforms of any rational point $x = p/q$, where $q = 2^\alpha \cdot q'$, α an integer and q' an odd integer, ultimately form an r-period with $r \leq \frac{1}{2}(q-1)$. In fact, if $x = p/q$, then $Tx = p_1/q_1$, where $p_1 \leq q_1$ and $q_1 = q$ or $q_1 = q/2$ according as q is odd or even. Thus, if $q = 2^\alpha \cdot q'$, then $T^\alpha x = p'/q'$, $p' \leq q'$. Further applications of T change p' to even integers less than q'. As these later are finite in number, the $T^n x$ for $n > \alpha$ form an r-period with $r \leq (q'-1)/2$. In the baker's language, if initially the salt is concentrated at a rational point then even after an infinite number of repetitions of T the salt will be found on at most a finite number of points.[1]

1.3. The transformation T has the property of complete mixing; i.e., if the function $P(x)$ has a support of nonzero measure—for example, if $P(x) \neq 0$ over an interval however small of (0, 1)—then

$$\lim_{n\to\infty} \mathbf{T}^n P(x) = \text{const} = \int_0^1 P(x)\,dx. \tag{1.4}$$

In the baker's language, if initially the salt covers a nonzero length, whatever its concentration, then repetitions of T ultimately spreads it out over the entire length (0, 1) uniformly.

To see this, write $P(x)$ as a Fourier series

$$P(x) = \sum_{m=-\infty}^{\infty} c_m \exp(2\pi i m x),$$
$$c_m = \int_0^1 P(x) \exp(-2\pi i m x) \tag{1.5}$$

and note that

$$\mathbf{T}^2[\exp(2\pi i m x)] = \mathbf{T}\cos(\pi m x)$$
$$= \begin{cases} \cos(\pi m x/2) & \text{if } m \text{ is even}, \\ 0 & \text{if } m \text{ is odd}. \end{cases} \tag{1.6}$$

Thus, if $m = 2^\alpha \cdot m'$, α, m' integers and m' odd, then $\mathbf{T}^{\alpha+2}\exp(2\pi i m x) = 0$, i.e.,

$$\lim_{n\to\infty} \mathbf{T}^n P(x) = \sum_{m=-\infty}^{\infty} c_m \lim_{n\to\infty} \mathbf{T}^n \exp(2\pi i m x)$$
$$= c_0 = \int_0^1 P(x)\,dx.$$

Using Eq. (1.3), the complete mixing property (1.4) may be rewritten as

$$\lim_{n\to\infty} \int_0^1 P(x)Q(T^n x)\,dx = \int_0^1 P(x)\,dx \int_0^1 Q(x)\,dx \tag{1.7}$$

for every pair of functions P and Q.

1.4. One may sometimes be satisfied with a lesser property. A transformation τ is said to have the property of mixing (and not of complete mixing!) if

$$\lim_{n\to\infty} \int_0^1 x^m \tau^n x\,dx = \int_0^1 x^m\,dx \int_0^1 x\,dx = 1/2(m+1) \tag{1.8}$$

for every positive integer m. Because of Carlson's theorem,[2] (1.8) is still true for every real positive m. Complete mixing and mixing are similar properties except for a certain class of functions, not regular enough to be expanded, in some way, as infinite sum of powers of x [for example, $P(x) = \sin(1/x)\cdots$].

For the baker's transformation T it is easy to verify that

$$\int_0^1 x T^n x\,dx$$
$$= \int_0^{2^{-n}} dx\, 2^n x \sum_{j=0}^{2^{n-1}-1} [(2^{-n+1}j + x) + (2^{-n+1}j - x)] \tag{1.9}$$
$$= \tfrac{1}{4}, \quad n \geq 1,$$

and, in general,

$$\int_0^1 x^m T^n x\,dx$$
$$= \int_0^{2^{-n}} 2^n x \sum_{j=0}^{2^{n-1}-1} [(2^{-n+1}j + x)^m + (2^{-n+1}j - x)^m]\,dx$$
$$= \frac{2^{-n(m+1)+1}}{(m+1)(m+2)} \sum_{j=0}^{2^{n-1}-1} [(2j)^{m+2} - (2j+1)^{m+2}]$$
$$+ \frac{2^n}{(m+1)(m+2)}$$
$$= \frac{2^{-n(m+1)+1}}{(m+1)(m+2)} \left(2^{m+3} \sum_{j=1}^{2^{n-1}-1} j^{m+2} - \sum_{j=1}^{2^n-1} j^{m+2} \right)$$
$$+ \frac{2^n}{(m+1)(m+2)}. \tag{1.10}$$

On using the formula

$$\sum_{j=1}^{n-1} j^p = \frac{1}{p+1} \sum_{j=1}^{p} \binom{p+1}{j} B_j n^{p+1-j}, \tag{1.11}$$

where B_j are Bernoulli numbers, we get after some arithmetic

$$\int_0^1 x^m T^n x\,dx = \frac{1}{m+1} \sum_{j=0}^{m} \binom{m+1}{j} \frac{2B_{j+2}}{(j+1)(j+2)} 2^{-nj}(2^{j+2} - 1) \tag{1.12}$$

$$= \frac{1}{2(m+1)} + O(2^{-2n}). \tag{1.13}$$

Of course, the complete mixing property implies the mixing property and therefore this verification is superfluous for T.

1.5. One may think of nonlinear transformations which are equivalent to baker's T. For example, let

$$H_1 x = \sin^2(\pi x/2), \quad H_1^{-1} x = (2/\pi) \sin^{-1}\sqrt{x};$$

then

$$F_1 x = H_1 T H_1^{-1} x = 4x(1-x)$$

is the quadratic transformation of Ulam and von Neuman.[3] Similarly

$$H_2 x = \sin(\pi x/2), \quad H_3 x = \tan(\pi x/4),$$
$$H_4 x = \tan^2(\pi x/4)$$

give us

$$F_j x = H_j T H_j^{-1} x,$$
$$F_2 x = 2x(1-x^2)^{1/2}, \quad 0 \le x \le 1,$$
$$F_3 x = \begin{cases} 2x/(1-x^2), & 0 \le x \le \sqrt{2}-1, \\ (1-x^2)/2x, & \sqrt{2}-1 \le x \le 1, \end{cases}$$

and

$$F_4 x = \begin{cases} 4x/(1-x)^2, & 0 \le x \le 3-\sqrt{8}, \\ (1-x)^2/4x, & 3-\sqrt{8} \le x \le 1. \end{cases}$$

These F_j all have the property of complete mixing just as T.

2. GENERALIZATIONS

2.1. A trivial generalization will be to fold the paste over exactly l times and then stretch it to its original length, where l is a positive integer;

$$T_l x = \begin{cases} -2j + lx, & 2j/l \le x \le (2j+1)/l, \\ & j = 0, 1, \ldots, [(l-1)/2], \\ 2j - lx, & (2j-1)/l \le x \le 2j/l, \\ & j = 1, 2, \ldots, [l/2], \end{cases} \tag{2.1}$$

and $[x]$ denotes the largest integer less than or equal to x. The transformation T considered above in paragraph 1.1 is T_2. A function $P(x)$ is changed by T_l to

$$\mathbf{T}_l P(x) = \frac{1}{l}\left[P\left(\frac{x}{l}\right) + P\left(\frac{2-x}{l}\right) + P\left(\frac{2+x}{l}\right) + P\left(\frac{4-x}{l}\right) + \cdots \right]$$
$$= \frac{1}{l}\left[\sum_{j=0}^{[(l-1)/2]} P\left(\frac{2j+x}{l}\right) + \sum_{j=1}^{[l/2]} P\left(\frac{2j-x}{l}\right) \right]. \tag{2.2}$$

It is easy to verify the complete mixing property when l is even. In fact from Eq. (2.2)

$$\mathbf{T}_l \exp(2\pi i m x) = \begin{cases} \cos(2\pi m x/l), & \text{if } 2m/l \text{ is integer}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.3}$$

and

$$\mathbf{T}_l \cos\left(\frac{2\pi m x}{l}\right) = \begin{cases} \cos(2\pi m x/l^2), & \text{if } 2m/l^2 \text{ is integer}, \\ 0, & \text{otherwise}. \end{cases} \tag{2.4}$$

Thus, as in paragraph 1.3, expanding $P(x)$ in a Fourier series, we see that, in the limit $n \to \infty$,

$$\lim_{n\to\infty} \mathbf{T}_l^n P(x) = \int_0^1 P(x)\,dx. \tag{2.5}$$

When l is odd, this procedure is not easy, and only the property of mixing is verified below.

We will follow paragraph 1.4 and calculate

$$\int_0^1 x^m T_l^n x\,dx$$
$$= \int_0^{l^{-n}} l^n x \left[\sum_{j=0}^{[(l^n-1)/2]} (2jl^{-n} + x)^m + \sum_{j=1}^{[l^n/2]} (2jl^{-n} - x)^m \right] dx$$
$$= \frac{l^{-n(m+1)}}{m+1} \left[\sum_{j=0}^{[(l^n-1)/2]} \left((2j+1)^{m+1} - \frac{1}{m+2}\,[(2j+1)^{m+2} - (2j)^{m+2}] \right) \right.$$
$$\left. - \sum_{j=1}^{[l^n/2]} \left((2j-1)^{m+1} - \frac{1}{m+2}\,[(2j)^{m+2} - (2j-1)^{m+2}] \right) \right]. \tag{2.6}$$

From here onwards it is better to separate the calculation for l even and l odd. Using Eq. (1.11), we get after some arithmetic

$$\int_0^1 x^m T^n x\,dx = \frac{1}{m+1} \sum_{j=0}^{m} \binom{m+1}{j} \frac{2B_{j+2}}{(j+1)(j+2)} (2^{j+2} - 1) l^{-nj} \tag{2.7}$$

when l is even and

$$\int_0^1 x^m T^n x\,dx$$
$$= \frac{2l^{2n}[(1+l^{-n})^{m+3} - 1]}{(m+1)(m+2)(m+3)} - \frac{2l^n(1+l^{-n})^{m+2}}{(m+1)(m+2)}$$
$$+ \frac{1}{6(m+1)}[4(1+l^{-n})^{m+1} + 5]$$
$$+ 2l^{-n} \sum_{j=0}^{m-1} \binom{m}{j} \frac{B_{j+3} l^{-nj}}{(j+1)(j+2)(j+3)} [2^{j+3}(1+l^{-n})^{m-j} - 1] \tag{2.8}$$

when l is odd. Thus whether l is even or odd, we get

$$\int_0^1 x^m T_l^n x\,dx = 1/2(m+1) + O(l^{-2n}) \tag{2.9}$$

and T_l has the property of mixing.

2.2. Another possible generalization of the baker's transformation is to fold and stretch unsymmetrically,

$$T_\alpha x = \begin{cases} x/\alpha & 0 < x < \alpha, \\ (1-x)/(1-\alpha), & \alpha < x < 1. \end{cases} \tag{2.10}$$

The corresponding $\mathbf{T}_\alpha$ is

$$\mathbf{T}_\alpha P(x) = \alpha P(\alpha x) + (1-\alpha)P[1-(1-\alpha)x]. \tag{2.11}$$

Already in this case it is not easy to confirm the complete mixing property. However, the mixing property can be verified as follows. Let

$$I(m,n) = \int_0^1 x^m T_\alpha^n x\,dx, \quad m \text{ positive integer}. \tag{2.12}$$

Then for $m \geq 1$, $n \geq 2$, we have

$$I(m,n) = \int_0^\alpha x^m T_\alpha^{n-1}\left(\frac{x}{\alpha}\right)dx + \int_\alpha^1 x^m T_\alpha^{n-1}\left(\frac{1-x}{1-\alpha}\right)dx.$$

A change of variables under the integral signs then gives the recurrence relation

$$I(m,n) = [\alpha^{m+1} - (\alpha-1)^{m+1}]I(m,n-1) - \sum_{j=0}^{m-1}\binom{m}{j}(\alpha-1)^{j+1}I(j,n-1), \tag{2.13}$$

$m \geq 1$, $n \geq 2$. The initial values can be easily calculated separately as

$$I(0,n) = \tfrac{1}{2}, \quad n \geq 1, \tag{2.14}$$

$$I(m,1) = \frac{1}{(m+1)(m+2)}\,\frac{1-\alpha^{m+1}}{1-\alpha}, \quad m \geq 0. \tag{2.15}$$

Equations (2.13), (2.15) determine $I(m,n)$ uniquely. As $I(m,n) = 1/2(m+1)$ is a solution of Eqs. (2.13) and (2.14), but not of (2.15), let us put

$$I(m,n) = 1/2(m+1) + F(m,n). \tag{2.16}$$

Then $F(m,n)$ satisfies the recurrence relation (2.13) with I replaced by F, the initial conditions now being

$$F(0,n) = 0, \quad F(m,1) = \frac{1}{(m+1)(m+2)} \times \frac{1-\alpha^{m+1}}{1-\alpha} - \frac{1}{2(m+1)}. \tag{2.17}$$

One can, in principle, determine all the $F(m,n)$ step by step, increasing m each time by one. For example,

$$F(1,n) = \tfrac{1}{12}[\alpha^2 - (\alpha-1)^2]^n, \tag{2.18}$$

$$F(2,n) = \frac{1}{6}\frac{[\alpha^2-(\alpha-1)^2]^n - [\alpha^3-(\alpha-1)^3]^n}{[\alpha^2-(\alpha-1)^2 - \alpha^3 + (\alpha-1)^3]} + \tfrac{1}{12}[\alpha^3-(\alpha-1)^3]^n(\alpha-1)(\alpha^3 - 2\alpha + 3). \tag{2.19}$$

In fact, for the mixing property we need only the result

$$\lim_{n\to\infty} F(m,n) = 0, \tag{2.20}$$

which can be seen by a recurrence on m, provided that $0 < \alpha < 1$.

2.3. There is no essential change in stretching and folding unsymmetrically at several points; for $0 = a_0 < a_1 < \cdots < a_l = 1$, let

$$Tx \equiv T(a_0, a_1, \ldots, a_l)x = \begin{cases} \dfrac{x - a_{2i}}{a_{2i+1} - a_{2i}}, & a_{2i} \leq x \leq a_{2i+1}, \quad i = 0, 1, \ldots, [(l-1)/2], \\ \dfrac{a_{2i} - x}{a_{2i} - a_{2i-1}}, & a_{2i-1} \leq x \leq a_{2i}, \quad i = 1, 2, \ldots, [l/2]. \end{cases} \tag{2.21}$$

The corresponding $\mathbf{T}$ is

$$\mathbf{T}(a_0, \ldots, a_l)P(x) = \sum_{i=0}^{[(l-1)/2]} (a_{2i+1} - a_{2i})P[a_{2i} + (a_{2i+1} - a_{2i})x] + \sum_{i=1}^{[l/2]} (a_{2i} - a_{2i-1})P[a_{2i} - (a_{2i} - a_{2i-1})x]. \tag{2.22}$$

As in paragraph 2.2 above, one may verify the mixing property of this T. Let

$$I(m,n) = \int_0^1 x^m T^n x\,dx.$$

Then just as in paragraph 2.2, we have for $m \geq 1$, $n \geq 2$,

$$I(m,n) = \left[\sum_{i=0}^{[(l-1)/2]} (a_{2i+1} - a_{2i})^{m+1} - \sum_{i=1}^{[l/2]} (a_{2i} - a_{2i-1})^{m+1}\right] \times I(m,n-1) + \sum_{j=0}^{m-1}\binom{m}{j}\left[\sum_{i=0}^{[(l-1)/2]} (a_{2i+1} - a_{2i})^{j+1} - \sum_{i=1}^{[l/2]} (a_{2i} - a_{2i-1})^{j+1}\right] a_{2i}^{m-j} I(j,n-1).$$

with the initial conditions (2.14) and easily calculated $I(m,1)$. Putting, as before,

$$I(m,n) = \frac{1}{2(m+1)} + F(m,n)$$

and observing that for any $j > 0$,

$$\left|\sum_{i=0}^{[(l-1)/2]} (a_{2i+1} - a_{2i})^j\right| \quad \text{and} \quad \left|\sum_{i=1}^{[l/2]} (a_{2i} - a_{2i-1})^j\right|$$

are both less than unity, a recurrence on m will show that

$$\lim_{n\to\infty} F(m,n) = 0, \quad m > 0.$$

As in (1.5), we may find nonlinear transformations which are equivalent to a transformation (2.21) or more generally to a broken linear transformation on $[\alpha, \beta]$ and such that both minimum α and maximum β are reached on each linear fragment.

For example, let $T_k(x)$ and $U_k(x)$ be the Tchebycheff polynomials[4] of order k, $V_k(x)$ being the related function $V_k(x) = (1-x^2)^{1/2}U_k(x)$; then, for

$$H(x) = \cos(\pi x/2), \quad H^{-1}(x) = (2/\pi)\cos^{-1}x,$$

$H^{-1}T_kH(x)$ and $H^{-1}V_kH(x)$ both are piecewise linear. Furthermore, in this peculiar case complete mixing can be proved, at least for even k.

2.4. Next one may consider a broken linear trans-

formation in which some of the linear parts do not map over the whole segment (0, 1). For example, for $0 < \alpha, \beta < 1$, let

$$Tx = \begin{cases} (1-\beta)x/\alpha + \beta, & 0 \le x \le \alpha, \\ (1-x)/(1-\alpha), & \alpha \le x \le 1. \end{cases} \tag{2.23}$$

This transformation T does not have the mixing property, as can be seen as follows. We have the recurrence relation:

$$I(m,n) = \int_0^1 x^m T^n x\, dx$$
$$= \left(\frac{\alpha}{1-\beta}\right)^{m+1} \int_\beta^1 (x-\beta)^m T^{n-1} x\, dx$$
$$+ (1-\alpha) \int_0^1 [1-(1-\alpha)x]^m T^{n-1} x\, dx.$$

A favorable case will be to take $\alpha = \beta$. Then

$$I(m,n) = \alpha^{m+1} \int_0^1 (1-x)^m T^{n-2} x\, dx$$
$$+ (1-\alpha) \int_0^1 [1-(1-\alpha)x]^m T^{n-1} x\, dx$$
$$= \sum_{j=0}^{m} \binom{m}{j} (-)^j [\alpha^{m+1} I(j, n-2) + (1-\alpha)^{j+1} I(j, n-1)].$$

These difference equations, along with the initial conditions

$$I(m,1) = [(m+1)(m+2)]^{-1}[\alpha^{m+2} + (1-\alpha^{m+1})/(1-\alpha)]$$

and

$$I(m,2) = \alpha/(m+1) + [(m+1)(m+2)]^{-1}$$
$$\times \{\alpha^{m+1} + \alpha^{-1} + [\alpha^{m+3} - (1-\alpha+\alpha^2)^{m+3}]/[\alpha(1-\alpha)^2]\},$$

give

$$\lim_{n\to\infty} I(m,n) = \frac{1}{2(m+1)} \frac{1-\alpha^3}{1-\alpha^2}, \quad 0 \ne \alpha \ne 1.$$

As $(1-\alpha^3)/(1-\alpha^2)$ is not equal to 1, the mixing property is absent.

The conclusion seems to be that a broken linear transformation has the property of mixing if and only if every linear portion of it maps on the whole segment (0, 1).

Actually, the definitions (1.7) and (1.8) can be enlarged to less trivial measures $\mu(x)$. First of all, we must set $\langle T^n(x) \rangle$ is invariant; then we shall say that there is mixing whenever

$$\lim_{n\to\infty} \int_0^1 x^m T^n x\, d\mu(x)$$
$$= \int_0^1 x^m\, d\mu(x) \int_0^1 T^n(x)\, d\mu(x)$$

and there is complete mixing if

$$\lim_{n\to\infty} \int_0^1 P(x) Q[T^n x]\, d\mu(x)$$
$$= \int_0^1 P(x)\, d\mu(x) \int_0^1 Q(x)\, d\mu(x)$$

for every pair of arbitrary functions P and Q.

In this new sense, mixing can be proved for the case $\alpha = \beta$ above when setting

$$d\mu(x) = \begin{cases} dx/(1+\alpha), & \text{if } 0 < x < \alpha, \\ dx/(1-\alpha^2), & \text{if } \alpha < x < 1. \end{cases} \tag{2.24}$$

It is presumably a limiting case between mixing and nonmixing, but we were not able to prove it. Experiments on computer[5] for $\alpha = \beta = \frac{1}{2}$ show that the total length of the segments of (0, 1) for which $T^n(x) \ge \frac{1}{2}$ [resp. $T^n(x) < \frac{1}{2}$] goes to a limit which is $\frac{2}{3}$ [resp. $\frac{1}{3}$] in agreement with (2.24). One may recover this result by a recurrence.[5] This is perhaps the way of testing whether or not there exists a measure constant by steps for which mixing holds in the more general case (2.23).

For the "mirror" transformation

$$T_1 x = \begin{cases} x/\alpha & \text{if } x < \alpha, \\ 1 + \alpha(1-\beta)/(1-\alpha) - [(1-\beta)/(1-\alpha)]x & \text{if } x > \alpha, \end{cases}$$

it is impossible to find a measure continuous by steps. This can be seen directly on the graph of $T^n(x)$. For $x > \alpha^n$, $T^{n+1}(x) \ge \beta$, and when n becomes very large, the origin becomes a singular point, the graph of $T^{n+1}(x)$ for $x > 0$ being concentrated in a band $\beta \le T^{n+1}(x) \le 1$. The only "measure" for which mixing is verified is $d\mu(x) = \delta(x - x_0)\, dx$, where δ is the Dirac δ distribution and x_0 is the invariant point $T(x_0) = x_0 = (1 - \alpha\beta)/(2 - \alpha - \beta)$.

The same thing occurs for the quadratic transformation

$$T_2 x = 4kx(1-x), \quad k \le 1,$$

and there is surely no mixing. The problem remains in both cases of the repartition in the plane of the extremities of the broken lines. This has not been done yet.

ACKNOWLEDGMENT

The authors are thankful to Dr. Y. Pomeau for many suggestions and comments.

*Present address: Japan Atomic Energy Research Institute, Tokaĭ Research Establishment, Tokaĭ-Mura, Naka-Gun, Ibaraki-Ken, Japan.

[1] E.N. Lorenz, J. Atmos. Sci. 20, 130 (1963).
[2] E.C. Tischmarsh, *Theory of Functions* (Oxford U.P., London and New York, 1939), p. 186.
[3] P.R. Stein and S. Ulam, in *Studies in Mathematical Physics*, edited by A.O. Barut (Reidel, Dordrecht, Holland, 1973), p. 268.
[4] See, for example, Bateman manuscript, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, p. 183.
[5] J. Ibanez and Y. Pomeau, private communication.

Bounds for ladder graph scattering amplitudes on the boundary of their cut*

G. Wanders and E. Morenzoni

Institut de Physique Théorique, Université de Lausanne, Lausanne, Switzerland †
(Received 14 February 1977)

The physical s channel of a ladder graph amplitude contains the cut of this analytic function. Exact bounds for the modulus of the sum of the ladder graph amplitudes of the scalar $g\phi^3$ theory are obtained in this channel. The derivation of these bounds is based on an analytic continuation of the Fourier transforms of the ladder graph amplitudes. The results are compatible with a smooth asymptotic Regge behavior, up to a logarithmic factor.

1. INTRODUCTION, OUTLINE OF THE METHOD AND STATEMENT OF THE RESULTS

At fixed momentum transfer and as a function of energy, a scattering amplitude coincides with the value taken by an analytic function on one of its cuts. This is true for each term of the perturbation series. As the evaluation of an analytic function on the boundary of its analyticity domain may be quite intricate, it is in general not easy to compute explicitly the contribution of a given Feynman diagram to the physical values of a scattering amplitude. For instance, the Regge behavior of the sum $T(s,t)$ of the ladder graphs shown in Fig. 1 can be obtained for $s \to \infty$ along a ray of the complex s plane (args = const $\neq 0$).[1] To our knowledge, the asymptotic behavior of $T(s,t)$ as $s \to \infty$ along the positive real axis has not been properly established.

In a previous work[2] the above-mentioned difficulty was circumvented by using the fact that the ladders of Fig. 1 have only a right-hand cut in the s plane. As the negative real s axis is inside the analyticity domain of $T(s,t)$, it was not too hard to derive upper and lower bounds of this amplitude in its physical u channel, where $s<0$. In fact, the methods developed in Ref. 2 allow the derivation of upper bounds for $|T(s,t)|$ not only on the negative real axis, but also on any ray of the complex s plane. Nontrivial extensions are needed on the right-hand cut $s>4\mu^2$. This article is a sequel to Ref. 2 and is devoted to this more difficult problem of getting informations on the behavior of $T(s,t)$ in its physical s channel. We show how upper bounds of the modulus $|T(s,t)|$ can be obtained there.

Whereas our bounds are valid at all energies, we shall be mainly interested in their implications on the high-energy behavior of $T(s,t)$. The question of this asymptotic behavior is complicated by the fact that the individual ladder $T_n(s,t)$ (Fig. 1) has its cut starting at $s=(n\mu)^2$ and has to be considered as a function of $s_n = s-(n\mu)^2$. For any s larger than $4\mu^2$ there are positive and negative s_n's. If s becomes large, there are always some s_n's which are small compared to s. Therefore, the asymptotic behavior of the sum $T(s,t)=\sum_n T_n(s,t)$ is not the result of a same asymptotic regime setting in for all the T_n's. For this reason, it is not immediately clear that the mass μ of the exchanged particle can be neglected. One may even wonder whether the power behavior $T(s,t)\approx s^\alpha$ obtained in the u channel will survive in the s channel. As a matter of fact, our results support the conjecture that the distribution of the branch points $s=(n\mu)^2$ over the positive real axis does not affect the leading part of the asymptotic behavior of $T(s,t)$. The reason for this is that the main contribution to $T(s,t)$ comes from the T_n's with $n\approx\log s$. The cuts of these T_n's are much below s and cannot prevent a smooth asymptotic behavior.

As in Ref. 2, our main tool will be the Fourier transform of the ladder graph $T_n(s,t)$. The expression of this Fourier transform has its simplest form in the case of forward scattering $t=0$. We restrict ourselves to this case and set $T(s)=T(s,0)$. After a redefinition of the function $f_n(z)$ used in Ref. 2, we write formula (2.3) of this reference as follows:

$$T_n(s) = i16\pi^2\lambda^n \int_0^\infty dz\, f_n(z)\exp(izs_n), \tag{1.1}$$

where $s_n = s-(n\mu)^2$ and $\lambda=(g/4\pi)^2$. If $s_n<0$, $T_n(s)$ is real and the techniques of Ref. 2 lead to upper and lower bounds of this quantity. Therefore, the really new problem in the derivation of a bound for

$$T(s)=\sum_{n=2}^{\infty} T_n(s), \tag{1.2}$$

when $s>4\mu^2$ comes from those terms which have $s_n>0$. We give a short outline of the method we shall apply in the treatment of these terms.

Suppose the Fourier transform $f_n(z)$ has an analytic continuation into a sector $0\leq \arg z \leq \theta$ of the upper half-plane which is suitably bounded. Then, if $s_n>0$, the path of integration in (1.1) can be shifted onto the ray $\arg z=\theta$ and we obtain an upper bound for $|T_n(s)|$ from

$$|T_n(s)| < 16\pi^2\lambda^n \int_0^\infty dr\, |f_n(re^{i\theta})| \exp(-rs_n\sin\theta), \tag{1.3}$$

if $s_n>0$.

All the present work is based on the fact that we succeeded in continuing $f_n(z)$ into a sector of the upper

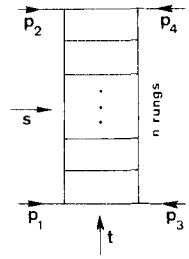


FIG. 1. The ladder graph $T_n(s,t)$. For simplicity, all particles are assumed to have the same mass $m=1$. The variables s and t are defined, as usual, by $s=(p_1+p_2)^2$, $t=(p_1+p_3)^2$.

half-plane and in establishing simple bounds for $|f_n(z)|$ in this sector. The only drawback of our undertaking is that the size of our sector is a function θ_n of n which tends to zero as $n \to \infty$. We do not know whether this decreasing θ_n is due to singularities or bad asymptotic behavior in the upper half-plane or if it is a manifestation of our inability.

The bound for $|f_n(re^{i\theta_n})|$ leads to a bound for $|T_n(s)|$ through (1.3) if $s_n > 0$. From our previous work,[2] we have also a bound for $T_n(s)$ if $s_n < 0$. A combined use of these bounds leads to an upper bound for $|T(s)|$ through (1.2). The s dependence of this bound has the form

$$|T(s)| < \text{const } (\log s)^{1/2} s^{\alpha_s}, \tag{1.4}$$

where

$$\alpha_s = -1 + \lambda. \tag{1.5}$$

This exponent is larger than the exponent α_u obtained in the u channel ($T(s) < \text{const } u^{\alpha_u}$ for $s \to -\infty$). It coincides with α_u in the weak coupling limit. The fact that we do not get $\alpha_s = \alpha_u$ for all coupling constants is no surprise because the majorizations performed here are cruder than those of Ref. 2. The factor $(\log s)^{1/2}$ in (1.4) is produced by our decreasing θ_n. We may also notice that the common weak coupling limit of α_s and α_u is identical with the weak coupling limit of the Regge pole of the ladder graphs obtained by other methods.[1] We conclude from these results that the sum of ladders is likely to exhibit the expected Regge behavior in the physical s channel, up to the uncertainty connected with our $(\log s)^{1/2}$ factor.

Tiktopoulos and Treiman[3] have established bounds for the absorptive part $A(s)$ of $T(s)$ which are consistent with (1.5). As these bounds have the form $A(s) < \text{const } s^{\alpha_s}$, we see that if our $(\log s)^{1/2}$ is really present, it affects only the real part of $T(s)$. Of course, $\mathrm{Re}T(s)$ and $A(s)$ are not independent: They are connected by a dispersion relation. However, it is impossible to get a bound for $\mathrm{Re}T(s)$ from a bound for $A(s)$ through a dispersion relation. Therefore, our results really complete those of Tiktopoulos and Treiman.

2. ANALYTIC CONTINUATION OF THE FOURIER TRANSFORM $f_n(z)$

Our starting point is an expression for the Fourier transform $f_n(z)$ which has been established in Ref. 4. Equation (2.6) of Ref. 4 gives, in the case of forward scattering,

$$f_n(z) = (iz)^{n-1} \int_0^\infty dq_1 \cdots \int_0^\infty dq_{n-1} \int_0^1 d\epsilon_2 \int_0^{\epsilon_2} d\epsilon_3 \cdots \int_0^{\epsilon_{n-2}} d\epsilon_{n-1} \times \prod_{k=1}^{n-1} (\epsilon_k - \epsilon_{k+1}) \int_0^\infty dx\, x^{2n-3} \exp[-iz(Ax^2 + Bx + C)], \tag{2.1}$$

where A, B, and C are positive functions of the ϵ's and q's:

$$A = \sum_{k=1}^{n-1} \left[\epsilon_k^2 q_k + \alpha_k \left(\frac{\alpha_k}{q_k} + \epsilon_k \right) \right],$$
$$B = \sum_{k=1}^{n-1} \left(\epsilon_k q_k + 2 \frac{\alpha_k \beta_k}{qk} + \alpha_k + \beta_k \epsilon_k \right), \tag{2.2}$$
$$C = \sum_{k=1}^{n-1} \left(q_k + \frac{\beta_k^2}{q_k} + \beta_k \right) - n^2 + 1.$$

For simplicity, we have assumed that the masses m and μ are equal and set $m = \mu = 1$. This assumption does not affect the generality of our results. The following notations have been used:

$$\alpha_k = \sum_{j=1}^{k-1} q_j \epsilon_j, \quad \beta_k = 1 + \sum_{j=1}^{k-1} q_j, \quad \epsilon_1 = 1, \ \epsilon_n = 0. \tag{2.3}$$

Whereas the holomorphy of $f_n(z)$ in $\mathrm{Im}z < 0$ is a direct consequence of (2.1), this expression, as it stands, does not allow a continuation of $f_n(z)$ into the upper half-plane. It has to be transformed suitably before such a continuation can be performed. Our aim is to continue $f_n(z)$ into some sector $0 \leq \arg z \leq \theta$ ($\theta < \pi/2$) and to have adequate bounds for $|f_n(z)|$ in this sector. This is achieved if the paths of integrations in (2.1) can be deformed continuously without changing the value of the integrals in such a way that $(Ax^2 + Bx + C)$ moves into the lower half-plane until $-\pi < \arg(Ax^2 + Bx + C) < -\theta$.

Our transformation of (2.1) proceeds as follows. First, the contour of the variable x is displaced onto a ray of the fourth quadrant, for instance $\arg x = -\pi/4$. This brings $(Ax^2 + Bx)$ into the lower half-plane. Then the paths of integration of the q's are deformed in such a way that

$$\arg C = -\theta. \tag{2.4}$$

The effects of these deformations on A and B have to be kept under control in order to be sure that:

$$-\pi < \arg(Ax^2 + Bx) < -\theta. \tag{2.5}$$

Once (2.4) and (2.5) are ensured, $f_n(z)$ can be continued into $0 \leq \arg z \leq \theta$.

As C is a nonlinear function of the q's, the manipulations leading to (2.4) are quite delicate. We shall not go into the details; we shall only show how the limitations of our procedure arise. According to its definition, C can be written

$$C = \sum_{k=1}^{n-1} (n-k) p_k, \tag{2.6}$$

where p_k is a function of the k first q's:

$$p_k = \tfrac{1}{2}[(n-k)\sqrt{q_k} - \beta_k/\sqrt{q_k}]^2. \tag{2.7}$$

Equation (2.6) tells us that a particular way of fulfilling (2.4) is to require

$$\arg p_k = -\theta \text{ for all } p\text{'s}. \tag{2.8}$$

To see if this is possible, we have to find out if the contours of the q's can be deformed continuously so as to bring all the p's onto the rays $\arg p_k = -\theta$. To do this, we have performed the change of variables $q_k \to p_k$ explicitly. An inspection of the singularities in the p's of the resulting integrand shows if it is possible to shift the path of each p_k onto $\arg p_k = -\theta$. This discussion is straightforward but tedious, and we shall not reproduce it. We shall only give a qualitative description in terms of the variables q_k.

Suppose we try to shift successively the p_k's onto $\arg p_k = -\theta$, starting with the last one, p_{n-1}. As q_{n-1} appears only in p_{n-1}, it is easy to bring p_{n-1} onto $\arg p_{n-1} = -\theta$ by deforming the path of q_{n-1}, all the other q_k's remaining on the positive real axis. For moderate

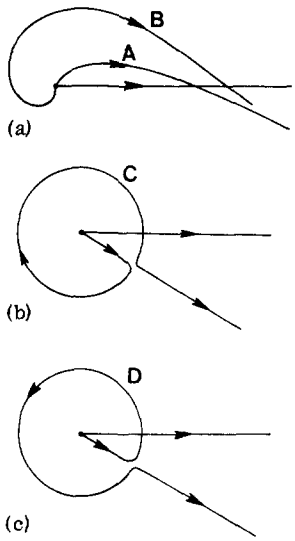


FIG. 2. The paths of integration C_{n-1} of the variable q_{n-1}. As $\arg\beta_{n-1}$ goes through increasing positive values, the positive real axis is deformed continuously into curves of the form A and B of Fig. 2(a). For $\arg\beta_{n-1}$, just below $(\pi-\theta)$ one gets the path C shown in Fig. 2(b) and as $\arg\beta_{n-1}$ goes through $(\pi-\theta)$ this path changes discontinuously into the curve D of Fig. 2(c). The difference between the integrals along C and D is twice the integral along the circle $|q_{n-1}| = |\beta_{n-1}|/(n-k)$; this integral is finite.

values of θ, the resulting path for p_{n-1} has the form A of Fig. 2a. In a second step we shift p_{n-2} onto $\arg p_{n-2} = -\theta$; to do this, we deform the contour of q_{n-2}. However, as q_{n-2} appears in β_{n-1}, the contour of q_{n-1} has to be readjusted in such a way that p_{n-1} remains on $\arg p_{n-1} = -\theta$. We may repeat this procedure step by step, deforming the contour of q_{n-k} and readjusting the contours of $q_{n-k-1},\ldots,q_{n-1}$ in the kth step.

The question is to know if these readjustments are always possible. Consider the case of q_{n-1}. The path C_{n-1} of q_{n-1} is entirely determined by the requirement $\arg p_{n-1} = -\theta$ and Eq. (2.7): It depends on the complex values of $q_1,\ldots,q_{n-1}$ on their respective paths through β_{n-1}. It turns out that there is a problem only if $\arg\beta_{n-1} > 0$. For moderate values of $\arg\beta_{n-1}$, C_{n-1} has the form A and B of Fig. 2. This form changes continuously if $\arg\beta_{n-1}$ increases until $\arg\beta_{n-1} = \pi - \theta$. If $\arg\beta_{n-1}$ happens to cross $(\pi-\theta)$ the form of C_{n-1} changes discontinuously from that of Fig. 2b to that of Fig. 2c. Due to the singularities of the integrand at $q=0$ and $q=\infty$, the integral along the contour 2c does not give the same result as the integral along 2b. Therefore, our procedure works only as long as the resulting complex values of β_{n-1} have arguments which do not exceed $(\pi-\theta)$.

A detailed study shows that $\arg\beta_{n-1}$ can indeed become positive and that it cannot be kept smaller than $(\pi-\theta)$ for large n if θ is a fixed constant. However, if a convenient n-dependence is introduced, one may ensure $\arg\beta_{n-1} < (\pi-\theta_n)$ and one can deform the paths of integrations in such a way that (2.5) as well as (2.4) are fulfilled. An adequate n-dependence of θ is

$$\theta \to \theta_n = \beta/n^\gamma \tag{2.9}$$

with $\gamma = 3\sqrt{3}/4\pi$ and $\beta < \pi/16$. Therefore, $f_n(z)$ *can be continued analytically into the sector* $0 \le \arg z \le \theta_n$, θ_n *being defined by* (2.9).

Our next task is to derive an upper bound of $|f_n(re^{i\theta})|$ for $0 \le \theta \le \theta_n$. According to our preceding discussion,

$$|f_n(re^{i\theta})| < r^{n-1}\int_0^1 d\epsilon_2\int_0^{\epsilon_2} d\epsilon_3\cdots\prod_{k=1}^{n-1}(\epsilon_k-\epsilon_{k+1})$$
$$\times\prod_{j=1}^{n-1}\int_{C_j}|dq_j|\int_0^\infty d|x|\;|x|^{2n-3}$$
$$\times\exp[r\,\mathrm{Im}(\exp(i\theta)(Ax^2+Bx+C))]. \tag{2.10}$$

The contours C_j join 0 to ∞, they are localized in known sectors:

$$q_j \in C_j \to -\theta_n < \arg q_j < \omega_j - \theta_n,\ \omega_j = \mathrm{const}\,\frac{1}{(n-j)^\gamma}. \tag{2.11}$$

Furthermore, the form of these contours are such that on C_j

$$|dq_j| < [1/\cos(\tfrac{1}{4}\omega_j)]d|q_j|. \tag{2.12}$$

The imaginary part in the exponential of the integrand is a sum of negative terms. Dropping some of them produces a majorization:

$$\mathrm{Im}[\exp(i\theta)(Ax^2+Bx+C)] < \sum_{k=1}^{n-1}\mathrm{Im}[\exp(i\theta)[x\epsilon_k(1+x\epsilon_k)q_k + x\epsilon_k]] < -\sum_{k=1}^{n-1}\{|x|\epsilon_k[|x|\epsilon_k\cos\omega_k + \sin(\pi/4-\omega_k)]|q_k| + \sin(\pi/4-\theta_n)|x|\epsilon_k\}. \tag{2.13}$$

The definitions (2.2), the inequalities (2.11) and the facts that $\theta \le \theta_n$ and $\arg x = -\pi/4$ have been taken into account.

After insertion of the bounds (2.12) and (2.13) into (2.10), the integrations over the $|q|$'s are trivial and one finds

$$|f_n(re^{i\theta})| < \frac{K_n}{(n-1)!}\left\{\int_0^\infty dx\,\frac{\exp[-rx\sin(\pi/4-\theta_n)]}{x+\nu_n}\right\}^{n-1} \tag{2.14}$$

with

$$\nu_n = \inf_{1\le k\le n-1}\frac{\sin(\pi/4-\omega_k)}{\cos\omega_k}$$
$$K_n = \prod_{k=1}^{n-1}\frac{1}{\cos(\tfrac{1}{4}\omega_k)\cos\omega_k}. \tag{2.15}$$

It turns out that the bound for $|T(s)|$ which emerges from (2.14) has its best behavior at large s if K_n is bounded for all n. Using (2.11), one finds

$$K_n < \exp[\mathrm{const}\,\beta^2 n^{1-2\gamma}]. \tag{2.16}$$

Therefore, all the K_n's have a common upper bound if β is made dependent on n:

$$\beta \to \frac{\pi}{16}\,\frac{1}{n^{1/2-\gamma}}. \tag{2.17}$$

Combining this with (2.9), we get our final width of the sector in which we continue $f_n(z)$:

$$\theta_n = \frac{\pi}{16}\,\frac{1}{n^{1/2}}. \tag{2.18}$$

Simple bounds for the integral appearing in (2.14) give

$$|f_n(re^{i\theta})| < \begin{cases}\dfrac{K}{(n-1)!}\left(\dfrac{r_0-1}{r}\right)^{n-1} & \text{if } r \ge 1,\\[2ex] \dfrac{K}{(n-1)!}\left(\log\dfrac{r_0}{r}\right)^{n-1} & \text{if } r<1.\end{cases} \tag{2.19}$$

Here

$$r_0 = \sup_{n\geq 2}\left[1+\frac{1}{\nu_n \sin(\pi/4+\theta_n)}\right] < 10. \tag{2.20}$$

The bound (2.19) is valid if $0 \leq \theta \leq \theta_n$, θ_n being defined in (2.18).

3. BOUNDS FOR THE LADDER GRAPHS

The ladder graph amplitude $T_n(s)$ depends on s through $s_n = s - n^2$ and the bound we shall establish has a different form if n is such that $s_n \geq 0$ or $s_n < 0$. We consider the case $s_n \geq 0$ first; it is precisely for this case that the machinery of the last section has been developed.

According to the results of Sec. 2, we are allowed to shift the path of integration in (1.1) onto $\arg z = \theta_n$ if $s_n \geq 0$. This gives

$$|T_n(s)| < 16\pi^2\lambda^n \int_0^\infty dr\, |f_n(re^{i\theta_n})| \exp(-rs_n \sin\theta_n). \tag{3.1}$$

Inserting the bounds (2.19) into this inequality, we get a sum of two integrals whose evaluation is sketched in the Appendix. The simple bounds which are obtained there have different expressions whether the quantity $(s_n \sin\theta_n)$ is larger or smaller than 1. It is readily seen that it exceeds 1 for nearly all allowed values of n if s is sufficiently large. Let $N = [\sqrt{s}\,]$ be the largest integer which is smaller than or equal to $\sqrt{s}$; the inequality $s_n \geq 0$ is equivalent to $n \leq N$. One finds

$$s_n \sin\theta_n > 1 \quad \text{for } n \leq N-1 \text{ if } s > 36.5.$$

The remaining $s_N \sin\theta_N$ can be larger or smaller than 1. Equations (3.1), (2.19), (A3) and (A4) lead to

$$|T_n(s)| < \text{const}\,\frac{\sqrt{n}}{s_n}\lambda^n\left[\sum_{k=0}^{n-1}\frac{1}{k!}\left(1+\frac{1}{e}\delta_{k,n-1}\right)\left(\log\frac{r_0 s_n}{\sqrt{n}}\right)^k + \frac{(r_0-1)^{n-1}}{(n-1)!}\right]. \tag{3.2}$$

This inequality holds if $n \leq N-1$; it holds also for $n = N$ if $s_N \sin\theta_N > 1$. If this last quantity is smaller than 1, one gets an upper bound for $|T_N(s)|$ from (3.2) by replacing the quantity $|s_N/\sqrt{N}|$ by 1.

If $s_n < 0$, or $n \geq N+1$, we apply the techniques of Ref. 2. The Fourier transform $f_n(z)$ being regular in the lower half-plane, the contour of integration in (1.1) can be shifted onto the negative imaginary axis

$$T_n(s) = 16\pi^2\lambda^n \int_0^\infty dz f_n(-iz)\exp(zs_n). \tag{3.3}$$

The definition (2.1) leads directly to an expression for $f_n(-iz)$. Performing a majorization similar to (2.13), one gets

$$f_n(-iz) < z^{n-1}\int_0^\infty dq_1 \cdots \int_0^\infty dq_{n-1}\int_0^1 d\epsilon_2 \cdots \int_0^{\epsilon_{n-2}} d\epsilon_{n-1} \times \epsilon_2 \cdots \epsilon_{n-1}\int_0^\infty dx\, x^{2n-3} \times \exp\left(-z\sum_{k=1}^{n-1}[x\epsilon_k(x\epsilon_k+1)q_k + x\epsilon_k]\right). \tag{3.4}$$

After integration over the q's, one obtains, in analogy to (2.19),

$$f_n(-iz) < \begin{cases} \dfrac{1}{(n-1)!} & \text{if } z \geq 1, \\ \dfrac{1}{z^{n-1}} & \\ \dfrac{1}{(n-1)!}\left(\log\dfrac{2}{z}\right)^{n-1} & \text{if } 0 \leq z < 1. \end{cases} \tag{3.5}$$

The expression (3.3) and the bounds (3.5) give

$$T_n(s) < \text{const}\,\frac{\lambda^n}{|s_n|}\left[\sum_{k=0}^{n-1}\frac{1}{k!}\left(1+\frac{1}{e}\delta_{k,n-1}\right)(\log(2|s_n|))^k + \frac{1}{(n-1)!}\right] \tag{3.6}$$

for $n \geq N+2$. If $n = N+1$, this inequality is also valid as it stands as long as $|s_{N+1}| \geq 1$. If $|s_{N+1}| < 1$, (3.6) gives an upper bound for $T_{N+1}(s)$ once s_{N+1} has been replaced by 1.

Finally, we have to evaluate the sums of the bounds (3.2) and (3.6). For simplicity we assume s such that both $s_N \sin\theta_N$ and $|S_{N+1}|$ are larger than. 1. We write

$$T(s) = T^{(1)}(s) + T^{(2)}(s) \tag{3.7}$$

with

$$T^{(1)}(s) = \sum_{n=2}^{N} T_n(s), \quad T^{(2)}(s) = \sum_{n=N+1}^{\infty} T_n(s). \tag{3.8}$$

An upper bound for $|T^{(1)}(s)|$ is obtained from (3.2). This bound contains a double sum $\tilde{T}(s)$ over n and k:

$$\tilde{T}(s) = \text{const}\sum_{n=2}^{N}\lambda^n\frac{\sqrt{n}}{s_n}\sum_{k=0}^{n-1}\frac{1}{k!}\left(\log\left(\frac{r_0 s_n}{\sqrt{n}}\right)\right)^k. \tag{3.9}$$

It is this double sum which provides the relevant s dependence of our bound. We show how it can be estimated. We assume $s > 36.5$; this ensures $\sqrt{s} > \log(r_0 s) > 1$. We decompose $\tilde{T}(s)$ into two parts; the sum $\tilde{T}_1(s)$ of terms which have $n \leq \log(r_0 s)+1$ and the sum $\tilde{T}_2(s)$ of the remaining terms, $n > \log(r_0 s)+1$.

In $\tilde{T}_1(s)$ we have $(\sqrt{n}/s_n) < (2(2\log r_0 s)^{1/2}/s)$ and $(s_n/\sqrt{n}) < s$. This gives

$$\tilde{T}_1(s) < \text{const}\,\frac{(\log(r_0 s))^{1/2}}{s}\sum_{n\leq \log(r_0 s)+1}\lambda^n\sum_{k=0}^{n-1}\frac{1}{k!}(\log(r_0 s))^k < \text{const}\,\frac{\lambda}{1-\lambda}(\log(r_0 s))^{1/2}s^{\lambda-1}. \tag{3.10}$$

The last inequality is valid for $\lambda < 1$. As the bound for $|T(s)|$ we are constructing turns out to diverge if λ exceeds 1, there is no need for a bound of $\tilde{T}_1(s)$ holding if $\lambda > 1$.

$\tilde{T}_2(s)$ is majorized in the following way, assuming $\lambda < 1$:

$$\tilde{T}_2(s) < \text{const}\sum_{n>\log(r_0 s)+1}\frac{\lambda^n\sqrt{n}}{s_n}\sum_{k=0}^{n-1}\frac{1}{k!}\left(\log\frac{r_0 s_n}{\sqrt{n}}\right)^k < \text{const}\,\frac{\lambda}{1-\lambda}\lambda^{\log(r_0 s)} < \text{const}\,\frac{\lambda}{1-\lambda}s^{-1+\lambda}. \tag{3.11}$$

After examination of the remaining simple sums contained in $T^{(1)}(s)$, one finds that we may write a bound for $|T^{(1)}(s)|$ which has the form (3.10):

$$|T^{(1)}(s)| < \text{const}\,\frac{\lambda}{1-\lambda}(\log(r_0 s))^{1/2}s^{\lambda-1}. \tag{3.12}$$

Our upper bound for $|T^{(2)}(s)|$ is obtained from (3.6). Here again we meet a double sum; it is this double sum which diverges if $\lambda > 1$. We perform the following majorization:

$$\sum_{n=N+1}^{\infty} \frac{\lambda^n}{|s_n|} \sum_{k=0}^{n-1} \frac{(\log(2|s_n|))^k}{k!} < \frac{\lambda}{1-\lambda} \lambda^N. \tag{3.13}$$

According to our assumptions $N > (\log r_0 s)$. This implies that the right-hand side of (3.13) is majorized by an expression which has the same form as (3.11). In fact, this is true for all terms of $T^{(2)}(s)$. Therefore,

$$T^{(2)}(s) < \text{const} \frac{\lambda}{1-\lambda} s^{-1+\lambda}. \tag{3.14}$$

Combining (3.12) and (3.14), we get our final result:

$$|T(s)| < \text{const} \frac{\lambda}{1-\lambda} (\log(r_0 s))^{1/2} s^{-1+\lambda}. \tag{3.15}$$

It is easily seen that (3.15) is still valid if $0 < s_N \sin\theta_N < 1$ or $0 > s_{N+1} > -1$. Therefore, the only limitation to the validity of (3.15) is

$$s > 36.5.$$

The divergence appearing in the bound (3.15) as $\lambda \to 1$ has already been encountered and commented in Ref. 2. The s dependence of (3.15) is discussed in Sec. 1.

APPENDIX

The two integrals we have to estimate are

$$I_1(x) = \int_0^1 dr \left(\log \frac{r_0}{r}\right)^{n-1} \exp(-rx) \tag{A1}$$

and

$$I_2(x) = \int_0^{\infty} dr \frac{1}{r^{n-1}} \exp(-rx). \tag{A2}$$

If $x > 1$, we transform I_1 as follows:

$$I_1(x) = \frac{1}{x} \int_0^x dr \left(\log \frac{xr_0}{r}\right)^{n-1} \exp(-r) < \frac{1}{x} \int_0^1 dr \left(\log \frac{xr_0}{r}\right)^{n-1} + \frac{(\log(xr_0))^{n-1}}{x} \int_1^x dr e^{-r} < \frac{(n-1)!}{x} \sum_{k=0}^{n-1} \frac{1}{k!} (\log(xr_0))^k + \frac{1}{e} \frac{(\log(xr_0))^{n-1}}{x}. \tag{A3}$$

If $x < 1$, we write

$$I_1(x) < \frac{1}{x} \int_0^x dr \left(\log \frac{xr_0}{r}\right)^{n-1} = (n-1)! \sum_{k=0}^{n-1} \frac{1}{k!} (\log(xr_0))^k. \tag{A4}$$

For $I_2(x)$, we use

$$I_2(x) < \int_1^{\infty} dr e^{-rx} = \frac{1}{x} \exp(-x) \tag{A5}$$

if $x > 1$ and

$$I_2(x) < \int_1^{\infty} dr \frac{1}{r^{n-1}} < 1 \tag{A6}$$

if $x < 1$ and $n \geq 3$. It is not necessary to discuss the case $n = 2$ because we are interested in large values of s. In our applications, $x = s_n \sin\theta_n$, and the situation $x < 1$ never appears for $n = 2$ if $s_2 \sin\theta_2 > 1$, i.e., $s > 11.3$.

*Work supported in part by the Fonds National Suisse de la Recherche Scientifique.

†Postal address: Bâtiment des Sciences Physiques, Dorigny, CH-1015 Lausanne (Switzerland).

[1]J.C. Polkinghorne, J. Math. Phys. **4**, 503 (1963); **5**, 431 (1964).
[2]G. Wanders, Phys. Lett. B **58**, 191 (1975); Helv. Physica Acta **49**, 415 (1976).
[3]G. Tiktopoulos and S. B. Treiman, Phys. Rev. **135**, B711 (1964); **137**, B1597 (1965).
[4]G. Wanders, Nuovo Cimento A **56**, 144 (1960).

King and domino polynomials for polyomino graphs

Akiko Motoyama and Haruo Hosoya*

Department of Chemistry, Ochanomizu University, Bunkyo-Ku, Tokyo 112, Japan
and The Institute for Molecular Science, Myohdaiji-machi, Okazaki, Aichi 444, Japan
(Received 14 December 1976)

For the purposes of treating several enumeration problems of lattice dynamics, king and domino polynomials are defined for a chessboard, polyomino, or square lattice of arbitrary size and shape. These polynomials are shown to be closely related to the partition function of the dimer statistics, the number of Kekulé structures, or maximum matching number. Several recursion formulas are found. Interpretation of these newly proposed quantities is given, and the possibility of extending them to the important physical models is discussed.

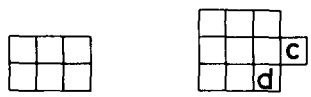
INTRODUCTION

Although the rook polynomial has been known as an effective means of considering the enumeration problems of chess and Japanese shohgi games, it apparently has not been utilized in the mathematics of the important models of physics and chemistry.[1,2]

We have proposed the king polynomial for enumerating the number of ways for putting nontaking kings on a given polyomino, square animal, or chessboard of an arbitrary size and shape. Interesting mathematical features including recursion formulas were found. It turned out that the king polynomial is closely related to several important enumeration problems, i.e., the number of maximum matching over a square lattice graph (paving dominoes), the partition function for the magnetic properties of transition metal crystal (dimer statistics),[3] the kinetics of adsorption of molecular oxygens onto a metal surface,[4] and the stabilities of unsaturated hydrocarbon molecules (Kekulé structures).[5,6]

1. POLYOMINOES AND SUBGRAPHS

The graphs with which the present paper is concerned are polyominoes (square animals) and their subgraphs derived by deleting squares and/or edges. A polyomino which is generated by the stacking of squares (hereafter called as cells) of equal size may also be called a chessboard although its size and shape are varied as graphs I and II.



I II

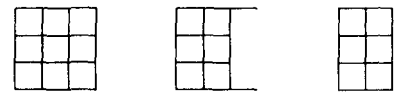
With respect to deletion of a particular cell c in graph G several subgraphs are defined as follows:

$G-c$ is obtained from G by deleting cell c, namely, by deleting those edges of c which belong only to c.

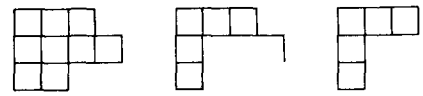
$G\ominus c$ is obtained from $G-c$ by deleting all the edges which were contained or incident to cell c in G.[7]

$G\boxminus c$ is obtained from $G\ominus c$ by deleting all the branches (let us call this process truncation of $G\ominus c$), or from G by deleting cell c together with all the adjacent cells.[8]

Examples of these subgraphs are shown for graph II and cells c and d.

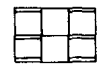


II$-c$ II$\ominus c$ II$\boxminus c$



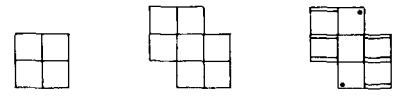
II$-d$ II$\ominus d$ II$\boxminus d$

Polyominoes can be classified into even and odd ones according to the number of the vertices but not that of the cells (see Table I). Graph III is odd. We call an odd polyomino as type D. An even graph with N ($=2m$) vertices is called a Kekulé graph if one can choose such a set of m disjoint edges that span all the N vertices as in I.a. We call this pattern a "Kekulé pattern.[9]" In a later



I.a

discussion Kekulé graphs will be classified into types A and B. A non-Kekulé polyomino is called a type C. Graphs I, II, and all the subgraphs of II explained above are Kekulé graphs, whereas even polyomino IV is type C (see IV.a). Note that a rectangular graph $[m,n]$[10] $\equiv[n,m]$ is a Kekulé graph unless both m and n are even.



III IV IV.a

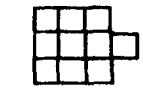
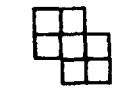
2. KING PATTERN AND KING POLYNOMIAL

A king can take on any of the eight neighboring cells. A number of distinct patterns as I.b, called by "king pattern,"

I.b

can be obtained by putting k nontaking kings (depicted

TABLE I. Classification of polyomino graphs.

		Example	Type	Pattern: King	Pattern: Domino
polyomino: even: Kekulé	$K_G(X) = D_G(X)$		A	○	○
polyomino: even: Kekulé	$K_G(X) \neq D_G(X)$		B	○	○
polyomino: even: non-Kekulé			C	○	×[a]
polyomino: odd			D	○	×

[a] × means that the domino polynomial is not defined.

with circles) for a given graph. Discussions in this section are applied to all types of polyominoes.

Define the nontaking number $r(G,k)$ as

$$r(G,k) = \begin{pmatrix} \text{number of king patterns with } k \\ \text{nontaking kings for graph } G \end{pmatrix}. \tag{1}$$

For the sake of the later discussion let us define

$$r(G,0) = 1 \tag{2}$$

for all the cases. This means that a polyomino graph itself can be interpreted as the king pattern with $k = 0$ (see Fig. 1).[11]

The king polynomial $K_G(X)$ for polyomino G is defined as

$$K_G(X) = \sum_{k=0}^{k_m} r(G,k) X^k, \tag{3}$$

where k_m is the maximum number for k, which does not exceed a quarter of the number of the vertices in the graph. For graphs I and II we have

$$K_I(X) = 1 + 6X + 4X^2,$$

$$K_{II}(X) = 1 + 10X + 22X^2 + 12X^3 + X^4.$$

Note that

$$K_G(1) = \sum_{k=0}^{k_m} r(G,k) = \begin{pmatrix} \text{number of the king} \\ \text{patterns for } G \end{pmatrix}. \tag{4}$$

Let g_c be the number of king patterns with a circle (king) assigned to a particular cell c. Then the total number $C(G)$ of the circles in the set of $K_G(1)$ king patterns for graph G with l cells is

$$C(G) = \sum_{c}^{l} g_c. \tag{5}$$

The number g_c is interpreted as a weight of cell c in the king pattern enumeration. Since each of the $r(G,k)$ king patterns with a given value k has k circles, $C(G)$ is expressed by

$$C(G) = \sum_{k=1}^{k_m} k \cdot r(G,k) = K_G'(1), \tag{6}$$

or

$$\sum_{c}^{l} g_c = K_G'(1), \tag{6'}$$

where K_G' is the first derivative of the king polynomial (3) with respect to X. These relations are illustrated below with graph I as an example.

3	1	3
3	1	3

I.c

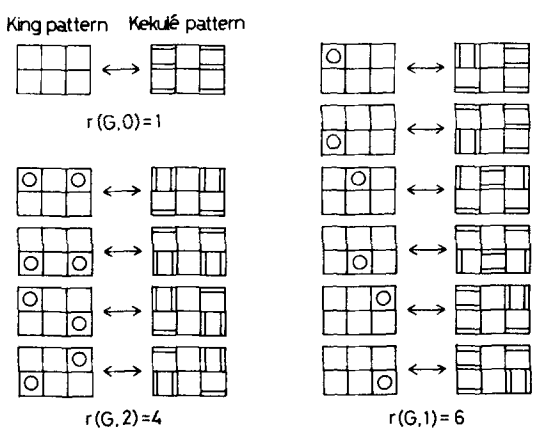


FIG. 1. One-to-one correspondence between the king and Kekulé patterns. The nontaking number $r(G,k)$ is also explained with G as I.

$$K'_{\Gamma}(X) = 6 + 8X,$$

$$C(\mathrm{I}) = 4 \times 3 + 2 \times 1 \quad [\text{from Eq. (5)}]$$
$$= 6 + 8 \quad [\text{from Eq. (6)}]$$
$$= 14.$$

It is evident from the definition of the king pattern that the g_c number is equal to the number of the king patterns of graph $G \boxminus c$, namely,

$$g_c = K_{G \boxminus c}(1). \tag{7}$$

By using the inclusion–exclusion principle the king polynomial for graph G can be obtained from those of the subgraphs of G as

$$K_G(X) = K_{G-c}(X) + X \cdot K_{G \boxminus c}(X). \tag{8}$$

In Tables II and III several examples of the king polynomials are given. Extensive tabulations of the king polynomials for smaller polyominoes and for typical series of polyominoes are given elsewhere.[14]

For a series of rectangular graphs $[m,n]$ the following recursive relations are obtained:

$$K_{[1,n]}(X) = K_{[1,n-1]}(X) + X \cdot K_{[1,n-2]}(X) \quad (n \geq 0), \tag{9}$$

$$K_{[2,n]}(X) = K_{[2,n-1]}(X) + 2X \cdot K_{[2,n-2]}(X) \quad (n \geq 0), \tag{10}$$

TABLE II. King polynomials and the number of the Kekulé patterns for several polyomino graphs.

G	$r(G,k)$ $k=0$	1	2	3	4	$K_G(1)$	$K(G)$
	1	1				2	2
	1	2				3	3
	1	3	1			5	5
	1	4	3			8	8
	1	4	2			7	7
	1	6	4			11	11
 [a]	1	9	16	8	1	35	36
 [b]	1	4	1			6	5

[a] Discrepancy between $K_G(1)$ and $K(G)$ is due to a degenerate king pattern. The domino polynomial is $D_G(X) = 1 + 10X + 16X^2 + 8X^3 + X^4$ and $D_G(1) = 36$.

[b] Discrepancy between $K_G(1)$ and $K(G)$ is due to an improper king pattern which has two kings at the terminal cells, $D_G(X) = 1 + 4X$ and $D_G(1) = 5$.

$$K_{[3,n]}(X) = (1+X)K_{[3,n-1]}(X) + (2X+X^2)K_{[3,n-2]}(X) - (X^2+X^3)K_{[3,n-3]}(X) \quad (n \geq 1), \tag{11}$$

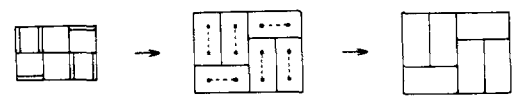
$$K_{[4,n]}(X) = (1+X)K_{[4,n-1]}(X) + (3X+4X^2)K_{[4,n-2]}(X) + (X^2-3X^3)K_{[4,n-3]}(X) - 3X^4 K_{[4,n-4]}(X) \quad (n \geq 2). \tag{12}$$

It is assumed here that $K_{[0,n]}(X) = K_{[-1,n]}(X) = 1$ and $K_{[-2,n]}(X) = 0$ for all $n \geq 0$. A general expression for $K_{[m,n]}(X)$ is not yet obtained.

3. KEKULÉ PATTERN AND DOMINO PATTERN

As will be explained later, one can make a one-to-one correspondence between the king and Kekulé patterns, as I.b and I.d. Further, if one puts a 1×2 rectangle (domino) onto each "double bond" of the Kekulé pattern, I.d, a "domino pattern," I.e., is obtained as shown below. The problem of enumerating the

I.b
king pattern



I.d
Kekulé pattern

I.e.
domino pattern

number of ways for "paving dominoes," "paving tatamis,[15]" or maximum matching is thus equivalent to the counting of the Kekulé patterns, which, together with king patterns, will mainly be the concern of this paper.

It is interesting to note that for a group of graphs there seems to be a one-to-one correspondence between the king and Kekulé patterns as seen in Fig. 1 with graph I as an example. This means

$$K_G(1) = K(G) \quad (G \in \text{type A}), \tag{13}$$

where $K(G)$ denotes the number of the Kekulé (domino) patterns or, in other words, the number of maximum matching for graph G. This is almost all the cases for smaller polyominoes as in Table II. Let us call a Kekulé graph which has property (13) type A, and call other Kekulé graphs type B (see Table I). Of the set of the rectangular graphs, the $[1,n]$ and $[2,3]$ graphs belong to type A, the graphs with even m and n belong to type D, and all others belong to type B.

Among a set of the Kekulé patterns for a given Kekulé graph of type A one can choose a standard pattern in which the largest number of horizontal (or vertical) double bonds are chosen as in I.a. By rotating a set of two or more double bonds circularly arranged one can get all other Kekulé patterns. For a rectangular graph as I one can make a one-to-one correspondence between the king and Kekulé patterns as in Fig. 1. The one-to-one correspondence (13) is not clearer but can be attained for other classes of Kekulé graphs of type A.

TABLE III. King and domino polynomials for rectangular graphs.[a]

		$r(G,k)$[d]							
G[b]	$k=0$	1	2	3	4	5	6	$K_G(1)$	$D_G(1)=K(G)$
[2,2]*[c]	1	4						5	—
[2,3]	1	6	4					11	11
[2,4]*	1	8	12					21	—
[2,5]	1	10	24 (22)	8				43	41
[2,6]*	1	12	40	32				85	—
[3,3]	1	9 (10)	16	8	1			35	36
[3,4]	1	12 (14)	37	34	9			93	95
[3,5]	1	15 (18)	67 (74)	105 (107)	65	15	1	269	281
[4,4]*	1	16	78	140	79			314	—
[4,5]	1	20 (26)	135 (155)	382 (378)	454 (410)	194 (186)	27	1213	1183

[a] For graphs $[1,n]$, see Table II.
[b] $[m,n]$ is an $m \times n$ rectangular polyomino.
[c] Graphs with an asterisk are odd and their domino polynomials are not defined.
[d] For the domino polynomial, only those coefficients which are different from the values of the corresponding king polynomial are given in parentheses.

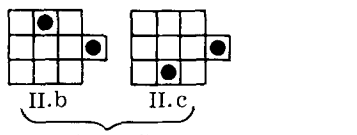
4. DOMINO POLYNOMIAL

Relation (13) does not hold in the Kekulé graphs of type B. For example, the number of the king patterns $K_{II}(1)=46$ exceeds the number of the Kekulé (domino) patterns $K(II)=42$. By taking II.a as the standard Kekulé pattern and by rotating sets of double bonds as in the

II.a

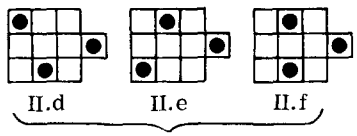
case of graph I, each king pattern finds its own counterpart in the set of the Kekulé patterns except in the following two cases.

Five king patterns II.b–II.f do not have their counterparts



II.b II.c

$(-2X^2)$

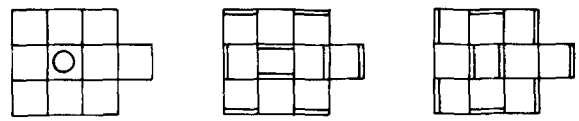


II.d II.e II.f

$(-3X^3)$

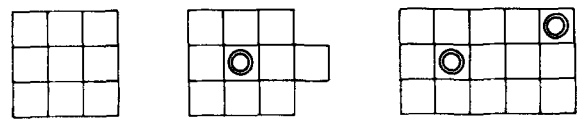
in the Kekulé patterns. We call them "improper (king) patterns" and darken the circles therein. In order to extend the applicability of relation (13) towards type B polyominoes, the terms arising from the solid circles should be subtracted from $K_G(X)$ as shown below patterns II.b–II.f.

Now for graph II we are left with one king pattern II.g and two Kekulé patterns II.h and II.i. Namely, two Kekulé



II.g II.h II.i

patterns correspond to one king pattern. It was found that this is always the case in which the rectangular graph [3, 3] or graph V is contained as a subunit of a Kekulé pattern.[16] We call a king pattern II.g which has a circle in the center of the rectangle [3, 3] a "degenerate (king) pattern"; it is depicted in II.j and VII. The correction terms for $K_G(X)$ should



V II.j VII

$(+X)$ $(+X^2)$

be as shown below these patterns, i.e., in a correction term a double circle acts just as a circle in the counting of $K_G(X)$. As the size of a graph increases additional corrections would be necessary for condensed degenerate patterns like VIII and multidegenerate patterns like IX, which we are not going to discuss in detail.

VIII IX

$(+X^2)$ $(+2X)$

Thus the domino polynomial $D_G(X)$ can be defined as

$$D_G(X)=K_G(X)-\begin{pmatrix}\text{correction terms from}\\ \text{improper king patterns}\end{pmatrix}+\begin{pmatrix}\text{correction terms from}\\ \text{degenerate king patterns}\end{pmatrix}=K_G(X)+\Delta_G(X). \tag{14}$$

For graph II the domino polynomial is obtained as

$$K_{II}(X)=1+10X+22X^2+12X^3+X^4 \quad \text{king polynomial}$$

$$\Delta_{II}(X)=\begin{cases}+\ X & \text{degenerate king pattern}\\ -2X^2-3X^3 & \text{improper king patterns (+}\end{cases}$$

$$D_{II}(X)=1+11X+20X^2+9X^3+X^4 \quad \text{domino polynomial}$$

By definition we have

$$D_G(1)=K(G) \quad (G\in\text{type A, B}). \tag{15}$$

A type A graph does not have any correction term in Eq. (14), and we have

$$D_G(X)=K_G(X) \quad (G\in\text{type A}). \tag{16}$$

Recursion formulas are also found for the domino polynomial but are not given here,[17] as they are not as simple as that of the king polynomial. Some special recursion formulas are given here for rectangular graphs:

$$D_{[1,n]}(X)=D_{[1,n-1]}(X)+X\cdot D_{[1,n-2]}(X) \quad (n\geq 0), \tag{17}$$

$$D_{[2,2n+1]}(X)=(1+3X)D_{[2,2n-1]}(X)+(X-2X^2)D_{[2,2n-3]}(X) \quad (n\geq 1) \tag{18}$$

$$D_{[3,n]}(X)=(1+X)D_{[3,n-1]}(X)+(3X+X^2)D_{[3,n-2]}(X)-(X+2X^2+X^3)D_{[3,n-3]}(X)+(X-2X^2-X^3)\times D_{[3,n-4]}(X)+(-X^2+X^3+X^4)D_{[3,n-5]}(X) \quad (n\geq 3). \tag{19}$$

It is assumed here that $D_{[0,n]}(X)=D_{[-1,n]}(X)=1$ and $D_{[-2,n]}(X)=0$ for all $n\geq 0$. Expression (19) can be shown not to be factorized. The domino polynomials for smaller polyominoes are given elsewhere.[14]

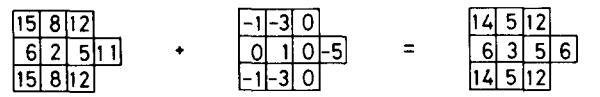
5. WEIGHT OF THE CELLS

As implied in Eq. (5), the g_c number, which is obtained from Eq. (7), is a weight of a particular cell c in the king pattern enumeration. An example is given for graph I as in I.c. Since relation (16) holds for a polyomino of type A, g_c is also a weight of cell c in the Kekulé (domino) pattern enumeration.

For a polyomino of type B, discrepancy arises between the king and domino polynomials, and also between the contribution of a particular cell in the enumeration of the numbers of king and Kekulé patterns. The former discrepancy could be remedied by the introduction of improper and degenerate patterns. The latter correction is made as follows. Let i_c be the number of improper king patterns with a (solid) cell c, and let d_c be that of degenerate patterns with a double circle. Then for each cell of a polyomino G of type B we have

$$g_c-i_c+d_c=K(G\ominus c), \tag{20}$$

where $K(G\ominus c)$ is the number of Kekulé patterns of subgraph $G\ominus c$. Relation (20) is shown below with graph II as an example.



$$\sum_c^I g_c=94 \qquad \sum_c^I (d_c-i_c)=-12 \qquad \sum_c^I K(G\ominus c)=82$$

By analogy with the derivation of Eq. (6′) the following relation is deduced from the procedure for getting the correction term $\Delta_G(X)$:

$$\sum_c^I (d_c-i_c)=\Delta_G'(1). \tag{21}$$

Recall the definition of $D_G(X)$,

$$D_G(X)=K_G(X)+\Delta_G(X) \tag{14}$$

and we have

$$D_G'(X)=K_G'(X)+\Delta_G'(X), \tag{22}$$

which gives

$$D_G'(1)=K_G'(1)+\Delta_G'(1) \tag{23}$$

$$=\sum_c^I (g_c+d_c-i_c), \quad \text{from Eqs. (6′) and (21)}$$

$$=\sum_c^I K(G\ominus c), \quad \text{from Eq. (20).} \tag{24}$$

In the above example we have

$$K_{II}'(X)=10+44X+36X^2+4X^3,$$

$$\Delta_{II}'(X)=1-4X-9X^2,$$

and

$$D_{II}'(X)=11+40X+27X^2+4X^3,$$

from which the numbers 94, −12, and 82 are obtained by the use of Eqs. (6′), (21), and (24), respectively.

6. DISCUSSION

The maximum matching problem over a square lattice, the domino paving problem, and the enumeration of the Kekulé structures are mathematically equivalent to the enumeration of the partition function of the dimer statistics. All these problems deal with the fully covered lattice and have been solved rigorously for an arbitrary two-dimensional $[n,m]$ lattice.[18-20] Although enumeration on a partially covered lattice is physically much more interesting,[21] breakthroughs for rigorous solutions are still being sought.[4,22-24] For characterizing the topological nature of arbitrary graphs the present authors have proposed the idea of nonadjacent numbers and topological index, which can also be applied to these problems.[7,12] The king pattern and king polynomial are

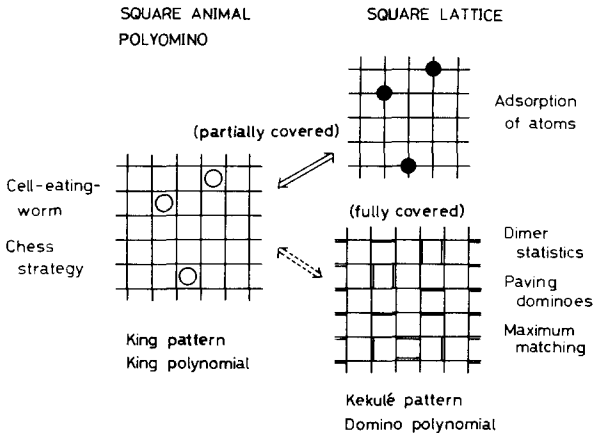
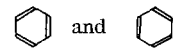
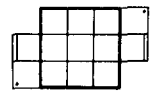


FIG. 2. Relation between the partially and fully covered square lattice problems through king and Kekulé patterns and through king and domino polynomials.

involved in the partially covered lattice problem but have a close relationship with the fully covered lattice problem as symbolically illustrated in Fig. 2. Note also that the king polynomial defined for a polyomino is related to different types of square lattice problems. Namely, the square lattices corresponding to the same polyomino are different in size as shown in Fig. 2.

Thus the king and domino polynomials, if extensively studied, might play an important role for correlating the fully and partially covered lattice problems. We are also attempting to extend the analysis of these polynomials to three-dimensional and infinitely large systems.

[1]J. Riordan, *An Introduction to Combinatorial Analysis* (Wiley, New York, 1958).
[2]C. L. Liu, *Introduction to Combinatorial Mathematics* (McGraw-Hill, New York, 1968).
[3]C. Domb and M. S. Green, *Phase Transitions and Critical Phenomena*, edited by M. S. Green (Academic, London, 1972), Vols. 1–3.
[4]D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967); R. B. McQuistan, Nuovo Cimento **58**, 86 (1968); J. Math. Phys. **14**, 1258 (1973).
[5]E. Clar, *The Aromatic Sextets* (Wiley, London, 1972).
[6]T. Yamaguchi, M. Suzuki, and H. Hosoya, Nat. Sci. Rep. Ochanomizu Univ. **26**, 39 (1975).
[7]H. Hosoya, K. Hosoi, and I. Gutman, Theor. Chim. Acta **38**, 37 (1975).
[8]In this paper two cells which share at least a vertex, as cells c and d in graph II, are "adjacent."
[9]The term Kekulé is taken from the Kekulé structures

and

for a benzene molecule used in chemistry. However, this terminology by no means implies the existence of a molecule corresponding to I.a.
[10]A rectangular polyomino with m rows and n columns.
[11]The idea of nontaking number is an extension of the nonadjacent number[12] and resonant-sextet number[13] for the various graphical enumeration problems.
[12]H. Hosoya, Bull. Chem. Soc. Jpn. **44**, 2332 (1971).
[13]H. Hosoya and T.Yamaguchi, Tetrahedron Lett. **1975**, 4659.
[14]A. Motoyama and H. Hosoya, Nat. Sci. Rep. Ochanomizu Univ. **27**, 107 (1976).
[15]A tatami is a grass-woven mat of 1×2 rectangular shape used as floor material for typical Japanese style rooms.
[16]In the case of graph VI the rectangular graph [3,3] cannot be a subunit of any Kekulé pattern.

VI
[17]A. Motoyama, thesis, Ochanomizu Univ. (1977).
[18]P. W. Kasteleyn, Physica **27**, 1209 (1961).
[19]N. H. V. Temperley and M. E. Fisher, Philos. Mag. **6**, 1061 (1961).
[20]J. K. Percus, "Combinatorial Methods," Lecture Notes of the Courant Institute, New York Univ. (1969).
[21]H. N. V. Temperley, in Ref. 3, p. 227.
[22]R. B. McQuistan, S. J. Lichtman, and L. P. Levine, J. Math. Phys. **13**, 242 (1972).
[23]R. C. Grimson, J. Math. Phys. **15**, 214 (1974).
[24]M. A. Rashid, J. Math. Phys. **15**, 474 (1974).

Bivector field theories, divergence-free vectors and the Einstein–Maxwell field equations

David Lovelock

Department of Mathematics, University of Arizona, Tucson, Arizona 85721
and Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada
(Received 21 January 1977)

A new derivation of the complete set of the Einstein–Maxwell field equations is presented which involves neither a variational principle nor the existence of a vector field (the so-called 4-potential). Unlike previous derivations, this derivation can therefore be used to motivate the Einstein–Maxwell field equations independently of the assumption of the existence or nonexistence of magnetic monopoles.

1. INTRODUCTION

In the general theory of relativity the field equations which purport to govern the interaction of the gravitational and electromagnetic fields, in regions devoid of sources, are the source-free Einstein—Maxwell field equations (with cosmological term), viz.,

$$aG^{ij}+bg^{ij}+2c[F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}(F^{rs}F_{rs})]=0 \tag{1.1}$$

and

$$F^{ij}{}_{|j}=0, \tag{1.2}$$

where a, b, c are constants, G^{ij} is the Einstein tensor, the vertical bar denotes covariant differentiation,

$$F_{ij}=\psi_{i,j}-\psi_{j,i}, \tag{1.3}$$

ψ_i is a vector field, and the comma denotes partial differentiation. It is well known[1] that the identity (1.3) is equivalent to the condition

$$\epsilon^{iabc}F_{ab|c}=0. \tag{1.4}$$

The full set of the Einstein—Maxwell equations is thus (1.1), (1.2), and (1.3), or, equivalently, (1.1), (1.2), and (1.4).

In the presence of sources these field equations are usually modified. If it is assumed that magnetic monopoles do not exist, then the right-hand sides of (1.1) and (1.2) are augmented by appropriate source terms, while (1.3) [or (1.4)] is unaffected. However, if magnetic monopoles are assumed to exist then (1.1), (1.2), and (1.4) are all augmented by appropriate source terms. Under these circumstances (1.3) is no longer a consequence of the augmented (1.4), which immediately implies that *we cannot infer the existence of a vector field* ψ_i for which (1.3) is valid.

It is well known that it is possible to derive (1.1) and (1.2) from a variational principle with a suitably chosen Lagrange density L *if there exists a vector field* ψ_i *for which* (1.3), and hence (1.4), *is valid.* This is usually accomplished in the following manner.

If L is a scalar density of the type

$$L=L(g_{ij};g_{ij,h_1};\cdots;g_{ij,h_1\cdots h_r};\psi_i;\psi_{i,j_1};\cdots;\psi_{i,j_1\cdots j_s}), \tag{1.5}$$

then we may associate with it two sets of Euler—Lagrange equations

$$E^{ij}(L)=0 \tag{1.6}$$

and

$$E^{i}(L)=0, \tag{1.7}$$

where

$$E^{ij}(L)=-\frac{\partial L}{\partial g_{ij}}+\sum_{k=1}^{r}(-1)^{k+1}\frac{\partial^k}{\partial x^{h_1}\cdots\partial x^{h_k}}\times\frac{\partial L}{\partial g_{ij,h_1\cdots h_k}} \tag{1.8}$$

and

$$E^{i}(L)=-\frac{\partial L}{\partial \psi_i}+\sum_{k=1}^{s}(-1)^{k+1}\frac{\partial^k}{\partial x^{j_1}\cdots\partial x^{j_k}}\frac{\partial L}{\partial \psi_{i,j_1\cdots j_k}}. \tag{1.9}$$

Since L is assumed to be a scalar density of the kind (1.5), the following is an identity[2]

$$2E^{ij}(L)_{|j}=-F^{i}{}_{h}E^{h}(L)-\psi^{i}E^{h}(L)_{|h}. \tag{1.10}$$

If, in particular, L is assumed to be of the special form

$$L=a\sqrt{g}R-2b\sqrt{g}+c\sqrt{g}F^{ij}F_{ij}, \tag{1.11}$$

then (1.6) and (1.7) reduce to (1.1) and (1.2) respectively.

In view of the physical significance of the Einstein—Maxwell field equations, an important problem is to determine those conditions which ensure the inevitability of (1.1)—(1.4).

We draw attention to the fact that L, given by (1.11), is a scalar density of the general kind

$$L=L(g_{ij};g_{ij,h};g_{ij,hk};\psi_{i,j}), \tag{1.12}$$

for which

$$E^{ij}(L)=E^{ij}(g_{ab};g_{ab,c};g_{ab,cd};\psi_{a,b}). \tag{1.13}$$

With these comments in mind, we cite the following results.[3] The only scalar density of the kind (1.12) for which (1.13) is valid is

$$L=a\sqrt{g}R+\alpha\sqrt{g}(R^2-4R_{ij}R^{ij}+R_{ijkh}R^{ijkh})+\beta\epsilon^{ijkh}R^{ab}{}_{ij}R_{abkh}+M, \tag{1.14}$$

where a, α, β are constants and M is a scalar density of the kind

$$M=M(g_{ab};\psi_{a,b}). \tag{1.15}$$

In this case (1.6) and (1.7) reduce to

$$a\sqrt{g}G^{ij}=\partial M/\partial g_{ij} \tag{1.16}$$

and

$$E^i(M)=0, \tag{1.17}$$

respectively. In order to restrict M we can proceed in two different ways. On the one hand we could demand that

$$\frac{\partial M}{\partial g_{ij}}=-b\sqrt{g}g^{ij}-2c\sqrt{g}[F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}(F^{rs}F_{rs})] \tag{1.18}$$

and then attempt to find M. This would ensure that we obtain (1.1). However, in this case (1.17) may not reduce to (1.2). On the other hand,[4] we could demand that

$$E^i(M)=4c(\sqrt{g}F^{ij})_{|j}, \tag{1.19}$$

in which case we would obtain (1.2) from (1.17), but (1.16) may not reduce to (1.1). It transpires that (1.18) and (1.19) are equivalent and they each imply that

$$M=c\sqrt{g}F^{ij}F_{ij}+\gamma\epsilon^{ijkh}F_{ij}F_{kh}-2b\sqrt{g}, \tag{1.20}$$

where c, γ, b are constants. Consequently, in either case *we inevitably obtain the Einstein–Maxwell field equations*, the coefficient of γ not contributing to either (1.16) or (1.17).

Recently Horndeski[5] has considered a slightly different problem, viz., to find all $E^{ij}(L)$, $E^i(L)$, where L is of the kind (1.5) for which

$$E^{ij}(L)=E^{ij}(g_{ab};g_{ab,c};g_{ab,cd};\psi_a;\psi_{a,b};\psi_{a,bc}), \tag{1.21}$$

$$E^{i}(L)=E^{i}(g_{ab};g_{ab,c};g_{ab,cd};\psi_a;\psi_{a,b};\psi_{a,bc}), \tag{1.22}$$

$$E^i(L)=\sqrt{g}F^{ij}{}_{|j} \quad \text{if} \quad R_{ijkh}=0, \tag{1.23}$$

and

$$E^i(L)_{|i}=0. \tag{1.24}$$

The motivation behind (1.21) and (1.22) is (1.1) and (1.2), while (1.23) is motivated by the experimentally accepted validity of Maxwell's equations in flat space–time. Finally (1.24) is motivated by the physically accepted evidence of conservation of charge. In fact the violation of (1.24) has been used to reject[6] various covariant generalizations of Maxwell's equations. Horndeski has shown that the general solution to this problem is

$$E^{ij}(L)=a\sqrt{g}G^{ij}+b\sqrt{g}g^{ij}+2c\sqrt{g}[F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}(F^{rs}F_{rs})]+\tau\sqrt{g}\,\delta^{iabc}_{defk}g^{dj}(F_{al}F^{el}R^{fk}{}_{bc}+F_{ab}{}^{|k}F^{ef}{}_{|c}), \tag{1.25}$$

and

$$E^i(L)=4c\sqrt{g}F^{ij}{}_{|j}+2\tau\sqrt{g}\,\delta^{iabc}_{defk}F^{de}{}_{|a}R^{fk}{}_{bc}, \tag{1.26}$$

where a, b, c, τ are constants, in which case (1.6) and (1.7) do not reduce to (1.1) and (1.2), and so (1.21)–(1.24) *do not uniquely characterize the Einstein–Maxwell field equations.*

However, as has been pointed out,[7] some physical field equations may not be derivable from a variational principle so that an assumption to the effect that the particular field equation we seek is obtainable from a variational principle, as was assumed heretofore, may not be justified *a priori*. Furthermore, the possible existence of magnetic monopoles has recently received considerable attention.[8] As we have seen, if magnetic monopoles do exist, we are not justified in inferring the existence of a vector field ψ_i for which (1.3) is valid, and, without a vector field, (1.9) is meaningless. Consequently, a motivation of the Einstein–Maxwell equations which hinges on the existence of a vector field cannot be used if monopoles exist.

With these two objections in mind, *the problem of determining an alternative set of conditions which ensure the inevitability of* (1.1), (1.2) *and* (1.4), *without the a priori assumption of the existence of a vector field for which* (1.3) *is valid, will be investigated.* In particular we shall be concerned with the following problem. Find all tensor densities B^i, C^i, A^{ij} for which

$$\text{(i)}\quad \begin{aligned} B^i&=B^i(g_{ab};g_{ab,c};g_{ab,cd};F_{ab|c}),\\ C^i&=C^i(g_{ab};g_{ab,c};g_{ab,cd};F_{ab|c}),\\ A^{ij}&=A^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab}), \end{aligned} \tag{1.27}$$

where $F_{ij}=-F_{ji}$;

$$\text{(ii)}\quad \text{if } R_{ijkh}=0, \quad \text{then } B^i=\sqrt{g}F^{ij}{}_{|j} \text{ and } C^i=\epsilon^{iabc}F_{ab|c}; \tag{1.28}$$

$$\text{(iii)}\quad B^i{}_{|i}=0 \text{ and } C^i{}_{|i}=0; \tag{1.29}$$

$$\text{(iv)}\quad A^{ij}{}_{|j}=\alpha^i{}_hB^h+\beta^i{}_hC^h, \tag{1.30}$$

where $\alpha^i{}_h$, $\beta^i{}_h$ are tensors and

$$\alpha^i{}_h=\alpha^i{}_h(g_{ab};F_{ab}), \quad \beta^i{}_h=\beta^i{}_h(g_{ab};F_{ab});$$

$$\text{(v)}\quad A^{ij}=A^{ji}. \tag{1.31}$$

The source-free field equations are then assumed to be of the form

$$A^{ij}=0, \quad B^i=0, \quad C^i=0. \tag{1.32}$$

Condition (ii) is motivated by the experimentally accepted validity of Maxwell's equations in special relativity. Condition (iii) is motivated by, and interpreted as, conservation of charge, both electric and magnetic. Condition (iv) is motivated by the requirement that the divergence of A^{ij} should vanish whenever "Maxwell's equations" $B^i=0$, $C^i=0$ are satisfied.[9] [This should also be compared with (1.10) when (1.24) is satisfied.] Condition (v) is motivated by the fact that the "Einstein equation" $A^{ij}=0$ is usually assumed to be symmetric.

In this paper we prove the following:

Theorem: If conditions (1.27)–(1.31) are satisfied,[10] then (1.32) are precisely the source-free Einstein–Maxwell field equations (1.1), (1.2) and (1.4).

This theorem can therefore be used to present a motivation for the Einstein–Maxwell field equations which is independent of the assumption of the existence or nonexistence of magnetic monopoles.

2. DIVERGENCE-FREE VECTORS

In order to solve (1.27)–(1.32) we shall first determine all vector densities A^i for which, in a four-dimensional space,

$$A^i=A^i(g_{ab};g_{ab,c};g_{ab,cd};F_{ab|c}) \tag{2.1}$$

and

$$A^i{}_{|i}=0. \tag{2.2}$$

We introduce the following notation which will be used repeatedly in the sequel, without specific mention. If $\alpha^{\cdots}_{\cdots}$ is any quantity, then

$$\alpha^{\cdots;ab,cd}_{\cdots} = \partial\alpha^{\cdots}_{\cdots}/\partial g_{ab,cd}$$

and

$$\alpha^{\cdots;ab,c}_{\cdots} = \partial\alpha^{\cdots}_{\cdots}/\partial F_{ab|c},$$

in which case we clearly have

$$\alpha^{\cdots;ab,cd}_{\cdots} = \alpha^{\cdots;ba,cd}_{\cdots} = \alpha^{\cdots;ab,dc}_{\cdots}$$

and

$$\alpha^{\cdots;ab,c}_{\cdots} = -\alpha^{\cdots;ba,c}_{\cdots}.$$

In view of (2.1) and the fact that A^i is a vector density, we see that $A^{i;ab,cd}$ and $A^{i;ab,c}$ are each tensor densities and, furthermore, satisfy the invariance identity[11]

$$A^{i;ab,cd} + A^{i;ad,bc} + A^{i;ac,db} = 0, \tag{2.3}$$

which implies that

$$A^{i;ab,cd} = A^{i;cd,ab}. \tag{2.4}$$

Written out in full, (2.2) reads

$$A^{i;ab,cd}g_{ab,cdi} + A^{i;ab,c}F_{ab|c,i} + (\partial A^i/\partial g_{ab,c})g_{ab,ci} + (\partial A^i/\partial g_{ab})g_{ab,i} = 0. \tag{2.5}$$

Differentiation of (2.5) with respect to $g_{ab,cdi}$ and $F_{ab,ci}$ gives rise to

$$A^{i;ab,cd} + A^{d;ab,ic} + A^{c;ab,di} = 0 \tag{2.6}$$

and

$$A^{i;ab,c} + A^{c;ab,i} = 0, \tag{2.7}$$

respectively.

We are now in a position to prove the following.

Lemma: In a four-dimensional space (2.1) and (2.2) imply that

$$A^{i;ab,cd;rs,tu} = 0, \tag{2.8}$$

$$A^{i;ab,c;rs,t;hk,l;pq,m} = 0, \tag{2.9}$$

and

$$A^{i;ab,c;rs,t;hj,kl} = 0. \tag{2.10}$$

Proof: The proof of (2.8) is essentially the same as that presented elsewhere,[12] and so will not be repeated here.

We establish (2.9) by noting that, from (2.7), $A^{i;ab,c;rs,t;hk,l;pq,m}$ is totally skew-symmetric in *ictlm*, which in a four-dimensional space, implies that it is identically zero.

To establish (2.10), we first note that (2.7) implies that $A^{i;ab,c;rs,t;hj,kl}$ is skew-symmetric in *ict*, so that we need only restrict our considerations to the case when *ict* are all distinct and equal to 1,2,3 (say). We now turn our attention to *hjkl*, and in particular to the number of 4's it contains. Clearly, by (2.3), if *hjkl* contains three or four 4's, then $A^{i;ab,c;rs,t;hj,kl}$ vanishes. We thus need consider only that situation in which *hjkl* contains at most two 4's, in which case the remaining two or more indices must be selected from 1's, 2's, and 3's. Furthermore, by (2.3), it is easily seen that, without loss of generality, we can always ensure that the *hj* indices in $A^{i;ab,c;rs,t;hj,kl}$ are taken entirely from 1's, 2's, and 3's. An elementary argument involving (2.6) and (2.7) now establishes (2.10), which completes the proof of the lemma.

From (2.8) and (2.10) it is clear that

$$A^{i;ab,cd;rs,t;hk,lm} = 0$$

and

$$A^{i;ab,cd;rs,t;hk,l} = 0,$$

in which case

$$A^{i;ab,cd;rs,t} = \alpha^{iabcdrst}(g_{hk};g_{hk,l}) \tag{2.11}$$

where $\alpha^{iabcdrst}$ has all the symmetry properties of $A^{i;ab,cd;rs,t}$ and is also a tensor density. By virtue of the tensorial character of $\alpha^{iabcdrst}$, it is easily established that[13]

$$\alpha^{iabcdrst} = \alpha^{iabcdrst}(g_{hk}). \tag{2.12}$$

Integration of (2.11) yields

$$A^{i;ab,cd} = \alpha^{iabcdrst}F_{rs|t} + \mu^{iabcd}(g_{hk};g_{hk,l}), \tag{2.13}$$

where μ^{iabcd} is a tensor density, in which case,[14] in V_4,

$$\mu^{iabcd} = 0.$$

Integration of (2.13) now gives rise to

$$A^i = \alpha^{iabcdrst}F_{rs|t}g_{ab,cd} + \mu^i(g_{hk};g_{hk,l};F_{hk|l}),$$

which by virtue of (2.3) can be rewritten in the form

$$A^i = \tfrac{2}{3}\alpha^{iabcdrst}F_{rs|t}R_{cabd} + \alpha^i(g_{hk};F_{hk|l}), \tag{2.14}$$

where α^i is a tensor density.

From (2.9) and (2.10) it is also clear that

$$A^{i;ab,c;rs,t;hk,l;pq,m} = 0$$

and

$$A^{i;ab,c;rs,t;hk,l;pq,uv} = 0,$$

which, when applied to (2.14), account being taken of (2.12), implies that

$$A^{i;ab,c;rs,t;hk,l} = \alpha^{i;ab,c;rs,t;hk,l} = \alpha^{i;ab,c;rs,t;hk,l}(g_{pq}). \tag{2.15}$$

In view of (2.7) it is easily seen that $A^{i;ab,c;rs,t;hk,l}$ is skew-symmetric in *ictl*, which, in a four-dimensional space, implies it is proportional to ϵ^{ictl}, i.e.,

$$\alpha^{i;ab,c;rs,t;hk,l} = \epsilon^{ictl}\alpha^{abrshk}, \tag{2.16}$$

where α^{abrshk} is a tensor, $\alpha^{abrshk} = \alpha^{abrshk}(g_{pq})$, and

$$\alpha^{abrshk} = -\alpha^{rsabhk} = -\alpha^{hkrsab} = -\alpha^{barshk}. \tag{2.17}$$

Integration of (2.16) then yields

$$\alpha^{i;ab,c;rs,t} = \epsilon^{ictl}\alpha^{abrshk}F_{hk|l} + \mu^{iabcrst}(g_{hk};g_{hk,l}),$$

which, in view of the fact that $\mu^{iabcdst}$ is a tensor density with an odd number of indices,[15] reduces to

$$\alpha^{i;ab,c;rs,t} = \epsilon^{ictl}\alpha^{abrshk}F_{hk|l}. \tag{2.18}$$

By integrating (2.18) twice and applying McKiernan's result, we thus find

$$\alpha^i = (1/3!)\epsilon^{ictl}\alpha^{abrshk}F_{hk|l}F_{rs|t}F_{ab|c} + \lambda^{iabc}F_{ab|c}, \tag{2.19}$$

where λ^{iabc} is a tensor density, $\lambda^{iabc}=\lambda^{iabc}(g_{rs})$, and

$$\lambda^{iabc}=-\lambda^{cabi}=-\lambda^{ibac}. \tag{2.20}$$

From (2.19) and (2.14) we thus find

$$A^i=\tfrac{2}{3}\alpha^{iabcdrst}F_{rs|t}R_{cabd}+(1/3!)\epsilon^{ictl}\alpha^{abrshk}F_{hk|l}F_{rs|t}F_{ab|c}+\lambda^{iabc}F_{ab|c}. \tag{2.21}$$

In order to obtain A^i explicitly, we shall now evaluate λ^{iabc}, α^{abrshk}, and $\alpha^{iabcdrst}$, all of which are tensorial and concomitants of g_{ij}.

The most general tensor density of type (4, 0) which is a concomitant of g_{ij} alone is[16]

$$\lambda^{iabc}=\alpha\sqrt{g}g^{ia}g^{bc}+\beta\sqrt{g}g^{ib}g^{ac}+\gamma\sqrt{g}g^{ic}g^{ab}+\delta\epsilon^{iabc}, \tag{2.22}$$

where $\alpha,\beta,\gamma,\delta$ are constants. In view of the fact that λ^{iabc} must also satisfy (2.20) we thus find

$$\gamma=0 \quad \text{and} \quad \beta=-\alpha,$$

in which case

$$\lambda^{iabc}F_{ab|c}=2\alpha\sqrt{g}F^{ij}{}_{|j}+\delta\epsilon^{iabc}F_{ab|c}. \tag{2.23}$$

It is easily shown that

$$(\lambda^{iabc}F_{ab|c})_{|i}=0. \tag{2.24}$$

We now turn to α_{abrshk} where the latter satisfies (2.17). In view of the fact that α_{abrshk} is a tensor and a concomitant of g_{ij} the associated invariance identity[16] gives rise, in the usual way, to

$$3\alpha_{abrshk}+\alpha_{barshk}+\alpha_{rbashk}+\alpha_{sbrahk}+\alpha_{hbrsak}+\alpha_{kbrsha}=g_{ab}g^{pq}\alpha_{pqrshk}+g_{ar}g^{pq}\alpha_{pbqshk}+g_{as}g^{pq}\alpha_{pbrqhk}+g_{ah}g^{pq}\alpha_{pbrsqk}+g_{ak}g^{pq}\alpha_{pbrshq},$$

which, in view of (2.17), reduces to

$$2\alpha_{abrshk}+\alpha_{rbashk}+\alpha_{sbrahk}+\alpha_{hbrsak}+\alpha_{kbrsha}=g_{ar}\mu_{bshk}-g_{as}\mu_{brhk}-g_{ah}\mu_{bkrs}+g_{ak}\mu_{bhrs}, \tag{2.25}$$

where

$$\mu_{bshk}=g^{pq}\alpha_{pbqshk}. \tag{2.26}$$

In (2.25) we interchange a and b and subtract the resulting equation from (2.25) to obtain

$$4\alpha_{abrshk}=g_{ar}\mu_{bshk}-g_{as}\mu_{brhk}-g_{ah}\mu_{bkrs}+g_{ak}\mu_{bhrs}-g_{br}\mu_{ashk}+g_{bs}\mu_{arhk}+g_{bh}\mu_{akrs}-g_{bk}\mu_{ahrs}. \tag{2.27}$$

Furthermore, from (2.26), we see that

$$\mu_{bshk}=\lambda\sqrt{g}(g_{bh}g_{sk}-g_{bk}g_{sh})+\mu g_{bi}g_{sj}g_{ht}g_{kl}\epsilon^{ijtl}, \tag{2.28}$$

where λ,μ are constants.

We shall now turn to $\alpha_{iabcdrst}$, where the latter is a tensor density, is a concomitant of g_{ab} and has all the symmetry properties of $A^{i;ab,cd;rs,t}$. In this case the appropriate invariance identity gives rise to

$$3\alpha_{iabcdrst}+\alpha_{aibcdrst}+\alpha_{baicdrst}+\alpha_{cabidrst}+\alpha_{dabcirst}+\alpha_{rabcdist}+\alpha_{sabcdrit}+\alpha_{tabcdrsi}=\lambda_{rabcdist}, \tag{2.29}$$

where

$$\lambda_{rabcdist}=g^{pq}[g_{ai}\alpha_{pqbcdrst}+g_{bi}\alpha_{paqcdrst}+g_{ci}\alpha_{pabqdrst}+g_{di}\alpha_{pabcqrst}+g_{ri}\alpha_{pabcdqst}+g_{si}\alpha_{pabcdrqt}+g_{ti}\alpha_{pabcdrsq}]. \tag{2.30}$$

In view of the symmetry properties of $\alpha_{iabcdrst}$, (2.29) reduces to

$$\alpha_{rabcdist}+\alpha_{sabcdrit}=\lambda_{rabcdist},$$

from which we obtain

$$2\alpha_{sabcdrit}=\lambda_{rabcdist}+\lambda_{rabcdsit}-\lambda_{iabcdrst}. \tag{2.31}$$

If we define

$$\rho_{bcdrst}=g^{pq}\alpha_{pqbcdrst} \tag{2.32}$$

and

$$\beta_{abcdst}=g^{pq}\alpha_{pabcdqst}, \tag{2.33}$$

we see that

$$\lambda_{rabcdist}=g_{ai}\rho_{bcdrst}+g_{bi}\rho_{acdrst}+g_{ci}\rho_{dabrst}+g_{di}\rho_{cabrst}+g_{ri}\beta_{abcdst}-g_{si}\beta_{abcdrt}. \tag{2.34}$$

From (2.31) and (2.34) it is easy to establish that

$$\alpha_{sabcdrit}F^{ri|t}R^{cabd}=2[2g_{ai}\rho_{bcdrst}+g_{as}\rho_{bcdrit}-g_{si}\beta_{abcdrt}]F^{ri|t}R^{cabd}. \tag{2.35}$$

We shall now evaluate ρ_{bcdrst}, which, by (2.32), may be expressed in the form

$$\rho_{bcdrst}=-\tfrac{1}{2}g^{pq}\alpha_{bpqcdrst}. \tag{2.36}$$

The invariance identity associated with ρ_{bcdrst} gives rise to

$$3\rho_{bcdrst}+\rho_{cbdrst}+\rho_{dcbrst}+\rho_{rcdbst}+\rho_{scdrbt}+\rho_{tcdrsb}=\lambda_{bcdrst}, \tag{2.37}$$

where

$$\lambda_{bcdrst}=g^{pq}[g_{bc}\rho_{pqdrst}+g_{bd}\rho_{pcqrst}+g_{br}\rho_{pcdqst}+g_{sb}\rho_{pcdrqt}+g_{bt}\rho_{pcdrsq}]. \tag{2.38}$$

In view of the symmetry properties of ρ_{bcdrst}, (2.37) reduces to

$$\rho_{bcdrst}+\rho_{rcdbst}+\rho_{scdrbt}=\lambda_{bcdrst},$$

from which we obtain

$$2\rho_{rcdbst}=\lambda_{bcdrst}-\lambda_{scdrbt}. \tag{2.39}$$

If we define

$$\alpha_{drst}=g^{pq}\rho_{pqdrst} \tag{2.40}$$

and

$$\beta_{cdst}=g^{pq}\rho_{pcdqst}, \tag{2.41}$$

we see that (2.38) can be expressed as

$$\lambda_{bcdrst}=g_{bc}\alpha_{drst}+g_{bd}\alpha_{crst}+g_{br}\beta_{cdst}-g_{bs}\beta_{cdrt}. \tag{2.42}$$

From (2.40) and (2.36) we note that

$$\alpha_{drst}=-\tfrac{1}{2}g^{pq}g^{hk}\alpha_{phkqdrst}=\tfrac{1}{4}g^{pq}g^{hk}\alpha_{dhkpqrst}, \tag{2.43}$$

in which case it is easily seen[17] that

$$\alpha_{drst}=a\sqrt{g}(g_{ds}g_{rt}-g_{dr}g_{st})+bg_{di}g_{rj}g_{sk}g_{tl}\epsilon^{ijkl}, \tag{2.44}$$

where a,b are constants. In a similar way (2.41) and (2.36) imply

$$\beta_{cdst} = -\tfrac{1}{2} g^{pq} g^{hk} \alpha_{phkcdqst}, \tag{2.45}$$

in which case

$$\beta_{cdst} = c\sqrt{g}\,(2g_{cd}g_{st} - g_{ct}g_{ds} - g_{dt}g_{cs}), \tag{2.46}$$

where c is a constant. Furthermore, from (2.43) and (2.45) we find

$$g^{dr}\alpha_{drst} = -\tfrac{1}{2} g^{cd}\beta_{cdst},$$

so that

$$a = c. \tag{2.47}$$

Substitution of (2.42), (2.44), (2.46), and (2.47) in (2.39) will thus determine ρ_{rcdbst}.

We shall now evaluate β_{abcdrt}, defined by (2.33). The invariance identity associated with β_{abcdrt} gives rise to

$$3\beta_{abcdst} + \beta_{bacdst} + \beta_{cbadst} + \beta_{dbcast} + \beta_{sbcdat} + \beta_{tbcdsa} = A_{abcdst} \tag{2.48}$$

and

$$3\beta_{sbcdat} + \beta_{sbcdta} + \beta_{sbctad} + \beta_{sbtdac} + \beta_{stcdab} + \beta_{tbcdas} = B_{sbcdat}, \tag{2.49}$$

where

$$A_{abcdst} = g^{hk}[g_{ab}\beta_{hkcdst} + g_{ac}\beta_{hbkdst} + g_{ad}\beta_{hbckst} + g_{as}\beta_{hbcdkt} + g_{at}\beta_{hbcdsk}] \tag{2.50}$$

and

$$B_{sbcdat} = g^{hk}[g_{at}\beta_{sbcdhk} + g_{dt}\beta_{sbchak} + g_{ct}\beta_{sbhdak} + g_{tb}\beta_{shcdak} + g_{ts}\beta_{hbcdak}]. \tag{2.51}$$

Because of the symmetry properties enjoyed by β_{abcdst}, it is easily seen that (2.48) and (2.49) reduce to

$$3\beta_{abcdst} + \beta_{sbcdat} + \beta_{tbcdsa} = A_{abcdst} \tag{2.52}$$

and

$$\beta_{sbcdat} + \beta_{sbcdta} = B_{sbcdat}, \tag{2.53}$$

respectively.

In (2.52) we interchange a and t, add the resulting equation to (2.52), and then subtract (2.53), to find

$$-4\beta_{atcdsb} = A_{abcdst} + A_{tbcdsa} - B_{sbcdat}. \tag{2.54}$$

From the latter equation we see that a knowledge of A_{abcdst} and B_{sbcdat} will determine β_{atcdsb}.

From (2.33) and (2.45) we observe that

$$g^{hk}\beta_{hkcdst} = g^{hk}g^{pq}\alpha_{phkcdqst} = -2\beta_{cdst}, \tag{2.55}$$

in which case

$$g^{hk}\beta_{hbkdst} = -\tfrac{1}{2} g^{hk}\beta_{hkbdst} = \beta_{bdst}, \tag{2.56}$$

and

$$g^{hk}\beta_{hbcdsk} = -\tfrac{1}{2} g^{hk}\beta_{hkcdsb} = \beta_{cdsb}. \tag{2.57}$$

In order to evaluate A_{abcdst} all that remains is to construct

$$g^{hk}\beta^{\bullet}_{hbcdkt} = g^{hk}g^{pq}\alpha_{phcdqkt},$$

which, being symmetric in cd and cyclic in bcd, must be given by

$$g^{hk}\beta_{hbcdkt} = \alpha\sqrt{g}\,(2g_{cd}g_{bt} - g_{ct}g_{bd} - g_{cb}g_{dt}), \tag{2.58}$$

where α is a constant. If we multiply (2.58) by g^{bt} and note (2.55), (2.46), and (2.47), we find

$$\alpha = c = a, \quad \text{i.e.,} \quad g^{hk}\beta_{hbcdkt} = \beta_{cdbt}. \tag{2.59}$$

From (2.50), (2.55)–(2.59) we thus find

$$A_{abcdst} = -2g_{ab}\beta_{cdst} + g_{ac}\beta_{bdst} + g_{ad}\beta_{bcst} + g_{as}\beta_{cdbt} + g_{at}\beta_{cdsb}. \tag{2.60}$$

Turning to B_{sbcdat} we note that, by (2.59) and (2.55),

$$g^{hk}\beta_{sbcdhk} = -g^{hk}\beta_{kscdhb} - g^{hk}\beta_{kbcdhs} = -2\beta_{cdsb} \tag{2.61}$$

and

$$g^{hk}\beta_{sbchak} = -\tfrac{1}{2} g^{hk}\beta_{hksbac} = \beta_{sbac}. \tag{2.62}$$

From (2.51), (2.61), and (2.62) we thus find

$$B_{sbcdat} = A_{tacdsb}, \tag{2.63}$$

where the right-hand side is given by (2.60).

We now substitute (2.39), (2.42), (2.44), (2.46), (2.47), (2.54), (2.60), and (2.63) in (2.35), which, after a lengthy, but straightforward, calculation gives rise to

$$\alpha^{sabcdrtt}F_{rt|t}R_{cabd} = 6a\sqrt{g}\,[2F^{ta}{}_{|t}R_a{}^s + F^{ca}{}_{|d}R^{sd}{}_{ca} + 2F_b{}^s{}_{|a}R^{ab} + F^{st}{}_{|t}R] + 3bF_{ca|t}\epsilon^{dbst}R^{ca}{}_{bd}. \tag{2.64}$$

From (2.64) it is easy to show that

$$(\alpha^{sabcdrtt}F_{rt|t}R_{cabd})_{|s} = 0. \tag{2.65}$$

We now substitute (2.23), (2.27), and (2.64) in (2.21) and observe (2.2), (2.24), (2.27), (2.28), and (2.65) to find

$$\epsilon^{ictl}\alpha^{abrshk}F_{hk|t}F_{rs|t}F_{ab|ct} = 0,$$

from which it is not difficult to deduce that

$$\alpha^{abrshk} = 0.$$

Consequently we have established the following:

Theorem: In a four-dimensional space the most general vector density $A^i = A^i(g_{ab}; g_{ab,c}; g_{ab,cd}; F_{ab|c})$ for which $A^i{}_{|i} = 0$ is

$$A^i = \alpha\sqrt{g}\,F^{ij}{}_{|j} + \beta\epsilon^{iabc}F_{ab|c} + \lambda\epsilon^{iabc}F_{hk|c}R^{hk}{}_{ab} + \mu\sqrt{g}\,[2F^{ta}{}_{|t}R_a{}^i + F^{ca}{}_{|d}R^{id}{}_{ca} + 2F_b{}^i{}_{|a}R^{ab} + F^{it}{}_{|t}R], \tag{2.66}$$

where $\alpha, \beta, \lambda, \mu$ are arbitrary constants.

We remark that a quantity formally similar to the expression in square brackets in (2.66) has arisen in the work of Horndeski[18] [compare with the coefficient of τ in (1.26)]. However, in (2.66) it is not assumed that the skew-symmetric tensor F_{ij} is obtained from a vector field in the usual manner. We also remark that the coefficients of λ and $\mu\sqrt{g}$ in (2.66) can each be expressed in the form

$$\delta^{iabc}_{rstu}H^{rs}{}_{|a}R^{tu}{}_{bc},$$

where H^{rs} is either $\epsilon^{rsjk}F_{jk}$ or F^{rs}.

We now introduce the notation

$$P^i = \sqrt{g}\,\delta^{iabc}_{rstu}F^{rs}{}_{|a}R^{tu}{}_{bc} \tag{2.67}$$

and

$$Q^i = \epsilon^{iabc} F_{hk|c} R^{hk}{}_{ab}. \tag{2.68}$$

From the previous theorem we immediately have the following:

Corollary: In a four-dimensional space the most general vector densities

$$B^i = B^i(g_{ab}; g_{ab,c}; g_{ab,cd}; F_{ab|c}),$$

$$C^i = C^i(g_{ab}; g_{ab,c}; g_{ab,cd}; F_{ab|c})$$

for which $B^i{}_{|i} = 0$, $C^i{}_{|i} = 0$, and

$$B^i = a\sqrt{g}\, F^{ij}{}_{|j} \quad \text{if} \quad R_{ijkh} = 0,$$

$$C^i = b\epsilon^{ijab} F_{ab|j} \quad \text{if} \quad R_{ijkh} = 0,$$

are

$$B^i = a\sqrt{g}\, F^{ij}{}_{|j} + cP^i + dQ^i, \tag{2.69}$$

$$C^i = b\epsilon^{ijab} F_{ab|j} + eP^i + fQ^i, \tag{2.70}$$

where a, b, c, d, e, f are constants, and P^i, Q^i are defined by (2.67) and (2.68)

3. THE EINSTEIN-MAXWELL FIELD EQUATIONS

In this section we shall obtain the most general quantities $A^{ij}, B^i, C^i, \alpha^i{}_h, \beta^i{}_h$ which satisfy the identity

$$A^{ij}{}_{|j} = \alpha^i{}_h B^h + \beta^i{}_h C^h, \tag{3.1}$$

where

(i) A^{ij} is a tensor density,

$$A^{ij} = A^{ij}(g_{ab}; g_{ab,c}; g_{ab,c}; g_{ab,cd}; F_{ab}), \tag{3.2}$$

$$A^{ij} = A^{ji}, \tag{3.3}$$

(ii) B^h is a vector density,

$$B^h = B^h(g_{ab}; g_{ab,c}; g_{ab,cd}; F_{ab|c}), \tag{3.4}$$

$$B^h = \sqrt{g}\, F^{hj}{}_{|j} \quad \text{if} \quad R_{ijhk} = 0, \tag{3.5}$$

(iii) C^h is a vector density,

$$C^h = C^h(g_{ab}; g_{ab,c}; g_{ab,cd}; F_{ab|c}), \tag{3.6}$$

$$C^h = \epsilon^{hijk} F_{ij|k} \quad \text{if} \quad R_{ijhk} = 0, \tag{3.7}$$

(iv) $\alpha^i{}_h, \beta^i{}_h$ are tensors, which are not both identically zero, and

$$\alpha^i{}_h = \alpha^i{}_h(g_{ab}; F_{ab}), \tag{3.8}$$

$$\beta^i{}_h = \beta^i{}_h(g_{ab}; F_{ab}). \tag{3.9}$$

As a consequence of (3.4)–(3.7), the corollary in the previous section is applicable and B^h and C^h are thus given by (2.69) and (2.70) respectively. By virtue of this comment together with (3.2), we see that (3.1) can be expressed in the form

$$A^{ij;ab,cd} g_{ab,cdj} + A^{ij;ab} F_{ab,j} + \frac{\partial A^{ij}}{\partial g_{ab,c}} g_{ab,cj} + \frac{\partial A^{ij}}{\partial g_{ab}} g_{ab,j} + \Gamma^i_{aj} A^{aj} = \alpha^i{}_h \sqrt{g}\, F^{hj}{}_{|j} + \beta^i{}_h \epsilon^{habc} F_{ab|c} + \lambda^i{}_h P^h + \mu^i{}_h Q^h, \tag{3.10}$$

where

$$A^{ij;ab} = \partial A^{ij}/\partial F_{ab}, \tag{3.11}$$

$$\lambda^i{}_h = c\alpha^i{}_h + e\beta^i{}_h,$$

$$\mu^i{}_h = d\alpha^i{}_h + f\beta^i{}_h. \tag{3.12}$$

By differentiating (3.10) with respect to $g_{ab,cdj}$ we find

$$A^{ij;ab,cd} + A^{id;ab,jc} + A^{ic;ab,dj} = 0,$$

which, when combined with (3.2) and (3.3) implies, in the usual way,[19] that

$$A^{ij;ab,cd} = a^{ijabcd}, \tag{3.13}$$

where a^{ijabcd} is a tensor density and

$$a^{ijabcd} = a^{ijabcd}(g_{rs}; F_{rs}). \tag{3.14}$$

With this in mind we return to (3.10) and differentiate it twice, first with respect to $F_{ab,j}$ and then with respect to $g_{rs,tu}$, to obtain

$$A^{ij;ab;rs,tu} = \lambda^i{}_h P^{h;ab,j;rs,tu} + \mu^i{}_h Q^{h;ab,j;rs,tu}. \tag{3.15}$$

We shall now show that both $\lambda^i{}_h$ and $\mu^i{}_h$ must vanish identically. From (2.67) and (2.68) it is easy to show that

$$P^{h;ab,j;rs,tu} = \sqrt{g}\, \tfrac{1}{2} \delta^{hjkl}_{iqmn} g^{ic} g^{ad} g^{am} g^{bn} D^{rs}_{cl} D^{tu}_{dk}$$

and (3.16)

$$Q^{h;ab,j;rs,tu} = \tfrac{1}{4} \epsilon^{hjkl} (g^{ac} g^{bd} - g^{bc} g^{ad})\, D^{rs}_{cl} D^{tu}_{dk},$$

where

$$D^{rs}_{cl} = \delta^r_c \delta^s_l + \delta^r_l \delta^s_c.$$

From (3.3) we obviously have

$$A^{ij;ab;rs,tu} g_{rs} g_{tu} = A^{ji;ab;rs,tu} g_{rs} g_{tu},$$

which, by (3.15) and (3.16), reduces to

$$2\sqrt{g}\, \lambda^{ib} g^{aj} - 2\sqrt{g}\, \lambda^{ia} g^{bj} + \mu^i{}_h \epsilon^{hjba} = 2\sqrt{g}\, \lambda^{jb} g^{ai} - 2\sqrt{g}\, \lambda^{ja} g^{bi} + \mu^j{}_h \epsilon^{hiba}. \tag{3.17}$$

If we multiply (3.17) by g_{aj}, we find

$$4\sqrt{g}\, \lambda^{ib} = -2\sqrt{g}\, (\lambda^{rs} g_{rs}) g^{bi} + \mu_{ah} \epsilon^{hiba},$$

from which it follows that

$$4\sqrt{g}\, \lambda^{ib} = \mu_{ah} \epsilon^{hiba}. \tag{3.18}$$

This equation clearly implies that λ^{ib} must be skew-symmetric.

We now return to (3.17) and multiply it by ϵ_{rjba} to obtain, by virtue of (3.18),

$$\mu^{ir} = -\mu^{ri}, \tag{3.19}$$

which, when taken in conjunction with (3.18), gives rise to

$$\mu_{rs} = -\sqrt{g}\, \epsilon_{rsib} \lambda^{ib}. \tag{3.20}$$

From (3.15), (3.16), and (3.20) we thus find

$$A^{ij;ab;rs,tu} g_{rs} g_{tu} = 4\sqrt{g}\, (\lambda^{ib} g^{aj} - \lambda^{ia} g^{bj} + \lambda^{ba} g^{ji} - \lambda^{ja} g^{bi} + \lambda^{jb} g^{ia}). \tag{3.21}$$

Equation (3.11) implies that

$$A^{ij;ab;cd} = A^{ij;cd;ab},$$

which, when imposed on (3.21), leads to

$$\lambda^{ib;cd}g^{aj}-\lambda^{ia;cd}g^{bj}+\lambda^{ba;cd}g^{ji}-\lambda^{ja;cd}g^{bi}+\lambda^{jb;cd}g^{ia}$$
$$=\lambda^{id;ab}g^{cj}-\lambda^{ic;ab}g^{dj}+\lambda^{dc;ab}g^{ji}-\lambda^{jc;ab}g^{di}+\lambda^{jd;ab}g^{ic}. \tag{3.22}$$

We shall now prove the following:

Lemma: If λ^{ij} is a skew-symmetric tensor which satisfies $\lambda^{ij}=\lambda^{ij}(g_{ab};F_{ab})$ and (3.22), then

$$\lambda^{ij}=\alpha F^{ij}, \tag{3.23}$$

where α is a constant.

Proof: We multiply (3.22) by g^{aj} and let

$$b^{cb}=\lambda^{jc;ab}g_{aj}$$

to find

$$3\lambda^{ib;cd}-\lambda^{id;cb}+\lambda^{ic;db}-\lambda^{dc;ib}=-b^{cb}g^{di}+b^{db}g^{ic}. \tag{3.24}$$

Multiplication of (3.24) by g_{ic} thus yields

$$b^{bd}=b^{db}. \tag{3.25}$$

We now multiply (3.24) by g_{cb} and note (3.25) to find

$$b^{id}=\beta g^{id}, \tag{3.26}$$

where

$$\beta=\tfrac{1}{4}g_{ij}b^{ij}=\beta(g_{ab};F_{ab}). \tag{3.27}$$

When (3.26) is substituted in (3.24), we see that

$$3\lambda^{ib;cd}-\lambda^{id;cb}+\lambda^{ic;db}-\lambda^{dc;ib}=\beta(g^{db}g^{ic}-g^{cb}g^{di}). \tag{3.28}$$

In (3.28) we cycle on cdi to find

$$\lambda^{ib;cd}+\lambda^{db;ic}+\lambda^{cb;di}=\lambda^{id;cb}+\lambda^{dc;ib}+\lambda^{ci;db}. \tag{3.29}$$

We subtract (3.29) from (3.28) and obtain

$$2\lambda^{ib;cd}-\lambda^{db;ic}-\lambda^{cb;di}=\beta(g^{db}g^{ic}-g^{cb}g^{di}). \tag{3.30}$$

The equation obtained from (3.30) by interchanging i with c is added to (3.30) to yield

$$3(\lambda^{ib;cd}-\lambda^{cb;di})=\beta(2g^{db}g^{ic}-g^{cb}g^{di}-g^{ib}g^{dc}). \tag{3.31}$$

The equation obtained from (3.31) by interchanging i with b is subtracted from (3.31) to give

$$2\lambda^{ib;cd}-\lambda^{cb;di}+\lambda^{ci;db}=\beta(g^{db}g^{ic}-g^{cb}g^{di}). \tag{3.32}$$

A comparison of (3.32) with (3.30) thus shows that

$$\lambda^{ci;db}=\lambda^{db;ci},$$

which, when substituted in (3.29), yields

$$\lambda^{ib;cd}+\lambda^{db;ic}+\lambda^{cb;di}=0. \tag{3.33}$$

Equation (3.33) is now substituted in (3.30) which gives rise to

$$3\lambda^{ib;cd}=\beta(g^{db}g^{ic}-g^{cb}g^{di}). \tag{3.34}$$

We now differentiate (3.34) with respect to F_{rs} to find

$$\beta^{;rs}(g^{db}g^{ic}-g^{cb}g^{di})=\beta^{;cd}(g^{sb}g^{ir}-g^{rb}g^{is}),$$

which, when multiplied by $g_{db}g_{ic}$ yields

$$\beta^{;rs}=0,$$

so that, by (3.27),

$$\beta=\beta(g_{ab}),$$

which, in turn implies[20]

$$\beta=\text{const.}$$

Integration of (3.34) thus yields

$$\lambda^{ib}=\alpha F^{ib}+\beta^{ib}, \tag{3.35}$$

where α is a constant and β^{ib} is a skew-symmetric tensor and

$$\beta^{ib}=\beta^{ib}(g_{rs}),$$

which implies

$$\beta^{ib}=0.$$

The latter completes the proof of the lemma in view of (3.35).

This lemma, together with (3.20), thus implies that

$$\lambda^{ij}=\alpha F^{ij},\quad \mu_{rs}=-\alpha\sqrt{g}\,\epsilon_{rsij}F^{ij}. \tag{3.36}$$

However, (3.11) implies that

$$A^{ij;ab;cd;rs,tu}R_{urst}g_{ac}=A^{ji;ab;cd;rs,tu}R_{urst}g_{ac}. \tag{3.37}$$

We now substitute (3.15), (3.16), and (3.36) in (3.37) to find, after some calculation, that the following must be an identity:

$$\alpha(g^{jb}R^{id}+g^{jd}R^{ib}-g^{ib}R^{jd}-g^{id}R^{jb})=0.$$

Since this has to be an identity we conclude that

$$\alpha=0,$$

which, from (3.36), implies that

$$\lambda^{ij}=\mu^{ij}=0, \tag{3.38}$$

identically.

When (3.38) is substituted in (3.15) we find

$$A^{ij;ab;rs,tu}=0,$$

which is now applied to (3.13) and (3.14) to obtain

$$A^{ij;ab,cd}=a^{ijabcd}(g_{rs}).$$

The latter equation is now integrated in the usual way[21] to give

$$A^{ij}=a\sqrt{g}\,G^{ij}+\theta^{ij}, \tag{3.39}$$

where $\theta^{ij}=\theta^{ij}(g_{ab};F_{ab})$ is a symmetric tensor density, a is a constant, and G^{ij} is the Einstein tensor. Our problem will be solved once we have determined θ^{ij}, which by (3.38), (3.39) and (3.10), must satisfy

$$\theta^{ij}{}_{|j}=\alpha^{i}{}_{h}\sqrt{g}\,F^{hj}{}_{|j}+\beta^{i}{}_{h}\epsilon^{habc}F_{ab|c}. \tag{3.40}$$

We shall now prove the following:

Theorem: The only tensor density θ^{ij} which satisfies

$$\theta^{ij}=\theta^{ij}(g_{ab};F_{ab}),\quad \theta^{ij}=\theta^{ji}, \tag{3.41}$$

and (3.40), where $\alpha^{i}{}_{h}$, $\beta^{i}{}_{h}$ satisfy (3.8) and (3.9), is

$$\theta^{ij}=-\lambda\sqrt{g}\,[F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}(F^{rs}F_{rs})]+\mu\sqrt{g}\,g^{ij}, \tag{3.42}$$

where λ, μ are constants.

Proof: From (3.40) it is easily shown that

$$\theta^{ij;ab}=\tfrac{1}{2}\sqrt{g}\,(\alpha^{ia}g^{jb}-\alpha^{ib}g^{ja})+\beta^{i}{}_{h}\epsilon^{hjab}. \tag{3.43}$$

By virtue of (3.41) we must have

$$\tfrac{1}{2}\sqrt{g}\,(\alpha^{ia}g^{jb}-\alpha^{ib}g^{ja})+\beta^{i}{}_{h}\epsilon^{hjab}$$
$$=\tfrac{1}{2}\sqrt{g}\,(\alpha^{ja}g^{ib}-\alpha^{jb}g^{ia})+\beta^{j}{}_{h}\epsilon^{hiab}.$$

We now apply to this equation the techniques used on (3.17) [which gave rise to (3.18), (3.19) and (3.20)] to find

$$\sqrt{g}\,\alpha^{ib}=\beta_{ah}\epsilon^{hiba},\quad \alpha^{ib}=-\alpha^{bi},$$
$$\beta_{ah}=-\beta_{ha},\quad \beta_{ah}=-\tfrac{1}{4}\sqrt{g}\,\epsilon_{ahib}\alpha^{ib}. \tag{3.44}$$

Substitution of (3.44) in (3.43) thus yields

$$\theta^{ij;ab}=\tfrac{1}{2}\sqrt{g}\,[\alpha^{ia}g^{jb}-\alpha^{ib}g^{ja}+\alpha^{ja}g^{ib}$$
$$-\alpha^{jb}g^{ia}+\alpha^{ab}g^{ij}]. \tag{3.45}$$

If the condition

$$\theta^{ij;ab;cd}=\theta^{ij;cd;ab}$$

is now applied to (3.45), we obtain (3.22) with $\lambda^{ib;cd}$ replaced by $\alpha^{ib;cd}$, in which case the lemma following (3.22) implies that

$$\alpha^{ij}=\lambda F^{ij}, \tag{3.46}$$

where λ is a constant. We now substitute this back into (3.45) to obtain

$$\theta^{ij;ab}=\tfrac{1}{2}\sqrt{g}\,\lambda(F^{ia}g^{jb}-F^{ib}g^{ja}+F^{ja}g^{ib}$$
$$-F^{jb}g^{ia}+F^{ab}g^{ij}). \tag{3.47}$$

If we define T^{ij} by

$$T^{ij}=-\lambda\sqrt{g}\,(F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}F^{rs}F_{rs}), \tag{3.48}$$

we find

$$T^{ij;ab}=-\tfrac{1}{2}\sqrt{g}\,\lambda(F^{jb}g^{ia}-F^{ja}g^{ib}+F^{ib}g^{ja}$$
$$-F^{ia}g^{jb}-F^{ab}g^{ij}). \tag{3.49}$$

A comparison (3.47) and (3.49) thus yields

$$\theta^{ij;ab}-T^{ij;ab}=0,$$

which implies that

$$\theta^{ij}=T^{ij}+\rho^{ij} \tag{3.50}$$

where ρ^{ij} is a symmetric tensor density and

$$\rho^{ij}=\rho^{ij}(g_{ab}).$$

In the usual way[22] we find

$$\rho^{ij}=\mu\sqrt{g}\,g^{ij}, \tag{3.51}$$

where μ is a constant. This proves the theorem in view of (3.48), (3.50), and (3.51).

It should be pointed out that this theorem is of interest in its own right since it augments some previous work[23] on uniquely characterizing the electromagnetic energy–momentum tensor.

We now apply this theorem to (3.39) which shows that θ^{ij} is given by (3.42) and, in view of (3.44) and (3.46),

$$\alpha^{ij}=\lambda F^{ij},\quad \beta_{ah}=-\tfrac{1}{4}\sqrt{g}\,\lambda\epsilon_{ahib}F^{ib}. \tag{3.52}$$

Substitution of (3.52) and (3.12) into the identifies (3.38) shows that

$$c=d=e=f=0, \tag{3.53}$$

which by (2.69) and (2.70) implies that

$$B^{i}=\sqrt{g}\,F^{ij}{}_{|j},\quad C^{i}=\epsilon^{ijab}F_{ab|j}.$$

We have thus proved the following:

Theorem: The only tensor densities A^{ij}, B^{i}, C^{i} which satisfy (3.1)–(3.9) are

$$A^{ij}=a\sqrt{g}\,G^{ij}-\lambda\sqrt{g}\,(F^{ih}F^{j}{}_{h}-\tfrac{1}{4}g^{ij}F^{rs}F_{rs})+\mu\sqrt{g}\,g^{ij}, \tag{3.54}$$

$$B^{i}=\sqrt{g}\,F^{ij}{}_{|j}, \tag{3.55}$$

and

$$C^{i}=\epsilon^{ijab}F_{ab|j}, \tag{3.56}$$

where a, λ, μ are constants.

The theorem stated in Sec. 1 is clearly an immediate consequence.

ACKNOWLEDGMENTS

I would like to thank G. W. Horndeski for various discussions. This work was supported in part by NSF MCS 76-06680 and NRC A7342.

[1]See, for example, D. Lovelock and H. Rund, *Tensors, Differential Forms and Variational Principles* (Wiley-Interscience, New York, 1975), pp. 146–47.
[2]G. W. Horndeski, Ph. D. thesis (University of Waterloo, Canada, 1973).
[3]D. Lovelock, Proc. Roy. Soc. London A **341**, 285 (1974).
[4]D. Lovelock, Gen. Rel. Grav. **5**, 399 (1974).
[5]G. W. Horndeski, J. Math. Phys. **17**, 1980 (1976).
[6]A. Trautman, "Foundations and Current Problems of General Relativity," in *Lectures on General Relativity* (Prentice-Hall, Englewood Cliffs, N. J., 1964), pp. 123–27.
[7]See, for example, C. Truesdell and R. Toupin, "The classical field theories", in *Handbuch der Physik, Vol. III/I* (Springer-Verlag, Berlin, 1960), p. 595.
[8]See, for example, J. Schwinger, Phys. Rev. D **12**, 3105 (1975).
[9]For the physical significance of this assumption, see J. L. Synge, "Introduction to General Relativity," in *Relativity, Groups and Topology* (Gordon and Breach, New York, 1964), p. 87.
[10]In fact this theorem is also true if the third equation in (1.27) is replaced by

$$A^{ij}=A^{ij}(g_{ab};g_{ab,c};g_{ab,cd};F_{ab};F_{ab,c}).$$

[11]See Ref. 1, Chap. 8.
[12]See D. Lovelock, J. Math. Phys. **13**, 874 (1972), or I. Anderson and D. Lovelock, J. Math. Phys. **17**, 1001 (1976).
[13]See Ref. 11.
[14]M. A. McKiernan, Demonstratio Math. **6**, 253 (1973).
[15]See Ref. 14.
[16]D. Lovelock, Arch. Ratl. Mech. Anal. **33**, 54 (1969).
[17]See Ref. 16.
[18]See Ref. 5.
[19]D. Lovelock, Aequationes Math. **4**, 127 (1970).
[20]See Ref. 16.
[21]See Ref. 19.
[22]See Ref. 16.
[23]C. D. Collinson, Proc. Cambridge Philos. Soc. **66**, 437 (1969); D. Lovelock, Int. J. Theoret. Phys. **10**, 59 (1974).

The classical action in nonrelativistic quantum mechanics

Aubrey Truman

Mathematics Department, Heriot-Watt University, Edinburgh, Scotland
(Received 20 October 1976)

We obtain some new results on the role played by the classical action in nonrelativistic quantum mechanics. The results are of the same genre as those derived previously by Nelson from the Trotter product formula. Here we work with the exact expression for the classical action not the approximate one as used by Nelson. Our results give a precise relationship between classical mechanics and quantum mechanics for a fairly wide class of potentials. The results are derived by using the properties of a new definition of the Feynman path integral $\mathcal{J}$ introduced in an earlier paper.

1. INTRODUCTION

In a previous paper we gave a new definition of the Feynman path integral $\mathcal{J}$ in nonrelativistic quantum mechanics and developed some of its properites.[1] We continue this investigation here exploring the connection between our definition of $\mathcal{J}$ and some of Feynman's original ideas on the path integral. It turns out that our definition of $\mathcal{J}$ is particularly well suited to answer one of the early questions posed by Feynman's work—the question of the precise relationship between classical mechanics and quantum mechanics. One aspect of this question is the problem of obtaining classical mechanics as the limiting case of quantum mechanics when $\hbar \to 0$. This aspect was discussed in our previous paper in terms of the quasiclassical representation, and we shall only mention this topic briefly here. There is, however, another related aspect which we now outline.

We restrict our attention to the mechanics of a single particle in one space dimension. Generalization of our work to a space of higher dimension than one is straightforward. We also choose units so that the particle mass $m=1$, and Planck's constant divided by 2π, $\hbar=1$. Let $(P_n\gamma)$ denote the polygonal path, defined for $\tau \in (0, t)$ by

$$(P_n\gamma)(\tau) = \gamma_j + (\tau - jt/n)(\gamma_{j+1} - \gamma_j)n/t,$$

$$jt/n \leq \tau < (j+1)t/n, \tag{1}$$

$j = 0, 1, 2, \ldots, n-1$, where initially γ_j are fixed and arbitrary, save for γ_n and $\gamma_n = 0$.[2] We denote by $(P_n\gamma + X)$ the polygonal path defined by

$$(P_n\gamma + X)(\tau) = (P_n\gamma)(\tau) + X, \quad \tau \in (0, t), \tag{2}$$

for a constant X, so that $(P_n\gamma + X)(\tau = t) = X$; and $S_{\mathrm{cl}}[P_n\gamma + X]$ denotes the classical action of a particle of mass unity in a potential V, traveling with a constant velocity along each of the segments of the path $(P_n\gamma + X)$,

$$S_{\mathrm{cl}}[P_n\gamma + X] = \sum_{j=0}^{n-1} \frac{(\gamma_{j+1} - \gamma_j)^2}{2\Delta t} - \int_0^t V[P_n\gamma + X]\, d\tau, \tag{3}$$

where $\Delta t = t/n$.

Now let $\psi(X, t)$ be the amplitude for the quantum mechanical particle of mass unity, moving in the potential V, to be at X at time t. Then, using earlier work of Dirac, Feynman conjectured that

$$\psi(X, t) = \lim_{n\to\infty} N_n \int d^n\gamma \exp\{iS_{\mathrm{cl}}[P_n\gamma + X]\}\, \psi(\gamma_0 + X, 0), \tag{4}$$

where $N_n = (2\pi i\, \Delta t)^{-n/2}$ is a normalization constant and $d^n\gamma = d\gamma_0 d\gamma_1 \cdots d\gamma_{n-1}$, each integration being from $-\infty$ to $+\infty$.[3]

In Eq. (4), $\psi(X, t)$ would be the solution of the Cauchy problem for the Schrödinger equation with potential V. The first rigorous proof of a result analogous to (4) was given by Kac for the heat equation with the potential V.[4] The corresponding result for the Schrödinger equation was given in a beautiful paper by Nelson.[5] Essentially, Nelson makes the replacement

$$\int_0^t V[P_n\gamma + X]\, d\tau$$
$$= \sum_{j=0}^{n-1} \int_{jt/n}^{(j+1)t/n} V[X + \gamma_j + (\tau - jt/n)(\gamma_{j+1} - \gamma_j)\Delta t^{-1}]\, d\tau$$
$$\to \sum_{j=0}^{n-1} V[\gamma_j + X]\Delta t, \tag{5}$$

corresponding to a Riemann sum approximation to the potential term and then, making elegant use of the Trotter product formula, he proves the validity of (4).[6] In this paper we make no such replacement, but, by using the properties of $\mathcal{J}$ deduced in our previous paper, we prove the validity of (4) for a wide class of potentials $V(\cdot)$ and a wide class of initial wavefunctions $\psi(\cdot, 0)$. In this way we establish a precise relationship between classical mechanics and quantum mechanics. This relationship is summarized in Theorems 6 and 7 of the Conclusion.

Our definition of $\mathcal{J}$ is based upon the properties of the underlying space of paths for spinless nonrelativistic quantum mechanical particles. Following previous authors we choose this path space to be H, the Hilbert space of continuous functions $\gamma(\tau)$, defined for $\tau \in (0, t)$, normalized so that $\gamma(t) = 0$, with weak derivative $d\gamma/d\tau \in L^2(0, t)$ and with inner product $(\gamma', \gamma) = \int_0^t (d\gamma'/d\tau) \times (d\gamma/d\tau)\, d\tau$. In Theorem 1 we prove that in the inner product norm topology H is a real separable Hilbert space, easily realized in terms of Fourier trigonometric series. We now give our definition of $\mathcal{J}$ for complex-valued functionals $f[\gamma]$ defined on the space of paths H.

We define the linear map $P_n : H \to H$ by identifying, in Eq. (1), γ_j with $\gamma(jt/n)$, $\gamma \in H$, so that, for $j = 0, 1, 2, \ldots, n-1$,

$$(P_n\gamma)(\tau) = \gamma(jt/n) + (\tau - jt/n)[\gamma(\overline{j+1}\,t/n) - \gamma(jt/n)]\Delta t^{-1},$$

$$jt/n \leq \tau < (j+1)t/n. \tag{6}$$

Let $f[\gamma]$ be a cylinder functional, i.e., a functional depending upon $\gamma(\sigma_j)$ for only a finite number of times $\sigma_j = jt/n$, $j = 0, 1, \ldots, n-1$, $f[\gamma] = f[\gamma_0, \ldots, \gamma_{n-1}]$. For such a cylinder functional we shall use the shorthand notation

$$\mathcal{J}_n[f] = N_n \int d^n\gamma \exp\left(i \sum_{j=0}^{n-1} \frac{(\gamma_{j+1} - \gamma_j)^2}{2\Delta t}\right) \times f[\gamma_0, \ldots, \gamma_{n-1}], \tag{7}$$

whenever this integral exists.

Definition: Let $f: H \to C$ be a functional with domain H and let $(f \circ P_n)$ be the composition with P_n. Then we define the Feynman path integral $\mathcal{J}[f]$ by

$$\mathcal{J}[f] = \lim_{n\to\infty} \mathcal{J}_n[f \circ P_n], \tag{8}$$

whenever this limit exists. We say that $f \in \mathcal{J}(P_\infty H)$ iff this last limit exists.

The above formulation is simpler than the one given in Ref. 1, which was introduced as an extension of $\mathcal{J}_{\mathrm{DAH}}$, the DeWitt and Albeverio, and Høegh-Krohn[7] definition of the Feynman integral (see definitions following Theorem 2). The formulation in Ref. 1 can be obtained by writing down the Parseval identity for the right-hand side of Eq. (7) before taking the limit $n \to \infty$ in Eq. (8). The definition given here avoids the complication of taking the Fourier transform of the functional $(f \circ P_n)$. However, we shall see that the new formulation is equivalent to that in Ref. 1 for all reasonable functionals. Because of its simplicity we shall use the new definition of $\mathcal{J}$. Moreover, the problem of establishing the relationship between classical mechanics and quantum mechanics is easily expressed in terms of this $\mathcal{J}$.

Let $\psi(X, t)$ be the solution of the Schrödinger equation

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2}\frac{\partial^2\psi}{\partial X^2} + V[X]\psi, \tag{9}$$

with Cauchy data $\psi(X, 0) = \phi(X)$. Then proving the validity of (4) is equivalent to proving that

$$\psi(X, t) = \mathcal{J}[\exp\{-i\int_0^t V[\gamma(\tau) + X]\,d\tau\}\phi[\gamma(0) + X]]. \tag{10}$$

In the following we prove the validity of (10) for a wide class of potentials V and a wide class of initial wavefunctions ϕ. Our proofs exploit the underlying Hilbert space character of the path space H and the important property that H has a reproducing kernel $G(\sigma, \cdot) \in H$, $G(\sigma, \tau) = t - \sigma v \tau$, where $\sigma v \tau$ denotes $\sup\{\sigma, \tau\}$, the reproducing kernel property being

$$(G(\sigma, \tau), \gamma(\tau)) = \gamma(\sigma), \qquad \forall \gamma \in H, \qquad \forall \sigma \in [0, t]. \tag{11}$$

This last fact, the elementary Lemma 1 of the next section, and the previously established properties of $\mathcal{J}_{\mathrm{DAH}}$ are the crucial elements in our proofs.

2. BOUNDED CONTINUOUS POTENTIALS

In this section we prove the validity of Eq. (10) for potentials V and initial wavefunctions ϕ which are the Fourier transforms of measures of bounded absolute variation. The Fourier transforms of measures of bounded absolute variation are necessarily bounded and uniformly continuous. When V is the Fourier transform of a measure of bounded absolute variation we shall simply refer to V as a bounded continous potential, even though this is not sufficient to characterize V completely. We establish this result by proving a theorem (Theorem 4) connecting our path integral $\mathcal{J}$ with the path integral $\mathcal{J}_{\mathrm{DAH}}$ of DeWitt, Albeverio, and Høegh-Krohn. We first give a concrete realization of the space of paths H and establish the reproducing kernel property for H in Theorem 1. We then prove a number of technical theorems and lemmas by exploiting the results of Theorem 1. These will lead naturally to Theorem 4.

Theorem 1: H is a real separable Hilbert space in the inner product norm topology. $\gamma \in H$ iff $\exists$ constants $\alpha_0, \alpha_n, \beta_n \in \mathcal{R}$ with $\sum_1^\infty(\alpha_n^2 + \beta_n^2) < \infty$, such that

$$\gamma(\tau) = \alpha_0(\tau - t) + \sum_{n=1}^{\infty} \frac{\alpha_n t}{2\pi n} \sin\left(\frac{2\pi n\tau}{t}\right) + \sum_{n=1}^{\infty} \frac{\beta_n t}{2\pi n}\left[1 - \cos\left(\frac{2\pi n\tau}{t}\right)\right], \quad \tau \in [0, t], \tag{12}$$

and

$$\|\gamma\|^2 = (\gamma, \gamma) = t\,\alpha_0^2 + \frac{t}{2}\sum_{n=1}^{\infty}(\alpha_n^2 + \beta_n^2) < \infty. \tag{13}$$

H has a reproducing kernel $G(\sigma, \tau) = t - \sigma v \tau$, where v denotes the maximum.

Proof: We prove first that if u is a continuous function on $(0, t)$ with weak derivative zero and $u(t) = 0$, then $u(\tau) = 0$, $\tau \in (0, t)$. If u has weak derivative zero on $(0, t)$, then

$$\langle u, \dot{v}\rangle = \int_{-\infty}^{\infty} u(\tau)\dot{v}(\tau)\,d\tau = 0, \quad \forall v \in \mathcal{D}(0, t).^9 \tag{14}$$

Let $\{h_j\}$ be an approximate identity on R^1, i.e., $\{h_j\}$ is a sequence of functions $h_j(\tau) = jh(j\tau)$, $j = 1, 2, \ldots$, $h \geq 0$, $h \in C_0^\infty$ and $\int h(\tau)\,d\tau = 1$. Define U^ϵ by $U^\epsilon(\tau) = \int_\epsilon^\tau u(\tau')\,d\tau'$, $\tau \in [\epsilon, t - \epsilon]$, $U^\epsilon(\tau) = 0$, otherwise; $t/2 > \epsilon > 0$. U^ϵ is continuously differentiable on $(\epsilon, t - \epsilon)$ and U^ϵ has compact support $\subset [\epsilon, t - \epsilon]$. Then, defining $v_j = (h_j * U^\epsilon)$, $v_j(\tau) \to U^\epsilon(\tau)$, as $j \to \infty$, $\tau \in (\epsilon, t - \epsilon)$ and, for sufficiently large j, $v_j \in \mathcal{D}(0, t)$. Also partial integration gives

$$\begin{aligned}\dot{v}_j(\tau) &= -\int_\epsilon^{t-\epsilon} \frac{d}{d\tau'} h_j(\tau - \tau')U^\epsilon(\tau')\,d\tau' \\ &= -[h_j(\tau - \tau')U^\epsilon(\tau)]_\epsilon^{t-\epsilon} + \int_\epsilon^{t-\epsilon} h_j(\tau - \tau')u(\tau')\,d\tau' \\ &= -h_j(\tau - t + \epsilon)U^\epsilon(t - \epsilon) + \int_\epsilon^{t-\epsilon} h_j(\tau - \tau')u(\tau')\,d\tau'.\end{aligned} \tag{15}$$

Hence, $v_j \in \mathcal{D}(0, t) \Rightarrow$

$$0 = \langle u, \dot{v}_j\rangle \to -u(t - \epsilon)U^\epsilon(t - \epsilon) + \int_\epsilon^{t-\epsilon} |u(\tau')|^2\,d\tau', \quad \text{as } j \to \infty. \tag{16}$$

Since u is continuous on $(0, t)$ and $\lim_{\tau\to t-} u(\tau) = u(t) = 0$, by choosing ϵ as small as we please, we obtain $\int_0^t |u(\tau')|^2\,d\tau' = 0 \Rightarrow u(\tau) = 0$ a.e. $(0, t)$. However, u is continuous $\Rightarrow S = \{\tau \in (0, t) \mid u(\tau) \neq 0\}$ is open. Thus, if $S \neq \Phi$, $\lambda[S] > 0$, where λ denotes Lebesgue measure. Therefore, $u(\tau) = 0$, $\tau \in (0, t)$.

The crux of the above is that continuous functions on the interval $(0, t)$, vanishing at one end, are uniquely determined by their weak derivatives. Therefore, $\gamma \in H$ is uniquely determined on $(0, t)$ by its weak derivative $d\gamma/d\tau \in L^2(0, t)$. However, any function $d\gamma/d\tau \in L^2(0, t)$ can be written as an a.e. convergent Fourier series

$$\frac{d\gamma}{d\tau}(\tau) = \alpha_0 + \sum_{n=1}^{\infty} \alpha_n \cos\left(\frac{2\pi n\tau}{t}\right) + \sum_{n=1}^{\infty} \beta_n \sin\left(\frac{2\pi n\tau}{t}\right), \quad \text{a.e. } \tau \in [0, t], \tag{17}$$

α_0, α_n, β_n being the usual Fourier coefficients with $\sum_1^\infty (\alpha_n^2 + \beta_n^2) < \infty$.

Consider the function γ defined by

$$\gamma(\tau) = \alpha_0(\tau - t) + \sum_{n=1}^{\infty} \frac{\alpha_n t}{2\pi n} \sin\left(\frac{2\pi n\tau}{t}\right) + \sum_{n=1}^{\infty} \frac{\beta_n t}{2\pi n}\left[1 - \cos\left(\frac{2\pi n\tau}{t}\right)\right], \quad \tau \in [0, t]. \tag{18}$$

Then the Cauchy–Schwarz inequality shows that the rhs is absolutely and uniformly convergent in $[0, t]$, and so $\gamma(\tau)$ must be continuous and clearly $\gamma(t) = 0$. We now show that γ has the weak derivative $d\gamma/d\tau$. Denote the Nth partial sum of the rhs by S_N. Then, using uniform convergence,

$$\langle \gamma, v\rangle = \lim_{N\to\infty} \langle S_N, v\rangle, \qquad \forall\, v \in \mathcal{D}(0, t)$$

$$\Rightarrow \langle \gamma, \dot{v}\rangle = \lim_{N\to\infty} \langle S_N, \dot{v}\rangle = -\lim_{N\to\infty} \langle \dot{S}_N, v\rangle, \qquad \forall\, v \in \mathcal{D}(0, t). \tag{19}$$

However, $\int_0^t |\dot{S}_N - d\gamma/d\tau|^2 d\tau \to 0$ as $N \to \infty$ and the Cauchy–Schwarz inequality for integrals yields

$$\lim_{N\to\infty} \langle \dot{S}_N, v\rangle = \langle d\gamma/d\tau, v\rangle, \qquad \forall\, v \in \mathcal{D}(0, t). \tag{20}$$

Combining the last two equations, we arrive at

$$\langle \gamma, \dot{v}\rangle = -\langle d\gamma/d\tau, v\rangle, \qquad \forall\, v \in \mathcal{D}(0, t), \tag{21}$$

and γ has weak derivative $d\gamma/d\tau$ as asserted. From the first part of the theorem γ is the unique continuous function on $(0, t)$ with $\gamma(t) = 0$ and weak derivative $d\gamma/d\tau$.

It follows trivially that

$$\|\gamma\|^2 = \int_0^t \frac{d\gamma}{d\tau}\frac{d\gamma}{d\tau} d\tau = t\alpha_0^2 + \frac{t}{2}\sum_1^\infty (\alpha_n^2 + \beta_n^2) < \infty. \tag{22}$$

This proves that H is a real separable Hilbert space in the inner product norm topology.

The reproducing kernel property follows because $(dG/d\tau)(\sigma, \tau)$, the weak derivative of $G(\sigma, \tau) = t - \sigma v\tau$, v being the maximum, is given by $(dG/d\tau)(\sigma, \tau) = -\theta(\tau - \sigma)$ and $\forall\, \gamma \in H$

$$(G(\sigma, \tau), \gamma(\tau)) = -\int_\sigma^t \frac{d\gamma}{d\tau} d\tau = \gamma(\sigma), \tag{23}$$

the last step following by integrating on a.e. convergent Fourier series term by term.[10] This proves the theorem. We also need Theorem 2.

Theorem 2: The linear map $P_n : H \to H$ is a projection, and, if $I : H \to H$ denotes the identity, then $P_n \xrightarrow{s} I$, in the strong operator topology on $\mathcal{L}(H, H)$, as $n \to \infty$.

Proof: From the definition of P_n it follows easily that $P_n^2 = P_n$. In terms of the reproducing kernel $G(\sigma, \tau) = t - \sigma v\tau$, v being the maximum, we have the important identity

$$(P_n\gamma)(\tau) = \sum_{j=0}^{n-1}\left[G\left(\frac{j+1}{n}t, \tau\right) - G\left(\frac{jt}{n}, \tau\right)\right][\gamma_{j+1} - \gamma_j]\Delta t^{-1}, \tag{24}$$

where $\gamma_j = \gamma(jt/n)$, $j = 0, 1, 2, \ldots, n$.

From the reproducing kernel property, $\forall\, \gamma,\ \gamma' \in H$,

$$(\gamma', P_n\gamma) = \sum_{j=0}^{n-1} (\gamma'_{j+1} - \gamma'_j)(\gamma_{j+1} - \gamma_j)\Delta t^{-1} = (P_n\gamma', \gamma). \tag{25}$$

The closed graph theorem now implies that P_n is a projection.[11]

To complete the proof, we are required to prove that $V = \{\gamma \in H \mid \|P_n\gamma - \gamma\| \to 0 \text{ as } n \to \infty\} = H$. For a proof of this result we refer to Theorem 4 of Ref. 1.

Following Albeverio and Høegh–Krohn,[12] we now introduce a space of measures on H. We choose as a convenient σ-field on H the σ-field generated by the subsets of H open in the inner product norm topology.

Definition: $M(H)$ is the space of complex-valued measures of bounded absolute variation on H, $\mu \in M(H)$, iff $\|\mu\| = \int |d\mu| < \infty$. $\|\ \|$ is a norm on $M(H)$.

We also require the space of functionals on H, each of which is the Fourier transform of a measure in $M(H)$.

Definition: The space of functionals $\mathcal{J}(H)$ is defined by $f \in \mathcal{J}(H)$ iff $f[\gamma] = \int \exp[-i(\gamma', \gamma)]\, d\mu(\gamma')$, $\mu \in M(H)$. We now define $\mathcal{J}_{\mathrm{DAH}}$ the path integral of DeWitt, Albeverio, Høegh–Krohn.

Definition: When $f \in \mathcal{J}(H)$, $f[\gamma] = \int \exp[-i(\gamma', \gamma)]\, d\mu(\gamma')$, $\mu \in M(H)$, $\mathcal{J}_{\mathrm{DAH}}(f)$ is defined by

$$\mathcal{J}_{\mathrm{DAH}}(f) = \int \exp[-(i/2)\|\gamma\|^2]\, d\mu(\gamma). \tag{26}$$

We remark here that $\mu \in M(H)$ ensures $\mathcal{J}_{\mathrm{DAH}}(f)$ exists. For the continuous function $\exp[-(i/2)\|\gamma\|^2]$ is Borel measurable and

$$|\mathcal{J}_{\mathrm{DAH}}(f)| \le \int |d\mu(\gamma')| = \|\mu\| \stackrel{\mathrm{def}}{=} \|f\|_0. \tag{27}$$

It is not difficult to establish that $\|\ \|_0$ is a norm on $\mathcal{J}(H)$. Also, the separability of H implies that if $f \in \mathcal{J}(H)$ is the Fourier transform of the measure $\mu \in M(H)$, then μ is uniquely determined by f. Thus, $\mathcal{J}_{\mathrm{DAH}}$ is well defined. The important properties of $M(H)$, $\mathcal{J}_{\mathrm{DAH}}$ are given in Refs. 1 and 6c. The property which we are particularly interested in is the content of the next theorem due to Albeverio and Høegh–Krohn.[13]

Theorem 3: The solution of the Schrödinger equation

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2}\frac{\partial^2\psi}{\partial X^2} + V[X]\psi \tag{28}$$

with Cauchy data $\psi(X, 0) = \phi(X) = \int \exp(i\alpha X) d\nu(\alpha) \in L^2(R^1)$, with a real-valued potential $V[X] = \int \exp(i\alpha X)\, d\mu(\alpha)$, μ, ν being of bounded absolute variation on R^1, is

$$\psi(X,t)=\mathcal{J}_{\mathrm{DAH}}[\exp\{-i\int_0^t V[\gamma(\tau)+X]\,d\tau\}\phi[\gamma(0)+X]], \tag{29}$$

$\exp\{-i\int_0^t V[\gamma(\tau)+X]\,d\tau\}\phi[\gamma(0)+X]\in\mathcal{J}(H)$.

Proof: See the proof of the Feynman–Itô formula in Ref. 1 or 6(c).

Before proving the main results of this section we need an elementary but important technical lemma.

Lemma 1: For real $a, b>0$ $\exists$ a finite constant $C(b)$ such that

$$\left|\int_0^a \exp(ibt^2)\,dt\right|\leq C(b), \tag{30}$$

uniformly $a\in(0,\infty)$.

Proof: Consider the simple closed contour C in the complex t plane:

$$C=\{t\,|\arg t=0,\ 0\leq|t|\leq a;\quad 0\leq\arg t\leq\pi/4,$$
$$|t|=a;\ \arg t=\pi/4,\quad 0\leq|t|\leq a\}.$$

Then Cauchy's theorem implies $\oint_c \exp(ibt^2)\,dt=0$. Denote the curved part of C by c; then we obtain

$$\int_0^a \exp(ibt^2)\,dt+\int_c \exp(ibt^2)\,dt-\int_0^a \exp(-bt^2)\,dt=0. \tag{31}$$

Clearly, for $a\in(0,\infty)$,

$$\left|\int_0^a \exp(-bt^2)\,dt\right|\leq\int_0^\infty \exp(-bt^2)\,dt=(\pi/4b)^{1/2}. \tag{32}$$

We must establish a similar bound for $\int_c \exp(ibt^2)\,dt$. Putting $t=a\exp(i\theta)$ gives

$$\left|\int_c \exp(ibt^2)\,dt\right|\leq a\int_0^{\pi/4}\exp(-ba^2\sin 2\theta)\,d\theta$$
$$=(a/2)\int_0^{\pi/2}\exp(-ba^2\sin u)\,du. \tag{33}$$

However, for $0\leq u\leq\pi/2$, $(\sin u)/u\geq 2/\pi$, and so we obtain

$$\left|\int_c \exp(ibt^2)\,dt\right|\leq(a/2)\int_0^{\pi/2}\exp[-2ba^2u/\pi]\,du$$
$$=(\pi/4ab)[1-\exp(-ba^2)]. \tag{34}$$

When $a\in(0,1)$, applying the mean-value theorem to the rhs of the above gives

$$\left|\int_c \exp(ibt^2)\,dt\right|\leq\pi a/4<\pi/4. \tag{35}$$

When $a\in[1,\infty)$, consider $f(a)=a^{-1}(1-\exp(-ba^2))$. Then $f(a)$ is continuously differentiable on $[1,\infty)$ and $f(a)\to 0$ as $a\to\infty$. Also, $f'(a)=0\Rightarrow\exp(-ba^2)=(2a^2b+1)^{-1}$. It follows that, for $a\in[1,\infty)$,

$$\left|\int_c \exp(ibt^2)\,dt\right|\leq(\pi/4ab)[2a^2b/(2a^2b+1)]<\pi/4b. \tag{36}$$

Putting $C(b)=(\pi/4b)^{1/2}+(\pi/4)v(\pi/4b)$, where v is the maximum, proves the lemma.

We now give the main result of this section.

Theorem 4: $\mathcal{J}(H)\subset\mathcal{J}(P_\infty H)$ and $\mathcal{J}$ is an extension of $\mathcal{J}_{\mathrm{DAH}}$.

Proof: Let $f\in\mathcal{J}(H)$ be given by $f[\gamma]=\int\exp[-i(\gamma',\gamma)]\times d\mu(\gamma')$, $\mu\in M(H)$; then we prove as a first step that

$$\mathcal{J}_n[f\circ P_n]=\int\exp[-(i/2)(\gamma,P_n\gamma)]\,d\mu(\gamma). \tag{37}$$

From Eq. (24) we observe that

$$[f\circ P_n][\gamma]=\int\exp\left(-i\sum_{j=0}^{n-1}(\gamma'_{j+1}-\gamma'_j)(\gamma_{j+1}-\gamma_j)\Delta t^{-1}\right)d\mu(\gamma'), \tag{38}$$

where $\gamma'_j=\gamma'(jt/n)$, $\gamma_j=\gamma(jt/n)$, $j=0,1,2,\ldots,n$, and $\Delta t=t/n$.

Fubini's theorem implies that

$$N_n\int_{-R}^{R}\cdots\int_{-R}^{R}\exp\left(\frac{i}{2}\sum_{j=0}^{n-1}(\gamma_{j+1}-\gamma_j)^2\Delta t^{-1}\right)[f\circ P_n][\gamma]\,d^n\gamma$$
$$=\int d\mu(\gamma')f_R[\gamma'], \tag{39}$$

where $f_R[\gamma']$ is given by

$$f_R[\gamma']=N_n\exp\left(-\frac{i}{2}\sum_{j=0}^{n-1}(\gamma'_{j+1}-\gamma'_j)^2\Delta t^{-1}\right)$$
$$\times\int_{-(R+\gamma'_0)}^{R-\gamma'_0}\int_{-(R+\gamma'_{n-1})}^{R-\gamma'_{n-1}}\exp\left(\frac{i}{2}\sum_{j=0}^{n-1}(\gamma''_{j+1}-\gamma''_j)^2\Delta t^{-1}\right)d^n\gamma'', \tag{40}$$

γ''_j being defined by $\gamma''_j=\gamma_j-\gamma'_j$, $j=0,1,2,\ldots,n-1$, $d^n\gamma''=d\gamma''_0\cdots d\gamma''_{n-1}$. However, from Lemma 1 we deduce $\exists$ a constant M, independent of R and γ', such that

$$|f_R[\gamma']|\leq M. \tag{41}$$

What is more, we easily see that

$$f_R[\gamma']\to\exp\left(-\frac{i}{2}\sum_{j=0}^{n-1}(\gamma'_{j+1}-\gamma'_j)^2\Delta t^{-1}\right)$$
$$=\exp\left(-\frac{i}{2}(\gamma',P_n\gamma')\right), \tag{42}$$

as $R\to\infty$. The dominated convergence theorem for the measure μ then yields

$$\mathcal{J}_n[f\circ P_n]=\lim_{R\to\infty}N_n\int_{-R}^{R}\cdots\int_{-R}^{R}\exp\left(\frac{i}{2}\sum_{j=0}^{n-1}(\gamma_{j+1}-\gamma_j)^2\Delta t^{-1}\right)$$
$$\times[f\circ P_n][\gamma]d^n\gamma=\int\exp\left(-\frac{i}{2}(\gamma',P_n\gamma')\right)d\mu(\gamma'), \tag{43}$$

as asserted.

We have already seen that, $\forall\,\gamma'\in H$, $\|P_n\gamma'-\gamma'\|^2=|(\gamma',\gamma')-(\gamma',P_n\gamma')|\to 0$, as $n\to\infty$. It follows that, $\forall\,\gamma'\in H$,

$$|\exp[-(i/2)(\gamma',P_n\gamma')]-\exp[-(i/2)(\gamma',\gamma')]|\to 0,$$
$$\text{as } n\to\infty. \tag{44}$$

Since $|\exp[-(i/2)(\gamma',P_n\gamma')]-\exp[-(i/2)(\gamma',\gamma')]|\leq 2$, a second application of the dominated convergence theorem for the measure μ yields

$$|\mathcal{J}_n[f\circ P_n]-\mathcal{J}_{\mathrm{DAH}}[f]|\leq\int|\exp[(-i/2)(\gamma',P_n\gamma')]$$
$$-\exp[(-i/2)(\gamma',\gamma')]|\,|d\mu(\gamma')|\to 0,$$
$$\text{as } n\to\infty. \tag{45}$$

Thus, $\mathcal{J}(H)\subset\mathcal{J}(P_\infty H)$ and $\mathcal{J}$ is an extension of $\mathcal{J}_{\mathrm{DAH}}$, proving Theorem 4.

Combining the last two theorems proves the validity of Eq. (10) for potentials V and initial wavefunctions ϕ which are the Fourier transforms of measures of bounded absolute variation. Hence, as promised, we have established a precise relationship between classical mechanics and quantum mechanics—the one given in Eq. (4) for the above potentials and initial wavefunctions.

The restriction in the above result to bounded continuous potentials seems somewhat unsatisfactory. However, Theorem 4 has an elementary corollary which suggests that this restriction is artificial.

Corollary: The solution of the Schrödinger equation

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2}\frac{\partial^2\psi}{\partial X^2} + V[X]\psi \tag{46}$$

with Cauchy data $\psi(X, 0) = \phi(X) = \int \exp(i\alpha X)\, d\nu(\alpha)$, with ν a measure of bounded absolute variation on $\mathcal{R}^1$ and a real-valued potential $V[X] = BX + C$, is given by

$$\psi(X, t) = \mathcal{F}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\, d\tau\}\phi[X + \gamma(0)]]. \tag{47}$$

Proof: A distinguished role is played by the classical path $\gamma^\alpha \in H$ defined by

$$\begin{aligned}\gamma^\alpha(\tau) &= (\alpha - Bt)(\tau - t) - (B/2)(\tau - t)^2 \\ &= -(\alpha - Bt)G(\tau, 0) - (B/2)(\tau - t)^2, \quad \tau \in (0, t),\end{aligned} \tag{48}$$

where $\dot\gamma^\alpha(\tau) = -B$, $\dot\gamma^\alpha(\tau = 0) = \alpha$, and $\gamma^\alpha(\tau = t) = 0$. [Note that the map $\mathcal{R} \to H$ defined by $\alpha \to \gamma^\alpha(\cdot)$ is continuous and $\{\alpha \mid \gamma^\alpha \in O \subset H,\ O \text{ open}\}$ is therefore open.]

The following identity obtained by partial integration explains why γ^α is important in this problem. For $V[X] = BX + C$,

$$\alpha[X + \gamma(0)] - \int_0^t V[X + \gamma(\tau)]\, d\tau = \alpha X - (BX + C)t - (\gamma^\alpha, \gamma), \tag{49}$$

where $(\ ,\ )$ is the inner product on H. We can now deduce that $f[\gamma] = \exp\{-i\int_0^t V[X + \gamma(\tau)]\, d\tau\}\phi[X + \gamma(0)] \in \mathcal{F}(H)$, when $\phi(X) = \int \exp(i\alpha X)\, d\nu(\alpha)$, ν being of bounded absolute variation. To see this, define the complex measure $\tilde\nu$ on $\mathcal{R}^1$ by

$$\tilde\nu(A) = \exp[-i(BX + C)t] \int_{\alpha \in A} \exp(i\alpha X)\, d\nu(\alpha), \tag{50}$$

for each Borel $A \subset \mathcal{R}^1$. Consider the measure μ_H defined by

$$\mu_H(A) = \tilde\nu[\{\alpha \mid \gamma^\alpha \in A\}], \tag{51}$$

for each Borel $A \subset H$. We shall show that f is the Fourier transform of the measure $\mu_H \in M(H)$.

Let $g[\ ]$ be any real bounded continuous functional $g: H \to \mathcal{R}^1$. Define $g_\gamma(\cdot)$ the real bounded continuous function $g_\gamma: \mathcal{R}^1 \to \mathcal{R}^1$ by

$$g_\gamma(\alpha) = g[\gamma^\alpha]. \tag{52}$$

Then, according to the definition of μ_H, we obtain

$$\begin{aligned}\sum_{m=-\infty}^{m=+\infty} \frac{m}{2^n}\mu_H\left[g^{-1}\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right)\right] &= \sum_{m=-\infty}^{m=+\infty} \frac{m}{2^n}\tilde\nu\left[\left\{\alpha \,\middle|\, \gamma^\alpha \in g^{-1}\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right)\right\}\right] \\ &= \sum_{m=-\infty}^{m=+\infty} \frac{m}{2^n}\tilde\nu\left[g_\gamma^{-1}\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right)\right].\end{aligned} \tag{53}$$

Hence, for all such real bounded continuous functionals g,

$$\int g[\gamma']\, d\mu_H(\gamma') = \int g_\gamma(\alpha)\, d\tilde\nu(\alpha). \tag{54}$$

We can then deduce that $\forall \gamma \in H$

$$\begin{aligned}\int \exp[-i(\gamma', \gamma)]\, d\mu_H(\gamma') &= \int \exp[-i(\gamma^\alpha, \gamma)]\, d\tilde\nu(\alpha) \\ &= \int \exp\{i[\alpha X - (BX + C)t \\ &\quad - (\gamma^\alpha, \gamma)]\}\, d\nu(\alpha).\end{aligned} \tag{55}$$

From Eq. (49), we obtain $\forall \gamma \in H$

$$\int \exp[-i(\gamma', \gamma)]\, d\mu_H(\gamma') = f[\gamma], \tag{56}$$

as asserted.

It is not difficult to show that $\mu_H \in M(H)$ follows from the fact that ν is of bounded absolute variation and so $f \in \mathcal{F}(H)$. Theorem 4 then gives

$$\begin{aligned}&\mathcal{F}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\, d\tau\}\phi[X + \gamma(0)]] \\ &\quad = \int \exp[-(i/2)(\gamma', \gamma')]\, d\mu_H(\gamma') \\ &\quad = \int d\nu(\alpha) \exp\{i[(X - \alpha t/2)(\alpha - Bt) - B^2t^3/6 - Ct]\},\end{aligned} \tag{57}$$

where in the last step we are using Eq. (54).

A simple Fourier transform shows that

$$\begin{aligned}&\int \exp(i\alpha\xi)G(X, \xi, t)\, d\xi \\ &\quad = \exp\{i[(X - \alpha t/2)(\alpha - Bt) - B^2t^3/6 - Ct]\},\end{aligned} \tag{58}$$

where $G(X, \xi, t) = (2\pi it)^{-1/2}\exp\{-i[(Bt/2)(X + \xi)^2 + B^2t^3/24 + Ct]\}$ is the Green's function of the original Schrödinger equation.[14] Taking Fourier transforms as in the last part of the proof of Theorem 5, we finally obtain

$$\begin{aligned}&\mathcal{F}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\, d\tau\}\phi[X + \gamma(0)]] \\ &\quad = \int G(X, \xi, t)\psi(\xi, 0)\, d\xi = \psi(X, t),\end{aligned} \tag{59}$$

proving the corollary.

In the last corollary the potential $V = (BX + C)$ is certainly not bounded, but we have seen that the precise relationship given in Eq. (4) between classical mechanics and quantum mechanics is still valid for this potential. This raises the question as to the types of potential for which Eq. (4) is true. In the next section we show that the result (4) also holds for the harmonic oscillator potentials $V = AX^2 + BX + C$, $A > 0$.

3. HARMONIC OSCILLATOR POTENTIALS

We consider the harmonic oscillator corresponding to the potential $V = AX^2 + BX + C$, $A > 0$. This potential V is unbounded and V is, therefore, not the Fourier transform of a measure of bounded absolute variation. We must now see whether the relationship (4) is valid for potentials V of this kind.

The first step in dealing with potentials V of this kind is to prove that if $G(X, \xi, t)$ is the Green's function of the corresponding Schrödinger equation with initial condition $\psi(X, 0) = \delta(X - \xi)$, then $\hat{G}(X, \alpha, t) = \lim_{\epsilon \to 0+} \int \exp(i\alpha\xi - \epsilon\xi^2) G(X, \xi, t)\, d\xi$ is given by

$$\hat{G}(X, \alpha, t) = \mathcal{J}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\, d\tau\} \times \exp\{i\alpha[X + \gamma(0)]\}]. \tag{60}$$

We have already seen that this identity is true if V is of the above type with $A = 0$. Equation (60) is also valid when V is the Fourier transform of a measure of bounded absolute variation. This follows from the results of the last section by considering the Cauchy problem with initial data $\psi(\xi, 0) = \phi_\epsilon(\xi) = \exp(i\alpha\xi - \epsilon\xi^2) \in L^2(\mathcal{R}^1)$, $\epsilon > 0$. Clearly $\phi_\epsilon(\xi)$ is also the Fourier transform of a measure of bounded absolute variation, $\epsilon \geq 0$. Therefore, if $\hat{G}_\epsilon(X, \alpha, t) = \int \exp(i\alpha\xi - \epsilon\xi^2) G(X, \xi, t)\, d\xi$, $\epsilon > 0$, we have

$$\hat{G}_\epsilon(X, \alpha, t) = \mathcal{J}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\, d\tau\} \phi_\epsilon[X + \gamma(0)]]. \tag{61}$$

In the Appendix we give a simple argument to show that $\|\phi_\epsilon[X + \gamma(0)] - \phi_0[X + \gamma(0)]\|_0 \to 0+$, as $\epsilon \to 0+$. For potentials V which are the Fourier transforms of measures of bounded absolute variation, Eq. (60) now follows from Eq. (27) and Theorem 4.

We now prove the validity of Eq. (60) for the harmonic oscillator potentials $V = AX^2 + BX + C$, $A > 0$. We then go on to prove that for potentials V of this kind, the solution of the corresponding Schrödinger equation with Cauchy data $\psi(X, 0) = \phi(X) = \int \exp(i\alpha X)\, d\nu(\alpha)$, ν again being a measure of bounded absolute variation on $\mathcal{R}^1$, is given by

$$\psi(X, t) = \mathcal{J}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\, d\tau\} \phi[X + \gamma(0)]]. \tag{62}$$

This is the result in Theorem 5. Before proving Theorem 5, however, we require a number of elementary technical lemmas—Lemmas 2, 3, 4.

Lemma 2: Let $D^{n-1}(wt/n)$ be the $(n-1)\times(n-1)$ determinant

$$D^{n-1}(wt/n) = \begin{vmatrix} 2 - 2w^2\Delta t^2/3 & -1 - w^2\Delta t^2/6 & 0 & 0 & \cdots & 0 & 0 \\ -1 - w^2\Delta t^2/6 & 2 - 2w^2\Delta t^2/3 & -1 - w^2\Delta t^2/6 & 0 & \cdots & 0 & 0 \\ 0 & -1 - w^2\Delta t^2/6 & 2 - 2w^2\Delta t^2/3 & -1 - w^2\Delta t^2/6 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & -1 - w^2\Delta t^2/6 & 2 - 2w^2\Delta t^2/3 \end{vmatrix}, \tag{63}$$

where $\Delta t = t/n$. Then, for sufficiently large n,

$$D^{n-1}(wt/n) = n(\sin wt)/wt + (wt/6)\sin wt + O(n^{-1}), \tag{64}$$

the constant in the term $O(n^{-1})$ being dependent on wt.

Proof: Expanding $D^r(wt/n)$, the corresponding $(r\times r)$ determinant $(n > r)$, by the elements of the first row,

$$D^r - aD^{r-1} + b^2D^{r-2} = 0, \quad r = 3, 4, 5, \ldots, \quad D^1 = a, \quad D^2 = a^2 - b^2, \tag{65}$$

where $a = 2 - 2w^2\Delta t^2/3$ and $b = -1 - w^2\Delta t^2/6$. Let u_+ and $u_- = u_+^*$ be the roots of the quadratic equation

$$u^2 - au + b^2 = 0. \tag{66}$$

Then, since $(a^2 - 4b^2) \neq 0$, $D^r = Au_+^r + Bu_-^r$, $r = 1, 2, \ldots$, where $Au_+ + Bu_- = a$, $Au_+^2 + Bu_-^2 = a^2 - b^2$.

It follows that

$$u_\pm = a \pm (a^2 - 4b^2)^{1/2}/2 = 1 - w^2\Delta t^2/3 \pm iw\Delta t(1 - w^2\Delta t^2/12)^{1/2}, \quad A = B^* = (-au_- + a^2 - b^2)/u_+(u_+ - u_-). \tag{67}$$

Then, for sufficiently large n,

$$u_\pm = 1 \pm iw\Delta t - w^2\Delta t^2/3 \mp i(w^3\Delta t^3)/24 + O(\Delta t^4) = (1 \pm iw\Delta t - w^2\Delta t^2/3 \mp iw^3\Delta t^3/24)[1 + O(\Delta t^4)]. \tag{68}$$

Denoting the principal branch of the logarithm by ln, we have in the cut complex z plane, the cut being $(-\infty, -1)$,

$$\ln(1 + z) = z - z^2/2 + z^3/3 - z^4/4 + \cdots, \quad |z| < 1, \tag{69}$$

where the above series is absolutely convergent in the disc, $|z| < 1$. It follows that, for sufficiently large n,

$$n\ln(u_\pm) = n[\pm iw\Delta t + w^2\Delta t^2/6 \mp iw^3\Delta t^3/24] + O(n^{-3}) = \pm iwt + w^2t^2/6n \mp iw^3t^3/24n^2 + O(n^{-3}) \tag{70}$$

and

$$u_\pm^n = \exp(\pm iwt)[1 + w^2t^2/6n + (1/n^2)(w^4t^4/72 \mp iw^3t^3/24)] + O(n^{-3}). \tag{71}$$

Also, from above, for sufficiently large n, we obtain

$$A = B^* = (1 + iw\Delta t + \tfrac{17}{24}w^2\Delta t^2)/2iw\Delta t + O(\Delta t^2). \tag{72}$$

Writing $D^{n-1}(wt/n) = (A/u_+)u_+^n + (B/u_-)u_-^n$, the required result follows from Eqs. (68), (71), and (72).

Lemma 3: Let M denote the $(n\times n)$ matrix

$$M=\begin{bmatrix} 1-w^2\Delta t^2/3 & -1-w^2\Delta t^2/6 & 0 & 0 & \cdots & 0 & 0 \\ -1-w^2\Delta t^2/6 & 2-2w^2\Delta t^2/3 & -1-w^2\Delta t^2/6 & 0 & \cdots & 0 & 0 \\ 0 & -1-w^2\Delta t^2/6 & 2-2w^2\Delta t^2/3 & -1-w^2\Delta t^2/6 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & -1-w^2\Delta t^2/6 & 2-2w^2\Delta t^2/3 \end{bmatrix}, \tag{73}$$

where $\Delta t=t/n$. Let M^{-1}_{ij} denote the (i,j)th entry of the matrix inverse M^{-1}. Then

$$M^{-1}_{11}=n(\tan wt)/wt+O(1),\quad M^{-1}_{1n}=M^{-1}_{n1}=1/\cos wt+O(n^{-1}) \tag{74}$$

and

$$M^{-1}_{nn}=1+(wt/n)\tan wt+O(n^{-2}),\quad wt\neq m\pi/2,\quad m=0,1,2,\cdots. \tag{75}$$

Proof: First of all we prove $\det M=\cos wt+O(n^{-1})$, for sufficiently large n. From the previous lemma

$$D^{n-1}=2\text{R. P.}[(\exp(iwt)/2iw\Delta t)(1+w^2\Delta t^2)(1+w^2t\Delta t)/6]+O(\Delta t),$$

$$D^{n-2}=2\text{R. P.}[(\exp(iwt)/2iw\Delta t)(1+w^2\Delta t^2)[(1+w^2t\Delta t)/6](1-iw\Delta t)]+O(\Delta t). \tag{76}$$

Splitting up the third bracket in D^{n-2}, gives

$$D^{n-2}=D^{n-1}-\text{R. P.}[\exp(iwt)(1+w^2\Delta t^2)(1+w^2t\Delta t/6)]+O(\Delta t)=D^{n-1}-\cos wt+O(\Delta t). \tag{77}$$

Expanding $\det M$ by the first row, we obtain

$$\det M=D^{n-1}-D^{n-2}+O(\Delta t)=\cos wt+O(n^{-1}). \tag{78}$$

We then have

$$M^{-1}_{11}=D^{n-1}/\det M=(n\tan wt)/wt+O(1),\quad M^{-1}_{1n}=M^{-1}_{n1}=(-1)^{n-1}(-1-w^2\Delta t^2/6)^{n-1}/\det M=1/\cos wt+O(n^{-1}), \tag{79}$$

$wt\neq(m+\frac{1}{2})\pi$, $m=0,1,2,\cdots$. Further, from the above expression for the matrix M, we obtain

$$M^{-1}_{nn}=[(1-w^2\Delta t^2/3)D^{n-2}-(1+w^2\Delta t^2/6)^2D^{n-3}]/[(1-w^2\Delta t^2/3)D^{n-1}-(1+w^2\Delta t^2/6)^2D^{n-2}]. \tag{80}$$

Substituting for D^{n-1} in terms of D^{n-2} and D^{n-3} from the recurrence relation for D^r, we obtain

$$M^{-1}_{nn}=[1-w^2\Delta t^2/3-(1+w^2\Delta t^2/6)^2D^{n-3}/D^{n-2}]/[2(1-w^2\Delta t^2/3)^2-(1+w^2\Delta t^2/6)^2-(1-w^2\Delta t^2/6)^2D^{n-3}/D^{n-2}]. \tag{81}$$

Arguing as in Eqs. (76) and (77), we easily deduce

$$D^{n-3}=D^{n-2}-\cos wt+\alpha(wt)w\Delta t+O(\Delta t^2), \tag{82}$$

where α is a finite real-valued function, whose detailed structure is unimportant. From the previous lemma, we obtain, for a finite real-valued β,

$$D^{n-3}/D^{n-2}=1-w\Delta t\cot wt+w^2\Delta t^2\beta(wt)+O(\Delta t^3),\quad wt\neq m\pi,\quad m=1,2,\cdots. \tag{83}$$

Substitution of this expression into Eq. (81) leads to the desired result for M^{-1}_{nn}.

Lemma 4: Let $f_\alpha[\gamma]=\exp\{-i\int_0^t(w^2/2)[X+\gamma(\tau)]^2\,d\tau\}\exp\{i\alpha[X+\gamma(0)]\}$. Then, for $wt\neq m\pi/2$, $m=1,2,\cdots$,

$$\mathcal{J}[f_\alpha[\gamma]]=\lim_{n\to\infty}\mathcal{J}_n[f_\alpha\circ P_n]=(\cos wt)^{-1/2}\exp\{-(i/2)[(\alpha^2/w+wX^2)\tan wt-2\alpha X\sec wt]\}. \tag{84}$$

Proof: First of all we have

$$\int_0^t[X+P_n\gamma(\tau)]^2\,d\tau=\sum_{j=0}^{n-1}\int_{jt/n}^{(j+1)t/n}\left[X+\gamma_j+\left(\tau-\frac{jt}{n}\right)(\gamma_{j+1}-\gamma_j)\frac{n}{t}\right]^2 d\tau=\tfrac{1}{3}\sum_{j=0}^{n-1}(\gamma_j'^2+\gamma_j'\gamma_{j+1}'+\gamma_{j+1}'^2)\frac{t}{n}, \tag{85}$$

where $\gamma_j'=X+\gamma_j$ and $\gamma_j=\gamma(jt/n)$, $j=0,1,2,\ldots,n$, γ being normalized so that $\gamma(t)=0$. Therefore, we arrive at

$$\mathcal{J}_n[f_\alpha\circ P_n]=N_n\int\exp\left(\frac{i}{2}\sum_{j=0}^{n-1}\frac{(\gamma_{j+1}'-\gamma_j')^2}{\Delta t}-\frac{iw^2}{2}\sum_{j=0}^{n-1}\frac{(\gamma_j'^2+\gamma_j'\gamma_{j+1}'+\gamma_{j+1}'^2)\Delta t}{3}+i\alpha\gamma_0'\right)d^n\gamma', \tag{86}$$

where $\Delta t=t/n$. Putting $\gamma_j'=\gamma_j(\Delta t)^{1/2}$, $j=0,1,2,\ldots,n-1$, gives

$$\mathcal{J}_n[f_\alpha\circ P_n]=N_n(\Delta t)^{n/2}\exp(-iw^2\Delta tx^2/6+ix^2/2\Delta t)\int\exp[(i/2)\gamma^TM\gamma]\exp(i\gamma^Tc)\,d^n\gamma, \tag{87}$$

where M is defined as in the previous lemma, $\gamma^T=(\gamma_0,\ldots,\gamma_{n-1})$ and

$$c^T=(\alpha(\Delta t)^{1/2},0,0,\ldots,-x/(\Delta t)^{1/2}-w^2(\Delta t)^{3/2}x/6).$$

For a real $(n\times n)$ nonsingular symmetric matrix M and a real $(n\times 1)$ column vector c,

$$\int \exp[(i/2)\,\gamma^T M\gamma]\exp(i\gamma^T c)\,d^n\gamma = (2\pi i)^{n/2}(\det M)^{-1/2}\exp[-(i/2)(c^T M^{-1}c)], \tag{88}$$

where on the rhs we choose the branch of the square root to be positive on the positive reals.

Thus, we arrive at

$$\mathcal{J}_n[f_\alpha \circ P_n] = (\det M)^{-1/2}\exp(-iw^2\Delta t x^2/6 + ix^2/2\Delta t)\,\exp[(-i/2)(c_1^2 M_{11}^{-1} + 2c_1c_n M_{1n}^{-1} + c_n^2 M_{nn}^{-1})]. \tag{89}$$

Using the previous lemma, for fixed values of X and t, we obtain

$$\mathcal{J}_n[f_\alpha \circ P_n] = (\cos wt)^{-1/2}\exp\{-(i/2)[(\alpha^2/w + wX^2)\tan wt - 2\alpha X \sec wt]\}\exp[O(n^{-1}) + O(\alpha^2 n^{-1}) + O(\alpha n^{-1})] \tag{90}$$

$wt \neq m\pi/2$, $m = 1, 2, \cdots$. This proves the lemma and Eq. (80) for the potential $V = w^2X^2/2$.

We now prove the main result of this section—Theorem 5.

Theorem 5: The solution of the Schrödinger equation

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2}\frac{\partial^2\psi}{\partial X^2} + V[X]\psi, \tag{91}$$

with Cauchy data $\psi(X, 0) = \phi(X) = \int \exp(i\alpha X)\,d\nu(\alpha)$, ν being a measure of bounded absolute variation, with a real-valued potential $V[X] = AX^2 + BX + C$, $A > 0$, is

$$\psi(X, t) = \mathcal{J}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\,d\tau\}\phi[X + \gamma(0)]], \tag{92}$$

$(2A)^{1/2}t \neq m\pi/2$, $m = 1, 2, \cdots$.

Proof: We write $V[X] = AX^2 + BX + C = A(X + B/2A)^2 + (C - B^2/4A) = w^2x'^2/2 + w_0$, where $X' = X + B/2A$, $w = (2A)^{1/2}$, and $w_0 = (C - B^2/4A)$. Then, putting $\psi'(X', t) = \exp(iw_0 t)\psi(X, t)$, reduces the Schrödinger equation to

$$i\frac{\partial\psi'}{\partial t} = -\frac{1}{2}\frac{\partial^2\psi'}{\partial X'^2} + \frac{w^2X'^2\psi'}{2}. \tag{93}$$

With Cauchy data $\psi'(X', 0) = \phi'(X') = \psi(X, 0) = \phi(X)$. We shall show that, for $wt \neq m\pi/2$, $m = 1, 2, \cdots$,

$$\psi'(X', t) = \mathcal{J}[\exp\{-i\int_0^t (w^2/2)[X' + \gamma(\tau)]^2\,d\tau\}\phi'[X' + \gamma(0)]], \tag{94}$$

for $\phi'(X') = \phi(X) = \int \exp(i\alpha X)\,d\nu(\alpha)$, ν being of bounded absolute variation on R^1. It will then follow that

$$\psi(X, t) = \exp(-iw_0 t)\psi'(X', t) = \mathcal{J}[\exp\{-i\int_0^t V[X + \gamma(\tau)]\,d\tau\}\phi[X + \gamma(0)]], \tag{95}$$

$\phi(X) = \int \exp(i\alpha X)\,d\nu(\alpha)$. In establishing the validity of equation (94) we drop all primes.

Define the functional $f_\alpha[\gamma]$ by $f_\alpha[\gamma] = \exp\{-i\int_0^t (w^2/2)[X + \gamma(\tau)]^2\,d\tau\}\exp\{i\alpha[X + \gamma(0)]\}$. We must prove first that

$$\int \mathcal{J}_n[f_\alpha \circ P_n]\,d\nu(\alpha) = \mathcal{J}_n[\int (f_\alpha \circ P_n)\,d\nu(\alpha)] = \mathcal{J}_n[\int f_\alpha\,d\nu(\alpha) \circ P_n]. \tag{96}$$

The rhs of the above identity is obvious. The equality on lhs is established in the Appendix by completing the square in the exponential.

From the previous lemma, for each real α, $\mathcal{J}_n[f_\alpha \circ P_n] \to \mathcal{J}[f_\alpha]$ as $n \to \infty$, $wt \neq m\pi/2$, $m = 1, 2, \cdots$, and

$$\int |\mathcal{J}_n[f_\alpha \circ P_n]|\,|d\nu(\alpha)| \leq (\det M)^{-1/2}\int |d\nu(\alpha)| < \infty. \tag{97}$$

The dominated convergence theorem then implies, for $wt \neq m\pi/2$, $m = 1, 2, \cdots$, $\int \mathcal{J}_n[f_\alpha \circ P_n]\,d\nu(\alpha) \to \int \mathcal{J}[f_\alpha]\,d\nu(\alpha)$ as $n \to \infty$.

Hence, letting $n \to \infty$ in Eq. (96), we obtain $wt \neq m\pi/2$, $m = 1, 2, \cdots$,

$$\int \mathcal{J}[f_\alpha]\,d\nu(\alpha) = \mathcal{J}[\int f_\alpha\,d\nu(\alpha)]. \tag{98}$$

The last equation is equivalent to

$$(\cos wt)^{-1/2}\int \exp\{-(i/2)[(\alpha^2/w + wX^2)\tan wt - 2\alpha X\sec wt]\}\,d\nu(\alpha) = \mathcal{J}[\exp\{-i\int_0^t (w^2/2)[X + \gamma(\tau)]^2\,d\tau\}\phi[X + \gamma(0)]], \tag{99}$$

where $\phi(X) = \int \exp(i\alpha X)\,d\nu(\alpha)$, for any measure ν of bounded absolute variation, $wt \neq m\pi/2$, $m = 1, 2, \cdots$.

We now take the Fourier transform and write

$$\begin{aligned}(\cos wt)^{-1/2}\exp\{-(i/2)[(\alpha^2/w + wX^2)\tan wt - 2\alpha X\sec wt]\} &= \lim_{R\to\infty}\left(\frac{w}{2\pi i\sin wt}\right)^{1/2}\int_{-R}^{R}\exp(i\alpha\xi) \\ &\quad\times\exp\{(iw/2\sin wt)[(X^2 + \xi^2)\cos wt - 2X\xi]\}\,d\xi \\ &= \lim_{R\to\infty} f_R(\alpha), \quad wt \neq m\pi/2, \quad m = 1, 2, \cdots.\end{aligned} \tag{100}$$

Therefore, $wt \neq m\pi/2$, $m = 1, 2, \cdots$,

$$\mathcal{J}[\exp\{-i\int_0^t (w^2/2)[X+\gamma(\tau)]^2 d\tau\}\phi[X+\gamma(0)]] = \int \lim_{R\to\infty} f_R(\alpha)\, d\nu(\alpha), \tag{101}$$

where, from Lemma 1, $wt \neq m\pi/2$, $|f_R(\alpha)| \leq M$, which is independent of R and α. Using the fact that ν is of bounded absolute variation, the dominated convergence theorem and Fubini's theorem give

$$\begin{aligned}\mathcal{J}[\exp\{-i\int_0^t (w^2/2)[X+\gamma(\tau)]^2 d\tau\}\phi[X+\gamma(0)]] &= \lim_{R\to\infty}\int f_R(\alpha)\, d\nu(\alpha)\\ &= \lim_{R\to\infty}\int_{-R}^{R} (w/2\pi i \sin wt)^{1/2} d\xi \exp\{(iw/2\sin wt)[(X^2+\xi^2)\cos wt - 2X\xi]\}\\ &\quad\times\int \exp(i\alpha\xi)\, d\nu(\alpha).\end{aligned} \tag{102}$$

However,

$$(w/2\pi i \sin wt)^{1/2}\exp\{(iw/2\sin wt)[(X^2+\xi^2)\cos wt - 2X\xi]\} = G(X,\xi,t), \tag{103}$$

where G is the well-known Green's function for the harmonic oscillator.[15] Finally then, we obtain

$$\mathcal{J}[\exp\{-i\int_0^t (w^2/2)[X+\gamma(\tau)]^2 d\tau\}\phi[X+\gamma(0)]] = \int G(X,\xi,t)\psi(\xi,0)\, d\xi = \psi(X,t), \quad (2A)^{1/2}t = wt \neq m\pi/2, \tag{104}$$

proving the theorem.

4. CONCLUSION

It is a simple matter to deduce from the above work the results corresponding to the case $\hbar \neq 1$, $m \neq 1$, in the Schrödinger equation. We summarize these results in two theorems.

Theorem 6: Consider a particle of mass m moving in the potential V. Let $\psi(x,t)$ be the quantum mechanical amplitude for observing the particle to be at x at time t. Let the polygonal path $(P_n\gamma + x)$ be defined as in the Introduction, and let $S_{cl}[P_n\gamma + x]$ denote the classical action of the particle of mass m moving in the potential V along the polygonal path $(P_n\gamma + x)$, the particle moving with constant velocity along each of the segments of $(P_n\gamma + x)$, so that

$$S_{cl}[P_n\gamma + x] = \frac{m}{2}\sum_{j=0}^{n-1}\frac{(\gamma_{j+1}-\gamma_j)^2}{\Delta t} - \int_0^t V[P_n\gamma + x]\, d\tau. \tag{105}$$

Then, if the potential V is either (an)harmonic, $V = Ax^2 + Bx + C$, $A > 0$, or V is the Fourier transform of a measure of bounded absolute variation and if $\psi(x,0) \in L^2(\mathcal{R}^1)$ is the Fourier transform of a measure of bounded absolute variation, we have

$$\psi(x,t) = \lim_{n\to\infty} N_n \int d^n\gamma \exp\{(i/\hbar)S_{cl}[P_n\gamma + x]\}\psi(\gamma_0 + x, 0), \tag{106}$$

where $N_n = (2\pi i\hbar t/nm)^{-n/2}$ is a normalization constant and $d^n\gamma = d\gamma_0\cdots d\gamma_{n-1}$, each integration being from $-\infty$ to $+\infty$.

In the case $V = Ax^2 + Bx + C$, we must stipulate $(2A)^{1/2}t \neq m\pi/2$ $(A > 0)$, $m = 1, 2, \cdots$, and we can lift the restriction $\psi(x,0) \in L^2(\mathcal{R}^1)$.

The analog of this result for the Green's function has a compelling simplicity.

Theorem 7: Let $G(x,\xi,t)$ be the Green's function for the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi, \tag{107}$$

with initial conditions $\psi(x,0) = \delta(x-\xi)$ and define $\hat{G}(x,\alpha,t) = \lim_{\epsilon\to 0+}\int \exp(i\alpha\xi - \epsilon\xi^2)G(x,\xi,t)\, d\xi$. Then, if the potential V is either (an)harmonic, or if V is the Fourier transform of a measure of bounded variation, we have

$$\begin{aligned}\hat{G}(x,\alpha,t) = \lim_{n\to\infty} N_n \int d^n\gamma \exp\{(i/\hbar)S_{cl}[P_n\gamma + x]\}\\ \times\exp[i\alpha(\gamma_0 + x)],\end{aligned} \tag{108}$$

where N_n is the above normalization constant and $S_{cl}[P_n\gamma + x]$ is the above classical action.

The above two theorems establish the promised exact relationship between classical mechanics and quantum mechanics. One could, of course, use them and some principle of stationary phase in ∞ dimensions to obtain classical mechanics from quantum mechanics in the limit as $\hbar \to 0$.[17] Personally we prefer to approach this problem by the quasiclassical representation of Ref. 1. In this connection it is worth remarking that the results of the quasiclassical representation are still valid for the new definition of $\mathcal{J}$ given here. The definition given here is less dependent on Fourier transform and is, therefore, easier to handle. Moreover, as we have seen, it is simply related to the underlying mechanics.

ACKNOWLEDGMENT

I was first introduced to reproducing kernel Hilbert spaces in a different context by Professor J.T. Lewis of the Dublin Institute of Advanced Study.[16] Later stimulus was provided by Dr. K.C. Hannabus of the Mathematical Institute, Oxford University, and Dr. A. L. Carey of the Department of Mathematical Physics, University of Adelaide.

APPENDIX A

Consider the functional $h_\epsilon[\gamma] = \phi_\epsilon[X + \gamma(0)] = \exp\{-\epsilon[X + \gamma(0)]^2\}$, $\epsilon \geq 0$. Then we prove here that $\|h_\epsilon - h_0\|_0 \to 0$, as $\epsilon \to 0+$, $\|\ \|_0$ being defined as in Eq. (27).

A Fourier transform shows that, for $\epsilon > 0$,

$$h_\epsilon[\gamma] = (4\pi\epsilon)^{-1/2} \int \exp[-i\alpha\gamma(0)] \exp(-\alpha^2/4\epsilon - i\alpha X)\, d\alpha. \tag{A1}$$

We now define the measure ν_ϵ on $\mathcal{R}^1$ by

$$\nu_\epsilon(B) = (4\pi\epsilon)^{-1/2} \int_{\alpha\in B} \exp(-\alpha^2/4\epsilon - i\alpha X)\, d\alpha, \tag{A2}$$

for each Borel $B \subset \mathcal{R}^1$. It is also convenient to introduce the continuous map $\pi : \mathcal{R}^1 \to H$ defined by

$$\pi(\alpha) = \gamma^\alpha(\cdot) \in H, \tag{A3}$$

where in this instance $\gamma^\alpha(\cdot) = \alpha G(0, \cdot)$, G being the reproducing kernel for H. Then, for $\epsilon > 0$, we define the measure μ_ϵ on H by

$$\mu_\epsilon(B) = \nu_\epsilon(\pi^{-1}B), \tag{A4}$$

for each Borel $B \subset H$. Not let g be any real-valued continuous bounded functional $g : H \to \mathcal{R}^1$. By definition of μ_ϵ

$$\begin{aligned}
&\sum_{m=-\infty}^{\infty} \frac{m}{2^n} \mu_\epsilon\left\{g^{-1}\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right)\right\} \\
&= \sum_{m=-\infty}^{\infty} \frac{m}{2^n} \nu_\epsilon\left\{(\pi^{-1}\circ g^{-1})\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right)\right\} \\
&= \sum_{m=-\infty}^{\infty} \nu_\epsilon\left\{(g\circ\pi)^{-1}\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right)\right\}.
\end{aligned}$$

It follows that, for each real bounded continuous functional g and each $\epsilon > 0$,

$$\int g[\gamma']\, d\mu_\epsilon(\gamma') = \int (g\circ\pi)(\alpha)\, d\nu_\epsilon(\alpha). \tag{A5}$$

Hence, $\forall\, \gamma \in H$, $\epsilon > 0$,

$$\begin{aligned}
&\int \exp[-i(\gamma', \gamma)]\, d\mu_\epsilon(\gamma') \\
&= \int \exp[-i(\gamma^\alpha, \gamma)]\, d\nu_\epsilon(\alpha) \\
&= (4\pi\epsilon)^{-1/2} \int \exp[-i\alpha\gamma(0)] \exp(-\alpha^2/4\epsilon - i\alpha X)\, d\alpha.
\end{aligned} \tag{A6}$$

Therefore, we obtain $\forall\, \gamma \in H$, $\epsilon > 0$,

$$\int \exp[-i(\gamma', \gamma)]\, d\mu_\epsilon(\gamma') = h_\epsilon[\gamma] \in \mathcal{F}(H). \tag{A7}$$

Also, from the above definition of μ_ϵ we see that, for each Borel $B \subset H$ and each $\epsilon > 0$,

$$|\mu_\epsilon(B)| = |\nu_\epsilon(\pi^{-1}B)| \le (4\pi\epsilon)^{-1/2} \int_{\alpha\in\pi^{-1}B} \exp(-\alpha^2/4\epsilon)\, d\alpha. \tag{A8}$$

The last inequality implies that, for each Borel $B \subset H$, $\mu_\epsilon(B) \to \mu_0(B)$, as $\epsilon \to 0+$, where for Borel $B \subset H$

$$\mu_0(B) = \begin{cases} 1, & \text{if } 0 \in B, \\ 0, & \text{otherwise,} \end{cases} \tag{A9}$$

and $h_0[\gamma] = 1 = \int \exp\{-i(\gamma', \gamma)\}\, d\mu_0(\gamma') \in \mathcal{F}(H)$.

It remains to prove that $\|\mu_\epsilon - \mu_0\| \to 0$, as $\epsilon \to 0+$. We define the measure ν_0 on $\mathcal{R}^1$ by

$$\nu_0(B) = \begin{cases} 1, & \text{if } 0 \in B, \\ 0, & \text{otherwise,} \end{cases} \tag{A10}$$

for each Borel $B \subset \mathcal{R}^1$. Then a simple argument using inequality (A8) gives

$$\|\mu_\epsilon - \mu_0\| = \int |d\mu_\epsilon - d\mu_0| = \int |d\nu_\epsilon - d\nu_0| \to 0, \tag{A11}$$

as $\epsilon \to 0+$, proving the result.

APPENDIX B

Here we show that the two integrals I and J, defined below, exist and are equal, for sufficiently large n, when $wt \neq m\pi/2$, $m = 1, 2, \cdots$:

$$I = \int \mathcal{F}_n[f_\alpha \circ P_n]\, d\nu(\alpha), \quad J = \mathcal{F}_n\left[\int (f_\alpha \circ P_n)\, d\nu(\alpha)\right], \tag{B1}$$

where f_α is the functional

$$\begin{aligned}
f_\alpha[\gamma] &= \exp\{-i\int_0^t (w^2/2)[X + \gamma(\tau)]^2\, d\tau\} \\
&\quad \times \exp\{i\alpha[X + \gamma(0)]\},
\end{aligned} \tag{B2}$$

and ν is a measure of bounded absolute variation on $\mathcal{R}^1$.

We consider the integral I first. From Lemma 4, we easily obtain

$$\begin{aligned}
I = (\det M)^{-1/2} &\exp\{-(i/2)[w^2\Delta t x^2/3 - x^2/\Delta t + M_{nn}^{-1}(x/\Delta t^{1/2} \\
&+ w^2\Delta t^{3/2}x/6)^2]\} \int \exp\{-(i/2)[\alpha^2\Delta t M_{11}^{-1} \\
&- 2\alpha(X + w^2\Delta t^2 x/6)M_{1n}^{-1}]\}\, d\nu(\alpha),
\end{aligned} \tag{B3}$$

where M is defined as in Lemma 3 and, for sufficiently large n, $wt \neq m\pi/2$, $m = 1, 2, \cdots$, $\det M \neq 0$.

Define $h(\gamma_0)$ by $h(\gamma_0) = \int \exp(i\alpha\Delta t^{1/2}\gamma_0)\, d\nu(\alpha)$. Then we easily obtain that the integral J is given by

$$\begin{aligned}
J = N_n(\Delta t)^{n/2} &\exp[-(i/2)(w^2\Delta t x^2/3 - x^2/\Delta t)] \\
&\times \int \exp[(i/2)\gamma^T M\gamma] \exp(i\gamma^T d) h(\gamma_0) d^n\gamma,
\end{aligned} \tag{B4}$$

where again M is defined as in Lemma 3, $\gamma^T = (\gamma_0, \ldots, \gamma_{n-1})$, and

$$d^T = (0, 0, \ldots, 0, -x/\Delta t^{1/2} - w^2\Delta t^{3/2}x/6).$$

We now complete the square in the exponential in the variables $\gamma_{n-1}, \gamma_{n-2}, \ldots, \gamma_0$, taken in that order, to obtain in order new "square variables" $z_1, z_2, \ldots, z_n$. Then we can see $\exists$ a real diagonal matrix Λ with entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ and a real upper left triangular matrix U with entries 1 down the secondary diagonal such that

$$U = \begin{bmatrix} U_{11} & U_{12} \cdots U_{1n-1} & 1 \\ U_{21} & U_{22} \cdots 1 & 0 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 1 & 0 \quad 0 \quad 0 & 0 \end{bmatrix},$$

$$U^{-1} = \begin{bmatrix} 0 & 0 \cdots 0 & 1 \\ 0 & 0 \cdots 1 & U_{2n}^{-1} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 1 & U_{n2}^{-1} \cdots U_{nn-1}^{-1} & U_{nn}^{-1} \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \cdots 0 & 0 \\ 0 & \lambda_2 \cdots 0 & 0 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 0 & 0 \cdots 0 & \lambda_n \end{bmatrix}. \tag{B5}$$

The matrices U and Λ are the matrices used to define the new "square variables" z given by $z^T = (z_1, \ldots, z_n)$ and $z = U\gamma$ so that

$$\gamma^T M \gamma = z^T \Lambda z, \qquad \forall \gamma \in R^n, \tag{B6}$$

$$\therefore \quad M = U^T \Lambda U. \tag{B7}$$

We shall also require the column vector $e = \Lambda^{-1}(U^T)^{-1}d$. A simple piece of algebra then yields

$$J = N_n(\Delta t)^{n/2} \exp[(-i/2)(w^2 \Delta t x^2/3 - x^2/\Delta t)]$$
$$\times \int \exp[(i/2)(z+e)^T \Lambda (z+e)] h(z_n)\, d^n z$$
$$\times \exp[(-i/2)\, d^T M^{-1} d]. \tag{B8}$$

The matrices U, Λ and e are complicated algebraic expressions in the entries of M. The properties we require are, however, easily deduced from the above.

First of all from above

$$\det M = (\det U)^2 \det \Lambda = (-1)^{n(n-1)} \det \Lambda = \lambda_1 \lambda_2 \cdots \lambda_n,$$
$$\therefore \quad \det M \neq 0 \Rightarrow \lambda_j \neq 0, \quad j = 1, 2, \ldots, n. \tag{B9}$$

Secondly,

$$\lambda_n^{-1} = M_{11}^{-1} \quad \text{and} \quad \lambda_n^{-1} U_{nn}^{-1} = M_{1n}^{-1}. \tag{B10}$$

For we have $M^{-1} = U^{-1}\Lambda^{-1}(U^T)^{-1}$ and so

$$M_{11}^{-1} = \sum_{j,k=1}^{n} U_{1j}^{-1} \Lambda_{jk}^{-1} U_{1k}^{-1} = [U_{1n}^{-1}]^2 \lambda_n^{-1} = \lambda_n^{-1},$$
$$M_{1n}^{-1} = \sum_{j,k=1}^{n} U_{1j}^{-1} \Lambda_{jk}^{-1} U_{nk}^{-1} = \lambda_n^{-1} U_{nn}^{-1}, \tag{B11}$$

where we are using $U_{1j}^{-1} = \delta_{jn}$ and $\Lambda_{jk}^{-1} = \lambda_j^{-1}\delta_{jk}$.

Thirdly,

$$e_n = M_{1n}^{-1} d_n. \tag{B12}$$

This follows because $e_n = \{\Lambda^{-1}(U^T)^{-1}d\}_n = \sum_{j,k=1}^{n} \Lambda_{nj}^{-1} U_{kj}^{-1} d_{k1}$ $= \lambda_n^{-1} U_{nn}^{-1} d_n = M_{1n}^{-1} d_n$, where we are using the fact that $d_{k1} = d_n \delta_{kn}$.

Combining these results, we arrive at

$$J = (\det M)^{-1/2} \exp\{-(i/2)[w^2 \Delta t x^2/3 - x^2/\Delta t$$
$$+ M_{nn}^{-1}(x/\Delta t^{1/2} + w^2 \Delta t^{3/2} x/6)^2]\}$$
$$\times (2\pi i/\lambda_n)^{-1/2} \int_{-\infty}^{\infty} h(z_n) \exp\{i/2 M_{11}^{-1}(z_n + M_{1n}^{-1} d_n)^2\}\, dz_n \tag{B13}$$

Comparing Eqs. (B3) and (B13), we see that proving the equality of I and J is equivalent to proving

$$\int h(u) \exp\{(i/2M_{11}^{-1})[u - M_{1n}^{-1}(x/\Delta t^{1/2} + w^2 \Delta t^{3/2} x/6)]^2\}\, du$$
$$= (2\pi i M_{11}^{-1})^{1/2} \int d\nu(\alpha) \exp\{-(i/2)[\alpha^2 \Delta t M_{11}^{-1}$$
$$- 2\alpha(X + w^2 \Delta t^2 x/6) M_{1n}^{-1}]\}. \tag{B14}$$

However, we have from Fubini's theorem

$$\text{lhs} = \lim_{R\to\infty} \int_{-R}^{R} du \exp\{(i/2M_{11}^{-1})[u - M_{1n}^{-1}(x/\Delta t^{1/2} + w^2 \Delta t^{3/2} x/6)]^2\}$$
$$\times \{\int \exp(i\alpha \Delta t^{1/2} u)\, d\nu(\alpha)\} = \lim_{R\to\infty} \int d\nu(\alpha) f_R(\alpha), \tag{B15}$$

where

$$f_R(\alpha) = \int_{-R}^{R} \exp\{(i/2M_{11}^{-1})[u - M_{1n}^{-1}(x/\Delta t^{1/2} + w^2 \Delta t^{3/2} x/6)]^2$$
$$+ i\alpha \Delta t^{1/2} u\}\, du.$$

From Lemma 1, $|f_R(\alpha)| < M$, where M is independent of α and R. Hence, the dominated convergence theorem implies

$$\text{lhs} = \int d\nu(\alpha) \lim_{R\to\infty} f_R(\alpha)$$
$$= \int d\nu(\alpha)(2\pi i M_{11}^{-1})^{1/2} \exp\{-(i/2)(\alpha^2 \Delta t M_{11}^{-1}$$
$$- 2\alpha(X + w^2 \Delta t^2 x/6) M_{1n}^{-1})\} = \text{rhs}. \tag{B16}$$

This proves that $I = J$.

[1] A. Truman, J. Math. Phys. 17, 1852 (1976).
[2] This is the convention for the Schrödinger equation. The usual convention for the heat equation is $\gamma_0 = 0$.
[3] (a) P.A.M. Dirac, *Quantum Mechanics* (Oxford U. P., London, 1930), p. 125; (b) R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[4] M. Kac, *Probability in the Physical Sciences* (Interscience, New York, 1959), Chap. IV.
[5] E. Nelson, J. Math. Phys. 5, 332 (1964).
[6] For the Trotter product formula, see M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972), p. 295.
[7] (a) C.M. DeWitt, Comm. Math. Phys. 28, 47 (1972). (b) C.M. DeWitt, Comm. Math. Phys. 37, 63 (1974). (c) C.M. DeWitt, Ann. Phys. 97, 367 (1976) and erratum 101, 682 (1976). (d) S. Albeverio and R. Høegh-Krohn, "Mathematical Theory of Feynman Path Integrals," University of Oslo preprint, Oct. 1974. (e) See also Ref. 1 above. For a review of work on the Feynman path integral and further references see J. Tarski, in *Functional Integration and Its Applications*, edited by A.M. Arthurs (Oxford U.P., London, 1975). (f) For a related definition of Feynman's path integral see Ph. Combe, G. Rideau, R. Rodriguez, and M. Sirugue-Collin, "On some mathematical problems in the definition of Feynman path integral," Marseille preprint, July 1976.
[8] E.g., $V[X] = \cos X^2$ is bounded and continuous, but is not the Fourier transform of a measure of bounded absolute variation.
[9] For an account of the test space $\mathcal{D}(0,t)$ see, for instance, W. Rudin, *Functional Analysis* (Tata McGraw-Hill, New Delhi, 1974).
[10] E.C. Titchmarsh, *The Theory of Functions* (Oxford U.P. London 1952), p. 419.
[11] See Ref. 6 above for the closed graph theorem.
[12] See Ref. 6 (c) above.
[13] See Ref. 6 (c) above and K. Itô, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1967), Vol. II, Part I, p. 145.
[14] D. ter Haar, *Selected Problems in Quantum Mechanics* (Infosearch, London, 1964), p. 142.
[15] See Ref. 13 above, p. 148.
[16] J.T. Lewis and D.E. Evans, lecture notes, Dublin Insitute of Advanced Study.
[17] *Note added in proof*: This is done rigorously by S. Albeverio and R. Høegh-Krohn in "Oscillatory Integrals and the Method of Stationary Phase in infinitely many dimensions with Applications to the Classical Limit of Quantum Mechanics," University of Oslo preprint, Sept. 1975-to appear in Inventiones Mathematuae. Also see A. Truman, "Classical Mechanics, the Diffusion (heat) Equation and Schrödinger's Equation" (to appear).

Maxwell's equations in axiomatic quantum field theory. I. Field tensor and potentials

P. J. M. Bongaarts

Instituut Lorentz, University of Leiden, Leiden, the Netherlands
(Received 13 September 1976)

An approach to the investigation of the Maxwell field in the framework of axiomatic quantum field theory is presented which employs Borchers' algebraic reformulation of Wightman theory in a modified form adapted to the special features of the electromagnetic field. This makes it possible to clarify the relation between tensor and potential field operators, the meaning and properties of different gauges, the sense in which field equations hold and the properties of state spaces with their special subspaces.

1. INTRODUCTION

In most of the standard texts on axiomatic field theory the Maxwell field is excluded right from the beginning. See, e.g., Ref. 1—3. This is because the Maxwell field in classical as well as in quantum theory exhibits complications not present in simpler cases like the Klein—Gordon and Dirac fields. In the first place there is a description in two types of variables, the field tensor $F_{\mu\nu}$ or the more convenient potential A_μ, which is, however, nonunique and admits gauge transformations. There are, furthermore, difficulties in the Lagrangian formulation and, associated with this, irregularities in the canonical quantization procedure. There is also the somewhat singular case of a zero mass representation of the Poincaré group. Last but not least there is the occurrence of an indefinite inner product in Lorentz covariant descriptions such as the Gupta—Bleuler formalism. In view of all this and of the fact that the general problems of quantum field theory are already severe enough, it is not surprising that rigorous investigations have tended to stay away from the special problems of the quantized Maxwell field. It is, nevertheless, possible to handle part of these problems in the framework of axiomatic field theory independently from the other more general problems of quantum field theory. It is to these problems that this and subsequent papers are devoted. Certain aspects of the mathematical theory of the quantized Maxwell field have been treated in earlier work. See, e.g., Refs. 4—6. (For an interesting recent approach in terms of Weyl systems, see Ref. 7.) Up till now however, the only systematic and detailed discussion of the Maxwell field in axiomatic field theory is the one given by Strocchi and Wightman.[8] (See also Rideau[9] for further developments.) The essential feature of their approach is the use of state spaces with two inner products, such as occur in the original Gupta—Bleuler formalism. One inner product is Lorentz covariant but in general not positive definite. The second inner product is positive definite but not necessarily invariant. (Situations like the Coulomb gauge appear as special cases in which the two inner products coincide.) The reason for the introduction of the second inner product seems to be the desire to keep the situation within the range of standard Hilbert space theory. A close inspection of some of the mathematical details not worked out by Strocchi and Wightman shows however that this advantage is largely illusory. The fields and other objects have rather awkward properties with respect to the Hilbert space topology. Typical of this is the fact that for the free field the operators representing the Lorentz group are not only nonunitary but even unbounded.

In this paper a different line of approach will be followed, based on an algebraic reformulation of Wightman theory developed by Borchers.[10] According to Wightman's reconstruction theorem any specific field theory is completely characterized by the vacuum expectation values of its products of field operators. In the formulation of Borchers this means that such a theory is described by a linear functional on a tensor algebra of test functions. This functional determines uniquely, through the GNS construction, all the mathematical properties of the corresponding space of state vectors and of the field operators acting in this space; in particular it provides convenient topologies in a natural manner. These features are fully retained in a slightly generalized version of the Wightman—Borchers formalism that will be employed in this paper. By relaxing the usual positivity requirement on the state functionals one obtains a mathematically satisfactory way of dealing with the occurrence of state spaces with indefinite metric. Furthermore, there will be two basic test-function algebras, one for the field tensor and a second one for the potential, related to each other by a natural homomorphism. In this framework a unified description of the quantized electromagnetic field can be given, in which it becomes clear what a choice of gauge means, why different gauges appear, what the relations between tensor field and potential field operators are, in what sense field equations are valid, and what the properties are of the representation spaces of state vectors, with their special subspaces. All the results and most of the assumptions of Strocchi and Wightman can be derived in full mathematical rigor, simply from the basic properties of Maxwell's equations. An amusing consequence will be that the classical Maxwell field is contained in the general formalism and appears as a rather special and somewhat trivial example.

In Sec. 2 a short review will be given of that part of the Borchers formalism that is needed for the purpose of this paper. In Sec. 3, the main part of the paper, a Borchers formalism for the electromagnetic field will be developed, consisting of separate tensor algebras for the $A_\mu(x)$ and $F_{\mu\nu}(x)$ fields, connected by a natural homomorphism. Properties of the corresponding field theories, as representations of these algebras, will be derived. Section 4 will contain various final remarks.

Subsequent papers will deal with the free field as a specific example and with the problem of gauge transformations.

2. BORCHERS' FORMULATION OF AXIOMATIC FIELD THEORY

In 1962 Borchers gave an algebraic formulation of axiomatic field theory.[10] It was further developed in Refs. 11 and 12. See also Refs. 13–15. For a single real scalar field it contains the following concepts:

a. A field algebra: It may be called the Borchers algebra, will be denoted as $\mathcal{A}$ and is defined as the tensor algebra over the space $\mathcal{S}(R_4)$ of complex valued Schwartz test functions. It is the topological direct sum $\sum_{n=0}^{\infty}\oplus\mathcal{S}(R_{4n})$; i.e., an element of $\mathcal{A}$ is a finite sequence $f_0, f_1, \ldots, f_k$, with $f_0 \in C$, $f_j \in \mathcal{S}(R_4)$, $j = 1, \ldots, k$. An *involution* is defined by $f_n(x_1, \ldots, x_n)^* = f_n(x_n, \ldots, x_1)$. In this way $\mathcal{A}$ becomes a topological $*$-algebra.

b. States: These are the positive and normalized elements ω from the dual $\mathcal{A}'$; i.e., $\omega(a^*a) \geq 0$, $\forall a \in \mathcal{A}$, $\omega(e) = 1$ (e the unit element in $\mathcal{A}$). Because $\mathcal{A}'$ is the topological product of the spaces $\mathcal{S}'(R_{4n})$, a state ω is given by an infinite sequence of tempered distributions in $4n$ variables, with $n = 0, 1, 2, \cdots$.

c. Transformations: These are described by bicontinuous $*$-automorphisms ψ of $\mathcal{A}$. Such a ψ induces by transposition an invertible bicontinuous map ψ' from $\mathcal{A}'$ onto itself, mapping the subset of states onto itself. Of special importance are the automorphisms ψ_T generated by invertible bicontinuous linear transformations T in the basic space $\mathcal{S}(R_4)$ according to

$$\psi_T : f_1 \otimes \cdots \otimes f_n \mapsto Tf_1 \otimes \cdots \otimes Tf_n \quad (f_k \in \mathcal{S}(R_4), \; k = 1, \ldots, n). \tag{1}$$

In this way one obtains, for instance, the action of the Poincaré group on $\mathcal{A}$ and $\mathcal{A}'$ by taking $(T(a, \Lambda)f)(x) = f(\Lambda^{-1}(x - a))$.

For further mathematical details, especially on the topological properties of $\mathcal{A}$ and $\mathcal{A}'$, see Refs. 11–15. For more general background material on topological vector spaces, see Refs. 16–18.

A state ω gives rise, by the GNS construction, to a representation ϕ of $\mathcal{A}$, in a space $\mathcal{H}$ which becomes by completion a Hilbert space. In $\mathcal{H}$ there is a cyclic vector Ω with the property

$$\omega(f_1 \otimes \cdots \otimes f_n) = (\Omega, \phi(f_1) \cdots \phi(f_n)\Omega), \tag{2}$$

$$f_k \in \mathcal{S}(R_4), \; k = 1, \ldots, n.$$

The operators $\phi(f)$, $f \in \mathcal{S}(R_4)$ are the field operators, Ω is the vacuum state. The GNS construction is such that if ω is invariant under a group of automorphisms, then this group will be represented by unitary operators in $\mathcal{H}$, leaving Ω invariant. In this way a Wightman field theory is given by a state ω that is Poincaré invariant and has further properties ensuring locality and a correct energy–momentum spectrum. For the description of fields different from a scalar field one substitutes for the basic space $\mathcal{S}(R_4)$ other spaces, e.g., spaces of spinor or vector valued Schwartz test functions.

Borchers' algebraic formulation of axiomatic field theory is completely equivalent to the original Hilbert space approach of Wightman. It is somewhat more abstract and the relation to conventional field theory is less obvious, but it can be argued that this is more than compensated for by a much greater transparency and internal simplicity. It also shows more explicitly a technical aspect of Wightman theory that is of importance for its application to the special case of the Maxwell field. This is the fact that there are two distinct topological structures that play a role in the theory. The first and most obvious one is the Hilbert space topology in the representation space. It comes from the positivity of the state functional and is important for the interpretation of the theory as a quantum mechanical theory. Because of the continuity of the functional there is, however, a second, much stronger locally convex topology in the representation space. It is directly connected with the topology of the algebra $\mathcal{A}$ which in turn comes from that of the test function space $\mathcal{S}(R_4)$ on which the algebraic formalism is based. It is on this topology that most of the technical developments of the theory depend. Not much of this is lost if one drops positively as a general requirement for state functionals. The essential properties of the GNS construction remain the same. For a linear functional ω which is continuous and real [i.e., $\omega(a^*) = \overline{\omega(a)}$, $\forall a \in \mathcal{A}$] the representation space $\mathcal{H}$ is the quotient space $\mathcal{A}/\mathcal{I}_\omega$, with $\mathcal{I}_\omega = \{a \in \mathcal{A} \mid \omega(ba) = 0, \; \forall b \in \mathcal{A}\}$. $\mathcal{I}_\omega$ is the closed left ideal of degeneration of ω and coincides with $\{a \in \mathcal{A} \mid \omega(a^*a) = 0\}$ when ω is positive. Because $\mathcal{A}$ is a nuclear space (in fact, a strict inductive limit of nuclear Fréchet spaces) the representation space $\mathcal{H}$ is nuclear; $\omega(a^*b)$ defines a (separately) continuous, nondegenerate sesquilinear form as inner product on $\mathcal{H}$. The field operators are continuous operators; they and other objects like the operators representing the Poincaré group when ω is invariant have quite convenient properties with respect to the topology in $\mathcal{H}$. There is, therefore, no need for an extra Hilbert space topology (in a separate publication more detailed results of this sort will be derived for certain classes of functionals). For these reasons it will be appropriate to allow in the description of the quantized Maxwell field state functionals that are continuous and real but not necessarily positive or normalized. Of course, at places in the theory where this is physically desirable, notions of positivity will reappear.

3. A WIGHTMAN-BORCHERS FORMULATION FOR THE MAXWELL FIELD

In classical electromagnetism the physical direct meaningful quantities are the electric and magnetic field strengths. These form together the antisymmetric field tensor $F_{\mu\nu}(x)$ which satisfies the two Maxwell's equations

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0, \tag{3}$$

$$\partial^\mu F_{\mu\nu} = J_\nu. \tag{4}$$

The first equation (3) is equivalent to the existence of a potential $A_\mu(x)$ from which $F_{\mu\nu}$ can be obtained according to

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu; \tag{5}$$

one may then restate the theory in terms of this auxiliary quantity, in particular the second equation (4), which connects the electromagnetic field with the current becomes

$$\partial^\mu \partial_\mu A_\nu - \partial_\nu \partial^\mu A_\mu = J_\nu. \tag{6}$$

The field $F_{\mu\nu}$ does not determine the potential A_μ uniquely, but only up to gauge transformations. For the quantized electromagnetic field a very similar line of reasoning can be followed. Before doing this in the rigorous form of the algebraic approach it may be helpful to give it first in a loose, nonrigorous version, using the language of standard Wightman theory. It then consists of the following steps:

a. Let there be given a Wightman theory for the tensor field $F_{\mu\nu}(x)$, i.e., one has a Hilbert space $\mathcal{H}^F$, a unit vector Ω^F, and field operators $F_{\mu\nu}(x)$, satisfying the first Maxwell equation (3) as an operator equation. (Further properties such as Lorentz covariance, locality, etc., are not of importance at this point.)

b. Because of a, one has a set of vacuum expectation values

$$\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}(x_1,\ldots,x_n) = (\Omega^F, F_{\mu_1\nu_1}(x_1)\cdots F_{\mu_n\nu_n}(x_n)\Omega^F),$$

$$n = 0,1,2,\cdots. \tag{7}$$

These have a positivity property because $\mathcal{H}^F$ is a Hilbert space and satisfy Eq. (3) in every variable separately.

One then proves a theorem stating that because the $\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}$ satisfy (4) in this sense, there exists a set of (generalized) functions $\omega^A_{\mu_1\cdots\mu_n}(x_1,\ldots,x_n)$, $n = 0,1,2,\cdots$, such that the given set $\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}$ can be obtained from it by repeated antisymmetric differentiation in every variable; e.g.,

$$\omega^F_{\mu_1\nu_1\mu_2\nu_2} = \partial^1_{\mu_1}\partial^2_{\mu_2}\omega^A_{\nu_1\nu_2} - \partial^1_{\mu_1}\partial^2_{\nu_2}\omega^A_{\nu_1\mu_2} - \partial^1_{\nu_1}\partial^2_{\mu_2}\omega^A_{\mu_1\nu_2} + \partial^1_{\nu_1}\partial^2_{\nu_2}\omega^A_{\mu_1\mu_2} \tag{8}$$

with $\partial^j_\mu = \partial/\partial(x_j)^\mu$. The functions $\omega^A_{\mu_1\cdots\mu_n}(x_1,\ldots,x_n)$ are not uniquely determined; this amounts to gauge freedom. It is important, furthermore, to note that they need not have the positivity property of the $\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}$; also they are not necessarily Poincaré covariant if the $\omega^F_{\mu_1\nu_1\cdots\mu_n\nu_n}$ are.

d. The reconstruction theorem in a slightly generalized form makes it possible to obtain from the $\omega^A_{\mu_1\cdots\mu_n}$ a field theory in terms of an operator field $A_\mu(x)$, such that the $\omega^A_{\mu_1\cdots\mu_n}$ become vacuum expectation values. One then has an inner product space $\mathcal{H}^A$, a unit vector Ω^A, and field operators $A_\mu(x)$.

In this way, by using essentially only the first Maxwell equation (3), one sees that a given theory for $F_{\mu\nu}(x)$, i.e., a triple $\{\mathcal{H}^F, \Omega^F, F_{\mu\nu}(x)\}$, gives rise to many different but physically equivalent theories in terms of a potential $A_\mu(x)$, i.e., triples $\{\mathcal{H}^A, \Omega^A, A_\mu(x)\}$. Any such $\{\mathcal{H}^A, \Omega^A, A_\mu(x)\}$ will be called a *gauge* for the given $\{\mathcal{H}^F, \Omega^F, F_{\mu\nu}(x)\}$. It is important to observe that $\mathcal{H}^A$ and $\mathcal{H}^F$ are by construction distinct spaces and that, therefore, at this point $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ does not make sense as an operator relation. It will be shown that there exist for every gauge a natural map from part of the space $\mathcal{H}^A$ onto $\mathcal{H}^F$. It will, however, in general not be possible to make, by means of this map, an identification between $\mathcal{H}^A$ and $\mathcal{H}^F$. This will be clear in the rigorous formulation to the Maxwell field that will now be given.

The field tensor description (F-description)

For a smeared field tensor operator one thinks of the symbolic expression $F(\psi) = \int F_{\mu\nu}(x)\psi^{\mu\nu}(x)\,d^4x$. Define, therefore, $\mathcal{A}^F$ as the Borchers algebra over the test function space $\mathcal{S}^{(2)}$ consisting of the complex, antisymmetric tensor valued functions $\psi^{\mu\nu}(x)$, with $\mathcal{S}(R_4)$ components. The basic space $\mathcal{S}^{(2)}$ is just as in the scalar field case a nuclear Fréchet space. The completed n-fold tensor product spaces $\hat{\otimes}^n \mathcal{S}^{(2)}$ are again nuclear Fréchet spaces, and can be identified with the space of tensor valued functions $\psi^{\mu_1\nu_1\cdots\mu_n\nu_n}(x_1,\ldots,x_n)$, antisymmetric in each pair of indices μ_j, ν_j separately. (Because of nuclearity the π and ϵ tensor product topologies and their completions coincide. Sec. Ref. 16, Theorems 43.9 and 50.9. $\mathcal{A}^F$ is the topological direct sum of the n-fold tensor products and is therefore a nuclear LF space. See Ref. 16, Theorem 50.8. Consider the linear subspace $\mathcal{S}_0^{(2)}$ of $\mathcal{S}^{(2)}$ consisting of all ψ that can be written as $\psi^{\mu\nu} = \partial_\rho \chi^{\mu\nu\rho}$, for some completely antisymmetric tensor function $\chi^{\mu\nu\rho}(x)$, with $\mathcal{S}(R_4)$ components. Consider in each $\hat{\otimes}^n \mathcal{S}^{(2)}$ the closed linear span of the tensor products $\mathcal{S}_0^{(2)} \otimes \mathcal{S}^{(2)} \otimes \cdots \otimes \mathcal{S}^{(2)}$, $\mathcal{S}^{(2)} \otimes \mathcal{S}_0^{(2)} \otimes \cdots \otimes \mathcal{S}^{(2)}, \ldots, \mathcal{S}^{(2)} \otimes \mathcal{S}^{(2)} \otimes \cdots \otimes \mathcal{S}_0^{(2)}$. The direct sum of these subspaces of $\hat{\otimes}^n \mathcal{S}^{(2)}$, for $n = 1,2,\cdots$, is a closed subspace in $\mathcal{A}^F$. It is in fact the closed 2-sided $*$-ideal generated by $\mathcal{S}_0^{(2)}$ in $\mathcal{A}^F$. It will be denoted as $\mathcal{J}(\mathcal{S}_0^{(2)})$. Let $\mathcal{N} = (\mathcal{J}(\mathcal{S}_0^{(2)}))^\perp$, the annihilator of $\mathcal{J}(\mathcal{S}_0^{(2)})$ in $(\mathcal{A}^F)'$, i.e., the linear subspace of $(\mathcal{A}^F)'$ defined as

$$\mathcal{N} = \{\omega \in (\mathcal{A}^F)' \mid \omega(a) = 0,\ \forall\, a \in \mathcal{J}(\mathcal{S}_0^{(2)})\}. \tag{9}$$

A field theory for the tensor field is then described by a positive, normalized functional ω^F in $\mathcal{N}$. [Note that $\omega^F \in \mathcal{N}$ is the rigorous form of the statement that the n-point functions satisfy (3) in every variable separately.] By the GNS construction this corresponds to a nuclear space $\mathcal{H}^F$ (which is at the same time a pre-Hilbert space), a cyclic unit vector Ω^F, and field operators $F(\psi)$ that satisfy $F(\psi) = 0$, $\forall\, \psi \in \mathcal{S}_0^{(2)}$. [This is, of course, the rigorous form of (3), as an operator equation, smeared with test functions.] Further requirements like Poincaré invariance of ω^F can be added in the usual manner.

The potential description (A-description)

For a smeared potential operator one thinks of the symbolic expression $A(f) = \int A_\mu(x) f^\mu(x)\,d^4x$. Define therefore $\mathcal{A}^A$ as the Borchers algebra over the space $\mathcal{S}^{(3)}$ consisting of all complex vector valued functions $f^\mu(x)$ with $\mathcal{S}(R_4)$ components. Again $\mathcal{A}^A$ is the topological direct sum of completed n-fold tensor products $\hat{\otimes}^n \mathcal{S}^{(3)}$ and as such a nuclear LF space. Let $\mathcal{S}_0^{(3)}$ be the linear subspace of $\mathcal{S}^{(3)}$, consisting of all f that can be written as $f^\mu = \partial_\nu \psi^{\mu\nu}$, for some $\psi \in \mathcal{S}^{(2)}$, and denote the closed subalgebra of $\mathcal{A}^A$ generated by $\mathcal{S}_0^{(2)}$ by $\mathcal{A}^A_{\text{ph}}$. (For reasons that will be clear further on it is appropriate to call this subalgebra the physical part of $\mathcal{A}^A$.) A field theory in terms of potentials is now described by a real functional ω^A from $(\mathcal{A}^A)'$. The functional ω^A need not to be positive

and Poincare covariance of the theory would not require it to be invariant. It will be shown that weaker conditions involving the subalgebra $\mathcal{A}^A_{\rm ph}$ characterize the functionals ω^A that describe physically acceptable theories. By the GNS construction a functional ω^A corresponds to a nuclear space H^A, provided with a (separately) continuous, nondegenerate but not necessarily positive definite inner product, a cyclic vector Ω^A, and field operators $A(f)$.

The relation between F- and A-description

There is a natural algebraic relation between the two Borchers algebras $\mathcal{A}^F$ and $\mathcal{A}^A$ and their duals. This is based on the linear map d from $\mathcal{S}^{(2)}$ into $\mathcal{S}^{(3)}$ defined by

$$d:\psi^{\mu\nu}\mapsto 2\partial_\nu\psi^{\mu\nu} \tag{10}$$

[in the language of differential forms over R_4, with $\mathcal{S}(R_4)$ components, the $\mathcal{S}^{(k)}$ are the spaces of k forms, written in dual representation; e.g., $\psi^{\mu\nu}=(2!)^{-1}\epsilon^{\mu\nu\rho\sigma}\psi_{\rho\sigma}$, $f^\mu=(3!)^{-1}\epsilon^{\mu\nu\rho\sigma}f_{\nu\rho\sigma}$. The $\mathcal{S}_0^{(k)}$ are the subspaces of exact forms; d is exterior differentiation, and the next theorem corresponds to Poincaré's lemma in this special context].

Theorem 1: The linear map $d:\mathcal{S}^{(2)}\to\mathcal{S}^{(3)}$ defined by $(d\psi)^\mu=2\partial_\nu\psi^{\mu\nu}$ is continuous. The kernel of d is $\mathcal{S}_0^{(2)}$; its image $\mathcal{S}_0^{(3)}$ consists of all f with $\partial_\mu f^\mu=0$.

Proof: Differentiation is a continuous operation in $\mathcal{S}(R_4)$ so that continuity of d is obvious. The nontrivial part of the statement about Ker$d$ is Ker$d\subset\mathcal{S}_0^{(2)}$. In terms of Fourier transforms and of components with lower indices, introduced by $\psi_{\mu\nu}=(2!)^{-1}\epsilon_{\mu\nu\rho\sigma}\psi^{\rho\sigma}$, $\chi_\mu=-(3!)^{-1}\epsilon_{\mu\nu\rho\sigma}\chi^{\nu\rho\sigma}$, the statement to be proved is: For a function $\psi_{\mu\nu}(k)$ such that $k_\mu\psi_{\nu\rho}(k)+k_\nu\psi_{\rho\mu}(k)+k_\rho\psi_{\mu\nu}(k)=0$, there exists a $\chi_\mu(k)$ such that $k_\mu\chi_\nu(k)-k_\nu\chi_\mu(k)=\psi_{\mu\nu}(k)$ [all functions have components in $\mathcal{S}(R_4)$]. For this the following division property is useful: Let $F(u_1,\ldots,u_n)$ be in $\mathcal{S}(R_n)$ with $F(0,u_2,\ldots,u_n)=0$, $\forall\, u_2,\ldots,u_n\in R_{n-1}$. Then $u_1^{-1}F(u_1,\ldots,u_n)$ is again a function in $\mathcal{S}(R_n)$. This can be verified, after Fourier transformation, as an integration property: For an $\mathcal{S}(R_n)$ function $\hat F(t_1,\ldots,t_n)$ for which $\int_{-\infty}^{+\infty}\hat F(t_1,\ldots,t_n)dt_1=0$, $t_2,\ldots,t_n\in R_{n-1}$, the function $\hat G(t_1,\ldots,t_n)=\int_{-\infty}^{t_1}\hat F(t,t_2,\ldots,t_n)dt$ is again in $\mathcal{S}(R_n)$. Suppose now $k_\mu\psi_{\nu\rho}+k_\nu\psi_{\rho\mu}+k_\rho\psi_{\mu\nu}=0$, $\forall k=(k_0,k_1,k_2,k_3)\in R_4$. For $k_0=0$ one has $k_j\psi_{0l}=k_l\psi_{0j}$, $j,l=1,2,3$. Then $\psi_{0l}(0,k_1,k_2,k_3)$ can be divided, as functions in $\mathcal{S}(R_3)$ by k_l, and give a single function in $\mathcal{S}(R_3)$; $h(k_1,k_2,k_3)=k_j^{-1}\psi_{0j}(0,k_1,k_2,k_3)$, $j=1,2,3$. Choose a function $\chi_0(k)$ in $\mathcal{S}(R_4)$ such that $\chi_0(0,k_1,k_2,k_3)=-h(k_1,k_2,k_3)$ and define functions $\chi_j(k)=k_0^{-1}(k_j\chi_0(k)+\psi_{0j}(k))$, $j=1,2,3$. These are also in $\mathcal{S}(R_4)$ and one verifies that $k_\mu\chi_\nu-k_\nu\chi_\mu=\psi_{\mu\nu}$. This proves the first statement. For the last part of the theorem one has to prove, again after Fourier transformation and using $f_{\mu\nu\rho}=-\epsilon_{\mu\nu\rho\sigma}f^\sigma$, the following statement: For a $f_{\mu\nu\rho}(k)$ such that $k_\mu f_{\nu\rho\sigma}-k_\nu f_{\rho\sigma\mu}+k_\rho f_{\sigma\mu\nu}-k_\sigma f_{\mu\nu\rho}=0$ there exists a $\psi_{\mu\nu}(k)$ such that $f_{\mu\nu\rho}=k_\mu\psi_{\nu\rho}+k_\nu\psi_{\rho\mu}+k_\rho\psi_{\mu\nu}$. [All functions in $\mathcal{S}(R_4)$.] For such a $f_{\mu\nu\rho}$ one has $k_2f_{013}=k_3f_{012}$, $\forall\, k_2,k_3\in R_2$ and $k_0=k_1=0$. Because of the division property the function $u(k_2,k_3)=k_2^{-1}f_{012}(0,0,k_2,k_3)=k_3^{-1}f_{013}(0,0,k_2,k_3)$ is well defined and in $\mathcal{S}(R_2)$. Choose $\psi_{01}(k)$ in $\mathcal{S}(R_4)$ such that $\psi_{01}(0,0,k_2,k_3)=u(k_2,k_3)$ and take $\psi_{10}=-\psi_{01}$. The function $k_2\psi_{01}(0,k_1,k_2,k_3)-f_{012}(0,k_1,k_2,k_3)$ can be divided by k_1; $u_2(k_1,k_2,k_3)=k_1^{-1}(k_2\psi^{01}(0,k_1,k_2,k_3)-f_{012}(0,k_1,k_2,k_3))$ is a well-defined function in $\mathcal{S}(R_3)$ and so is $u_3(k_1,k_2,k_3)=k_1^{-1}(k_3\psi_{01}(0,k_1,k_2,k_3)-f_{013}(0,k_1,k_2,k_3))$. Choose for $j=2,3$ functions ψ_{0j} in $\mathcal{S}(R_4)$ such that $\psi_{0j}(0,k_1,k_2,k_3)=u_j(k_1,k_2,k_3)$ and define $\psi_{j0}=-\psi_{0j}$. One verifies that $k_j\psi_{0l}-k_l\psi_{0j}+f_{0jl}=0$, for $k_0=0$, $\forall\, k_1,k_2,k_3\in R_3$ and $j,l=1,2,3$. [In particular, for $k_0=0$, $k_2\psi_{03}-k_3\psi_{02}=k_2k_1^{-1}(k_3\psi_{01}-f_{013})-k_3k_1^{-1}(k_2\psi_{01}-f_{012})=k_1^{-1}(-k_2f_{013}+k_3f_{012})=-f_{023}$).] This allows the definition of $\mathcal{S}(R_4)$ functions ψ_{jl}, $j,l=1,2,3$ as $\psi_{jl}=k_0^{-1}(k_j\psi_{0l}-k_l\psi_{0j}+f_{0jl})$. Finally one verifies that the $\psi_{\mu\nu}$ satisfy the relation $f_{\mu\nu\rho}=k_\mu\psi_{\nu\rho}+k_\nu\psi_{\rho\mu}+k_\rho\psi_{\mu\nu}$. QED

Corollary: The subspaces $\mathcal{S}_0^{(j)}$ of $\mathcal{S}^{(j)}$ are closed, for $j=2,3$.

Theorem 2: The linear map d defines by extension of $\psi_1\otimes\psi_2\otimes\cdots\otimes\psi_n\mapsto d\psi_1\otimes d\psi_2\otimes\cdots\otimes d\psi_n$ a continuous (algebraic) $*$-homomorphism Θ_d of $\mathcal{A}^F$ into $\mathcal{A}^A$, with Ker$\Theta_d=\mathcal{J}(\mathcal{S}_0^{(2)})$ and Im$\Theta_d=\mathcal{A}^A_{\rm ph}$.

Proof: The spaces $\mathcal{S}^{(2)}$ and $\mathcal{S}^{(3)}$ are Fréchet spaces, the continuous linear map d has as its image the closed subspace $\mathcal{S}_0^{(3)}$; therefore, d is a homomorphism, in the sense of linear maps between topological vector spaces. See Ref. 16, Chap. 17. The space $\mathcal{S}^{(2)}$ and $\mathcal{S}^{(3)}$ are also nuclear; the π and ϵ topologies on the n-fold tensor products coincide and give rise to the same completed tensor product. See Ref. 16, Theorem 50.1. One then combines Theorem 43.9 of Ref. 16 and the corollary of Theorem 43.7 of Ref. 15, uses the associativity of tensor products to extend all results from 2 factors to n factors, and obtains that the linear map $\otimes^n d$ from $\otimes^n\mathcal{S}^{(2)}$ into $\otimes^n\mathcal{S}^{(3)}$, determined by $\psi_1\otimes\psi_2\otimes\cdots\otimes\psi_n\mapsto d\psi_1\otimes d\psi_2\otimes\cdots\otimes d\psi_n$, can be uniquely extended to a map $\hat\otimes^n d$ from $\hat\otimes^n\mathcal{S}^{(2)}$ into $\hat\otimes^n\mathcal{S}^{(3)}$. It is continuous and, moreover, a homomorphism in the sense of topological vector spaces. Its image is $\hat\otimes^n\mathcal{S}_0^{(3)}$, considered as closed subspace of $\hat\otimes^n\mathcal{S}^{(3)}$. The kernel of $\hat\otimes^n d$ is the linear span of the closed subspaces of $\hat\otimes^n\mathcal{S}^{(2)}$: $(\hat\otimes^k\mathcal{S}^{(2)})\hat\otimes\mathcal{S}_0^{(2)}\hat\otimes(\hat\otimes^{n-k-1}\mathcal{S}^{(2)})$, $k=0,1,2,\ldots,n-1$. See Ref. 16, exercise 43.2. The algebras $\mathcal{A}^F$ and $\mathcal{A}^A$ are the topological direct sums of the $\hat\otimes^n\mathcal{S}^{(2)}$ and $\hat\otimes^n\mathcal{S}^{(3)}$. Using the properties of such direct sums and of countably strict inductive limits in general (see Ref. 16, Chaps. 13 and 50 and Ref. 18, Chap. V), one obtains that $\Theta_d=\sum_{n=0}^\infty\oplus(\hat\otimes^n d)$ is a continuous linear map from $\mathcal{A}^F$ into $\mathcal{A}^A$ with Ker$\Theta_d=\sum_{n=0}^\infty\oplus$Ker$(\hat\otimes^n d)=\mathcal{J}(\mathcal{S}_0^{(2)})$ and Im$\Theta_d=\sum_{n=0}^\infty\oplus$Im$(\hat\otimes^n d)=\mathcal{A}^A_{\rm ph}$. One easily verifies, moreover, that Θ_d is a $*$-homomorphism in the sense of maps of $*$-algebras. QED

Theorem 3: The transpose Θ_d' is a linear map from $(\mathcal{A}^A)'$ into $(\mathcal{A}^F)'$, continuous with respect to weak and strong topologies and with Ker$\Theta_d'=(\mathcal{A}^A_{\rm ph})^\perp$ and Im$\Theta_d'=\mathcal{H}$.

Proof: The continuity of Θ_d' is a well-known property of transposed maps. See, e.g., Ref. 16, corollary of Theorem 19.5. One has Ker$\Theta_d'=($Im$\Theta_d)^\perp$ and Im$\Theta_d=\mathcal{A}^A_{\rm ph}$ (Theorem 2), so Ker$\Theta_d'=(\mathcal{A}^A_{\rm ph})^\perp$. To prove that Im$\Theta_d'=\mathcal{H}$, it is sufficient to show that for the transposed maps $(\hat\otimes^n d)'$: $(\hat\otimes^n\mathcal{S}^{(3)})'\to(\hat\otimes^n\mathcal{S}^{(2)})'$, $n=0,1,2,\ldots$, one has Im$(\hat\otimes^n d)'=($Ker$\hat\otimes^n d)^\perp$ [as subspaces of $(\hat\otimes^n\mathcal{S}^{(2)})'$]. The map $\hat\otimes^N d$ is a homomorphism of topological vector spaces of the Fréchet space $\hat\otimes^n\mathcal{S}^{(2)}$ into the Fréchet

space $\hat{\otimes}^n \mathcal{S}^{(3)}$, with as image the closed subspace $\hat{\otimes}^n \mathcal{S}_0^{(3)}$. (Proof of Theorem 2.) It can therefore be written as composition of the canonical map from $\hat{\otimes}^n \mathcal{S}^{(2)}$ onto the quotient space $\hat{\otimes}^n \mathcal{S}^{(2)}/\mathrm{Ker}\hat{\otimes}^n d$ and an isomorphism (of topological vector spaces) of $\hat{\otimes}^n \mathcal{S}^{(2)}/\mathrm{Ker}\hat{\otimes}^n d$ onto $\hat{\otimes}^n \mathcal{S}_0^{(3)}$. The transposed map associated with this isomorphism is a bijective linear map from $(\hat{\otimes}^n \mathcal{S}_0^{(3)})'$ onto $(\hat{\otimes}^n \mathcal{S}^{(2)}/\mathrm{Ker}\hat{\otimes}^n d)'$. Composing this map with the canonical map from $(\hat{\otimes}^n \mathcal{S}^{(3)})'/(\mathrm{Im}\hat{\otimes}^n d)^\perp$ onto $(\mathrm{Im}\hat{\otimes}^n d)' = (\hat{\otimes}^n \mathcal{S}_0^{(3)})'$, and with the canonical map from $(\hat{\otimes}^n \mathcal{S}^{(2)}/\mathrm{Ker}\hat{\otimes}^n d)'$ onto $(\mathrm{Ker}\hat{\otimes}^n d)^\perp$, one obtains a linear bijection from $(\hat{\otimes}^n \mathcal{S}^{(3)})'/(\mathrm{Im}\hat{\otimes}^n d)^\perp$ onto $(\mathrm{Ker}\hat{\otimes}^n d)^\perp$. This in turn gives a linear map from $(\hat{\otimes}^n \mathcal{S}^{(3)})'$ into $(\hat{\otimes}^n \mathcal{S}^{(2)})'$ with $(\mathrm{Im}\hat{\otimes}^n d)^\perp$ as kernel and $(\mathrm{Ker}\hat{\otimes}^n d)^\perp$ as image, which can be identified with the map $(\hat{\otimes}^n d)'$. QED

The map Θ_d' is, of course, nothing but the repeated antisymmetric differentiation of n-point functions $\omega^A_{\mu_1\cdots\mu_n}(x_1,\ldots,x_n)$, employed under point c in the beginning of this section, but now properly formulated. Theorem 3 is the basic theorem for the present formalism. It says roughly that every set of n-point functions in the A-description goes by antisymmetric differentiation over to a set of n-point functions in the F-description, satisfying the first Maxwell equation in every variable separately and, moreover, that each such F-description set comes from some A-description set.

It should be noted that if a state ω^F in $\mathcal{M}$ is normalized then every ω^A such that $\omega^F = \Theta_d'\omega^A$ is also normalized; if ω^F is real, ω^A need not be real; however, an equivalent real state ω_1^A can immediately be obtained as $\omega_1^A(a) = \frac{1}{2}[\omega^A(a) + \overline{\omega^A(a^*)}]$, $\forall a \in \mathcal{A}^A$. The situation with respect to positivity and Lorentz invariance is less simple. At this point it is not clear whether for an arbitrary positive, respectively Lorentz invariant state ω^F in $\mathcal{M}$ the inverse image $(\Theta_d')^{-1}\omega^F$ contains a positive, respectively Lorentz invariant state ω^A. For the free field one can derive explicitly the different gauges, as will be done in a subsequent paper. In that case there exist gauges corresponding with positive states on $\mathcal{A}^A$ and also gauges corresponding with Lorentz invariant states, but there is no gauge having both properties. It is expected that this is typical for the general case.

Relation between representations in F- and A-description

Let ω^F be a state in $\mathcal{A}^F$, positive, normalized and in $\mathcal{M}$. Let $\{\mathcal{H}^F, \Omega^F, \Pi^F\}$ the corresponding field theory obtained as the GNS representation associated with ω^F. Suppose ω^A to be a real state on $\mathcal{A}^A$ in $(\Theta_d')^{-1}\omega^F$ and $\{\mathcal{H}^A, \Omega^A, \Pi^A\}$ the corresponding field theory. In the terminology adopted in this paper $\{\mathcal{H}^A, \Omega^A, \Pi^A\}$ is a gauge for $\{\mathcal{H}^F, \Omega^F, \Pi^F\}$. [$\Pi^F(a)$ is the operator in $\mathcal{H}^F$ representing a general element a in $\mathcal{A}^F$. The field operator is then $F(\psi) = \Pi^F(\psi)$, $\forall \psi \in \mathcal{S}^{(2)}$; in the same way $A(f) = \Pi^A(f)$, $\forall f \in \mathcal{S}^{(3)}$. It should, furthermore, be noted that the completion of the pre-Hilbert space $\mathcal{H}^A$ will not be considered in this paper.] Define $\mathcal{H}^A_{\mathrm{ph}}$ as the subspace of $\mathcal{H}^A$ consisting of all vectors $\Pi^A(a)\Omega^A$, $a \in \mathcal{A}^A_{\mathrm{ph}}$. The spaces $\mathcal{H}^A$ and $\mathcal{H}^F$ are by construction distinct; there is, however, a natural map from the subspace $\mathcal{H}^A_{\mathrm{ph}}$ onto $\mathcal{H}^F$.

Theorem 4: The relation $\Pi^A(\Theta_d a)\Omega^A \to \Pi^F(a)\Omega^F$, $\forall a \in \mathcal{A}^F$, defines a linear isometric map W from $\mathcal{H}^A_{\mathrm{ph}}$ onto $\mathcal{H}^F$ having the following properties:

a. $W\Omega^A = \Omega^F$.

b. $W\Pi^A(\Theta_d a) = \Pi^F(a)W$, $\forall a \in \mathcal{A}^F$ [in particular $WA(d\psi) = F(\psi)W$, $\forall \psi \in \mathcal{S}^{(2)}$].

c. The action of W on state vectors is consistent with the action of Θ_d' on state functionals in the following sense: A vector Ω_1^A in $\mathcal{H}^A_{\mathrm{ph}}$ and its image $\Omega_1^F = W\Omega_1^A$ give rise to functionals ω_1^A and ω_1^F on $\mathcal{A}^A$ and $\mathcal{A}^F$ by the formula $\omega_1^{F,A}(a) = (\Omega_1^{F,A}, \Pi^{F,A}(a)\Omega_1^{F,A})$, $\forall a \in \mathcal{A}^{F,A}$. Then one has $\omega_1^F = \Theta_d'\omega_1^A$.

Proof:

$$(\Pi^F(a)\Omega^F, \Pi^F(b)\Omega^F) = \omega^F(a^*b) = (\Theta_d'\omega^A)(a^*b) = \omega^A((\Theta_d a^*)(\Theta_d b)) = (\Pi^A(\Theta_d a)\Omega^A, \Pi^A(\Theta_d b)\Omega^A), \quad \forall a, b \in \mathcal{A}^F.$$

So $\Pi^A(\Theta_d a)\Omega^A = 0$ implies $\Pi^F(a)\Omega^F = 0$, and therefore W is well defined. The isometry $(W\Pi^A(\Theta_d a)\Omega^A, W\Pi^A(\Theta_d b)\Omega^A) = (\Pi^A(\Theta_d a)\Omega^A, \Pi^A(\Theta_d b)\Omega^A)$ follows immediately. The proof of a and b is obvious. For c one has: $\Omega_1^A \in \mathcal{A}^A_{\mathrm{ph}}$, then $\exists b \in \mathcal{A}^A$ such that $\Omega_1^A = \Pi^A(\Theta_d b)\Omega^A$. Then

$$\omega_1^A(a) = (\Pi^A(\Theta_d b)\Omega^A, \Pi^A(a)\Pi^A(\Theta_d b)\Omega^A) = \omega^A((\Theta_d b)^* a(\Theta_d b)), \quad \forall a \in \mathcal{A}^A.$$

Also

$$\omega_1^F(c) = (\Omega_1^F, \Pi^F(c)\Omega_1^F) = (\Pi^F(b)\Omega^F, \Pi^F(c)\Pi^F(b)\Omega^F) = \omega^F(b^*cb) = (\Theta_d'\omega^A)(b^*cb) = \omega^A((\Theta_d b)^*(\Theta_d c)(\Theta_d b)) = \omega_1^A(\Theta_d c) = (\Theta_d'\omega_1^A)(c), \quad \forall c \in \mathcal{A}^F.$$ QED

Corollary: The inner product in $\mathcal{H}^A$ is definite positive on the subspace $\mathcal{H}^A_{\mathrm{ph}}$.

Because of this and the other properties of the correspondence between $\mathcal{H}^A_{\mathrm{ph}}$ and $\mathcal{H}^F$ one may call $\mathcal{H}^A_{\mathrm{ph}}$ the subspace of physical photon states. The null space of W may be denoted as $\mathcal{H}^A_0$, it consists of all null-length vectors in $\mathcal{H}^A_{\mathrm{ph}}$. Because of the Schwartz inequality, valid on $\mathcal{H}^A_{\mathrm{ph}}$ one has $\Psi \in \mathcal{H}^A_0 \Leftrightarrow (\Psi, \Psi^1) = 0$, $\forall \Psi^1 \in \mathcal{H}^A_{\mathrm{ph}}$. The space of null-photon states $\mathcal{H}^A_0$ defines an equivalence relation for physical photon states: for $\Psi_1, \Psi_2 \in \mathcal{H}^A_{\mathrm{ph}}$, $\Psi_1 \cong \Psi_2 \Leftrightarrow \Psi_1 - \Psi_2 \in \mathcal{H}^A_0 \Leftrightarrow W\Psi_1 = W\Psi_2$. The subspaces $\mathcal{H}^A_0$, $\mathcal{H}^A_{\mathrm{ph}}$ and their properties are, of course, well known from the Gupta–Bleuler formalism for the free field. In the general case such a structure is postulated by Strocchi and Wightman. It must, however, be emphasized that it appears in this formalism in quite a general way as a rigorous consequence of the fact that the electromagnetic field tensor satisfies the first Maxwell equation as an operator equation.

There is also a natural correspondence between those operators in $\mathcal{H}^A$ that represent the physical algebra $\mathcal{A}^A_{\mathrm{ph}}$ and the operators in $\mathcal{H}^F$ that represent $\mathcal{A}^F$. The relation $\Pi^A(\Theta_d a) \to \Pi^F(a)$ defines in fact a $*$-homomorphism from $\Pi^A(\mathcal{A}^A_{\mathrm{ph}})$ onto $\Pi^F(\mathcal{A}^F)$ [because: $\Pi^A(\Theta_d a) = 0 \Rightarrow \omega^A(\Theta_d(bac)) = 0$, $\forall b, c \in \mathcal{A}^F \Rightarrow \omega^F(bac) = 0$, $\forall b, c \in \mathcal{A}^F \Rightarrow \Pi^F(a) = 0$.] The operators in $\Pi^A(\mathcal{A}^A_{\mathrm{ph}})$ may be called physical operators. They correspond to what are sometimes called gauge invariant operators. One verifies easily that they leave $\mathcal{H}^A_{\mathrm{ph}}$ and $\mathcal{H}^A_0$ invariant and have the property

$$(\Psi_1, \Pi^A(\Theta_d a)\Psi_2) = (\Psi_1', \Pi^A(\Theta_d a)\Psi_2'), \quad \forall \Psi_j, \Psi_j' \in \mathcal{H}^A_{\mathrm{ph}},\ \Psi_j \cong \Psi_j',\ j = 1, 2,\ a \in \mathcal{A}^F. \tag{11}$$

An important special case occurs when it is possible to choose a gauge such that $\mathcal{H}^A_{ph} = \mathcal{H}^A$, in this case $\mathcal{H}^A_0 = 0$, W becomes a unitary transformation from $\mathcal{H}^A$ onto $\mathcal{H}^F$, which can be used to identify the two representation spaces. This leads to a much simpler situation with a single prehilbert space as representation space for the operators $A(f)$ and $F(\psi)$. They are then connected by the relation $F(\psi) = A(d\psi)$, $\forall \psi \in \mathcal{S}^{(2)}$, i.e., the relation $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is valid as an operator relation. An example of such a situation is the Coulomb gauge for the free field. The price to be paid for the simplification of the description is well known in this case. Manifest Lorentz covariance is lost, in the sense that the potential $A_\mu(x)$ no longer transforms as a vector under Lorentz transformations; i.e., $U(\Lambda)A_\mu(x)U(\Lambda)^{-1} \neq (\Lambda^{-1})_\mu{}^{\mu'} A_{\mu'}(\Lambda x)$.

From the results in this section it is clear how axiomatic field theory in terms of the field $A_\mu(x)$ should be formulated and how the usual axioms of the Wightman–Borchers formulation should be modified:

a. The vacuum state should be a continuous, real, normalized linear functional ω^A on the Borchers algebra $\mathcal{A}^A$. It has to satisfy a restricted positivity condition; it should be positive on the subalgebra $\mathcal{A}^A_{ph}$. (From this the properties of the representation space, the special subspaces, the equivalence relation between physical vectors, etc., will follow.)

b. Lorentz invariance, locality and spectral properties should be required also only with respect to the subalgebra $\mathcal{A}^A_{ph}$.

The vacuum state will not be unique; two such states will be physically equivalent when having the same restriction to $\mathcal{A}^A_{ph}$. They correspond then to different gauges of a single physical theory. One may make a choice from the available gauge by strengthening the requirements for ω^A, e.g., into full positivity on $\mathcal{A}^A$ or full Lorentz invariance on $\mathcal{A}^A$: The crucial point is, however, that in general these stronger requirements cannot be fulfilled simultaneously.

4. MISCELLANEOUS FINAL REMARKS

The formalism given in this paper can be applied to specific cases. It provides in particular a simple point of view for the discussion of the free field. This has many apparently very different realizations. Although these gauges and their properties are, in a more or less rigorous form, well known (the paper of Strocchi and Wightman contains an admirable review), the situation as a whole can be much better understood in terms of the algebraic formalism. The different gauges appear there in a natural manner, and their properties can be derived in a systematic and rigorous way. A detailed discussion of this kind will be given in a subsequent paper.

The situation of a classical Maxwell field can be easily accommodated in the algebraic scheme as a special, rather trivial case. Let $f_{\mu\nu}(x)$ be a real-valued classical solution of Eq. (3) and $\alpha_\mu(x)$ a potential corresponding with $f_{\mu\nu}(x)$ according to relation (5). (The functions are supposed to be C^∞ and to have a behavior at infinity such that they define tempered distributions in R_4.) One obtains real, normalized functionals $\omega^{F,A}$ on $\mathcal{H}^{F,A}$ by defining

$$\begin{aligned}
&\omega^{F,A}(e) = 1,\\
&\omega^F(\psi_1 \cdots \psi_n) = \int f_{\mu_1\nu_1}(x_1)\cdots f_{\mu_n\nu_n}(x_n)\\
&\qquad\qquad \times \psi^{\mu\ \nu}(x_1)\cdots\psi^{\mu_n\nu_n}(x_n)\,d^4x_1\cdots d^4x_n,\\
&\omega^A(f_1\cdots f_n) = \int \alpha_{\mu_1}(x_1)\cdots\alpha_{\mu_n}(x_n)\\
&\qquad\qquad \times f^{\mu_1}(x_1)\cdots F^{\mu n}(x_n)\,d^4x_1\cdots d^4x_n,
\end{aligned} \tag{12}$$

$$\forall \psi_1,\ldots,\psi_n \in \mathcal{S}^{(2)},\ \forall f_1,\ldots,f_n \in \mathcal{S}^{(3)},\ n = 1,2,3,\cdots.$$

The field operators in the representation spaces $\mathcal{H}^{F,A}$ associated with $\omega^{F,A}$ commute. They act on the vacuum states $\Omega^{F,A}$ by scalar multiplication; e.g., $A(f)\Omega^A = [\int \alpha_\mu(x) f^\mu(x)\,d^4x]\Omega^A$. The spaces $\mathcal{H}^{F,A}$ are therefore one-dimensional and can be identified by means of the map W; finally (5) holds as an operator relation. Less trivial is the description of stochastic classical Maxwell fields. These are given by functionals $\omega^{F,A}$ for which the n-point functions are symmetric. The field operators will again commute; however, $\mathcal{H}^{F,A}$ will be nontrivial. The $\omega^{F,A}$ can be considered to give infinite dimensional moment problems. Under appropriate conditions the GNS construction will provide solutions in the form of infinite systems of classical random variables on probability spaces.

The main theme of this paper is the investigation of the consequences of Eq. (3) as an operator equation for $F_{\mu\nu}$, in particular, the existence of a field A_μ such that (5) holds in some sense. It is obvious that the same methods could be used to study a problem discussed by Pohlmeyer[19] from a different angle: Find, for a given vector field operator $\phi_\mu(x)$ with $\partial_\mu\phi_\nu - \partial_\nu\phi_\mu = 0$, a scalar field ϕ such that $\phi_\mu = \partial_\mu\phi$. (In fact this problem will play a role in the discussion of gauge transformations to be given in a subsequent paper.)

The formalism, as developed so far, contains only the quantized Maxwell field. This means that it describes rigorously situations in which the Maxwell field is free, interacts with given external currents (classical or quantized) or has some form of self-interaction. The structure of the formalism depends, however, only on the first Maxwell equation (3); the second equation (4), characterizing the interaction, is not used. One may therefore expect that a very similar structure will emerge in an extended formalism in which the Maxwell field appears coupled to a quantized Dirac spinor field or a charged scalar field. The development of such an extension is the main task in this or any other rigorous approach to the formulation of quantum electrodynamics in axiomatic field theory.

ACKNOWLEDGMENTS

The author is grateful to Professor E.M. de Jager for valuable conversations on topics from distribution theory and to Professor N.H. Kuiper for the hospitality extended to him at the Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, where part of this work was done. The investigation is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (F.O.M.)", which is financially supported

by the "Organisatie voor Zuiver Wetenschappelijk Onderzoek (Z.W.O.)."

[1]R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965).
[2]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[3]N. N. Bogolubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, New York, 1975).
[4]A. S. Wightman and L. Gårding, Arkiv Fysik 28, 129 (1964).
[5]D. Kastler, *Introduction à l'électrodynamique quantique* (Dunod, Paris, 1961).
[6]P. K. Rasevskiĭ, Amer. Math. Soc. Transl. Ser. 2, 23, 231 (1963) [Usp. Mat. Nauk (N.S.) 13, 3 (1958)].
[7]A. L. Carey, J. M. Gaffney, and C. A. Hurst, preprint (Adelaide University, 1976).
[8]F. Strocchi and A. S. Wightman, J. Math. Phys. 15, 2198 (1974).
[9]G. Rideau, Lett. Math. Phys. 1, 17 (1975).
[10]H. J. Borchers, Nuovo Cimento 24, 214 (1962).
[11]H. J. Borchers, lecture, in *Europhysics Conference on Statistical Mechanics and Field Theory, Haifa* 1971 (Israel U. P., Jerusalem, 1972).
[12]H. J. Borchers, lecture, in *International Symposium on Mathematical Problems in Theoretical Physics, Kyoto* 1975 (Springer-Verlag, Berlin, 1975).
[13]K. Maurin, Bull. Acad. Pol. Sci. XI, 115 (1963).
[14]K. Maurin, Bull. Acad. Pol. Sci. XI, 121 (1963).
[15]W. Wyss, in *Lectures in Theoretical Physics, Vol. XI-D* (*Boulder* 1968) (Gordon and Breach, New York, 1969).
[16]F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).
[17]J. L. Kelley and I. Namioka, *Linear Topological Spaces* (Van Nostrand, Princeton, N.J., 1963).
[18]A. P. Robertson, W. J. Robertson, *Topological Vector Spaces* (Cambridge U. P., Cambridge, 1964).
[19]K. Pohlmeyer, Commun. Math. Phys. 25, 73 (1972).

$D(1,0)$ Killing structures and $\mathcal{E}$ potentials

S. Hacyan

Instituto de Astronomía, Universidad Nacional Autónoma de México, A.P. 70-264, México 20, D. F., Mexico

J. F. Plebański*

Departamento de Física, Centro de Investigación y de Estudios Avanzados, Instituto Politécnico Nacional, A.P. 14-740, México 14, D. F., Mexico

(Received 4 February 1977)

Type D metrics in vacuum or with a special electromagnetic field are studied using the fact that they admit a $D(1,0)$ Killing spinor. It is shown that the $\mathcal{E}$ and Φ potentials of Ernst can be obtained in a straightforward way for these types of metrics.

1. INTRODUCTION

Type D metrics have some interesting properties which distinguish them among other types of metrics. Of great interest is a result of Carter[1] regarding separability of variables of the Hamilton–Jacobi equations, and the relation of this property with the existence of a Killing tensor. As it was later shown by Walker and Penrose,[2] Carter's result is related to the fact that vacuum D metrics, and the Kerr–Newman metric as well, always admit a Killing spinor, from which a Killing tensor can be constructed. This result was later generalized by Hughston *et al.*[3]

Some properties of Killing spinors of a general type were studied by the present authors,[4] the results being valid either in a real or a complex Riemannian space. Our interest in complex spaces was originated by recent developments in the theory of "heavens[5–7]" and "hyperheavens.[8]" Our study[4] of spinorial Killing structures was a starting point for the recent progress of Ernst and one of us[9] in the theory of complex $\mathcal{E}$ potentials,[10,11] a technique which has been the basis for obtaining some new solutions.[12,13] The main idea in Ref. 9 was to integrate direct square products of members of a $D(\frac{1}{2},\frac{1}{2})$ Killing structure,[4] thereby obtaining the potentials. Whether similar results may be obtained from other types of Killing structures is an interesting problem; the objective of this paper is to test this idea for the case of a $D(1,0)$ Killing structure in vacuum (with possible cosmological constant $\lambda \neq 0$), and with an electromagnetic field aligned with respect to the curvature.

The basic idea is the following. Let h_{AB} be a $D(1,0)$ Killing spinor, that is, it satisfies the equation

$$\nabla^{\dot{A}}_{(A} h_{BC)} = 0 \tag{1.1}$$

(the formalism and notation of Ref. 7 will be used throughout this paper). Then, in general, one has[4,14]

$$\nabla^{\dot{A}}_{A} h_{BC} = \epsilon_{A(B} K^{\dot{A}}_{C)}, \tag{1.2a}$$

$$\nabla^{\dot{A}}_{A} K^{\dot{B}}_{B} = 4C_{(A}{}^{N\dot{A}\dot{B}} h_{B)N} + \epsilon_{AB}\bar{l}^{\dot{A}\dot{B}} + \epsilon^{\dot{A}\dot{B}} l_{AB}, \tag{1.2b}$$

$$\nabla^{\dot{A}}_{A} \bar{l}^{\dot{B}\dot{C}} = 2C^{\dot{A}\dot{B}\dot{C}}{}_{N} K_{A}{}^{\dot{N}} + 6C_{A}{}^{R(\dot{A}\dot{B}} K^{\dot{C})}_{R} - 4h_{(A}{}^{R}\nabla^{N(\dot{A}} C_{N)R}{}^{\dot{B}\dot{C})} - \tfrac{2}{3}\{h^{NS}\nabla^{P(\dot{C}} C_{NSAP} - \tfrac{1}{6}h^{S}_{A}\nabla_{S}{}^{(\dot{C}} R + \tfrac{1}{4}K_{A}{}^{(\dot{C}}\}\epsilon^{\dot{B})\dot{A}}, \tag{1.2c}$$

where

$$l_{AB} \equiv C_{AB}{}^{PQ} h_{PQ} + (R/6) h_{AB}, \tag{1.3}$$

C_{ABCD} and $C_{AB\dot{A}\dot{B}}$ are the spinorial images of the Weyl tensor and the traceless Ricci tensor, respectively, and R is the Ricci scalar. In the case of a real V_4 with signature $(+++-)$, $\bar{l}_{\dot{A}\dot{B}}$ is the complex conjugate of l_{AB} if the vector $K_{A\dot{A}}$ is real. In the general case of a complex V_4, the Killing structure is closed with respect to differentiation through Eqs. (1.2) and (1.3), *without* postulating the existence of a $D(0,1)$ Killing spinor $h_{\dot{A}\dot{B}}$—which, of course, is present in the real case.

The integrability condition of (1.2a) gives[4]

$$C^{N}{}_{(ABC} h_{D)N} = 0, \tag{1.4}$$

which is a very strong constraint: The metric must be of type D—if $\det(h_{AB}) \neq 0$— or type N—if $\det(h_{AB}) = 0$— (see Sommers[15]). In the second case, it further follows that the metric must be that of a plane wave.[15] In the following we shall only consider the case when the metric is of type D. A direct consequence of (1.4) is that the conformal curvature C_{ABCD} must be proportional to $h_{(AB} h_{CD)}$.

2. $\mathcal{E}$ POTENTIALS IN VACUUM

In D-vacuum $C_{AB\dot{A}\dot{B}} = 0$ and $R = -4\lambda$ (cosmological constant), and Walker and Penrose[2] have shown that a nontrivial $D(1,0)$ Killing spinor always exists. Furthermore, from such a Killing spinor, h_{AB}, a vector $K_{A\dot{A}}$ can be constructed through our Eq. (1.2a) and, as Sommers[15] has shown, $K_{A\dot{A}}$ is a Killing vector. This can be seen immediately by specializing Eq. (1.2b) to the vacuum case

$$\nabla^{\dot{A}}_{A} K^{\dot{B}}_{B} = \epsilon_{AB}\bar{l}^{\dot{A}\dot{B}} + \epsilon^{\dot{A}\dot{B}} l_{AB}. \tag{2.1}$$

This is precisely the definition of a Killing vector in spinorial language (see Ref. 4). Furthermore, if we define

$$h_{PQ} h^{PQ} \equiv -2\psi^{-2}, \tag{2.2}$$

then[2]

$$l_{AB} = -\tfrac{2}{3}(2\psi^{3} + \lambda) h_{AB}, \tag{2.3}$$

$$C_{ABCD} = \psi^{5} h_{(AB} h_{CD)}, \tag{2.4}$$

this last relation being a consequence of Bianchi iden-

tities. Notice that, in general, $K_{A\dot{A}}$ is a complex vector.

By direct differentiation of Eq. (2.2), and use of Eqs. (1.2a), (2.1), and (2.3), one obtains

$$h_{AN}K^{N\dot{A}} = \nabla_A^{\dot{A}}\psi^{-2} \tag{2.5}$$

$$\Rightarrow K_A^{\dot{A}} = -2\psi^{-1}h_{AN}\nabla^{N\dot{A}}\psi, \tag{2.6}$$

and this last equation gives $K_A^{\dot{A}}$ in an explicit form assuming that h_{AB} is known. Following Ernst and Plebański,[9] we now define a vector

$$G_A^{\dot{A}} := 2K^{N\dot{A}}l_{NA} \tag{2.7}$$

which turns out to be the gradient of an $\mathcal{E}$ potential[9]:

$$G_A^{\dot{A}} = \nabla_A^{\dot{A}}\mathcal{E}. \tag{2.8}$$

Using (2.1), (2.3), and (2.6) it easily follows that

$$\mathcal{E} = \tfrac{16}{3}\psi - (4\lambda/3)\psi^{-2} + \text{const.} \tag{2.9}$$

One can substitute this into Ernst's equation[9] or, alternatively, take directly the derivative of Eq. (2.5) and, using formulas (1.2), obtain the master equation

$$(\nabla_A^{\dot{A}}\nabla_{\dot{A}}^{A} - \tfrac{4}{3}\lambda)\psi = \tfrac{8}{3}\psi^4 \tag{2.10}$$

which is a wavelike nonlinear equation. The function ψ has a direct geometrical meaning since, according to Eq. (2.4), the curvature invariant $C^{(3)}$ is proportional to ψ^3. For instance, for the seven parameter metric of Plebański and Demiański (see, e.g., Plebański[16]) specialized to the vacuum case we find

$$\psi \propto \frac{p+q}{1-ipq} \tag{2.11}$$

in the notation of Ref. 16 (the proportionality constant is irrelevant since any Killing spinor is defined with precision up to a multiplicative constant).

We note that the above analysis, which applies to real D metrics, is also valid in a complex V_4. The difference is that condition (1.4) is less restrictive: The space–time must be of type $D\otimes$anything.

3. $\mathcal{E}$ AND Φ POTENTIALS IN ELECTROVACUUM

Let us now consider an electrovac structure such that the two (different) eigenvectors of the electromagnetic field are parallel to the two Debever–Penrose vectors of the type D metric. Let the Maxwell field be

$$f_{AB} = \psi^3 h_{AB}, \tag{3.1}$$

while the curvature is given by

$$C_{ABCD} = \varphi h_{(AB}h_{CD)}, \tag{3.2a}$$

$$C_{AB\dot{A}\dot{B}} = -8f_{AB}f_{\dot{A}\dot{B}}, \tag{3.2b}$$

and we postulate that h_{AB} is the Killing spinor introduced in Sec. 1. Then, the Maxwell equations without currents,

$$\nabla^{N\dot{A}}f_{NA} = 0 = \nabla_{A\dot{N}}f^{N\dot{A}}, \tag{3.3}$$

imply that

$$\nabla_A^{\dot{A}}f_{BC} = -\tfrac{3}{2}\psi^5 h_{(AB}h_{CD)}K^{D\dot{A}} \tag{3.4}$$

and it follows from the Bianchi identities that

$$\nabla_A^{\dot{A}}(\varphi\psi^{-5}) + 12f_{AN}K^{A\dot{N}} = 0. \tag{3.5}$$

A basic result of Hughston *et al.*[3] is that nontrivial Einstein–Maxwell fields such as the one described above do exist (the Kerr–Newman metric is an example).

Now, the vector $K_{A\dot{A}}$ defined by Eq. (1.2a) is also a Killing vector in this case; this is due to the fact that the first term on the right-hand side of Eq. (1.2b) vanishes for a Maxwell field such as the one described above. The analog of Eq. (2.3) in this case is

$$l_{AB} = (-\tfrac{4}{3}\varphi\psi^{-2} - \tfrac{2}{3}\lambda)h_{AB}, \tag{3.6}$$

and Eqs. (2.2) and (2.4) are still valid.

Following Ernst and Plebański,[9] we again define a vector $G_A^{\dot{A}}$ as in Eq. (2.7), and two additional vectors, $F_A^{\dot{A}}$ and $\bar{F}_A^{\dot{A}}$, which are the gradients of two potentials, Φ and $\bar{\Phi}$, namely

$$-4K^{N\dot{A}}f_{NA} =: F_A^{\dot{A}} = \nabla_A^{\dot{A}}\Phi, \tag{3.7a}$$

$$4K_{A\dot{N}}f^{\dot{N}\dot{A}} =: \bar{F}_A{}^{\dot{A}} = \nabla_A^{\dot{A}}\bar{\Phi}. \tag{3.7b}$$

From Eqs. (2.2), (3.1), (2.6), and (3.5) it follows immediately that, in our case,

$$F_A{}^{\dot{A}} = \nabla_A^{\dot{A}}(8\psi), \tag{3.8a}$$

$$\bar{F}_A{}^{\dot{A}} = \nabla_A^{\dot{A}}(\tfrac{1}{3}\varphi\psi^{-5}), \tag{3.8b}$$

which identifies the Φ and $\bar{\Phi}$ potentials. Now, it can be shown that[9]

$$G_A^{\dot{A}} = \nabla_A^{\dot{A}}\mathcal{E} + 2\bar{\Phi}\nabla_A^{\dot{A}}\Phi, \tag{3.9}$$

which is the generalization of Eq. (2.8). Using the values of Φ and $\bar{\Phi}$ as implied by Eqs. (3.8), it follows that

$$\mathcal{E} = -\tfrac{8}{3}\lambda\psi^{-2} + (\text{const})\psi + \text{const.} \tag{3.10}$$

Furthermore, f defined by

$$f := -\tfrac{1}{2}K^N{}_{\dot{N}}K^{\dot{N}}{}_N \tag{3.11}$$

is, according to Eq. (2.6),

$$f = 2\psi^{-4}(\nabla_N^{\dot{N}}\psi)(\nabla^N{}_{\dot{N}}\psi), \tag{3.12}$$

and therefore Eqs. (3.22) and (3.23) of Ref. 9 reduce to the single equation

$$(\nabla^N{}_{\dot{N}}\nabla^{\dot{N}}{}_N - \tfrac{4}{3}\lambda)\psi = \tfrac{8}{3}\varphi\psi^{-1}. \tag{3.13}$$

According to equation (3.10), when $\lambda = 0$, the $\mathcal{E}$ potential can be made zero. What happens, actually, is that the nontrivial potential in this case is Φ, which was null in the vacuum case. The relevant function is ψ, which satisfies the wavelike equation (3.13).

For the seven parameter metric,[16] which is of the type studied in this section, the ψ and φ functions are

$$\psi \propto \frac{p+q}{1-ipq}, \tag{3.14a}$$

$$\varphi \propto \left(\frac{p+q}{1-ipq}\right)^5\left[(m_0+in_0)+(e_0^2+g_0^2)\left(\frac{p+q}{1+ipq}\right)\right], \quad (3.14b)$$

again in the notation of Ref. 16.

The above analysis is also valid in a complex V_4, where $f_{\dot{A}\dot{B}}^*$ is not necessarily the complex conjugate of f_{AB}; notice that in no moment have we assumed anything about $f_{\dot{A}\dot{B}}^*$, except that it satisfies Maxwell equations.

ACKNOWLEDGMENTS

The interest of Professor F.J. Ernst in this work is gratefully acknowledged.

*On leave of absence from the University of Warsaw, Warsaw, Poland.

[1]B. Carter, Phys. Rev. 174, 1559 (1968); also, see I. Hauser and R.J. Malhiot, J. Math. Phys. 15, 816 (1974); 17, 1306 (1976) for a detailed study of rank two Killing tensors.
[2]M. Walker and R. Penrose, Commun. Math. Phys. 18, 265 (1970).
[3]L.P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. 27, 303 (1972).
[4]S. Hacyan and J.F. Plebański, J. Math. Phys. 17, 2203 (1976).
[5]E.T. Newman in Tel Aviv GR7 Lecture (1974).
[6]R. Penrose, "The Nonlinear Graviton," First Award Gravity Research Foundation (1975).
[7]J.F. Plebański, J. Math. Phys. 16, 2395 (1975).
[8]J.F. Plebański and I. Robinson, Phys. Rev. Lett. 37, 493 (1976); A. García, J.F. Plebański, and I. Robinson, "Null Strings and Complex Einstein-Maxwell Fields with λ," Nuovo Cimento (to be published); J.D. Finley, III and J.F. Plebański, J. Math. Phys. 17, 2207 (1976).
[9]F.J. Ernst and J.F. Plebański, "Killing Structures and Complex $\mathcal{E}$-Potentials," Ann. Phys. (to be published).
[10]F.J. Ernst, Phys. Rev. 168, 1415 (1968).
[11]F.J. Ernst, J. Math. Phys. 15, 1409 (1974).
[12]W. Kinnersley, J. Math. Phys. 14, 651 (1973); also, see F.P. Esposito and L. Witten, preprint, University of Cincinnati.
[13]A. Tomimatsu and H. Sato, Progr. Theor. Phys. 50, 95 (1973); also, see M. Yamazaki, preprint, Kanasawa University.
[14]Equation (3.10b) of Ref. 4, which corresponds to Eq. (1.2b) of this paper, is valid for Hermitian $K_{A\dot{A}}$ only (i.e., a real vector K_α); this was not mentioned explicitly in Ref. 4.
[15]P. Sommers, thesis, The University of Texas at Austin (1973).
[16]J.F. Plebański, Ann. New York Acad. Sci. 262, 246 (1975).

ERRATA

Erratum: Higher indices of group representations [J. Math. Phys. 17, 1972 (1976)]

J. Patera, R. T. Sharp, and P. Winternitz

Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada

(Received 29 March 1977)

In formula (51) please insert $]^{-1}$ after the factor $(2n+3)$; in (52) please replace $|\epsilon_i^{j-1}|$ by $|\epsilon_i^{n-j}|$. Insert [after N_λ in (62) and] after the entire formula; the denominator of the last term in (62) should be $n(2n-1)^2$. The factor $1/n$ in the second line of (72) should be $1/n^2$.

All numerical calculations in the article were performed using correct formulas and are valid. We thank W. McKay for having performed these calculations and for calling to our attention the errors in the published formulas.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*

$$\varphi \propto \left(\frac{p+q}{1-ipq}\right)^5\left[(m_0+in_0)+(e_0^2+g_0^2)\left(\frac{p+q}{1+ipq}\right)\right], \quad (3.14b)$$

again in the notation of Ref. 16.

The above analysis is also valid in a complex V_4, where $f_{\dot{A}\dot{B}}^*$ is not necessarily the complex conjugate of f_{AB}; notice that in no moment have we assumed anything about $f_{\dot{A}\dot{B}}^*$, except that it satisfies Maxwell equations.

ACKNOWLEDGMENTS

The interest of Professor F.J. Ernst in this work is gratefully acknowledged.

*On leave of absence from the University of Warsaw, Warsaw, Poland.

[1]B. Carter, Phys. Rev. 174, 1559 (1968); also, see I. Hauser and R.J. Malhiot, J. Math. Phys. 15, 816 (1974); 17, 1306 (1976) for a detailed study of rank two Killing tensors.
[2]M. Walker and R. Penrose, Commun. Math. Phys. 18, 265 (1970).
[3]L.P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. 27, 303 (1972).
[4]S. Hacyan and J.F. Plebański, J. Math. Phys. 17, 2203 (1976).
[5]E.T. Newman in Tel Aviv GR7 Lecture (1974).
[6]R. Penrose, "The Nonlinear Graviton," First Award Gravity Research Foundation (1975).
[7]J.F. Plebański, J. Math. Phys. 16, 2395 (1975).
[8]J.F. Plebański and I. Robinson, Phys. Rev. Lett. 37, 493 (1976); A. García, J.F. Plebański, and I. Robinson, "Null Strings and Complex Einstein-Maxwell Fields with λ," Nuovo Cimento (to be published); J.D. Finley, III and J.F. Plebański, J. Math. Phys. 17, 2207 (1976).
[9]F.J. Ernst and J.F. Plebański, "Killing Structures and Complex $\mathcal{E}$-Potentials," Ann. Phys. (to be published).
[10]F.J. Ernst, Phys. Rev. 168, 1415 (1968).
[11]F.J. Ernst, J. Math. Phys. 15, 1409 (1974).
[12]W. Kinnersley, J. Math. Phys. 14, 651 (1973); also, see F.P. Esposito and L. Witten, preprint, University of Cincinnati.
[13]A. Tomimatsu and H. Sato, Progr. Theor. Phys. 50, 95 (1973); also, see M. Yamazaki, preprint, Kanasawa University.
[14]Equation (3.10b) of Ref. 4, which corresponds to Eq. (1.2b) of this paper, is valid for Hermitian $K_{A\dot{A}}$ only (i.e., a real vector K_α); this was not mentioned explicitly in Ref. 4.
[15]P. Sommers, thesis, The University of Texas at Austin (1973).
[16]J.F. Plebański, Ann. New York Acad. Sci. 262, 246 (1975).

ERRATA

Erratum: Higher indices of group representations [J. Math. Phys. 17, 1972 (1976)]

J. Patera, R. T. Sharp, and P. Winternitz

Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada

(Received 29 March 1977)

In formula (51) please insert $]^{-1}$ after the factor $(2n+3)$; in (52) please replace $|\epsilon_i^{j-1}|$ by $|\epsilon_i^{n-j}|$. Insert [after N_λ in (62) and] after the entire formula; the denominator of the last term in (62) should be $n(2n-1)^2$. The factor $1/n$ in the second line of (72) should be $1/n^2$.

All numerical calculations in the article were performed using correct formulas and are valid. We thank W. McKay for having performed these calculations and for calling to our attention the errors in the published formulas.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*